Intégration

Poèmes mathématiques
de Fourier, Lebesgue et Banach

Cours d'Université enseigné à Lyon et à Rennes

Version provisoire du 1er janvier 2026

Chapitres 1 à 8 (nouvelle numérotation)

(17 chapitres prévus au total)

Cédric Villani

Université Claude Bernard Lyon 1 & Académie des sciences

À la mémoire de Haïm Brezis
Maître et ami
Et contradicteur bienveillant

Table des matières

Avant-Propos

Entre 2003 et 2008 j'ai enseigné à l'École normale supérieure de Lyon un cours de licence intitulé “Intégration et Analyse de Fourier” ; puis à partir de 2025 à l'Unversité Rennes I un cours intitulé “Intégrale de Lebesgue”. Le second tour de piste a été l'occasion, avec une vingtaine d'années de recul, de reprendre et améliorer le jeu de notes incomplètes issues du premier tour.

Sans chercher à être un ouvrage de référence, l'ouvrage constitue cependant une synthèse ambitieuse puisqu'il entremêle trois courants de pensée scientifiques majeurs, résonnant bien au-delà de la pensée mathématique : le calcul intégral, l'analyse harmonique et l'analyse fonctionnelle. L'analyse harmonique, ou analyse en fréquences, quantifie la régularité au moyen du calcul intégral, et l'analyse fonctionnelle est née avec les espaces d'intégration avant de se développer dans l'analyse de la régularité : ce sont donc des domaines fortement interconnectés, au regard de l'histoire comme du présent.

Comme toujours en sciences, ces théories ont été des œuvres collectives, influencées par les autres développements scientifiques et même philosophiques, pleines d'aller-retour complexes ; mais trois mathématiciens en particulier en sont devenus les figures emblématiques : Joseph Fourier (1768-1830) a fondé l'analyse en fréquences, Henri Lebesgue (1875-1941) a établi la théorie moderne de l'intégration, Stefan Banach (1892-1945) incarne le nouveau vent d'analyse fonctionnelle abstraite qui a soufflé de l'Europe de l'est au mitan du vingtième siècle. Nés dans des milieux modestes, orphelins de père, ou de mère, ou des deux, ces trois là ne sont pas partis gâtés par le sort, et pourtant leurs noms résonnent aujourd'hui avec la plus grande force dans l'océan agité des idées et des techniques. Le traité de Fourier sur l'équation de la chaleur (1811), la note de Lebesgue sur l'intégration (1901), l'ouvrage de Banach sur les opérations linéaires (1932) sont des jalons majeurs dans l'histoire des sciences. Et n'importe où dans le monde, quand on parle de “mesure de Lebesgue”, “analyse de Fourier” ou “espaces de Banach” cela évoque des champs gigantesques, pleins de glorieux accomplissements – et, pour les deux derniers, de problèmes ouverts. La puissance visionnaire de ces pionniers justifie que j'évoque leurs “Poèmes mathématiques” – une expression employée par Lord Kelvin, le plus puissant physicien de son temps, pour qualifier l'œuvre de Fourier.

Les théories de Lebesgue, Fourier et Banach ne sont pas seulement faites de définitions et théorèmes, ce sont aussi des méthodes et des points de vue, avec lesquelles toute analyste doit se familiariser. Tous les problèmes mathématiques que j'ai rencontrés dans ma vie de chercheur faisaient intervenir une combinaison de ces trois visions. Pour les présenter dans cet ouvrage, j'ai longuement remanié les présentations et les preuves. Un premier défi était de rassembler de façon cohérente les bases de trois théories souvent dispersées, et dont j'ai moi-même appris les fondamentaux dans les excellents ouvrages de Gramain, Brezis, Rudin et Körner. Le second défi était d'aller suffisamment loin pour couvrir tous les outils, parfois subtils, utiles à la pratique mathématique courante, et dans le bon niveau de généarlité, sans sacrifier à la pédagogie. Enfin, par cohérence avec mes propres choix, je souhaitais que le tout fût assemblé sans recours à l'axiome du choix – ne serait-ce que parce que cet axiome, du moins dans sa version générale, est inutile pour développer ces sujets, et plus généralement à l'essentiel de l'analyse.

La bibliographie est volontairement réduite à un petit nombre d'ouvrages et articles : ceux que j'ai moi-même utilisés régulièrement au cours de ma carrière. Quelques points de repère historiques seront fournis à différents endroits du traité.

Je remercie ceux qui m'ont aidé, au long des années, à la rédaction de ces notes par leurs commentaires, rectifications et aides ponctuelles, en particulier Luigi Ambrosio, Guillaume Aubrun, Haïm Brézis, Nassif Ghoussoub, Étienne Ghys, Baptiste Huguet, François Japiot, Sébastien Martineau, Julien Melleray, Quentin Mérigot, Forte Shinko, Jean-Claude Sikorav.

Choix de présentation

Ce bref chapitre est destiné aux lectrices déjà familières avec la matière enseignée et qui souhaiteraient savoir les choix pédagogiques de ce cours. Il peut être omis sans conséquence.

Touit au long de l'ouvrage, la lectrice est encouragée à se méfier de l'axiome du choix, et à éviter son usage. Aucun des théorèmes du cours ne l'utilise. *Toute l'analyse classique peut se construire sans la forme forte de l'axiome du choix.* Un bref chapitre préliminaire est consacré à une mise au point sur ce sujet.

Il existe deux grands cadres pour développer la théorie de la mesure : celui des espaces polonais (métriques séparables complets) et celui des espaces localement compacts. Le premier choix est le cadre pertinent pour la théorie des probabilités [Billingsley, Dudley, Parthasarathy], le second est privilégié par les admirateurs des espaces topologiques [Bourbaki, Halmos, Rudin]. Sans surprise, j'ai choisi le point de vue polonais, beaucoup plus important pour les applications, tout en gardant pourtant une place pour quelques énoncés emblématiques dans le cadre localement compact non métrique. Les démonstrations seront complètes dans le cas métrique, et seulement esquissées dans le cas non métrique (par exemple le théorème de Tychonov et le lemme d'Urysohn non métriques seront admis).

La régularité des mesures est abordée dès le premier chapitre, de même que diverses propriétés reliant topologie et théorie de la mesure. 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 être utilisable 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. En effet, les démonstrations classiques du Théor ème de Riesz, soit reprennent en fait des arguments du Théorème de Carathéodory, soit s'appuient explicitement sur la forme classique de ce dernier mais doivent alors y ajouter de délicats ingrédients supplémentaires ; la présentation adoptée ici évite cet écueil.

La linéarité de l'intégrale de Lebesgue est d'ordinaire établie comme corollaire du Théorème de convergence monotone ; cette approche est économe, 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 établ 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 duThéorème de convergence monotone est reprise par la suite, dans un nouveau chapitre 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. En fait, comme rappelé dans ces notes, on pourrait choisir le théorème d'Egorov comme point de départ de la théorie des passages à la limite ; mais il est plus naturel de réserver ce rôle au théorème de convergence monotone, pour son analogie formelle avec la propriété d'additivité dénombrable.

Les liens entre théorie de la mesure et logique axiomatique d'une part, théorie de la mesure et théorie des probabilités d'autre part, sont esquissés. Cela dit, je présente et démontre les principaux résultats techniques qui sont utiles en probabilités, y compris les subtils théorèmes d'existence de Kolmogorov et de Ionescu Tulcea, ou la loi du 0-1 de Hewitt et Savage (et celles de Kolmogorov et de Borel).

Je ne recommande pas la complétion de Lebesgue, qui n'apporte de justifications qu'à la marge (sauf peut-être dans la théorie des processus stochastiques), et au prix d'une augmentation considérable de la tribu. Ne pas compléter demande un peu d'attention aux négligeables (qui ne sont plus des ensembles de mesure nulle, mais des *parties incluses dans* des ensembles de mesure nulle). L'opération de complétion est donc présentée mais non recommandée en général. Ce choix est similaire à celui de Carlen et Loss dans leur traité d'analyse.

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, aussi parce que la construction est facilitée par le concept de fonction mesurable. Pour compenser cette faiblesse de plan, j'ai annoncé dès le début les rudiments sur les tribus produits, et le résultat d'existence de mesure produit dès après le théorème de Carathéodory, avec dans un cas particulier une démonstration qui annonce également le théorème d'existence de Kolmogorov.

Après ce passage en revue des résultats majeurs, un chapitre d'approfondissement, plus avancé, est consacré à une introduction à la théorie descriptive des ensembles ; vifs remerciements à Julien Melleray et à Forte Shinko pour m'avoir patiemment permis de me retrouver dans le jardin foisonnant que constitue ce sujet. Le chapitre se termine par un exposé sur la sélection mesurable, incluant la preuve des théorèmes de sélection dans les ensembles à coupes dénombrables, à coupes ouvertes et à coupes compactes. Pour la première fois en version cours, un argument relativement élémentaire est présenté pour le théorème de Lusin–Novikov de sélection mesurable dans les sections dénombrables, suivant un travail très récent de Shinko.

Ensuite un chapitre spécifique est consacré à la mesure de Lebesgue dans l'espace euclidien. La théorie de Riemann y est rappelée et les liens avec celle de Lebesgue sont explorés. L'intégrale de Riemann ne doit pas être sous-estimée, c'est le plus souvent elle qui permet d'effectuer les calculs pratiques. Ce chapitre accueille aussi une discussion assez précise sur la mesurabilité et la non-mesurabilité, en rapport avec les paradoxes de Banach–Tarski et l'axiome du choix.

Les mesures de Hausdorff sont absentes de la plupart des traités introductifs (à l'exception notable de [Billingsley]). Mais l'importance de ce concept dans de nombreuses branches des sciences, et la popularité du concept de mesure fractale ou de dimension fractale, motivent amplement leur étude dès ce niveau.

L'analyse fonctionnelle est développée dans les espaces de Banach, avec des hypothèses systématiques de séparabilité ou de dénombrabilité. Ainsi, le théorème de Hahn–Banach n'est démontré que dans le cadre des espaces vectoriels normés séparables – un cadre "restreint" mais suffisant pour toutes les applications que j'en connais, et encore une fois cela évite le recours à l'axiome du choix dans sa version forte. On pourra voir dans ce choix une régression assumée : par endroits, la présentation s'apparente plus au traitement originel de Banach, dans les années 1930, qu'à la théorie des espaces vectoriels topologiques développée dans les années 1950 et 1960.

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. Je propose au passage une nouvelle démonstration très simple des inégalités de Hanner, par application directe de l'inégalité de Jensen.

Dans le chapitre d'analyse fonctionnelle, j'ai également inclus et démontré des énoncés de compacité dont je sais par expérience que la démonstration est difficile à trouver dans les ouvrages de référence, malgré leur importance. Outre le classique Théorème de compacité de Prokhorov dans l'espace des mesures, j'ai donc inclus les théorèmes de compacité de Dunford–Pettis et de Schur, qui ont trait à la compacité faible L^1. Le critère de compacité pour des mesures signées (une question pourtant naturelle) semble n'être traité que dans les ouvrages pointus de Bogachev (merci à Luigi Ambrosio pour cette référence) ; j'en ai extrait une version simplifiée qui couvre le cadre des espaces métriques localement compacts.

L'analyse hilbertienne est d'abord traitée comme un cas particulier de l'analyse de Banach, avant de faire l'objet d'un chapitre d'approfondissement spécifique.

L'analyse harmonique est introduite en même temps que l'analyse par convolution, et dans le cadre des groupes localement compacts, métrisables et σ-compacts. C'est dans ce contexte que sont énoncés les grands résultats comme le Théorème de Pontryagin ou celui de Haar. Cet entre-deux, plus général que celui des espaces modèles et plus restreint que celui des groupes topologiques localement compacts, m'a semblé le bon dosage pour couvrir les problèmes que l'on rencontre d'ordinaire en analyse.

L'analyse de Fourier dans les espaces modèles, avec son cortège de points de vue et ses identités merveilleuses, fait ensuite l'objet d'un chapitre spécifique d'approfondissement.

L'approximation est l'un des piliers de l'analyse, et toutes les techniques de ce cours (Lebesgue, Fourier, Banach, Hilbert) viennent avec leur point de vue en la matière. Un chapitre est dédié à cette question fondamentale, sans aucune prétention d'exhaustivité.

La fonction maximale de Hardy–Littlewood n'est pas toujours introduite dans les cours de théorie de la mesure (mais elle est bien traitée dans celui de Rudin), et j'ai souhaité lui donner une place encore plus visible dans un chapitre spécifique qui prépare la discussion de la désintégration.

Le chapitre sur la désintégration referme le mouvement ouvert par l'intégration. Il traite aussi bien du Théorème de Radon–Nikodym (comme il est d'usage) que du problème plus général de désintégration de la mesure (beaucoup moins traité et d'ordinaire réservé aux ouvrages de probabilité avancée). J'ai choisi des conventions faisant apparaître Radon–Nikodym comme un cas particulier de cette désintégration plus générale. Ce chapitre continue avec des énoncés de reconstruction de la densité de Radon–Nikodym dans des espaces localement compacts ; on y trouve aussi bien le théorème classique de densité de Lebesgue (reconstruction ponctuelle), que des énoncés légèrement moins précis mais plus simples de reconstruction L^1. Les probabilités conditionnelles y trouvent aussi leur place.

La théorie de Brunn–Minkowski, et les inégalités géométriques, au confluent de l'analyse et de la géométrie, sont d'habitude réservés aux ouvrages spécialisés ; mais il s'agit de matériaux importants, en ligne directe avec le problème historique de l'isopérimétrie, dont tout le monde devrait avoir des notions, ne serait-ce que pour la culture générale ; cela justifiait aussi un chapitre spécifique.

Le dernier chapitre est un chapitre d'ouverture : moins axé sur la rigueur et la généralité, il présente des théories, problèmes et applications issues des analyses de Lebesgue, Fourier et Banach. Il pourra être consulté pour la culture scientifique comme pour les techniques qui y sont exposées.

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 $x \notin A$
proj : la projection ; si $(x, y) \in X \times Y$, alors $\mathrm{proj}_X(x, y) = x$.
Si P est une propriété dépendant d'une variable x, $\{P\}$ pourra désigner $\{x;\ P(x)\}$ (par exemple $\{f = 0\}$ désignera le lieu d'annulation de f).

Ensembles :
$\mathbb{N}$: l'ensemble des nombres entiers naturels *non nuls* : $\mathbb{N} = \{1, 2, 3, \ldots\}$
$\mathbb{N}_0$: l'ensemble des nombres entiers naturels ou nuls : $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$
$\mathbb{Z}$: 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
$\mathrm{Int}(A)$: 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. $\inf\{d(x, y);\ y \in A\}$
$\mathrm{osc}_x(f)$: oscillation de f en x, i.e. la limite du diamètre de $f(B_r(x))$ quand $r \to 0$

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

$C(A_1, \ldots, A_N)$: cylindre de base $A_1, \ldots, A_N$
Si μ est une mesure 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)
(Je ne distingue pas typographiquement entre l'espace des fonctions définies partout, et celui des fonctions définies presque partout)
$\|f\|_\infty$: supremum essentiel de $|f|$

Calcul différentiel :
∇T : matrice Jacobienne de $T : \mathbb{R}^n \to \mathbb{R}^m$ (gradient de T si $m = 1$)
$\nabla^2 f$: matrice Hessienne de f
Δf : Laplacien de f (trace de $\nabla^2 f$)

“la lectrice” = “le lecteur ou la lectrice” (même convention que Körner)

Je désignerai les 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 dans la bibliographie en fin d'ouvrage.

Mise au point axiomatique

L'analyse classique, comme l'essentiel de l'édifice mathématique, repose sur les axiomes de Ernst Zermelo et Abraham Adolf Fraenkel, soit le système de Zermelo–Fraenkel (ZF), qui permet de définir et manipuler des ensembles, de les comparer et de les mettre en correspondance avec des propriétés, de construire les entiers, et partant les rationnels et les réels, et tout ce qui s'ensuit. Personne n'a jamais réussi à démontrer que ZF est exempt de contradiction, mais c'est un acte de foi dont s'accommodent toutes les sciences mathématiques à ce jour (même si quelques non-croyants existent, comme le prouve le fait que le grand théoricien Edward Nelson a sérieusement cru trouver une absurdité dans ces axiomes, avant de se rétracter).

Il y a cependant débat pour un autre axiome, l'axiome du choix, ajouté à ZF pour former l'axiomatique ZFC. Cet axiome dit que *Pour toute famille d'ensembles non vides, il existe une fonction de choix, c'est à dire une façon de choisir un élément dans chacun des membres de la famille.* Autrement dit, si $(A_x)_{x\in\mathcal{X}}$ est une famille d'ensembles, tous non vides, indexée par l'ensemble $\mathcal{X}$, alors il existe une application f qui à tout $x \in \mathcal{X}$ associe $f(x) \in A_x$. Autrement dit encore : Si tous les A_x sont non vides, alors leur produit est non vide.

On peut se représenter les A_x de façon imagée comme des arbres, chacun pousse au point x et chacun a au moins une feuille, on veut cueillir une collection mathématique de feuilles, exactement une pour chaque arbre.

FIGURE

Ainsi présenté, l'axiome semble naturel. Il pose pourtant de nombreux problèmes. D'abord, il implique le principe de bon ordre, et lui est même équivalent : *Tout ensemble peut être muni d'un bon ordre*, c'est à dire un ordre total et strict dans lequel toute partie non vide admet un élément minimal (comme dans $\mathbb{N}$ où toute partie non vide admet un plus petit élément) ; l'idée que l'on puisse ordonner ainsi tous les éléments de $\mathbb{R}$ défie l'imagination. L'axiome du choix est également équivalent au plus obscur Lemme de Zorn : *Tout ensemble partiellement ordonné, dans lequel toute chaîne totalement ordonnée admet au moins un majorant, possède au moins un élément maximal.* Selon une célèbre plaisanterie du mathématicien Jerry Bona, "l'axiome du choix est évidemment vrai, le principe de bon ordre évidemment faux, et qui peut dire ce qu'il en est du lemme de Zorn ?" (évidemment une blague puisque les trois énoncés sont équivalents)

Ensuite, l'axiome du choix est "hautement non constructif", car il n'y a aucune indication, en pratique, de la recette que l'on pourrait suivre pour construire la fonction f. Par définition, dire qu'un ensemble est non vide, c'est qu'il contient au moins un élément et alors on peut le choisir ; face à une collection finie d'ensembles non vides, on applique ce raisonnement pour chacun des ensembles considérés. Mais si la collection est infinie, on ne peut réitérer le raisonnement une infinité de fois, on a besoin d'un axiome supplémentaire pour effectuer cette infinité de choix, et plus l'ensemble $\mathcal{X}$ est de grande cardinalité, plus l'acte de foi est exigeant.

Enfin l'axiome du choix implique des paradoxes choquants, comme on le discutera avec Banach–Tarski ; le petit ouvrage de Wagon présente très bien ces questions.

Mais si l'on s'interdit tout axiome du choix pour des ensembles infinis, on se heurte rapidement á d'autres paradoxes choquants ; et bien des raisonnements par récurrence, pourtant naturels, cessent d'être valables. On est ainsi amené à demander au minimum qu'un produit dénombrable d'ensembles non vides soit non vide (axiome du choix dénombrable). Mais cela n'est pas encore tout à fait suffisant pour mettre en forme les raisonnements habituels par récurrence : la bonne hypothèse, introduite par le mathématicien allemand-suisse Paul Bernays en 1942, est l'**axiome du choix dépendant** : Si X est un ensemble non vide, et pour tout $x \in X$ on se donne une partie non vide $F(x)$ de X, alors il existe une suite $(x_n)_{n\in\mathbb{N}} \in X^{\mathbb{N}}$ telle que pour tout n, $x_{n+1} \in F(x_n)$.

Voici une reformulation commode. Soit $(A_n)_{n\in\mathbb{N}}$ une famille dénombrable d'ensembles non vides ; pour tout n et tous $x_1, \ldots, x_{n-1}$ dans $A_1 \times \ldots \times A_{n-1}$ on se donne une partie $F_n(x_1, x_2, \ldots, x_{n-1}) \subset A_n$. On suppose que $x_i \in F_i(x_1, \ldots, x_{i-1})$ pour tout $i \leq n$ implique $F_n(x_1, \ldots x_{n-1})$ non vide. Alors on peut trouver une fonction f définie sur $\mathbb{N}$, telle que $f(n) \in F_n(f(1), f(2), \ldots, f(n-1))$ pour tout $n \in \mathbb{N}$. Autrement dit : Si à chaque étape on peut choisir un élément de A_n en fonction des choix précédents, alors il existe une suite faite de tels choix successifs.

Tout en évitant les paradoxes, la théorie résultante, ZF+CD (Zermelo–Fraenkel avec Choix Dépendant) permet de construire *toute l'analyse réelle classique*, y compris la théorie de la mesure, la théorie des équations aux dérivées partielles, la théorie des probabilités, le calcul des variations, la géométrie différentielle, l'analyse non lisse moderne, l'analyse fonctionnelle dans les espaces de Banach séparables, etc. C'est le cadre que j'ai moi-même adopté à partir du milieu des années 2000 pour tous mes articles et ouvrages, et que je recommande vivement. Qu'on aime ou pas les raisonnements non constructifs, *le simple fait que l'on puisse se passer de l'axiome du choix incite à ne pas l'utiliser.*

Et donc, **cet ouvrage n'utilise pas l'axiome du choix, mais seulement l'axiome du choix dépendant**.

Comme la plupart des ouvrages de référence utilisent l'axiome du choix complet (ZFC), certains de leurs énoncés deviennent indémontrables dans ZF+CD ; mais comme on le verra dans le présent cours, cela ne présente aucune gravité et n'entrave en rien nos objectifs et applications.

Pour la lectrice, ne pas utiliser l'axiome du choix rend certains énoncés un peu plus exigeants, mais ne complique en rien les démonstrations, au contraire. Utiliser ZF+CD veut dire que l'on s'autorise tous les raisonnements intuitifs faisant intervenir des suites avec choix successifs ; en pratique on n'y pensera même pas – alors que l'usage intelligent du lemme de Zorn est subtil, voire obscur.

D'autres axiomes sont parfois ajoutés à ZF : axiomes de grand cardinal en particulier. Ils n'offrent pas d'intérêt particulier pour le présent cours, mais seront parfois cités pour expliquer tel ou tel contre-exemple.

CHAPITRE I

Introduction : Aperçu historique et motivation

Commençons par une présentation à grands traits de la théorie de l'intégration à travers les âges, et ses plus célèbres contributeurs. J'insisterai plus particulièrement sur quelques concepts et destins.

I-1. D'Archimède à Lebesgue

Dans l'histoire des sciences et des concepts, l'intégration précède de beaucoup la dérivation. La quantité de corde nécessaire à une clôture, de petits pavés pour quarrer une mosaïque, de métal pour fondre une pièce de ferronnerie, demandaient des calculs additifs sur des longueurs, aires et volumes, en lien avec les mesures physiques. Les longueurs et les aires sont déjà présentes dans les très anciens théorèmes de l'Égypte et de la Mésopotamie anciennes : théorème de Pythagore, nombre π, formules d'Héron d'Alexandrie, calculs de volumes de polyèdres... Les Grecs anciens connaissaient aussi la solution du **problème isopérimétrique**, d'ailleurs présente dans la légende de Didon d'Énée : à périmètre donné, c'est le cercle qui a la plus grande surface (et dans un demi-espace, c'est le demi-cercle).

Cette tradition culmine avec Archimède, le grand génie des sciences de l'Antiquité. Il était si fier de l'un de ses calculs d'aire qu'il en a fait graver le croquis sur sa tombe : une sphère inscrite dans un cylindre, dont la hauteur est égale au diamètre, rappelant que l'aire de la sphère ($4\pi R^2$, si R est le rayon) est aussi la surface latérale du cylindre, ou encore le double de la surface des disques en haut et en bas. Ce même dessin a d'ailleurs été reproduit au revers de la médaille Fields !

C'est aussi à Archimède que l'on doit la plus ancienne évocation du très important problème de **passage à la limite** dans une suite de calculs intégraux. Il s'agit d'approcher π comme limite du périmètre de polygones réguliers quand le nombre de côtés augmente. Il en tire une valeur approchée de π à 0,01% près (via un polygone à 96 côtés) et la formule de l'aire du disque. Quelque temps plus tard, Zénodore s'appuie sur les raisonnements d'Archimède pour démontrer des énoncés de nature isopérimétrique : à nombre de côtés et périmètre fixés, ce sont les polygones réguliers qui ont la plus grande surface ; et à surface fixée, c'est la sphère qui enclôt le plus grand volume.

Les techniques de l'Antiquité sont reprises et amplifiées par les mathématiciens du monde arabo-perse, qui dominent la pensée scientifique du Moyen-Âge. Au 9e siècle, Thabit ibn Qurra (Thabet) propose une méthode de calcul d'aire proche de la sommation en tranches, avec épaisseur variables, qui sera la base de l'intégrale de Riemann.

À l'époque de la grande renaissance mathématique, au début du 17e siècle, les précurseurs du calcul intégral comme Bonaventura Cavalieri, Gilles Personne de Roberval, Evangelista Torricelli, Isaac Barrow, René Descartes, Pierre de Fermat, Blaise Pascal rivalisèrent d'ingéniosité pour mettre au point des méthodes générales et calculer les longueurs et aires de courbes notables. La cycloïde alors fascinait les scientifiques. Galilée détermina expérimentalement (au poids !) que l'aire sous le graphe en vaut $3\pi R^2$, ce que Roberval fut le premier à démontrer (R le rayon de la roue qui tourne pour engendrer la cycloïde). Christopher Wren prouva que la longueur du graphe est de $8\pi R$. Pascal en fit un défi sous le pseudonyme de Dettonville. La mode était aux calculs explicites et ingénieux.

C'est sous la plume de Pascal que l'on trouve la plus ancienne trace d'un signe intégral : un S allongé, S comme *somme* ou *summa*, notation qui fut reprise par Leibniz et connut une extraordinaire popularité. Moi-même, j'ai écrit ce symbole des centaines de milliers de fois dans ma carrière, et il s'est si bien incrusté dans mon cerveau que sa forme s'impose machinalement à mon esprit dès que je réfléchis à un raisonnement...

Les calculs d'aires, de longueurs et de volumes se développèrent au cours des siècles. Dès que le calcul différentiel fut établi, on identifia la recherche de primitives comme une méthode de calcul d'intégrale. L'intégration joue un rôle majeur dans la théorie de la gravitation universelle d'Isaac Newton. Elle apparaît à d'innombrables reprises dans les travaux de Leonhard Euler, y compris dans ses célèbres fonctions Beta et Gamma. L'heure est alors toujours à la recherche de relations et d'identités remarquables.

Un jour de 1800, le génie allemand Carl Friedrich Gauss calcule la moyenne arithmétique-géométrique de 1 et $\sqrt{2}$... avec 18 décimales ! Pour rappel, cette moyenne s'obtient en partant des deux nombres 1 et $\sqrt{2}$ à l'étape 1, et à chaque nouvelle étape on calcule les moyennes arithmétique et géométrique des deux nombres de l'étape précédente. Pressentant quelque chose d'intéressant dans ce nombre, Gauss joue avec... et se convainc *numériquement* que c'est exactement $2\pi/L$, où L est la longueur de la lemniscate de Bernoulli, courbe d'équation polaire $\rho = \sqrt{\cos(2\theta)}$. C'est une magnifique relation, qu'il mettra quelques mois à démontrer en la mettant en correspondance aussi avec les intégrales d'Euler. Encore et toujours des relations exactes pour des courbes particulières... Mais c'est fascinant de se dire que Gauss connaissait par cœur la longueur de la lemniscate !

Une grande mutation s'accomplit dans l'analyse tout au long du dix-neuvième siècle : le passage du particulier au général, du lisse au non-lisse. Cauchy, puis Weierstrass reprennent l'analyse sur ses bases ; Cantor d'eveloppera la théorie des ensembles et des cardinaux ; Poincaré initiera la théorie et la classification des équations différentielles générales. La fonction de Weiestrass (1872), continue mais nulle part différentiable, est emblématique de ce nouvel état d'esprit où les constructions reposent sur l'axiomatique plutôt que sur les formules. Dans ce vaste mouvement, deux contributions à la théorie de l'intégration sont particulièrement importantes : celles de Joseph Fourier (1811) et Bernhard Riemann (1854).

Né en 1768 à Auxerre, **Joseph Fourier** est le douzième enfant d'un père tailleur. Orphelin à neuf ans, il manifeste des talents prodigieux en mathématique, mais sa basse extraction l'empêche de les développer et il se destine à la religion. La Révolution française lui offre l'opportunité de reprendre les sciences, tout en menant aussi une carrière dans la politique et l'administration. Un temps emprisonné dans les convulsions de la Terreur, il échapppe de peu à la mort, enseigne dans des institutions d'enseignement supérieur, entre à l'Académie des sciences, participe activement à l'expédition d'Égypte, et devient sous Napoléon le second préfet de l'Isère, où il se révèle extraordinairement actif. Coïcidence ou pas, lui qui était légendairement frileux – plus couvert en plein été qu'un explorateur des régions polaires, nous dit un collègue dans son éloge funèbre – se passionna pour la modélisation de la chaleur, le grand œuvre de sa vie. Ce n'était pas le seul de ses centres d'intérêt puisqu'il fut également, avec un siècle d'avance, un pionnier de la programmation linéaire.

Les contributions de Fourier à la modélisation de la chaleur lui valent le titre de père de la physique mathématique [Dhombres]. On lui doit les premières discussions claires des échanges de chaleur par diffusion, convection et par effet de serre. Il établit l'équation de la chaleur par raisonnement phénoménologique, détermina ce que nous appelons maintenant les fonctions propres, et montra comment exprimer *n'importe quelle* fonction dans ce que nous nommons maintenant séries et intégrales de Fourier, grâce aux *formules intégrales*

$$\widehat{f)}(\xi) = \int e^{-2i\pi\xi\cdot x}\, dx.$$

C'est un acte fondateur pour l'étude des équations aux dérivées partielles linéaires, mais aussi le premier lien entre régularité et intégration, et les prémices de l'analyse fonctionnelle. Son mémoire, soumis en 1811 avec la devise *Et ignem regunt numeri* ("même le feu est régi par les nombres"), paru en 1822 après bien des déboires politico-universitaires, est un jalon important dans l'histoire des sciences. C'est la naissance de l'**analyse de Fourier** qui révolutionnera les sciences et les techniques, aujourd'hui l'un des piliers de toute l'industrie numérique.

Quant à **Bernhard Riemann**, c'est un fils de pasteur allemand, prodigieusement doué ; exactement comme son maître Gauss, il se destine à la théologie, avant d'être séduit par la beauté mathématique. Timide et fragile, peu doué pour les affaires humaines, il plane haut au-dessus de ses contemporains. Sa carrière ne dure qu'une dizaine d'années, mais l'ensemble des concepts nommés en son honneur donne le vertige : surfaces de Riemann, sphère de Riemann, géométrie riemannienne, courbure de Riemann, fonction zêta de Riemann, Hypothèse de Riemann, problème de Riemann... et bien sûr l'**intégrale de Riemann**, qui permet d'intégrer nombre de fonctions, régulières ou non. La mort de Riemann dans sa quarantième année contribue à en faire un personnage romantique inspirant, objet d'une forme de dévotion, au point que des intellectuels et artistes se recueillent à l'occasion sur sa tombe.

Après Riemann, l'intégration est solidement définie et prend toute sa part dans le formidable développement de l'analyse réelle et complexe, et bien sûr de l'analyse fréquentielle à la Fourier. Mais dès la fin du 19e siècle, les limitations de la théorie de Riemann sont apparentes et il devient nécessaire de la généraliser. Thomas Joannes Stieltjes, Camille Jordan, William Young font des propositions, mais c'est finalement un groupe de jeunes mathématiciens français, pasionné par la description des ensembles et des fonctions, qui décroche le gros lot : **René Baire, Émile Borel et Henri Lebesgue**. Et au sein de ce groupe, c'est Lebesgue qui aura l'honneur de donner à l'intégration sa forme moderne.

Né dans une modeste famille picarde touchée par le malheur et la maladie, fils d'un ouvrier typographe d'une mère institutrice, Henri Lebesgue n'a que trois ans quand la tuberculose emporte son père et ses deux sœurs, et le laisse invalide. Il manifeste de vrais dons pour les sciences et sa mère se démène pour qu'il puisse suivre des études dans les meilleurs établissements. À l'École normale supérieure, il retrouvera d'autres jeunes passionnés comme Baire et Borel, issus de milieux modestes mais accédant à la recherche internationale.

Une génération après le contre-exemple de Weierstrass, l'heure est aux fonctions irrégulières, un concept qui va dominer toute la science du 20e siècle. Dans quelques années, d'ailleurs, Louis Bachelier en fera la base de son étude mathématique des cours de la bourse, et Albert Einstein et Marian Smoluchowski feront le lien entre les trajectoires hautement chaotiques du mouvement brownien, et la théorie atomique de Ludwig Boltzmann. Aux yeux de la jeune garde, l'irrégulier est la règle et non l'exception. Pour protester contre la théorie traditionnelle, lisse, des surfaces développables (surfaces isométriques au plan euclidien), le jeune Lebesgue froisse son mouchoir et constate à haute voix que le résultat ne satisfait pas les conclusions des théorèmes du cours !

Bien plus exigeante sur la rigueur, cette nouvelle génération apprend à manier avec virtuosité les concepts de la théorie des ensembles et de la topologie naissante – ouverts, fermés, continuité, opérations dénombrables. Pour mesurer un ensemble, Baire inaugure la voie topologique (intersections d'ouverts et unions de fermés). Borel et Lebesgue, quant à eux, attaquent la voie métrique (recouvrement par des boules et des pavés), qui s'avère bien plus efficace. Leur démarche aboutit aux deux concepts fondamentaux sur lesquels est bâti l'édifice moderne : **mesure de Borel** (1895) et **intégrale de Lebesgue** (1901). La note fondatrice de Lebesgue aux Comptes rendus de l'Académie des sciences, développée dans le Cours Peccot, est un acte fondateur pour l'ensemble de la communauté scientifique qui se réappropriera cet outil.

À dire vrai, il est difficile de démêler les mérites respectifs de Borel et Lebesgue, d'autant que la coopération enthousiaste de leurs jeunes années laissera la place à d'amères querelles de priorité, envenimées par la grande déprime de la guerre, et par l'inégalité considérable de carrières. En effet, Borel a tous les honneurs de la société, allié à des personnalités influentes (Appell, Painlevé), et menant une carrière universitaire et politique qui le verra maire, député et ministre, et militant de la cause européenne ; évoluant au cœur de la vie intellectuelle de son temps, il dirige l'École normale supérieure, fonde l'Institut Henri Poincaré et aura une réelle influence en France sur la recherche industrielle, sur le développement des probabilités, des statistiques, des sciences de la décision – alors que Lebesgue, méprisant les mondanités, mène une carrière bien plus modeste et laborieuse. Baire pour sa part voit son parcours prématurément interrompu par la maladie et finira sa vie dans la détresse matérielle et psychologique.

La théorie de Borel et Lebesgue est restée remarquablement stable depuis : si l'on compare la note de Lebesgue de 1901 à un cours moderne d'intégration, on reconnaît les mêmes angles d'approche et les mêmes architectures de raisonnements. En effet, cette théorie a tenu toutes les promesses : elle a facilement accueilli dans son écrin trois siècles d'analyse des infiniment petits, de calcul des probabilités, d'analyse des variations – dont bien sûr l'analyse fréquentielle de Fourier.

C'est tout naturellement que la nouvelle théorie intègre aussi les travaux menés à la fin du 19e siècle sur les volumes et les surfaces des solides convexes, dans la veine du problème isopérimétrique, par les chercheurs du monde germanique : le Suisse Jakob Steiner (reconnu en son temps comme le plus universel des géomètres), l'Allemand Hermann Brunn et le Juif polonais-lithuanien-russe-allemand Hermann Minkowski (génie profond et original, camarade proche de David Hilbert, également père de la géométrie de l'espace-temps relativiste).

I-2. Le nouveau découpage en tranches

Le jeune Lebesgue était connu pour son style imagé et son humour, et la citation que voici l'illustre bien.

Je dois payer une certaine somme ; je fouille dans mes poches et j'en sors des pièces et des billets de différentes valeurs. Je les verse à mon créancier dans l'ordre où elles se présentent jusqu'à atteindre le total de ma dette. C'est l'intégrale de Riemann. Mais je peux opérer autrement. Ayant sorti tout mon argent, je réunis les billets de même valeur, les pièces semblables, et j'effectue le paiement en donnant ensemble les signes monétaires de même valeur. C'est mon intégrale.

Ainsi présentée, 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 très différente : Riemann en fait un découpage en tranches verticales, et Lebesgue en tranches horizontales. Pour Riemann ce sont les variations de la fonction sur son domaine de définition qui comptent, pour Lebesgue ce sont les variations des valeurs. Dans le cas de Riemann, la base (le côté horizontal du rectangle) est toute simple (un petit segment, disons) et la fonction définit la hauteur (le côté vertical) ; pour Lebesgue, au contraire la hauteur est toute simple et c'est la fonction qui définit la base ; en fait cette base peut être tourmentée, de sorte 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.

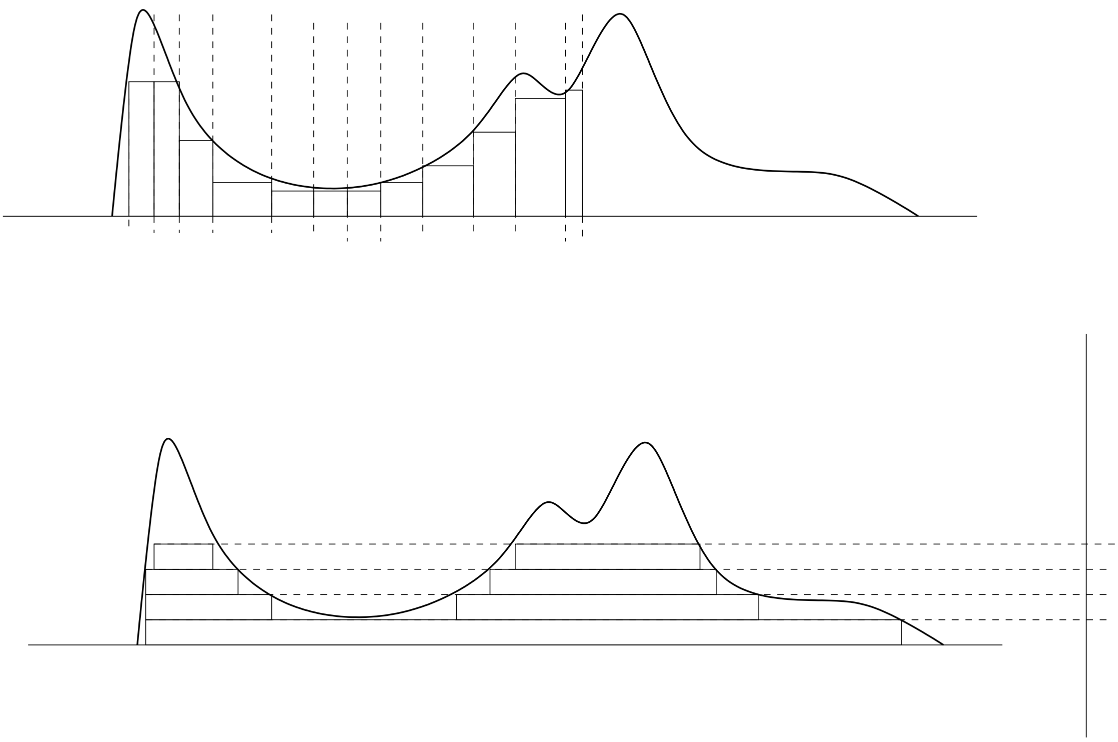

Figure 1. Procédé de Riemann vs. procédé de Lebesgue

I-3. Forces et faiblesses de l'intégrale de Lebesgue

Extrême généralité

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, mais Riemann lui-même a conscience que les conditions à imposer sur les fonctions sont encore relativement fortes. Il démontre en effet qu'une fonction $f : [a, b] \to \mathbb{R}$ 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]. L'idée de Lebesgue n'est pas très loin...

De fait, Lebesgue montre qu'une fonction $f : [a, b] \to \mathbb{R}$ 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. Plus généralement, un ensemble est mesurable au sens de Riemann (c'est à dire, sa fonction indicatrice est intégrable) si et seulement si sa frontière est de mesure nulle. Or dans de nombreux problèmes d'analyse, on rencontre des fonctions dont les discontinuités ne peuvent être négligées, et des ensembles dont la frontière est "étalée".

Mais dans la théorie de Lebesgue, la classe des fonctions intégrables est beaucoup plus grande. En fait, comme Robert Solovay le démontrera en 1970, il est impossible de construire (sauf à utiliser l'axiome du choix dans sa version complète) une fonction bornée non intégrable, ou un ensemble non mesurable. Cette grande généralité est l'un des plus grands avantages de la théorie de Lebesgue.

L'exemple de la fonction indicatrice de $\mathbb{Q} \cap [0, 1]$ 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.

De façon similaire, quand on s'intéresse à l'analyse de Fourier, le cadre de Borel et Lebesgue apporte une extraordinaire généralité : en fait, comme le montreront Lennart Carleson (1966) et Richard Hunt (1968), toute fonction p-sommable au sens de Lebesgue ($1 < p < \infty$) est la limite de sa série de Fourier en dehors d'un ensemble de mesure de Lebesgue nulle.

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

$$f(b) - f(a) \;=\; \int_a^b f'(t)\,dt \tag{1}$$

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.

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.

Au contraire, comme le comprend le mathématicien autrichien Johann 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 maniables. Le cas échéant, cet espace pourra même être un espace fonctionnel de dimension infinie. En 1921 un exemple spectaculaire et d'importance considérable est fourni par le mathématicien Norbert Wiener (américain d'origine juive polonaise, enfant prodige et père de la cybernétique, par ailleurs militant de la paix et de la cause animale). Wiener construit une mesure sur l'espace $C([0, 1], \mathbb{R}^d)$ des fonctions continues sur $[0, 1]$ à valeurs dans $\mathbb{R}^d$; bien sûr c'est un espace de dimension infinie, séparable mais pas localement compact. Cette **mesure de Wiener** est une mesure "gaussienne", d'importance capitale en probabilités et en physique où elle modélise le mouvement brownien.

Fluidité du passage à la limite

Peut-on échanger les opérations limite et intégrale ? C'est un problème classique, déjà en germe dans Archimède, et source de milliers d'exercices dans le cadre de l'intégrale de Riemann. Mais dans la théorie de Riemann on ne peut même pas 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

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

Cela ne dit pas si une limite de fonctions continues peut ne pas être Riemann-intégrable invalide, mais cela invalide, par l'absurde, l'hypothèse que la classe des fonctions Riemann-intégrable est fermée sous l'action de la limite.

Au contraire, l'intégrabilité au sens de Lebesgue est effectivement 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. Il suffit par exemple que ces fonctions soient définies sur un intervalle fixé et uniformément bornées. 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* $[a,b] \to [0,1]$*, convergeant ponctuellement (simplement) vers 0. Alors* $\int_a^b f_n \longrightarrow 0$.

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 ces théories alternatives ne se sont-elles pas imposées face à 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, que l'on peut traiter à la main. D'autre part, parce que ces théories alternatives sont moins souples que l'intégrale de Lebesgue, et plus exigeantes sur la topologie de l'espace de départ.

Difficuté de calcul

Par nature, les tranches sont beaucoup moins explicites dans la théorie de Lebesgue que dans celle de Riemann ; de sorte que c'est cette dernière théorie qui prévaut dans le domaine de l'analyse numérique, et en général des calculs explicites. Cela n'amène pas à abandonner l'intégrale de Lebesgue pour les applications, mais à travailler en va-et-vient entre les deux notions : on profite de la nature constructive de l'intégrale de Riemann, de la généralité et de la fluidité de l'intégrale de Lebesgue, et on passe de l'une à l'autre en utilisant leur compatibilité, c'est à dire le fait que l'intégrale de Lebesgue se ramène à l'intégrale de Riemann dans la classe des fonctions Riemann-intégrables.

I-4. Un grand édifice

Une fois les fondations établies, de nombreux mathématiciens perfectionnent et généralisent la théorie de Borel et Lebesgue au cours du vingtième siècle, en particulier (par ordre chronologique approximatif) Beppo Levi, Pierre Fatou, Guido Fubini, Leonida Tonelli, Dimitri Egorov, Mikhaïl Souslin, Constantin Carathéodory, Giuseppe Vitali, Nikolaï Nikolaïevitch Lusin, Johann Radon, Frigyes et Marcel Riesz, Ernst Sigismund Fischer, Felix Hausdorff, Andreï Nikolaïevich Kolmogorov, et Abram Samoïlovitch Besicovitch.

Mais ce sont aussi de nouveaux points de vue et théories qui émergent. 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.

Pour commencer, la généralité de l'intégrale de Lebesgue et sa stabilité par passage à la limite en font un cadre idéal pour formaliser les **espaces de fonctions**, des espaces géométriques dans lesquels chaque point est une fonction. C'est la naissance de l'**analyse fonctionnelle**, qui vise à soutenir l'étude des fonctions par des raisonnements de nature géométrique, analytique ou synthétique se tenant dans des espaces de fonctions. L'idée était déjà en germe dans Fourier ; mais ce sont des disciples de Lebesgue, l'**École mathématique de Lwów** (aujourd'hui Lviv en Ukraine), qui vont la mettre au point.

L'école de Lwów désigne un groupe de plusieurs dizaines de mathématiciens polonais travaillant dans l'entre-deux guerres à Lwów, également en collaboration étroite avec leurs compatriotes de Cracovie et Varsovie. L'acte fondateur du groupe est la rencontre fortuite, à l'université de Cracovie, entre l'enseignant Hugo Steinhaus et l'étudiant Stefan Banach, qui discutait avec son camarade Otto Nikodym précisément sur la mesure de Lebesgue. C'était le point de départ d'une série de recherches qui ont placé la Pologne au plus haut niveau de l'analyse fonctionnelle avant que son école soit écrasée par la Seconde Guerre mondiale, prise en étau entre les mâchoires allemande et russe.

Parmi eux, le plus brillant, reconnu par tous comme le chef de file, était le phénomène **Stefan Banach**. Né en 1892 à Cracovie dans une famille pauvre, abandonné par sa mère à sa naissance, il poursuit des études supérieures dans la Pologne tout juste indépendante et devient en 1919 l'un des fondateurs de la Société mathématique de Pologne. Gagnant rapidement ses galons de résolveur de problème, il devient au sein du groupe de Lwów le plus ardent animateur d'un joyeux séminaire qui se tient au Café écossais et où l'on enchaîne problèmes de mathématique et boissons. En 1924 il publie avec Alfred Tarski ce que l'on appelle désormais le paradoxe de Banach–Tarski ; en 1927 les théorèmes de Hahn–Banach (avec Hans Hahn) et Banach–Steinhaus (avec Hugo Steinhaus). Puis en 1932 qu'il publie son grand œuvre, *Théorie des opérations linéaires*, ouvrage majeur de l'histoire mathématique où l'analyse fonctionnelle est développée selon une forme toujours d'actualité. Dans le même temps, sous l'impulsion en particulier de Kazimierz Kuratowski, Alfred Tarski et Wacław Sierpiński, ce groupe développe la théorie de la mesure dans les espaces métriques séparables complets, avec une profondeur telle qu'on les appelle depuis "espaces polonais".

Dans la foulée, en 1933 le mathématicien russe Andreï Nikolaïevich Kolmogorov publie son ouvrage *Fondations de la théorie des probabilités* : il y montre que la théorie de Lebesgue est un parfait écrin pour refonder tout le calcul des probabilités, initié au 17e siècle par Pascal et Fermat, et perfectionné ensuite par les Daniel et Jacob Bernoulli, Abraham De Moivre, Pierre-Simon de Laplace, Henri Poincaré, etc. Le mathématicien polonais Stanislaw Ulam, également de l'École de Lwów (plus tard l'un des pères de la Bombe H) était arrivé aux mêmes conclusions que Kolmogorov ; et le terrain avait bien sûr été préparé par la découverte de la mesure de Wiener. Aujourd'hui encore, l'axiomatique probabiliste est restée la même que celle de Kolmogorov, basée sur la théorie de la mesure de Borel et Lebesgue.

Presque en même temps, le Français Jean Leray (en 1933, dans le cas particulier des équations de Navier–Stokes incompressibles) et le Russe Sergueï Lvovitch Sobolev (en 1935) utilisent la théorie de Lebesgue pour donner un sens aux dérivées partielles de fonctions non nécessairement différentiables. L'idée est d'exploiter la formule d'intégration par parties

$$\int \frac{\partial f}{\partial x_j}\,\varphi\,dx = -\int f\,\frac{\partial \varphi}{\partial x_j}\,dx$$

pour *définir* $\partial_j f$ comme fonctionnelle linéaire sur l'espace des fonctions φ continûment différentiables et à support compact. L'objectif est de construire ainsi, *via l'intégration*, des solutions généralisées d'équations aux dérivées partielles. Cette approche est très naturelle du point de vue physique (en témoigne le fait que James Clerk Maxwell l'avait utilisée à l'occasion en théorie cinétique des gaz dès les années 1860). Leray raconte que le vieillissant Lebesgue avait voulu le dissuader de se lancer dans des recherches aussi ardues... mais son article est entré dans l'histoire. Ces travaux annoncent la théorie des distributions de Laurent Schwartz, cadre conceptuel utilisé aujourd'hui presque universellement dans le domaine des équations aux dérivées partielles.

Simultanément à l'étude par Leray de Navier–Stokes incompressible, l'analyste suédois Torsten Carleman exploite toute la puissance de l'intégration de Lebesgue pour effectuer la première étude mathématique de l'équation de Boltzmann (en l'occurrence, pour un gaz spatialement homogène de sphères dures).

Il se sera donc écoulé exactement une génération entre la découverte de l'intégrale de Lebesgue, et la naissance de trois grandes branches qui s'appuient dessus : analyse fonctionnelle (Banach), théorie moderne des probabilités (Kolmogorov), solutions généralisées des équations de la physique mathématique (Leray, Sobolev, Carleman).

À cette liste on peut ajouter une quatrième branche qui s'est développée vers l'intérieur de la théorie ensembliste de Borel et Lebesgue, pour décrire la nature et la structure des ensembles mesurables, sous l'influence des écoles russe et polonaise : c'est la théorie descriptive des ensembles. Les initiateurs en sont deux mathématiciens russes influencés par l'école française de Darboux, Borel, Hadamard, Poincaré : il s'agit de Dimitri Egorov et son élève Nikolaï Nikolaevitch Luzin. (Tous deux seront persécutés pour leurs convictions religieuses, l'un mourra d'une grève de la faim, l'autre sera la cible d'un violent procès politico-scientifique qui aurait pu lui être fatal.) C'est en 1930 que Luzin publie en français le premier traité consacré à ce sujet : *Leçons sur les ensembles analytiques et leurs applications.*

Par-dessus la guerre est passée, marquant tous nos protagonistes. Banach s'accommode de l'occupation soviétique, mais après l'invasion nazie en est réduit à héberger des poux pour les recherches médicales... il meurt d'un cancer du poumon juste après la fin de la guerre. Kolmogorov (re)devient une véritable institution dans son pays, et l'un des mathématiciens les plus féconds du 20e siècle pour ses travaux en probabilité, mécanique, théorie de la complexité. Également en Russie, Sobolev est impliqué dans le programme atomique soviétique et travaille au développement universitaire de la Sibérie. Après avoir été fait prisonnier, Leray devient un pionnier de la toplogie algébrique et l'une des figures tutélaires de la mathématique française. Quant à Carleman, adhérent aux idées racistes propagées par les mouvements fascistes, il termine sa carrière dans l'isolement.

Après guerre commence une nouvelle période mathématique, sous l'effet des forces puissantes qui accélèrent le progrès scientifique : grands programmes de recherche nationaux, guerre froide, essor des grandes universités américaines, course à l'espace, essor de l'informatique, recherche industrielle, grands programmes d'équipement civils et militaires, etc. Toutes les branches mathématiques se développent à grande vitesse. C'est le cas en particulier pour celles qui sont directement liées à l'intégration : équations aux dérivées partielles, systèmes dynamiques, analyse numérique, analyse fonctionnelle, analyse harmonique, calcul des variations, théorie géométrique de la mesure, analyse convexe, analyse non lisse dans l'espace euclidien et dans les espaces métriques... À travers ces sujets en plein essor, l'intégration de Lebesgue devient ainsi l'un des piliers de tout l'édifice mathématique moderne.

Parmi ces directions variées, l'analyse harmonique a été significativement transformée avec la théorie des ondelettes, qui renouvelle la théorie de Fourier. Initiée par l'intuition révolutionnaire de l'ingénieur français Jean Morlet dans les années 80, cette théorie s'est développée en un corpus complet allant des applications les plus terre à terre (la motivation initiale était la prospective pétrolière) au cœur de l'analyse réelle. Deux "maîtres ondelettistes" en particulier ont engrangé quantité de récompenses internationales au 21e siècle : le français Yves Meyer et la belge Ingrid Daubechies (hélas elle n'est que la première femme citée dans cette histoire, dont la prédominance masculine reflète des biais de genre et stéréotypes qui ont imprégné nos sociétés depuis des siècles ! mais elle est une protagoniste majeure, et fut même présidente de l'Union mathématique internationale).

Impossible de boucler ce tour d'horizon sommaire sans mentionner que la logique également a été durablement influencée par l'analyse de Borel et Lebesgue. On a déjà mentionné les noms du logicien polonais Alfred Tarski (paradoxe de Banach–Tarski, 1924) ; après-guerre, ce sera l'Américain Richard Solovay (théorème de Solovay sur les ensembles mesurables, 1970). Ainsi la théorie de l'intégration s'est-elle retrouvée à jouer un rôle transversal et universel en mathématique, depuis les couches les plus théoriques jusqu'aux plus appliquées.

CHAPITRE II

Mesures

Ce chapitre est consacré aux mesures, selon le point de vue ensembliste de Borel : une mesure sur un ensemble X est une *fonction σ-additive d'ensembles*, définie sur une tribu $\mathcal{A}$ de parties de X, dites parties mesurables ; le triplet $(X, \mathcal{A}, \mu)$ est appelé *espace mesuré* (section II-1). Après leur définition abstraite, une liste d'exemples est présentée (section II-2.

Les mesures qui nous intéresseront le plus sont les mesures boréliennes, sur la tribu engendrée par la topologie ; pour aller plus loin nous aurons besoin de quelques rappels de topologie (section II-3).

Viennent alors les notions majeures qui constituent le cœur du chapitre : (a) sous des hypothèses minimales, les mesures boréliennes sont *régulières*, c'est à dire qu'on peut bien approcher les ensembles, au sens de la mesure, soit par des ouverts soit par des compacts (section II-4) ; (b) on peut définir le support d'une mesure et qualifier sa concentration ou sa diffusivité ; (c) de puissants théorèmes permettent d'*étendre* de façon unique des mesures définies seulement sur des parties simples (théorèmes d'extension de Carathéodory et de Kolmogorov, section II-6, la démonstration du théorème de Kolmogorov étant remise au Chapitre IV), (d) on peut *compléter* les espaces de mesure en ajoutant, si on le souhaite, les ensembles négligeables.

Cette théorie générale est appliquée à la droite réelle pour obtenir la mesure originelle de Lebesgue sur la droite réelle et sur l'espace euclidien (section **??**).

Je terminerai par l'étude de recouvrements par de petites boules (section II-8), une technique dont la pertinence sera démontrée bien plus tard, dans le Chapitre **??**.

Seule les sections II-1 et II-2 sont indispensables à la compréhension de la suite du cours, car elle présentent les concepts clés et les exemples majeurs.

II-1. Espaces mesurables et mesurés

En quête de définition

Les notions d'"intégrale" et de "volume" vont de pair. Si l'on sait intégrer des fonctions, alors on peut définir le volume d'un ensemble A comme l'intégrale de l'indicatrice de A. Et réciproquement, si l'on sait définir les volumes des ensembles, alors on peut définir l'intégrale d'une fonction en additionnant les volumes des "petites tranches superposées" sous le graphe de la fonction.

FIGURE de découpage en tranches

En d'autres termes, une bonne théorie d'intégration doit mener aux deux identités liées

$$\begin{cases} \mathrm{vol}(A) = \displaystyle\int 1_A \\ \\ \displaystyle\int f = \int_0^{+\infty} \mathrm{vol}(\{f \geq t\})\, dt \end{cases}$$

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

La théorie d'intégration de Lebesgue se développe sur le concept de **mesure**, introduit par Borel quelques années avant les travaux de Lebesgue pour quantifier les tailles des ensembles. En fonction du contexte, on peut penser à une mesure comme à un volume, ou une surface, ou une longueur, ou quelque chose d'autre. Dans tout ce chapitre, on se concentrera donc sur la mesure des ensembles, et ce n'est que dans le chapitre suivant qu'on considèrera l'autre versant, celui des fonctions.

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 une "masse" positive (finie ou infinie), notée $\mu[A]$ (ma convention) ou $\mu(A)$.

C'est bien le minimum d'imposer qu'une telle fonction 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 fondamentale implique les règles de calcul habituel des longueurs, des surfaces ou 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 la formule d'usage courant

$$\mu[A \cap B] < \infty \Longrightarrow \quad \mu[A \cup B] = \mu[A] + \mu[B] - \mu[A \cap B]. \tag{2}$$

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, ou plus généralement par isométrie.

Ce cahier des charges paraît raisonnable, et on aimerait le prendre pour base de notre étude. Le théorème suivant (qui découle des travaux de Banach, Tarski et Solovay) pourra donc apparaître comme un choc décourageant : *Il est impossible de démontrer l'existence d'une fonction d'ensembles* $\mu : \mathcal{P}(\mathbb{R}^3) \to [0, +\infty]$ *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. La mesure μ n'est pas définie sur les parties non mesurables, l'expression $\mu[A]$ n'a même pas de sens si A n'est pas mesurable.

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.

Définir les mesures sur des algèbres donne beaucoup moins de contraintes, donc beaucoup plus de flexibilité. Mais alors on en a même trop ; même pour des situations toutes naturelles, dans l'espace euclidien par exemple, on manque de bons résultats d'unicité, et il n'y a pas de consensus sur leur utilisation [Dudley, p 112]. En outre il est difficile, dans ce cadre, de démontrer les passages à la limite que l'on rencontre sans cesse dans les intégrations, et qui passionnaient déjà Archimède.

Le remède proposé par Borel et Lebesgue, aussi bien pour donner plus de rigidité à la théorie, que pour traiter plus facilement les passages à la limite, consiste à imposer *dès la définition* que ces mesures soient non seulement additives, mais aussi *dénombrablement additives*. Autrement dit, on requiert que la mesure d'une union dénombrable de parties disjointes (au sens fort, c'est à dire deux à deux disjointes) soit égale à la somme des mesures de toutes les parties.

Pourquoi s'arrêter en si bon chemin et ne pas imposer cette relation d'additivité pour une union *quelconque* de parties, pas forcément dénombrable ? En fait, une telle théorie serait tout simplement triviale. Par exemple, supposons que la mesure des singletons soit nulle ; comme un ensemble est l'union de ses éléments, tout ensemble serait alors de mesure nulle.

De façon remarquable, entre l'additivité pour les unions finies qui est trop lâche, et l'additivité pour les unions quelconques, qui est triviale, il y a une ligne de crête qui chemine à merveille, c'est l'additivité dénombrable. Pour lui donner sens, il va falloir renforcer la définition des algèbres, en imposant de plus la stabilité par les opérations ensemblistes dénombrables ; c'est ce que l'on appelle les *σ-algèbres*, ou *tribus*.

II-1.1. Algèbres.

Définition II-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) $A \in \mathcal{A} \Longrightarrow X \setminus A \in \mathcal{A}$ *;*
(iii) $A, B \in \mathcal{A} \Longrightarrow A \cup B \in \mathcal{A}$.

Remarque II-2. D'autres 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 II-3. Une algèbre est automatiquement stable par union finie, intersection finie, différence et différence symétrique. (Pour rappel, la différence de deux ensembles A et B est $A \setminus B$, tandis que leur différence symétrique est $(A \setminus B) \cup (B \setminus A)$.) En somme, une algèbre est un ensemble de parties dans lequel on peut effectuer toutes les opérations ensemblistes classiques.

Exemples II-4. (i) Algèbres triviales. Si X est un ensemble quelconque, alors on peut toujours définir la plus petite algèbre de parties de X, savoir $\{\emptyset, X\}$; et la plus grande, qui est tout simplement $\mathcal{P}(X)$. Ni l'une, ni l'autre ne sont fort intéressantes, on les appelle souvent triviales.

(ii) Algèbre engendrée par une partie. Si $A \subset X$, la plus petite algèbre contenant A est $\{\emptyset, X, A, X \setminus A\}$.

(iii) 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. C'est bien sûr une généralisation de (ii), qui correspond à une partition en 2 éléments. On peut montrer que toute algèbre finie est associée à une partition : pour cela, on identifie les éléments de Π comme les éléments minimaux de $\mathcal{A}$.

(iv) 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 $\mathcal{F}$). Soit X un ensemble quelconque, et soit $\mathcal{F}$ une famille de parties de X, qui (a) est stable par intersection, (b) contient X et (c) 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 (exercice).

(v) Algèbre engendrée par les **pavés**. C'est un cas particulier de (iv). 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

$$\begin{aligned}(X_1 \times X_2) \setminus (A_1 \times A_2) = &(X_1 \setminus A_1) \times (X_2 \setminus A_2) \\ &\cup\ (X_1 \setminus A_1) \times A_2 \qquad \cup A_1 \times (X_2 \setminus A_2).\end{aligned}$$

D'après (iv), on sait alors que l'ensemble des unions finies de pavés forme une algèbre de parties de X : voir l'illustration sur la figure 1, dans l'espace produit $X = \mathbb{R}^2$.

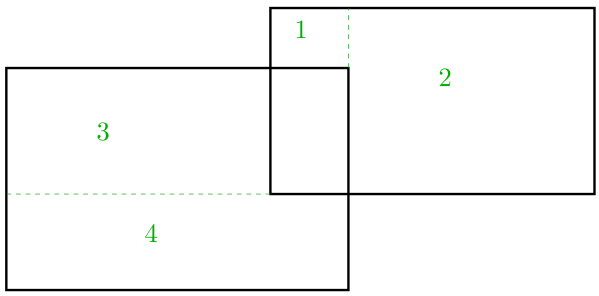

FIGURE 1. La différence symétrique de deux pavés est une union finie de pavés

(vi) Algèbre engendrée par les **cylindres**. C'est encore un cas particulier de (iv), et une généralisation de (v), 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 d sa propre algèbre. Je vais commencer par considérer le cas simple d'un produit dénombrable d'ensembles finis X_k munis

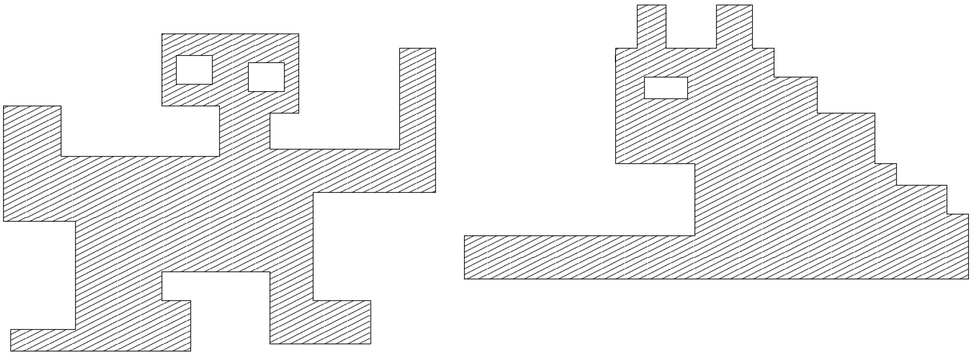

FIGURE 2. Deux membres de l'algèbre engendrée par les pavés dans $\mathbb{R} \times \mathbb{R}$

de leur algèbre $\mathcal{P}(X_k)$. (Le choix "trivial" où chaque X_k est l'espace $\{0, 1\}$ est déjà hautement 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 dépend que d'un nombre fini de coordonnées" : pour tout K et tout choix de $(a_1, \ldots, a_K) \in X_1 \times \ldots \times X_K$, on posera donc

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

On vérifie facilement (exercice) 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 (iv), 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, et où les X_k ne sont pas forcément finis. 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 $(t_1, \ldots, t_K)$ dans T^K, pour tout $j \in \{1, \ldots, K\}$ on choisit A_{t_j} dans $\mathcal{A}_{t_j}$ et on définit le cylindre élémentaire $C_{(t_1,\ldots,t_K)}(A_{t_1}, \ldots, A_{t_K})$ par la formule

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

Ici la base du cylindre C dans les variables $(t_1, \ldots, t_K)$ est $A_{t_1} \times \ldots \times A_{t_K}$. L'ensemble des unions finies de cylindres élémentaires est alors une algèbre.

EXERCICE II-5. Soit $X = \{a, b, c\}$ un ensemble à 3 éléments. Déterminer toutes les algèbres de parties de X.

II-1.2. Sigma-algèbres. Les algèbres sont stables par toutes les opérations ensemblistes classiques (union, intersection, différence) appliquées à des familles finies de parties ; pour les σ-algèbres c'est la même chose, mais appliquées à des familles dénombrables. (Pour mémoire, j'appelle "dénombrable" un ensemble qui est soit fini, soit en bijection avec $\mathbb{N}$; d'autres auteurs appellent "dénombrable" un ensemble qui est en bijection avec $\mathbb{N}$; c'est une question de convention sans importance.)

Définition II-6 (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) $A \in \mathcal{A} \Longrightarrow X \setminus A \in \mathcal{A}$ *;*

(iii) $[\forall k \in \mathbb{N},\ A_k \in \mathcal{A}] \Longrightarrow \displaystyle\bigcup_{k\in\mathbb{N}} A_k \in \mathcal{A}$.

Remarque II-7. Une σ-algèbre est une algèbre : on le voit en prenant $A_1 = A$, $A_2 = B$, $A_k = \emptyset$ pour tout $k \geq 3$. Une σ-algèbre est automatiquement stable par **intersection dénombrable**, comme on le voit en passant aux complémentaires dans (iii). En fait une σ-algèbre est le cadre

Exemple II-8. Les algèbres triviales sont aussi des σ-algèbres. Toute algèbre finie est une σ-algèbre ; c'est le cas en particulier de celles qui sont associées à une partition.

Exemple II-9. L'ensemble des unions finies d'intervalles de $\mathbb{R}$ est une algèbre, mais ce n'est pas une σ-algèbre. Idem pour l'ensemble des unions finies de pavés de $\mathbb{R}^2$.

Exercice II-10. Soit X un ensemble infini. Montrer que l'ensemble des parties finies ou cofinies (cofini = dont le complémentaire est fini) est une algèbre, mais pas une σ-algèbre.

Remarque II-11. La lectrice familière 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) $O_1, O_2 \in \mathcal{O} \Longrightarrow O_1 \cap O_2 \in \mathcal{O}$, (iii) $\forall i, O_i \in \mathcal{O} \Longrightarrow \cup O_i \in \mathcal{O}$. Noter que dans (iii), la famille I indexant les O_i est arbitraire (pas nécessairement dénombrable). Autrement dit, une topologie est stable par intersection finie et union quelconque, alors qu'une σ-algèbre est stable par intersection dénombrable et union dénombrable – et passage au complémentaire.

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 II-3.

Dans la suite, je privilégierai la dénomination de “tribu” pour désigner les σ-algèbres.

Définition II-12 (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 d'une tribu triviale.

Considérer des σ-algèbres plutôt que des algèbres est séduisant car on obtient ainsi des familles riches qui se comportent bien vis-à-vis des unions infinies, limites, etc. Mais en pratique, les tribus ne seront pas explicites ; on ne les maniera pas directement, on préfèrera les voir comme des “limites”, dont le comportement est dicté par une famille “dense” beaucoup plus simple, et on travaillera seulement sur cette dernière famille. Cela n'a de sens que si la tribu est complètement déterminée par cette fameuse famille ; c'est l'objet du concept qui suit, à la fois élémentaire et fondamental à la théorie.

Proposition II-13 (tribu engendrée par une famille). *Soit X un ensemble quelconque, et soit $\mathcal{F}$ 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 $\sigma(\mathcal{F})$.*

La démonstration de cette proposition est laissée en exercice. (Ne pas oublier de montrer que l'intersection apparaissant dans l'énoncé n'est pas vide.)

Exercice II-14. Soit $\mathcal{F}$ une famille de parties d'un ensemble X, stable par passage au complémentaire. Montrer que $\sigma(\mathcal{F})$ est la plus petite classe contenant $\mathcal{F}$ qui soit stable par intersection dénombrable et union dénombrable. Indication : Si $\mathcal{A}$ est une classe vérifiant les propriétés précédentes, on pourra montrer que $\{A \in \mathcal{A};\ X \setminus A \in \mathcal{A}\} \supset \sigma(\mathcal{F})$.

Exemple II-15 (Tribu engendrée par les intervalles). Tout ouvert de $\mathbb{R}$ peut s'écrire comme union disjointe dénombrable d'intervalles ouverts, qui sont ses composantes connexes (cette union est dénombrable car chacun de ces intervalles contient au moins un rationnel). 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 borélienne** ou **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'unions dénombrables d'intersections dénombrables d'ensembles ouverts ou fermés, etc. — et plus encore.

Cette richesse est donc à la fois une force et une faiblesse : il est en pratique impossible de “décrire” ce qu'est un ensemble borélien “générique” dans $\mathbb{R}$. On peut se les représenter par le procédé itératif suivant, transfini, c'est à dire qu'il fait intervenir des cardinaux plus grands que celui de $\mathbb{N}$ (d'habitude noté $\aleph_0$). À l'étape 0, on considère les ouverts et les fermés. À l'étape 1, les unions dénombrables d'intersections dénombrables d'ouverts et de fermés. À l'étape j, les unions dénombrables d'intersections dénombrables des ensembles apparaissant à l'étape $j-1$. Et l'on continue... une infinité de fois ; en fait, on s'arrête dès que l'on atteint un ordinal non dénombrable.

C'est d'une complexité insaisissable... Pourtant, la tribu borélienne réelle n'est pas si peuplée : elle a “seulement” la puissance du continu, son cardinal est $c = 2^{\aleph_0}$, la puissance du continu (qui est au moins égal à $\aleph_1$ le premier cardinal strictement plus grand que $\aleph_0$; l'égalité entre c et $\aleph_1$ étant le plus célèbre des problèmes indécidables). Au sens de la théorie des ensembles, *la tribu borélienne est aussi peuplée que la droite réelle*, et il n'y a donc “pas plus” de boréliens que d'ouverts – alors que l'ensemble des parties de $\mathbb{R}$ est phénoménalement plus grand, de cardinal 2^c. Les boréliens, bien que très complexes, gardent une certaine régularité, bien plus (“infiniment plus”) que des parties quelconques de $\mathbb{R}$.

Voici les exemples les plus importants de tribus :

Exemples II-16. (i) **Tribu induite par restriction** : Soit $(X, \mathcal{A})$ un espace mesurable, et $Y \in \mathcal{A}$. Alors l'ensemble des éléments de $\mathcal{A}$ qui sont inclus dans A définit une tribu de parties de A, appelé tout naturellement la restriction de $\mathcal{A}$ à A.

(ii) **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. Et par l'Exercice II-14, c'est la plus petite classe stable par union et intersection dénombrable qui contienne les ouverts et les fermés.

(iii) **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 dyadiques fermés $\prod [m_k 2^{-\ell_k}, (m_k+1)2^{-\ell_k}]$, $m_k, \ell_k \in \mathbb{N}$ (ou ouverts, ou semi-ouverts) ;
- les boules ouvertes (ou fermées) dans $\mathbb{R}^n$.

En dimension $n = 1$, toutes ces familles se ramènent à une seule : la famille des **intervalles** de $\mathbb{R}$.

(iv) **Tribu produit** : Cette construction suit celle de l'Exemple II-4(v). 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 $\mathcal{A} \otimes \mathcal{B}$. Cette construction se généralise facilement au produit d'un nombre *fini* d'espaces mesurables. La tribu produit est facile à définir, mais on ne peut guère la décrire explicitement : elle contient les pavés, les intersections dénombrables d'unions dénombrables de pavés, etc.

(v) **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 tribu engendrée par les cylindres élémentaires, suivant la construction de l'Exemple II-4(vi). Il s'agit en fait de la généralisation naturelle du concept de tribu produit. C'est la tribu classique que l'on utilise d'ordinaire sur un produit infini. Ce n'est pas la seule tribu possible, et certaines constructions alternatives font l'objet de recherches récentes (contributions de Boris Tsirelson, par exemple) ; mais la tribu cylindrique est bien celle que l'on utilise dans la quasi-totalité des cas.

Exemple II-17 (Boréliens de $\overline{\mathbb{R}}$). Soit $\mathbb{R} = [-\infty, +\infty]$ la droite réelle complétée, c'est à dire à laquelle on adjoint $-\infty$ et $+\infty$. On en fait un espace topologique en considérant la topologie engendrée par tous les intervalles ouverts et par $\{-\infty\}$ et $\{+\infty\}$. Alors 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 $]a, +\infty]$; ce sont aussi les boréliens de $\mathbb{R}$ auxquels on s'autorise à adjoindre $-\infty$ ou $+\infty$.

Démonstration.]Preuve de l'Exemple II-17] Vérifions-le par exemple pour la famille $\mathcal{F}$ des intervalles $[a, +\infty]$. On a $\{+\infty\} = \cap_{n\in\mathbb{N}}[n, +\infty]$, donc $\{+\infty\} \in \sigma(\mathcal{F})$. Le complémentaire de $[a, +\infty]$ est $[-\infty, a[$ donc tous les intervalles de la forme $[-\infty, b[$ appartiennent aussi à $\sigma(\mathcal{F})$. C'est donc aussi le cas de $\{-\infty\} = \cap_{n\in\mathbb{N}}[-\infty, n[$. De même pour un intervalle $]a, b[$ puisqu'il peut s'écrire $\cup_{n\in\mathbb{N}}[a+1/n, b[$. Donc $\sigma(\mathcal{F})$ contient les unions dénombrables d'intervalles ouverts, c'est à dire les ouverts de $\mathbb{R}$; et comme il contient aussi $\{-\infty\}$ et $\{+\infty\}$, par adjonction $\sigma(\mathcal{F})$ contient tous les ouverts de $\overline{\mathbb{R}}$, et donc tous les boréliens de $\overline{\mathbb{R}}$, ce qui conclut la preuve. □

Exemple II-18. 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

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

Mais aussi, tout sous-ensemble dénombrable de $\{0, 1\}^{\mathbb{N}}$ est mesurable. Ou encore, par exemple, si on se donne $\alpha \in [0, 1]$, l'ensemble des $(x_n)_{n\in\mathbb{N}}$ tels que le nombre moyen d'occurrences de 0 dans $(x_n)_{1\leq n\leq N}$ tende vers α (c'est à dire les $(x_n)_{n\in\mathbb{N}}$ tels que $N^{-1}\#\{j \in \{1, \ldots, N\};\ x_j = 0\} \to \alpha$ quand $N \to \infty$) est encore mesurable, de même que tous les autres ensembles auxquels vous pourrez penser. En fait, on peut montrer qu'il est impossible de décrire une partie de $\{0, 1\}^{\mathbb{N}}$ qui ne soit pas mesurable pour la tribu cylindrique.

Remarque II-19. La construction de la tribu produit est très similaire à celle de la topologie produit, qui sera rappelée dans la section II-3.4. 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 IV. Pour se donner une idée des problèmes proprement vertigineux que l'on peut rencontrer, la lectrice 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 ?

Exercice II-20. Démontrer l'assertion faite dans l'Exemple II-16(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).

Exercice II-21. Soient $m, n \in \mathbb{N}$. En utilisant le résultat de l'exercice II-20, montrer que $\mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}^m \times \mathbb{R}^n)$.

II-1.3. Mesures. Les tribus sont le cadre naturel pour définir les mesures, introduites par Borel : ce sont des fonctions d'ensembles vérifiant l'axiome de σ-additivité.

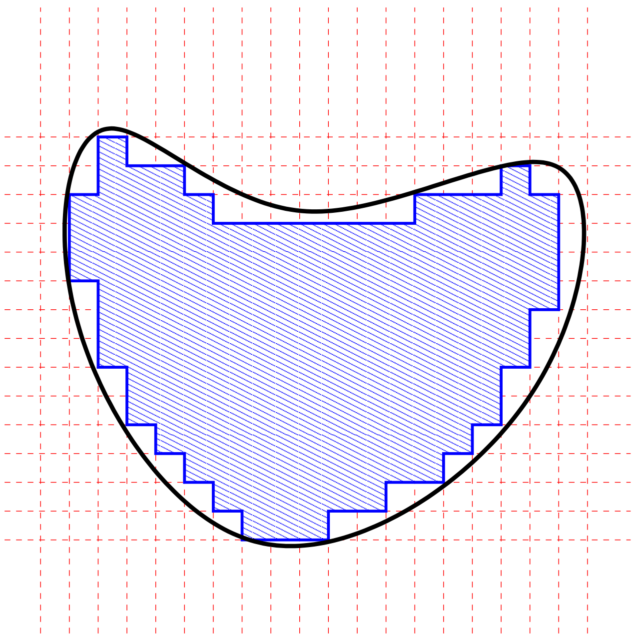

Figure 3. Approximation d'un ouvert de $\mathbb{R}^n$ par une union de petits pavés

Définition II-22. *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* $[0, +\infty]$*, telles que*

(i) $\mu[\emptyset] = 0$ *;*

(ii) Pour toute famille dénombrable $(A_k)_{k\in\mathbb{N}}$ *d'ensembles mesurables disjoints,*

$$\mu\Big[\bigcup_{k\in\mathbb{N}} A_k\Big] = \sum_{k\in\mathbb{N}} \mu[A_k]. \tag{3}$$

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

Remarques II-23. La somme apparaissant au membre de droite de (3) converge toujours dans $[0, +\infty]$. À la place de l'axiome $\mu[\emptyset] = 0$, on aurait pu imposer que μ n'est pas identiquement $+\infty$; en effet, cela impose $a = \mu[\emptyset] < +\infty$, mais alors en choisissant $A_k = \emptyset$ pour tout k, on a $a = \infty \cdot a$, d'où $a = 0$. Bien sûr, les mesures vérifient toutes les règles habituelles des fonctions additives d'ensembles, comme discuté en début de chapitre ; en particulier, une mesure μ est une fonction croissante d'ensembles, et pour toutes parties mesurables A, B,

$$\mu[A\cap B] < \infty \Longrightarrow \quad \mu[A\cup B] = \mu[A] + \mu[B] - \mu[A\cap B].$$

Remarque II-24. Dans la définition on a imposé que les A_k soient disjoints ; par là on entend *deux à deux disjoints*, c'est à dire que pour tous $i, j \in \mathbb{N}$ on a $A_i \cap A_j = \emptyset$. L'exercice suivant aborde le cas d'ensembles non disjoints.

Exercice II-25. Soit μ une mesure. Montrer que pour toute famille dénombrable (A_k) de parties mesurables,

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

Pourquoi l'inégalité doit-elle être en général stricte ?

Si $(X, \mathcal{A}, \mu)$ est 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.

Notons tout de suite certaines opérations simples qui peuvent être effectuées sur les mesures :

Proposition II-26 (Opérations sur les mesures). *(i) Si μ_1 et μ_2 sont des mesures définies sur le même espace mesurable, alors $\mu_1 + \mu_2$ est une mesure.*

(ii) Si μ est une mesure et α un nombre réel positif, alors $\alpha\mu$ est une mesure.

(iii) Si $(X, \mathcal{A}, \mu)$ est un espace mesuré, et Y une partie mesurable de X, alors μ induit sur Y une mesure par restriction à la tribu des parties mesurables de X qui sont incluses dans Y. Cette mesure ν, restriction de μ à Y, est notée

$$\nu = \mu|_Y \qquad \textit{ou} \qquad \nu = \mu\lfloor_Y.$$

(iv) Si $(\mu_n)_{n\in\mathbb{N}}$ est une suite croissante de mesures définie sur un même espace mesurable (c'est à dire : pour tout A et pour tout n, $\mu_n[A] \leq \mu_{n+1}[A]$), alors $\lim \mu_n$ *est une mesure.*

Exercice II-27. Démontrer la propriété (iv) ci-dessus, et montrer par un contre-exemple que la conclusion est en général fausse pour une suite décroissante.

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 suite $(A_k)_{k\in\mathbb{N}}$ 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 II-28. *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré.*

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

$$\mu\Big[\bigcup_{k\in\mathbb{N}} A_k\Big] = \lim_{k\to\infty} \mu[A_k] = \sup_{k\in\mathbb{N}} \mu[A_k].$$

(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

$$\mu\Big[\bigcap_{k\in\mathbb{N}} A_k\Big] = \lim_{k\to\infty} \mu[A_k] = \inf_{k\in\mathbb{N}} \mu[A_k].$$

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

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

La série de droite est donc convergente, et

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

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 $(A_0 \setminus A_k)$. □

Exercice II-29. 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.

Je conclurai cette section avec une liste de mesures importantes. Attention : si la construction des σ-algèbres est souvent un exercice facile grâce à la notion de tribu engendrée, la construction des mesures peut poser des problèmes bien plus considérables. Dans la liste ci-dessous, les deux premiers exemples sont faciles à construire, mais les suivants sont beaucoup plus subtils ; pour l'instant, j'admettrai leur existence, et ce n'est que dans la suite du chapitre que viendront les outils puissants qui permettent de les réaliser.

II-2. 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 un point de X, on note δ_x la mesure définie par

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

Autrement dit $\delta_x[A] = 1_A(x)$. On peut se représenter cette mesure comme une "masse ponctuelle" située au point x_0. Cette mesure est nommée en l'honneur de Paul Dirac, grand théoricien de la physique quantique, qui la maniait dans de nombreux calculs formels. Malgré sa simplicité, la masse de Dirac peut être considérée est une sorte de "brique élémentaire" des mesures : de larges classes de mesures peuvent en effet être "approchées" par des combinaisons de masses de Dirac. On peut les interpréter comme les points extrémaux de l'espace convexe des mesures de probabilité ; et toute mesure peut être vue comme une combinaison convexe de masses de Dirac.

(ii) La **mesure de comptage** n'est autre que la fonction "cardinal", à valeurs dans $\mathbb{N} \cup \{+\infty\}$. C'est aussi $\sum_{x\in X} \delta_x$.

(iii) La **mesure de Lebesgue** λ_n 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 ($n = 3$), et les généralise en toute dimension. On peut la définir en spécifiant le volume des pavés, ou celui des boules, selon les formules vues précédemment. Par exemple on peut dire que λ_n est l'unique mesure borélienne telle que $\mu\big[\prod_i [a_i, b_i]\big] = \prod_i (b_i - a_i)$. (Attention, à ce stade je n'ai prouv́e ni l'existence ni l'unicité, et cele n'a rien d'évident ; j'approfondirai la construction de la mesure de Lebesgue dans la section II-8, puis dans le chapitre VI.)

(iv) Les **mesures à densité** : Soit f une fonction continue par morceaux, positive, des variables $x_1, \ldots, x_n$; cela induit sur $\mathbb{R}^n$ une unique mesure μ telle que pour tout pavé $P = \prod[a_i, b_i]$,

$$\mu[P] = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n)\, dx_1 \dots dx_n.$$

On verra plus tard que l'on peut aussi considérer des fonctions f bien moins régulières.

(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 $\tau_a : x \longmapsto (a.x)$. 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 ces notions !), 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, dans l'espace métrique (M, d), où d est la distance géodésique associée à g. De façon équivalente, on peut écrire $\mathrm{vol}\,(dx) = \sqrt{\det g}\, dx^1 \dots dx^n$ dans une carte. Si M est une sous-variété de codimension 1 dans $\mathbb{R}^{n+1}$, on peut utiliser encore une autre définition équivalente : pour $X \subset M$ on considère $X_\varepsilon = \{x \in \mathbb{R}^{n+1}; d(x, X) \leq \varepsilon\}$ pour $\varepsilon > 0$, et on définit $\mathrm{vol}\,[X]$ comme la limite de $|X_\varepsilon|/\varepsilon$ quand $\varepsilon \to 0$, où $|X_\varepsilon|$ est la mesure $(n + 1)$-dimensionnelle de X_ε.

(vii) La **mesure de Wiener**, définie sur $X = C([0, T], \mathbb{R}^d)$ (muni de la topologie de la convergence uniforme), où $d \in \mathbb{N}$ et $T > 0$ est fixé. Cette mesure est intimement liée au mouvement brownien, qui gouverne les trajectoires très chaotiques de petites particules dans un fluide. 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 $[0, T]$; alors 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 qui possèdent cette propriété. La mesure de Wiener est l'unique mesure de Borel $\mathcal{W}$ sur X qui vérifie la condition cylindrique suivante : Pour tout $K \in \mathbb{N}$ et toute suite strictement croissante $(t_1, \dots, t_K)$ dans $(0, T]$, pour tous pavés $P_1, \dots, P_K$ dans $\mathbb{R}^d$,

$$\mathcal{W}\Big[\big\{x \in X;\ x(t_1) \in P_1, \dots, x(t_K) \in P_K\big\}\Big]$$
$$= \int_{P_1} dx_1 \dots \int_{P_K} dx_K \exp\left[-\Big(\frac{|x_1|^2}{2t_1} + \frac{|x_2 - x_1|^2}{2(t_2 - t_1)} + \dots + \frac{|x_K - x_{K-1}|^2}{2(t_K - t_{K-1})}\Big)\right].$$

La construction de cette mesure en 1921 par Norbert Wiener marquait l'aboutissement d'une démarche sur l'intégration en dimension infinie entamée quelques années plus tôt par René Gateaux et Paul Lévy. La mesure de Wiener a été l'un des premiers grands succès de la théorie de Lebesgue appliquée à des espaces de dimension infinie.

II-3. Rappels de topologie

Dans la section II-1, les concepts topologiques ne jouaient aucun rôle, sauf pour la définition des tribus et mesures de Borel. Mais dans toute la suite du cours, on maniera des mesures de Borel de façon plus subtile, et la topologie s'invitera avec insistance. Pour préparer cela, je vais brièvement passer en revue les notions utiles ; la lectrice pourra compléter elle-même les démonstrations, ou bien les retrouver dans les ouvrages d'introduction à la topologie.

Préliminaires

Soit X un ensemble quelconque ; on le munit d'une topologie 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. On peut se représenter un ouvert comme un ensemble qui "entoure" chacun de ses points ; ou encore, comme un ensemble dont aucun point n'est frontière.

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

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

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, $\mathrm{Int}(A)$, comme le plus grand ouvert contenu dans A, c'est-à-dire l'union de tous les ouverts contenus dans A. La **frontière** de A est $\overline{A} \setminus \mathrm{Int}(A)$; il est commode pour l'intuition de se représenter l'adhérence de A comme "A avec toute sa frontière".

Une partie A est dite dense dans X si $\overline{A} = X$.

Si X est un espace topologique, un sous-ensemble arbitraire A de X devient lui-même 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 $n \geq N \Longrightarrow x_n \in V$. 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 $f : X \to Y$ 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 voisinage est une abstraction extrême de la notion de "proximité" : des voisinages emboîtés autour d'un point x, définissent les points qui sont de plus en plus proches de x. Une fonction est continue si elle préserve la proximité. Quant aux composantes connexes, ce sont en quelque sorte les 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) ; par exemple, il suffit de remplacer une distance d par $g(d)$, où g est n'importe quelle fonction strictement croissante ; ce qui laisse penser que le concept de topologie est plus satisfaisant que celui de métrique.

Cependant,

(a) en général, une topologie abstraite non métrique peut présenter des propriétés "pathologiques" du point de vue de l'analyse ;

(b) une métrique permet de quantifier la notion de "proximité" (la distance donne une valeur), ce qui est précieux dans bien des problèmes pratiques ;

(c) en analyse, on peut (presque) toujours se ramener à des topologies métriques. Tous les problèmes que j'ai jamais rencontrés, que ce soit en physique mathématique, en probabilité, en géométrie différentielle, et en tout cas en théorie de la mesure, peuvent se formuler en n'utilisant que des topologies métriques. Cela n'est pas vrai dans d'autres domaines mathématiques : par exemple, la topologie de Zariski, très utile en géométrie algébrique, est résolument non métrique. Mais dans un cours d'intégration, il est tout à fait légitime de se restreindre aux topologies métriques.

II-3.1. Espaces métriques. Soit X un ensemble quelconque ; on dit qu'une application $d : X \times X \to \mathbb{R}_+$ 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.

Un exemple très particulier est la **topologie triviale**, où toutes les parties de X sont ouvertes ; elle correspond à la métrique triviale $d(x,y) = 1_{x \neq y}$.

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

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

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 ;

- V est voisinage de x si et seulement si il existe $r > 0$ tel que $B_r(x) \subset V$;

- O est ouvert si et seulement si il est voisinage de tous ses points ;

- $f : (X,d) \to (X',d')$ définie entre deux espaces métriques est continue si et seulement si

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

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

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

- 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 $x_n \in F$, $x_n \longrightarrow x \in X$ impliquent $x \in F$;

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

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.

II-3.2. 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 $n \to k(n)$, définie de $\mathbb{N}$ dans $\mathbb{N}$, telle que $x_{k(n)} \longrightarrow x$.

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

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

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 $X \subset \cup_{j\in J} O_j$. 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** [1] si son adhérence $\overline{A}$ est compacte.

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 ; une partie fermée d'un compact est compacte ;

- 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 parlera plus tard des 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

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

(en effet 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, d) est compact si et seulement si toute suite $(x_n)_{n\in\mathbb{N}}$ 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

$$\forall x, y \in X \qquad d(x, y) \leq \delta \Longrightarrow\ d'\big(f(x), f(y)\big) \leq \varepsilon.$$

On peut alors définir le module de continuité de f par

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

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

---

1. J'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 est minime : en effet, 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 II-39.

Exemples II-30. 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 arbitrairement près). 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 II-31 (σ-compacité). *Un espace topologique X est dit σ-compact s'il est union dénombrable de compacts.*

Définition II-32 (compacité locale). *Un espace topologique X est dit* **localement compact** *si tout $x \in X$ admet un voisinage compact.*

Définition II-33 (espace polonais). *Un espace topologique X est dit* **polonais** *si c'est un espace métrique séparable et complet.*

Remarque II-34. Un espace polonais localement compact est automatiquement σ-compact (exercice).

Exemple II-35. 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, $C([0,1],\mathbb{R}^n)$ (muni de la norme du supremum), est polonais mais n'est ni localement compact, ni σ-compact ; en fait tous les compacts y sont d'intérieur vide.

Remarque II-36. 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.

II-3.3. 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 $f(\lim x_n) = \lim f(x_n)$.

- 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 que pour tous a_0, a_1 réels distincts, 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 $f : A \to \mathbb{R}$ admet un prolongement lipschitzien à X

tout entier (a priori non unique). Pour le voir, il suffit de poser

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

- Si (X,d) est un espace métrique et F est un fermé de X, alors toute fonction continue $f : F \to \mathbb{R}$ 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 vai 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 $|f - (g_1 + \ldots + g_n)| \leq 2^{-n}$ 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 $\widetilde{f} := f/(1+f)$, construisant ainsi une fonction continue $\widetilde{g}$ sur X, à valeurs dans $[0,1]$. L'ensemble $F' := \{\widetilde{g} = 1\}$ 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 $h\widetilde{g}$ est alors à valeurs dans $[0,1[$, et $g := (h\widetilde{g})/(h\widetilde{g} - 1)$ est une extension continue de f. □

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 $O_0 = \{x;\ f(x) < 1/3\}$, $O_1 = \{x;\ f(x) > 2/3\}$.

II-3.4. 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

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

On peut alors définir la **topologie produit** comme la topologie engendrée par les ouverts cylindriques, i.e. les ouverts de la forme $C(O_{t_1}, \ldots, O_{t_K})$, 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 $C(A_1, \ldots, A_n)$, 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

$k \to \infty$. Il est alors facile de vérifier que X est un espace métrique quand on le munit de la distance

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

(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 II-37 (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 II-38. 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 car on peut utiliser à la fois la métrique et la dénombrabilité.

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 $(x^n_k)_{k \in \mathbb{N}}$, avec $x^n_k \in X_k$. De la famille des $(x^n_1)_{n \in \mathbb{N}}$, 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 $\varphi_k : \mathbb{N} \to \mathbb{N}$, telles que la suite $x_k^{\varphi_1 \circ \ldots \circ \varphi_k(n)}$ est convergente dans X_k. On pose alors

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

Pour tout $k \leq n$, on peut écrire

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

où ψ_k est une fonction croissante. Il s'ensuit que, pour $n \geq k$, la suite $(x_k^{\varphi(n)})$ est extraite de $x_k^{\varphi_1 \circ \ldots \circ \varphi_k(n)}$, 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

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

métrise la topologie produit de X et le rend complet.

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

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

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

II-3.5. 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" pour topologie (par exemple un ensemble gras au sens de Baire, i.e. une intersection dense d'ouverts) et très petit pour la mesure (i.e. inclus dans une famille de boules de volume total arbitarirement petit) ; on y reviendra.

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) va de pair 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 don 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 excellents ont été établis pour ces espaces ; parmi les plus remarquables se trouvent le Théorème de Représentation de Riesz (Chapitre III) ; et le Théorème de Haar (Chapitre **??**). 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 parallélement au cours du vingtième siècle, avec des points communs et des divergences : les espaces localement compacts d'une part, les espaces polonais d'autre part. 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.

L'expérience a montré 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 est clair maintenant que le “bon” contexte mathématique de la théorie de la mesure est celui des espaces polonais. Cela n’empêche pas que des hypothèses *additionnelles* de compacité locale aient des conséquences fort commodes.

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 ; la lectrice intéressée pourra en reconstituer les détails.

Qu’est-ce qui fait le succès de ces deux catégories d’espace ? Pour simplifier,

- dans les espaces polonais, tout se ramène à des boules ;

- dans les espaces localement compacts, tout se ramène à des fonctions continues.

Je vais en dire un peu plus dans la fin de cette section.

II-3.6. 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 II-39 (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 quelconque de nombres positifs décroissant vers 0. Soit $\mathcal{B}_k$ l’ensemble des boules ouvertes de centre x_n et 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, K est compact si et seulement si il existe (N_k) et $(n(k,j))$ ($k\in\mathbb{N}$, $1\le j\le N_k$) tels que

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

Remarque II-40. La formule (4) dit que l’on peut bien approcher K en le recouvrant par un grand nombre de petites boules que l’on ferme. Noter que l’union de boules fermées apparaissant au membre de droite n’est pas forcément compacte ; c’est l’intersection qui l’est.

Preuve du Théorème II-39. (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 $d(x_n, x) < \varepsilon_k$. On va 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 $d(x, y) \le d(x, x_n) + d(x_n, y) < \varepsilon_k + \varepsilon_k < r$, 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. ☐

II-3.7. 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 II-41 (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 II-42 (partition de l'unité). *Soient X un espace séparé localement compact, K un compact de X, et $O_1, \ldots, O_n$ une collection finie d'ouverts, tels que $K \subset \cup O_k$. Alors il existe des fonctions continues $f_1, \ldots, f_n$, à valeurs dans $[0, 1]$, telles que chaque f_k ait son support compact et inclus dans O_k, et*

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

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

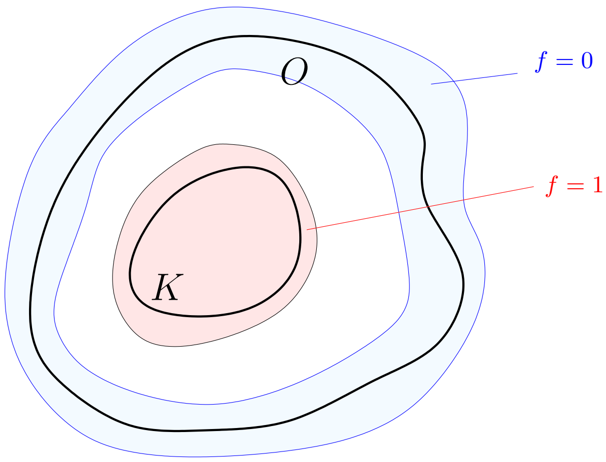

Figure 4. Le lemme d'Urysohn

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 II-43 (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 II-43 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 $V_{x_1}, \ldots, V_{x_K}$, 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. □

Preuve du Théorème II-41 dans le cas métrique. Soient K' et O' comme dans le Lemme II-43. 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). □

Preuve du Théorème II-42 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
$$K \cap F_j \subset \Big(\bigcap_i O_i\Big) \setminus \Big(\bigcap_{i \neq j} O_i\Big) \subset O_j.$$
Grâce au Théorème II-41, 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

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

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

Voici pour finir une proposition 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) :

Proposition II-44. *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 $C_1 \subset O_1 \cup O_2 \cup \ldots \cup O_{N_1}$. En particulier, le compact $C'_1 \cup \ldots \cup C'_{N_1}$ est voisinage de C_1. Il suffit alors de poser $K_2 := C'_1 \cup \ldots \cup C_{\max(2,N_1)}$.

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 $C_1 \cup \ldots \cup C_{n+1}$. 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. □

II-4. Régularité des espaces mesurés

De nombreux résultats de théorie de la mesure s'appliquent dans un cadre très général, sans aucune hypothèse additionnelle. Mais en pratique il est très commode de pouvoir s'appuyer sur certaines propriétés très utiles, dites de "régularité", qui mêlent des hypothèses de topologie et de théorie de la mesure.

II-4.1. Vocabulaire de base.

Définition II-45 (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 II-46. Certains théorèmes importants d'intégration sur les espaces produits utiliseront crucialement des hypothèses de σ-finitude.

Définition II-47 (probabilité). *Un espace mesuré (X, μ) est appelé espace de probabilité si $\mu[X] = 1$.*

Remarque II-48. 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 indépendamment à Kolmogorov et Ulam.

Définition II-49 (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 II-50 (concentration, négligeabilité). *Soit $(X, \mathcal{A}, \mu)$ 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 (ou 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 $\mu[A] > 0$.

Remarque II-51. 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 A si et seulement si $\mu[A] > 0$.

Exemple II-52. 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 II-53 (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 II-54. (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.

II-4.2. Mesures de Borel et régularité.

Définition II-55 (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”.

Définition II-56 (régularité). *Soit X 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*[2] *si elle vérifie la propriété caractéristique suivante : pour tout ensemble mesurable $A \in \mathcal{A}$ on a*

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

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 II-57 (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,*

(i) 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 mesurable négligeable ;

(ii) A 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 mesurable 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 $\mu[K_j] \to \mu[E]$, $\mu[O_j] \to \mu[E]$. 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 $F = \cup K_j$ et $G = \cap O_j$. La conclusion découle de la σ-additivité de μ. $\square$

La régularité va souvent de pair avec la propriété suivante :

Définition II-58 (finitude sur les compacts). *Une mesure de Borel μ sur un espace topologique X est dite finie sur les compacts*[3] *si pour tout compact K de X on a $\mu[K] < +\infty$.*

Exemple II-59. 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 II-60. *Soient $\mathcal{A}$ et $\mathcal{B}$ deux algèbres, avec $\mathcal{A} \subset \mathcal{B}$, et $\mu : \mathcal{B} \to \mathbb{R}_+$ une fonction additive d'ensembles, vérifiant la propriété de régularité intérieure partielle*

$$\forall B \in \mathcal{B} \quad \mu[B] = \sup\Big\{\mu[K];\ K \text{ compact},\ K \subset B,\ K \in \mathcal{A}\Big\}. \tag{5}$$

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

Remarque II-61. La Remarque VIII-67 montrera que l'hypothèse 5 est cruciale.

2. Cette propriété est parfois appelée du nom de Radon, et le terme "régularité" désigne parfois la propriété d'approximation par des fermés, plutôt que par des compacts [Bogachev].

3. Une 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].

Preuve de la Proposition II-60. Soit μ vérifiant les hypothèses du théorème. Par aditivité on a, pour toute famille $(A_k)_{k\in\mathbb{N}}$ d'ensembles mesurables disjoints, $\sum_{1\le k\le N}\mu[A_k]\le\mu[\cup_{k\in\mathbb{N}}A_k]$, d'où

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

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

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

d'où $\mu[\cup_{k\ge N+1}A_k]\ge\delta$ (ici on utilise le fait que μ est à valeurs finies). Posons $C_\ell=\cup_{k\ge\ell+1}A_k$: alors $(C_\ell)_{\ell\in\mathbb{N}}$ est une suite décroissante de compacts d'intersection vide, et $\mu[C_\ell]\ge\delta$ 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 $\mu[C_\ell\setminus K_\ell]\le 2^{-(\ell+1)}\delta$. Alors $K_1\cap\ldots\cap K_\ell\in\mathcal{A}$ et

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

en particulier $K_1\cap\ldots\cap K_\ell$ est non vide. Les compacts $(K_1\cap\ldots\cap K_\ell)_{\ell\in\mathbb{N}}$ 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. ☐

II-4.3. 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 II-62 (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 II-62.

Corollaire II-63 (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 II-62, mentionnons sans démonstration une variante qui s'applique à des espaces topologiques localement compacts [Rudin, Théorème 2.18], et implique également le Corollaire II-63.

Théorème II-64 (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 II-65. 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 II-64 n'est pas forcément vraie sans l'hypothèse quelque peu étrange de σ-compacité des ouverts.

Voici maintenant une preuve du Théorème II-62, 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 II-77), mais bien sûr il n'y a pas de cercle vicieux !

Preuve du théorème II-62. 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 $S := \cup_{k\in\mathbb{N}} K_{1/k}$.

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

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

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

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

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

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

Posons

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

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

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

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,

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

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

$$\mu[O] - \varepsilon \leq \mu[A] \leq \mu[F] + \varepsilon. \tag{6}$$

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 II-77, 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 $\mu[F_k] \to \mu[A]$. Introduisons, pour $k \in \mathbb{N}$,

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

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 $\mu[F_k] \to \mu[A]$. Or chaque F_k est fermé, puisque image réciproque de l'intervalle fermé $[1/k, +\infty[$ par l'application continue (et même 1-lipschitzienne) $x \longmapsto d(x, A)$.

4. Éliminons 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. Éliminons 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 $O = \cup O_n$: alors O est un ouvert contenant $\cup(A \cap X_n) = A$, et

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

Pour l'autre sens, distinguons deux cas. Si $\mu[A] = +\infty$, alors $\mu[A \cap X_n] \to +\infty$, et l'inégalité $\mu[K_n] \geq \mu[A \cap X_n] - \varepsilon$ implique $\mu[K_n] \to \infty$. 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. □

II-5. Concentration

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 puissante théorie de la "concentration de la mesure" étudie et quantifie cela pour bon nombre de mesures apparaissant dans des problèmes variés de géométrie, statistique ou physique. Ici nous allons simplement passer en revue quelques notions de base sur le support et la diffusivité (le caractère diffus) d'une mesure.

II-5.1. Support. Le support est le plus petit fermé sur lequel μ est concentrée.

Théorème II-66 (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.*

En combinant ce théorème avec les théorèmes de régularité automatique II-62 et II-64, on obtient le

Corollaire II-67. *Si μ est une mesure de Borel sur un espace topologique X, et que*

- soit X est polonais et μ est σ-finie,

- soit X est séparé, localement compact et tous les ouverts y sont σ-compacts, et μ est finie sur les compacts,

alors on peut définir le support de μ.

Remarque II-68. 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. Si f est une mesure à densité continue (Exemple (iv) dans la section II-2), les deux notions coïncident.

Démonstration du Théorème II-66. 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 $x_1, \ldots, x_J \in K$ 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 II-69. Dans le cas polonais, au lieu d'obtenir le corollaire II-67 en combinant les théorèmes II-66 et II-62), on peut aussi raisonner directement en notant qu'un espace polonais admet une base $\mathcal{B}$ dénombrable d'ouverts (comme dans le Théorème II-39). 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 $\mu[\omega] = 0$.

Faisons enfin le lien avec la notion d'atome :

Proposition II-70 (une mesure insécable est un atome). *(i) Soit $(X, \mathcal{A}, \mu)$ un espace séparé mesuré, tel que $\mu[X] > 0$ et tel qu'on ne peut séparer X en ensembles mesurables disjoints X_1 et X_2 de mesure positive. Alors μ est un atome : $\mu = m\delta_x$ pour un certain $x \in X$ et $m > 0$ (éventuellement $m = \infty$).*

(ii) Plus généralement, si l'on ne peut pas trouver plus de K parties mesurables disjointes de mesures positives, alors μ est une combinaison de mesures de Dirac placées en K points : $\mu = \sum_{i=1}^{K} m_i \delta_{x_i}$.

Preuve de la Proposition II-70. Démontrons (i). Si $\operatorname{Spt} \mu$ est un singleton $\{x\}$, alors μ est de la forme $m\delta_x$; sinon on peut trouver x_1 et x_2 distincts dans $\operatorname{Spt} \mu$; alors si B_1 et B_2 sont deux boules disjointes centrées en x_1 et x_2 respectivement, on a $\mu[B_1] > 0$ et $\mu[B_2] > 0$, donc on sépare X en deux parties disjointes de masses positives via $X_1 = B_1$ et $X_2 = X \setminus X_1$.

La preuve de (ii) suit le même raisonnement. $\square$

II-5.2. Diffusivité. Comment traduire l'idée qu'une mesure sur un espace X est “diffuse”, “bien répartie”, qu'elle “charge tout l'espace” ? Plusieurs notions coexistent, les plus simples sont les suivantes.

- On dit que μ est **sans atomes** si il n'existe aucun x tel que $\mu[\{x\}] > 0$.
- On dit que μ est **de plein support** si son support est égal à X tout entier.
- On dit que μ est **doublante sur les boules** si sa restriction à chaque boule $B(x_0, R)$ est doublante (avec une constante de doublement $C = C(R, x_0)$), au sens de la définition suivante.

Définition II-71 (doublement). *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 II-72. Il est équivalent de définir ce concept en termes de boules ouvertes ou de boules fermées. En effet, par σ-additivité,

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

Exemples II-73. On verra plus tard que la mesure de Lebesgue en dimension n est 2^n-doublante. La mesure de volume sur une variété riemannienne compacte est également doublante. La mesure de volume sur une variété riemannienne complète non bornée n'est pas forcément doublante, mais elle est doublante sur les boules. La mesure $\sum_{n\in\mathbb{N}} \delta_n$ sur $\mathbb{N}$ (ou $\mathbb{R}$) n'est pas doublante, ni doublante sur les boules : 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.

De ces différentes notions, celle de doublement est la plus précise :

Théorème II-74. *Soit μ une mesure régulière doublante sur les boules d'un espace métrique (X,d), et non identiquement égale à 0 ou $+\infty$. Alors*
- *pour tout $x \in X$ et $r > 0$ on a $0 < \mu[B_r(x)] < +\infty$;*
- *μ est de plein support ;*
- *la mesure μ ne peut avoir d'atome en dehors des points isolés de X.*

Démonstration. Il suffit de montrer ces propriétés pour la boule $B(x_0, R)$, avec R arbitrairement grand. On peut donc supposer que μ est C-doublante pour un certain $C > 0$. Supposons d'abord qu'il existe $x \in X$ et $r > 0$ avec $\mu[B_r(x)] = 0$. Alors $\mu[B_{2r}(x)] \le C\mu[B_r(x)] = 0$. 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 $\mu[B_\varepsilon(x)] \le \alpha + \delta$. 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 $d(x,z) \le d(x,y) + d(y,z) < r + r < \varepsilon$). Comme par ailleurs $B_r(y)$ ne contient pas x, on a

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

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

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

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

II-6. 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**[4], ou mesure de Carathéodory.

Ce théorème subtil commence par utiliser un outil simple et très efficace pour montrer que deux mesures coïncident sur une tribu entière, ou plus généralement qu'une propriété est vraie pour toute une tribu. C'est le **lemme de classe monotone** (déjà utilisé dans la preuve du Théorème II-62).

II-6.1. Lemme de classe monotone.

Définition II-75 (classe monotone). *On appelle classe monotone une famille $\mathcal{C}$ de parties d'un ensemble X, stable par limite croissante et par différence :*

$$\big[\forall k \in \mathbb{N},\ A_k \in \mathcal{C},\ A_k \subset A_{k+1}\big] \Longrightarrow \bigcup_{k\in\mathbb{N}} A_k \in \mathcal{C},$$

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

Remarque II-76. 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. L'énoncé suivant fournit une condition suffisante pour qu'il y ait identité.

Théorème II-77 (Lemme de classe monotone). *Soit $\mathcal{F}$ une famille de parties d'un ensemble X, stable par intersection finie. Soit $\mathcal{C} = \mathrm{CM}(\mathcal{F})$ 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

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

4. que certains auteurs appellent tout simplement “mesure” [Evans-Gariepy]

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

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 μ est une mesure, en général on ne sait pas en général calculer $\mu[A \cap B]$, ou $\mu[A \cup B]$, 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, en général on ne sait pas calculer $\mu[\cup A_k]$ en fonction des $\mu[A_k]$; mais si la suite (A_k) est croissante, alors $\mu[\cup A_k] = \lim \mu[A_k]$.

II-6.2. 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'élements 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

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

Le théorème suivant dit que cette condition nécessaire est également suffisante, pourvu qu'on lui adjoigne une autre condition de σ-finitude.

Théorème II-78 (théorème de prolongement de Carathéodory). *Soient X un ensemble, $\mathcal{A}$ une algèbre de parties de X, et μ une fonction positive sur $\mathcal{A}$, telle que*

(a) μ est σ-additive sur $\mathcal{A}$;

(b) 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 II-79. 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 $\mathcal{A}$,

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

Si $\mu[X] = +\infty$, il existe aussi une reformulation analogue, mais elle est un 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 $\lim_{k \to \infty} \mu[B \cap X_k] = \mu[B] = +\infty$. 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 $\mathcal{A}$,

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

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

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

Remarque II-80. On peut faire l'analogie avec le célèbre théorème de prolongement unique des fonctions continues. Ce théorème énonce que si f est une fonction uniformément continue, définie sur une partie quelconque E d'un espace complet Y, alors f se prolonge de façon unique en une fonction continue sur $\overline{E}$. En théorie de la mesure, ce qui joue le rôle de f c'est la fonction d'ensemble, ce qui joue le rôle de E c'est la famille $\mathcal{F}$, et ce qui joue le rôle de $\overline{E}$ c'est $\sigma(\mathcal{F})$. Montrer l'unicité est facile, c'est juste un passage à la limite dans le cas topologique, et une application du lemme de classe monotone dans le cas de la théorie de la mesure. Montrer l'existence est plus délicat, et requiert l'uniforme continuité dans le cas topologique, et un raisonnement bien plus complexe pour la théorie de la mesure.

Remarque II-81. Le théorème de Carathéodory, ou ses variantes, est le cœur de toute la théorie de Lebesgue. En général, quand on doit construire une mesure, on ne sait l'évaluer que sur certaines parties simples : ainsi pour calculer l'aire d'une figure dans le plan on cherche à l'approcher par une union de rectangles disjoints, dont on sait calculer l'aire. On veut ainsi passer à la limite dans le calcul de l'aire d'une union finie de rectangles. Mais comment montrer que ce calcul converge ? et qu'il converge quelle que soit le procédé d'approximation ? Le théorème de prolongement, en passant de la mesure définie sur l'algèbre des unions finies de pavés, à la σ-algèbre de toutes les limites d'unions finies de pavés, garantira la convergence et l'unicité de la limite. De la même façon que si une fonction F dans un espace métrique est continue mais qu'on ne sait la calculer que sur une partie dense, disons D, on pourra toujours poser $F(x) = \lim F(x_n)$, où $x_n \in D$ et $x_n \to x$, et la valeur sera bien définie et indépendante de la suite x_n choisie.

On peut trouver des preuves du Théorème II-78 dans diverses sources, par exemple [Bony, section 1.6], Gramain, section VI.1] ou [Dudley, Théorème 3.1.4]. Mais je vais plutôt adapter ces preuves pour démontrer un énoncé plus général, qui contient le Théorème II-78 comme cas particulier. L'intérêt propre de l'énoncé généralisé apparaîtra par la suite.

Théorème II-82 (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 μ^* le prolongement de μ défini ainsi : pour toute partie A de X,

$$\mu^*[A] := \inf \left\{ \sum_{k=1}^{\infty} \mu[A_k];\ A_k \in \mathcal{F};\ A \subset \cup A_k \right\}. \tag{7}$$

On suppose que

$$\forall A, B \in \mathcal{F}, \quad \mu[A \cap B] + \mu^*[A \setminus B] = \mu[A]. \tag{8}$$

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

$$\mathcal{M} := \big\{ A \subset X;\ \forall B \subset X,\ \mu^*[B \cap A] + \mu^*[B \setminus A] = \mu^*[B] \big\}. \tag{9}$$

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* (8) *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*

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

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

Définition II-83 (mesure extérieure). *La fonction μ^* apparaissant dans* (7) *est appelée mesure extérieure associée à μ (et à la famille $\mathcal{F}$).*

Remarques II-84. (i) Bien noter que $\mu^*[A]$ 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,

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

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) L'intuition est la suivante : la mesure extérieure cherche à mesurer un ensemble A en l'approchant "par l'extérieur", et en se ramenant à la famille de référence $\mathcal{F}$. Si la frontière de A n'est pas trop affreuse, on devrait retomber sur la valeur attendue pour $\mu[A]$. En outre, si B est un autre ensemble quelconque, mesurable ou pas, la mesure $\mu^*[B]$ ne devrait pas changer si l'on coupe B selon $B\cap A$ et $B\setminus A$. (B n'est pas forcément mesurable, mais comme A l'est, découper B selon A, en plus des autres découpages envisagés, ne perturbera pas l'évaluation de la mesure extérieure de B.)

IMAGE

(iv) Dans le cas où X est de masse totale finie ($\mu[X] < +\infty$), on pourrait se contenter de définir les parties mesurables A (la tribu $\mathcal{M}$) 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 $X \setminus A$. Lebesgue utilisait déjà cette construction.

Voici maintenant quelques remarques sur l'énoncé du Théorème II-82, qui est le cœur même de la théorie de Lebesgue.

Remarques II-85. (i) Sans l'hypothèse de "σ-finitude" faite en (i) au Théorème II-82, 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].

(ii) Une partie $\mathcal{F}$ qui vérifie les hypothèses de la partie (iii) du Théorème II-82, 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 II-78.

Démonstration du Théorème II-82. Si la preuve de la partie (ii) est subtile, la partie (i) en revanche est une application simple du Lemme de classe monotone.

Commençons par démontrer (i) sous l'hypothèse $\mu[X] < +\infty$. Le Lemme de classe monotone (Théorème II-77) 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 $\widetilde{\mu}$ deux prolongements possibles. On pose

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

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

$$A \subset B \Longrightarrow \mu[B \setminus A] = \mu[B] - \mu[A], \qquad \textit{etc.}, \tag{10}$$

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 (10), 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 $\widetilde{\mu}$ deux prolongements possibles ; par le raisonnement précédent on sait que $\mu[X_k \cap A] = \widetilde{\mu}[X_k \cap A]$ 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 $k \to \infty$ et obtenir $\mu[A] = \widetilde{\mu}[A]$, ce qui conclut la preuve de (i).

Attelons-nous maintenant à la démonstration de (ii). Comme dans l'énoncé, on définit

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

Montrons 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 (8). On divise la preuve en sept étapes.

1. <u>μ^* *est croissante*</u>. 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. <u>μ^* *est dénombrablement sous-additive*</u>. 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

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

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

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

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. <u>$\mathcal{M}$ *est une algèbre*</u>. 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

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

$$= \mu^*[B \cap A_1] + \mu^*[B \setminus A_1] = \mu^*[B],$$

ce qui conclut l'argument.

4. <u>μ^* *est additive sur* $\mathcal{M}$</u>. 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 $n \to \infty$, et en utilisant la sous-additivité, on trouve

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

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. *$(X, \mathcal{M}, \mu^*)$ 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 (8) va intervenir.

Posons $B = \emptyset$ dans (8), 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 (8), on trouve

$$\mu^*[B \cap A] + \mu^*[B \setminus A] \leq \sum_k \mu[A_k \cap A] + \mu^*[B \setminus A]$$
$$\leq \sum_k \big(\mu[A_k \cap A] + \mu^*[A_k \setminus A]\big) = \sum_k \mu[A_k].$$

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 II-82. Si le critère (8) 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 (8). On procède en trois étapes.

1. $\underline{\mu \text{ est croissante sur } \mathcal{F}}$. 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. $\underline{\mu \text{ coïncide avec } \mu^* \text{ sur } \mathcal{F}}$. 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

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

Pour chaque j, les B_{j,i_j} sont disjoints ; cela entraîne que pour deux choix différents du multi-indice $(i_1, \ldots, i_j)$, les parties $A \cap A_j \cap B_{1,i_1} \cap \ldots \cap B_{j-1,i_{j-1}}$ 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

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

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 (8) 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 (8), 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 (8) et du Théorème II-82. □

Exercice II-86. Adapter, en la simplifiant, la preuve du Théorème II-82 pour démontrer directement le cas particulier du Théorème II-78.

Je vais présenter dès à présent deux applications importantes du Théorème de prolongement de Carathéodory. Dans l'immédiat, je n'en fournirai donnera que des preuves partielles ; les preuves complètes viendront plus tard dans le cours.

II-6.3. Produits infinis. Le théorème suivant permet de construire des mesures sur des produits infinis :

Théorème II-87 (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* $A_1, \ldots, A_m$ *de parties mesurables de* $X_1, \ldots, X_m$ *respectivement, on pose*

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

On définit alors

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

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* $C(A_1, \ldots, A_m)$*,* $m \in \mathbb{N}$*.*

Remarque II-88. 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 $\mu_k[X_k]$.

Dans le cas où les X_k sont des ensembles finis (au sens de : ensembles de cardinal fini, et pas : ensembles de mesure finie), la démonstration du Théorème II-87 est très simple et je vais la présenter tout de suite. Le cas général [Dudley, p. 257-259] est plus subtil, et nécessitera d'être plus aguerri : Cf. Théorème IV-104 au Chapitre IV.

Démonstration du Théorème II-87 pour des espaces de cardinal fini. Sans perte de généralité, on suppose que

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

et chaque X_k est muni de la tribu triviale $\mathcal{P}(X_k)$. Les cylindres sont de la forme $C = A_1 \times A_2 \times \ldots \times A_k \times X_{k+1} \times X_{k+2} \times \ldots$; on définit alors $\mu^\infty[C] = \prod_{j=1}^{k} \mu_k[A_k]$. Il est facile de vérifier que les cylindres forment une algèbre, et que la fonction μ^∞ est additive sur cette algèbre. En outre, $\mu^\infty[X] = 1$. Pour appliquer le Théorème II-78 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. Soit donc $(C_n)_{n\in\mathbb{N}}$ une famille de cylindres disjoints, dont l'union est un cylindre $C = \prod A_k$. Par le Lemme II-89 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^∞. □

Lemme II-89 (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* $\bigcup C_n = \prod X_k$*. Alors il n'y a qu'un nombre fini de* C_n *non vides.*

Je vais fournir deux démonstrations de ce lemme. La première est simple et élémentaire, le principe rappelle un peu celui qui permettra dans la suite de démontrer le Théorème II-87 dans un cadre général. La deuxième, plus compacte, illustrera l'intérêt d'un raisonnement topologique, et préparera la voie à la démonstration du Théorème II-90 ci-dessous.

Première démonstration. Chacun des cylindres C_n est union disjointe de “cylindres élémentaires”, de la forme $(a_1, a_2, \ldots, a_K) \times \prod_{k \geq K+1} X_k$; on dit qu'un tel cylindre est d'ordre K et a pour base $\{a_1, \ldots, a_K\}$. 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 $Y_1 = \{u_1, \ldots, u_\ell\}$ 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 $j \in \mathbb{N}$, $\{y_1, y_2, \ldots, y_j\}$ 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 du Théorème II-37 où l'on considère un produit dénombrable). 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. $\square$

II-6.4. 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 II-90 (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 $F = \{t_1, \ldots, t_K\} \subset T$, on définit $X_F := X_{t_1} \times \ldots \times X_{t_K}$; et pour tout borélien A_F de X_F, on définit le cylindre

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

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 II-91. (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 $t' \in T' \backslash T$. 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 $\mathbb{N}$ (temps discret) 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) Ce théorème sera démontré au Chapitre IV (Théorème IV-119) après un passage en revue des propriétés principales de l'intégration produit.

(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.

II-6.5. 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 II-92 (Critère de Carathéodory). *Soit (X, d) un espace métrique, et soit μ^* une mesure extérieure sur X, au sens de la Remarque II-84 (i) ; on définit la tribu $\mathcal{M}$ comme dans l'énoncé du Théorème II-82. Si, pour toutes parties A et B de X telles que*

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

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

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

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

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

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. □

II-7. Complétion de mesures

Le prolongement d'une mesure est une opération délicate mais inoffensive ; au contraire, la complétion est une opération très simple mais risquée.

Théorème II-93 (complétion d'une mesure). *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et soit $\overline{\mathcal{A}}$ la famille des parties E de X qui s'écrivent $A \cup N$, où $A \in \mathcal{A}$ et N est μ-négligeable, c'est à dire inclus dans une partie $Z \in \mathcal{A}$, telle que $\mu[Z] = 0$. 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.*

Définition II-94 (ensembles μ-mesurables). *On dit que la tribu $\overline{\mathcal{A}}$ apparaissant dans le Théorème II-93 est la tribu des ensembles (ou parties) μ-mesurables.*

Remarques II-95. (i) De façon équivalente, les ensembles μ-mesurables sont les parties E telles qu'il existe A et A' dans $\mathcal{A}$, telles que $A \subset E \subset A'$ et $\mu[A' \setminus A] = 0$. Ou encore, ce sont les parties E telles qu'il existe B dans $\mathcal{A}$ et N négligeable, telles que $E = B \setminus N$.

(ii) Cette notion dépend bien sûr de μ, mais aussi de $\mathcal{A}$. Souvent, par défaut $\mathcal{A}$ sera la tribu borélienne d'un espace topologique.

Preuve du Théorème II-93. Avec les notations de l'énoncé, il suffit de poser $\overline{\mu}[E] = \mu[A]$, et de vérifier les axiomes de σ-additivité (exercice). La seule subtilité consiste en fait à montrer que cette définition est licite, c'est à dire qu'elle ne dépend pas du choix de A et N. Pour cela, on pourra remarquer que (avec des notations évidentes) si $E = A_1 \cup N_1 = A_2 \cup N_2$, alors $A_1 \setminus A_2 \subset E \setminus A_2 \subset N_2 \subset Z_2$ est de mesure nulle, donc $\mu[A_1] = \mu[(A_1 \cap A_2)] + \mu[A_1 \setminus A_2] = \mu[A_1 \cap A_2] \leq \mu[A_2]$; et par symétrie, $\mu[A_1] = \mu[A_2]$. □

Remarque II-96. La simplicité de l'énoncé et de sa preuve masque le fait que les ensembles ainsi complétés peuvent être extrêmement compliqués. L'étude de l'intégration sur des espaces produits montrera bien que l'opération de complétude n'est pas inoffensive.

Le Théorème II-82 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 II-93 car une mesure admet en général plusieurs prolongements complets. Il y a cependant unicité quand on impose certaines condition de régularité.

Théorème II-97 (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 II-93 est l'unique prolongement complet de $(X, \mathcal{A}, \mu)$ en un espace mesuré complet et régulier.*

Démonstration. Soit $(X, \widetilde{\mathcal{A}}, \widetilde{\mu})$ un prolongement complet régulier de $(X, \mathcal{A}, \mu)$. Comme X est σ-fini, tout élément A de $\widetilde{\mathcal{A}}$ peut s'écrire comme union dénombrable d'éléments A_k de $\widetilde{\mathcal{A}}$ avec $\widetilde{\mu}[A_k] < +\infty$. Par la Proposition II-57, pour chaque A_k il existe des Boréliens B_k et C_k tels que $\widetilde{\mu}[B_k] = \widetilde{\mu}[A_k] = \widetilde{\mu}[C_k]$, 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 $\widetilde{\mu}$-négligeable, ce qui montre que $\widetilde{\mathcal{A}}$ contient $\overline{\mathcal{A}}$. On conclut que les deux tribus sont égales, et la fin de la démonstration en découle aisément. □

En conclusion : tant que l'on travaille avec des mesures régulières, il n'y a pas de difficulté à définir *la* complétion.

II-8. Construction de la mesure de Lebesgue

Le théorème suivant est une application du théorème de prolongement de Carathéodory ; c'est aussi l'acte fondateur de la théorie de Lebesgue. On munira bien sûr $\mathbb{R}$ de sa topologie habituelle.

Théorème II-98 (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 $\overline{\lambda}$ 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*

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

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

Démonstration du Théorème II-98. La famille des intervalles est stable par intersection finie (l'intersection de deux intervalle 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 II-82(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 II-82 (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*

$$\sum_{k\in\mathbb{N}} |I_k| = |I|. \tag{11}$$

Admettons provisoirement ce résultat ; on peut alors appliquer le Théorème II-82 (iii) pour construire la mesure de Lebesgue via le concept de mesure extérieure de Carathéodory. Les propriétés de la complétion resultent alors des Théorèmes II-93 et II-97. □

Démonstration de (11). Si l'on sait prouver (11) 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

$$|I| = \sum_\ell |I^\ell|, \qquad |I_k| = \sum_\ell |I_k^\ell|.$$

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 (11) 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,

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

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 $A = I_1 \cup \ldots \cup I_n$, alors $|A| \leq |I_1| + \ldots + |I_n|$, 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 $[0, 1] \subset J_1 \cup J_2 \cup \ldots \cup J_K$. En particulier,

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

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

II-9*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 ; c'est le cœur de la procédure de "désintégration" étudiée dans le Chapitre **??**. Dans cette optique, il est 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 II-99 (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 $\widetilde{\mathcal{B}}$ de boules disjointes telles que*

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

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

$$\mu\Big[\bigcup_{B\in\widetilde{\mathcal{B}}} B\Big] \geq C^{-2}\,\mu\Big[\bigcup_{B\in\mathcal{B}} B\Big].$$

Avant d'aborder la preuve du Théorème II-99 proprement dite, commençons par deux lemmes très simples :

Lemme II-100. *Soient B et B' deux boules fermées d'un espace métrique, de rayons respectifs r et r', telles que*

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

Alors, avec les notations du Théorème II-99, on a $B \subset 4B'$.

Lemme II-101. *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.*

Remarque II-102. Cet énoncé de théorie des ensembles relève de l'axiome du choix (ou de manière équivalente, du principe de maximalité de Hausdorff) ; mais comme on va le voir, l'hypothèse de séparabilité permet d'éviter le recours à l'axiome du choix général, pour ne garder que l'axiome du choix dépendant – même si la famille $\mathcal{B}$ est non dénombrable.

Preuve du Lemme II-100. On écrit $B = B_{r]}(x)$, $B' = B_{r']}(x')$. 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 !) □

Preuve du Lemme II-101. 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 $\widetilde{\mathcal{B}}_1 = \{B_1\}$. Dans le cas contraire, on pose $\widetilde{\mathcal{B}}_1 = \emptyset$.

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

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

Soit $\mathcal{M} = \{B_{i_1}, B_{i_2}, \ldots, B_{i_j}, \ldots\}$ 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 $\widetilde{\mathcal{B}}_{\ell-1}$, 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 $\widetilde{\mathcal{B}}_{m-1}$.

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

Démonstration du Théorème II-99. 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 : $r(B_1) \geq r(B_2) \geq r(B_3) \geq \ldots$. On construit alors l'ensemble $\widetilde{\mathcal{B}}$ selon la même procédure que précédemment : au début on pose $\widetilde{\mathcal{B}}_1 = \{B_1\}$, puis si B_2 n'intersecte pas B_1 on pose $\widetilde{\mathcal{B}}_2 = \{B_1, B_2\}$, sinon on pose $\widetilde{\mathcal{B}}_2 = \{B_1\}$; et ainsi de suite. À l'étape k, si B_k n'intersecte aucun élément de $\widetilde{\mathcal{B}}_{k-1}$ on pose $\widetilde{\mathcal{B}}_k = \widetilde{\mathcal{B}}_{k-1} \cup \{B_k\}$, et sinon on pose $\widetilde{\mathcal{B}}_k = \widetilde{\mathcal{B}}_{k-1}$.

On définit alors $\widetilde{\mathcal{B}} = \cup \widetilde{\mathcal{B}}_k$. Soit $B = B_k$ un élément quelconque de $\mathcal{B}$. S'il n'est pas dans $\widetilde{\mathcal{B}}$, c'est qu'il intersecte une boule B_j de $\widetilde{\mathcal{B}}_{k-1}$ 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 II-100) $r' \geq r$; d'où $B \subset 3B'$ par un argument similaire à celui du Lemme II-100.

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 faut donc 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 $j \in \mathbb{N}$, soit

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

On choisit grâce au Lemme II-101 une famille dénombrable disjointe maximale $\widetilde{\mathcal{M}}_1$ dans $\mathcal{B}_1$.

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

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

On pose enfin $\widetilde{\mathcal{B}} = \cup \mathcal{M}_k$. 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 $B' \in \widetilde{\mathcal{B}}$.

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

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

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 II-100 pour conclure que $B \subset 4B'$. Ceci conclut la preuve de (i).

Passons maintenant à (ii) : Pour cela on écrit

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

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 II-103. *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*

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

Exemple II-104. On verra 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 II-103. 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 II-99, il existe une famille dénombrable disjointe $\widetilde{\mathcal{B}} \subset \mathcal{B}$ telle que

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

d'où

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

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

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

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 $r_k \to \infty$ ($k \in \mathbb{N}$) 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. □

Remarque II-105. 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. En voici deux parmi les plus intéressants :

• le **Lemme de recouvrement de Besicovitch** [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* $\mathcal{B}_1, \ldots \mathcal{B}_K$ *de* $\mathcal{B}$*, qui recouvrent l'ensemble* C*. En particulier, de* $\mathcal{B}$ *on peut extraire un sous-recouvrement* $\widetilde{\mathcal{B}}$ *tel que chaque* $x \in C$ *appartient à au moins une, et au plus* K *boules de* $\mathcal{B}$*.*

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 II-103 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, 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, et dont les diamètres sont à peu près proportionnels à leur distance au bord de O :

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

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 III

Intégration selon Lebesgue et selon Riesz

Ce chapitre définit l'intégrale de Lebesgue pour une large classe de fonctions, dites sommables au sens de Lebesgue ; ce cadre abstrait inclut les fonctions continues à support compact comme cas particulier. Le point de départ sera la notion de fonction mesurable (section III-1) ; de là découlera la définition de l'intégrale de Lebesgue (section III-2), et l'on vérifiera qu'elle constitue une forme linéaire (section III-3).

Une autre route consiste à partir des fonctions continues et à définir l'intégrale comme une forme linéaire continue sur cet espace. Le **théorème de Riesz** (section III-4) assure que ces deux points de vue sont équivalents, modulo quelques subtilités, sous certaines hypothèses topologiques. La plus importante de ces hypothèses est la condition de **compacité locale**, qui est vraie en dimension finie (dans $\mathbb{R}^n$ ou une autre variété différentielle complète de dimension finie), mais en général fausse en dimension infinie, comme dans l'espace de Wiener.

Ce chapitre se conclut par quelques mots sur l'intégration à valeurs vectorielles (section III-5), qui sera abordée plus en détail dans un chapitre ultérieur.

III-1. Fonctions mesurables

On cherche à définir une large classe de fonctions susceptibles d'être intégrées, c'est à dire évaluées par l'action d'une mesure : les **fonctions mesurables**. Cette classe de fonctions devrait comprendre à tout le moins les fonctions indicatrices d'ensembles mesurables. Dans un cadre topologique, elle devrait aussi comprendre les fonctions continues.

En topologie, où les parties ouvertes jouent un rôle privilégié, on dit qu'une fonction 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.

III-1.1. Définition.

Définition III-1 (fonctions mesurables). *Soient $(X.\mathcal{A})$ et $(Y,\mathcal{B})$ deux espaces mesurables. Une fonction $f : X \to Y$ est dite mesurable si l'image réciproque de n'importe quelle partie. mesurable de Y est une partie mesurable de X :*

$$\forall B \in \mathcal{B}, \qquad f^{-1}(B) \in \mathcal{A}.$$

Exemple III-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. Cela est vrai quelle que soit la tribu dont on munit l'espace d'arrivée $\mathbb{R}$. En effet, 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, si l'on travaille avec la tribu borélienne, la mesurabilité est la règle plutôt que l'exception (Remarque III-26), et la mesurabilité est beaucoup, beaucoup plus générale que la continuité.

Remarques III-3. (i) Soient $(X, \mathcal{A})$ un espace mesurable, Y un ensemble quelconque, et $f : X \to Y$. 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 de Y comme une partie dont l'image réciproque est mesurable. Cela définit bien une tribu sur Y (exercice). Cette tribu est **la plus grande** qui rende f mesurable : toute tribu plus petite a la même propriété, et aucune tribu strictement plus grande ne l'a. On l'appelle **tribu image** de $\mathcal{A}$ par f et on la note $f_\#\mathcal{A}$ (ou $f_*\mathcal{A}$, ou $f\mathcal{A}$).

(ii) Soient maintenant X un ensemble quelconque, $(Y, \mathcal{B})$ un espace mesurable, et $f : X \to Y$ 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, et aucune tribu strictement plus petite ne l'a (exercice). On l'appelle **tribu engendrée** par f et on la note $\sigma(f)$ (ou $f^*\mathcal{B}$ ou $f^{-1}\mathcal{B}$) ; on reviendra sur cette notion majeure. Intuitivement, $\sigma(f)$ est la tribu des ensembles que l'on peut décrire “au moyen seulement de la fonction f”, ou encore “au travers des seules valeurs de la fonction f”. En particulier, $\sigma(\mathrm{Id}) = \mathcal{B}$ (la tribu engendrée par la fonction identité est la tribu $\mathcal{B}$ tout entière).

(iii) Plus généralement, si une famille de fonctions $(f_i)_{i\in I}$ est donnée, $f_i : X \to Y_i$, on note $\sigma((f_i)_{i\in I})$ la plus petite tribu qui rende mesurable toutes les f_i, ou de façon équivalente, qui rende mesurable l'application produit $f = \prod_{i\in I} f_i$, avec la tribu produit sur $\prod Y_i$. Intuitivement, par exemple $\sigma(f, g)$ (abréviation pour $\sigma((f, g))$) est la tribu des événements que l'on peut décrire en utilisant seulement les valeurs de f et g.

(iv) 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 *a priori* pas mesurable. C'est cependant le cas si $f : X \to Y$ 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. En particulier, la réciproque d'une bijection mesurable entre espaces polonais est automatiquement mesurable (Théorème III-24 plus loin).

Le critère pratique qui suit est d'usage constant.

Proposition III-4 (critère pratique de mesurabilité). *Soient $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ deux espaces mesurables, et $f : X \to Y$. 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. Si f est mesurable, bien sûr $f^{-1}(F)$ doit être mesurable pour tout $F \in \mathcal{F}$. Réciproquement, supposons que $f^{-1}(F)$ est mesurable pour tout $F \in \mathcal{F}$, et définissons

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

On a déjà noté dans la Remarque III-3(ii) que $\mathcal{C}$ est une tribu ; cela découle en fait des formules habituelles

$$f^{-1}(Y \setminus B) = f^{-1}(Y) \setminus f^{-1}(B), \qquad f^{-1}\Big(\bigcup B_k\Big) = \bigcup f^{-1}(B_k).$$

Mais comme elle contient $\mathcal{F}$, elle contient également $\sigma(\mathcal{F}) = \mathcal{B}$, ce qui prouve la mesurabilité. □

Exemples III-5. (i) L'exemple le plus courant est le suivant : si X et Y sont deux espaces topologiques, munis de leur tribu borélienne, et $f : X \to Y$ 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 $f : X \to Y$ 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 $Y = \mathbb{R}$, pour montrer la mesurabilité 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 fontion 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, pour montrer la mesurabilité il suffit de vérifier que l'image réciproque de tout pavé est mesurable. Et si Y est 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 les liens entre mesurabilité et continuité sont étroits. D'une part, 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, qui seront évoqués section III-4.4, en sont une bonne illustration. Ensuite, l'exercice suivant montre que sous des hypothèses assez générales, toute fonction borélienne peut être considérée comme une fonction continue pour une topologie plus fine, sans toucher aux boréliens. Il est donc naturel que ces deux classes de fonctions, les boréliennes et les continues, vérifient des énoncés assez similaires.

Exercice III-6 (Une fonction borélienne est toujours continue en un sens). Soit X un espace polonais.

(a) On dit qu'une partie A de X vérifie la propriété (B) si il existe une topologie polonaise qui raffine la topologie de X tout en préservant la classe des boréliens, et pour laquelle A est ouvert. Montrer que l'ensemble des parties vérifiant (B) est une σ-algèbre contenant les ensembles ouverts. (Noter que dans la propriété (B) on peut même imposer que A soit ouvert et fermé.)

(b) En déduire que tous les boréliens vérifient la propriété (B).

(c) Si Y est un espace polonais et $f : X \to Y$ est borélienne, montrer que tous les $f^{-1}(B)$, pour B borélien de Y, vérifient la propriété (B).

(d) Montrer qu'il existe une topologie polonaise sur X, plus fine que la topologie initiale mais avec les mêmes ensembles boréliens, et qui rend f continue.

III-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 III-7 (restriction). *(i) Soient $(X,\mathcal{A})$ et $(Y,\mathcal{B})$ deux espaces mesurables, et A une partie mesurable de X. Soit $f : X \to Y$ 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.*

(ii) La même propriété reste vraie si A n'est pas mesurable, la tribu induite par $\mathcal{A}$ étant alors l'ensemble de toutes les intersections d'éléments de $\mathcal{A}$ avec A.

Démonstration. C'est un simple jeu de maniement des axiomes. □

Proposition III-8 (recollement). *(i) Dans un espace mesurable $(X,\mathcal{A})$, soit $(A_k)_{k\in\mathbb{N}}$ une famille dénombrable de parties mesurables disjointes, telle que $X = \cup A_k$. Soit également $(Y,\mathcal{B})$ un espace mesurable. Sur chaque A_k (considéré comme espace mesurable), on se donne une fonction mesurable $f_k : A_k \to Y$. Soit f la fonction qui pour tout k coïncide avec f_k sur A_k. Alors f est mesurable.*

(ii) Dans un espace mesurable $(X,\mathcal{A})$, soit A une partie quelconque de X, f est une fonction mesurable de A (muni de la tribu induite) dans Y, espace polonais, alors f est la restriction à A d'une fonction mesurable $X \to Y$.

Démonstration. Pour (i), 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.

Pour (ii), on procède par approximation ; l'argument est remis à la section III-2.2. □

Proposition III-9 (produit infini de fonctions mesurables). *Soient $(X_t,\mathcal{A}_t)_{t\in T}$ et $(Y_t,\mathcal{B})_{t\in T}$ des espaces mesurables, 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 $f_1,\ldots,f_k$ sont des applications mesurables définies sur $X_1,\ldots,X_k$ respectivement, à valeurs dans $Y_1,\ldots,Y_k$ respectivement, alors l'application $f = (f_1,\ldots,f_k)$ est mesurable de $X_1\times\ldots\times X_k$ dans $Y_1\times\ldots\times Y_k$.

Démonstration. Par le critère III-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 $B_T = (B_{t_1},\ldots,B_{t_K})$, $T = (t_1,\ldots,t_K)$:

$$C_T(B_T) = \Big\{y\in Y;\ \forall k\in\{1,\ldots,K\},\ y_{t_k}\in B_{t_k}\Big\}.$$

Ici K est arbitraire et chaque B_{t_k} est une partie mesurable de Y_{t_k}. Alors

$$f^{-1}(C_T(B_T) = C_T(A_T),$$

où

$$A_T = (A_{t_1},\ldots,A_{t_K}),\qquad A_{t_k} = f_k^{-1}(B_{t_k}).$$

C'est en particulier un ensemble mesurable. □

Proposition III-10 (composition). *Soient $f : X \to Y$ et $g : Y \to Z$ deux applications mesurables entre espaces mesurables, alors leur composition $g \circ f$ est mesurable.*

Démonstration. C'est une conséquence immédiate de la définition. □

Corollaire III-11. *Soient X, Y et Z des espaces mesurables, tels que Y et Z sont des espaces topologiques munis de leur tribu borélienne, soient $f : X \to Y$ une application mesurable, et $\varphi : Y \to Z$ une application continue. Alors $\varphi \circ f : X \to Z$ est une application mesurable.*

Démonstration. C'est une conséquence de la proposition précédente, combinée avec l'exemple III-5(i). □

Proposition III-12 (graphe). *Soient $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ des ensembles mesurables, $f : X \to Y$ une fonction, et $G(f) = \{(x, f(x));\ x \in X\} \subset X \times Y$ son graphe. On munit $X \times Y$ de la tribu produit. Alors, si f est mesurable et que la diagonale $\Delta = \{(x, x);\ x \in X\}$ est mesurable, alors $G(f)$ est mesurable.*

Démonstration. Le graphe de f est l'image réciproque de la diagonale par l'application mesurable $(x, y) \longmapsto (f(x), y)$. □

En anticipant sur la proposition IV-39, on en déduit

Corollaire III-13. *Si X et Y sont des espaces métriques séparables complets, et que $f : X \to Y$ est mesurable, alors son graphe est mesurable dans $X \times Y$. Cela reste vrai si f n'est définie que sur une partie mesurable de X.*

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 III-14 (opérations élémentaires). *Soient $(X, \mathcal{A})$ un espace mesurable, 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 III-11 avec les applications continues addition, soustraction, etc. Noter que $\min(f, g) = (f + g)/2 - |f - g|/2$. □

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 $]a, +\infty[$ et $]-\infty, a[$ 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 ; en pratique ce sont juste les boréliens de $\mathbb{R}$, auxquels on s'autorise d'ajouter $+\infty$ et/ou $-\infty$ (Exemple II-17). La motivation pour considérer $\overline{\mathbb{R}}$ plutôt que $\mathbb{R}$ est simple : la convergence dans $\overline{\mathbb{R}}$ est plus aisée que dans $\mathbb{R}$.

Remarque III-15. Attention aux opérations dans $\overline{\mathbb{R}}$: l'opération $(x,y) \to x+y$ 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 l'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 III-16 (stabilité de la mesurabilité réelle par limite). *Soit* $(f_n)_{n\in\mathbb{N}}$ *une famille de fonctions mesurables sur un espace mesurable* $(X,\mathcal{A})$*, à valeurs dans* $\overline{\mathbb{R}}$*. Alors*

(i) les fonctions $\inf f_n$, $\sup f_n$, $\liminf f_n$ *et* $\limsup f_n$ *sont mesurables ;*

(ii) la fonction $\lim f_n$ *est mesurable sur son domaine, c'est à dire sur l'ensemble mesurable* C *où la suite de fonctions* (f_n) *converge simplement ;*

(iii) si g *est une fonction mesurable quelconque de* X *dans* $\mathbb{R}$ *(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* $\overline{\mathbb{R}}$*.*

Remarque III-17. 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 $\inf f_n$ est mesurable. On ne va pas travailler avec les boréliens directement, mais on va choisir une famille génératrice commode ; par exemple, les intervalles $[\alpha, +\infty]$ (Cf Exemple II-17). Soit donc $A_\alpha = \{\inf f_n \geq \alpha\}$, avec $\alpha \in \mathbb{R}$, le but est de montrer que A_α est mesurable. Or $A_\alpha = \{x; \forall n \in \mathbb{N},\ f_n(x) \geq \alpha\}$, donc

$$A_\alpha = \cap_{n\in\mathbb{N}} f_n^{-1}\big([\alpha, +\infty]\big),$$

qui est mesurable comme intersection d'ensembles mesurables ; cela achève la preuve de la mesurabilité de $\inf f_n$. La mesurabilité de $\sup f_n$ s'en déduit, puisque $\sup f_n = -\inf(-f_n)$. (Exercice : Refaire la preuve avec les $]\alpha, +\infty]$; puis avec les $[-\infty, \alpha]$; puis avec les $[-\infty, \alpha[$.)

Montrons ensuite, par exemple, que $\limsup f_n$ est mesurable. Idem, il suffit de montrer que $B_\alpha = \{\limsup f_n \geq \alpha\}$ est mesurable, pour tout $\alpha \in \mathbb{R}$. Or dire que x appartient à B_α, c'est dire que pour tout $\varepsilon > 0$, $f_n(x)$ prend une infinité de fois une valeur supérieure ou égale à $\alpha - \varepsilon$. Les $\varepsilon > 0$ forment une famille non dénombrable, mais on ne perd rien en la renplaçant par la famille dénombrable des $1/k$: C_α est l'ensemble des x tels que :pour tout k, pour tout n il existe $m \geq n$ tel que $f_m(x) \geq \alpha - 1/k$. Autrement dit,

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

donc C_α est bien mesurable, ce qui conclut la démonstration du point (i). (Exercice : Refaire la preuve avec les $]\alpha, +\infty]$; puis avec les $[-\infty, \alpha]$; puis avec les $[-\infty, \alpha[$.)

(ii) L'ensemble C des points de convergence est l'union de trois ensembles qui d'après (i) sont tous mesurables :

- l'ensemble où $f_n \to +\infty$, i.e. $\liminf f_n = +\infty$;
- l'ensemble où $f_n \to -\infty$, 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 (ii) grâce à la Proposition III-7.

(iii) découle de (ii) et de la Proposition III-8. □

Remarque III-18. Pour traiter $\pm\infty$, on aurait aussi pu dire que $\{f_n \to +\infty\} = \{\liminf f_n = +\infty\} \cap \{\liminf f_n = +\infty\}$ est mesurable comme intersection de mesurables. Mais en tout cas on ne peut pas traiter de façon unifiée $\pm\infty$ et $\mathbb{R}$, car l'expression $\limsup f_n - \liminf f_n$ n'est pas définie si les deux sont égales à $+\infty$ ou $-\infty$; et on ne peut pas non plus passer par $C = \cup_{\alpha\in\mathbb{R}}\{\liminf f_n = \alpha\} \cap \{\limsup f_n = \alpha\}$ car la somme n'est pas dénombrable !

Exemples III-19. (i) Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction dérivable, alors sa dérivée est mesurable. En effet, c'est la limite simple de la suite de fonctions continues

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

Bien noter que la dérivée n'est pas forcément continue.

(ii) Une série de Fourier définit une fonction mesurable sur son domaine de convergence ; de même pour une série entière. Rappelons que de telles séries peuvent ne pas converger, et même quand elles convergent, leur valeur limite n'est pas forcément continue.

(iii) L'intégrale de Riemann étant définie par un procédé de limite, les fonctions définies comme des intégrales de Riemann à paramètre sont mesurables. On verra au Chapitre IV que cet énoncé se généralise à des intégrales à paramètre définies dans la théorie de Lebesgue.

Exercice III-20. Redémontrer les points (ii) et (iii) du Théorème III-16 quand les f_n sont à valeurs dans $\mathbb{R}$, sans faire référence à des lim inf et lim sup, via le critère suivant : *une suite de nombres réels converge si et seulement si elle est de Cauchy.* En déduire que les points (ii) et (iii) du Théorème III-16 restent vrais quand l'espace d'arrivée $\overline{\mathbb{R}}$ est remplacé par un espace métrique complet arbitraire (Commencer par prouver la mesurabilité de l'ensemble des points x pour lesquels la suite $(f_n(x))$ est de Cauchy.)

En s'insipirant de l'exercice III-20 on pourra démontrer la généralisation suivante :

Théorème III-21 (stabilité de la mesurabilité par limite, cas général). *Soit $(f_n)_{n\in\mathbb{N}}$ une famille de fonctions mesurables sur un espace mesurable $(X, \mathcal{A})$, à valeurs dans un espace polonais Y muni de sa tribu borélienne. Alors la fonction* $\lim f_n$ *(éventuellement étendue arbitrairement hors du domaine de convergence) est mesurable. Plus précisément,*

(i) l'ensemble C des points de convergence de (f_n) est mesurable ;

(ii) la fonction $\lim f_n$ *est mesurable de C dans Y ;*

(iii) si g est une fonction mesurable quelconque de X dans Y, 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 Y.

Pour conclure cette section, voici, sans démonstration complète, deux résultats frappants sur les isomorphismes mesurables, c'est à dire les applications à la fois mesurables et inversibles.

Définition III-22 (isomorphisme mesurable). *Soient $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ deux espaces mesurables. On dit que f réalise un isomorphisme mesurable entre X et Y si f est bijective et bimesurable, c'est à dire mesurable et d'inverse mesurable, $f_\#\mathcal{A} = \mathcal{B}$ et $(f^{-1})_\#\mathcal{B} = \mathcal{A}$.*

Quand deux espaces mesurables sont mesurablement isomorphes, tout énoncé de mesurabilité prouvé dans l'un s'appliquera aussi à l'autre.

Théorème III-23 (théorème d'isomorphisme dans les espaces polonais). *Soient $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ deux espaces polonais munis de leurs tribus boréliennes respectives, et soient A un borélien de X, B un borélien de Y. Alors A et B sont mesurablement isomorphes, si et seulement si ils ont même cardinalité.*

Théorème III-24 (théorème de l'inverse mesurable). *Soient X et Y deux espaces polonais, munis de leurs tribus boréliennes respectives, A et B des boréliens de X et Y respectivement, et f une bijection mesurable de A dans B. Alors f^{-1} est mesurable.*

Remarques III-25. (i) Les deux théorèmes précédents sont liés : le premier dit que s'il existe une bijection quelconque entre A et B, alors il existe aussi une bijection bimesurable ; le second (parfois appelé Théorème de Kuratowski) dit que s'il existe une bijection mesurable entre A et B, alors automatiquement elle est bimesurable.

(ii) Il n'y a en fait que trois cas de figure : Soit A et B sont finis, soit ils sont infinis dénombrables, soit ils ont le même cardinal que $\mathbb{R}$.

Je ne présenterai pas ici la preuve du Théorème III-23 ; on la trouvera en grand détail dans [Parthasarathy, sections I.2]. Pour le Th éorème III-24 (objet de [Parthasarathy, section I.3]), une preuve sera présentée pour les lectrices averties, dans le Chapitre V ; ce théorème y sera replacé dans le cadre plus général de la théorie descriptive des ensembles et des théorèmes de **sélection mesurable**.

Remarques III-26 (Mesurabilité, règle ou exception ?). (i) En conséquence des théorèmes de stabilité précédents (énoncés III-7 à III-24), la plupart (sinon la totalité) des fonctions que l'on rencontre ou que l'on construit pour résoudre des problèmes d'analyse réelle sont mesurables. Par analyse réelle j'entends : quand le cadre est $\mathbb{R}$ ou $\mathbb{R}^n$ ou une variété riemannienne de dimension finie, munie de la tribu borélienne. Le plus souvent, on ne se donne même pas la peine de le mentionner, car la mesurabilité est presque automatique. En toute rigueur, il faut cependant démontrer cette mesurabilité. En pratique il y a (à peu de choses près) une seule situation où cela s'avère délicat : quand on prend un supremum ou un infimum sur une famille de fonctions non dénombrables. On reviendra sur cette question dans le chapitre VI.

(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 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, pour encoder de l'information par rapport à la dépendance aux variables aléatoires. Dans ce cas les tribus sont bien plus grossières que la tribu borélienne, et la mesurabilité 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 par 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 (et ouvertes). Il serait naturel d'introduire une fois pour toutes la tribu produit sur X, qui rend mesurables toutes les applications coordonnées $\sigma_k : x \longmapsto x_k$; c'est aussi la tribu borélienne sur le produit. 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 $\sigma_1, \ldots, \sigma_n$ (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 l'on s'autorise à oublier un ensemble de mesure nulle : c'est un corollaire du **théorème de Lusin**, présenté à la fin de ce chapitre. Dans le même esprit, le **théorème de Vitali–Carathéodory** montrera que l'on peut approcher une fonction mesurable par des fonctions semi-continues, par au-dessus ou par en-dessous, au prix d'une erreur arbitrairement petite sur l'intégrale.

III-1.3. Tribu engendrée par une fonction mesurable.

Théorème III-27 (tribu engendrée par une fonction). *(i) Soient X un espace quelconque et $(Y, \mathcal{B})$ 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, $(Y_t, \mathcal{B}_t)_{t\in T}$ une famille d'espaces mesurables indexés par un ensemble T arbitraire ; pour tout $t \in T$ on se donne une fonction $f_t : X \to Y_t$. 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

$$f^{-1}\Big(\bigcap B_k\Big) = \bigcap f^{-1}(B_k); \qquad f^{-1}\Big(\bigcup B_k\Big) = \bigcup f^{-1}(B_k).$$

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 III-28 (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 $\Phi(y) = z$. $\square$

On trouvera un peu plus loin une variante un peu moins ambitieuse, mais rigoureuse (Théorème III-40).

III-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 III-29 (mesurabilité pour la tribu complétée). *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et soit $\overline{\mathcal{A}}$ 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.*

La preuve de ce théorème est remise à la section III-2.2 ; elle reposera sur une très efficace méthode d'approximation des fonctions mesurables.

III-2. L'intégrale selon Lebesgue

De la même façon que les ensembles mesurables sont ceux dont on définit la mesure, les fonctions réelles mesurables sont celles dont on espère définir l'intégrale.

On a vu dans la section précédente l'extrême généralité de la notion de fonction mesurable, et en conséquence on pourra intégrer de très nombreuses fonctions dans la théorie de Lebesgue. Le principal prix à payer sera la renonciation aux compensations. Dans le cas de l'intégrale de Riemann, on a diverses recettes pour traiter des intégrales semi-convergentes, dans lesquelles de grandes valeurs positives et négatives proches se compensent, dans des procédés de limite ad hoc, qui font intervenir la structure de l'espace de définition. Par exemple, au voisinage d'un point de $\mathbb{R}$ on a de grandes valeurs positives d'un côté et négatives de l'autre, et on peut définir des approximations bien choisies à l'approche de ce point singulier. En théorie de Lebesgue, on pense aussi peu que possible à l'ensemble de départ (qui peut être très général), on se concentre sur les valeurs ; cela rend impossible en pratique un bon traitement des compensations. Le plan est donc de séparer tout simplement parties positive et négative, décrire $\int f = \int f_+ - \int f_-$ et de définir séparément les deux intégrales ; la question est alors d'intégrer les fonctions positives.

Il peut subsister un problème si l'intégrale diverge (vaut formellement $+\infty$), mais ce n'est pas bien grave : attribuons dans ce cas explicitement la valeur $+\infty$ quand cela se présente. On peut même, si cela est commode, autoriser la fonction intégrée à prendre ses valeurs dans $[0, +\infty]$. Si les deux intégrales $\int f_+$ et $\int f_-$ sont infinies, on ne parviendra pas à dire la valeur de $\int f$; il faut peut-être alors, en amont de la procédure de Lebesgue, introduire un redécoupage de l'intégrale, ou définir un procédé limite ; c'est une autre histoire.

Dans cette section, on commencera donc par définir l'intégrale des fonctions positives, en les approchant par des fonctions très simples : prenant seulement un nombre fini de valeurs. Ensuite, quand cela aura un sens, on généralisera aux fonctions signées (i.e. à valeurs positives ou négatives).

III-2.1. Fonctions simples. Comme dans la théorie de Riemann, on définit 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. Mais au lieu des fonctions constantes par morceaux, c'est une classe bien plus grande qui jouera un rôle privilégié : les fonctions mesurables prenant un nombre fini de valeurs, aussi appelées **fonctions simples**.

Définition III-30 (fonction simple). *Soit $(X, \mathcal{A})$ un espace mesurable. On appelle fonction simple positive (ou juste fonction simple, ou fonction étagée) une fonction de la forme*

$$f = \sum_{k=1}^{N} \alpha_k 1_{A_k},$$

où les α_k sont des nombres positifs (éventuellement $+\infty$) et les A_k sont des parties mesurables formant une partition de X.

Les deux critères suivants, presque évidents, sont laissés en exercice.

Proposition III-31 (reformulation de la simplicité). *(i) Soit $(A_k)_{1\leq k\leq N}$ une famille finie de parties mesurables, pas forcément une partition, et soit $(\alpha_k)_{1\leq k\leq N}$ une famille de nombres réels positifs (éventuellement $+\infty$) ; alors $f = \sum \alpha_k 1_{A_k}$ est simple.*

(ii) Une fonction positive f est simple si et seulement si elle est mesurable et prend un nombre fini de valeurs.

Remarques III-32. (i) La condition de positivité est imposée ici uniquement parce que nous avons pour but de définir d'abord l'intégrale des fonctions positives. Mais la définition des fonctions simples s'étend à des fonctions à valeurs dans un espace mesurable quelconque, pas forcément $\mathbb{R}_+$: on dira juste qu'une fonction est simple si elle est mesurable et ne prend qu'un nombre fini de valeurs.

(ii) Simples au sens de leurs valeurs (un ensemble fini), les “fonctions simples” peuvent néanmoins être très complexes dans leurs variations : aussi complexes que peuvent l’être les ensembles mesurables.

La définition de l’intégrale d’une fonction simple sous le sens :

Définition III-33 (intégrale d’une fonction simple). *Une fonction simple positive f étant donnée sur l’espace mesuré $(X, \mathcal{A}, \mu)$, avec les notations de la Définition III-30, on pose*

$$\int f\,d\mu = \sum_{k=1}^{N} \alpha_k\,\mu[A_k],$$

avec la convention $0 \times (+\infty) = 0$.

Proposition III-34 (Invariance de l’intégrale des fonctions simples). *(i) Si $f = \sum \alpha_k 1_{A_k}$ est une fonction simple, avec les α_k positifs mais les A_k ne formant pas forcément une partition, alors on a toujours*

$$\int f\,d\mu = \sum_{k=1}^{N} \alpha_k\,\mu[A_k].$$

(ii) Si f est une fonction simple prenant les valeurs $\alpha_1, \ldots, \alpha_K \in [0, +\infty]$, alors

$$\int f\,d\mu = \sum_{k=1}^{N} \alpha_k\,\mu\big[f^{-1}(\alpha_k)\big].$$

Cette dernière proposition, intuitive, est laissée en exercice (fastidieux et pas si simple !).

Proposition III-35 (additivité de l’intégrale des fonctions simples). *Soient f et g deux fonctions simples positives, alors pour tous $\alpha, \beta \in \mathbb{R}_+$, la fonction $\alpha f + \beta g$ est simple, et*

$$\int (\alpha f + \beta g)\,d\mu = \alpha \int f\,d\mu + \beta \int g\,d\mu.$$

Démonstration. Ecrivons $f = \sum a_j 1_{A_j}$, $g = \sum b_k 1_{B_k}$, 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

$$\begin{aligned}\sum_{jk} (\alpha a_j + \beta b_k)\mu[A_j \cap B_k] &= \alpha \sum_j a_j \Big(\sum_k \mu[A_j \cap B_k]\Big) + \beta \sum_k b_k \Big(\sum_j \mu[A_j \cap B_k]\Big)\\ &= \alpha \sum_j a_j \mu[A_j \cap (\cup B_k)] + \beta \sum_k b_k \mu[(\cup A_j) \cap B_k]\\ &= \alpha \sum_j a_j \mu[A_j] + \beta \sum_k b_k \mu[B_k].\end{aligned}$$

(On peut aussi déduire cet énoncé de la Proposition III-34.) □

III-2.2. Approximation des fonctions mesurables. C’est un résultat élémentaire et fondamental que toute fonction mesurable positive peut être approchée par des fonctions simples :

Théorème III-36 (approximation par des fonctions simples). *Soit f une fonction mesurable sur un espace mesurable $(X, \mathcal{A})$, à valeurs dans $[0, +\infty]$. Alors il existe une suite croissante $(\varphi_n)_{n\in\mathbb{N}}$ de fonctions simples positives, 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 $\widetilde{\varphi}_n = \varphi_n 1_{X_n}$ pour prouver la dernière partie de l'énoncé. □

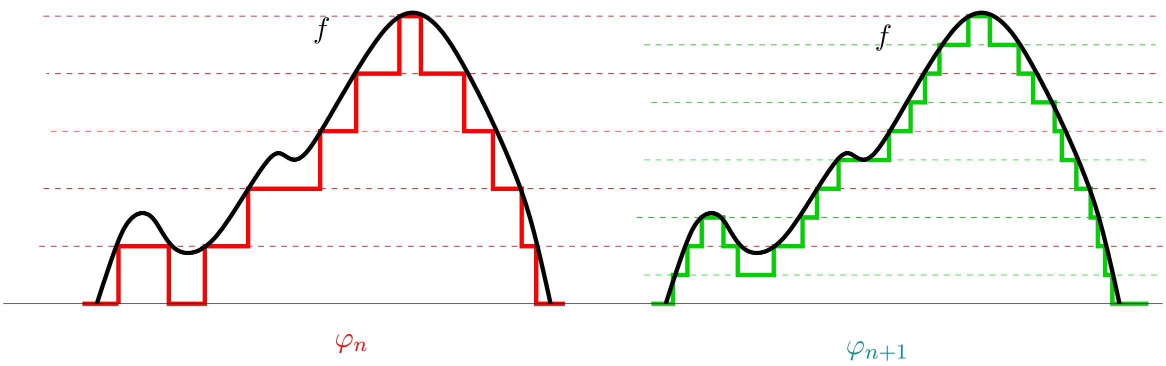

Figure 1. Approximation d'une fonction mesurable par des fonctions simples

Corollaire III-37 (une fonction mesurable est combinaison dénombrable de fonctions indicatrices). *Soit f une fonction mesurable, à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$. Alors il existe des nombres réels 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 III-38. 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).

On peut étendre ce résultat à des espaces bien plus généraux :

Théorème III-39 (approximation par des fonctions simples, encore). *Soit f une fonction mesurable entre espaces mesurables $(X, \mathcal{A})$ et $(Y, \mathcal{B})$. On suppose que Y est un espace polonais et $\mathcal{B}$ sa tribu borélienne ; on note $(y_k)_{k\in\mathbb{N}}$ une suite dense dans Y. Alors il existe une suite $(\varphi_n)_{n\in\mathbb{N}}$ de fonctions simples, prenant ses valeurs dans $\{y_k\}_{k\in\mathbb{N}}$, telle que φ_n converge simplement vers f.*

La démonstration du Théorème III-39 est laissée en exercice.

Pour illustrer l'intérêt de ces théorèmes d'approximation, voyons comment les utiliser pour résoudre le problème abordé dans le paragraphe III-1.3, dans un cadre légèrement restreint ; et dans la foulée pour démontrer le Théorème III-29.

Théorème III-40 (fonctions mesurables pour $\sigma(f)$). *Soient X un espace quelconque, Y un espace mesurable, et Z un espace polonais, 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 Z. Comme g prend un nombre fini de valeurs, la tentative de démonstration présentée au paragraphe III-1.3 aboutit (pourquoi ?) et permet de construire une fonction Φ mesurable, telle que $g = \Phi \circ f$.

2. Considérons pour commencer le cas où $Z = \mathbb{R}_+$. Par le Théorème III-36 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. Si maintenant $Z = \mathbb{R}$, écrivons $g = g_+ - g_-$ avec $g_\pm \geq 0$; par l'étape 2, $g_\pm = \Phi_\pm \circ f$, d'où $g = (\Phi_+ - \Phi_-) \circ f$, ce qui conclut la preuve.

4. Enfin si Z est un espace polonais quelconque, on peut raisonner de même en appliquant l'Exercice III-20 et le Théorème III-39. $\square$

Démonstration du Théorème III-29. 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 III-37, 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 $N := \cup N_k$, et $g = \sum c_k 1_{B_k}$. $\square$

Exercice III-41. En utilisant le même argument, démontrer la Proposition III-8(ii).

Maintenant le théorème d'approximation des fonctions mesurables par des fonctions simples permettra de construire l'intégrale des fonctions mesurables positives à partir de l'intégrale des fonctions simples.

III-2.3. Intégrale des fonctions positives.

Définition III-42 (intégrale d'une fonction positive). *Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, et f une fonction mesurable sur X, à valeurs dans $[0, +\infty]$. On appelle intégrale de f pour la mesure μ, et on note*

$$\int f \, d\mu$$

$$\left(ou \int_X f\,d\mu \ ou \int_X f(x)\,d\mu(x) \ ou \int_X f(x)\,\mu(dx)\right)$$

la quantité

$$\sup\left\{\int g\,d\mu;\quad g \ simple;\ 0\le g\le f\right\}\in[0,+\infty].$$

Remarque III-43. (i) Le supremum est pris sur une classe non vide, puisque la fonction nulle est admissible.

(ii) Si f est une fonction simple, cette définition coïncide avec la Définition III-33.

(iii) On verra dans le chapitre suivant que

$$\int f\,d\mu = \lim_{n\to\infty}\sum_{k\in\mathbb{N}}\frac{1}{2^n}\,\mu\left[\left\{x;\ f(x)\ge\frac{k}{2^n}\right\}\right],$$

ce qui justifie l'intuition suggérée par la figure 1 au Chapitre I. En ce sens, l'intégration 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 ; ou de façon équivalente, comme l'intégrale de la fonction mesurable $f1_A$; ou de façon équivalente comme l'intégrale de la restriction de f à A par rapport à la restriction de μ à A. On note que si $\mu[A]=0$, alors la restriction de μ à A est la mesure nulle, et en particulier $\int_A f\,d\mu=0$.

Exercice III-44. Soit C la mesure de comptage sur $\mathbb{N}$, muni de la tribu triviale $\mathcal{P}(\mathbb{N})$. Qu'est-ce qu'une fonction étagée sur $\mathbb{N}$? Montrer que si $f:\mathbb{N}\to[0,+\infty]$, alors $\int f\,dC=\sum_{n\in\mathbb{N}}f(n)$.

Définition III-45 (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 (ceux qui sont inclus dans un ensemble 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 III-46 (presque partout). *Soit (X,μ) un espace mesuré. On dit qu'une propriété est vraie μ-presque partout (ou $d\mu$-presque partout, ou $d\mu$-p.p., ou p.p.) si l'ensemble des éléments de X qui ne vérifient pas cette propriété est négligeable.*

Proposition III-47 (Propriétés élémentaires de l'intégrale des fonctions positives). *Soient $(X,\mathcal{A},\mu)$ un espace mesuré, et f,g deux fonctions positives mesurables sur X.*

(i) Si $\int f\,d\mu<+\infty$, alors f est finie μ-presque partout ;

(ii) Si $\int f\,d\mu=0$, alors f est nulle μ-presque partout ;

(iii) Si $f\le g$ μ-presque partout, alors $\int f\le\int g$. En particulier, si g est sommable, alors f l'est aussi.

(iv) Si $f=g$ μ-presque partout, alors $\int f=\int g$.

(v) 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.

(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.

(iv) résulte de (iii).

(v) est laissé en exercice. □

III-2.4. Intégrale des fonctions sommables.

Définition III-48 (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 III-49. 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 conséquence facile de la définition de l'intégrale et des propriétés précédentes.

Proposition III-50 (propriétés élémentaires de l'intégrale des fonctions sommables). *Soient $(X, \mathcal{A}, \mu)$ 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 III-51. 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 $c\,\mu[X]$.

III-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 évident que l'intégrale d'une fonction sommable positive est positive. Il est à peine moins évident que $\int(\lambda f) = \lambda \int f$ pour toute fonction f sommable et pour tout scalaire λ. En. revanche, la relation capitale $\int(f+g) = (\int f) + (\int g)$ est beaucoup plus subtile !

III-3.1. Addition des fonctions positives.

Théorème III-52 (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 $\int(f+g) = +\infty$ 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 $0 \le \varphi \le f$, $0 \le \psi \le g$, et $\int f \le \int \varphi + \varepsilon$, $\int g \le \int \psi + \varepsilon$. D'après la Proposition III-35, $\varphi + \psi$ est simple, et

$$\int f + \int g \le \int \varphi + \int \psi + 2\varepsilon = \int (\varphi+\psi) + 2\varepsilon \le \int (f+g) + 2\varepsilon.$$

En faisant tendre ε vers 0, on obtient

$$\int f + \int g \le \int (f+g).$$

2. Par la Proposition III-36, on peut trouver des suites croissantes (φ_k) et (ψ_k) de fonctions simples telles que $0 \le \varphi_k \le f$, $0 \le \psi_k \le g$, convergeant simplement vers f et g respectivement. Soit $\delta \in (0,1)$ arbitraire, on pose

$$A_k := \Big\{x;\ \varphi_k(x) \ge (1-\delta)\,f(x),\ \psi_k(x) \ge (1-\delta)\,g(x)\Big\}.$$

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,

$$\int f + \int g \ge \int \varphi_k + \int \psi_k \ge \int_{A_k} \varphi_k + \int_{A_k} \psi_k = \int_{A_k} (\varphi_k + \psi_k).$$

Par la définition de A_k et la positivité de l'intégrale, on a

$$\int_{A_k} (\varphi_k + \psi_k) \geq \int_{A_k} (1-\delta)\chi = (1-\delta) \int_{A_k} \chi.$$

Écrivons χ sous la forme $\sum_{1\leq j\leq J} \alpha_j 1_{B_j}$. Puisque les A_k forment une famille croissante et que $\mu[X \setminus (\cup A_k)] = 0$, on a

$$\int_{A_k} \chi = \sum_j \alpha_j\, \mu[A_k \cap B_j] \xrightarrow[k\to\infty]{} \sum_j \alpha_j\, \mu[(\cup A_k) \cap B_j] = \sum_j \alpha_j\, \mu[(\cup A_k) \cap B_j] = \int \chi.$$

En passant à la limite quand $k \to \infty$ 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 III-53. Dans le chapitre suivant, un raisonnement exactement similaire permettra de 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.

III-3.2. Généralisation : fonctions sommables. Le Théorème III-52 entraîne facilement la linéarité de l'intégrale des fonctions sommables.

Théorème III-54 (linéarité de l'intégrale). *Soient (X,μ) un espace mesuré ; f, g deux fonctions sommables sur X, à valeurs dans $\overline{\mathbb{R}}$; 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 $\int (\alpha f)\, d\mu = \alpha(\int f\, d\mu)$, et $\int (\beta g)\, d\mu = \beta(\int g\, d\mu)$. Il suffit donc de montrer que si f et g sont sommables *de signe quelconque*, alors $\int (f+g)\, d\mu = (\int f\, d\mu) + (\int g\, d\mu)$. 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 III-52 :

$$\int (f+g)_+\,d\mu + \int f_-\,d\mu + \int g_-\,d\mu = \int f_+\,d\mu + \int g_+\,d\mu + \int (f+g)_-\,d\mu.$$

Toutes ces quantités sont finies puisque $|f|$ et $|g|$ sont intégrables, on en déduit donc

$$\int (f+g)_+\,d\mu - \int (f+g)_-\,d\mu = \int f_+\,d\mu - \int f_- d\,d\mu + \int g_+\,d\mu - \int g_-\,d\mu,$$

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. Enfin la propriété de positivité est évidente : si f est mesurable positive, alors $\int f\,d\mu \geq 0$. □

Définition III-55 (espace de Lebesgue). *Si $(X,\mathcal{A},\mu)$ est un espace mesuré, l'espace vectoriel des fonctions sommables est noté $L^1(X,\mu)$ (ou $L^1(X,\mathcal{A},\mu)$, ou $L^1(X,d\mu)$, ou $L^1(X)$, ou $L^1(\mu)$, ou $L^1(d\mu)$) et appelé espace de Lebesgue d'exposant 1.*

Remarque III-56. Soient f et g deux fonctions sommables, et $\lambda \in \mathbb{R}$. Alors

$$\int |\lambda f| = \int |\lambda||f| = |\lambda| \int |f|;$$

$$\int |f+g| \leq \int |f| + |g| = \int |f| + \int |g|.$$

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 $\int |f| = 0 \Longrightarrow f = 0$. Mais cette dernière identité est évidemment fausse : on sait que $\int |f| = 0$ si et seulement si la fonction f est nulle hors d'un ensemble μ-négligeable, ce qui n'impose pas à f d'être identiquement nulle, mais seulement nulle presque partout, ou "presque 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". Attention : 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.

III-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 III-57 (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 III-58. La nouveauté par rapport au Théorème III-54 est la continuité de μ. On rappelle qu'une forme linéaire L est dite continue si

$$\|L\| = \sup_{x\neq 0} \frac{|Lx|}{\|x\|} < +\infty.$$

Preuve de la Proposition III-57. Bien sûr, toute fonction continue est borélienne, donc mesurable, et toute fonction mesurable bornée est intégrable. En outre,

$$\left|\int f\right| \leq \int |f| \leq \int \|f\|_\infty = \mu[X]\,\|f\|_\infty.$$

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 $\mu[X]$. Le choix $f = 1$ atteint la borne, d'où $\|\mu\|_{C_b(X)^*} = \mu[X]$. □

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 $X = \mathbb{R}^d$. 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. La prochaine section sera l'occasion de démontrer un énoncé un peu plus général.

Exercice III-59. (i) Admettons pour quelques instants que toute forme linéaire continue sur un sous-espace fermé d'un espace vectoriel E peut se prolonger en une forme linéaire continue sur E tout entier (cette version du théorème de Hahn–Banach exige l'axiome du choix général). Étendre l'application "limite à l'infini" sur les fonctions continues $\mathbb{R} \to \mathbb{R}$ qui convergent à l'infini, en une forme linéaire continue sur l'espace de toutes les fonctions continues bornées sur $\mathbb{R}$, convergentes ou non. Montrer que cette forme linéaire est finiment additive, mais pas σ-additive, et n'est donc pas une mesure.

(ii) Retournant maintenant à l'axiomatique de ce cours qui ne comprend pas l'axiome du choix général, montrer qu'il est soit faux, soit indécidable, que $C_b(\mathbb{R})^*$ soit constitué de mesures.

III-4. L'intégrale selon Riesz

Le résultat central de cette section, le théorème fondamental de Riesz, s'applique quand l'espace est localement compact, et atteint alors deux objectifs simultanément :

- il identifie le dual de l'espace des fonctions continues à support compact ;
- il fournit une autre construction, alternative à celle de Lebesgue mais équivalente, de l'intégration.

III-4.1. Enoncé du théorème.

Commençons par quelques définitions.

Définition III-60 (espaces de fonctions continues). *Soit X un espace topologique arbitraire. Si $f : X \to \mathbb{R}$ 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,*

$$C_c(X) \subset C_0(X) \subset C_b(X) \subset C(X).$$

L' inclusion est stricte en général, sauf quand X est compact, auquel cas tous ces espaces coïncident. 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 III-61. 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. En fait 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 convention qui suit, interne à cette section, sera utile pour abréger quelques formulations.

Définition III-62 (pré-régularité). *Soit μ une mesure de Borel sur un espace topologique X. On dira que μ est pré-régulière si pour tout Borélien A de X,*

$$\mu[A] = \inf \big\{\mu[O];\ \ O \text{ ouvert};\ A \subset O\big\}$$

(régularité extérieure) et pour tout **ouvert** *B de X,*

$$\mu[B] = \sup \big\{\mu[K];\ K \text{ compact};\ K \subset B\big\}.$$

Théorème III-63 (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, pré-régulières et finies sur les compacts ;

via les formules

$$\begin{cases} \Lambda f & = \int f\, d\mu, \text{ pour tout } f \in C_c(X), \\ \mu[O] & = \sup\Big\{\Lambda f,\ f \in C_c(X),\ 0 \leq f \leq 1_O\Big\}, \text{ pour tout ouvert } O. \end{cases}$$

Avant de continuer, voici une liste de commentaires sur cet énoncé, qui admet quelques variantes plus ou moins subtiles.

Remarques III-64. (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 vérifiera en exercice que ses compacts (décrits par le théorème d'Ascoli) 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 décrite dans la section II-2.

(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 $C_c(X)$ pour la norme de la convergence uniforme.

(iii) Dans la définition de "pré-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 VIII-66 ; 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 III-67 et III-68 présentés dans la sous-section suivante.

(iv) On peut, si on le souhaite, compléter la mesure μ grâce au Théorème II-93, 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 ($f \leq g \Longrightarrow \Lambda f \leq \Lambda g$) 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 VIII-66.

(vi) Si l'on réfléchit un peu à l'énoncé, on a l'impression que l'hypothèse de prérégularité 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 pr̂e-régularité, rien ne garantit a priori l'*unicité* de la mesure μ correspondant à Λ.

Voici maintenant deux remarques d'ordre plus général :

Remarque III-65. 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 III-66. 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 VIII). 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 III-63, je vais maintenant donner deux énoncés simplifiés.

III-4.2. Enoncés 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 III-67 (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 les formules

$$\begin{cases} \Lambda f &= \int f\,d\mu, \ \ \textit{pour tout } f \in C_c(X) \\ \mu[B] &= \sup\Big\{\Lambda f,\ f \in C_c(X),\ 0 \leq f \leq 1_B\Big\}, \ \textit{pour tout borélien } B. \end{cases}$$

Ces mesures sont automatiquement régulières.

Théorème III-68 (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 les formules

$$\begin{cases} \Lambda f &= \int f\,d\mu, \text{ pour tout } f \in C(X) \\ \mu[B] &= \sup\Big\{\Lambda f,\ f \in C_c(X),\ 0 \le f \le 1_B\Big\}, \text{ pour tout borélien } B. \end{cases}$$

Ces mesures sont automatiquement régulières, et ces formes linéaires sont automatiquement continues ; on a alors

$$\|\Lambda\| = \mu[X]. \tag{12}$$

Démonstration. Les Théorèmes III-67 et III-68 s'obtiennent en combinant le théorème de représentation de Riesz avec les théorèmes de régularité II-64 et II-62, respectivement. En ce qui concerne (12), pour l'égalité il suffit de choisir $f = 1$ dans le calcul de $\sup \|\Lambda f\|/\|f\|_\infty$. □

Passons maintenant à la démonstration du Théorème de Riesz. Il découlera assez simplement du Théorème de Carathéodory généralisé établi au Chapitre II, Théorème II-82. D'autres approches sont possibles. Une démonstration compacte (!), assez délicate, est proposée dans [Rudin, pp. 40-47] ; mais elle reprend plusieurs des arguments utilisés dans la preuve du Théorème II-82. On voit ici l'intérêt du Théorème II-82 : démontrer le Théorème de Riesz via le Théorème de Carathéodory lui-même, sous la forme du Théorème II-78, est un formidable casse-tête ! Une variante de cette dernière démarche est menée à bien dans [Dudley], via un intermédiaire délicat, le 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.

III-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 prérégulière, au sens de la Définition III-62, on sait qu'il existe un ouvert O contenant K tel que $\mu_1[K] \ge \mu_1[O] - \varepsilon$, 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

$$\mu_2[K] = \int 1_K\,d\mu_2 \le \int \varphi\,d\mu_2 = \int \varphi\,d\mu_1 \le \int 1_O\,d\mu_1 = \mu_1[O] \le \mu_1[K] + \varepsilon.$$

On conclut en faisant tendre ε vers 0 que $\mu_2[K] \le \mu_1[K]$, 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 prérégulières, elles coïncident également sur les ouverts, et par suite (toujours par prérégularité) 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

$$\mu[O] = \sup \Big\{ \Lambda f;\ f \in C_c(X);\ 0 \leq f \leq 1_O \Big\}.$$

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 II-82(ii) assure l'existence d'un tel prolongement si la condition (8) 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 $O := \cup O_k$, 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 II-42 de partition de l'unité pour trouver des fonctions continues $\chi_{i_1}, \ldots, \chi_{i_n}$, 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,

$$\mu^*[A] = \inf \big\{ \mu[O];\ O \text{ ouvert},\ A \subset O \big\}.$$

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,*

$$\mu[A \cap B] + \mu^*[A \setminus B] \leq \mu[A]. \tag{13}$$

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 (13). 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 II-43, 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 III-64 ne nécessitent pas de justification, sauf le point (iii) que je vais maintenant considérer.

Démonstration de la Remarque III-64(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

$$\mathcal{C} := \big\{A \in \mathcal{A};\ A \cap C \in \mathcal{B} \text{ pour tout compact } C\big\}.$$

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 II-77 (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 $K_1 \cup \ldots \cup K_\ell$, 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. □

III-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 approfondie, et s'il existe 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 III-69 (Théorème de Lusin). *Soit X un espace topologique séparé localement compact, et soit μ une mesure de Borel régulière sur X. Soit $f : X \to \mathbb{R}$ 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 de Lusin le corollaire suivant :

Corollaire III-70 (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*

$$\int |f_n - f| \, d\mu \xrightarrow[n\to\infty]{} 0.$$

Démonstration du Théorème III-69. Démontrons les deux énoncés (i) et (ii) 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 II-41, 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

$$\{x;\ \varphi_{k,n}(x) \neq g_k(x) - g_{k-1}(x)\} \subset A_{k,n}; \qquad \mu[A_{k,n}] \leq 2^{-k}/n.$$

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 $\mu[E_m] \to 0$ quand $m \to \infty$. 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 III-70. 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 IV-12 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 $\int |f - f_n| \to 0$. $\square$

Remarque III-71. Ce théorème de densité est très général, mais pas très explicite. On verra 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 $f_\varepsilon \le f$, 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 III-72 (Théorème de Vitali-Carathéodory). *Soit X un espace topologique localement compact, et soit μ une mesure de Borel régulière sur X. Soit $f : X \to \mathbb{R}$ une fonction mesurable. Alors, pour tout $\varepsilon > 0$ il existe des fonctions f^+ et f^-, telles que $f^- \le f \le f^+$, f^- est semi-continue supérieurement et majorée, f^- est semi-continue inférieurement et minorée, et*

$$\int f^+ \, d\mu - \varepsilon \le \int f \le \int f^- \, d\mu + \varepsilon.$$

Remarque III-73. Attention, ici f^+ et f^- n'ont rien à voir avec f_+ (partie positive) et f_- (partie négative).

Démonstration. Quitte à séparer f en parties positive et négative, on peut supposer $f \ge 0$. En approchant f par une suite de fonctions simples, on constate que l'on peut écrire

$$f = \sum_{i=1}^{\infty} c_i 1_{E_i}, \qquad \sum_{i=1}^{\infty} c_i \mu[E_i] < +\infty.$$

Pour chaque i on choisit un ouvert O_i contenant E_i, et un compact K_i contenu dans E_i, tels que

$$\mu[O_i \setminus K_i] \le \varepsilon/2^{i+1}.$$

On pose alors

$$f^- = \sum_{i=1}^{\infty} c_i 1_{K_i}, \qquad f^+ = \sum_{i=1}^{N} c_i 1_{O_i},$$

où N est choisi de telle sorte que

$$\sum_{i=N+1}^{\infty} c_i \mu[E_i] \le \varepsilon/2.$$

On vérifie facilement que f^+ et f^- vérifient toutes les conditions requises. $\square$

III-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 III-74 (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 $f : X \to E$ 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 demi-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 – par exemple $f(t,x) = f(t)(x)$ – 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 nettement plus délicat de définir l'intégrale de f :

- soit on peut la définir composante par composante, sous de bonnes hypothèses de séparabilité, en particulier ;

- soit, si l'on intègre sur un espace de fonctions, on reprend la théorie en distinguant intégration de la partie positive et intégration de la partie négative.

- soit on reprend la théorie directement dans un cadre fonctionnel élargi, en ne considérant comme fonctions simples que des combinaisons linéaires de fonctions indicatrices d'ensembles *de mesure finie* ;

Le Théorème de Désintégration de la mesure, au Chapitre **??**, fournira un bon exemple.

L'adaptation de la théorie de Lebesgue à l'intégration sur des espaces fonctionnels est la théorie de **l' intégrale de Bochner**. Les principaux résultats en 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

$$\int_X \bigl(\sup_Y |u(x,y)|\bigr)\, d\mu(x) < +\infty;$$

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, la lectrice devrait pouvoir facilement interpréter les principaux résultats de l'intégration à valeurs vectorielles.

Sans développer une théorie complète, je montrerai au chapitre **??** comment définir une théorie simple d'intégration à valeurs vectorielles qui couvre la plupart des espaces habituels. Pour cela, attendons d'être un peu plus aguerris en analyse fonctionnelle.

CHAPITRE IV

Théorèmes fondamentaux d'intégration

Maintenant que l'intégrale est définie, on va établir ses propriétés fondamentales : celles qui servent constamment et qui ont fait le succès de la théorie de Lebesgue.

Ce chapitre passe donc en revue, dans un cadre très général, les outils-clés suivants : (i) des théorèmes de passage à la limite sous l'intégrale, (ii) des théorèmes de changement de variable abstrait, (iii) des théorèmes d'intégration produit, et (iv) des inégalités contrôlant les expressions intégrales. Tout cela occupe les sections IV-1 à IV-4, formant peut-être la partie la plus importante du cours.

Les sections IV-4.5 et IV-5.2 traitent de sujets plus avancés : d'une part les notions d'équi-intégrabilité et de tension, en lien avec la compacité des familles de mesures ; d'autre part, la construction de mesures produits avec un nombre infini de facteurs.

Dans tout ce chapitre on travaillera avec des mesures "individuellement" : typiquement, un théorème fera intervenir une mesure fixée. Le Chapitre VIII, au contraire, considérera des *espaces* de mesures, étudiées collectivement.

Certains théorèmes ou contre-exemples se baseront sur la mesure de Lebesgue dans $\mathbb{R}$, dont l'existence a été établie dans la Section II-8 ; ici 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. Plus tard, dans le Chapitre VI, on se plongera plus en détail dans les propriétés de cette mesure particulière.

IV-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. Peut-on passer à la limite dans l'intégrale des f_n ? On va passer en revue quatre problèmes différents :

$\star$ 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 $\int$? Le **Théorème de convergence monotone de Beppo Levi** assure que c'est toujours possible.

$\star$ 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** l'autorise.

$\star$ Puis on considère le cas où la suite (f_n) ne converge pas nécessairement ; tout au moins, on sait alors que sa limite inférieure et sa limite supérieure 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**.

$\star$ Quand on s'intéresse aux fonctions continues, une hypothèse rès forte que l'on utilise souvent est la convergence **uniforme**, qui permet en particulier de passer à la limite dans l'intégrale de Riemann. La théorie de Lebesgue ne contient pas de théorème de limite sous hypothèse de convergence uniforme, car la notion plus faible de convergence simple lui suffit bien. On peut se demander à quel point la notion de convergence uniforme est plus forte que la notion de convergence simple. Dans le cadre des fonctions continues, la nuance est considérable. Mais le **Théorème d'Egorov** implique que, du point de vue de la mesure, la différence est fine. On peut ainsi parfois ramener un problème de convergence simple à un problème de convergence uniforme.

IV-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, et qui généralise un résultat antérieur de Lebesgue. 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 IV-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 $(X,\mathcal{A},\mu)$, à valeurs dans $[0,+\infty]$. Alors*

$$\int_X \lim_{n\to\infty} f_n\,d\mu = \lim_{n\to\infty}\int_X f_n\,d\mu. \tag{14}$$

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.

Les deux corollaires qui suivent s'obtiennentt en remarquant que les sommes partielles d'une série à termes positifs forment une famille croissante.

Corollaire IV-2 (interversion de série et sommation pour des fonctions positives). *Soit $(f_n)_{n\in\mathbb{N}}$ une famille de fonctions mesurables, définies sur un espace mesuré $(X,\mathcal{A},\mu)$, à valeurs dans $[0,+\infty]$. Alors*

$$\int\left(\sum_{n\in\mathbb{N}} f_n\right)d\mu = \sum_{n\in\mathbb{N}}\int f_n\,d\mu. \tag{15}$$

Corollaire IV-3 (interversion de limite croissante et série). *Soit $(a_{jm})_{j\in\mathbb{N},m\in\mathbb{N}}$ un tableau dénombrable de nombres réels positifs, croissante en m. Alors*

$$\lim_{m\to\infty}\sum_{j\in\mathbb{N}} a_{jm} = \sum_{j\in\mathbb{N}}\lim_{m\to\infty} a_{jm}. \tag{16}$$

Exemple IV-4. 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

$$\lim_{m\to\infty}\mu^m[A] = \sum_{j\in\mathbb{N}}\big(\lim_{m\to\infty}\mu^m[A_j]\big).$$

Il s'ensuit qu'*une limite croissante de mesures est une mesure.* Ou encore : *une somme dénombrable de mesures est une mesure.*

Remarques IV-5. (i) Soit $(A_n)_{n\in\mathbb{N}}$ une famille croissante d'ensembles mesurables, $A = \cup A_n$, et soit $f_n = 1_{A_n}$; alors la fonction indicatrice 1_A est la limite croissante des f_n, et la formule (14) devient donc

$$\mu[\cup A_k] = \lim_{n\to\infty}\mu[A_n].$$

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 (15) se transforme en

$$\mu\big[\cup A_n\big] = \sum \mu[A_n].$$

On retrouve donc en cas particulier du théorème de convergence monotone 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) Clairement, les énoncés précédents sont également valables si les conditions de croissance ou de décroissance ne sont vérifiées que μ-presque partout.

Démonstration du Théorème IV-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

$$\int \lim_{n\to\infty} f_n = \int \lim_{n\to\infty}(f_n - g) + \int g = \lim_{n\to\infty}\int (f_n - g) + \int g = \lim_{n\to\infty}\int f_n.$$

On traite l'autre cas en changeant f_n en $-f_n$. Il suffit donc d'établir (i).

Soit $f = \lim f_n$; par hypothèse $f_n \leq f$, et 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 reste à établir l'inégalité inverse, qui est le coeur du problème. On va 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

$$A_n = \big\{x \in X;\ f_n(x) \geq (1-\delta)\chi(x)\big\}.$$

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 que la réunion des A_n est l'espace X tout entier. Si l'on écrit $\chi = \sum_{1\leq j\leq J}\alpha_j 1_{B_j}$,

$$\int \chi 1_{A_n} = \int \sum_{j=1}^{J}\alpha_j 1_{A_n\cap B_j} = \sum_{j=1}^{J}\alpha_j\mu[A_n\cap B_j] \xrightarrow[n\to\infty]{} \sum_{j=1}^{J}\alpha_j\mu[X\cap B_j] = \int \chi.$$

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

$$\lim_{n\to\infty} \int f_n \geq (1-\delta) \int \chi.$$

En prenant le supremum sur toutes les fonctions simples χ minorant f, et en faisant tendre δ vers 0, on aboutit bien à

$$\lim_{n\to\infty} \int f_n \geq \int f.$$

□

Exercice IV-6. Retrouver l'additivité de l'intégrale en combinant le théorème de convergence monotone et la Proposition III-36.

Voici maintenant une conséquence simple et importante du théorème de convergence monotone.

Proposition IV-7 (l'intégrale restreinte définit une mesure). *Soit f une fonction mesurable définie sur un espace mesuré $(X, \mathcal{A}, \mu)$, à valeurs dans $[0, +\infty]$. Alors la fonction d'ensembles $f\mu$ définie par*

$$f\mu[A] = \int_A f\,d\mu = \int_X f\,1_A\,d\mu \tag{17}$$

est une mesure sur la σ-algèbre $\mathcal{A}$. En outre, elle vérifie

$$\mu[A] = 0 \Longrightarrow f\mu[A] = 0. \tag{18}$$

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 donc

$$f1_A = \sum (f1_{A_k}).$$

Le Corollaire IV-2 implique donc

$$\sum_k \int (f1_{A_k})\,d\mu = \int \left(\sum_k f1A_k\right)\,d\mu = \int (f1_A)\,d\mu,$$

soit

$$\sum_k \int_{A_k} f\,d\mu = \int_{\cup A_k} f\,d\mu.$$

Cette propriété de σ-additivité montre que $f\mu$ est bien une mesure. □

La propriété (18) est importante et mérite un nom :

Définition IV-8 (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 toute partie A mesurable,*

$$\mu[A] = 0 \Longrightarrow \nu[A] = 0.$$

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 IV-9 (changement de densité de référence). *Sur* $(X, \mathcal{A}, \mu)$ *un espace mesuré,*

(i) Soient f *et* g *deux fonctions mesurables à valeurs dans* $[0, +\infty]$*. Alors*

$$\int fg\,d\mu = \int f\,d(g\mu);$$

(ii) Soient h *et* g *deux fonctions mesurables à valeurs dans* $[0, +\infty]$*, telles que*

$$g(x) \in \{0, +\infty\} \Longrightarrow h(x) = 0.$$

Alors, avec les conventions $1/0 = +\infty$*,* $0 \times (+\infty) = 0/0 = (+\infty)/(+\infty) = 0$*, on a*

$$\int h\,d\mu = \int \frac{h}{g}\,d(g\mu). \tag{19}$$

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 $\int fg\,d\mu$ et $\int f_n\,d(g\mu)$ vers $\int f\,d(g\mu)$. Il suffit donc de prouver (19) 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) Les hypothèses faites sur g et h garantissent que

$$h = \left(\frac{h}{g}\right) g,$$

ce qui permet d'appliquer (i) avec $f = h/g$. □

IV-1.2. Lemme de Fatou.

Théorème IV-10 (Lemme de Fatou). *(i) Soient* $(f_n)_{n\in\mathbb{N}}$ *une suite de fonctions définies sur un espace mesuré* $(X, \mathcal{A}, \mu)$*, à valeurs dans* $[0, +\infty]$*. Alors*

$$\int (\liminf_{n\to\infty} f_n)\,d\mu \le \liminf_{n\to\infty} \int f_n\,d\mu.$$

(ii) Cette conclusion est toujours valable si les f_n *sont à valeurs dans* $\overline{\mathbb{R}}$ *et toutes minorées par une fonction sommable.*

(iii) Symétriquement, si $(f_n)_{n\in\mathbb{N}}$ *est une suite de fonctions à valeurs dans* $\overline{\mathbb{R}}$*, toutes majorées par une fonction sommable, alors*

$$\int (\limsup_{n\to\infty} f_n)\,d\mu \ge \limsup_{n\to\infty} \int f_n\,d\mu.$$

Démonstration. Là encore, l'énoncé (iii) découle de (ii) via un changement de signe, et l'énoncé (ii) découlera de l'énoncé (i), on se concentre donc sur ce dernier.

Soit $g_n := \inf_{k\ge n} f_k$. On vérifie facilement que g_n est mesurable, et définit une suite **croissante** qui converge partout vers $f := \liminf f_k$. Bien sûr, $g_n \le f_n$. En appliquant le théorème de convergence monotone et en passant à la $\liminf$, on trouve

$$\int f = \lim_{n\to\infty} \int g_n \le \liminf \int f_n.$$

□

Remarque IV-11. Il est facile de construire des exemples où

$$\int \liminf f_n < \liminf \int f_n,$$

même si la convergence a lieu partout : cela fournit ainsi des contre-exemples au passage à la limite sous l'intégrale. 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 ; quitte à la multiplier par une constante, supposons que $\int \varphi = 1$. 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 on montre, par des changements de variables élémentaires, 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 de fonctions se concentre près de 0), la suite (g_n) un phénomène d'**évanescence** (toute la masse part à l'infini de manière diffuse), et la suite (h_n) un comportement de **bosse glissante** (la masse “glisse” à l'infini, sans s'étaler). Concentration, évanescence et glissade sont les trois obstructions archétypes au passage à la limite sous l'intégrale.

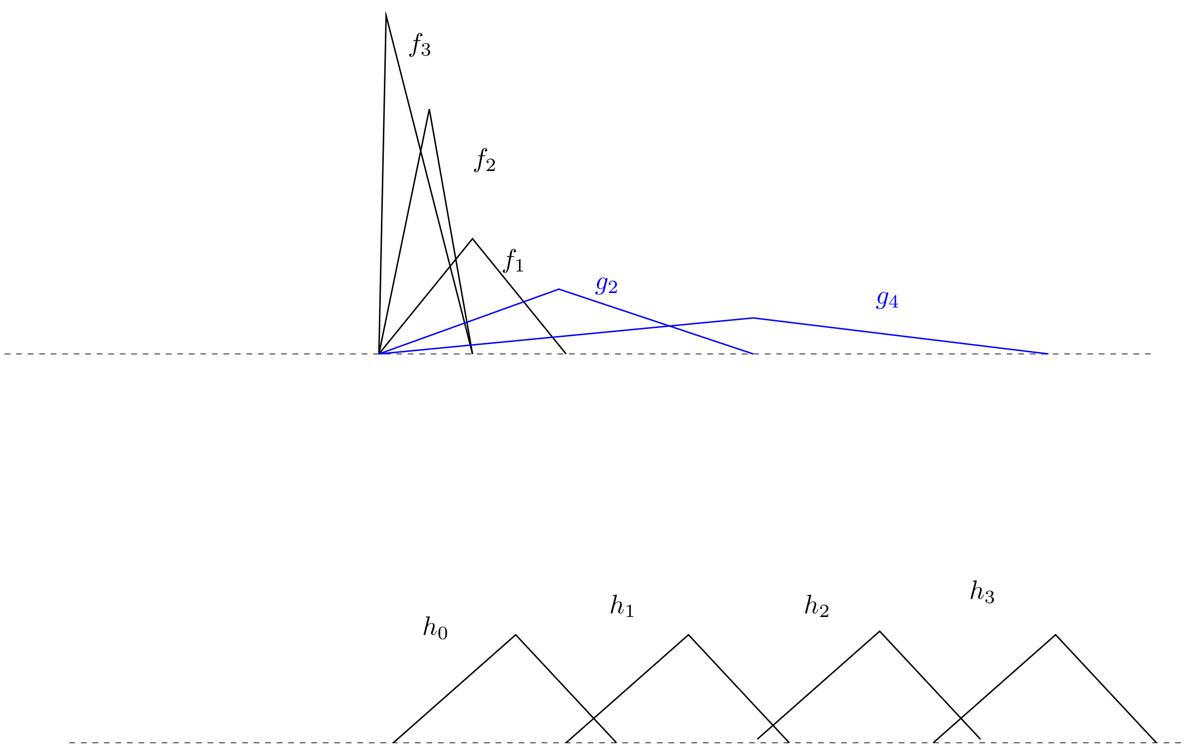

Figure 1. concentration des f_n, évanescence des g_n, bosse glissante h_n

IV-1.3. Convergence dominée.

THÉORÈME IV-12 (théorème de convergence dominée de Lebesgue). *Soit* $(f_n)_{n\in\mathbb{N}}$ *une famille de fonctions définies sur un espace mesuré* $(X, \mathcal{A}, \mu)$*, 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

$$\int_X \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int_X f_n \, d\mu \leq \limsup_{n\to\infty} \int_X f_n \, d\mu \leq \int_X \limsup_{n\to\infty} f_n \, d\mu. \tag{20}$$

En particulier, si (f_n) *est dominée et vérifie*

(ii) f_n *converge presque partout vers* f*,*
alors f *est intégrable et*

$$\int_X f \, d\mu = \lim_{n\to\infty} \int_X f_n \, d\mu. \tag{21}$$

En outre, on a l'énoncé plus précis

$$\lim_{n\to\infty} \int_X |f_n - f| \, d\mu = 0. \tag{22}$$

Enfin, 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 tel que* $\int (\lim g_n) = \lim \int g_n < +\infty$*.*

En introduisant les sommes partielles de séries de fonctions, on déduit de ce théorème le corollaire suivant :

COROLLAIRE IV-13 (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* $(X, \mathcal{A}, \mu)$*, 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 IV-14. (i) La version courte du Théorème IV-12 est la suivante : *Dès que* $(f_n)_{n\in\mathbb{N}}$ *est dominée et converge presque partout, on a*

$$\int_X \lim f_n = \lim \int_X f_n. \tag{23}$$

Au plan formel, c'est donc juste une interversion entre les opérations de limite et d'intégration.

(ii) La fonction f dans (21) (ou la fonction $\lim f_n$ dans (23)) n'est définie qu'en dehors d'un ensemble de mesure nulle ; en toute rigueur, pour que la formule ait un sens, il faut soit la restreindre à l'ensemble mesurable C où la suite de fonctions converge, soit étendre la fonction limite en une fonction mesurable sur X tout entier (également appelée f par abus de notation) grâce au Théorème III-16(iii) ; dans ce dernier cas, peu importent les valeurs du prolongement hors de l'ensemble de convergence.

(iii) La condition (i) dans le Théorème IV-12 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 IV-11 : en cherchant une fonction dominante on serait tombé essentiellement, respectivement sur : la fonction $1/x$ sur $[0,1]$, la fonction $1/x$ sur $[1,+\infty[$, et la fonction constante 1 sur $[0,+\infty[$, toutes trois non sommables.

(iv) 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 (u_n) à valeurs réelles est telle que toute suite extraite admet une sous-suite convergeant vers $\ell \in \mathbb{R}$, alors la suite (u_n) tout entière converge vers ℓ (en l'occurrence, $\ell = \int f$).

Exemple IV-15. Sachant que l'intégrale de Lebesgue généralise l'intégrale de Riemann, on déduit facilement du Théorème IV-12 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.

Remarque IV-16. L'énoncé IV-15 a beau s'exprimer en termes de concepts classiques − fonctions continues par morceaux, intégrale de Riemann − il est fort difficile à démontrer avec des outils classiques (même quand les fonctions f_n sont continues), alors qu'il tombe comme un fruit mûr dans le jardin de Lebesgue.

Démonstration du Théorème IV-12. Partons de l'hypothèse plus générale (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 à démontrer est une conséquence simple du Lemme de Fatou. En effet, la fonction $g_n + f_n$ est positive, donc

$$\int \liminf (g_n + f_n) \le \liminf \int (g_n + f_n).$$

En combinant cela avec l'hypothèse (i'), on obtient

$$\begin{aligned}\int g + \int (\liminf f_n) = \int (g + \liminf f_n) &= \int \liminf (g_n + f_n) \\ &\le \liminf \int (g_n + f_n) = \int g + \liminf \int f_n.\end{aligned}$$

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 (20) sont donc bien satisfaites.

Supposons maintenant que f_n converge presque partout vers f ; alors $\liminf f_n$ et $\limsup f_n$ coïncident avec f, d'où

$$\int f \leq \liminf \int f_n \leq \limsup \int f_n \leq \int f,$$

il y a donc égalité partout, ce qui démontre (21).

Montrons enfin (22), à savoir que $\int |f_n - f| \longrightarrow 0$; puisque f est intégrable, cela impliquera à nouveau, par inégalité triangulaire, que $\int f_n$ converge bien vers $\int f$:

$$\left| \int f_n \, d\mu - \int f \, d\mu \right| = \left| \int (f_n - f) \, d\mu \right| \leq \int |f_n - f| \, d\mu \xrightarrow[n \to \infty]{} 0.$$

Pour prouver (22), on applique la conclusion précédente (21) en remplaçant f_n et g_n par $\widetilde{f_n} = |f_n - f|$ et $\widetilde{g}_n = g_n + |f|$. L'hypothèse (i') est bien vérifiée pour $\widetilde{f_n}$ (et d'ailleurs aussi l'hypothèse (i), si (f_n) la vérifie). D'où

$$0 \leq \limsup \int |f_n - f| \leq \int \limsup |f_n - f| = 0,$$

ce qui prouve bien (22). □

Exercice IV-17. 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 $\sigma : \mathbb{N} \to \mathbb{N}$, et la valeur de la somme ne dépend pas de σ.

Remarque IV-18. Ici j'ai déduit le Théorème de convergence dominée du Lemme de Fatou, qui lui-même découlait du Théorème de convergence monotone. Mais à partir du Théorème de convergence dominée on peut aussi retrouver le Théorème de convergence monotone, au moins quand X est σ-fini ; de sorte que les trois énoncés sont quasiment équivalents. En effet, supposons que $f_n \geq 0$ converge en croissant vers f. Si f est sommable, alors f_n est dominée par f, et on peut passer à la limite dans l'intégrale par convergence dominée. Et si f n'est pas sommable, alors on peut trouver une famille $(\chi_m)_{m \in \mathbb{N}}$ de fonctions étagées, telles que $0 \leq \chi_m \leq f$ et $\int \chi_m \to +\infty$; quitte à remplacer χ_m par $\chi_m 1_{X_k}$, où $\mu[X_k] < +\infty$ et $\cup X_k = X$, on peut supposer χ_m intégrable ; alors $\min(f_n, \chi_m)$ est dominée par χ_m et converge vers χ_m pour $n \to \infty$, d'où $\liminf \int f_n \geq \int \lim \min(f_n, \chi_m) = \int \chi_m$, et en faisant tendre m vers l'infini on conclut que $\int f_n \to +\infty$.

En guise d'application du théorème de convergence dominée, voici un théorème simple de continuité des intégrales à paramètre.

Théorème IV-19 (continuité des intégrales à paramètre). *Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, et Z un espace métrique. On se donne $f : X \times Z \to [-\infty, +\infty]$ une fonction telle que* *(i) pour tout* $z \in Z$, $x \longmapsto f(x,z)$ *est mesurable ;*

(ii) pour tout $x \in X$, $z \longmapsto f(x,z)$ *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* $\int g(x)\,\mu(dx) < +\infty$.

Alors l'application $\varphi(z) = \int f(x,z)\,\mu(dx)$ *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 $\varphi(z_n) \to \varphi(z)$. 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). □

Remarque IV-20. Les propriétés de *mesurabilité* et d'*intégrabilité* des intégrales à paramètre, à dépendance pas forcément continue, seront étudiées plus loin dans ce chapitre ; voir le Théorème IV-56.

Voici un corollaire pratique du Théorème IV-19 :

Corollaire IV-21 (Dérivation des intégrales à paramètre). *Soient* $(X,\mathcal{A},\mu)$ *un espace mesuré,* I *un intervalle de* $\mathbb{R}$, *et* $f : X \times I \to \mathbb{R}$. *On suppose que*

(i) Pour tout $t \in I$, $x \longmapsto f(x,t)$ *est mesurable et* $\int |f(x,t)|\,\mu(dx) < +\infty$ *;*

(ii) Pour tout $x \in X$, $t \longmapsto f(x,t)$ *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* $\int g(x)\,\mu(dx) < +\infty$.

Alors $F : t \longmapsto \int f(x,t)\,\mu(dx)$ *est dérivable sur* I, *et pour tout* $t \in I$ *on a*

$$F'(t) = \int_X \frac{\partial f}{\partial t}(x,t)\,\mu(dx).$$

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

$$h(x,s) = \begin{cases} \dfrac{f(x,t-s) - f(x)}{s} & \text{si } 0 < |s| < \varepsilon \\[2ex] \dfrac{\partial f}{\partial t}(x,t) & \text{si } s = 0. \end{cases}$$

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 IV-19, $\lim_{s\to 0} \int h(x,s)\,\mu(dx) = \int h(x,0)\,\mu(dx)$, ce qui équivaut au résultat recherché. □

Exercice IV-22. Énoncer et prouver une variante du Corollaire IV-21 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.

IV-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 IV-11 montrent qu'on ne peut l'éliminer purement et simplement de l'énoncé du Théorème IV-12. Peut-on cependant la remplacer par une hypothèse moins contraignante ? Existe-t-il des situations où la convergence des intégrales est vraie sans qu'il y ait domination ? Voici deux exemples de telles situations, faisant intervenir la mesure de Lebesgue sur $\mathbb{R}$:

EXEMPLES IV-23 (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 (exercice).

(ii) Soit f_n définie sur $\mathbb{R}$ par

$$f_n(x) = \begin{cases} -n & \text{si } -\dfrac{1}{n} \le x < 0; \\ \\ +n & \text{si } 0 < x \le \dfrac{1}{n}; \\ \\ 0 & \text{sinon.} \end{cases}$$

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 $(f_n 1_{x\ge 0})$, et l'intégrale de f_n sur $\mathbb{R}_+$ convergerait vers 0 ; or elle est toujours égale à 1.

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 IV-24 (en l'absence de fortes compensations, la domination est nécessaire pour passer à la limite). *Soit (f_n) une famille de fonctions mesurables, définies sur un espace mesuré $(X, \mathcal{A}, \mu)$, 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 $|f_{n'}| \le g$ presque partout.

(ii) Si

$$\lim_{n\to\infty} \int |f_n| = \int |\lim_{n\to\infty} f_n|,$$

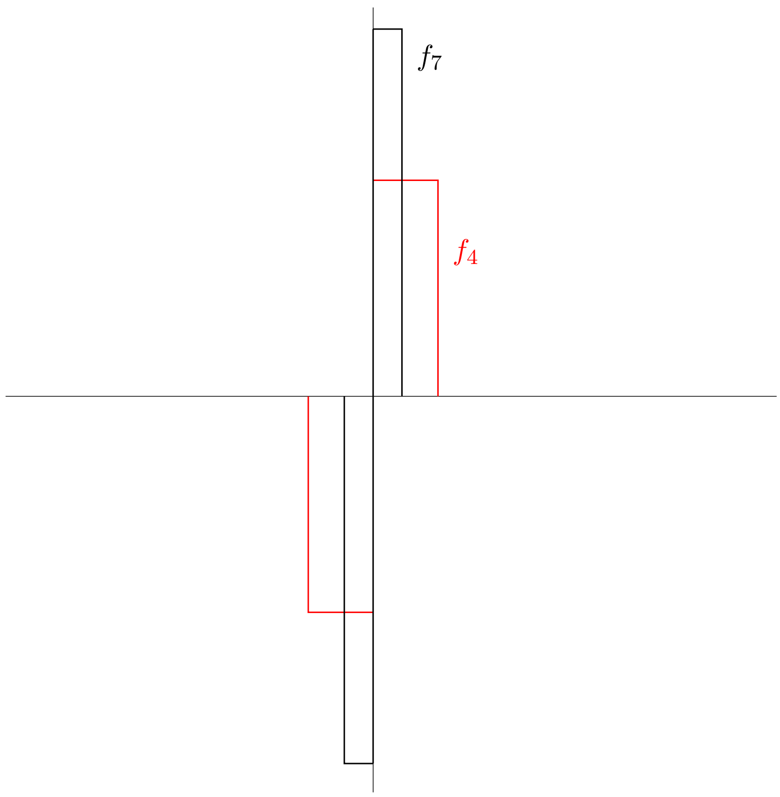

FIGURE 2. compensation entre grandes valeurs positives et négatives

alors il existe une sous-suite extraite de (f_n) *qui vérifie l'hypothèse de domination.*

Il est clair que l'énoncé (ii) découle de (i). On démontrera ce théorème dans la section suivante, comme conséquence du Théorème d'Egorov.

REMARQUE IV-25. L'énoncé (ii) du Théorème IV-24 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 on reparlera au Chapitre VIII.

En combinant les Théorèmes IV-12 et IV-24, on obtient facilement le corollaire suivant.

COROLLAIRE IV-26 (en l'absence de fortes compensations, l'échange limite-somme est quasiment équivalent à la domination). *Soient* $(X, \mathcal{A}, \mu)$ *un espace mesuré, et* $(f_n)_{n\in\mathbb{N}}$ *une famille de fonctions mesurables, définies de* $(X, \mathcal{A}, \mu)$ *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*

$$\text{“} \quad \lim \int f_n \, d\mu = \int (\lim f_n) \, d\mu \quad \text{”}$$

et

“De toute suite extraite $(f_{n'})$ *on peut extraire une sous-suite dominée”*

sont équivalents.

IV-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. Mais du point de vue de la théorie de la mesure, la différence n'est pas si grande, au sens de l'énoncé suivant.

Théorème IV-27 (Théorème d'Egorov). *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, tel que $\mu[X] < +\infty$, 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 IV-28. Un exemple classique de suite qui converge simplement mais non uniformément est la famille des fonctions $f_n : x \longmapsto x^n$ 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 $[0, 1 - \varepsilon]$, si petit que soit $\varepsilon > 0$.

Remarque IV-29. Puisque la convergence uniforme laisse stable la classe des fonctions continues, le théorème d'Egorov admet, dans un cadre topologique, le corollaire suivant :

Corollaire IV-30 (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 $f : X \to \mathbb{R}$. Alors f est continue en-dehors d'un ensemble de mesure arbitrairement petite.*

On retrouve ainsi un énoncé très proche du Théorème III-69 de Lusin. Le Corollaire IV-30 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 IV-31. On ne peut se passer de l'hypothèse de finitude de μ dans le Théorème IV-27 ; 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

$$S_{n,k} := \bigcap_{i,j\geq n} \Big\{x \in X;\ |f_j(x) - f_i(x)| \leq 1/k\Big\}.$$

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 tout $j \geq n_k$, $|f_j(x) - f(x)| \leq 1/k$. En d'autres termes, (f_n) converge uniformément sur S. D'autre part,

$$\mu[X \setminus S] \leq \sum_{k=1}^{\infty} \mu[X \setminus S_{n_k,k}] < \varepsilon\Big(\sum_{k \in \mathbb{N}} 2^{-k}\Big) = \varepsilon.$$

L'ensemble $A_\varepsilon = X \setminus S$ vérifie donc la conclusion du théorème. □

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 IV-24. En fait on aurait pu présenter toute la théorie du passage à la limite en prenant comme point de départ le théorème d'Egorov plutôt que le théorème de convergence monotone.

Nouvelle démonstration du Théorème IV-12. 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 IV-9 :

$$\int f_n\, d\mu = \int h_n\, d\nu, \qquad \int f\, d\mu = \int h\, d\nu,$$

où

$$h_n = \frac{f_n}{g}, \qquad h = \frac{f}{g}, \qquad \nu = g\mu.$$

L'ensemble des points où g s'annulle 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

$$\left| \int_{X \setminus A_\varepsilon} h_n\, d\nu \right| \leq \varepsilon, \qquad \left| \int_{X \setminus A_\varepsilon} h\, d\nu \right| \leq \varepsilon.$$

On en déduit

$$\left| \int h_n\, d\nu - \int h\, d\nu \right| \leq \left| \int_{X \setminus A_\varepsilon} (h_n - h)\, d\nu \right| + 2\varepsilon.$$

Pour tout ε fixé, grâce à la convergence uniforme on a

$$\left| \int_{X \setminus A_\varepsilon} (h_n - h)\, d\nu \right| \leq \left(\sup_{x \in X \setminus A_\varepsilon} |h_n - h| \right) \nu[X] \xrightarrow[n \to \infty]{} 0,$$

donc
$$\limsup_{n\to\infty}\left|\int h_n\,d\nu-\int h\,d\nu\right|\le 2\varepsilon.$$
On conclut en faisant tendre ε vers 0. □

Démonstration du Théorème IV-24. Comme on l'a déjà 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 $\int f_n\to\int f$.

Soit A un ensemble mesurable arbitraire. En appliquant le Lemme de Fatou et l'inégalité $\liminf a_n+\liminf b_n\le\liminf(a_n+b_n)$, laissée en exercice, on a
$$\int_X f=\int_A f+\int_{X\setminus A}f\le\liminf_{n\to\infty}\int_A f_n+\liminf_{n\to\infty}\int_{X\setminus A}f_n\le\liminf_{n\to\infty}\int_X f_n=\int_X f.$$
Les deux membres étant égaux, il y a égalité à chaque étape, d'où
$$\int_A f\,d\mu=\liminf_{n\to\infty}\int_A f_n\,d\mu. \tag{24}$$

Soit $\nu=f\mu$; comme f est sommable, la mesure ν est finie. Pour tout $k\in\mathbb{N}$, on pose
$$B_k=\{x;\ f(x)\le 1/k\}.$$
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 $\nu[B_k]\le\varepsilon$. 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 $n\ge m$ assez grand, on aura
$$x\in X\setminus C\Longrightarrow f_n(x)\le 2f(x).$$

D'après (24), appliqué à $A=C$, on sait que $\liminf\int_A f_n\le 2\varepsilon$. En particulier, on peut trouver $N\ge m$ tel que
$$\int_C f_N\le 4\varepsilon.$$

Récapitulons : pour tout $\varepsilon>0$, pour tout $p\in\mathbb{N}$, nous pouvons construire un ensemble C et un entier $N\ge p$ tels que $\int_{C_\varepsilon}f_N\le 4\varepsilon$, et $f_N\le 2f$ en-dehors de C. On répète cette construction avec $\varepsilon=2^{-k}$: n_k étant donné, on construit $C=C_k$ et $N=n_{k+1}\ge n_k$ tels que
$$\int_{C_k}f_{n_k}\le 4\cdot 2^{-k};\qquad x\in X\setminus C_k\Rightarrow f_{n_k}(x)\le 2f(x).$$
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
$$\int\sum_{k\in\mathbb{N}}f_{n_k}1_{C_k}=\sum_{k\in\mathbb{N}}\int_{C_k}f_{n_k}\le 4\sum_{k\in\mathbb{N}}2^{-k}<+\infty.$$

□

IV-1.6. Les quatre faces de la convergence. Les énoncés de Beppo Levi, Fatou, Lebesgue et Egorov sont finalement très proches les uns des autres ; c'est seulement l'usage qui leur a attribué des statuts différents de lemme ou théorème, et chacun d'entre eux pourrait être choisi comme base dans un exposé sur le passage à la limite dans la théorie de Lebesgue. C'est en fonction de la situation que l'on choisit d'appliquer l'une ou l'autre de ces quatre faces du problème de la convergence des intégrales.

IV-1.7. Formule de sommation par tranches. Voici maintenant une application importante et parlante des résultats de la section précédente. Soit λ la mesure de Lebesgue sur $\mathbb{R}$, définie dans la section II-8. Si f est mesurable, on notera $\{f > t\} = \{x; f(x) > t\}$.

Théorème IV-32 (Formule de sommation par tranches). *Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, et f une fonction mesurable positive ; alors*

$$\begin{aligned}\int_X f(x)\,\mu(dx) &= \int_{\mathbb{R}_+} \mu[\{f > t\}]\,\lambda(dt)\\ &= \lim_{n\to\infty} \sum_{k\in\mathbb{N}} \frac{1}{2^n}\,\mu\left[\left\{x;\ f(x) \geq \frac{k}{2^n}\right\}\right].\end{aligned}$$

Remarque IV-33. Cet énoncé justifie en un sens le dessin de la figure 1 dans l'introduction.

Démonstration du Théorème IV-32. 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 IV-32 sont encore égales.

Soit enfin f une fonction mesurable positive. Par le Théorème III-36, 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

$$\int_X f_n\,d\mu = \int_{\mathbb{R}} \mu[\{f_n > t\}]\,\lambda(dt).$$

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,

$$\mu[\{f > t\}] = \lim_{n\to\infty} \mu[\{f_n > t\}].$$

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

$$\int_{\mathbb{R}} \mu[\{f_n > t\}]\,\lambda(dt) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} \mu[\{f > t\}]\,\lambda(dt).$$

On conclut que

$$\int_X f\,d\mu = \int_{\mathbb{R}} \mu[\{f > t\}]\,\lambda(dt).$$

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 $0 \le \phi_n \le \phi$, et on vérifie facilement que ϕ_n converge en croissant vers ϕ. On peut donc encore appliquer le Théorème de convergence monotone pour obtenir

$$\int_{\mathbb{R}} \phi(t)\,\lambda(dt) = \lim_{n\to\infty} \int_{\mathbb{R}} \phi_n(t)\,\lambda(dt),$$

ce qui revient à

$$\int_{\mathbb{R}} \mu[\{f > t\}]\,\lambda(dt) = \lim_{n\to\infty} \sum_{k\in\mathbb{N}} \frac{1}{2^n}\mu\left[\left\{x;\ f(x) \ge \frac{k}{2^n}\right\}\right].$$

□

La formule de sommation par tranches admet une généralisation importante :

Théorème IV-34 (Sommation par tranches, encore). *Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, f une fonction mesurable positive, et ν une mesure de Borel sur $\mathbb{R}_+$. Pour tout $r \ge 0$, on définit $\Phi(r) = \nu[\,[0, r[\,]$. Alors,*

$$\begin{aligned}\int_X \Phi(f(x))\,\mu(dx) &= \int_{\mathbb{R}} \mu[\{f > t\}]\,\nu(dt)\\ &= \lim_{n\to\infty} \sum_{k\in\mathbb{N}} \nu\,](k-1)2^{-n}, k2^{-n}]\,\mu\left[\left\{x;\ f(x) \ge \frac{k}{2^n}\right\}\right]\end{aligned}$$

Remarque IV-35. On retrouve le Théorème IV-32 via le cas particulier $\nu = \lambda$.

Exemple IV-36. Soit ϕ une fonction positive continue par morceaux sur $\mathbb{R}_+$, et Φ sa primitive (avec $\Phi(0) = 0$). Alors

$$\int_X \Phi(f(x))\,\mu(dx) = \int_{\mathbb{R}} \mu[\{f > t\}]\,\phi(t)\,\lambda(dt).$$

Par exemple,

$$\int_X |f|^p\,d\mu = \int_{\mathbb{R}} \mu[\{f > t\}]\,pt^{p-1}\,dt. \tag{25}$$

Je démontrerai le Théorème IV-34 plus tard. Il est clair qu'il suffit d'établir l'égalité $\int_X \Phi(f(x))\,\mu(dx) = \int_{\mathbb{R}} \mu[\{f > t\}]\,\nu(dt)$; 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 ; c'est la preuve la plus simple. Le cas général, sans hypothèse de σ-finitude, sera ensuite prouvé grâce à un théorème de changement de variables.

IV-2. Intégration sur les espaces produits

La théorie abstraite de l'intégrale de Lebesgue aborde efficacement les intégrales multiples, pourvu que l'on prenne garde à quelques subtilités.

IV-2.1. Rappels et compléments sur la tribu produit.

Définition IV-37 (tribu produit). *Soient $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ deux espaces mesurables. 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, c'est à dire les parties de la forme $A \times B$, où A et B sont des parties mesurables de X et Y respectivement.*

Proposition IV-38 (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,*

$$\sigma(\mathcal{F} \otimes \mathcal{G}) = \sigma(\mathcal{F}) \otimes \sigma(\mathcal{G}).$$

Preuve de la Proposition IV-38. 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 $B \in \sigma(\mathcal{G})$. Donc

$$\sigma(\mathcal{F} \otimes \mathcal{G}) \subset \sigma(\mathcal{F}) \otimes \sigma(\mathcal{G}).$$

C'est l'inclusion réciproque qu'il faut établir. Pour cela, il suffit de montrer que

$$\forall A \in \sigma(\mathcal{F}), \quad \forall B \in \sigma(\mathcal{G}), \qquad A \times B \in \sigma(\mathcal{F} \otimes \mathcal{G}).$$

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 un schéma classique, déjà utilisé dans la preuve du Théorème II-77. On montre dans un premier temps que $A \times B \in \sigma(\mathcal{F} \otimes \mathcal{G})$ pour tous $A \in \sigma(\mathcal{F})$ et $B \in \mathcal{G}$; 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 $\mathcal{F}$. Dans un second temps on montre que $A \times B \in \sigma(\mathcal{F} \otimes \mathcal{G})$ pour tous $A \in \sigma(\mathcal{F})$ et $B \in \sigma(\mathcal{G})$, par un argument similaire. ☐

Dans un cadre abstrait, la tribu produit peut être très difficile à décrire. Mais pour les tribus boréliennes, le problème se simplifie grâce à la proposition suivante.

Proposition IV-39 (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 IV-38 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 O est réunion dénombrable de telles boules : comme dans la preuve du Théorème II-39, il suffit d'écrire

$$O = \bigcup_{B(x_k,1/n)\subset O} B(x_k, 1/n).$$

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. Donc 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. □

Exemple IV-40. $\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n)$.

Remarque IV-41. 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 II-93. Soit d'autre part $\overline{A \otimes B}$ la complétion de la tribu produit $\mathcal{A} \otimes \mathcal{B}$. En général,

$$\overline{\mathcal{A}} \otimes \overline{\mathcal{B}} \neq \overline{(A \otimes B)}.$$

On verra au Chapitre VI que même dans le cas simple où $\mathcal{A} = \mathcal{B}$ est la tribu borélienne sur $[0, 1]$, la complétion de la tribu produit n'est pas identique au produit des tribus complétées (ou tout au moins qu'il est impossible de prouver cette identité).

IV-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 universelle, mais sera bien commode pour préciser les idées.

Définition IV-42 (section). *Soit C un ensemble mesurable dans un espace produit $X \times Y$, muni de la tribu produit. Pour tout $x \in X$, on appelle section (ou coupe, ou tranche) de C en x le long de Y l'ensemble*

$$C_x = \{y \in Y;\ (x, y) \in C\}.$$

Proposition IV-43 (les sections 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 section 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 $\varphi_x : y \longmapsto (x, y)$. 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. Donc $\mathcal{C}$ contient tous les pavés, et partant, toute la tribu produit. □

Remarque IV-44. 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 seraitce que la tribu des ensembles Lebesgue-mesurables dans $\mathbb{R} \times \mathbb{R}$, comme on le verra au chapitre suivant.

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 IV-45 (mesurabilité par rapport à une composante). *Soient X et Y deux espaces mesurables, et $f : X \times Y \to \overline{\mathbb{R}}$ une fonction mesurable pour la tribu produit sur $X \times Y$. Alors, pour tout $x \in X$, la fonction $y \longmapsto f(x,y)$ est mesurable de X dans $\overline{\mathbb{R}}$. Le même résultat reste vrai si $\overline{\mathbb{R}}$ est remplacé par un espace métrique complet.*

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 $y \longmapsto f_n(x,y)$ n'est autre que $\sum \lambda_k 1_{(C_k)_x}$. Par la Proposition IV-43, 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. $\square$

IV-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, et c'est la base de la mesure d'aire dans le plan. On généralise cette démarche à un cadre abstrait avec la notion de pavés.

Théorème IV-46 (mesure produit). *(i) Soient $(X,\mathcal{A},\mu)$ et $(Y,\mathcal{B},\nu)$ deux espaces mesurés, σ-finis. On munit $X \times Y$ de la tribu produit $\mathcal{A} \otimes \mathcal{B}$. Alors il existe une unique mesure θ sur $X \times Y$ telle que*

$$\forall (A,B) \in \mathcal{A} \times \mathcal{B}, \qquad \theta[A \times B] = \mu[A] \times \nu[B].$$

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. $\mathcal{B} = \sigma(\mathcal{G})$), et si 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

$$\forall (A,B) \in \mathcal{F} \times \mathcal{G}, \qquad \theta[A \times B] = \mu[A] \times \nu[B].$$

Remarque IV-47. La notation de produit tensoriel traduit l'idée que les variables $x \in X$ et $y \in Y$ sont "indépendantes". Quand on considère $\mu \otimes \nu$, on conserve toute l'information sur μ et toute l'information sur ν, on les apparie ensemble de façon bilinéaire.

Notation IV-48 (intégrale produit). Soient $(X,\mathcal{A},\mu)$ et $(Y,\mathcal{B},\nu)$ sont deux espaces mesurés, et f une fonction mesurable de $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ dans $\overline{\mathbb{R}}$. Alors, dès que $\int f\, d(\mu \otimes \nu)$ est bien défini, on notera indifféremment

$$\begin{aligned}\int_{X\times Y} f\, d(\mu\otimes\nu) &= \int_{X\times Y} f(x,y)\, d(\mu\otimes\nu)(x,y)\\ &= \iint_{X\times Y} f(x,y)\,(\mu\otimes\nu)(dx\,dy) = \iint_{X\times Y} f(x,y)\,\mu(dx)\,\nu(dy),\end{aligned}$$

la valeur de cette *intégrale produit.* (On peut aussi utiliser le symbole $\iint$ dans les deux premières expressions si l'on souhaite insister sur la nature produit de cette intégrale.)

Démonstration. C'est, comme d'habitude, le Théorème de prolongement II-82 qui permettra de construire la mesure θ. Il va s'agir de prolonger la mesure produit de l'algèbre faite des unions finies de pavés, à la σ-algèbre engendrée par ces pavés.

On définit donc $\mathcal{P}$ comme l'ensemble de tous les pavés $P = A \times B$, où $A \in \mathcal{A}$ et $B \in \mathcal{B}$, et $\theta[P] = \mu[A] \times \nu[B]$. Clairement, $\mathcal{P}$ 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 II-82.

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 II-82 sont donc vérifiées, et il ne reste à vérifier que la σ-additivité de θ sur $\mathcal{P}$.

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 IV-2.3) ; notre but est de prouver que $\theta[P] = \sum \theta[P_k]$.

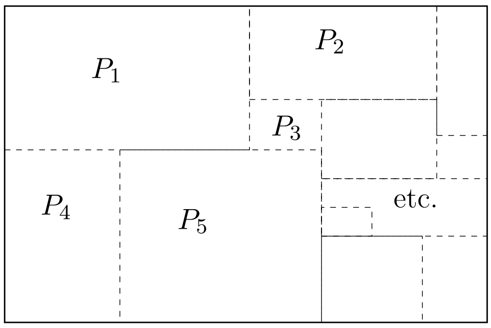

Figure 3. Recouvrement (infini) de P par des pavés P_k

Pour tout k on définit sur X

$$f_k(x) = \nu[B_k]\, 1_{A_k}(x).$$

Clairement, f_k est une fonction mesurable positive, et

$$\int_X f_k(x)\, \mu(dx) = \nu[B_k]\, \mu[A_k] = \theta[P_k]. \tag{26}$$

En outre pour tout x, la fonction $1_{(x,y) \in P_k}$ est clairement mesurable sur B (si $x \in A_k$, c'est la fonction indicatrice de A_k, sinon c'est la fonction nulle), et

$$f_k(x) = \int_Y 1_{B_k}(y)\, 1_{A_k}(x)\, \nu(dy) = \int_Y 1_{P_k}(x,y)\, \nu(dy).$$

Pour tout x fixé, par convergence monotone, appliquée à la mesure ν,

$$\sum_k f_k(x) = \sum_k \int_Y 1_{P_k}(x,y)\, \nu(dy) = \int_Y \sum_k 1_{P_k}(x,y)\, \nu(dy). \tag{27}$$

Puisque chaque (x, y) de P appartient à un et un seul des P_k,

$$\sum_k 1_{P_k} = 1_P,$$

donc (27) devient

$$\sum_k f_k(x) = \int_Y 1_P(x,y)\,\nu(dy) = \int_Y 1_A(x)1_B(y)\,\nu(dy) = \nu[B]\,1_A(x).$$

En appliquant à nouveau la convergence monotone, cette fois pour la mesure μ, on échange à nouveau somme et intégrale :

$$\sum_k \int_X f_k\,d\mu = \int_X \sum_k f_k\,d\mu = \int \nu[B]1_A\,d\mu = \nu[B]\,\mu[A].$$

En appliquant (26) et la définition de la mesure produit, cette dernière égalité devient

$$\sum_k \theta[P_k] = \theta[P],$$

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 IV-38, et que $X \times Y$ est réunion croissante d'une suite d'éléments de cette famille. On peut alors appliquer le Théorème II-82 (ii) pour conclure à l'unicité d'une mesure satisfaisant aux conditions requises. $\square$

Remarque IV-49. 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

$$\Big(\sum_k \mu[A'_k]\Big)\Big(\sum_\ell \nu[B'_\ell]\Big).$$

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 IV-50. 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, il est conforme à l'intuition de multiplier les probabilités respectives de ces deux événements. Et finalement c'est la définition de l'indépendance : si A et B sont deux ensembles mesurables (appelés événements) dans l'espace Ω des possibles, et $\mathbb{P}$ une mesure de probabilité, on dira que A et B sont indépendants si $\mathbb{P}[A\cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$. Dans le cas où $\mathbb{P}[B] > 0$, cela se reforme de façon encore plus parlante avec la notion de probabilité conditionnelle : la probabilité que A se réalise "sachant que B est vrai" (ou simplement "sachant B") vaut $\mathbb{P}[A \cap B]/\mathbb{P}[B]$, dite probabilité de A conditionnée à B. Avec cette notion, dire que A et B sont indépendants, c'est dire que la probabilité de A conditionnée B est égale à la probabilité de A ; et aussi, si $\mathbb{P}[\Omega \setminus B] > 0$, à la probabilité de A conditionnée au complémentaire de B. Autrement dit, la probabilité de A reste la même que B soit vrai ou pas, la même indépendamment de la réalisation de B. Et plus généralement, en probabilité on dit que deux fonctions mesurables (deux variables aléatoires) f et g sur un espace de probabilité $(X, \mathcal{C}, \pi)$ sont indépendantes si $(f,g)_\#\pi = (f_\#\pi) \otimes (g_\#\pi)$.

Exemple IV-51. 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*. On reviendra par la suite sur les propriétés de cette mesure et de ses analogues en dimension plus grande.

IV-2.4. Généralisation : mesures dépendant d'un paramètre.

Proposition IV-52 (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 $x \longmapsto \nu_x$ est mesurable, au sens où pour tout $B \subset Y$ l'application*

$$x \longmapsto \nu_x[B]$$

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

$$(\mu \otimes \nu_x)[A] = \int_X \nu_x[A_x]\,\mu(dx).$$

Remarque IV-53. Dans cette notation la variable x au membre de gauche est formelle. Il serait plus correct mais moins parlant de noter $\mu \otimes \nu_{\cdot}$ cette mesure.

Preuve de la Proposition IV-52. Il y a deux choses à vérifier : (i) que la fonction $x \longmapsto \nu_x[A_x]$ 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 sections $(Y_k)_x$ disjointes, et

$$\nu_x[A_x] = \nu_x\big[\cup_k (A\cap Y_k)_x\big] = \sum_k \nu_x[(A\cap Y_k)_x].$$

Il suffit donc de vérifier que chaque application $x \longmapsto \nu_x[(A\cap Y_k)_x]$ 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 sections distinctes A^1_x et A^2_x le long de Y, pour chaque x, d'où $\nu_x[(A^1 \cup A^2)_x] = \nu_x[A^1_x] + \nu_x[A^2_x]$. 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 (Théorème II-77) 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 sections A^n_x sont disjointes, d'où

$$\int_X \nu_x[(\cup A^n)_x]\,\mu(dx) = \int_X \sum_x \nu_x[A^n_x]\,\mu(dx) = \sum_n \int_X \nu_x[A^n]\,\mu(dx).$$

□

Remarque IV-54. Si $\mu[A] = 0$ alors $(\mu\otimes\nu_x)[A\times Y] = 0$; je montrerai dans la Section **??** que sous certaines hypothèses peu contraignantes, cette propriété caractérise les mesures qui peuvent s'écrire sous la forme $\mu\otimes\nu_x$.

Exemple IV-55. 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 qui sont rigoureusement 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 $\nu_x(dy) = f(x,y)\,\lambda(dy)$.

IV-2.5. Théorème de Fubini–Tonelli–Lebesgue. On nomme théorème de Fubini, de façon générique, tout énoncé permettant d'échanger des opérations d'intégration, ou plus généralement de définir des intégrales multiples, que ce soit dans la théorie de Riemann, dans celle de Lebesgue ou dans une autre. On utilise parfois le nom de théorème de Tonelli quand on considère des fonctions positives mesurables. Accoler les trois noms de Fubini, Tonelli et Lebesgue est donc le plus juste pour le présent cours ; en pratique et par commodité, on dit le plus souvent “théorème de Fubini”.

Théorème IV-56 (Théorème de Fubini–Tonelli–Lebesgue). *Soient $(X,\mathcal{A},\mu)$ et $(Y,\mathcal{B},\nu)$ deux espaces mesurés, σ-finis. On munit $X\times Y$ de la tribu produit $\mathcal{A}\otimes\mathcal{B}$. Alors*

(i) Pour toute fonction mesurable f définie sur $X\times Y$, à valeurs dans $[0,+\infty]$, les fonctions

$$x\longmapsto \int_Y f(x,y)\,\nu(dy), \qquad y\longmapsto \int_X f(x,y)\,\mu(dx)$$

sont mesurables sur X et Y respectivement. En outre,

$$\iint_{X\times Y} f(x,y)\,(\mu\otimes\nu)(dx\,dy) = \int_X\left(\int_Y f(x,y)\,\nu(dy)\right)\mu(dx) = \int_Y\left(\int_X f(x,y)\,\mu(dx)\right)\nu(dy).$$

(ii) Soit f une fonction mesurable définie sur $X\times Y$, à valeurs dans $\overline{\mathbb{R}}$. Si

$$\iint_{X\times Y} |f(x,y)|\,(\mu\otimes\nu)(dx\,dy) < +\infty,$$

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

$$\varphi : y\longmapsto \int_X f(x,y)\,\mu(dx)$$

est alors ν-sommable sur l'ensemble S_Y des y tels que $f(\cdot, y)$ est μ-sommable ; et la fonction

$$\psi : x \longmapsto \int_Y f(x,y)\,\nu(dy)$$

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

$$\begin{aligned} \iint_{X\times Y} f(x,y)\,(\mu\otimes\nu)(dx\,dy) &= \iint_X \left(\int_Y f(x,y)\,\nu(dy)\right)\,\mu(dx) \\ &= \int_Y \left(\int_X f(x,y)\,\mu(dx)\right)\,\nu(dy). \end{aligned} \tag{28}$$

Remarque IV-57. 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 déjà vus en section IV-2 comme corollaire des théorèmes de convergence monotone, et de convergence dominée. Ce n'est cependant pas vraiment une nouvelle démonstration car la convergence monotone joue un rôle clé dans la construction de l'intégrale produit.

Démonstration du Théorème IV-56. Il est 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

$$x \longmapsto \int_Y |f(x,y)|\,\nu(dy), \qquad y \longmapsto \int_X |f(x,y)|\,\mu(dx)$$

sont sommables ; en particulier, elles sont finies presque partout, donc 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

$$x \longmapsto \int_X |f(x,y)|\,\nu(dy)$$

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

$$\left|\int_Y f(x,y)\,\nu(dy)\right| \leq \int_Y |f(x,y)|\,\nu(dy)$$

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 (28) 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 en est assez laborieuse et utilise des schémas déjà rencontrés : remplacer les fonctions mesurables par des fonctions indicatrices, remplacer les ensembles mesurables par des pavés. On va 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 $x \longmapsto \int f(x,y)\,\nu(dy)$ 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

$$\int_Y 1_{\cup A_n}(x,y)\,\nu(dy) = \lim_{k\to\infty} \int_Y 1_{A_k}(x,y)\,\nu(dy),$$

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 (Théorème II-77). 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 IV-58. Dans l'énoncé, j'ai 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 sont définies qu'en-dehors d'un ensemble négligeable. Dans le contexte présent, peu importe, tant qu'on a les idées claires ! Dans d'autres situations, travailler avec des fonctions définies partout (et pas seulement presque partout) peut éviter certaines confusions.

Voici maintenant quelques remarques sur le Théorème de Fubini–Tonelli–Lebesgue, que l'on pourra illustrer grâce à la mesure de Lebesgue λ sur $\mathbb{R}$.

Remarque IV-59. La σ-finitude de X et Y est une hypothèse importante dans le Théorème IV-56. Un contre-exemple classique consiste à considérer la mesure de Lebesgue λ sur $[0,1]$ d'une part, la mesure de comptage C 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

$$\forall y, \quad \int 1_\Delta(x,y)\,\lambda(dx) = 0; \qquad \forall x, \quad \int 1_\Delta(x,y)\,C(dy) = 1.$$

En particulier,

$$\int\int 1_\Delta(x,y)\,\lambda(dx)\,C(dy) = 0; \qquad \int\int 1_\Delta(x,y)\,C(dy)\,\lambda(dx) = 1.$$

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 $\lambda \otimes C$ soit bien définie ; est bien définie en revanche la mesure extérieure $(\lambda \otimes C)^*$ associée aux recouvrements par des pavés. En l'occurrence, on se convainc facilement que $(\lambda \otimes C)^*[\Delta] = +\infty$.

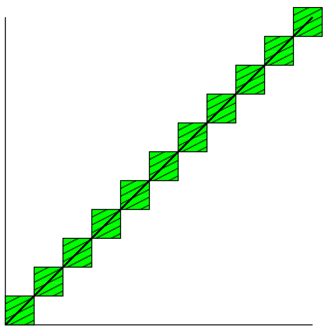

FIGURE 4. La diagonale est limite d'une union de petits carrés

REMARQUE IV-60. Il est également important que la fonction f soit *mesurable pour la tribu produit* ! Un contre-exemple surprenant dû à Sierpiński [Rudin, p. 167], sous hypothèse d'axiome du choix, montre que les quantités

$$\int_{[0,1]} \left(\int_{[0,1]} f(x,y)\,\lambda(dy) \right) \lambda(dx); \qquad \int_{[0,1]} \left(\int_{[0,1]} f(x,y)\,\lambda(dx) \right) \lambda(dy)$$

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 $f : \mathbb{R}^2 \to \mathbb{R}$ est telle que les applications partielles $f(x,\cdot)$ et $f(\cdot,y)$ sont respectivement Borel-mesurables en y pour tout x, et *continues* en x pour tout y, alors f est automatiquement Borel-mesurable [Rudin, p. 176].

Avec le Théorème de Fubini, on peut démontrer la formule de sommation par tranches généralisée qui avait été annoncée dans la Section IV-1.7, du moins sous hypothèse de σ-finitude :

Démonstration du Théorème IV-34 quand X est σ-fini. Puisque $\Phi(f) = \nu[[0,f[]$, on peut écrire, en utilisant Fubini-Tonelli-Lebesgue,

$$\begin{aligned}\int_X \Phi(f(x))\,d\mu(x) &= \int_X \int_{\mathbb{R}} 1_{[0,f(x)[}(t)\,\nu(dt)\,d\mu(x)\\ &= \int_{\mathbb{R}} \left(\int_X 1_{[0,f(x)[}(t)\,d\mu(x)\right)\nu(dt)\\ &= \int_{\mathbb{R}} \mu[\{f>t\}]\,\nu(dt).\end{aligned}$$

□

On verra plus tard (Section IV-3.3) comment se passer aussi de l'hypothèse de σ-finitude.

IV-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 IV-61 (produits multiples et intégrales multiples). *(i) Soient* $(X_1, \mathcal{A}_1, \mu_1)$, *...,* $(X_n, \mathcal{A}_n, \mu_n)$ *des espaces mesurés. Alors la tribu* $(((\mathcal{A}_1 \otimes \mathcal{A}_2) \otimes \mathcal{A}_3) \ldots \otimes \mathcal{A}_n)$ *est la tribu engendrée par les pavés multiples, de la forme* $A_1 \times \ldots \times A_n$*, où* $A_i \in \mathcal{A}_i$ *pour tout* i*. On l'appelle tribu produit de* $\mathcal{A}_1, \ldots, \mathcal{A}_n$ *et on la note*

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \ldots \otimes \mathcal{A}_n.$$

(ii) Soient $X_1, \ldots, X_n$ *des ensembles quelconques, et* $\mathcal{F}_1, \ldots, \mathcal{F}_n$ *des familles de parties de* $X_1, \ldots, X_n$ *respectivement. On suppose que* X_i *est réunion dénombrable d'éléments de* $\mathcal{F}_i$*, pour tout* $i \in \{1, \ldots, n\}$*. Alors la tribu produit* $\sigma(\mathcal{F}_1) \otimes \sigma(\mathcal{F}_2) \ldots \otimes \sigma(\mathcal{F}_n)$ *est engendrée par les pavés de la forme* $A_1 \times \ldots \times A_n$*, où* $A_i \in \mathcal{F}_i$ *pour tout* i*.*

(iii) Soient $X_1, \ldots, X_n$ *des espaces métriques séparables, munis de leurs tribus boréliennes respectives. Alors la tribu produit sur* $X_1 \times \ldots \times X_n$ *coïncide avec la tribu borélienne sur* $X_1 \times \ldots \times X_n$*.*

(iv) Soient $(X_1, \mathcal{A}_1), \ldots, (X_n, \mathcal{A}_n)$ *des espaces mesurables, et soit* $A \subset \prod X_i$ *un ensemble mesurable pour la tribu produit. Alors pour tout* k *les* $(n-k)$*-sections*

$$\{(x_1, \ldots, x_{n-k}) \in X_1 \times \ldots \times X_{n-k};\ (x_1, \ldots, x_n) \in A\}$$

sont mesurables pour la tribu produit $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_{n-k}$*.*

(iv) Soient $(X_1, \mathcal{A}_1, \mu_1), \ldots, (X_n, \mathcal{A}_n, \mu_n)$ *des espaces mesurés* σ*-finis. Alors sur* $X_1 \times \ldots \times X_n$*, muni de la tribu produit, il existe une unique mesure* μ *telle que pour tout pavé* $P = A_1 \times \ldots \times A_n$*, où* $A_i \in \mathcal{A}_i$ *pour tout* i*,*

$$\mu[P] = \prod_{i=1}^n \mu_i[A_i].$$

Cette mesure coïncide avec $((\mu_1 \otimes \mu_2) \otimes \ldots) \otimes \mu_n$ *; on l'appelle mesure produit de* $\mu_1, \ldots, \mu_n$ *et on la note*

$$\mu_1 \otimes \mu_2 \otimes \ldots \otimes \mu_n.$$

Si $\mathcal{F}_1, \ldots, \mathcal{F}_n$ sont des familles de parties de $X_1, \ldots, X_n$ 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

$$\forall i,\ A_i \in \mathcal{F}_i \Longrightarrow \mu[A_1 \times \ldots \times A_n] = \prod_{i=1}^{n} \mu_i[A_i].$$

(v) Soient $(X_0, \mathcal{A}_0), \ldots (X_n, \mathcal{A}_n)_{n\in\mathbb{N}}$ des espaces mesurables. On se donne une mesure μ_0 sur X_0 ; et pour tout $j \in \{1, \ldots, n-1\}$ 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,

$$\mu[\prod A_i] = \int_{A_0} \int_{A_1} \cdots \int_{A_{n-1}} \int_{A_n} \nu_{x_{n-1}}(dx_n)\, \nu_{x_{n-2}}(dx_{n-1}) \ldots \nu_{x_0}(dx_1)\, \mu(dx_0).$$

On la note $\mu_0 \otimes \nu_{x_0} \otimes \ldots \otimes \nu_{x_{n-1}}$ (étant entendu que dans cette notation les x_i sont des symboles formels rappelant juste la dépendance en la variable).

(vi) (Fubini pour des intégrales multiples) Soient $(X_1, \mathcal{A}_1, \mu_1), \ldots, (X_n, \mathcal{A}_n, \mu_n)$ des espaces mesurés σ-finis. On munit $X^k := X_1 \times \ldots \times X_k$ de la tribu produit $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_k$. Alors, pour tout $k \in \{1, \ldots, n-1\}$ et toute fonction mesurable f définie sur X^n, à valeurs dans $[0, +\infty]$, la fonction

$$(x_{k+1}, \ldots, x_n) \longmapsto \int_{X_1\times\ldots\times X_k} f(x_1, \ldots, x_n)\, d(\mu_1 \otimes \ldots \otimes \mu_k)(x_1, \ldots, x_k)$$

est mesurable sur $X_{k+1} \times \ldots \times X_n$. En outre

$$\int_{X^n} f(x_1, \ldots, x_n)\, d(\mu_1 \otimes \ldots \otimes \mu_n)(x_1, \ldots, x_n) = \int_{X_n} \cdots \int_{X_1} f(x_1, \ldots, x_n)\, \mu_1(dx_1) \ldots \mu_n(dx_n),$$

où le membre de droite peut être vu soit comme une suite d'intégrations successives par rapport aux mesures μ_i, soit comme une seule intégration par rapport à la mesure $\mu_1 \otimes \ldots \otimes \mu_n$; le résultat peut aussi être dénoté par

$$\int \cdots \int_{X^n} f(x_1, \ldots, x_n)\, \mu_1 \otimes \ldots \otimes \mu_n(dx_1 \ldots dx_n)$$

En outre, si f est une fonction mesurable définie sur X^n, à valeurs dans $\overline{\mathbb{R}}$, telle que

$$\int_{X^n} |f(x_1, \ldots, x_n)|\, d(\mu_1 \otimes \ldots \otimes \mu_n)(x_1, \ldots, x_n) < +\infty,$$

alors pour chaque k la fonction

$$(x_{k+1}, \ldots, x_n) \longmapsto \int_{X_1\times\ldots\times X_k} f(x_1, \ldots, x_n)\, d(\mu_1 \otimes \ldots \otimes \mu_k)(x_1, \ldots, x_k)$$

est bien définie et sommable hors d'un ensemble négligeable Z_k ; quitte à la redéfinir arbitrairement sur Z_k, on a

$$\int_{X^n} f(x_1, \ldots, x_n)\, d(\mu_1 \otimes \ldots \otimes \mu_n)(x_1, \ldots, x_n) = \int_{X_n} \cdots \int_{X_1} f(x_1, \ldots, x_n)\, \mu_1(dx_1) \ldots \mu_n(dx_n)$$

$$= \int_{X_{\sigma(n)}} \cdots \int_{X_{\sigma(1)}} f(x_1, \ldots, x_n)\, \mu_{\sigma(1)}(dx_{\sigma(1)}) \ldots \mu_{\sigma(n)}(dx_{\sigma(n)})$$

pour toute permutation σ de $\{1,\dots,n\}$.

Exemple IV-62. On peut définir la mesure de Lebesgue en dimension n par $\lambda_n = \lambda^{\otimes n}$.

Remarque IV-63. On verra en fin de chapitre que l'on peut, sous certaines conditions, définir aussi des *produits infinis* de mesures. Cette opération n'est pas toujours permise, ainsi le produit $\lambda^{\otimes\infty}$ n'a pas de sens.

IV-3. Changement de variable

Le changement de variable est le remplacement d'un espace d'intégration par un autre. Les théorèmes classiques de changement de variable s'écrivent dans un cadre différentiable : $\mathbb{R}^n$, ou un ouvert de $\mathbb{R}^n$, ou une variété riemannienne. L'un d'entre eux dit que si φ est un C^1-difféomorphisme entre ouverts O et U de $\mathbb{R}^n$, alors

$$\int_U f(y)\,dy = \int_O f(\varphi(x))\,|\det\nabla\varphi(x)|\,dx,$$

où dx désigne la mesure de Lebesgue dans la variable x, et $\nabla\varphi$ est la matrice jacobienne de φ. Cette formule permet de passer d'une intégrale dans la variable y à une intégrale dans la variable x, où $y = \varphi(x)$.

Mais le changement de variable peut aussi se formuler dans le cadre bien plus général des espaces mesurés et des fonctions mesurables. Il ne sera plus question alors de difféomorphisme ou de déterminant jacobien, qui n'ont pas forcément de sens. C'est donc une formule bien plus abstraite qui sera au cœur de cette section, basée sur la notion importante de **mesure image**. On verra plus tard comment faire le lien avec les formules classiques de changement de variable dans $\mathbb{R}^n$.

IV-3.1. Image d'une mesure par une fonction mesurable. La proposition qui suit se contente de rappeler une notion introduite dans la Remarque III-3(i).

Proposition IV-64 (Tribu image). *Soient $(X,\mathcal{A})$ un espace mesurable, Y un ensemble quelconque, et $f : X \to Y$. On peut définir une tribu, notée $f_\#\mathcal{A}$ (ou $f\#\mathcal{A}$, ou $f_*\mathcal{A}$, ou $f\mathcal{A}$) sur Y, par*

$$f_\#\mathcal{A} = \left\{B \subset Y;\ f^{-1}(B) \in \mathcal{A}\right\}.$$

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 si f est une application mesurable, alors $\mathcal{B} \subset f_\#\mathcal{A}$.

Définition IV-65 (Mesure image). *Soit $(X,\mathcal{A},\mu)$ un espace mesuré, et $f : X \to Y$. 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 $f_*\mu$).

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.

REMARQUE IV-66. On rencontre parfois la notation $f\mu$ pour la mesure image de f par μ, mais il y a alors risque de confusion avec la notion très différente de mesure de densité f par rapport à μ.

IV-3.2. Théorème de changement de variable.

THÉORÈME IV-67. *Soient* $(X, \mathcal{A})$ *et* $(Y, \mathcal{B})$ *deux espaces mesurables, et* $\varphi : X \to Y$ *une application mesurable. Soit* μ *une mesure sur l'espace mesurable* X*. Alors*

(i) Pour toute fonction f *mesurable sur* Y*, à valeurs dans* $[0, +\infty]$*,*

$$\int f\, d(\varphi_\# \mu) = \int (f \circ \varphi)\, d\mu. \tag{29}$$

(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 (i) implique (ii) ; on va donc se contenter de démontrer (i). Si f est une fonction simple, l'égalité (29) découle de la définition de $\varphi_\# \mu$. En effet, quand B est une partie mesurable, et que f est la fonction indicatrice de B, alors les deux membres de (29) se ramènent à $\mu[f^{-1}(B)]$.

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 passe à la limite par le Théorème de convergence monotone de Beppo Levi (Théorème IV-1). □

REMARQUE IV-68. Il peut se produire que $f \circ \varphi$ soit mesurable sans que f le soit. Par exemple c'est le cas, dès que $\varphi(X)$ n'est pas mesurable, pour la fonction $f = 1_{\varphi(X)}$.

IV-3.3. Morphismes d'espaces mesurés. La formule de changement de variables vue précédemment ne suppose aucune régularité et s'applique donc dans des problèmes théoriques abstraits.

Soit la situation où $(X, \mathcal{A}, \mu)$ est un espace mesuré, φ une application $X \to Y$, 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 parle de fonction "bimesurable"), alors f^{-1} réalisera également un morphisme entre Y et X, et les énoncés de théorie de la mesure faisant intervenir $(X, \mathcal{A}, \mu)$ seront équivalents aux énoncés correspondants faisant intervenir $(Y, f_\# \mathcal{A}, f\#\mu)$. 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 ; c'est tout simplement un changement de variable abstrait.

A titre d'exemple, voici un surprenant résultat de classification selon lequel tout espace polonais est isomorphe à $[0, 1]$.

THÉORÈME IV-69 (représentation des espaces polonais). *Soit* $(X, \mathcal{A}, \mu)$ *un espace polonais muni d'une mesure de Borel finie. Soit* I *l'intervalle* $[0, 1]$ *muni de sa tribu borélienne* $\mathcal{B}$*. Alors* *(i) Il existe une mesure finie λ sur I, et une application mesurable $f : I \to X$ qui réalise un morphisme entre $(I, \mathcal{B}, \lambda)$ et $(X, \mathcal{A}, \mu)$. Autrement dit, toute mesure finie sur un espace 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 espace polonais est isomorphe à une mesure finie sans atome sur $[0,1]$.

Remarque IV-70. On rappelle que la réciproque d'une bijection mesurable entre espaces polonais est automatiquement mesurable (Théorème III-24).

Exemple IV-71. Comme on le verra au Chapitre VI, 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) ν obtenue par produit tensoriel dénombrable de la mesure de Bernoulli sur $\{0,1\}$, i.e. la mesure qui attribue un poids identique $1/2$ à $\{0\}$ et à $\{1\}$:

$$\nu = \left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right)^{\otimes \mathbb{N}}.$$

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 tous ses points, il y en a une infinité non dénombrable). On voit sur cet exemple que la théorie de la mesure selon Lebesgue est insensible à la topologie.

On va maintenant appliquer le théorème de changement de variable pour prouver le Théorème IV-34 dans le cas général (rappelons que ce théorème a déjà été démontré dans le cas où X est σ-fini à l'aide du Théorème de Fubini).

Démonstration du Théorème IV-34. Appliquons le Théorème IV-32 à la fonction positive $\Phi \circ f$: ainsi

$$\int_X \Phi \circ f \, d\mu = \int_{\mathbb{R}_+} \mu[\{\Phi \circ f > t\}] \, \lambda(dt).$$

De par sa définition, la fonction Φ est croissante et continue à gauche (en effet, $\nu[0,x[\, = \lim_{k\to\infty} \nu[0, x - k^{-1}[$. On définit son inverse généralisé par la formule

$$\Phi^{-1}(t) := \inf\big\{s \geq 0; \quad \Phi(s) > t\big\}.$$

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

$$\int_{\mathbb{R}} \mu[\{\Phi \circ f > t\}] \, \lambda(dt) = \int_{\mathbb{R}} \mu[\{f > \Phi^{-1}(t)\}] \, \lambda(dt) = \int_{\mathbb{R}} \mu[\{f > s\}] \, \big[(\Phi^{-1})_{\#}\lambda\big](ds).$$

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 $\Phi(s)$. $\square$

IV-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. Les trois inégalités fondamentales, valables en toute généralité, sont les inégalités de **Tchebychev**, **Jensen** et **Hölder**. Deux autres inégalités viennent compléter le tableau : les inégalités de **Young** intégrées, plus générales que celle de Hölder ; et les inégalités de **Minkowski**, qui joueront un rôle majeur au Chapitre VIII.

Toutes ces inégalités, pour l'essentiel antérieures à la théorie de Lebesgue, ont été découvertes, redécouvertes et améliorées par des analystes et statisticiens actifs durant la seconde moitié du dix-neuvième siècle : Viktor Bouniakovski, Hermann Amandus Schwarz, Leonard James Rogers, Otto Hölder, Johan Jensen, Pafnouti Tchebychev, William Henry Young, Andrey Andreyevitch Markov, Irénée-Jules Bienaymé, Hermann Minkowski... La paternité est particulièrement brouillée : par exempe, l'inégalité de Jensen est d'abord établie par Hölder, et l'inégalité de Hölder par Rogers... En outre elles sont toutes étroitement liées, et relèvent d'une même philosophie : utiliser la convexité pour borner une intégrale faisant intervenir un produit de deux fonctions, par des intégrales faisant intervenir chaque fonction séparément. Elles s'appliquent pareillement à des sommes discrètes (et dans ce cadre remontent au moins à Augustin-Louis Cauchy) et à toute notion acceptable d'intégrale ; de sorte que la théorie de Lebesgue n'a eu aucun souci à les inclure quand elle s'est développée.

Avec le développement de la théorie de l'information, sont venues s'ajouter à la liste les inégalités intégrales **entropiques**, qui sont à la fois un cas particulier des inégalités de Young et un cas limite des inégalités de Hölder ;

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.

IV-4.1. Inégalité de Thebychev. L'inégalité de Tchebychev (un nom que l'on orthographie de multiples autres manières, comme Chebisheff) est aussi élémentaire qu'utile, particulièrement dans le domaine des probabilités.

Théorème IV-72 (inégalité de Tchebychev). *(i) Soient $(X,\mathcal{A},\mu)$ un espace mesuré et $f : X \to \mathbb{R}_+$ une fonction mesurable positive. Alors, pour tout $a > 0$,*

$$\mu\Big[\{x\in X;\ f(x)\geq a\}\Big] \leq \frac{1}{a}\int_X f\,d\mu. \tag{30}$$

(ii) Soient $(X,\mathcal{A},\mu)$ un espace mesuré, $f : X \to \mathbb{R}_+$ une fonction mesurable positive, et $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction mesurable croissante. Alors, pour tout $a \geq 0$,

$$\mu\Big[\{x\in X;\ f(x)\geq a\}\Big] \leq \frac{1}{\Phi(a)}\int_X \Phi(f(x))\,\mu(dx).$$

Remarques IV-73. (a) Dans le cas dégénéré où $a = 0$ et f est nulle presque partout, l'inégalité (30) est a priori fausse (sous la convention habituelle $0/0 = 0$).

(b) L'énoncé (i) est souvent appelé inégalité de Markov ; l'énoncé (ii) est souvent appelé inégalité de Bienaymé–Tchebychev quand $\Phi(r) = r^2$, et inégalité de Tchebychev exponentielle quand $\Phi(r) = e^{\alpha r}$. 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 $a \to \infty$. En fait, dans la plupart des situations concrètes, on obtient des estimations de décroissance presque optimales par un choix convenable de Φ.

(c) En corollaire de l'inégalité de Tchebychev, si $\int |f|\, d\mu < +\infty$ on a

$$\forall \delta > 0 \qquad \mu[\{|f| \geq \delta\}] < +\infty. \tag{31}$$

Quand f vérifie (31) on dit parfois que f "s'annule à l'infini". La même conclusion est vraie si $\int \Phi(f)\, d\mu < +\infty$ pour Φ une fonction $\mathbb{R} \to \mathbb{R}_+$ dont les restrictions à $\mathbb{R}_+$ et $\mathbb{R}_-$ sont strictement croissantes ; par exemple si $\int |f|^p\, d\mu < +\infty$ pour un certain $p \in]0, +\infty[$.

Démonstration du Théorème IV-72. Posons $A := \{x \in X;\ f(x) \geq a\}$. Comme f est positive, on a

$$f \geq a 1_A. \tag{32}$$

L'ensemble A est mesurable puisque f l'est ; donc $a1_A$ est étagée, et son intégrale est $a\mu[A]$. La définition même de l'intégrale implique donc $\int f\, d\mu \geq a\mu[A]$, d'où (i).

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,

$$\{x;\ f(x) \geq a\} \subset \Big\{x;\ \Phi(f(x)) \geq \Phi(a)\Big\}.$$

□

Remarque IV-74. Il est facile de vérifier que l'énoncé est en général faux si f n'est pas positive ! (exercice) Dans la pratique, on cherchera donc toujours à se ramener à des fonctions f positives, 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.

IV-4.2. Inégalité de Jensen.

Théorème IV-75 (inégalité de Jensen dans $\mathbb{R}^n$). *Soient $(X, \mathcal{A})$ un espace mesuré équipé d'une mesure de probabilité μ, $f : X \to \mathbb{R}^n$ une fonction mesurable dont chaque composante est μ-sommable, et $\Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe semi-continue inférieurement. 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 IV-76 (inégalité de Jensen pour des puissances). *Soient* $(X, \mathcal{A})$ *un espace mesurable,* μ *une mesure de probabilité sur* X*,* $f : X \to \mathbb{R}^n$ *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 IV-77. (i) Si Φ est continue à valeurs réelles, elle est automatiquement continue, l'hypothèse de semi-continuité inférieure dans le Théorème IV-75 devient donc superflue.

(ii) 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 $X = \{1, \ldots, N\}$, $\mu = \sum \lambda_i \delta_i$ et $f(i) = x_i$, l'inégalité de Jensen se réduit à l'inégalité

$$\sum_{i=1}^N \lambda_i = 1 \Longrightarrow \quad \Phi\bigl(\sum \lambda_i x_i\bigr) \leq \sum \lambda_i \Phi(x_i),$$

que l'on peut également adopter comme définition de la convexité. L'inégalité de Jensen **n'est donc qu'une "version continue" ou "limite continue"** de l'inégalité ci-dessus.

(iii) 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 **??**. 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 $y_1, \ldots, y_k \in \mathbb{R}^n$, 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,

$$\Phi\left(\int f\,d\mu\right) = \Phi\bigl(\sum \alpha_k y_k\bigr) \leq \sum \alpha_k\,\Phi(y_k) = \int \Phi \circ f\,d\mu.$$

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

$$\int g^{(\ell)}\,d\mu \xrightarrow[\ell \to \infty]{} \int f\,d\mu,$$

donc

$$\Phi\left(\int g^{(\ell)}\,d\mu\right) \xrightarrow[\ell \to \infty]{} \Phi\left(\int f\,d\mu\right);$$

et par lipschitzianité de Φ,

$$\left|\int \Phi(f(x))\,\mu(dx) - \int \Phi(g^{(\ell)}(x))\,\mu(dx)\right| \leq \int |\Phi(f(x)) - \Phi(g^{(\ell)}(x))|\,\mu(dx)$$
$$\leq \|\Phi\|_{\mathrm{Lip}} \int |f(x) - g^{(\ell)}(x)|\,\mu(dx) \xrightarrow[\ell\to\infty]{} 0.$$

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 devrons 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

$$\Phi(\xi) \leq \int \Phi\,d\mu, \qquad \xi = \int x\,\mu(dx).$$

(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 IV-79 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 **??**). 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

$$\int \Phi\,d\mu - \Phi(\xi) - \int \langle y, z-\xi\rangle\,\mu(dz) = \left(\int \Phi\,d\mu - \Phi(\xi)\right) - \int \langle y, \xi-\xi\rangle$$
$$= 0 - 0 = 0,$$

elle est forcément nulle. On conclut que

$$\Phi(z) = \Phi(\xi) + \langle y, z-\xi\rangle,$$

μ-presque partout, ce qui démontre la conclusion. □

Remarque IV-78. 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 IV-79. *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 $\int |x|\,\mu(dx) < +\infty$. On note E l'espace affine euclidien engendré par le support de μ, et $\xi = \int x\,\mu(dx)$ 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 **??**) 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. □

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

$$\forall \lambda > 0, \quad \forall x \in \mathbb{R}^n, \qquad \Phi(\lambda x) = \lambda \Phi(x). \tag{33}$$

Théorème IV-80 (inégalité de Jensen pour des fonctions convexes 1-homogènes). *Soient X un espace mesuré, μ une mesure sur X, $f : X \to \mathbb{R}^n$ une fonction mesurable dont chaque composante est μ-sommable, et $\Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe, semi-continue inférieurement, homogène de degré 1 au sens de* (33). *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 IV-81. 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 IV-80. 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 Tchebychev 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,

$$\Phi\left(\int_{A_k} f\, d\mu\right) \leq \int_{A_k} \Phi \circ f\, d\mu. \tag{34}$$

En distinguant les cas $f(x) = 0$ et $f(x) \neq 0$, on voit que $f 1_{A_k}$ converge partout vers f quand $k \to \infty$. Par convergence dominée,

$$\int_{A_k} f\, d\mu \xrightarrow[k\to\infty]{} \int_X f\, d\mu,$$

et par semi-continuité inférieure de Φ,

$$\Phi\left(\int f\, d\mu\right) \leq \liminf_{k\to\infty} \int_{A_k} \Phi\circ f\, d\mu. \tag{35}$$

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, par convergence dominée,

$$\int_{A_k} f\, d\mu \xrightarrow[k\to\infty]{} \int_X f\, d\mu.$$

Cela conclut l'argument. □

Remarque IV-82. 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) ; et elle est homogène de degré 1.

IV-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

$$f(x)\,g(x) \leq \Phi(f(x)) + \Phi^*(g(x)),$$

où Φ^* est la transformée de Legendre de Φ. 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

$$\int \langle f,g\rangle\, d\mu \leq \int_X \Phi\circ f\, d\mu + \int_X \Phi^*\circ g\, d\mu, \tag{36}$$

dès que ces intégrales sont bien définies dans $\overline{\mathbb{R}}$.

Ces inégalités aussi élémentaires que cruciales peuvent parfois être améliorées, en particulier quand Φ est **homogène**, comme on va le voir.

IV-4.4. Inégalité de Hölder. Très souvent, quand on majore des produits de fonctions, on cherche à exploiter des intégrabilités différentes pour les deux facteurs. Un exemple évident est l'inégalité utile

$$|f| \leq C \ \mu\text{-presque partout} \quad \Longrightarrow \left|\int fg\, d\mu\right| \leq C\int |g|\, d\mu.$$

L'inégalité de Hölder affine cette inégalité grâce à l'usage de fonctions puissances.

Théorème IV-83 (inégalité de Hölder). *Soit* $(X, \mathcal{A}, \mu)$ *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*

$$\left|\int fg\,d\mu\right| \le \left(\int |f|^p\,d\mu\right)^{1/p} \left(\int |g|^{p'}\,d\mu\right)^{1/p'}, \tag{37}$$

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* $\varepsilon\, g$ *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 IV-84. (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, et cette inégalité revient à dire que la fonction $f \longmapsto \int |f|^2$ est une forme quadratique positive. On reviendra sur cela dans les chapitres VIII et **??**.

(ii) L'**inégalité de Minkowski** fait aussi intervenir des fonctions puissance et de la convexité ; elle sera introduite dans le Chapitre VIII,

(iii) L'inégalité de Hölder reste vraie pour $p = 1$ ou $p = \infty$, si l'on convient de poser

$$\left(\int |g|^\infty\,d\mu\right)^{1/\infty} := \inf\Big\{C \in \mathbb{R};\ |g| \le C\ \mu\text{-presque partout}\Big\}.$$

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 $0 < p \le 1$, il existe une inégalité de Hölder renversée : l'exposant p' est remplacé par $q = p/(1-p) = -p'$, le signe d'inégalité dans (37) est renversé, et les fonctions f et g sont supposées positives :

$$f, g \ge 0 \implies \int fg\,d\mu \ge \left(\int f^p\,d\mu\right)^{1/p} \left(\int g^q\,d\mu\right)^{1/q}. \tag{38}$$

Cette inégalité, d'usage beaucoup moins fréquent que (37), repose sur les mêmes bases que l'inégalité habituelle, et elle est laissée en exercice.

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 alors $\widetilde{f} = f/(\int |f|^p)^{1/p}$, $\widetilde{g} = g/(\int |g|^{p'})^{1/p'}$, on a alors $\int |\widetilde{f}|^p = \int |\widetilde{g}|^{p'} = 1$ et on doit prouver

$$\left|\int \widetilde{f}\widetilde{g}, d\mu\right| \le 1.$$

On sait déjà que

$$\left|\int \widetilde{f}\widetilde{g}\,d\mu\right| \le \int |\widetilde{f}\widetilde{g}|\,d\mu,$$

et si le membre de droite est fini, l'égalité n'est possible que si, μ-presque partout, $\widetilde{f}\widetilde{g} = |\widetilde{f}\widetilde{g}|$, c'est-à-dire si $\widetilde{f}$ et $\widetilde{g}$ ont même signe, ou de façon équivalente, si f et g ont même signe.

Pour récapituler, on s'est ramené, par cet argument d'homogénéité, au cas particulier suivant : montrer que, si f et g sont deux fonctions positives,

$$\int f^p = \int g^{p'} = 1 \quad \Longrightarrow \int fg \leq 1,$$

avec égalité si et seulement si $f = g$ presque partout. On écrit alors l'inégalité du Lemme IV-126 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, ce qui conclut l'argument. $\square$

L'inégalité de Hölder admet plusieurs avatars simples et intéressants.

Théorème IV-85 (variantes de l'inégalité de Hölder). *(i) Soit $(X, \mathcal{A}, \mu)$ 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 $\lambda > 0$,*

$$\left|\int fg\, d\mu\right| \leq \frac{\lambda^p}{p} \int |f|^p\, d\mu + \frac{1}{\lambda^{p'} p'} \int |g|^q\, d\mu,$$

où l'on convient que le membre de gauche vaut $+\infty$ si fg n'est pas intégrable.

(ii) Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, soient $p_1, \ldots, p_k \in]1, +\infty[$ tels que

$$\sum_{i=1}^{k} \frac{1}{p_i} = 1,$$

et soient $f_1, \ldots, f_k$ des fonctions mesurables sur X, à valeurs dans $\overline{\mathbb{R}}$. Alors

$$\left|\int (\prod_i f_i)\, d\mu\right| \leq \prod_i \left(\int |f_i|^{p_i}\, d\mu\right)^{1/p_i},$$

où l'on convient que le membre de gauche vaut $+\infty$ si $\prod f_i$ n'est pas intégrable.

(iii) Soient $(X, \mathcal{A}, \mu)$ un espace mesuré σ-fini et $(Y, \mathcal{B}, \pi)$ un espace de probabilité. Alors, pour toute fonction F mesurable de $X \times Y$ dans $\mathbb{R}_+ \cup \{+\infty\}$, on a

$$\int_X \exp\left(\int_Y \log F(x,y)\, \pi(dy)\right) \mu(dx) \leq \exp\left(\int_Y \log\left(\int_X F(x,y)\, \mu(dx)\right) \pi(dy)\right).$$

(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 $f : X \to E^*$ et $g : X \to E$ des fonctions mesurables, soit $p \in]1,+\infty[$ et $q := p/(p-1)$. Alors

$$\left|\int \langle f,g\rangle_{E^*\times E}\,d\mu\right| \leq \left(\int \|f\|_{E^*}^p\right)^{1/p}\left(\int \|g\|_E^{p'}\right)^{1/p'},$$

où l'on convient que le membre de gauche vaut $+\infty$ si $\langle f,g\rangle$ n'est pas intégrable.

(vi) Soit $(X,\mathcal{A},\mu)$ un espace mesuré, $f,g : X \to \mathbb{C}$ deux fonctions mesurables à valeurs complexes, $p \in]1,+\infty[$ et $q := p/(p-1)$. Alors

$$\left|\int fg\,d\mu\right| \leq \left(\int |f|^p\,d\mu\right)^{1/p}\left(\int |g|^{p'}\,d\mu\right)^{1/p'x},$$

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

$$\inf_{\lambda\in[0,1]}\left\{\frac{\lambda^p}{p}\int |f|^p\,d\mu + \frac{1}{\lambda^{p'}p'}\int |g|^{p'}\right\} = \left(\int |f|^p\,d\mu\right)^{1/p}\left(\int |g|^{p'}\,d\mu\right)^{1/p'}.$$

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 IV-126 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

$$\forall y \in Y, \qquad \int_X F(x,y)\,d\mu(x) = 1,$$

auquel cas l'inégalité à établir est

$$\int_X \exp\left(\int_Y \log F(x,y)\,\pi(dy)\right)\mu(dx) \leq 1.$$

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 $\langle f,g\rangle_{E^*\times E} \leq \|f\|_{E^*}\|g\|_E$. □

Remarque IV-86. 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 $Y := \{0,1\}$, $\pi := (1/p)\delta_0 + (1/p')\delta_1$, $F(x,0) := f(x)^p$, $F(x,1) := g(x)^{p'}$. 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.

IV-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 d'autres fonctions, par commodité ou par nécessité. L'une de ces fonctions est l'**entropie** d'une mesure ou d'une densité, un concept introduit au dix-neuvième siècle par Ludwig Boltzmann, le héros de la théorie atomique, et redécouvert dans les années 1940 par Claude Shannon, père fondateur de la théorie de l'information et de la communication, et l'un des pères de l'informatique. Les propriétés numériques de l'entropie ont été étudiées au milieu du vingtième siècle par des spécialistes d'information et statistique comme Mark Semonovitch Pinsker, Richard Leibler, Imre Csiszár, Johannes Kemperman, Solomon Kullback.

Définition IV-87 (entropie). *Soient* (X, μ) *un espace mesuré et* $f : X \to \mathbb{R}_+ \cup \{+\infty\}$ *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 IV-88. La fonction $f \longmapsto f \log f - f + 1$ est positive, et donc l'information de Kullback toujours positive, alors que l'entropie peut prendre l'un ou l'autre 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, aussi appelée divergence de Kullback–Leibler, 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 IV-89 (inégalité de convexité pour l'entropie). *Soit* $(X, \mathcal{A}, \mu)$ *un espace de probabilité, et soient* $f, g : X \to \mathbb{R}_+ \cup \{+\infty\}$ *deux fonctions mesurables positives. Alors,*

$$\int fg \, d\mu \leq \int (f \log f - f + 1) \, d\mu + \log \int e^g \, d\mu.$$

Démonstration du Théorème IV-89. Par homogénéité, on peut se ramener au cas où $\int e^g \, d\mu = 1$, et il s'agit alors de prouver que

$$\int fg \, d\mu \leq \int (f \log f - f + 1) \, d\mu.$$

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

$$\int fg \, d\mu \leq \int (f \log f - f + 1) \, d\mu + \int (e^g - 1) \, d\mu,$$

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 répond de manière assez précise à cette question.

Théorème IV-90 (inégalité de Pinsker). *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*

$$\int_X |f-1|\, d\mu \le \sqrt{2\int_X (f\log f - f + 1)\, d\mu}.$$

Remarque IV-91. Cette inégalité se rencontre aussi sous le nom de Csiszár–Kullback–Pinsker, ou diverses combinaisons de ces trois noms.

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 $t \in [0,1]$,

$$\left(\int |f-1|\, d\mu\right)^2 \le \left(\int_X \frac{(f-1)^2}{1+t(f-1)}\, d\mu\right)\left(\int (1+t(f-1))\, d\mu\right) = \int_X \frac{(f-1)^2}{1+t(f-1)}\, d\mu,$$

puisque μ et $f\mu$ sont toutes deux des mesures de probabilité. On conclut que

$$\left(\int_0^1 (1-t)\, dt\right)\left(\int |f-1|\, d\mu\right)^2 \le \int (f\log f - f + 1)\, d\mu,$$

ce qui est équivalent à la conclusion souhaitée. $\square$

IV-5*Équi-intégrabilité et tension

Cette section plus technique aborde les deux critères majeurs de compacité liés à l'intégration de Lebesgue ; elle pourra être omise en première lecture.

IV-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 $f \in \mathcal{F}$. Écrit autrement :

$$\forall \varepsilon > 0\ \exists \delta > 0;\ \forall x, y \in X,\ \forall f \in \mathcal{F} \quad d(x,y) \le \delta \implies |f(y) - f(x)| \le \varepsilon.$$

Bien noter que dans cette écriture δ est indépendant de $f \in \mathcal{F}$: c'est en ce sens que le module de continuité est dit uniforme. Bien évidemment, si $\mathcal{F}$ est équicontinu, alors tout $f \in \mathcal{F}$ est uniformément continu ; et la réciproque est fausse.

La notion d'équicontinuité joue un rôle important, par exemple dans l'étude de la compacité dans des espaces de fonctions continues. Ainsi, le théorème d'Ascoli indique si K est un espace métrique compact, alors les ensembles précompacts dans $C(K)$ sont exactement les ensembles équicontinus.

En théorie de la mesure, un concept analogue est l'*équi-intégrabilité* :

Définition IV-92 (Équi-intégrabilité). *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré.*

(i) On dit qu'un ensemble $\mathcal{F}$ de fonctions $f : X \to \mathbb{R}$ est équi-intégrable si pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que

$$\mu[A] \leq \delta \Longrightarrow \qquad \forall f \in \mathcal{F}, \quad \int_A |f| \, d\mu \leq \varepsilon. \tag{39}$$

(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

$$\forall f \in \mathcal{F}, \quad \int_{X \setminus Y_\varepsilon} |f| \, d\mu \leq \varepsilon. \tag{40}$$

L'équi-intégrabilité se prouve le plus souvent grâce au critère suivant :

Proposition IV-93 (Équi-intégrabilité : reformulation). *Soient $(X, \mathcal{A}, \mu)$ 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 $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$, telle que

$$\lim_{r \to \infty} \frac{\Phi(r)}{r} = +\infty \qquad \text{et} \qquad \sup_{f \in \mathcal{F}} \int_X \Phi(|f|) \, d\mu < +\infty. \tag{41}$$

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 une fonction $\varphi : X \to \mathbb{R}_+$ telle que

$$\forall r > 0, \quad \mu[\{\varphi \leq r\}] < +\infty; \qquad \text{et} \qquad \sup_{f \in \mathcal{F}} \int_X |f| \, \varphi \, d\mu < +\infty.$$

Démonstration. Commençons par la propriété (i). Supposons (41) 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

$$r \geq R \Longrightarrow \qquad \frac{\Phi(r)}{r} \geq \frac{C}{2\varepsilon}.$$

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 Tchebychev, pour tout $N \geq 1$ on a

$$\mu\big[\{|f| \geq N\}\big] \leq \frac{1}{N} \int |f| \, d\mu \leq \frac{C}{N}.$$

Pour $\varepsilon > 0$, soit δ comme dans (39), 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,

$$\forall \varepsilon > 0, \quad \exists N(\varepsilon) \geq 1; \qquad \forall f \in \mathcal{F}, \quad \int_{|f|\geq N(\varepsilon)} |f|\, d\mu \leq \varepsilon.$$

On pose $N_0 = 0$ et on construit par récurrence une suite d'entiers $(N_k)_{k\geq 1}$ tels que pour tout $k \geq 1$,

$$N_k > 2N_{k-1}, \qquad \text{et} \qquad \sup_{f\in\mathcal{F}} \int_{|f|\geq N_k} |f|\, d\mu \leq 2^{-k}.$$

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 $\Phi(x)/x \longrightarrow +\infty$ quand $x \to \infty$. 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 exercice d'analyse réelle classique, laissé à la lectrice.

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 Tchebychev 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

$$\int_{X \setminus Y} |f|\, d\mu = \lim_{k \to \infty} \int_{X \setminus Y_k} |f|\, d\mu = 0.$$

Il s'ensuit que f est nulle 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, via le critère d'Ascoli, revient à dire qu'un singleton est bien compact !) Un énoncé analogue est valable dans le cadre de l'équi-intégrabilité :

Proposition IV-94. *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,

$$\eta(M) := \int_{|f| \geq M} |f_M - f|\, d\mu \xrightarrow[M \to \infty]{} 0.$$

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 Tchebychev 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,

$$\int_{X \setminus A_k} |f|\, d\mu = \int 1_{|f| < k^{-1}} |f|\, d\mu \xrightarrow[k \to \infty]{} 0.$$

On peut donc trouver k assez grand pour que cette quantité soit majorée par ε. □

Remarque IV-95. 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 $\Phi(r)/r \to \infty$, 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 IV-96. On verra au Chapitre **??** 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 ; et approfondit le parallèle avec la propriété d'équi-intégrabilité.

IV-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 IV-97 (tension). *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*

$$\sup_{\mu \in \mathcal{M}} \mu[X \setminus K_\varepsilon] \le \varepsilon.$$

La notion de tension n'est pas sans rapport avec l'équi-intégrabilité à l'infini : si ν est une mesure de référence qui attribue une mesure finie aux compacts, alors la tension d'une famille de mesures de la forme $f\nu$ implique son équi-intégrabilité à l'infini ; et en fait les deux concepts sont équivalents modulo le remplacement des compacts par les ensembles de mesure finie. Le lien entre les deux notions est clarifié par la formulation équivalente de la tension que voici :

Proposition IV-98 (tension : reformulation). *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 $\varphi : X \to \mathbb{R}_+$, tendant vers l'infini à l'infini (au sens où pour tout $r > 0$ il existe un compact K_r tel que $x \notin K_r \Longrightarrow \varphi(x) \ge r$), 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 $\mu \in \mathcal{M}$, $\int \varphi\, d\mu \le C$. On pose $r = C/\varepsilon$: par inégalité de Tchebychev,

$$\mu[\{\varphi \ge r\}] \le C/r = \varepsilon.$$

Soit K_r comme dans l'énoncé ; on a alors $X \setminus K_r \subset \{\varphi \ge r\}$, d'où $\mu[K_r] \le \varepsilon$.

Réciproquement, soit $\mathcal{M}$ un ensemble tendu. Par récurrence, on peut construire une suite croissante de compacts K_n tels que pour tout $\mu \in \mathcal{M}$,

$$\mu[K_n] \le 2^{-n}.$$

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 IV-93(ii).

La dernière assertion (on peut choisir φ continue si X est localement compact) est laissée en exercice. □

IV-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. Les outils de la section IV-2 permettent de démontrer les principaux résultats de ce sujet.

IV-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 IV-99 (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 $C_F(A_F)$, aussi noté abusivement $C(A_F)$, par*

$$C(A_F) = \big\{(x_t)_{t\in T};\ (x_t)_{t\in F} \in A_F\big\}.$$

Remarques IV-100. 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 IV-101 (tribu produit infini). *Soit T un ensemble arbitraire et $(X_t, \mathcal{A}_t)_{t\in T}$ une famille d'espaces 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, rappelée ci-après :

Définition IV-102 (topologie produit infini). *Soient 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 IV-103. (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.

IV-6.2. Produit infini de mesures. Dans le Chapitre II, nous avions prouvé le Théorème II-87 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, pp. 257-259] pour une variante.

Théorème IV-104 (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, $1 \leq i \leq n$,

$$\mu^\infty[A_1 \times \ldots \times A_n \times \prod_{i \geq n+1} X_i] = \mu_1[A_1] \times \ldots \times \mu_n[A_n] \times \prod_{i \geq n+1} \mu_i[A_i].$$

On a alors, pour toutes parties mesurables A_k de X_k,

$$\mu^\infty[\prod A_k] = \prod \mu_k[A_k].$$

Exemple IV-105. 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 $\mathbb{R}^\infty$ (que l'on peut identifier à ℓ^2).

Démonstration du Théorème IV-104. Si $A_1, \ldots, A_n$ sont des parties mesurables de $X_1, \ldots, X_n$ respectivement, on note

$$C(A_1 \times \ldots \times A_n) = \big\{x \in X;\ \forall i \leq n,\ x_i \in A_i\big\}$$

le cylindre produit de $A_1, \ldots, A_n$, et on note $\mathcal{C}$ l'ensemble de tous ces cylindres, pour toutes valeurs de $n \in \mathbb{N}$.

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 II-82(i) garantit l'unicité du prolongement éventuel. Pour prouver l'existence de ce prolongement, en vertu du Théorème II-82(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

$$\sum \mu^\infty[C_k] < \mu[X]. \tag{42}$$

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 $C(A_1, \ldots, A_K)$; et sa *composante* d'ordre ℓ comme A_ℓ. On vérifie facilement, grâce à l'existence de la mesure produit $\mu_1 \otimes \ldots \otimes \mu_K$, 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 (42) 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,

$$\int_{Z_1} \phi_1(x_1)\, \mu_1(dx_1) < \mu_1[Z_1] = M_2\, \varepsilon_1.$$

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 $\mu'[C_k'] = \prod_{j=2}^\infty [A_k]$, 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 et x_2 appartiennent respectivement 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 $A_1 \times C_k'$, et que x_2 appartient à la seconde 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 $x = (x_1, x_2, \ldots)$ telle que pour tout K, $(x_1, \ldots, x_K)$ 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 $A_1 \times \ldots \times A_n \times \prod_{j \geq n+1} X_j$ décroissent vers $\prod A_k$ quand $n \to \infty$, et que $\mu[X] < +\infty$. □

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 IV-106 (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*

$$C(A_F) := \big\{x \in X;\ \forall t \in F,\ x_t \in A_t\big\}.$$

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 IV-61 (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 II-78.

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. (Comme toujours en théorie de la mesure, on se ramène à la dénombrabilité.) La σ-additivité est alors une conséquence du Théorème IV-104. □

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 IV-106.

Proposition IV-107 (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

$$\prod_{k\in\mathbb{N}} \mu_k = \mu_1 \otimes \mu_2 \otimes \ldots \otimes \mu_N \otimes \left(\prod_{k\geq N+1} \mu_k\right),$$

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.

IV-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 IV-108 (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*

$$\mu[C_n \setminus A] + \mu[A \setminus C_n] \xrightarrow[n\to\infty]{} 0. \tag{43}$$

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 (43). Il est clair que $\mathcal{C}$ contient les cylindres ; pour conclure il 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

$$\mu\Big[A \setminus (\cup_{1\leq k\leq N} A^k)\Big] \leq \varepsilon/2.$$

D'autre part, puisque $\mathcal{C}$ est stable par union finie, il existe un cylindre C tel que

$$\mu\Big[C \setminus (\cup_{1\le k\le N} A^k)\Big] + \mu\Big[(\cup_{1\le k\le N} A^k) \setminus C\Big] \le \varepsilon/2.$$

On en déduit la conclusion souhaitée. □

IV-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 trois lois du 0-1 célèbres ; la plus ancienne est celle d'Émile Borel (1909), défrichant alors l'application des mesures au calcul des probabilités ; la seconde, la plus connue, est celle d'Andreï Kolmogorov (1933), partie de son grand œuvre sur l'axiomatisation des probabilités ; et la troisième, la plus générale, due aux mathématiciens américains Edwin Hewitt (spécialiste de théorie des représentations) et Leonard Savage (théoricien des statistiques) en 1955. On commencera par établir la loi de Hewitt–Savage pour en déduire celle de Kolmogorov, qui à son tour impliquera celle de Borel.

Définition IV-109 (permutation finie). *Soit $\sigma : \mathbb{N} \to \mathbb{N}$ 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 IV-110 (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 IV-111 (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 $n \ge m$. On définit la tribu asymptotique, $\mathcal{A}_\infty$, comme l'intersection de toutes les tribus $\mathcal{A}_m$.*

Remarque IV-112. 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 IV-113. 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 $\sum n x_n = 1$ 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 IV-114 (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 $m \ge n$; 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 IV-114, 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 $\{0, \ldots, n-1\}$. 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 $\{0, \ldots, n\}$. 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 $\{0, \ldots, n\}$. 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. $\square$

Considérons maintenant les lois du 0-1 proprement dites.

Théorème IV-115 (loi du 0-1 de Hewitt–Savage). *Soit $(X, \mathcal{A}, \mu)$ 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 IV-114, on obtient le

Corollaire IV-116 (loi du 0-1 de Kolmogorov). *Soit $(X, \mathcal{A}, \mu)$ 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 IV-115. Soit $\varepsilon > 0$; par le Théorème IV-108, il existe un cylindre C tel que

$$\mu^{\otimes\mathbb{N}}[A \setminus C] + \mu^{\otimes\mathbb{N}}[C \setminus A] \leq \varepsilon. \tag{44}$$

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 $\{1, \ldots, n\}$ et $\{n+1, \ldots, 2n\}$. 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 IV-107) 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 (44), 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

$$\mu^{\otimes n}[B]\, (1 - \mu^{\otimes n}[B]) \leq \varepsilon.$$

Posons $b := \mu^{\otimes n}[B]$, $a := \mu^{\otimes \mathbb{N}}[A]$. On a donc

$$a(1-a) \le (b+\varepsilon)(1-b+\varepsilon) = b(1-b) + \varepsilon + \varepsilon^2 \le 2\varepsilon + \varepsilon^2.$$

En faisant tendre ε vers 0, on trouve $a(1-a) = 0$, ce qui conclut la preuve. □

Remarque IV-117. 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, de sorte qu'en un sens ces deux lois sont équivalentes [Dudley, p. 272].

Concluons avec la loi de Borel, dans un énoncé qui va un peu plus loin que juste la règle 0-1. Elle repose sur la notion de limite supérieure d'une famille d'ensembles : si $(A_n)_{n\in\mathbb{N}}$ sont des ensembles, alors $\limsup A_n$ est l'ensemble des éléments x qui appartiennent à une infinité de A_n.

Théorème IV-118 (Loi du 0-1 de Borel). *Soient $(A_n)_{n\in\mathbb{N}}$ une famille de parties mesurables indépendantes d'un espace de probabilité $(X, \mathcal{A}, \mu)$. Alors $\mu[\limsup A_n] \in \{0,1\}$. En outre, cette probabilité vaut 1 si et seulement si $\sum \mu[A_n] = +\infty$.*

Preuve du Théorème IV-118. Soit $f_n = 1_{A_n}$; la mesure produit $(f_n)_\#\mu$ est la mesure sur $\{0,1\}$ qui attribue $\mu[A_n]$ à $\{1\}$, et $1-\mu[A_n]$ à $\{0\}$. Par indépendance, la mesure image de μ par $f = (f_n)_{n\in\mathbb{N}}$ est le produit infini des $(f_n)_\#\mu$, que l'on peut aussi noter $f_\#\mu$. Dire que $x \in \limsup A_n$, c'est dire que la suite $f(x)$ prend une infinité de fois la valeur 1. Dans $\{0,1\}^{\mathbb{N}}$, l'événement B défini par "$(y_n)_{n\in\mathbb{N}}$ appartient à B si et seulement si y_n prend une infinité de fois la valeur 1" est bien évidemment asymptotique, la loi du 0-1 de Kolmogorov implique donc que $\mu[\limsup A_n] = (f_\#\mu)[B] \in \{0,1\}$.

Jusqu'ici l'énoncé est juste un corollaire de la loi de Kolmogorov. Mais on va maintenant aller un peu plus loin en précisant l'alternative via la nature de la série $\sum \mu[A_n]$. Supposons d'abord que $\sum \mu[A_n]$ converge, alors pour tout $k \in \mathbb{N}$,

$$\mu\big[\limsup A_n\big] \le \mu\Big[\bigcup_{n\ge k} A_n\Big] \le \sum_{n\ge k} \mu[A_n].$$

En faisant tendre n vers l'infini, on trouve

$$\mu\big[\limsup A_n\big] \xrightarrow[n\to\infty]{} 0.$$

Supposons maintenant que $\sum \mu[A_n]$ diverge. Dire que $x \notin \limsup A_n$, c'est dire qu'il existe un entier k au-delà duquel x n'appartient à aucun X_n. D'où l'union croissante

$$X \setminus \limsup A_n = \bigcup_{k\in\mathbb{N}} \bigcap_{n\ge k} (X \setminus A_n).$$

Par convergence monotone et indépendance,

$$\mu\big[X \setminus \limsup A_n\big] = \lim_{k\to\infty} \prod_{n\ge k} \mu[X \setminus A_n].$$

D'où

$$\begin{aligned}\log \mu\big[X \setminus \limsup A_n\big] &= \lim_{k\to\infty} \sum_{n\geq k} \log \mu[X \setminus A_n] \\ &= \lim_{k\to\infty} \sum_{n\geq k} \log\big(1-\mu[A_n]\big) \\ &\leq \limsup_{k\to\infty} \sum_{n\geq k} -\mu[A_n] = -\infty,\end{aligned}$$

où l'on a utilisé $\log(1-u) \leq -u$ et $\sum \mu[A_n] = +\infty$. Soit $\mu[\limsup A_n] = 0$. □

IV-6.5. Théorème de prolongement de Kolmogorov. Pour conclure cette section sur les produits infinis, démontrons le Théorème II-90. Pour le confort de la lectrice, je vais commencer par rappeler son énoncé, sous une forme équivalente mais légèrement différente et un peu plus précise.

Théorème IV-119 (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 $F = \{t_1, \ldots, t_K\} \subset T$, on définit $X_F := X_{t_1} \times \ldots \times X_{t_K}$; et pour tout borélien A_F de X_F, on définit le cylindre $C_F(A_F)$, noté abusivement $C(A_F)$, par

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

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,

$$\big[C(A_F) = C(B_{F'})\big] \Longrightarrow \ \mu_F[A_F] = \mu_{F'}[B_{F'}].$$

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

$$A_F = A_{t_1} \times \ldots \times A_{t_K},$$

chacun des A_t appartenant à $\mathcal{F}_t$.

Preuve du Théorème IV-119. 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 II-78, 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 $(A_k)_{k\geq 1}$ des cylindres disjoints, dont l'union $A = \cup A_k$ est aussi un cylindre ; on doit prouver que

$$\mu[A] = \sum_{k\geq 1} \mu[A_k].$$

Pour tout N, les cylindres $A, A_1, \ldots, A_N$ 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 $N \to \infty$, 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 $T = \{1, 2, \ldots\}$. On définit alors $\mu_n = \mu_F$ pour $F = \{1, \ldots, n\}$. 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 $X^n = X_1 \times \ldots \times X_n$.

L'espace X^{n_0} est produit fini d'espaces polonais, donc polonais lui-même. Par le Théorème de régularité II-62, 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

$$\mu_n[K_{n_0} \times \ldots \times K_n] \geq \mu[A] - 2\varepsilon.$$

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 $U_1, \ldots, U_N$. Leur union est un cylindre, d'ordre fini M ; comme ils recouvrent K, ils recouvrent également le cylindre d'ordre M

$$C_M(K) := K_{n_0} \times K_{n_1} \times \ldots \times K_M \times X_{M+1} \times X_{M+2} \times \ldots$$

En particulier, $\mu[C_M(K)] \le \sum_{k=1}^N \mu[U_k]$. Mais, par construction,

$$\mu[C_M(K)] = \mu_M[K_{n_0} \times \ldots \times K_M] \ge \mu[A] - 2\varepsilon$$

et

$$\sum_k \mu[U_k] \le \sum_k \mu[A_k] + \varepsilon.$$

On conclut que

$$\mu[A] \le \sum_k \mu[A_k] + 3\varepsilon.$$

Comme ε est arbitraire, il s'ensuit que $\mu[A] \le \sum_k \mu[A_k]$, ce qui était notre but. □

Remarque IV-120. 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 légèrement plus souples, sans gain de généralité significatif toutefois.

Exercice IV-121. Montrer que le Théorème de Kolmogorov implique le résultat suivant, qui fonde la théorie des chaînes de Markov en probabilités :

Théorème IV-122 (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* $C := A_0 \times A_1 \times A_2 \times \ldots \times A_n \times \left(\prod_{j\ge n+1} X_j\right)$ *de* X*,*

$$\mu[C] = \int_{A_0}\int_{A_1} \ldots \int_{A_{n-1}}\int_{A_n} \nu_{x_{n-1}}(dx_n)\, \nu_{x_{n-2}}(dx_{n-1}) \ldots \nu_{x_1}(dx_0)\, \mu(dx_0).$$

Voici maintenant un exemple où l'espace T n'est pas dénombrable.

Théorème IV-123 (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* $(\gamma_{t,x})_{t\ge 0,\, x\in X}$*, 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* $\mathbb{R}_+ \times \mathbb{R}_+$*,* $0 \le s \le t$*, et pour tout* $x \in X$*. Pour toute famille finie* $F = \{t_0, \ldots, t_N\}$*,* $0 = t_0 \le t_1 \le t_2 \le \ldots \le t_N$*, pour toutes parties mesurables* $A_0, A_1, \ldots, A_N$ *de* X*, on défnit le cylindre*

$$C_F(A_0, \ldots, A_N) = \left\{x \in X^{\mathbb{R}_+};\ \forall k \in \{0, \ldots, N\},\ x_{t_k} \in A_k\right\}.$$

Alors il existe une unique mesure de probabilité μ *sur* $X^{\mathbb{R}_+}$ *telle que pour tout cylindre* $C_F(A_0, \ldots, A_N)$*,*

$$\mu\big[C_F(A_0, \ldots, A_N)\big] = \int_{A_0\times\ldots\times A_N} \mu_{t_0}(dx_0)\, \gamma_{t_1-t_0,x_0}(dx_1)\, \gamma_{t_2-t_1}(dx_2) \ldots \gamma_{t_N-t_{N-1}}(dx_N).$$

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 $C_F(A_0, \ldots, A_N)$, où F est une partie finie arbitraire de T, et les A_j sont des éléments arbitraires de $\mathcal{F}$.

Exemple IV-124. L'exemple le plus célèbre est 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, qui sont les de $\mathbb{R}_+$ dans $\mathbb{R}$.

Le théorème précédent est cependant 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 bien 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, jouent un rôle aussi important dans des espaces vectoriels généraux, que les fonctions croissantes sur $\mathbb{R}$. Elles apparaissent naturellement dès que l'on veut estimer numériquement des intégrales. 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" $(1-\lambda)x + \lambda y$ est un élément de C. On dit que $\Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est une fonction convexe si, pour tous $x, y \in X$, $\lambda \in [0,1]$,

$$\Phi\big((1-\lambda)\, x + \lambda\, y\big) \leq (1-\lambda)\, \Phi(x) + \lambda\, \Phi(y).$$

Du point de vue géométrique, cela veut dire que pour tous 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 dimposer que pour tout $N \in \mathbb{N}$, et tous $x_1, \ldots, x_N \in \mathbb{R}^n$, $\lambda_1, \ldots, \lambda_N \geq 0$, $\sum \lambda_i = 1$,

$$\Phi\Big(\sum \lambda_i x_i\Big) \leq \sum \lambda_i \Phi(x_i).$$

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

(a) Φ ne soit pas identiquement $+\infty$;

(b) 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 $\Phi(x_t) \longrightarrow \Phi(x_1)$ quand $t \to 1$, où $x_t := (1-t)\,x_0 + t\,x_1$.

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

$$\forall y \in \mathbb{R}^n, \qquad \Phi(y) \geq \Phi(x) + \langle z, y - x \rangle. \tag{45}$$

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 vecteurs z admissibles dans (45) 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,

$$\forall y \in \mathbb{R}^n, \qquad \Phi(y) \geq \Phi(x) + \langle \nabla\Phi(x), y - x \rangle.$$

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 $\nabla\Phi(x) = z$.

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** en tout point de Ω. C'est bien sûr la positivité au sens des matrices symétriques : explicitement, cela veut dire

$$\forall x \in \Omega, \quad \forall \xi \in \mathbb{R}^n, \qquad \langle \nabla^2\Phi(x)\xi, \xi \rangle \geq 0.$$

C'est le critère que l'on utilise le plus souvent, en pratique, pour vérifier la convexité.

Une fonction $\Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ étant donnée, on peut définir sa **transformée de Legendre** :

$$\Phi^*(y) := \sup_{x \in \mathbb{R}^n} \Big(\langle x, y \rangle - \Phi(x) \Big).$$

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

$$\Phi^{**} = \Phi.$$

(Réciproquement, si $\Phi^{**} = \Phi$, alors Φ est convexe et semi-continue inférieurement.) 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

$$\langle x, y \rangle = \Phi(x) + \Phi^*(y) \Longleftrightarrow y \in \partial\Phi(x).$$

En particulier, si Φ est différentiable en x, alors
$$\langle x, \nabla\Phi(x)\rangle = \Phi(x) + \Phi^*(\nabla\Phi(x)),$$
et $\nabla\Phi(x)$ 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 IV-125. Vérifier que, pour tout $p \in [1, +\infty[$,
$$\Phi(x) = \frac{|x|^p}{p} \Longrightarrow \qquad \Phi^*(y) = \frac{|y|^{p'}}{p'}, \qquad p' = \frac{p}{p-1}.$$
Cette identité est vraie dans $\mathbb{R}$ ou plus généralement dans $\mathbb{R}^n$. Vérifier également que, sur $\mathbb{R}_+$,
$$\Phi(x) = x\log x - x \Longrightarrow \qquad \Phi^*(y) = e^y.$$

La proposition qui suit est une conséquence immédiate de l'exercice.

Proposition IV-126 (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 IV-127 (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 IV-128. L'exposant p' est souvent noté q, ce qui a l'inconvénient de ne pas imposer de lien notationnel avec p ; en outre, en théorie des probabilités, il est également fréquent de noter $q = 1 - p$ si $p \in [0, 1]$... C'est pourquoi je recommande d'utiliser la notation p' (également très courante). Bien sûr, $(p')' = p$.

La deuxième partie de l'exercice IV-125 mène à l'inégalité suivante, également fort utile.

Proposition IV-129 (inégalité de convexité de Young logarithmique). *Soient a, b deux nombres réels positifs. Alors*
$$ab \leq (a\log a - a + 1) + (e^b - 1),$$
avec égalité seulement pour $b = \log a$.

CHAPITRE V

Théorie descriptive des ensembles

Ce chapitre d'approfondissement, à réserver à une seconde lecture, aborde des concepts plus techniques, indispensables dans certains problèmes d'analyse et de probabilité. Il est motivé par quelques questions naturelles et liées entre elles, que nous avons commencé à rencontrer dans les Chapitres II (nature des ensembles boréliens), III (stabilité des fonctions boréliennes) et IV (boréliens dans les espaces produits) ; par exemple

• Comment décrire, aussi constructivement que possible, les ensembles boréliens ?

• Si C est borélien dans un espace produit $X \times Y$, on sait que ses sections selon X ou Y sont mesurables (Proposition IV-43), mais que dire de ses projections ?

• Et peut-on choisir mesurablement une application $y = y(x)$ telle que $(x, y(x)) \in C$ pour tout x dans la projection de C ? De façon équivalente, peut-on trouver un graphe borélien inclus dans C et avec même projection ?

Ces questions ont initié la **théorie descriptive des ensembles**, une vingtaine d'années après les avancées de Baire, Borel et Lebesgue : tout un paysage mathématique, à l'interface entre la logique et l'analyse, dont on ne soupçonnait pas la richesse.

La naissance de la théorie est quelque peu dramatique et pleine de rebondissements. Lebesgue en 1905 avait cru prouver que la projection d'un borélien de $\mathbb{R}^2$ est un borélien de $\mathbb{R}$. Mais en 1917, un jeune mathématicien russe, Mikhaïl Souslin, fils de paysans pauvres, étudiant alors à Moscou sous la direction du célèbre professeur Nikolai Luzin, découvrit une faille dans l'argument de Lebesgue (même les meilleurs font des erreurs...). Pour résoudre le problème qui avait berné Lebesgue, Souslin introduisit une panoplie d'outils conceptuels (aujourd'hui appelés problème de Souslin, schéma de Souslin, opération de Souslin, droite de Souslin, propriété de Souslin, représentation de Souslin, arbre de Souslin etc), donnant naissance au concept d'ensemble analytique et ouvrant tout un champ mathématique nouveau. Souslin meurt prématurément de maladie, dans les affres de la guerre civile russe, sans avoir eu le temps de rien publier ou presque ; Luzin et d'autres continuent ses travaux.

En écho dramatique, en 1936 le traitement réservé aux travaux de Souslin fait partie du dossier monté à charge contre Luzin par ses jeunes collègues et anciens étudiants (dont Andreï Kolmogorov) dans le plus brûlant épisode de procès politique stalinien touchant la communauté mathématique – épisode historico-scientifique d'une grande complexité qui voit toute la communauté se déchirer, et contribue à isoler la Russie (parmi les nombreux griefs faits à Luzin il y avait celui de trop publier en langues étrangères...). Au-delà de ces convulsions, la théorie de Souslin a effectivement prospéré, en particulier grâce aux communautés mathématiques russe, polonaise et japonaise, jusqu'à aboutir dans les années 1970 à un paysage cohérent et puissant, riche de problèmes intrinsèques et toujours actif, la **théorie descriptive des ensembles**.

Le concept central dans cette théorie est celui d'**ensemble analytique**, parfois appelé ensemble souslinien : une notion assez proche mais plus générale que celle d'ensemble borélien, et très stable. (Ce concept n'a rien à voir avec celui de fonction analytique d'une variable réelle ou complexe.) Il en existe plusieurs définitions équivalentes dont voici la plus commode :

Définition V-1 (ensemble analytique). *On dit qu'un ensemble A dans un espace polonais est analytique s'il peut s'écrire sous la forme $f(\mathbb{N}^{\mathbb{N}})$, où $\mathbb{N}$ est muni de la topologie discrète, $\mathbb{N}^{\mathbb{N}}$ de la topologie produit, et f est continue.*

On rappelle qu'un espace est dit polonais quand il est métrique, séparable et complet ; c'est le cadre dans lequel se déploie presque entièrement la théorie des ensembles analytiques.

On peut se représenter un ensemble analytique A comme étant "descriptible" au moyen de $\mathbb{N}^{\mathbb{N}}$: pour repérer un point $y \in A$, il suffit de se donner une suite à valeurs entières : en gros, chacune des valeurs de cette suite indique l'appartenance de y à l'un des ensembles d'une famille dénombrable de plus en plus fine. C'est ainsi une suite de choix qui permet de repérer le point, un peu comme quand on cherche un mot dans un dictionnaire en déterminant d'abord la première lettre, puis la seconde et ainsi de suite (tenter de se représenter une infinité dénombrable de lettres, des mots d'une longueur infinie, et une quantité non dénombrable de mots !).

Dans ce chapitre on va passer en revue certains des concepts centraux de la théorie, concluant par les importants théorèmes de sélection. Certaines preuves ne seront pas données en intégralité ; on renvoie à [Dellacherie, Kechris, Melleray, Miller, Parthasarathy, Tserunyan] pour des exposés plus complets.

V-1.*Description d'un espace polonais

Si (X, d) est un espace polonais, alors on peut le recouvrir par une famille dénombrable de boules (ouvertes ou fermées) de rayon $r > 0$ arbitrairement petit, disons $(B_n)_{n\in\mathbb{N}}$. On peut aussi transformer ce recouvrement en partition en posant $B_1' = B_1$, $B_2' = B_2 \setminus B_1$, $B_3' = B_3 \setminus (B_1 \cup B_2)$, etc. Les ensembles B_i' ne seront plus alors fermés, mais ils seront disjoints deux à deux. Si $x \in B_n$ (ou $x \in B_n'$), l'indice n donnera une première indication grossière de la position de x.

On peut ensuite itérer en raffinant ce recouvrement : chaque B_i pourra être recouvert par des boules fermées $B_{i,j}$ de rayon $r/2$ (chaque B_{ij} est une boule dans l'espace fermé B_i ; cela n'en fait pas une boule dans X, mais tout de même un ensemble fermé de diamètre au plus r). Ou bien chaque B_i' peut être recouvert par des boules ouvertes $\widetilde{B}_{ij}$ dans X, de rayon $r/2$, et en intersectant ces boules avec B_i', puis en les transformant en partition par le procédé habituel, on aura alors recouvert B_i' par des ensembles B_{ij}' pas forcément fermés, mais disjoints et de diamètre au plus r.

On peut ainsi définir inductivement une famille d'ensembles $A_{n_1,n_2,\ldots,n_k}$, où chaque indice n_i est un entier et k est un entier arbitraire ; tous ces ensembles appartenant à l'algèbre engendrée par les boules ouvertes ; leur union, pour chaque k fixé, recouvrant X ; et leur diamètre, pour k fixé, étant au plus de 2^{-k}.

Il y a bien des façons de varier la construction. On est parti des boules, on aurait pu choisir une autre famille dénombrable recouvrant X ; ou choisir une autre suite de diamètres tendant vers 0 ; mais ce qui compte le plus, ce sont les deux propriétés caractéristiques

- $A_{n_1,n_2,\ldots,n_{k+1}} \subset A_{n_1,n_2,\ldots,n_k}$;
- $\operatorname{diam}(A_{n_1,n_2,\ldots,n_k}) \xrightarrow[k\to\infty]{} 0$ pour toute suite $(n_1, n_2, \ldots)$,

qui font des A_n un **système souslinien régulier**. Ici n varie dans l'espace $\mathbb{N}^{<\mathbb{N}}$ de toutes les suites finies d'entiers.

Et selon que l'on travaille avec les familles B_s ou B'_s ci-dessus évoquées, on peut en outre imposer que pour chaque $(n_1, \ldots, n_k)$ les $(A_{n_1,\ldots,n_{k+1}})_{n_{k+1}\in\mathbb{N}}$ forment une partition de $A_{n_1,\ldots,n_k}$ (c'est alors un **système de Lusin**) ; ou encore que tous les ensembles $(A_{n_1,\ldots,n_k})$ soient fermés.

Que l'on travaille avec l'une ou l'autre de ces variantes, ou d'autres encore, cela permet de définir sans ambiguïté tout point $x \in X$ par une liste des indices des parties auxquelles il appartient : une suite $(n_1, n_2, \ldots, n_k, \ldots) \in \mathbb{N}^{\mathbb{N}}$, telle que $x \in A_{n_1,\ldots,n_k}$ (que l'on notera aussi $A_{n_1 n_2 \ldots n_k}$ pour tout k). Avec un système souslinien de fermés, l'intersection sera non vide pour tout choix de suite (n_i) (grâce au théorème des fermés emboîtés), mais un même x pourra éventuellement être représenté par plusieurs suites. Si pour chaque k fixé les ensembles $A_{n_1\ldots n_k}$ sont distincts l'intersection pourra être parfois vide, mais chaque x correspond à une seule suite (n_i). Et si c'est un système lusinien, on a finalement une bijection entre $\mathbb{N}^{\mathbb{N}}$ et X. Dans tous les cas, cette technique de description permet de paramétrer X par une partie de $\mathbb{N}^{\mathbb{N}}$, et le repérage de ce point s'apparente à une suite dénombrable de choix dénombrables (on verra des énoncés plus précis dans quelques instants).

FIGURE

L'application qui à $n = (n_1, n_2, \ldots)$ associe x est continue si $\mathbb{N}^{\mathbb{N}}$ est muni de la topologie produit (chaque facteur $\mathbb{N}$ est muni de la topologie discrète). Cet espace emblématique, appelé **espace de Baire** est une figure centrale de la théorie.

Définition V-2 (espace de Baire). *On appelle espace de Baire l'ensemble* $\mathcal{N} = \mathbb{N}^{\mathbb{N}}$ *de toutes les suites entières, muni de la topologie produit ; c'est un espace polonais totalement discontinu.*

Bien sûr on n'a pas défini explicitement de métrique sur $\mathcal{N}$, mais on a de multiples choix induisant la même topologie, par exemple

$$d(m, n) = \sum_{k\in\mathbb{N}} 2^{-k} \frac{|n_k - m_k|}{1 + |n_k - m_k|},$$

ainsi la convergence de n^j vers n signifie que pour tout k il existe J tel que $n^j_k = n_k$ pour $j \geq J$; et $(\mathcal{N}, d)$ est bien un espace séparable et complet. Il s'ensuit naturellement une base dénombrable de voisinages : tous les $\{m_1, \ldots, m_k\} \times \mathcal{N}$, où $k \in \mathbb{N}$ et $m_i \in \mathbb{N}$. L'espace dénombrable qui les indexe naturellement est l'ensemble des suites finies d'entiers, qui jouera un rôle important dans la suite :

$$\mathbb{N}^{<\mathbb{N}} = \bigcup_{k\in\mathbb{N}} \mathbb{N}^k. \tag{46}$$

On peut se représenter le processus de description (la correspondance avec l'espace de Baire) de plusieurs façons équivalentes : avec des unions et intersections de parties comme on vient de le faire ; ou bien par un processus de repérage le long des branches d'un arbre (chaque nouvel indice correspond à un branchement de l'arbre), ou encore par des familles de conditions logiques (conjonction pour l'intersection de parties, etc). Dans ce chapitre on se contentera de la première approche, mais les arbres et les familles de conditions font aussi partie de la panoplie de la théorie descriptive de la mesure, de même que les jeux ensemblistes (on imagine que deux joueurs, en compétition l'un avec l'autre, font des choix successifs d'ensembles de façon à atteindre un certain objectif, et on se demande si l'un ou l'autre a une stratégie gagnante).

Pour les espaces métriques compacts (qui sont a fortiori polonais), la même construction s'applique évidemment ; mais en outre, par compacité, on peut se limiter à un nombre fini de parties à chaque étape. Quitte à ajouter des étapes intermédiaires, on peut en fait se ramener à la situation où à chaque étape l'ensemble $A_{n_1,\dots,n_k}$ est subdivisé en seulement deux parties $A_{n_1,\dots,n_{k+1}}$. Ainsi, pour décrire un compact, on peut se limiter à un espace plus petit que $\mathcal{N}$, qui lui ressemble beaucoup mais qui est compact : c'est l'**espace de Cantor**.

DÉFINITION V-3 (espace de Cantor). *On appelle espace de Cantor l'ensemble* $\mathcal{C} = \{0,1\}^{\mathbb{N}}$ *de toutes les suites de 0 et de 1, muni de la topologie produit ; c'est un espace métrique compact totalement discontinu.*

Une base de voisinages de $\mathcal{C}$ est fournie par les ensembles $\{a_1,\dots,a_k\} \times \mathcal{C}$ où les a_i appartiennent à l'ensemble $\{0,1\}$, que l'on peut aussi noter 2 ; ainsi l'espace dénombrable qui indexe cette base de voisinages est

$$2^{<\mathbb{N}} = \bigcup_{k\in\mathbb{N}} \{0,1\}^k. \tag{47}$$

L'espace de Baire et l'espace de Cantor ont même cardinalité (il n'est pas difficile de trouver une surjection de $\mathcal{C}$ dans $\mathcal{N}$), à savoir c, la même que $\mathbb{R}$ (c comme "continu"). Par opposition, $\mathbb{N}^{<\mathbb{N}}$ et $2^{<\mathbb{N}}$ ont la cardinalité de $\mathbb{N}$, soit $\aleph_0$.

Nous avons maintenant tous les outils conceptuel pour démontrer le

THÉORÈME V-4 (description des espaces polonais).
(a) Tout espace polonais est image par une application continue de l'espace de Baire $\mathcal{N}$ *;*
(b) Tout espace polonais est image par une bijection continue d'un fermé de $\mathcal{N}$ *;*
(c) Tout espace métrique compact est image par une application continue de l'espace de Cantor $\mathcal{C}$*.*

REMARQUES V-5. (i) La bijection en (b) est continue mais son inverse a priori ne l'est pas : il s'agit donc d'un plongement continu, mais pas d'un homéomorphisme. Par exemple, $\mathbb{R}$ n'est homéomorphe à aucun sous-espace de l'espace de Baire (pourquoi ?).

(ii) Tout espace métrique compact n'est pas image par une bijection continue d'un fermé de $\mathcal{C}$. En effet, si cette bijection f existait, d'un sous-espace fermé F de $\mathcal{C}$ dans X, alors F serait compact, donc f enverrait les fermés de F dans les fermés de X, donc les ouverts aussi ; donc f^{-1} serait continue, et f serait un homéomorphisme, préservant toutes les propriétés topologiques de $\mathcal{C}$. Or $\mathcal{C}$ est totalement discontinu, ce qui n'est pas le cas des compacts en général. Noter que l'application naturelle de $\mathcal{C}$ dans $[0,1]$, à savoir $x \longmapsto \sum 2^{-n}x_n$, est continue mais non bijective (pourquoi ?)

EXERCICE V-6. Dans le cas où $X = \mathbb{R}$, construire une bijection satisfaisant (b).

Exercice V-7. Construire une application mesurable surjective de $\mathcal{C}$ dans $\mathcal{N}$; vérifier directement qu'elle est discontinue.

Preuve du Théorème V-4. Ces résultats sont obtenus par des variantes de schémas de Souslin.

Pour (a) on recouvre le polonais X par des boules ouvertes B_i ; quitte à répéter une infinité de fois certaines boules, on peut supposer que c'est une famille dénombrable infinie ; puis on recouvre chacune de ces boules par des boules ouvertes B_{ij} telles que $\overline{B_{ij}} \subset B_i$, et ainsi de suite ; à chaque étape k on impose que le rayon de ces boules soit majoré par 2^{-k}. Ainsi le théorème des fermés emboîtés garantit que l'intersection des $B_{i_1\ldots i_k}$ est réduit à un point, et ce pour tout suite d'entiers $(i_1, i_2, \ldots)$. Pour $n \in \mathcal{N}$ on définit alors $f(n)$ par

$$\{f(n)\} = \bigcap_{k\in\mathbb{N}} B_{n_1 n_2\ldots n_k}.$$

Pour (b) on modifie la construction en prenant une partition borélienne à chaque étape ; il y a une subtilité. Voici une première tentative. Pour la première étape : si (B_i) est un recouvrement par des boules de diamètre au plus 1, on pose $B'_1 = B_1$, $B'_i = B_i \setminus (B'_1 \cup B'_2 \ldots \cup B'_{i-1})$. On peut continuer ainsi en recouvrant chaque B'_i en boules ouvertes de rayon plus petit et en transformant ce recouvrement à nouveau en partition B'_{ij} de B'_i. Et ainsi de suite, de sorte que $B'_{i_1 i_2\ldots i_k} = \bigcup_{\ell\in\mathbb{N}} B'_{i_1 i_2\ldots i_{k+1}}$, et on impose de même que le diamètre des ensembles $B'_{i_1\ldots i_k}$ soit au plus 2^{-k}. Ainsi tout $x \in X$ est repéré par une suite unique $(i_1, i_2, \ldots)$, et l'on peut définir

$$F = \Big\{ n \in \mathcal{N};\ \bigcap_{k\in\mathbb{N}} B'_{n_1 n_2\ldots n_k} \neq \emptyset \Big\}.$$

Mais on se retrouve alors dans une impasse pour prouver que F est fermé !

On modifie donc la construction des partitions pour imposer la condition supplémentaire

$$B'_{i_1\ldots i_k} = \bigcup_{\ell\in\mathbb{N}} \overline{B'_{i_1\ldots i_k \ell}}.$$

Pour cela on suppose par induction que chaque $B'_{i_1\ldots i_k}$ est une union dénombrable de fermés (un F_σ). La propriété élémentaire que l'on utilise est que : *Si C est un F_σ, alors il peut s'écrire comme union dénombrable disjointe de F_σ de diamètre arbitrairement petit, dont l'adhérence est incluse dans C.*

Pour cela on écrit $C = \bigcup F_j$, où chaque F_j est un fermé ; sans perte de généralité cette union est croissante ; alors C est l'union disjointe des $C_{j+1} \setminus C_j$, qui à leur tour peuvent s'écrire comme l'union des fermés $C_{j+1} \setminus C_j^{(m)}$, où $C_j^{(m)}$ est le voisinage ouvert de C_j fait des boules ouvertes de rayon $1/m$ centrées en C_j ; on décompose chacun de ces fermés en une quantité dénombrable de fermés de diamètre au plus ε par le procédé habituel, on les transforme en partition constituée de F_σ par le procédé habituel, les adhérences de tous les ensembles ainsi formés sont inclus dans C_j, et donc dans C. On ne garde bien sûr que les ensembles non vides.

Une fois que cette condition supplémentaire est assurée, l'argument va de soi. Pour tout $n \in F$, l'intersection $\bigcap_{k\in\mathbb{N}} B'_{n_1\ldots n_k}$ a un diamètre nul, donc elle est réduite à un point, que l'on note $f(n)$. Par construction $f(F) = X$, et il est facile de vérifier que f est une bijection continue. Par construction $\bigcap B'_{i_1\ldots i_k} = \bigcap \overline{B'}_{i_1\ldots\ldots i_k}$, cela et le théorème des fermés emboîtés montre que toute suite de Cauchy d'éléments de F converge dans $\mathcal{N}$, ainsi F est bien fermé.

Enfin pour (c) on raisonne comme en (a), mais par compacité chaque étape k ne fait intervenir qu'un nombre fini d'ouverts, disons N_k ; on choisit alors m_k tel que $2^{m_k} \geq N_k$, et on choisit une surjection de $\{0,1\}^{m_k}$ dans $\{1,\ldots,N_k\}$. Alors la construction précédente réalise X comme image continue de $\prod_{k\in\mathbb{N}}\{0,1\}^{m_k}$, mais cela est en fait la même chose que $\mathcal{C}$. □

On traduit le Théorème V-4 en disant qu'un espace polonais, et a fortiori un espace métrique compact, sont **analytiques**. Cette notion sera développée dans la section suivante, mais notons d'ores et déjà un corollaire à la fois surprenant et utile sur la nature des boréliens :

Corollaire V-8 (Boréliens comme plongements de fermés). *Tout borélien d'un espace polonais X est l'image d'un fermé F de $\mathcal{N}$ par une injection continue de F dans X.*

Preuve du Corollaire V-8. Soit B un borélien de X. Par l'Exercice III-6, B est image d'un fermé F_1 (de X muni d'une structure polonaise enrichie) par une injection continue. Ce fermé F_1 est lui-même un espace polonais, donc image d'un fermé de $\mathcal{N}$ par une injection continue, d'après le Théorème V-4. □

V-2*Ensembles analytiques

Définition V-9 (analyticité). *Soit X un espace polonais. Une partie A de X est dite analytique s'il existe une application continue $f : \mathcal{N} \to X$ telle que $f(\mathcal{N}) = A$.*

Plus généralement, un espace topologique Y est dit analytique (ou souslinien) s'il est homéomorphe à une partie analytique d'un espace polonais.

Au vu du Théorème V-4, il est équivalent de définir les ensembles (ou les espaces) analytiques comme les images des applications continues définies entre espaces polonais ; de sorte que le choix de l'espace $\mathcal{N}$ dans la définition précédente n'est pas aussi arbitraire qu'il semble.

On peut décrire les ensembles analytiques par des systèmes de Souslin réguliers $(P_n)_{n\in\mathbb{N}^{<\mathbb{N}}}$, avec un procédé similaire à celui de la Section V. Et réciproquement, si l'on se donne un système de Souslin régulier dans X, on peut en reconstituer un ensemble analytique, par l'opération $(\mathcal{A})$ de Souslin :

$$\mathcal{A}(P) = \bigcup_{n\in\mathcal{N}} \bigcap_{k\in\mathbb{N}} P_{n_1,\ldots,n_k}. \tag{48}$$

Noter que même si les P_n sont ouverts (ou fermés), $\mathcal{A}(P)$ n'est pas a priori borélien, puisque l'union est prise sur un ensemble non dénombrable. Il existe une quantité vertigineuse de systèmes (P) qui aboutissent au même ensemble $\mathcal{A}(P)$, parfois appelé le noyau du système.

Il y a de la flexibilité dans les hypothèses sur les parties P_n, que l'on peut choisir dans une classe commode Γ : par exemple l'ensemble des boules ouvertes, ou l'ensemble des boules fermées, ou l'ensemble des ouverts, ou l'ensemble des fermés, ou l'ensemble des intersections dénombrables d'ouverts, ou même l'ensemble des parties analytiques : tout cela mène toujours, via l'opération $(\mathcal{A})$, aux ensembles analytiques. En fait l'opération de Souslin est idempotente : partant d'une famille quelconque Γ, on a

$$\mathcal{A}\mathcal{A}(\Gamma) = \mathcal{A}(\Gamma).$$

Cela vient de ce que si l'on itère l'opération $(\mathcal{A})$, on va être amené à choisir, pour chaque $n \in \mathcal{N}$, une suite de parties indexées elles-mêmes par $\mathcal{N}$; autrement dit le résultat sera indexé par $\mathcal{N} \times \mathcal{N}^{\mathbb{N}}$; mais cet ensemble (encore un paradoxe de l'infini !) est lui-même homéomorophe à $\mathcal{N}$.

On a aussi de la flexibilité dans la convergence de ces suites de parties. Quand on se donne un système régulier de Souslin, on impose seulement que le diamètre de $P_{n_1,\ldots,n_k}$ tende vers 0 le long de chaque $n \in \mathbb{N}$; mais quitte à subdiviser, on peut aussi imposer une convergence uniforme en k, par exemple $\operatorname{diam}(P_{n_1,\ldots,n_k}) \le 2^{-k}$. Si les P_n sont fermés, l'intersection est automatiquement non vide par principe des fermés emboîtés. Même sans l'hypothèse de fermeture, on peut choisir les P_n de façon à ce que l'intersection soit non vide : si f est une application continue surjective de $\mathcal{N}$ dans A, il suffit de choisir

$$P_{n_1,\ldots,n_k} = f\big(\{n_1,\ldots,n_k\} \times \mathcal{N}\big); \tag{49}$$

ainsi on a bien

$$\forall n \in \mathcal{N}, \qquad f(n) \in \bigcap_{k\in\mathbb{N}} P_{n_1,\ldots,n_k}.$$

Noter que dans ce cas le diamètre de $P_{n_1,\ldots,n_k}$ tend bien vers 0 (mais pas forcément uniformément) car les $\{n_1,\ldots,n_k\} \times \mathcal{N}$ constituent une base de voisinages de n pour la topologie produit.

En maniant les propriétés de restriction et de recollement des fonctions boréliennes induits (Proposition III-8), on démontre que la notion d'analyticité est intrinsèque : Si A est une partie quelconque d'un espace polonais (X,d), c'est équivalent de dire que A est un espace souslinien, ou qu'il est analytique dans (X,d), ou qu'il est obtenu par application du schéma $(\mathcal{A})$ de Souslin à partir d'une famille de parties fermées (ou ouvertes, ou boréliennes, ou analytiques) dans X.

On peut avoir une autre intuition de l'opération $(\mathcal{A})$ en examinant le cas particulier d'un système lusinien : alors (exercice)

$$\mathcal{A}(P) = \bigcap_{k\in\mathbb{N}} \bigcup_{m\in\mathbb{N}^k} P_{m_1,\ldots,m_k}, \tag{50}$$

qui est une intersection décroissante d'unions d'ensembles élémentaires de plus en plus fins (imaginer qu'on commence par une description grossière de l'ensemble, puis de plus en plus fine, et ainsi de suite). Noter que si les P_m sont boréliens, alors (50) définit un ensemble borélien ; mais cela est lié à l'hypothèse que le système est lusinien.

La force des ensembles analytiques tient à ce que, tout en étant plus généraux et plus stables que les boréliens, ils en restent “assez proches”, tant par les liens entre les deux notions que par la similitude de leurs propriétés.

Théorème V-10 (Propriétés des ensembles analytiques). *(a) Si X est un espace polonais, les parties analytiques de X forment une famille de cardinalité au plus c, la puissance du continu.*

(b) Si $f : X \to Y$ est une application continue entre espaces polonais, alors les images par f des ensembles analytiques sont analytiques, et les images réciproques par f des ensembles analytiques sont analytiques ; de même si f est seulement supposée borélienne et définie sur une partie borélienne B de X.

(c) Une union dénombrable de parties analytiques est analytique ; de même pour une intersection dénombrable de parties analytiques.

(d) Si X est un espace polonais, tout borélien de X est aussi analytique ; mais dès que X est non dénombrable, il existe des ensembles analytiques non boréliens.

(e) Si X et Y sont deux espaces polonais, et B un borélien de $X \times Y$, alors la projection de B sur X est analytique ; en outre les parties analytiques de X sont exactement les projections sur X des boréliens de $X \times \mathcal{N}$.

Remarque V-11. (i) L'énoncé (e) implique en particulier que les projections des boréliens de $\mathbb{R}^2$ dans $\mathbb{R}$ sont des ensembles analytiques, à défaut d'être boréliens ; c'était le point de départ de la réflexion de Souslin ;

(ii) On a déjà vu (Exercice III-6) que tout borélien est image d'un fermé, et même d'un ouvert fermé, par une injection continue. On a maintenant une généralisation partielle : de par leur définition, tous les analytiques sont des images continues d'ensembles ouverts fermés.

(iii) En résumé, la propriété d'analyticité est stable par union, intersection dénombrable, image réciproque (par une fonction continue, ou borélienne), comme les boréliens ; mais elle est aussi stable par image directe. En revanche, elle *n'est pas stable* par passage au complémentaire. On appelle **coanalytiques** les complémentaires des ensembles analytiques ; ils vérifient des propriétés bien différentes des analytiques et ce n'est pas la même intuition qui leur est associée. Boréliens, analytiques et coanalytiques sont les trois premières classes de la classification de Lusin : Classe 0, les boréliens ; classe 1, les images continues des boréliens (donc les analytiques) ; classe 2, les complémentaires des analytiques (donc les coanalytiques) ; classe 3, les images continues de la classe 2 ; classe 4, les complémentaires de la classe 3 ; etc.

Preuve (pas tout à fait complète) du Théorème V-10. 1. Un espace polonais admet une base dénombrable de voisinages, disons $(V_k)_{k\in\mathbb{N}}$. Un ouvert O est décrit par les voisinages V_k qu'il contient ; la cardinalité de la topologie de l'espace est donc au plus celle de l'ensemble $\{0,1\}^{\mathbb{N}}$. Cela est vrai en particulier des ouverts de $\mathcal{N}$.

2. Une fonction continue de $f : \mathcal{N} \to X$ est déterminée par les ouverts $f^{-1}(V_k)$, où les V_k forment une base dénombrable de voisinages de X. Donc il y a au plus autant de fonctions continues de $\mathcal{N}$ dans X, que de suites à valeurs dans les ouverts, ou de façon équivalente à valeurs dans $\{0,1\}^{\mathbb{N}}$. Mais cela est aussi c, la puissance du continu (car $\{0,1\}^{\mathbb{N}\times\mathbb{N}} \simeq \{0,1\}^{\mathbb{N}}$). Cela prouve (a).

3. Il est évident, par la Définition V-9, que l'image continue d'un ensemble analytique est aussi analytique.

4. Maintenant si $f : X \to Y$ est seulement borélienne, on peut toujours enrichir la topologie de X en lui ajoutant tous les $f^{-1}(V_n)$, pour V_n dans une base de voisinages de Y. L'espace ainsi enrichi par la topologie engendrée reste polonais, et la fonction f devient continue grâce à sa nature borélienne (c'est l'Exercice III-6). Par l'étape 3, il en resulte que l'image directe d'un ensemble analytique par f est analytique.

5. Si les $A_i = f_i(\mathcal{N})$ ($i \in \mathbb{N}$) sont analytiques, on se donne des copies disjointes $\mathcal{N}_i$ de $\mathcal{N}$, alors $\cup f_i$ (définie sur l'union des $\mathcal{N}_i$ et coïncidant avec f_i sur chaque $\mathcal{N}_i$) envoie continûment $\cup\mathcal{N}_i$ sur $\cup A_i$. Mais $\cup\mathcal{N}_i \simeq \mathbb{N}\times\mathcal{N}$ peut aussi s'écrire comme image continue de $\mathcal{N}$. Cela prouve la stabilité par union dénombrable. De façon similaire, on définit $\cap f_i$ sur l'espace polonais $Y = \{(x_1, x_2, \ldots); f_1(x_1) = f_2(x_2) = \ldots\}$ par

$(\cap f_i)(x_1,\ldots) = f_1(x_1)$ (qui est en fait $f_k(x_k)$ pour tout k) et cela envoie continûment Y sur $\cap A_i$. Cela conclut le point (c).

6. Les analytiques sont stables par union et intersection dénombrables, et contiennent bien sûr les ensembles fermés (qui sont des espaces polonais !) et les ensembles ouverts (union dénombrables de boules fermées). Par l'Exercice II-14, la classe des ensembles analytiques contient la classe des boréliens. Cela montre la première partie de (d) ; je ne prouverai pas la seconde partie, mais citerai des exemples particuliers en fin de section.

7. La projection est continue de $X \times Y$ dans X (ou dans Y), les projections sur X (ou sur Y) des boréliens dans l'espace produit sont donc analytiques.

8. Si f est continue $\mathcal{N} \to X$, son graphe est borélien, et la projection de ce graphe est $f(\mathcal{N})$: cela prouve que tout analytique de X est la projection d'un graphe borélien dans $\mathcal{N} \times X$ (ou d'un borélien dans $X \times \mathcal{N}$, en inversant les facteurs).

9. Soit f mesurable. Par le Corollaire III-13, son graphe $G(f)$ est mesurable dans $X \times Y$. Si A est analytique, $f^{-1}(A)$ est la projection sur X de l'ensemble analytique $G(f) \cap X \times A$, c'est donc un ensemble analytique. Cela avec l'étape 4 achève de prouver (b). $\square$

Comme on l'a vu, les ensembles analytiques forment une classe plus générale que les ensembles boréliens, mais ils ne sont pas plus nombreux ; pour exhiber des ensembles analytiques non boréliens, on ne peut donc se contenter d'un argument de cardinalité. On a déjà mentionné en passant un contre-exemple :

Exemple V-12 (analytiques non boréliens). L'ensemble de Lusin du Théorème VI-47 est analytique dans $\mathbb{R}$, non borélien. On notera que sa définition fait intervenir des suites infinies de suites, c'est à dire une paramétrisation par $\mathcal{N}$; le développement en fraction rationnelle étant le moyen de mettre $\mathbb{R}$ en correspondance avec $\mathcal{N}$.

Bien sûr, cet ensemble est artificiellement conçu et sa définition assez tourmentée. En fait il est difficile de trouver des contre-exemples à la fois explicites et assez naturels. La liste ci-dessous regroupe quasiment tous ceux qui sont connus ; elle est entièrement due à l'école de topologie polonaise de l'entre-deux guerres, très active à Varsovie et Wrocław (Stefan Mazurkiewicz, Witold Hurewicz, Zygmunt Janiszewski, Edward Marczewski (Szpilrajn), Kazimierz Kuratowski, ...)

Exemple V-13 (analytiques non boréliens, encore). (i) (Mazurkiewicz) L'ensemble des fonctions dérivables de $[0,1]$ dans $\mathbb{R}$ est coanalytique, non borélien dans $C([0,1];\mathbb{R})$.
(ii) (Hurewicz) L'ensemble des compacts non dénombrables de $[0,1]$ est analytique, non borélien dans l'espace de Hausdorff des compacts de $[0,1]$.
(iii) (Kuratowski–Marcewski) L'ensemble des compacts de $[0,1]$ contenus dans $\mathbb{Q}$ est coanalytique, non borélien dans l'espace de Hausdorff.

V-3*De l'analyticité à la mesurabilité

Explorons maintenant les liens entre analyticité et mesurabilité. Pour commencer, l'énoncé qui suit indique que l'on peut distinguer les ensembles analytiques via des boréliens, bien que la structure de ces derniers soit moins fine.

Théorème V-14 (Théorème de séparation de Lusin). *Soient (X,d) un espace polonais et A^1, A^2 deux ensembles analytiques disjoints. Alors il existe des boréliens disjoints B^1 et B^2 tels que $A^i \subset B^i$ pour $i = 1,2$. De même si X est souslinien.*

Remarques V-15. (i) Cet énoncé se généralise facilement à des familles dénombrables d'ensembles analytiques deux à deux disjoints ($\forall i \neq j,\ A_i \cap A_j = \emptyset$) : on peut alors les inclure dans une famille dénombrable d'ensembles boréliens deux à deux disjoints (Exercice). C'est plus délicat de le généraliser pour des familles globalement disjointes ($\bigcap_{j\in\mathbb{N}} A_j = \emptyset$) ; on y reviendra.

(ii) Ce théorème est faux si l'on remplace "analytiques" par "coanalytiques" (d'après un contre-exemple du mathématicien russe Sergueï Novikov, un des plus importants contributeurs à la théorie descriptive des ensembles).

Preuve du Théorème V-14. Le raisonnement est le même que X soit supposé polonais ou plus généralement souslinien. Soient A^1 et A^2 deux ensembles analytiques disjoints dans X ; on les suppose tous deux non vides, sans quoi le résultat est évident. Écrivons-les sous forme de schémas de Souslin réguliers :

$$A^i = \bigcup_{n\in\mathcal{N}} \bigcap_{k\in\mathbb{N}} P^i_{n_1\ldots n_k}$$

où pour tout $n \in \mathcal{N}$ l'intersection $\bigcap_{k\in\mathbb{N}} P^i_{n_1\ldots n_k}$ se réduit à exactement un élément de A^i (Cf (49)).

Supposons par l'absurde qu'il n'existe aucune paire (B^1, B^2) de boréliens disjoints qui sépare (A^1, A^2), au sens où $A^i \subset B^i$. Pour $i = 1, 2$ et $n_1 \in \mathbb{N}$ on pose alors

$$A^i_{n_1} = \bigcup_{(n_2,n_3,\ldots)\in\mathcal{N}} \bigcap_{k\in\mathbb{N}} P^i_{n_1 n_2\ldots n_k},$$

de sorte que

$$A^1 = \bigcup_{n_1\in\mathbb{N}} A^1_{n_1}, \qquad A^2 = \bigcup_{m_1\in\mathbb{N}} A^2_{m_1}.$$

Si tous les couples $(A^1_{n_1}, A^2_{m_1})$ pouvaient être séparés par des boréliens disjoints $C^1_{n_1}$ et $C^2_{m_1}$, alors on pourrait séparer A^1 et A^2 par les boréliens disjoints $D^1_1 = \bigcup_{n_1\in\mathbb{N}} C^1_{n_1}$ et $D^2_1 = \bigcup_{m_1\in\mathbb{N}} C^2_{m_1}$; mais cela contredirait notre hypothèse. Il doit donc exister au moins un n_1 et un m_1 tels que $A^1_{n_1}$ et $A^2_{m_1}$ ne peuvent être séparés par aucun couple de boréliens disjoints. Comme le système est régulier on a $A^1_{n_1} \subset P^1_{n_1}$ et $A^2_{m_1} \subset P^2_{m_1}$, donc $P^1_{n_1} \cap P^2_{m_1} \neq \emptyset$.

On peut alors recommencer le raisonnement, et par récurrence construire $(n_1, n_2, \ldots)$ et $(m_1, m_2, \ldots)$, et

$$A^1_{n_1 n_2\ldots n_\ell} = \bigcap_{s\in\mathcal{N}} \bigcap_{k\in\mathbb{N}} P^1_{n_1\ldots n_\ell s_1\ldots s_k} \subset P^1_{n_1 n_2\ldots n_\ell},$$

$$A^2_{m_1 m_2\ldots m_\ell} = \bigcap_{s\in\mathcal{N}} \bigcap_{k\in\mathbb{N}} P^2_{m_1\ldots m_\ell s_1\ldots s_k} \subset P^2_{m_1 m_2\ldots m_\ell},$$

tels que (a) $A^1_{n_1\ldots n_\ell}$ et $A^2_{m_1\ldots m_\ell}$ ne peuvent être séparés par aucune paire de boréliens disjoints, et (b) $P^1_{n_1\ldots n_\ell} \cap P^2_{m_1\ldots m_\ell} \neq \emptyset$.

Par construction il existe $a^1 \in A^1$ et $a^2 \in A^2$ tels que

$$\bigcap_{\ell\in\mathbb{N}} P^1_{n_1\ldots n_\ell} = \{a^\ell_1\} \qquad \bigcap_{\ell\in\mathbb{N}} P^2_{m_1\ldots m_\ell} = \{a^\ell_2\}.$$

En particulier $a^1 \neq a^2$ puisque les A^i sont disjoints. Soient alors B^1 et B^2 des boules ouvertes disjointes telles que $a^i \in B^i$. Pour ℓ assez grand on a $A^1_{n_1 n_2\ldots n_\ell} \subset B^1$ et

$A^2_{m_1 m_2 \dots m_\ell} \subset B^2$. Mais cela contredit le fait que $A^1_{n_1 \dots n_\ell}$ et $A^2_{m_1 \dots m_\ell}$ ne sont séparables par aucune paire de boréliens, concluant le raisonnement par l'absurde. $\square$

EXERCICE V-16. Mettre en œuvre le même argument sans schéma de Souslin, en partant de la définition des analytiques comme images continues de $\mathcal{N}$.

V-3.1. Caractérisations et inclusions. Ce cours est maintenant mûr pour les deux principaux résultats liant analyticité et mesurabilité.

THÉORÈME V-17 (Caractérisation de Souslin des boréliens). *Soit (X, d) un espace polonais. Alors $A \subset X$ est borélien si et seulement si il est analytique et coanalytique dans X. De même si X est un espace souslinien.*

THÉORÈME V-18 (mesurabilité universelle des analytiques). *Soit (X, d) un espace polonais et μ une mesure borélienne σ-finie sur X. Alors tout ensemble analytique $A \subset X$ est μ-mesurable.*

On rappelle qu'une partie A est μ-mesurable si elle appartient à la tribu borélienne complétée pour μ, c'est à dire si A s'écrit commel 'union d'un borélien et d'un ensemble μ-négligeable. Le Théorème V-18 dit que l'ensemble A sera μ-mesurable pour *toute* mesure μ σ-finie : on dit alors que A est **universellement mesurable**. On a de la flexibilité sur l'hypothèse de σ-finitude :

DÉFINITION V-19 (mesurabilité universelle). *Soit (X, d) un espace polonais. On dit que $A \subset X$ est universellement mesurable s'il est μ-mesurable pour toute mesure μ borélienne σ-finie sur X, ou de façon équivalente pour toute mesure de probabilité borélienne sur X.*

EXERCICE V-20. Prouver que les deux formulations sont effectivement équivalentes.

Le tableau est donc le suivant (pour toute mesure σ-finie μ, par exemple la mesure de Lebesgue sur $\mathbb{R}^n$) :

$$\{\text{Ouverts, Fermés}\} \subset \{\text{Boréliens}\} \subset \{\text{Analytiques}\} \subset \{\mu\text{-mesurables}\}$$
$$\{\text{Ouverts, Fermés}\} \subset \{\text{Boréliens}\} \subset \{\text{Coanalytiques}\} \subset \{\mu\text{-mesurables}\}$$
$$\{\text{Boréliens}\} = \{\text{Analytiques}\} \cap \{\text{Coanalytiques}\}$$

REMARQUE V-21. Dans $\mathbb{R}^n$ il y a $c = 2^{\aleph_0}$ ensembles boréliens et 2^c ensembles Lebesgue-mesurables ; combien d'ensembles universellement mesurables ? On ne sait pas exactement, mais un résultat récent de Larson–Neeman–Shelah nous dit qu'il n'y a pas de contradiction à supposer que c'est c. La classe des universellement mesurables reste donc en un sens “bien plus petite” que la classe des mesurables.

PREUVE DU THÉORÈME V-17. Si A et $X \setminus A$ sont analytiques, alors par le Théorème V-14 de séparation de Lusin, on trouve des boréliens disjoints B et B' tels que $A \subset B$ et $X \setminus A \subset B'$. Mais alors

$$B \subset X \setminus B' \subset A \subset B,$$

et donc $A = B$. $\square$

La preuve du Théorème V-18 demandera quelques préparatifs impliquant la notion d'**enveloppe**.

Définition V-22 (enveloppe). *Soit* $(X, \mathcal{B}, \mu)$ *un espace mesuré et* $A \subset X$ *un ensemble quelconque. On dit que* $B \in \mathcal{A}$ *est une* μ*-enveloppe de* A *si (i)* $A \subset B$*, (ii) pour tout* $B' \in \mathcal{A}$ *contenant* A*, l'ensemble* $B \setminus B'$ *est* μ*-négligeable.*

Bien sûr la notion d'enveloppe dépend de la tribu considérée. L'intuition est la suivante : l'enveloppe d'un ensemble non mesurable A est une partie mesurable B qui épouse A de si près qu'aucune autre partie mesurable ne peut mieux faire. Une enveloppe n'est en général pas unique, mais deux enveloppes coïncident à un ensemble négligeable près. On a déjà vu (Remarque II-95 (i)) que si μ est une mesure borélienne, alors toute partie μ-mesurable s'écrit sous la forme $A = B \setminus N$, où B est borélien et N est μ-négligeable ; alors B est une μ-enveloppe de A. Mais un résultat bien plus général est vrai, du moins sous une hypothèse de σ-finitude :

Proposition V-23 (existence d'enveloppe). *Sur un espace polonais* (X, d)*, soit* μ *une mesure borélienne* σ*-finie, alors toute partie* A *de* X *admet une* μ*-enveloppe borélienne.*

Voici maintenant l'autre ingrédient majeur en vue de la mesurabilité.

Lemme V-24 (Lemme de Szpilrajn/Marcewski). *Soit* $\mathcal{T}$ *une tribu d'un espace polonais* (X, d) *vérifiant la propriété suivante :*

(Sz) Pour toute partie $A \subset X$ il existe $B \in \mathcal{T}$ tel que $A \subset B$ et tel que pour toute partie $B' \in \mathcal{T}$ contenant A, tout sous-ensemble de $B \setminus B'$ appartient à $\mathcal{T}$.

Alors l'opération $(\mathcal{A})$ *de Souslin préserve* $\mathcal{T}$*. Plus explicitement, si toutes les parties intervenant dans le schéma de Souslin appartiennent à* $\mathcal{T}$*, alors la partie ainsi définie appartient aussi à* $\mathcal{T}$*.*

Pour rappel, Edward Szpilrajn et Edward Marcewski sont une seule et même personne. Le Théorème V-18 découlera immédiatement de la Proposition V-23 et du Lemme V-24 :

Preuve du Théorème V-18. Soit μ une mesure σ-finie sur X. Par la Propriété V-23, toute partie $A \subset X$ admet une μ-enveloppe borélienne, donc un $B \in \mathcal{B}$ tel que pour tout $B' \in \mathcal{B}$, $B' \supset A$, on a $\mu[B \setminus B'] = 0$. Soit maintenant $\mathcal{T}$ la tribu des ensembles μ-mesurables, c'est à dire de la forme $B \cup N$ où B est borélien et N négligeable. Pour tout $T \in \mathcal{T}$ contenant A, on peut inclure T dans un borélien B' tel que $\mu[B' \setminus T] = 0$, et alors $\mu[B \setminus T] \leq \mu[B \setminus B'] + \mu[B' \setminus T] = 0$; donc tous les sous-ensembles de $B \setminus T$ sont négligeables et en particulier appartiennent à $\mathcal{T}$. On conclut que $\mathcal{T}$ vérifie la propriété **(Sz)** ; l'application du schéma de Souslin à la tribu $\mathcal{B} \subset \mathcal{T}$ des boréliens induit donc un élément de $\mathcal{T}$. $\square$

Voici maintenant les preuves des deux énoncés clé intervenant dans l'argument ci-dessus. (On pourra apprécier en particulier l'élégance de la preuve du Lemme V-24, petit bijou de théorie des ensembles où tous les arguments s'enchaînent harmonieusement.)

Preuve de la Proposition V-23. Soit $A \subset X$. Soit $(X_n)_{n \in \mathbb{N}}$ une partition borélienne de X avec $\mu[X_n] < +\infty$. Pour chaque X_n on cherche une μ-enveloppe borélienne B_n de $A_n = A \cap X_n$. Une fois cela accompli, on posera $B = \bigcup B_n$, de sorte que B contiendra $\bigcup A_n = A$, et si B' est un borélien contenant A, alors $B'_n = B' \cap X_n$ contient A_n, donc $\mu[B_n \setminus B'_n] = 0$, et de même pour $\mu[B \setminus B']$ par σ-additivité.

Il suffit donc de traiter le cas où μ est finie. On pose alors

$$I = \inf \Big\{ \mu[B];\ B \in \mathcal{B},\ B \supset A \Big\} \in \mathbb{R}_+.$$

Soit $(B_\ell)_{\ell \in \mathbb{N}}$ une suite de boréliens telle que $\mu[B_\ell] \longrightarrow I$; quitte à remplacer B_ℓ par $\widetilde{B}_\ell = B_1 \cap \ldots \cap B_\ell$, on peut la supposer décroissante. On pose $B_\infty = \bigcap B_\ell$, c'est un borélien qui contient A et dont la mesure vaut I.

Si maintenant B' est un borélien contenant A, alors $B_\infty \cap B'$ contient A, donc

$$I = \mu[B_\infty] = \mu[B_\infty \cap B'] + \mu[B_\infty \cap B'] \geq I + \mu[B_\infty \setminus B'],$$

donc $\mu[B_\infty \setminus B] = 0$. $\square$

Preuve du Lemme V-24. Soit donc $(P_s)_{s \in \mathbb{N}^{<\mathbb{N}}}$ une famille d'éléments de $\mathcal{T}$, indexée par $\mathbb{N}^{<\mathbb{N}}$. On pose

$$A = \mathcal{A}(P) = \bigcup_{n \in \mathcal{N}} \bigcap_{k \in \mathbb{N}} P_{n_1 \ldots n_k}. \tag{51}$$

Le but est de prouver, utilisant la propriété **(Sz)**, que $A \in \mathcal{T}$.

1. Quitte à remplacer $P_{n_1 \ldots n_k}$ par $Q_{n_1 \ldots n_k} = \bigcap_{j \leq k} P_{n_1 \ldots n_j}$, on peut supposer que $P_{m_1 \ldots m_{k+1}} \subset P_{m_1 \ldots m_k}$ pour tous m et k, c'est à dire que le système est régulier.

2. Pour tout $m = (m_1, \ldots, m_k) \in \mathbb{N}^{<\mathbb{N}}$, on pose

$$B_{m_1 \ldots m_k} = \bigcup_{q \in \mathcal{N}} \bigcap_{\ell \in \mathbb{N}} P_{m_1 \ldots m_k q_1 \ldots q_\ell}.$$

C'est l'opération $(\mathcal{A})$ de Souslin appliquée à toutes les parties qui “raffinent” P_m dans le schéma (51). Par l'étape 1, toutes les opérations se passent dans P_m. Par ailleurs

$$\bigcap_{\ell \in \mathbb{N}} P_{m_1 \ldots m_k q_1 \ldots q_\ell} = \bigcap_{\ell \geq 2} P_{m_1 \ldots m_k q_1 \ldots q_\ell},$$

donc en renommant $(q_1, q_2, \ldots) = (r, q_1', q_2', \ldots)$ on trouve

$$B_{m_1 \ldots m_k} = \bigcup_{r \in \mathbb{N}} \bigcup_{q' \in \mathcal{N}} \bigcap_{\ell \in \mathbb{N}} P_{m_1 \ldots m_k r q_1' \ldots q_\ell'},$$

ce qui est l'union de tous les $B_{m_1 \ldots m_k r}$. Cela fonctionne aussi quand $k = 0$, pour A tout entier. Pour récapituler,

(a) $B_m \subset P_m$

(b) $\forall m \in \mathbb{N}^{<\mathbb{N}}, \quad B_m = \bigcup_{r \in \mathbb{N}} B_{mr}$

(c) $A = \bigcup_{r \in \mathbb{N}} B_r$.

Si $x \in A$, par (c) on peut trouver m_1 tel que $x \in B_{m_1}$, puis par (b) on trouve m_2 tel que $x \in B_{m_1 m_2}$, et ainsi de suite, de sorte qu'il existe $m \in \mathcal{N}$ tel que $x \in \bigcap_{k \in \mathbb{N}} B_{m_1 \ldots m_k}$. En d'autres termes, $A \subset \mathcal{A}(B)$. Réciproquement, comme $B_m \subset P_m$ on a $\mathcal{A}(B) \subset \mathcal{A}(P) = A$. Donc

(d) $A = \mathcal{A}(B)$.

Le système (B_m) est donc un système alternatif pour A, très régulier en un sens ; mais ses parties ne sont pas a priori mesurables.

3. Pour tout $m \in \mathbb{N}^{<\mathbb{N}}$, on introduit une $\mathcal{T}$-enveloppe de B_m : c'est à dire un ensemble $B_m^* \in \mathcal{T}$ tel que

(e) $\forall m \in \mathbb{N}^{<\mathbb{N}}, \quad B_m \subset B_m^*$

(f) Pour toute partie $B' \in \mathcal{T}$ contenant B_m, tous les sous-ensembles de $B_m^* \setminus B'$ appartiennent à $\mathcal{T}$.

(Noter que la défintion des B^*_m n'implique qu'une quantité dénombrable de choix puisque $\mathbb{N}^{<\mathbb{N}}$ est dénombrable.) Alors

$$\begin{aligned}\mathcal{A}(B^*) \setminus \mathcal{A}(P) &= \Big[\bigcup_{n\in\mathcal{N}} \bigcap_{k\in\mathbb{N}} B^*_{n_1\dots n_k}\Big] \setminus \Big[\bigcup_{m\in\mathcal{N}} \bigcap_{\ell\in\mathbb{N}} P_{m_1\dots m_\ell}\Big] \\ &= \bigcup_{n\in\mathcal{N}} \bigcap_{k\in\mathbb{N}} \Big[B^*_{n_1\dots n_k} \setminus \bigcup_{m\in\mathcal{N}} \bigcap_{\ell\in\mathbb{N}} P_{m_1\dots m_\ell}\Big] \\ &\subset \bigcup_{n\in\mathcal{N}} \bigcap_{k\in\mathbb{N}} (B^*_{n_1\dots n_k} \setminus P_{n_1\dots n_k}) \\ &\subset \bigcup_{m\in\mathbb{N}^{<\mathbb{N}}} (B^*_m \setminus P_m).\end{aligned}$$

Soit $C = \bigcup_{m\in\mathbb{N}^{<\mathbb{N}}} B^*_m \setminus P_m \in \mathcal{T}$ (union dénombrable d'éléments de $\mathcal{T}$). Pour tout $D \subset C$,

$$D = D \cap \big(\bigcup_{m\in\mathbb{N}^{<\mathbb{N}}} B^*_m \setminus P_m\big) = \bigcup_{m\in\mathbb{N}^{<\mathbb{N}}} D \cap (B^*_m \setminus P_m);$$

par (a) et (f) on a $D \cap (B^*_m \setminus P_m) \in \mathcal{T}$; donc $D \in \mathcal{T}$, et finalement

$$\mathcal{A}(B^*) \setminus \mathcal{A}(P) \in \mathcal{T}. \tag{52}$$

4. Reste à prouver que $\mathcal{A}(B^*) \in \mathcal{T}$. Si l'on a gagné la mesurabilité et la propriété d'enveloppe en passant des B_m aux B^*_m, on a perdu la propriété (b). Pour la regagner, on va "compléter" le schéma. On définit donc $(\widehat{B}_m)_{m\in\mathbb{N}^{<\mathbb{N}}}$ par

$$\begin{gathered}\forall \ell \in \mathbb{N} \qquad \widehat{B}_\ell = B^*_\ell \\ \forall k \in \mathbb{N}\ \forall n \in \mathbb{N}^k\ \forall \ell \geq 2 \qquad \widehat{B}_{n_1\dots n_k\ell} = B^*_{n_1\dots n_k(\ell-1)} \\ \widehat{B}_{n_1\dots n_k 1} = \widehat{B}_{n_1\dots n_k} \setminus \bigcup_{\ell\in\mathbb{N}} B^*_{n_1\dots n_k\ell}.\end{gathered}$$

Ainsi les $\widehat{B}_m$ sont $\mathcal{T}$-mesurables, et par construction

$$\widehat{B}_{n_1\dots n_k} = \bigcup_{\ell\in\mathbb{N}} \widehat{B}_{n_1\dots n_k\ell}.$$

Par le même raisonnement qu'en 1,

$$\mathcal{A}(\widehat{B}) = \bigcup_{\ell\in\mathbb{N}} \widehat{B}_\ell = \bigcup_{\ell\in\mathbb{N}} B^*_\ell. \tag{53}$$

Par ailleurs toutes les intersections impliquées dans le schéma des B^* se retrouvent dans le schéma des $\widehat{B}$ (quitte à changer d'indices pour suivre les décalages), de sorte que $\mathcal{A}(B^*) \subset \mathcal{A}(\widehat{B})$. Et en désignant par $m\ell$ la concaténation de la suite finie m avec l'entier ℓ,

$$\mathcal{A}(\widehat{B}) \setminus \mathcal{A}(B^*) \subset \bigcup_{m\in\mathbb{N}^{<\mathbb{N}}} B^*_m \setminus \big(\bigcup_{\ell\in\mathbb{N}} B^*_{m\ell}\big) \tag{54}$$

Par (b) et (e), $B_m \subset \bigcup_{\ell\in\mathbb{N}} B^*_{m\ell} \in \mathcal{T}$, on peut alors appliquer (f) pour conclure que chacun des ensembles apparaissant au membre de droite de (54) a tous ses sous-ensembles mesurables ; le membre de gauche est donc lui aussi mesurable :

$$\mathcal{A}(\widehat{B}) \setminus \mathcal{A}(B^*) \in \mathcal{T}. \tag{55}$$

5. On récapitule :

$$A = \mathcal{A}(P) = \mathcal{A}(B) \subset \mathcal{A}(B^*) \subset \mathcal{A}(\widehat{B}) = \bigcup B^*_\ell$$

et

$$A = \left(\bigcup_{k\in\mathbb{N}} B^*_k\right) \setminus \left[\mathcal{A}(\widehat{B}) \setminus \mathcal{A}(B^*)\right] \setminus \left[\mathcal{A}(B^*) \setminus \mathcal{A}(B)\right]$$

appartient à la tribu $\mathcal{T}$ au vu de (52) et (55), ce qui conclut la preuve. □

V-3.2. Quelques conséquences. Concluons cette section avec deux résultats intéressants qui découlent comme fruit mûr de la théorie. Bien que ces énoncés ne fassent intervenir que des boréliens, leur démonstration passe par les ensembles analytiques. Le premier avait été énoncé sans preuve en tant que Théorème III-24.

Théorème V-25 (Théorème de l'inverse mesurable, revu). *Soient X et Y deux espaces polonais munis de leurs tribus boréliennes respectives. Alors*

(i) Si E est une partie borélienne de X et f une application injective mesurable de E dans Y, alors $f(E)$ est un borélien de Y et f^{-1} est mesurable de $f(E)$ dans E ;

(ii) En particulier, si $f : X \to Y$ est bijective mesurable, alors son inverse f^{-1} est mesurable de Y dans X.

Théorème V-26 (mesurabilité des images mesurables). *Soient X et Y deux espaces polonais munis de leurs tribus boréliennes respectives, A une partie borélienne de X et $f : X \to Y$ une application mesurable. Alors $f(A)$ est universellement mesurable.*

Remarque V-27. Attention, ce résultat dit que l'image d'un ensemble borélien est mesurable (au sens de : μ-mesurable, pour toute mesure σ-finie μ), mais pas que l'image d'un ensemble mesurable est mesurable !

Preuve du Théorème V-25. Toute partie borélienne A de E est analytique, et son image $f(A)$ par f est analytique dans Y (Théorème V-10(d) et (b)). Idem pour $f(E)\setminus f(A) = f(E\setminus A)$. L'ensemble $f(A)$ est à la fois analytique et coanalytique dans Y, il est donc borélien. Cela vaut en particulier pour $f(E)$. Mais alors $(f^{-1})^{-1}(A) = f(A)$ est borélien, pour tout borélien A ; l'application f^{-1} est donc bien mesurable. (Ou de façon équivalente : Comme f est borélienne, son graphe est borélien dans $X \times Y$ (Corollaire III-13) ; le graphe obtenu en échangeant Y et X est donc toujours borélien ; mais c'est aussi le graphe de f^{-1}.) □

Preuve du Théorème V-26. Il suffit d'enchaîner les Théorèmes V-10(d) (les boréliens sont analytiques), V-10(b) (les analytiques sont stables par les applications mesurables) et V-18 (les analytiques sont universellement mesurables). □

V-4*Classification borélienne des espaces polonais

Nous voici en mesure de démontrer un résultat qui avait été évoqué sans preuve dans les sections III-1 et IV-3 : la classification des espaces polonais, modulo bijection bimesurable. Ici plus besoin d'hypothèse topologique, seul compte le cardinal. Le gros du travail est effectué par le théorème de bijection mesurable ci-suit :

Théorème V-28 (Théorème de Cantor–Bernstein borélien). *Soient X et Y deux espaces polonais, ou sousliniens. On suppose qu'il existe une injection borélienne*

$f : X \to Y$ et une injection borélienne $g : Y \to X$. Alors il existe une bijection borélienne $F : X \to Y$, d'inverse borélienne.

Remarque V-29. Ce théorème n'est autre que la version borélienne du théorème de Cantor–Bernstein, selon lequel si X et Y sont deux ensembles, chacun s'injectant dans l'autre, alors ils sont en bijection. Ce pilier de la logique, énoncé sans preuve par Georg Cantor, a été démontré par Felix Bernstein à la toute fin du 19e siècle. C'est Borel qui a le premier publié la preuve de Bernstein, ce qui montre bien le lien étroit entre ce théorème et la naissance de la théorie de la mesure. Le nom d'Ernst Schröder est aussi souvent associé, même si sa preuve était erronée ; d'autres mathématiciens impliqués ont été Alwin Korselt, Richard Dedekind, Ernst Zermelo et Julius König, dont l'argument s'est imposé comme le plus populaire.

Preuve du Théorème V-28. Quitte à remplacer X par $X \times \{0\}$ et Y par $Y \times \{1\}$ on peut supposer X et Y disjoints.

Soit $f : X \to Y$ une injection borélienne ; en particulier $f(X)$ est borélien et f^{-1} est une bijection borélienne de $f(X)$ dans X (Théorème V-25). De même, soit $g : Y \to X$ une injection borélienne, alors $g(Y)$ est borélien et g^{-1} est une bijection borélienne de $g(Y)$ dans Y.

À tout $x \in X$ on va associer une suite finie ou infinie $(x_0, x_1, \ldots)$, prenant ses valeurs alternativement dans X et dans Y, ainsi. On pose $x_0 = x$. Si $x_0 \in g(Y)$, on pose $x_1 = g^{-1}(x_0)$, et sinon on ne définit pas x_1. Si x_1 est défini et appartient à $f(X)$, on pose $x_2 = f^{-1}(x_1)$, et sinon on ne définit pas x_2. On continue ainsi indéfiniment.

À chaque étape, chaque composante de la suite $S(x)$ ainsi définie est une fonction mesurable de x. La longueur $|S(x)|$ de cette suite est mesurable aussi : par exemple $\{x;\ |S(x)| \geq 2\} = g(Y)$ qui est bien borélien, $\{x;\ |S(x)| \geq 3\} = g(Y) \cap g(f(X))$ qui est borélien car g est injective sur Y et donc a fortiori sur $f(X)$, etc. Et par différence, les ensembles $|S| = n$ sont tous mesurables, pour tout n fini ou infini.

On partitionne alors X, et on définit F, ainsi :

• si $|S(x)|$ est fini et pair (c'est à dire que le dernier élément x_m appartient à Y) on pose $F(x) = f(x)$;

• si $|S(x)|$ est fini et impair (c'est à dire que le dernier élément x_m appartient à X) on pose $F(x) = g^{-1}(x)$;

• si $|S(x)| = \infty$ (c'est à dire que la suite ne s'arrête jamais), on pose $F(x) = f(x)$.

Il est clair que cette fonction est mesurable. On vérifie explicitement (exercice) que si l'on réalise la même construction symétrique, avec des notations similaires, sur Y, alors la réciproque de F est

• là où $|S(y)|$ est fini et impair, $f^{-1}(y)$;

• là où $|S(y)|$ est fini et pair, $g(y)$;

• là où $|S(y)| = \infty$, $f^{-1}(y)$. □

Remarque V-30 (Présentation alternative). Voici une autre façon [Kechris] de présenter la preuve du Théorème V-28. On définit deux suites d'ensembles boréliens X_n, Y_n comme suit : $X_0 = X$, $Y_0 = Y$, $X_{n+1} = g \circ f(X_n)$, $Y_{n+1} = f \circ g(Y_n)$, et $X_\infty = \bigcap_{n \geq 0} X_n$, $Y_\infty = \bigcap_{n \geq 0} Y_n$. Alors $f(X_\infty) = Y_\infty$, $f(X_n \setminus g(Y_n)) = f(X_n) \setminus Y_{n+1}$, $g(Y_n \setminus f(X_n)) = g(Y_n) \setminus X_{n+1}$. On pose alors $A = X_\infty \cup \bigcup_{n \geq 0}(X_n \setminus g(Y_n))$, $B = \bigcup_{n \geq 0}(Y_n \setminus f(X_n))$. Tous ces ensembles sont boréliens et f induit une bijection de A sur $Y \setminus B$ tandis que g induit une bijection de B sur $X \setminus A$. On définit alors F comme étant égale à f sur A et g^{-1} sur $X \setminus A$.

Maintenant voici le résultat annoncé :

THÉORÈME V-31 (Classification borélienne selon le cardinal). *Soit A un ensemble borélien d'un espace polonais X. Alors*

- *si A a cardinal $k \in \mathbb{N}$, il existe une bijection bimesurable de $\{1, \ldots, k\}$ sur A ;*
- *si A est infini dénombrable, il existe une bijection bimesurable de $\mathbb{N}$ sur A ;*
- *si A est infini non dénombrable, il existe une bijection bimesurable de $\mathcal{C}$ sur A.*

REMARQUE V-32. En particulier, dans la classe des ensembles boréliens des espaces polonais, il n'existe pas de cardinal intermédiaire : si un ensemble de cette classe n'est pas dénombrable, alors il a même cardinalité que $\mathbb{R}$, la puissance du continu. La preuve montrera en fait qu'il contient une réplique de (= un ensemble homéomorphe à) $\mathcal{C}$.

COROLLAIRE V-33. *Soient X et Y deux espaces polonais, et $A \subset X$, $B \subset Y$ des boréliens. Alors A et B sont en bijection mesurable si et seulement si ils ont même cardinal. En particulier, deux espaces polonais sont isomorphes en tant qu'espaces mesurables, si et seulement si ils ont même cardinal.*

PREUVE DU THÉORÈME V-31. Si A est dénombrable dans un espace polonais, toutes ses sous-parties sont dénombrables ; ainsi toute bijection entre espaces dénombrables, continue ou non, sera automatiquement mesurable. Seul le cas non dénombrable nécessite donc du travail.

Soit donc A un borélien non dénombrable dans X. D'après le Théorème V-28 il suffit de trouver une injection mesurable de A dans $\mathcal{C}$, et vice-versa.

1. Pour construire une injection de A dans $\mathcal{C}$, l'observation clé est la suivante :

Si A est un borélien non dénombrable, alors on peut le partager en deux boréliens disjoints non dénombrables.

Pour voir cela, on recouvre alors A par des boules $(B_\ell)_{\ell \in \mathbb{N}}$ de rayon $1/2$, et on en déduit via le procédé habituel une partition de A par des ensembles boréliens $(B'_\ell)_{\ell \in \mathbb{N}}$ de diamètre au plus 1. Si deux de ces ensembles sont non dénombrables, disons B'_m et B'_n, le but recherché est atteint, puisque A est l'union disjointe des deux boréliens non dénombrables B'_m et $A \setminus B'_m \supset B'_n$. Sinon, cela veut dire qu'un seul de ces ensembles, disons B'_{m_1} est non dénombrable, tous les autres le sont. Alors on subdivise à nouveau B'_{m_1} en le recouvrant par des boules fermées $(B_{m_1 \ell})_{\ell \in \mathbb{N}}$ de rayon $1/4$, et on recommence le raisonnement : si après transformation en partition on a trouvé deux boréliens non dénombrables disjoints, on a gagné, et sinon on subdivise encore. Si le processus n'aboutit jamais, c'est que l'on a une suite de boules fermées emboîtées non vides, B_{m_1}, $B_{m_1 m_2}$, etc. dont le diamètre tend vers 0, et tels que tous leurs complémentaires dans A sont dénombrables. Mais l'intersection de ces boules est réduite à un point ; on conclut que A est dénombrable, ce qui est une contradiction. Donc le processus aboutit toujours.

À partir de là, il est facile de construire une injection mesurable de A dans $\mathcal{C}$. Dans un premier temps, on raffine inductivement A en (A_i), $i \in \{0, 1\}$, puis en (A_{ij}), $(i, j) \in \{0, 1\}^2$, et ainsi de suite, construisant un système lusinien de boréliens $(A_s)_{s \in 2^{<\mathbb{N}}}$. Ainsi tout $x \in A$ appartient de façon unique à une intersection de parties $A_{i_1 \ldots i_k}$, pour tout k, et on peut lui associer la suite $(i_1, i_2, \ldots)$.

2. Pour construire une injection de $\mathcal{C}$ dans A, on commence par invoquer le Théorème V-4 (b), selon lequel A est image d'un espace polonais Z par une bijection

continue (a fortiori borélienne) ; il suffit donc de construire une injection de $\mathcal{C}$ dans Z, qui est forcément non dénombrable. L'observation clé est alors la suivante :

Si Z est un espace polonais non dénombrable, on peut trouver dans Z deux fermés disjoints non dénombrables de diamètre arbitrairement petit

Pour prouver cette assertion, on se donne $\varepsilon > 0$ arbitraire, et on écrit Z comme union dénombrable de boules fermées de diamètre au plus ε ; puis on transforme ce recouvrement en partition, dont chaque élément $B'_i = B_i \setminus (B_1 \cup B_{i-1})$ est la différence d'un fermé et d'un autre fermé, donc s'écrit comme une union dénombrable de fermés $(F_i^\ell)_{\ell\in\mathbb{N}}$. Si deux des ensembles B'_i, disons B'_{i_1} et B'_{i_2} sont non dénombrables, alors on peut trouver ℓ_1 et ℓ_2 tels que $F_{i_1}^{\ell_1}$ et $F_{i_2}^{\ell_2}$ sont non dénombrables (sinon B'_{i_1} ou B'_{i_2} serait union dénombrable d'ensembles dénombrables, donc dénombrable) et la conclusion est assurée. Et si on ne peut trouver deux tels ensembles, c'est qu'il existe $i_1 \in \mathbb{N}$ tel que $X \setminus F_{i_1}$ est dénombrable, a fortiori $X \setminus B_{i_1}$. On recommence alors à décomposer B_{i_1} en boules fermées $(B_{i_1\ell})_{\ell\in\mathbb{N}}$ de diamètre au plus $\varepsilon/2$, à en faire une partition, et ainsi de suite : si ce procédé itératif n'aboutit pas, c'est qu'on a une suite de fermés emboîtés $B_{i_1\ldots i_k}$, tous non vides (car non dénombrables), dont le diamètre tend vers 0, et dont le complémentaire est dénombrable ; mais alors leur intersection est réduite à un point et c'est X tout entier qui est dénombrable, une contradiction.

Une fois l'observation établie, on trouve dans X, par induction, deux fermés disjoints non dénombrables F_0 et F_1, de diamètre au plus 1 ; puis on trouve dans chacun d'entre eux deux fermés disjoints non dénombrables F_{i0} et F_{i1}, de diamètre au plus 1, et ainsi de suite. Il ne reste plus qu'à associer à $(i_1, i_2, \ldots) \in \mathcal{C}$ l'unique intersection des fermés emboîtés $F_{i_1 i_2 \ldots i_k}$. □

Preuve du Corollaire V-33. Il suffit d'appliquer deux fois le Théorème V-31 en prenant l'espace de référence, savoir $\{1, \ldots, k\}$ ou $\mathbb{N}$ ou $\mathcal{C}$ comme intermédiaire entre A et B. □

V-5*Sélection mesurable

On peut vivre sans connaître les ensembles analytiques (voire !), mais on ne peut pas faire de théorie de la mesure un tant soit peu avancée sans rencontrer le problème de la sélection mesurable. Voici une situation classique : il se présente une famille $(C_x)_{x\in X}$ de parties indexées par x et on doit choisir un y dans chacun des C_x, de façon mesurable. Cela revient à rechercher un graphe de fonction mesurable dans la réunion des $\{x\} \times C_x$. La situation est similaire à celle de l'axiome du choix, mais ce dernier ne dit de toute façon rien sur la mesurabilité de la fonction de choix. En fait on peut, en supposant seulement l'axiome du choix dépendant, couvrir toutes les situations d'intérêt que l'on rencontre classiquement.

Bien sûr, il faut des hypothèses sur la façon donc C_x varie avec x ; le plus souvent, c'est un ensemble de couples admissibles $\{(x, y)\}$ admissibles qui est fourni. On supposera donc que $C = \bigcup_{x\in X} \{x\} \times C_x$ est borélien dans $X \times Y$, et les C_x seront les **sections** ou **coupes** de C. Trouver un graphe de fonction mesurable dans une partie mesurable d'un espace produit, avec pleine projection sur la base : cette opération s'appelle **sélection mesurable** dans C (ou section mesurable de C), ou encore **uniformisation** de C. Ici "uniforme" désigne la propriété de fonction : l'objet recherché est une fonction $x \longmapsto f(x)$, fournissant une seule valeur pour chaque

x de la projection de C (au contraire de la “fonction multivaluée” ou “fonction multiforme” $x \longmapsto C_x$).

Un résultat marquant de la théorie descriptive des ensembles est que *dès que* X *et* Y *sont des espaces polonais non dénombrables, il existe un borélien* C *de* $X \times Y$*, qui n'admet aucune uniformisation borélienne.* La sélection borélienne doit donc être justifée via des hypothèses sur C. Il existe deux grandes familles de théorèmes en la matière : à sections grandes, à sections petites. L'intuition dans la première catégorie est que si les sections sont assez grandes, il est d'autant plus facile d'y trouver un graphe ; et dans la seconde, que si les sections sont petites la mesurabilité est facile à assurer.

V-5.1. Sélections élémentaires. Voici le plus élémentaire des théorèmes de sélection à grandes sections ; il semble absolument trivial, mais il est plus difficile qu'il n'en a l'air :

Proposition V-34 (cylindres mesurables). *Soient* X *et* Y *deux espaces mesurables, et* $C \subset X \times Y$ *mesurable tel que* $C_x = Y$ *pour tout* $x \in X$*. Alors* $C = A \times Y$*, où* A *est mesurable, et en particulier toute fonction borélienne sur* A *(par exemple constante) définit une sélection mesurable dans* C*.*

Preuve de la Proposition V-34. L'ensemble A est analytique comme projection du borélien C ; et de même pour $X \setminus A$ comme projection de $(X \times Y) \setminus C$. Par le Théorème V-17 de Souslin, A est borélien. □

Quant au plus élémentaire des théorèmes de sélection à petite section, il a l'air tout aussi trivial que le précédent et il est encore plus délicat.

Proposition V-35 (un graphe borélien est le graphe d'une borélienne). *Soit* $C \subset X \times Y$ *un ensemble mesurable dont chaque section non vide est un point :* $C_x = \{y\}$*. Alors* C *est le graphe d'úne fonction borélienne définie sur une partie borélienne de* X*.*

Preuve de la Proposition V-35. L'application $F : C \to X$ définie par $F(x, y) = x$ est borélienne (toujours) et injective (par l'hypothèse sur les sections). Par le Théorème V-25, son image est un borélien B. À tout $x \in B$ correspond un unique y tel que $(x, y) \in C$, appelons-le $g(x)$. Alors l'application $(x, g(x))$, de B dans C, est l'inverse de F : par le Théorème V-25 à nouveau, $(x, g(x))$ est borélienne, donc sa seconde composante $g(x)$ est une fonction borélienne de x. □

En conséquence de ces énoncés on peut adopter la terminologie suivante sans ambiguïté :

Définition V-36 (cylindre et graphe boréliens). *Soient* X *et* Y *deux espaces polonais. Alors*

(i) On dit que $C \subset X \times Y$ *est un cylindre borélien si* C *est de la forme* $A \times Y$ *avec* A *borélien de* X*, ou de façon équivalente avec* $A \times Y$ *borélien de* $X \times Y$*.*

(ii) On dit que $C \subset X \times Y$ *est un graphe borélien si c'est le graphe d'une fonction borélienne, ou de façon équivalente si c'est un ensemble borélien dont les coupes sont soit l'ensemble vide, soit un singleton.*

Exercice V-37 (Multigraphes). Sur X un espace polonais, soient f et g deux fonctions boréliennes à valeurs réelles, définies sur des domaines boréliens. Soit C l'union du graphe de f et du graphe de g. Recouvrir C par des cylindres $B_\ell \times F_\ell$

($\ell \in \mathbb{N}$) où B_ℓ est borélien et F_ℓ fermé, tels que $C \cap (B_\ell \times F_\ell)$ est un graphe. On pourra commencer par noter que les ensembles $\{f = g\}$ et $\{f \neq g\}$ sont des boréliens disjoints. Généraliser à un nombre fini arbitraire de fonctions et indiquer pourquoi cela ne s'applique pas à un ensemble dénombrable de fonctions.

V-5.2. Théorèmes classiques de sélection. Dans l'exemple de la Proposition V-35 chaque section est réduite à un point. Pour élargir en direction d'ensembles "un peu moins petits", deux choix naturels se présentent : des sections compactes (petitesse au sens topologique) ; ou des sections (au plus) dénombrables (petitesse au sens du cardinal). Les deux théorèmes suivants traitent ces situations, respectivement ; le premier est particulièrement utile.

Théorème V-38 (sélection mesurable de Novikov dans des sections compactes). *Soient X et Y des espaces polonais, et $C \subset X \times Y$ un ensemble mesurable. On suppose que pour tout $x \in X$, la section C_x est compacte. Alors la projection B de C sur X est borélienne, et il existe une application mesurable $f : B \to Y$ telle que $f(x) \in C_x$ pour tout $x \in B$.*

Théorème V-39 (sélection mesurable de Lusin–Novikov dans des sections dénombrables). *Soient X et Y des espaces polonais, et $C \subset X \times Y$ un ensemble mesurable. On suppose que pour tout $x \in X$, la section C_x est dénombrable. Alors il existe des boréliens $B_n \subset X$ et des fonctions mesurables $f_n : B_n \to Y$ telles que $C = \bigcup_{k\in\mathbb{N}} \{(x, f_n(x));\ x \in B_n\}$. En d'autres termes, C est une union dénombrable de graphes boréliens. En particulier, la projection B de C sur X est mesurable et il existe une fonction borélienne $f : B \to Y$ telle que $f(x) \in C_x$ pour tout $x \in B$.*

Un puissant et difficile théorème généralise à la fois les deux derniers énoncés : il a d'abord été découvert par les efforts partiellement indépendants de V.Ya. Arsenin, E.A. Čegolkov (russes), Mitrofan Cioban (moldave) et Kinjirô Kunugui (japonais), avant la version aboutie de Jean Saint-Raymond (français) :

Théorème V-40 (sélection mesurable dans des sections K_σ). *Soient X et Y des espaces polonais, et $C \subset X \times Y$ un ensemble borélien dont les coupes C_x sont des unions dénombrables de compacts, pour tout $x \in X$. Alors C est une union dénombrable d'ensembles boréliens à coupes compactes ; en particulier, la projection B de C sur X est un borélien et il existe une fonction borélienne $f : B \to Y$ telle que $f(x) \in C_x$ pour tout $x \in B$.*

D'autre part, dans la catégorie des théorèmes de sélection à grandes sections, voici un énoncé simple à sections "moins grandes" que l'énoncé maximal de la Proposition V-34 :

Théorème V-41 (Théorème de sélection de Kunugui–Novikov à coupes ouvertes). *Soient X et Y des espaces polonais, et $C \subset X \times Y$ un ensemble mesurable. On suppose que pour tout $x \in X$, la section C_x est ouverte, et on se donne $(V_n)_{n\in\mathbb{N}}$ une base dénombrable d'ouverts dans Y. Alors il existe des boréliens $B_n \subset X$ tels que $C = \bigcup_{k\in\mathbb{N}} B_n \times V_n$. En particulier, la projection B de C sur X est un borélien, et on peut trouver une fonction mesurable $f : B \to Y$, ne prenant qu'une quantité dénombrable de valeurs, telle que $f(x) \in C_x$ pour tout $x \in X$.*

Dans la suite de cette section je vais fournir des preuves des Théorèmes V-39, V-38 et V-41 ; je laisserai de côté la difficile preuve du Théorème V-40, renvoyant à [Kechris] pour cela et bien davantage.

V-5.3. Sélection dans les coupes dénombrables. Le Théorème V-39 est un résultat célèbre dans la théorie analytique des ensembles, avec de nombreux développements [Kechris, Melleray]. On commence par se ramener au cas d'un ensemble fermé, grâce au Lemme qui suit.

Lemme V-42. *Si $C \subset X \times Y$ est un borélien à coupes dénombrables, il existe un espace polonais Z, un ensemble fermé $\widetilde{C} \subset X \times Z$, à coupes dénombrables, une application continue $\psi : Z \to Y$ telle que $(Id, \psi)(\widetilde{C}) = C$; en particulier, les projections de C et de $\widetilde{C}$ sur X coïncident. En outre, il existe $\theta : Z \to X$ continue telle que $\widetilde{C} = \{(\theta(z), z);\ z \in Z\}$; en particulier, pour tout fermé F de Z, $\mathrm{proj}_Z(\widetilde{C} \cap (X \times F)) = F$ et plus généralement la projection sur Z de tout fermé de $\widetilde{C}$ est un fermé.*

Remarque V-43. On dit que $\widetilde{C}$ est l'**antigraphe** de la fonction continue θ.

Preuve du Lemme V-42. Par le Théorème V-4 (b), il existe un espace polonais Z (sous-espace fermé de $\mathcal{N}$) et une bijection continue $\varphi : Z \to C$. On définit $\widetilde{C} \subset X \times Z$ par

$$\widetilde{C} = \Big\{ \big(x, \varphi^{-1}(x,y)\big);\ (x,y) \in C \Big\}.$$

Il n'est pas évident a priori que $\widetilde{C}$ est borélien (l'application φ^{-1} est mesurable mais cela ne suffit pas) ; mais nous allons voir dans un instant qu'il est fermé. Pour commencer, il est évident que $(x, z) \longmapsto \varphi(z)$ est continue de $\widetilde{C}$ dans C ; c'est une surjection par construction ; et c'est aussi une injection car $\varphi(z) = \varphi(z')$ avec $z = \varphi^{-1}(x, y)$ et $z' = \varphi^{-1}(x', y')$ implique $x = x'$ et $y = y'$. Donc φ est une bijection continue de $\widetilde{C}$ dans C. Et si $(x, z) \in \widetilde{C}$, soit y l'unique élément de Z tel que $z = \varphi^{-1}(x, y)$; alors $\varphi(z) = (x, y)$, donc y est la seconde composante de $\varphi(z)$, que l'on note $\psi(z)$, de sorte que $(x, \psi(z)) = (x, y)$; et réciproquement, si $(x, y) \in C$ alors $(x, \psi(\varphi^{-1}(x, y)) = (x, y)$, de sorte que l'application continue (Id, ψ) est une surjection de $\widetilde{C}$ dans C.

De façon tautologique, la projection de $\widetilde{C}$ sur X coïncide avec celle de C sur X.

Pour tout $x \in X$, la coupe $\widetilde{C}_x$ est $\{\varphi^{-1}(x, y);\ (x, y) \in C\}$, en bijection donc avec C_x, et donc dénombrable.

Appelons maintenant θ la première composante de φ, de sorte que $\varphi(z) = (\theta(z), \psi(z))$. Par construction on a bien $\widetilde{C} = \{(\theta(z), z),\ z \in Z\}$. La fin de l'énoncé en découle facilement. $\square$

Passons maintenant au cœur de la démonstration du Théorème de Lusin–Novikov. La preuve ci-dessous, évitant le recours à des outils sophistiqués, est due à Forte Shinko, jeune spécialiste japonais de théorie descriptive des ensembles.

Preuve du Théorème V-39. 1. Le Lemme V-42 fournit un espace polonais Z et un fermé $\widetilde{C} \subset X \times Z$ à coupes dénombrables. Supposons démontré l'énoncé du Théorème V-39 pour $\widetilde{C}$. Alors la projection B de $\widetilde{C}$ est borélienne ; mais c'est aussi la projection de C. En outre il existe des fonction boréliennes $f_n : D_n \to Z$, D_n borélien de X, telles que $\widetilde{C}$ est l'union des graphes des f_n ; alors, avec la fonction ψ fournie par le Lemme V-42, C est l'union des graphes des fonctions $\psi \circ f_n$, chacune étant borélienne comme composition d'une foction borélienne et d'une fonction continue.

Il suffit donc de prouver le Théorème quand l'ensemble C dans $X \times Y$ vérifie les hypothèses que l'on a conférées à $\widetilde{C}$ dans $X \times Z$, c'est à dire quand C est l'antigraphe d'une fonction continue, avec les bonnes conséquences que cela entraîne, en particulier que C est fermé et se projette sur Y tout entier.

2. Soit A un fermé de X et soit $\mathcal{F} = \{F_1, \ldots, F_m\}$ une famille finie de fermés deux à deux disjoints de Y. Disons que $(A, \mathcal{F})$ vérifie la propriété **(P)** si *il n'existe aucune famille finie $\{B_1, \ldots, B_m\}$ de boréliens de A, dont l'union recouvre A tout entier, telle que chaque $C \cap (B_j \times F_j)$ $(1 \leq j \leq m)$ est une union dénombrable de graphes de fonctions boréliennes* (donc des graphes de fonctions $f_j^r : D_j^r \to F_j$ $(r \in \mathbb{N})$, où chaque D_j^r est un borélien de B_j).

Quelques remarques sur cette propriété. Le choix $B_i = A$, $B_j = \emptyset$ pour $j \neq i$, montre que pour tout i dans $\{1, \ldots, m\}$, $C \cap (A \times F_i)$ n'est pas union dénombrable de graphes. Mais la liberté laissée au choix du recouvrement borélien rend la condition a priori bien plus forte que cette condition énoncée sur les F_i séparément. On note au passage que si **(P)** est vraie, tous les ensembles $C \cap (A \times F_i)$ doivent bien sûr être non vides.

Le but est de montrer que C est union dénombrable de graphes boréliens, c'est à dire que $(X, \{Y\})$ *ne vérifie pas* **(P)**. Supposant par l'absurde que $(X, \{Y\})$ vérifie **(P)**, on va aboutir à une contradiction. Pour cela on va construire par induction une suite de fermés $(A_k)_{k \in \mathbb{N}}$ de Z, et une suite $(\mathcal{F}_k)_{k \in \mathbb{N}}$ de familles finies de fermés disjoints dans Y, vérifient, pour tout $k \in \mathbb{N}$,

$$\begin{cases} A_{k+1} \subset A_k \text{ et } \operatorname{diam}(A_k) \leq 1/k; \\ \mathcal{F}_k \text{ est constituée de } 2^k \text{ parties fermées } F_s \text{ indexées par } \{0,1\}^k \\ \qquad \text{et pour tout } s = (s_1, \ldots, s_k), \text{ on a } F_{s_1 \ldots s_{k+1}} \subset F_{s_1 \ldots s_k}, \\ \qquad \text{et } \operatorname{diam}(F_{s_1 \ldots s_k}) \leq 2^{-k}; \\ (A_k, \mathcal{F}_k) \text{ vérifie } \mathbf{(P)}. \end{cases}$$

Ces propriétés s'étendent à $k = 0$ si l'on pose $A_0 = X$, $F_\emptyset = Y$ (quitte à remplacer la distance de Y par une distance équivalente bornée par 1).

Supposons construits A_k et $\mathcal{F}_k$ pour $k \in \mathbb{N}_0$, on va construire A_{k+1} et $\mathcal{F}_{k+1}$.

3. On commence par raffiner chaque ensemble F_s $(s \in \{0,1\}^k)$ à tour de rôle, le remplaçant par deux fermés disjoints F_{s0} et F_{s1} inclus dans F_s et de petit diamètre (les ensembles $F_{s'}$ pour $s' \leq s$ ayant déjà été remplacés par $F_{s'0}$ et $F_{s'1}$) tout en préservant la propriété **(P)**. Cela se fera (quitte à renuméroter) grâce à la propriété d'héritage que voici :

$\mathbf{(H)_I}$ *Si* $(A, \{F_1, F_2, \ldots, F_m\})$ *vérifie* **(P)** *et si* $\varepsilon > 0$, *alors il existe deux fermés disjoints* $F_{10}, F_{11} \subset F_1$, *de diamètre au plus* ε, *tels que* $(A, \{F_{10}, F_{11}, F_2, \ldots, F_m\})$ *vérifie* **(P)**.

Pour prouver $\mathbf{(H)_I}$, on commence par définir $\Delta(F_1) = \{(x,x),\ x \in F_1\}$ (diagonale du carré ensembliste $F_1 \times F_1$) et on recouvre $(F_1 \times F_1) \setminus \Delta(F_1)$ par des produits de boules fermées de diamètre au plus ε :

$$(F_1 \times F_1) \setminus \Delta(F_1) = \bigcup_{n \in \mathbb{N}} F_{10}^n \times F_{11}^n,$$

où chaque fermé F_{1i}^n est donc inclus dans F_1 et de diamètre au plus ε ; par construction F_{10}^n et F_{11}^n sont disjoints pour tout n.

Si la conclusion de $(\mathbf{H})_{\mathbf{I}}$ est en défaut, alors pour tout $n \in \mathbb{N}$ on peut trouver des boréliens de A, notés $B_{10}^n, B_{11}^n, B_2^n, \ldots, B_m^n$, recouvrant A tout entier, tels que $C \cap (B_{10}^n \times F_{10}^n)$, $C \cap (B_{11}^n \times F_{11}^n)$, $C \cap (B_j^n \times F_j^n)$, pour toutes valeurs de $j \in \{2, \ldots, m\}$ et $n \in \mathbb{N}$, sont tous des unions de graphes de fonctions boréliennes $f_{10}^{n,r} : D_{10}^{n,r} \to F_{10}^n$, $f_{11}^{n,r} : D_{11} \to F_{11}^{n,r}$, $f_j^{n,r} : D_j^{n,r} \to F_j^n$, et $D_s^{n,r} \subset B_s^n$ ($j \in \{2, \ldots, m\}$, $n \in \mathbb{N}$, $r \in \mathbb{N}$).

On pose $B_j = \bigcup_{n \in \mathbb{N}} B_j^n$; ainsi, pour tout $j \geq 2$, $C \cap (B_j \times F_j)$ est l'union des $C \cap (B_j^n \times F_j)$ et c'est donc l'union de tous les graphes des $f_j^{n,r}$, dont les domaines sont tous inclus dans B_j.

On pose ensuite $B_1 = A \setminus (B_2 \cap \ldots \cup B_m)$. Comme tous les B_j^n sont disjoints de B_1, les boréliens $\{B_{10}^n, B_{11}^n\}$ recouvrent B_1 ; quitte à remplacer B_{10}^n par $B_{10}^n \cap B_1$ et B_{11}^n par $B_{11}^n \cap (B_1 \setminus B_{10}^n)$ (et à restreindre les domaines $D_{10}^{n,r}, D_{11}^{n,r}$ en conséquence), on peut supposer que pour tout n, B_1 est l'union disjointe de B_{10}^n et B_{11}^n.

Considérons alors

$$\Gamma = \left[C \cap (B_1 \times F_1)\right] \setminus \Bigg[\Big(\bigcup_{n \in \mathbb{N}} \bigcup_{r \in \mathbb{N}} \{(x, f_{10}^{n,r}(x)),\ x \in D_{10}^{n,r}\}\Big) \bigcup \Big(\bigcup_{n \in \mathbb{N}} \bigcup_{r \in \mathbb{N}} \{(x, f_{11}^{n,r}(x)),\ x \in D_{11}^{n,r}\}\Big)\Bigg].$$

Soient $(\overline{x}, y_0)$ et $(\overline{x}, y_1)$ deux éléments de Γ partageant la première coordonnée. Si $y_0 \neq y_1$, soit $n \in \mathbb{N}$ tel que $(y_0, y_1) \in F_{10}^n \times F_{11}^n$. Comme $\overline{x} \in B_1$, on a soit $\overline{x} \in B_{10}^n$ soit $\overline{x} \in B_{11}^n$; par exemple $\overline{x} \in B_{10}^n$. Mais alors $(\overline{x}, y_0) \in C \cap [B_{10}^n \times F_{10}^n]$, donc il est de la forme $(\overline{x}, f_{10}^{n,r}(\overline{x}))$ pour un certain $r \in \mathbb{N}$; ce qui est exclu par définition de Γ. On conclut que $y_0 = y_1$; autrement dit, Γ *est un graphe.*

Mais Γ est par ailleurs borélien, comme différence d'un borélien et d'une union dénombrable de boréliens ; par le Théorème V-35 c'est le graphe d'une fonction borélienne $\gamma : D_1 \to F_1$ où D_1 est un borélien de B_1.

Ainsi $C \cap [B_1 \times F_1]$ est à son tour une union dénombrable de graphes boréliens, en contradiction avec l'hypothèse de $(\mathbf{H})_{\mathbf{I}}$.

On peut donc appliquer la propriété $(\mathbf{H})_{\mathbf{I}}$ à 2^k reprises, pour passer de la famille $\mathcal{F}_k$ à la famille $\mathcal{F}_{k+1}$.

4. On va maintenant affiner A_k en A_{k+1}, tout en préservant la propriété $(\mathbf{P})$. Cela se fera par une seconde propriété d'héritage :

$(\mathbf{H})_{\mathbf{II}}$ *Si $(A, \mathcal{F})$ vérifie* $(\mathbf{P})$ *et si $\delta > 0$, alors il existe un fermé $A' \subset A$, de diamètre au plus δ, tels que $(A', \mathcal{F})$ vérifie (P).*

Pour prouver $(\mathbf{H})_{\mathbf{II}}$, on écrit $A = \bigcup_{n \in \mathbb{N}} A^n$, où chaque A^n est une boule finie de diamètre au plus δ. Écrivons aussi $\mathcal{F} = \{F_1, \ldots, F_m\}$. Si la conclusion de $(\mathbf{H})_{\mathbf{II}}$ est fausse, alors aucun des $(A^n, \mathcal{F})$ ne vérifie $(\mathbf{P})$; donc pour tout $n \in \mathbb{N}$ on peut trouver des boréliens $B_1^n, \ldots, B_m^n$ recouvrant A^n et tels que $C \cap (B_j^n \times F_j)$ est union dénombrable de graphes boréliens. On pose alors $B_j = \bigcup_{n \in \mathbb{N}} B_j^n$, ainsi chaque $C \cap (B_j \times F_j)$ est l'union des $C \cap (B_j^n \times F_j)$, et donc une union de graphes boréliens. En outre

$$\bigcup_{1 \leq j \leq m} B_j = \bigcup_{n \in \mathbb{N}} \bigcup_{1 \leq j \leq m} B_j^n = \bigcup_{n \in \mathbb{N}} A^n = A.$$

Donc $(A, \mathcal{F})$ viole (P), en contradiction avec l'hypothèse de $(\mathbf{H})_{\mathbf{II}}$.

5. Une fois la famille $(A_k, \mathcal{F}_k)$ construite, pour tout $c \in \{0,1\}^{\mathbb{N}}$ la famille $C \cap (A_k \times F_{n_1 \ldots n_k})$ est faite de fermés emboîtés, non vides (car $(A_k, \mathcal{F}_k)$ vérifie $(\mathbf{P})$, Cf.

la remarque faite en 2), et de diamètre tendant vers 0. Par théorème des fermés emboîtés, l'intersection est un singleton, de la forme $\{(\overline{x}, \chi(c))\}$, où $\{\overline{x}\}$ est la limite décroissante des A_k ; et bien sûr la fonction χ est injective car si c et c' diffèrent en position k, alors $\chi(c)$ et $\chi(c')$ appartiennent aux fermés disjoints $F_{c_1\dots c_k}$ et $F_{c'_1\dots c'_k}$ respectivement.

En conclusion la section $C_{\overline{x}}$ contient une réplique de l'ensemble $\mathcal{C}$ de Cantor, elle n'est donc pas dénombrable, ce qui contredit l'hypothèse et achève la preuve du Théorème V-39. □

Remarques V-44. (i) La preuve montre en fait que si $C \subset X \times Y$ est borélien, alors soit c'est une union dénombrable de graphes boréliens, soit l'une de ses fibres contient une réplique de l'ensemble $\mathcal{C}$ de Cantor. Bien sûr, cet énoncé contient comme cas très particulier la Remarque V-32 (quand X est réduit à un point !). Rétrospectivement, on comprend un petit peu mieux l'esprit de la preuve : s'il n'existait qu'une fibre $\overline{x} \times Y$ contenant une réplique de $\mathcal{C}$, c'est cette fibre qui ferait obstruction au recouvrement par une famille dénombrable de graphes boréliens, et toujours elle si l'on restreint les valeurs pour isoler les points de $\mathcal{C}$.

(ii) En conséquence du Théorème V-39, si $C \subset X \times Y$ est un borélien à coupes dénombrables, alors l'ensemble des points d'unicité de C,

$$U(C) = \{x \in X;\ \exists! y \in Y; (x, y) \in C\}$$

est borélien (exercice). Un résultat plus général, dû à Lusin, affirme que si C est borélien alors U est coanalytique : voir Théorème V-56 plus loin. Comme on le verra alors, on peut en déduire une autre preuve, moins élémentaire mais plus compacte, de ce qu'un borélien à coupes dénombrables admet une uniformisation borélienne.

V-5.4. Théorème de séparation de Novikov et conséquences. Le Théorème de séparation de Novikov, qui renforce le Théorème de séparation V-14 de Lusin, servira à prouver les Théorèmes V-41 et V-38.

Théorème V-45 (Théorème de séparation de Novikov). *Dans un espace polonais, soient $(A_k)_{k\in\mathbb{N}}$ une famille dénombrable de parties analytiques telles que $\bigcap A_k = \emptyset$. Alors il existe une famille de boréliens $(B_k)_{k\in\mathbb{N}}$ tels que $A_k \subset B_k$ pour tout k, et $\bigcap B_k = \emptyset$.*

Preuve du Théorème V-45. L'argument qui suit est dû à Gabriel Mokobodzki (spécialiste français de théorie des ensembles, fils de juif polonais déporté, élève de Gustave Choquet).

Pour chaque A_k on introduit un système de Souslin $(P^{(k)}_n)_{n\in\mathbb{N}^{<\mathbb{N}}}$, en imposant que pour tout $n \in \mathbb{N}^{<\mathbb{N}}$, $P^{(k)}_{n_1\dots n_r} = \cup_{\ell\in\mathbb{N}} P^{(k)}_{n_1\dots n_r \ell}$, que pour tout $k \in \mathbb{N}$ et $n \in \mathcal{N}$, $\operatorname{diam}(P^{(k)}_{n_1\dots n_r}) \to 0$ quand $r \to \infty$, et que pour toute suite $(n_\ell)_{\ell\in\mathbb{N}}$ l'intersection $\bigcap_{\ell\in\mathbb{N}} P^{(k)}_{n_1\dots n_\ell}$ soit constituée d'un point unique. On pose $P^{(k)}_\emptyset = \bigcup_{\ell\in\mathbb{N}} P^{(k)}_\ell = A_k$. Comme les A_k sont globalement disjoints, il en va de même de leurs sous-parties $P^{(k)}_{s_k}$, pour tout choix d'indices $s_k \in \mathbb{N}^{<\mathbb{N}}$.

Disons qu'une suite de multi-indices $(s_1, s_2, \dots)$, avec $s_i \in \mathbb{N}^{<\mathbb{N}}$ pour tout i, vérifie la propriété **(N)** si *les parties $P^{(1)}_{s_1}, P^{(2)}_{s_2}$, etc. peuvent s'inclure dans des boréliens ${Bs_1}^{(1)}, B^{(2)}_{s_2}, \dots$ globalement disjoints.*

Le but est donc de montrer que $(\emptyset, \emptyset, \emptyset, \ldots)$ vérifie la propriété **(N)**.

Supposons par l'absurde que tel n'est pas le cas. Si les suites $(\ell, \emptyset, \emptyset, \ldots)$, pour $\ell \in \mathbb{N}$, vérifiaient toutes la propriété **(N)**, on aurait des boréliens $B_\ell^{(1)}, B_\emptyset^{(2)}, B_\emptyset^{(3)}, \ldots$ contenant $P_\ell^{(1)}, P_\emptyset^{(2)}, P_\emptyset^{(3)}, \ldots$. globalement disjoints ; alors en posant $B^{(1)} = \bigcup B_\ell^{(1)}$ on aurait des boréliens $B_\emptyset^{(1)}, B_\emptyset^{(2)}, \ldots$ assurant la propriété **(N)** pour $(\emptyset, \emptyset, \ldots)$, contrairement à notre hypothèse. Il existe donc $\ell \in \mathbb{N}$ tel que $(\ell, \emptyset, \emptyset, \ldots)$ ne vérifie pas **(N)**.

Et on peut continuer ainsi par induction : chaque fois que $(s_1, s_2, \ldots)$ ne vérifie pas **(N)**, pour tout r on peut trouver ℓ tel que $(s_1, s_2, \ldots, s_{r-1}, s'_r, s_{r+1}, \ldots)$ ne vérifie pas **(N)**, avec $s_r = (a_1, \ldots, a_k)$ et $s'_r = (a_1, \ldots, a_k, \ell)$. On peut ainsi accroître à volonté la longueur de n'importe quel indice dans la suite mettant **(N)** en défaut.

On définit ainsi par récurrence une suite $(n_1, n_2, \ldots)$ délélements de $\mathcal{N}$ (chaque n_k est une suite d'entiers) tels que pour tout $\ell \in \mathbb{N}$, $(n_1|\ell, n_2|\ell, \ldots, n_\ell|\ell, \emptyset, \emptyset, \ldots)$ met **(N)** en défaut, où $n_k|\ell$ est la sous-suite finie constituée des ℓ premiers éléments de n_k.

Pour chaque $k \in \mathbb{N}$, l'intersection des $P_{n_k|\ell}^{(i)}$ est réduite à un élément, appelons le p_k ; il appartient à A_k. Ces parties étant globalement disjointes, il existe au moins deux indices i et j dans $\mathbb{N}$ tels que $p_i \neq p_j$. Soient B_i et B_j des boules disjointes contenant p_i et p_j respectivement : pour ℓ assez grand, on a $P_{n_i|\ell}^{(i)} \subset B_i$, $P_{n_j|\ell}^{(j)} \subset B_j$; sans perte de généralité $\ell > \max(i, j)$. Mais alors $(X, \ldots, X, B_i, X \ldots, X, B_j, X, X, \ldots)$ est une suite de boréliens globalement disjoints et séoarant $P_{n_1|\ell}^{(1)}, \ldots, P_{n_\ell|\ell}^{(\ell)}, A_{\ell+1}, A_{\ell+2}, \ldots$, donc $(n_1|\ell, n_2|\ell, \ldots, n_\ell|\ell, \emptyset, \emptyset, \ldots)$ vérifie la propriété **(N)**, en contradiction avec notre construction. $\square$

En supposant l'un des éléments borélien et disjoint de tous les autres, en passant aux complémentaires, on obtient le

Corollaire V-46 (Recouvrement de borélien par des coanalytiques). *Soit A une partie borélienne d'un espace polonais (X, d), et soit $(Z_k)_{k\in\mathbb{N}}$ une famille dénombrable d'ensembles coanalytiques dont l'union est égale à A ; alors on peut trouver une suite de boréliens $(B_k)_{k\in\mathbb{N}}$ tels que $B_k \subset Z_k$ pour tout k, dont l'union est toujours égale à A.*

Ces résultats de séparation et recouvrement sont des ingrédients efficaces pour prouver les Théorèmes V-38 et V-41, comme on va le voir maintenant.

Preuve du Théorème V-41. On note proj_X la projection sur X. On fixe $(V_n)_{n\in\mathbb{N}}$ une base dénombrable de voisinages de Y, de sorte que tout C_x est une union de certains des V_n. Pour tout n,

$$X_n = \{x \in X;\ V_n \subset C_x\} = X \setminus \mathrm{proj}_X\Big[C \cap (X \times (Y \setminus V_n))\Big]$$

est un ensemble coanalytique (car complémentaire de la projection d'un borélien). Donc $\bigcup Z_n = X_n \times V_n$ est également coanalytique. Comme l'union des Z_n recouvre C, par le Corollaire V-46 on peut trouver des boréliens $(Q_n)_{n\in\mathbb{N}}$ de $X \times Y$ recouvrant C, avec $Q_n \subset Z_n$. Si A_n est la projection de Q_n sur X, A_n est analytique et $A_n \subset X_n$. Par le Théorème V-14 de séparation de Lusin (appliqué aux ensembles analytiques disjoints A_n et $X \setminus X_n$) on peut trouver un borélien B_n tel que $A_n \subset B_n \subset X_n$. Alors $\mathrm{proj}_X Q_n \subset B_n$ et $\mathrm{proj}_Y Q_n \subset V_n$, d'où $Q_n \subset B_n \times V_n$; a fortiori $C \subset \bigcup B_n \times V_n$.

Mais par définition de X_n, $B_n \times V_n \subset X_n \times V_n \subset C$, donc l'inclusion réciproque $\bigcup B_n \times V_n \subset C$ est vraie aussi. $\square$

Passons maintenant au Théorème V-38. C'est en fait un résultat intermédiaire plus précis qui sera démontré, mais il faudra au préalable quelques notions topologiques plus avancées sur les compacts.

Si X est un espace polonais, on note $K(X)$ l'espace des compacts de X ; on le munit de la **distance de Hausdorff** :

$$d_{\mathcal{H}}(K, L) = \max\Big(\sup_{x\in K} d(x, L), \sup_{y\in L} d(y, K)\Big),$$

où $d(a, B) = \inf_{b\in B} d(a, b)$. Cette distance fait de $K(X)$ un espace métrique polonais (c'est un excellent exercice, pas si simple). Si $D = \{x_n, n \in \mathbb{N}\}$ est une partie dense, alors l'ensemble des sous-ensembles finis de D est une famille dénombrable dense dans $K(X)$. La topologie induite sur $K(X)$ par la distance de Hausdorff est la **topologie de Vietoris** (du nom du topologiste autrichien Leopold Vietoris, célèbre aussi pour sa longévité puisqu'il publiait encore à 100 ans passés...) : c'est la topologie qui est engendrée par les ouverts de la forme $\{K;\ K \subset U\}$, où U est un ouvert donné, et ceux de la forme $\{K;\ K \cap U \neq \emptyset\}$. Cette topologie est l'analogue pour les compacts de la convergence uniforme pour les fonctions.

Si X est un espace polonais, on peut le **compactifier**, c'est à dire l'injecter continûment dans un espace compact dont il sera une partie dense. Il existe de nombreuses façons de compactifier, la plus simple étant la compactification d'Alexandrov (on ajoute un point à l'infini, dont les voisinages sont les complémentaires des compacts) ; mais ce qui est important dans ce contexte est de préserver la nature polonaise, et ce ne sera pas le cas du compactifié d'Alexandrov (sauf si X est localement compact : exercice). Voici une compactification qui s'applique à tout espace polonais et préserve la nature polonaise de la topologie : on commence par remplacer la distance d par une distance topologiquement équivalente à valeurs dans $[0, 1]$, comme $d/(1 + d)$; on choisit une suite $(x_n)_{n\in\mathbb{N}}$ dense dans X ; alors l'application $F : x \longmapsto (d(x, x_n))_{n\in\mathbb{N}}$ est une injection continue de X dans l'espace métrique compact $[0, 1]^{\mathbb{N}}$; on vérifie que F^{-1} est continue de $F(X)$ (muni de la topologie induite par $[0, 1]^{\mathbb{N}}$) dans X, de sorte que X est homéomorphe à $F(X)$. L'adhérence $Y = \overline{F(X)}$ (adhérence de $F(X)$ dans $[0, 1]^{\mathbb{N}}$) est un espace métrique compact dans lequel $F(X)$ est dense. En outre, $Y \setminus F(X)$ est l'ensemble des $y \in Y$ tels que $\lim_{r\to 0} \operatorname{diam}(F^{-1}(B_r(y))) > 0$ (sinon la famille des $F^{-1}(z)$, $z \to y$, convergerait vers un certain x et on aurait $y = F(x)$) ; donc $F(X)$ coïncide avec l'ensemble des y vérifiant $\lim_{r\to 0} \operatorname{diam}(F^{-1}(B_r(y))) = 0$ (cette dernière quantité est en fait l'oscillation de la fonction F de X dans Y) ; c'est l'intersection des ouverts O_k définis par $\lim_{r\to 0} \operatorname{diam}(F^{-1}(B_r(y))) < 1/k$. Tout cela nous dit que modulo l'homéomorphisme F, X est une intersection dénombrable d'ouverts de Y, c'est à dire ce qu'on appelle un G_δ ; et en particulier c'est un borélien de Y.

Maintenant, modulo l'isomorphisme F, $K(X)$ est l'ensemble des compacts de Y qui sont inclus dans X, c'est à dire dans l'intersection des O_k. Mais pour chaque k, l'ensemble des compacts de Y inclus dans O_k est un ouvert de $K(Y)$; leur intersection $K(X)$ est donc un G_δ de $K(Y)$, et en particulier un borélien.

Avec ces notions, le Théorème V-38 est une conséquence immédiate des deux énoncés qui suivent :

Théorème V-47 (structure des boréliens à coupes compactes). *Si X et Y sont des espaces polonais et C est un borélien de $X \times Y$ à coupes C_x compactes, alors $x \longmapsto C_x$ est une application borélienne à valeurs dans $K(Y)$.*

Proposition V-48 (représentant continu d'un compact). *Si Y est un espace polonais, il existe une application continue $I : K(Y) \to Y$.*

Preuve du Théorème V-47. 1. Soit $\overline{Y}$ une compactification polonaise de Y ; ainsi Y est homéomorphe à un G_δ dense de $\overline{Y}$; par abus de langage on fera comme si $Y \subset \overline{Y}$. Fixons une base dénombrable $(V_n)_{n\in\mathbb{N}}$ de voisinages de $\overline{Y}$. Les coupes de C sont compactes dans Y, et donc dans $\overline{Y}$ aussi. Le complémentaire de C dans $X \times \overline{Y}$ est à coupes ouvertes, et donc (par le Théorème V-41) de la forme $\cup B_n \times V_n$ où les B_n sont boréliens dans X. Soit $K_n = \overline{Y} \setminus V_n$, et soit b la fonction de X dans $\mathcal{P}(\mathbb{N})$ (l'ensemble des parties de $\mathbb{N}$) définie par

$$b(x) = \Big\{n \in \mathbb{N};\ x \in B_n\Big\}.$$

Ainsi

$$\overline{Y} \setminus C_x = \bigcup_{n\in b(x)} V_n,$$

soit

$$C_x = \bigcap_{n\in b(x)} K_n,$$

ou encore $C_x = \Phi \circ b(x)$, où

$$\Phi(B) = \bigcap_{n\in B} K_n.$$

On munit $\mathcal{P}(\mathbb{N}) \simeq \{0,1\}^{\mathbb{N}}$ de la topologie produit, c'est à dire la convergence successive des termes de la suite. La fonction b est mesurable comme supremum d'une infinité dénombrable de fonctions mesurables (la k-ème est la fonction qui sur B_k vaut $(0,\ldots,0,1,0,\ldots,0)$ avec un 1 en k-ème position, et hors de B_k vaut 0). Vérifions maintenant que Φ est mesurable.

2. Soit U un ouvert de $\overline{Y}$, par abus de notation on écrira $K(U)$ pour l'ensemble des compacts de $\overline{Y}$ qui sont inclus dans U ; on sait que c'est un ouvert de $K(\overline{Y})$. Si $\Phi(B) \in K(U)$, c'est à dire si $\Phi(B) \subset U$, alors il existe N tel que pour $n \geq N$, $\bigcap_{n\in B, n\leq N} K_n \subset U$ (ici on utilise que l'intersection des compacts K_n avec le compact $\overline{Y}\setminus U$ est vide, donc il existe une sous-famille finie de ces compacts dont l'intersection est vide). Ainsi pour tout B' tel que $B \cap \{1,\ldots,N\} = B' \cap \{1,\ldots,N\}$ on aura $\Phi(B') \subset \bigcap_{n\leq N} \bigcap_{n\in B, n\leq N} K_n \subset U$. Cela montre que l'image réciproque par Φ de l'ouvert $K(U)$ est un ouvert. (Cela suggère la continuité de Φ, mais il y a une deuxième sorte d'ouverts à considérer pour couvrir la topologie de $K(\overline{Y})$.)

3. Soit à nouveau U un ouvert de $\overline{Y}$; on considére $\mathcal{K}^U$ l'ensemble des compacts K tels que $K \cap U \neq \emptyset$. C'est le complémentaire dans $K(\overline{Y})$ de l'ensemble $K(L)$ des compacts inclus dans le compact $L = \overline{Y} \setminus U$. Soit $Q_\ell = \{y;\ d(y,L) < 1/\ell\}$, où d désigne la distance dans $\overline{Y}$; alors avec des notations évidentes $K(L)$ est l'intersection des $K(Q_\ell)$, qui sont des ouverts. Donc $K(L)$ est le complémentaire d'une intersection dénombrable d'ouverts de la forme $K(Q_\ell)$. Ainsi les ouverts de la forme $K(U)$ suffisent à engendrer tous les boréliens de $K(\overline{Y})$, et l'étape 2 montre que l'image réciproque que Φ de tous ces ouverts est un borélien. Finalement Φ est bien borélienne.

4. À ce stade on a montré que $x \longmapsto C_x$ est borélienne de X dans $K(\overline{Y})$. Mais par hypothèse elle est à valeurs dans le borélien $K(Y)$; elle est donc borélienne de X dans $K(Y)$. □

Passons maintenant à la Proposition V-48. Pour se faire une intuition de la construction, on peut commencer par considérer le cas particulier représentatif $Y = \mathbb{R}$: il suffit alors de choisir $I(K) = \inf K$ (exercice).

Preuve de la Proposition V-48. On écrit Y comme une union dénombrable de boules ouvertes $B_1, B_2, \ldots$ de diamètre au plus $1/2$, non nécessairement disjointes. Puis on subdivise chaque boule B_k en ouverts $B_{k1}, B_{k2}, \ldots$ de diamètre au plus $1/4$ et de sorte que $\overline{B}_{ki} \subset B_k$; on continue de façon inductive, définissant un schéma de Souslin régulier d'ouverts, recouvrant Y tout entier à chaque étape k, avec des ouverts de diamètre au plus 2^{-k}, que l'on note $B_{i_1\ldots i_k}$, $k \in \mathbb{N}$, et tels que $\overline{B}_{i_1\ldots i_{k+1}} \subset B_{i_1\ldots i_k}$.

Si K est un compact de Y on définit une suite d'indices $i_1, i_2, \ldots$ en choisissant pour i_1 le plus petit indice i tel que $B_i \cap K \neq \emptyset$, puis pour i_2 le plus petit indice i tel que $B_{i_1 i} \cap K \neq \emptyset$, etc. La construction assure que la suite est bien définie, et

$$K \bigcap \bigcap_{\ell \in \mathbb{N}} B_{i_1\ldots i_\ell} = K \bigcap \bigcap_{\ell \in \mathbb{N}} \overline{B}_{i_1\ldots i_\ell}$$

est réduit à un point par le théorème des fermés emboîtés. C'est celui-ci que l'on note $I(K)$. On vérifie (exercice) que l'application I est alors continue de $K(Y)$ dans Y. □

Remarque V-49. On peut également donner du Théorème V-40 une version précisée sous la forme suivante, due à Jean Saint Raymond : *Tout borélien de $X \times Y$, à coupes K_σ, s'écrit comme une union dénombrable de boréliens à coupes compactes – et donc comme une union dénombrable de graphes boréliens à valeurs dans $K(Y)$.*

V-5.5. Deuxième théorème de séparation. Les théorèmes de séparation de Lusin et Novikov admettent des généralisations importantes et délicates, regroupées sous le nom de “deuxième théorème de séparation”. En voici la version la plus simple, historiquement la première :

Théorème V-50 (Deuxième théorème de séparation de Lusin). *Soient A_1 et A_2 des ensembles analytiques d'un espace polonais X. Alors il existe deux ensembles coanalytiques C_1 et C_2 disjoints tels que $A_1 \setminus A_2 \subset C_1$ et $A_2 \setminus A_1 \subset C_2$.*

En posant $A_1' = (X \setminus C_2) \cup (A_1 \cap A_2)$ et $A_2' = (X \setminus C_1) \cup (A_1 \cap A_2)$, on obtient une autre forme qui ressemble beaucoup au premier théorème de séparation :

Corollaire V-51 (Deuxième théorème de séparation de Lusin reformulé). *Soient A_1 et A_2 des ensembles analytiques d'un espace polonais X. Alors il existe deux ensembles analytiques A_1' et A_2' tels que $A_1 \subset A_1'$, $A_2 \subset A_2'$, $A_1' \cap A_2' = A_1 \cap A_2$.*

C'est donc le même énoncé que le premier théorème (Théorème V-14)... sauf que l'intersection $A_1 \cap A_2$ n'est plus forcément supposée vide.

À première vue, pour obtenir le Théorème V-51, il suffit d'appliquer le premier théorème de séparation dans l'espace $X \setminus (A_1 \cap A_2)$, obtenant ainsi deux boréliens disjoints $B_1 \supset A_1$ et $B_2 \supset A_2$, puis d'ajouter l'analytique $A_1 \cap A_2$ à l'un et l'autre. Mais ce raisonnement ne fonctionne pas du tout, car $X \setminus (A_1 \cap A_2)$ est coanalytique, ce n'est donc a priori ni un espace polonais, ni même un espace souslinien.

Il s'agit donc en pratique de travailler dans un espace métrique séparable non souslinien ; cela rend la preuve du Théorème V-50 beaucoup plus sophistiquée que celle du Théorème V-14. En plus des schémas de Souslin, elle fait intervenir la structure des arbres qui les représentent, et des indices à valeur dans les ordinaux, pour représenter la complexité des parties prenantes [Dellacherie,Kechris]. Ensuite, comme dans le passage du Théorème de séparation de Lusin à celui de Novikov, on peut aussi considérer une infinité dénombrable d'ensembles. Après Lusin et Novikov, des auteurs comme Gabriel Mokobodzki, Claude Ambrose Rogers, Alexander (Alekos) Kechris, Benjamin Miller et d'autres ont contribué à réécrire, améliorer et augmenter ces résultats. Voici deux énoncés représentatifs :

THÉORÈME V-52 (réduction coanalytique, version à deux ensembles). *Dans un espace X polonais ou souslinien,*
(a) Soient A_1 et A_2 deux ensembles analytiques. Alors il existe deux ensembles analytiques A_1' et A_2' tels que $A_1 \subset A_1'$, $A_2 \subset A_2'$, $A_1' \cap A_2' = A_1 \cap A_2$, $A_1' \cup A_2' = X$;
(b) Soient C_1 et C_2 deux parties coanalytiques, alors il existe deux ensembles coanalytiques C_1' et C_2' tels que $C_1' \subset C_1$, $C_2' \subset C_2$, $C_1' \cap C_2' = \emptyset$, $C_1' \cup C_2' = C_1 \cup C_2$.

THÉORÈME V-53 (réduction coanalytique, version générale). *Dans un espace X polonais ou souslinien,*
(a) Si $(A_n)_{n\in\mathbb{N}}$ est une suite de parties analytiques, alors il existe une suite $(A_n')_{n\in\mathbb{N}}$ de parties analytiques telles que $A_n' \supset A_n$ pour tout n, $A_n' \cup A_m' = X$ pour tous entiers n, m distincts, et $\bigcap_{n\in\mathbb{N}} A_n' = \bigcap_{n\in\mathbb{N}} A_n$;
(b) Si $(C_n)_{n\in\mathbb{N}}$ est une suite de parties coanalytiques, alors il existe une suite $(C_n')_{n\in\mathbb{N}}$ de parties coanalytiques telles que $C_n' \subset C_n$ pour tout n, les C_n' sont deux à deux disjoints, et $\bigcup_{n\in\mathbb{N}} C_n' = \bigcup_{n\in\mathbb{N}} C_n$.

REMARQUES V-54. (i) Dans l'un et l'autre théorème, les énoncés (b) et (a) sont identiques, modulo passage au complémentaire.
(ii) L'énoncé du Théorème V-52 est précisé par rapport au Théorème V-50, seulement par la condition supplémentaire $A_1' \cup A_2' = X$.
(iii) La preuve du Théorème V-52 n'est guère plus simple que celle du Théorème V-53 ; dans un cas comme dans l'autre, on a besoin d'outils conceptuels nettement plus sophistiqués que pour les Théorèmes de séparation V-14 ou V-45.
(iv) L'énoncé du Théorème V-53(b) permet de comprendre la terminologie de **réduction coanalytique** : partant d'une famille $(C_n)_{n\in\mathbb{N}}$ dont l'union est un certain ensemble C, on a trouvé une nouvelle famille $(C_n')_{n\in\mathbb{N}}$ d'ensembles inclus dans les C_n, et qui forme une *partition* de C. Ce processus de réduction des ensembles C_n en C_n' est donc un analogue sophistiqué, dans la classe des coanalytiques, du procédé familier de réduction des boréliens où à partir de $(B_n)_{n\in\mathbb{N}}$ on pose $B_n' = B_n \setminus (B_1 \cup \ldots \cup B_{n-1})$. On dit que la classe des coanalytiques, comme la classe des boréliens, possède la **propriété de réduction** ; ce n'est pas le cas en revanche de la classe des analytiques. [Kechris]

V-5.6. Ensemble d'unicité. Le Théorème V-52 était originellement motivé par l'étude d'un ensemble que l'on a déjà rencontré en passant.

DÉFINITION V-55 (ensemble d'unicité). *Soit $C \subset X \times Y$ un borélien dans le produit de deux espaces X et Y. On appelle ensemble d'unicité de C*

$$U(C) = \{x \in X;\ \exists! y \in Y,\ (x, y) \in C\}.$$

Autrement dit, $U(C)$ est l'ensemble des x dont la section C_x est un singleton. Il est naturel de s'intéresser à cet ensemble sur lequel C devient un graphe. Le principe de réduction coanalytique permet de démontrer un résultat de structure simple et frappant :

Théorème V-56 (coanalyticité de l'ensemble d'unicité). *Soient X et Y deux espaces polonais, et $C \subset X \times Y$ un borélien ; alors $U(C)$ est coanalytique.*

Remarque V-57. Il est faux en général que $U(C)$ soit borélien. En effet, soient A un ensemble analytique non borélien dans X (j'ai admis son existence dans le Théorème V-10(d)) et f une application continue $\mathcal{N} \to X$ telle que $f(\mathcal{N}) = A$. Soit z_0 un point extérieur à $\mathcal{N}$ et isolé ; de sorte que $Z = \mathcal{N} \cup \{z_0\}$ est encore un espace polonais (pour construire cela rigoureusement, par exemple on considère $\mathcal{N} \times \{0\} \cup \{(0,1)\}$. Soit maintenant $C \subset X \times Z$ défini par $\{(f(z), z);\ z \in Z\} \cup \{(x, z_0);\ x \in X\}$ (on a simplement ajouté un point sur chaque section de l'antigraphe de f). L'ensemble d'unicité $U(C)$ est alors le complémentaire de l'image de f, c'est donc bien un coanalytique, mais pas un borélien !

Preuve du Théorème V-56. L'argument qui suit est dû au jeune logicien canadien Ronnie Chen (élève d'Alekos Kechris).

1. Par le Lemme V-42, il suffit de traiter le cas où $C = \{(f(z), z); z \in Z\}$ pour un certain espace polonais Z. Par le Théorème V-31 il existe une bijection mesurable φ de $\mathcal{C}$ dans Z, de sorte que $C = \{(f(\varphi(c)), \varphi(c)),\ c \in \mathcal{C}\}$, et l'enemble d'unicité de C est le même que celui de $\widetilde{C} = \{(f \circ \varphi(c), c)\} \subset X \times \mathcal{C}$, l'antigraphe de la fonction $\widetilde{f} = f \circ \varphi$. Tout cela pour dire qu'il suffit de prouver la coanalyticité de l'ensemble d'unicité de l'antigraphe d'une fonction borélienne définie sur $\mathcal{C}$. Dans la suite on notera C cet antigraphe et f cette fonction ; l'ensemble d'unicité $U(C)$, c'est alors l'ensemble des valeurs qui sont atteintes une fois et une seule par f.

2. Soit X_2 l'ensemble des $x \in X$ qui sont images par f d'au moins deux points distincts. On note $\Delta(\mathcal{C}) = \{(z, z);\ z \in \mathcal{C}\}$ la diagonale de $\mathcal{C}$, et similairement $\Delta(X)$ la diagonale de X ; on pose aussi $F(z, z') = (f(z), f(z'))$. Alors $X_2 = f(F^{-1}(\Delta(X)) \cap (\mathcal{C} \setminus \Delta(\mathcal{C}))$ (vérifier !) ; c'est donc l'image d'un borélien par une application borélienne (vérifier !), et donc un ensemble analytique par le Théorème V-10(b). Par passage au complémentaire, l'ensemble des valeurs qui sont atteintes au plus une fois est un coanalytique.

3. Pour $z \in \mathcal{C}$ on note $z = (z_n)_{n \in \mathbb{N}}$. Pour tout $n \in \mathbb{N}$ fixé, les ensembles $A_n^0 = f(\{z_n = 0\})$ et $A_n^1 = f(\{z_n = 1\})$ sont analytiques ; par le Théorème de réduction coanalytique V-50 il existe des ensembles analytiques ${A'_n}^0$ et ${A'_n}^1$, contenant A_n^0 et A_n^1 respectivement, tels que ${A'_n}^0 \cup {A'_n}^1 = X$ et ${A'_n}^0 \cap {A'_n}^1 = A_n^0 \cap A_n^1$.

4. Soit

$$Q = \Big\{x \in X;\ \forall z \in \mathcal{C},\ \big[(\forall n \in \mathbb{N}\ x \in {A'_n}^{z_n}) \Rightarrow f(z) = x\big]\Big\}.$$

Alors le complémentaire de Q est la projection sur X de l'ensemble $R(x, z)$ défini par

$$\big[\forall n \in \mathbb{N},\ (z_n = 0) \text{ ou } (x \in {A'_n}^0)\big] \text{ et } \big[\forall n \in\ N,\ (z_n = 1) \text{ ou } (x \in {A'_n}^1)\big] \text{ et } [f(z) \neq x].$$

Ainsi l'ensemble R est l'intersection de trois ensembles, chacun des deux premiers est une intersection dénombrable d'unions d'un borélien avec un analytique, et le troisième est un borélien ; finalement R est analytique et sa projection aussi (on a enchaîné les propriétés énumérées au Théorème V-10). L'ensemble Q est donc coanalytique aussi.

5. Maintenant pour tout $x \in X$ et pour tout $n \in \mathbb{N}$, il existe $i_n \in \{0,1\}$ tel que $x \in {A'_n}^{i_n}$; en posant $z = (i_1, i_2, \ldots)$ on voit que la relation définissant Q est vérifiée pour (x, z), de sorte que $x = f(z)$. En particulier, Q est inclus dans $f(Z)$.

6. Pour tout $x \in X_2$ on a $|f^{-1}(x)| \leq 1$ et pour tout $x \in Q$ o a $|f^{-1}(x)| \geq 1$; donc l'ensemble d'unicité de C est exactement $X_2 \cap Q$, intersection de deux coanalytiques, et donc lui-même coanalytique. □

Remarque V-58. Pour manier les analytiques et coanalytiques, les experts en théorie descriptive des ensembles sont souvent guidés par une intuition appuyée sur les prédicats logiques apparaissant dans les formules. Par exemple, pour l'ensemble X_2 on dira en considérant la formule $X_2 = \{x;\ \forall z, z' \in X,\ [f(z) = f(z') = x \implies z = z']\}$ que le prédicat $[f(z) = f(z') = x \implies z = z']$ est borélien en (x, z, z') et que le quantificateur universel $\forall$ appliqué aux variables (z, z') fait du résultat un ensemble coanalytique.

En guise d'application du Théorème V-56, voici une nouvelle preuve de l'uniformisation pour un borélien C à coupes dénombrables. Pour rappel, on a démontré au Théorème V-39 qu'un tel borélien est union dénombrable de graphes ; par la nouvelle approche on va seulement retrouver le corollaire utile de sélection mesurable, à savoir : *Soit $C \subset X \times Y$ un borélien à sections dénombrables, alors la projection B de C sur X est mesurable et il existe une fonction borélienne $f : B \to Y$ telle que $f(x) \in C_x$ pour tout $x \in B$.*

Preuve alternative de la sélection mesurable de Lusin–Novikov. Par le Lemme V-42 il suffit de traiter le cas où C est fermé, donc chaque C_x aussi. Dans ce cas, pour tout $x \in B$ l'ensemble C_x est fermé, non vide, dénombrable, il admet donc un point isolé w, c'est à dire que $B_r(w) \cap C_x = \{w\}$; soit alors $(V_n)_{n \in \mathbb{N}}$ une base de voisinages fermés de Y : pour tout x il existera donc n tel que $V_n \cap C_x$ est un singleton, de sorte que B est égal à l'union des $U(C \cap (X \times V_n))$, qui est coanalytique par le Théorème V-56. Mais B est aussi analytique comme projection d'un borélien ; il est donc borélien par le Théorème V-17 de Souslin. En outre, pour tout n, la restriction de $C \cap (X \times V_n)$ à son ensemble d'unicité est un graphe (par définition de l'ensemble d'unicité) borélien (car intersection de C avec les deux boréliens $X \times V_n$ et $U(C) \times Y$). Ainsi l'on trouve une suite de graphes boréliens $f_n : D_n \to Y$, où chaque D_n est borélien, avec $\bigcup D_n = B$ et le graphe de f_n est inclus dans C. Pour conclure il suffit de transformer $(D_n)_{n \in \mathbb{N}}$ en partition de B par le procédé habituel. □

CHAPITRE VI

La mesure de Lebesgue

Jusqu'à présent, on a étudié la théorie de Lebesgue dans le cadre abstrait développé par Radon et ses successeurs. Dans ce chapitre et le suivant, l'accent portera sur des mesures particulières dans l'espace euclidien $\mathbb{R}^n$, muni de sa topologie habituelle : ce sont la mesure de Lebesgue (volume n-dimensionnel en dimension n) ; et ses généralisations appelées mesures de Hausdorff d'autre part, qui concernent toutes les dimensions (y compris fractionnaires) entre 0 et n.

La mesure de Lebesgue est celle que l'on utilise couramment, "par défaut", dans $\mathbb{R}^n$, le plus souvent sans le préciser. Elle correspond à la notion intuitive de volume n-dimensionnel et elle est invariante aussi bien pour la structure euclidienne de $\mathbb{R}^n$, que pour sa structure de groupe ; l'intégrale qui lui est associée prolonge le concept d'intégrale de Riemann. Il est vital, en analyse réelle, d'être bien au fait de ses principales propriétés.

Après avoir expliqué comment on peut construire la mesure de Lebesgue (c'est à dire prouver son existence, ce qui est délicat, et son unicité, ce qui est très facile), je passerai en revue quelques-unes des propriétés d'invariance qui la rendent si naturelle. Puis je montrerai que l'intégrale associée à la mesure de Lebesgue généralise le concept d'intégrale de Riemann, et enchaînerai avec ses propriétés les plus remarquables, en particulier la formule de changement de variables. Enfin je reviendrai sur la mesurabilité des parties de $\mathbb{R}^n$ au sens de Lebesgue, en lien avec l'axiomatique.

VI-1. Construction de la mesure de Lebesgue, encore

On a déjà rencontré la mesure de Lebesgue dans $\mathbb{R}$ (section II-8). On va la passer à nouveau en revue (un peu de répétition ne fera pas de mal) et la généraliser à $\mathbb{R}^n$.

Définition VI-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 $P = [a_1, b_1] \times \ldots \times [a_n, b_n] \subset \mathbb{R}^n$ $(-\infty < a_i \leq b_i < +\infty)$ on ait

$$\lambda_n[P] = \prod_{j=1}^{n} (b_j - a_j). \tag{56}$$

Cette mesure est appelée **mesure de Lebesgue n-dimensionnelle** *et notée λ_n (ou $\mathcal{L}_n$, ou $\mathcal{L}^n$, ou λ, ou $\mathcal{L}$). On note également*

$$\lambda_n[A] = |A|_n = |A|.$$

(ii) Si $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est une fonction borélienne λ_n-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 la mesure de Lebesgue complétée, ou tout simplement 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*

$$A \subset E \subset B; \qquad \lambda_n[B \setminus A] = 0.$$

Une fonction $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, mesurable pour cette tribu, est dite Lebesgue-mesurable.

La définition de la tribu complétée en (iii) suit celle du Théorème II-93 ; on reparlera dans la section VI-4.3 de la structure des ensembles Lebesgue-mesurables.

Pour l'instant, commençons par vérifier que la Définition VI-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 II-82(i). En revanche, établir l'*existence* de la mesure de Lebesgue nécessite un peu plus de travail. 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.

VI-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 IV-39 n'est autre que $\mathcal{B}(\mathbb{R}^n)$. Par définition de la mesure produit, cette mesure vérifie (56), 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

$$\lambda_m \otimes \lambda_n = \lambda_{m+n}, \tag{57}$$

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, et cela impliquera la construction de λ_n. Je passerai en revue trois arguments différents. En exercice, on pourra adapter les preuves pour obtenir des constructions directes de λ_n.

VI-1.2. Via le théorème de prolongement de Carathéodory. C'est la démonstration qui a déjà été présentée dans la section II-8 ; elle repose sur la σ-additivité de la fonction longueur sur la famille des intervalles.

VI-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,

$$x = \sum_{k \geq 1} x_k \, 2^{-k}, \qquad x_k \in \{0, 1\}.$$

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 $x \in \{0,1\}^{\mathbb{N}}$. Par exemple, 5/8 s'écrira $(1,0,1,0,0,0....)$.

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 II-87. On peut alors transporter la mesure $\beta^{\otimes \mathbb{N}}$ sur l'intervalle $[0,1]$, via l'application

$$\varphi : (x_k)_{k \in \mathbb{N}} \longmapsto \sum_{k \geq 1} x_k 2^{-k}.$$

L'ensemble des nombres de $[0,1]$ dont le développement en base 2 commence par une suite donnée $(x_1, \ldots, x_k)$ 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 (correspondant aux "écritures impropres" : 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. 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 $(0,0,1,1,1,1,1,\ldots)$ (écriture impropre de 1/4) et du point $(0,1,1,0,0,0,0,0,\ldots)$ (écriture propre de 3/8).

Comme la mesure ainsi définie coïncide avec la mesure de Lebesgue sur les intervalles dyadiques, et que les intervalles dyadiques engendrent tous les boréliens de $[0,1]$, on conclut que λ est bien la mesure de Lebesgue.

VI-1.4. Via le théorème de représentation de Riesz. Une troisième façon de construire la mesure de Lebesgue consiste à faire appel au Théorème de Riesz III-63, ou même à sa version simplifiée III-67. Dans ce cas, la forme linéaire positive à 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.

Pour 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

$$f \leq 1_I \leq g, \qquad \int f\,d\lambda = \int_{\textbf{Riem}} f = (b-a)-\varepsilon, \qquad \int g\,d\lambda = \int_{\textbf{Riem}} g = (b-a)+\varepsilon,$$

où $\int_{\textbf{Riem}}$ 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.

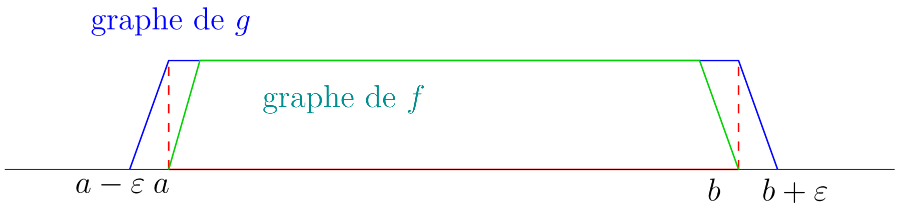

FIGURE 1. Fonctions continues approchant $1_{[a,b]}$

REMARQUE VI-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.

VI-2. Propriétés fondamentales de la mesure de Lebesgue

Cette section est consacrée à 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 euclidienne ;

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), j'utiliserai la notion de "mesure doublante" (Définition II-71) ; la propriété souhaitée découlera alors de c). En chemin, on reviendra sur la notion de Lebesgue-négligeabilité.

VI-2.1. Invariance par translation. Si P est un pavé de $\mathbb{R}^n$ et $\tau : x \to x+h$ 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 $\tau(P)$. Cela, 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 VI-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 II-16 (ii)). Il s'ensuit que $\mu = \mu[C]\,\lambda_n$. □

Remarques VI-4. (i) On pourrait dans le Théorème VI-3 remplacer l'hypothèse "finie sur les compacts" par "finie sur l'intervalle $[0, 1]$" (exercice).

(ii) Le résultat n'est plus vrai sans une hypothèse de finitude. Par exemple, la mesure de comptage est une mesure de Borel invariante par translation. On verra d'autres exemples dans le Chapitre VII.

VI-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 de dimension n, $\mathbb{T}^n$, comme le quotient de $\mathbb{R}^n$ par $\mathbb{Z}^n$, autrement dit par la relation déquivalence : $x\mathcal{R}y \Leftrightarrow x - y \in \mathbb{Z}^n$. 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 $\mathbb{C}$. 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 VI-5 (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 effectivement 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. □

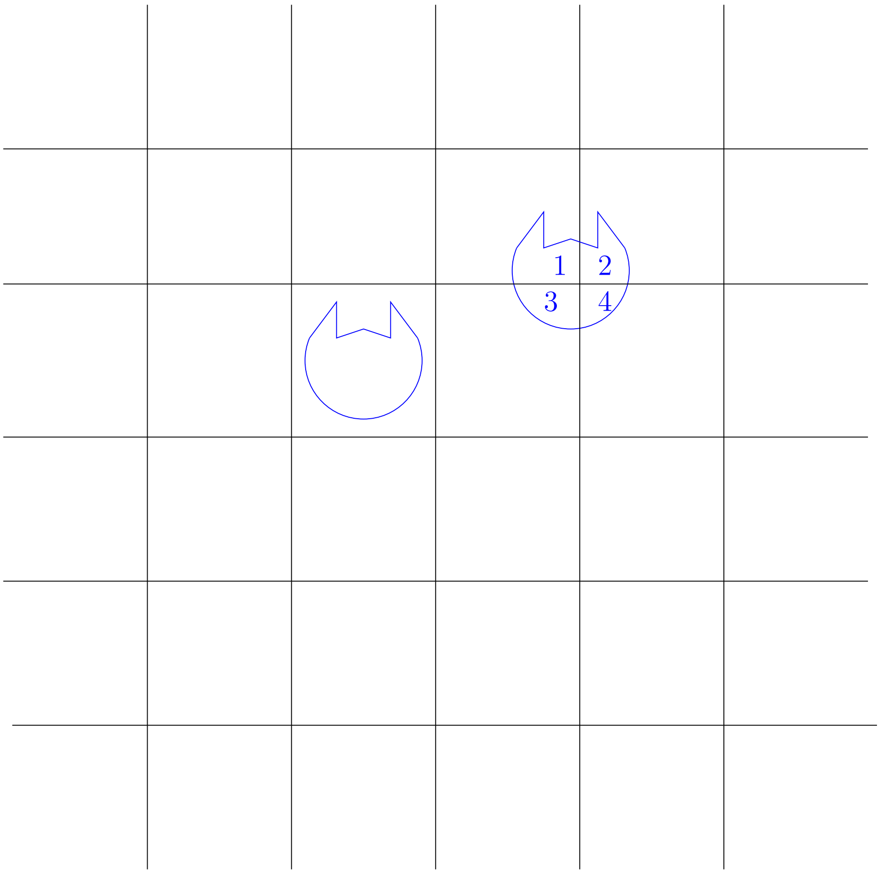

Figure 2. Invariance de la mesure de Lebesgue par translation dans le tore. La somme des aires des quatre morceaux du chat translaté coïncide avec l'aire totale du chat initial.

VI-2.3. Action des homothéties. Soit $f : x \to \alpha x$, 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 $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\lambda_n[\alpha A] = |\alpha|^n \lambda_n[A].$$

VI-2.4. Régularité et diffusivité.

Proposition VI-6. *La mesure de Lebesgue sur $\mathbb{R}^n$ est régulière, sans atomes 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 II-64. Par construction, elle attribue bien sûr aux singletons la mesure nulle. Par ailleurs, si on se donne $x \in \mathbb{R}^n$ 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,

$$\lambda_n[B[x,2r]] \leq 2^n \lambda_n[B[x,r]].$$

$\square$

VI-2.5. Diffusivité de la restriction. Voici maintenant une question plus subtile : que dire de la restriction de la mesure de Lebesgue à un ensemble mesurable ? Est-elle bien répartie, diffuse, doublante ? Cela dépend de la forme de l'ensemble.

PROPOSITION VI-7 (restriction à un domaine convexe). *Soit C un domaine convexe de $\mathbb{R}^n$; alors la restriction $\lambda_n \lfloor_C$ de la mesure de Lebesgue à C est 2^n-doublante.*

DÉMONSTRATION. Soit $x \in C$. La boule $B_r(x)$ dans l'espace métrique C n'est autre que $B_r(x) \cap C$. Comme C est convexe, il est étoilé par rapport à x ; en utilisant l'invariance par translation on peut supposer que $x = 0$, de sorte que $C \subset \lambda C$ pour tout $\lambda \geq 1$. Alors $B_{2r}(0) \cap C \subset B_{2r}(0) \cap (2C) = 2(B_r(0) \cap C)$, et $\lambda_n[B_{2r}(0) \cap C] \leq 2^n \lambda_n[B_r(0) \cap C]$. $\square$

COROLLAIRE VI-8. *La restriction de la mesure de Lebesgue λ_n à une boule, à un cube, à un cône sont 2^n-doublantes.*

Voici maintenant en exercice deux situations typiques.

EXERCICE VI-9. (i) Soit D un domaine constitué d'une boule et d'une union finie de cônes pleins (une sorte de "hérisson" mathématique). Montrer que la mesure de Lebesgue restreinte à D est doublante.

(ii) Soit D le domaine du plan (x, y) délimité par les conditions $0 \leq x \leq 1$, $0 \leq y \leq x^2$. Montrer que la restriction de λ_2 à D n'est pas doublante. On pourra considérer $x = (0, 0)$.

L'intuition qui se dégage de ces exemples est la suivante : pour qu'un ensemble, disons ouvert, induise une mesure de Lebesgue doublante, il ne doit pas présenter de pointes trop effilées. Voici le concept naturel :

DÉFINITION VI-10 (domaine lipschitzien). *On dit que O, ouvert de $\mathbb{R}^n$, est un domaine lipschitzien si son bord peut s'écrire comme une union finie de graphes d'applications lipschitziennes $\mathbb{R}^{n-1} \to \mathbb{R}$.*

La reformulation suivante sera admise ici, Cf par exemple [Grisvard].

PROPOSITION VI-11 (domaine lipschitzien, reformulation). *Un ouvert O de $\mathbb{R}^n$ est lipschitzien si et seulement si il satisfait une condition de cône intérieur uniforme : près de tout point x de ∂O on peut faire bouger un cône fini, d'ouverture uniformément minorée, à l'intérieur de O jusqu'à toucher x.*

FIGURES

Le théorème qui suit, également admis, est une première réponse à la question qui a motivé cette incursion dans les mesures restreintes.

THÉORÈME VI-12 (doublement sur les domaines lipschitziens). *Soit O un ouvert lipschitzien de $\mathbb{R}^n$; alors la restriction de la mesure λ_n à O est doublante.*

VI-2.6. 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 particulier, 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 VI-13. 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 VI-14. 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 VI-15 (les graphes mesurables sont négligeables). *Soient $D \subset \mathbb{R}^n$, et $f : D \to \mathbb{R}^m$ une application mesurable, avec $m \geq 1$. Alors le graphe de f est de mesure de Lebesgue nulle dans $\mathbb{R}^{n+m}$.*

Proposition VI-16 (l'image lipschitzienne d'un ensemble négligeable est négligeable). *Soient A un borélien de mesure de Lebesgue nulle dans $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^m$ 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 VI-15. Pour l'instant je vais me limiter au cas où f est continue ; le cas général viendra 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 $\delta \to 0$. $\square$

Preuve de la Proposition VI-16. 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. (Noter que rien ne garantit que $f(A)$ soit un borélien.) $\square$

La Proposition VI-16 admet un corollaire intéressant :

Corollaire VI-17 (Les applications lipschitziennes préservent la Lebesgue--mesurabilité). *Soit Ω un ouvert de $\mathbb{R}^n$, et $f : \Omega \to \mathbb{R}^m$ 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. On se rappelle aussi que les images par les applications continues des ensembles *boréliens* sont Lebesgue-mesurables (Théorème V-26).

Démonstration du Corollaire VI-17. 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 VI-16 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 etre **gras** au sens de la topologie, c'est-à-dire intersection dénombrable d'ouverts denses.

Exemple VI-18. 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. 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.

Remarque VI-19. Cette construction se généralise facilement en remplaçant I par $\mathbb{R}$ tout entier, ou $\mathbb{R}^n$, ou n'importe quel ouvert de $\mathbb{R}^n$.

Remarque VI-20. 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.

Et par ailleurs, comme je l'ai déjà mentionné, les ensembles Lebesgue-négligeables sont bien plus nombreux que les ensembles boréliens. En fait on peut se permettre de modifier *arbitrairement* ces ensembles sans remettre en cause leur Lebesgue-mesurabilité.

Exercice VI-21. (i) Soit C l'ensemble triadique de Cantor. Montrer que toute partie de C est Lebesgue-mesurable : cela fournit une famille de parties mesurables, de cardinalité 2^c.

(ii) Soit C' l'ensemble des $\sum_{k\in\mathbb{N}} \alpha_k 3^{-k}$, où $\alpha_k \in \{0,1\}$. (C'est semblable à l'ensemble triadique de Cantor, mais les chiffres sont dans $\{0,1\}$ plutôt que dans $\{0,2\}$.) Montrer que C' est de mesure nulle, et montrer que pourtant $C' + C' = [0,1]$. (Attention, c'est une addition ensembliste, pas une union !)

(iii) Pour tout k, soit D_k un sous ensemble à deux éléments de $\{0, ..., 9\}$. On appelle X l'ensemble de tous les nombres de $[0,1]$ dont la décimale de rang k appartient à D_k. Quelle est la mesure de X ? Le cardinal de X ? (On rappelle le théorème de Cantor–Bernstein : S'il existe une injection de E dans F et une autre de F dans E alors E et F sont en bijection.) On note maintenant $D = (D_k)_{k\in\mathbb{N}}$; à chaque D est associé un $X = X_D$. Construire une famille non dénombrable de suites D telle que les X_D sont deux à deux disjoints, chacun de mesure nulle, et leur union est égale à $[0,1]$.

VI-2.7. Action des contractions. Il est intuitif que la contraction des longueurs induit une contraction du volume. Le théorème suivant précise cette idée.

Théorème VI-22 (Réduire les distances réduit les volumes). *Soit Ω un ouvert de $\mathbb{R}^n$, et $f : \Omega \to \mathbb{R}^n$ une application 1-lipschitzienne :*

$$\forall x, y \in \mathbb{R}^n, \qquad |f(x) - f(y)| \le |x - y|.$$

Alors $f(\Omega)$ est Lebesgue-mesurable, et

$$\lambda_n[f(\Omega)] \le \lambda_n[\Omega]. \tag{58}$$

Remarques VI-23. (i) On peut énoncer ainsi ce théorème en termes boréliens : $f(\Omega)$ est l'union d'un ensemble borélien B et d'un ensemble néligeable, tel que $\lambda_n[B] \le \lambda_n[\Omega]$.

(ii) 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.

(iii) Le Théorème VI-22 admet une généralisation immédiate au cas où f est seulement supposée L-lipschitzienne, avec L éventuellement différent de 1 : il suffit de remplacer (58) par

$$\lambda_n[f(\Omega)] \le L^n\, \lambda_n[\Omega].$$

Le concept de mesure de Hausdorff permettra de démontrer un énoncé encore bien plus général : voir la Proposition VII-6.

Preuve du Théorème VI-22. Notons pour commencer que $f(\Omega)$ est Lebesgue-mesurable en vertu de la Proposition VI-17 ; 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

$$\lambda_n[f(A)] \le \lambda_n[A].$$

La mesure de Lebesgue étant 2^n-doublante, on peut appliquer le Corollaire II-103 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 VI-16, $f(N)$ est négligeable. Donc

$$\begin{aligned} \lambda_n[f(\Omega)] = \lambda_n \left[f\left(\bigcup_{j\in\mathbb{N}} B_j \right) \right] &= \lambda_n \left[\bigcup_{j\in\mathbb{N}} f(B_j) \right] \\ &\le \sum_{j\in\mathbb{N}} \lambda_n[f(B_j)] \le \sum_{j\in\mathbb{N}} \lambda_n[B_j] \\ &= \lambda_n \left[\bigcup_{j\in\mathbb{N}} B_j \right] = \lambda_n[\Omega], \end{aligned}$$

où l'on a utilisé le fait que les boules B_j sont disjointes. □

Corollaire VI-24 (les isométries préservent le volume). *Soit $f : \mathbb{R}^n \to \mathbb{R}^n$ 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 VI-22 à f et à f^{-1}. □

Remarque VI-25. Une isométrie de $\mathbb{R}^n$ est forcément une application affine ; on peut donc aussi 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 VI-26. 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 VI-26 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).

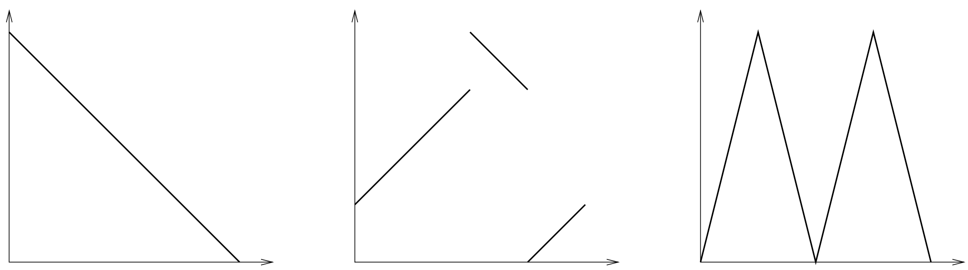

Figure 3. Quelques graphes de transformations préservant la mesure de Lebesgue

En plusieurs dimensions, la classe des applications préservant la mesure de Lebesgue est d'une très grande importance dans de nombreux domaines de la mathématique. 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.

VI-2.8. 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 VI-27 (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,*

$$\lambda_n[P] = |\det A|. \tag{59}$$

Remarque VI-28. Il s'agit ici de volume non orienté.

D'abord, pourquoi ce résultat est-il naturel ? Considérons une application linéaire de la forme $T(x_1, \ldots, x_n) = (\alpha_1 x_1, \ldots, \alpha_n x_n)$, 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 VI-24), 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.

Preuve du Théorème VI-27. 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 (59) sont donc nuls.

Dans le cas où A est inversible (et donc T est bijective $\mathbb{R}^n \to \mathbb{R}^n$), on va établir l'énoncé plus général

$$T_\# \lambda_n = |\det A|^{-1} \lambda_n. \tag{60}$$

Montrons que (59) et (60) sont équivalents. L'équation (60) 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 (59) par le choix $B = C$. Réciproquement, si (59) 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 VI-3, on conclut que $\mu = \lambda_n$, ce qui revient à (60).

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 (60), soit sous la forme (59), 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 :

$$Ax = (\alpha x_1, x_2, \ldots, x_n),$$

alors l'image de C est un pavé de côtés $|\alpha|, 1, \ldots, 1$; donc de volume $|\alpha| = |\det A|$.

(III) si A est de la forme

$$Ax = (x_1 + x_2, x_2, \ldots, x_n),$$

alors $AC = P \times [0,1]^{n-2}$, où P est le parallélogramme (2-dimensionnel) de sommets $(0,0)$, $(0,1)$, $(1,1)$ et $(1,0)$. Or 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.

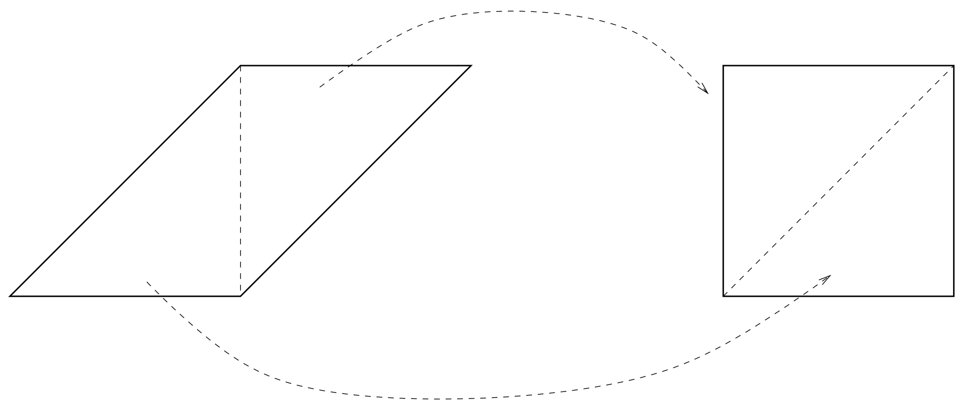

Figure 4. Le parallélogramme a même aire que le carré

On note ensuite que la formule (60) 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 (60) pour n'importe quel A inversible. □

VI-3. L'intégrale de Lebesgue généralise l'intégrale de Riemann

J'ai déjà mentionné sans preuve que l'intégrale de Lebesgue généralise l'intégrale de Riemann. Ce fait est majeur à plusieurs titres : non seulement il assure la cohérence entre les deux plus importantes théories d'intégration ; mais en outre, l'intégrale de Riemann est une notion simple, avec laquelle la lectrice est sans doute familière ; et tous les procédés habituels d'intégration numérique de fonctions (et donc de calcul numérique d'aires ou de volumes de formes délimitées par des graphes) se ramènent en pratique à des variantes de l'intégrale de Riemann : méthode des rectangles, des trapèzes, etc. *Même quand on utilise toute la force de la théorie de Lebesgue, le plus souvent on réalise les calculs numériques ou pratiques par la méthode de Riemann.*

Je me limiterai ici à la dimension 1, même si les démonstrations se généralisent sans autre problème que la lourdeur des notations. Et pour simplifier, je me limiterai à des fonctions positives, le cas général étant conçu pour s'y ramener. Pour commencer, rappelons 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 bornées sur tout intervalle compact $[a,b] \subset J$ (par exemple, les fonctions $x \longmapsto x$ ou $x \longmapsto 1/\sqrt{x}$ sont localement bornées sur $]0,+\infty[$). On appellera **subdivision** d'un intervalle $[a,b]$ une famille de sous-intervalles $I_0,\ldots,I_K$ de $[a,b]$ vérifiant $I_k = [a_{k-1},a_k]$, avec $a = a_0 < a_1 < \ldots < a_K = b$.

Définition VI-29 (Riemann-intégrabilité). *Soit J un intervalle de $\mathbb{R}$, et $f : J \to \mathbb{R}_+$ une fonction localement bornée sur J. Une subdivision σ de l'intervalle $[a,b] \subset J$ en sous-intervalles $I_1,\ldots,I_K$ étant donnée, on définit $m_k(f) := \inf_{I_k} f$, $M_k(f) := \sup_{I_k} f$, et*

$$I^-(f,\sigma) := \sum_k |I_k|\, m_k(f), \qquad I^+(f,\sigma) := \sum_k |I_k|\, M_k(f).$$

On pose alors

$$\mathcal{R}^-_{[a,b]}(f) := \sup_{\sigma\in\Sigma} I^-(f,\sigma), \qquad \mathcal{R}^+_{[a,b]}(f) := \inf_{\sigma\in\Sigma} I^+(f,\sigma),$$

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.

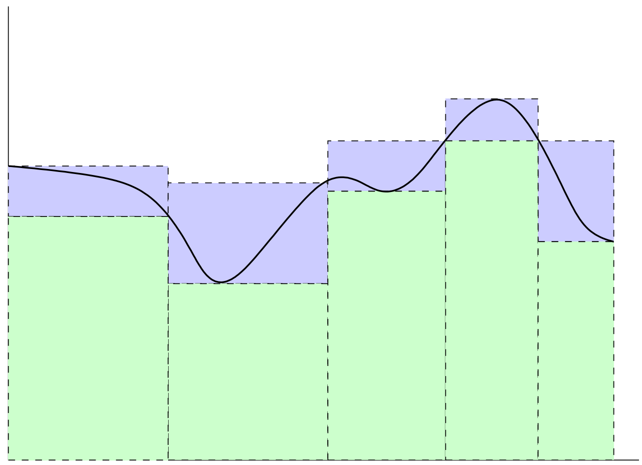

Figure 5. L'intégrale de Riemann définie par encadrements : $I^-(f,\sigma)$ est la somme des aires des rectangles inférieurs, $I^+(f,\sigma)$ l'aire totale hachurée.

Voici maintenant le principal résultat de cette section.

Théorème VI-30 (L'intégrale de Lebesgue généralise celle de Riemann). *Soient J un intervalle de $\mathbb{R}$ et $f : J \to \mathbb{R}_+$ une fonction bornée sur les compacts de J. Alors f est Riemann-intégrable si et seulement si*

(i) elle est Lebesgue-mesurable ;

(ii) l'ensemble de ses points de discontinuité est négligeable.

Dans ce cas, l'intégrale de Riemann de f est égale à l'intégrale de Lebesgue de f.

Remarque VI-31. Dans la Définition VI-29, la fonction f n'est pas *a priori* supposée mesurable. Si elle l'est, alors l'ensemble D de ses points de discontinuité est automatiquement mesurable (exercice), et donc de mesure nulle. Dans le cas contraire, la négligeabilité a le sens habituel : on peut inclure D dans un ensemble négligeable de $\mathbb{R}$, ou encore, on peut inclure D dans une union finie d'intervalles de longueur totale arbitrairement petite.

Remarque VI-32. La fonction $1_{[0,1]\cap\mathbb{Q}}$ est un exemple de fonction Lebesgue-mesurable bornée qui n'est pas Riemann-intégrable. (Que valent $\mathcal{R}^+_{[0,1]}f$ et $\mathcal{R}^-_{[0,1]}f$?)

Preuve du Théorème VI-30. 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 $\int_J f = \lim_{k\to\infty} \int_{[a_k,b_k]} f$, 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

$$\sum_k |I_k|\, m_k(f) \geq \mathcal{R}(f) - \varepsilon, \qquad \sum_\ell |I'_\ell|\, M_\ell(f) \leq \mathcal{R}(f) + \varepsilon.$$

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

$$\sum_k |I_k|\, m_k(f) \geq \mathcal{R}(f) - \frac{1}{m}, \qquad \sum_k |I_k|\, M_k(f) \leq \mathcal{R}(f) - \frac{1}{m}.$$

On définit alors les fonctions f_m^- et f_m^+ par

$$x \in \mathrm{Int}(I_k) \Longrightarrow \quad f_m^-(x) = m_k(f), \quad f_m^+(x) = M_k(f),$$

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 $\sigma_1, \ldots, \sigma_m$, 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,

$$\int f^+ = \lim \int f_m^+, \qquad \int f^- = \lim \int f_m^-,$$

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 E l'ensemble de toutes les extrémités des sous-intervalles I_k des subdivisions σ_m ; comme E est dénombrable, il est mesurable et de mesure nulle. Pour presque tout $x \in [a,b] \setminus E$, on a $f^+(x) = f^-(x)$, ce qui veut dire que x est point intérieur d'une famille d'intervalles J_m, décroissante, vérifiant

$$\lim_{m\to\infty}\big(\inf_{J_m} f\big) = \lim_{m\to\infty}\big(\sup_{J_m} f\big) = f(x).$$

Il s'ensuit que f est continue en x. Cela prouve que l'ensemble des points de discontinuité de f est négligeable.

3. Réciproquement, soit f une fonction bornée, Lebesgue-intégrable sur $[a,b]$, positive, dont l'ensemble des points de discontinuité est de mesure nulle ; si l'on montre que f est Riemann-intégrable, alors on saura par l'étape 2 que la valeur de l'intégrale de Riemann de f coïncide avec celle de l'intégrale de Lebesgue. 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 E l'ensemble de toutes les extrémités de ces intervalles, et soit x un point de continuité de f n'appartenant pas à E. 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 $m \to \infty$, autrement dit

$$\sup_{I_k^m} f - \inf_{I_k^m} f \xrightarrow[m\to\infty]{} 0,$$

soit encore $f_m^+(x) - f_m^-(x) \longrightarrow 0$. 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 $\int f_m^+ - \int f_m^- \longrightarrow 0$, ce qui signifie exactement que f est Riemann-intégrable. ☐

VI-4. Règles de calcul associées à l'intégrale de Lebesgue

VI-4.1. Dérivation et intégration dans $\mathbb{R}$. J'ai déjà mentionné au Chapitre I que l'une des motivations de Lebesgue était de construire une théorie dans laquelle intégration et dérivation seraient toujours des opérations inverses l'une de l'autre, offrant une solution abstraite générale au **problèmes des primitives**. Le cadre naturel de son principal résultat en la matière est celui des **applications absolument continues**.

Définition VI-33 (absolue continuité). *Soient I un intervalle de $\mathbb{R}$, et $f : I \to \mathbb{R}$ 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 $(1 \leq k \leq N)$,*

$$\sum_{1\leq k\leq N} |b_k - a_k| \leq \delta \Longrightarrow \sum_{1\leq k\leq N} |f(b_k) - f(a_k)| \leq \varepsilon.$$

Il est clair qu'une fonction lipschitzienne est absolument continue : dans la Définition VI-33 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, de dérivée bornée, est lipschitzienne, et donc absolument continue.

Il est clair par ailleurs que l'absolue continuité implique l'uniforme continuité (pourquoi ?) ; le concept d'absolue continuité est donc intermédiaire entre celui d'uniforme continuité et celui de lipschitzianité.

Remarque VI-34. 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 VI-35 (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*

$$f(x) - f(a) = \int_a^x f'(t)\, dt. \tag{61}$$

En conséquence de quoi, *pour retrouver la primitive d'une fonction dérivée, il suffit de l'intégrer*. Je démontrerai ce résultat plus tard, dans le Chapitre **??**.

Remarque VI-36. La formule (61) reste vraie si f est continue sur $[a, b]$ et dérivable *partout, sauf en au plus un ensemble dénombrable* (en particulier si f est différentiable au sens classique du terme). La théorie de Lebesgue n'est pas assez fine pour démontrer ce résultat, qui s'inscrit dans l'intégrale de Denjoy : voir [Gordon, Théorème 6.27] (The Integrals of Lebesgue, Denjoy, Perron and Henstock). (On peut consulter aussi Bruckner, Differentiation of real functions.)

VI-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 dans la section VI-1.1,

$$\lambda_m \otimes \lambda_n = \lambda_{m+n}.$$

Remarque VI-37. 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 ; on y reviendra dans le Chapitre **??**.

À titre d'illustration, on va donner deux applications du théorème de Fubini : la démonstration générale de la Proposition VI-15 (qui dans la section VI-2.6 avait été démontrée seulement pour des graphes de fonctions continues) ; puis une démonstration alternative du Théorème VI-27.

Preuve de la Proposition VI-15. Notons d'abord que le graphe de f est mesurable, car image réciproque de 0 par l'application mesurable $(x, y) \longmapsto y - f(x)$. En outre, l'application indicatrice du graphe de f vaut $1_{f(x)=y}(x, y)$. Ensuite, par Fubini,

$$\begin{aligned}\int_{\mathbb{R}^{n+m}} 1_{f(x)=y}\,(x, y)\, d\lambda_n(x)\, d\lambda_m(y) &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} 1_{y=f(x)}\,(x, y)\, d\lambda_m(y) \right) d\lambda_n(x) \\ &= \int_{\mathbb{R}^n} 0\, d\lambda_n(x) = 0.\end{aligned}$$

Ceci prouve que le graphe de f est négligeable. □

Preuve alternative du Théorème VI-27. Comme dans la preuve vue en section VI-2.8, on se ramène au cas où A est inversible, on montre l'équivalence entre (59) et (60) 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 de la mesure de Lebesgue par permutation) on peut supposer que $Ax = (A'(x_1,\ldots,x_n),x_n)$, et la sous-matrice A_{n-1} formée des $n-1$ premières lignes et colonnes de A est inversible :

$$A = \begin{pmatrix} & & & 0 \\ & A_{n-1} & & \vdots \\ & & & 0 \\ * & \ldots & * & 1 \end{pmatrix}$$

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 section de P selon $x_n = y$; cette section 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

$$(A^{-1})_\#\lambda_n = \bigl((A_{n-1}^{-1})_\#\lambda_{n-1}\bigr) \otimes \lambda_1.$$

En particulier, si la formule (60) 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 (60) est vraie pour A' et A'', elle est donc aussi vraie pour A grâce à l'invariance par produit. □

VI-4.3. Changement de variable. Dans la section IV-3 on a rencontré 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 qui suit est un bon compromis entre généralité et simplicité. On notera $d_x\varphi$ la différentielle de φ en x, que l'on peut identifier à la matrice des dérivées partielles :

$$[d_x\varphi]_{ij} = \left(\frac{\partial \varphi_i}{\partial x_j}\right).$$

La notation $\lambda_n\lfloor_A$ désignera la restriction de λ_n au borélien A.

Théorème VI-38 (Changement de variable C^1 dans $\mathbb{R}^n$). *Soient U un ouvert de $\mathbb{R}^n$ et $\varphi : U \to V$ 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 $V \to \mathbb{R}$ (ou pour toute fonction f mesurable positive $V \to [0,+\infty]$),

$$\int_{\varphi(U)} f\,d\lambda_n = \int_U (f\circ\varphi)|\det d\varphi|\,d\lambda_n;$$

(iii) $\varphi_\#(\lambda_n\lfloor_U) = m\,\lambda_n\lfloor_{\varphi(U)}$*, où* m *est la fonction définie par*

$$m(y) = \frac{1}{\left|\det d_{\varphi^{-1}(y)}\varphi\right|}.$$

Remarque VI-39. 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 VI-40. Les erreurs classiques dans l'application du Théorème VI-38 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 VI-41. Le Théorème VI-27 est un cas particulier du Théorème VI-38, correspondant au cas où φ est affine bijective.

Remarque VI-42. À son tour, le Théorème VI-38 se généralise considérablement :

(a) L'hypothèse de régularité C^1 peut être assouplie en régularité Lipschitz, ou même en des hypothèses encore beaucoup plus générales telles que la différentiabilité presque partout (mais la Remarque VI-39 n'est plus forcément valide).

(b) L'hypothèse de bijectivité de φ peut être remplacée par l'injectivité en-dehors d'un ensemble négligeable ; mais 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 $\mathbb{R}^n \to \mathbb{R}^m$ 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 VII.

Ces diverses généralisations sont abordées dans le livre [Evans–Gariepy] ; et dans divers articles de recherche dont certains sont très récents. La section **??** sera l'occasion de revenir sur ces généralisations.

Revenant au Théorème VI-38, on peut le prouver de plusieurs manières. L'équivalence entre les énoncés (i)–(iii) est une conséquence facile de la définition de la mesure image (Définition IV-65), du théorème abstrait de changement de variable (Théorème IV-67) 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 VI-4.2 pour (re)démontrer le Théorème VI-27 ; la seconde, au contraire, considère le Théorème VI-27 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 VI-22, qui lui-même fait appel au lemme de recouvrement de Vitali. Les deux preuves utilisent un argument de localisation.

Première preuve du Théorème VI-38. 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 $[a, b] \subset U$,

$$\lambda[\varphi(I)] = \int_a^b |\varphi'(x)|\, dx. \tag{62}$$

Étant un difféomorphisme, φ est soit strictement croissante, soit strictement décroissante sur $[a, b]$. Dans le premier cas, la formule (62) devient $\varphi(b)-\varphi(a) = \int_a^b \varphi'(x)\, dx$, ce qui est évidemment vrai. Dans le deuxième cas, (62) 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 VI-43. *Soient U et V deux ouverts de $\mathbb{R}^n$ et $\varphi : U \to V$ un difféomorphisme. Alors localement φ s'écrit comme composition de permutations des coordonnées, et de difféomorphismes préservant au moins une coordonnée.*

Preuve du Lemme VI-43. 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 VI-38. 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 :

$$\begin{aligned}(f^z)_\#(\lambda_{n-1}\lfloor_{U^z}) &= |\det d_{x'} f^z|^{-1}\, (\lambda_{n-1}\lfloor_{\varphi^z(U^z)}) \\ &= |\det d_{(x',z)} f|^{-1}\, (\lambda_{n-1}\lfloor_{\varphi^z(U^z)}).\end{aligned}$$

En appliquant le théorème de Fubini, on en déduit, exactement comme dans la démonstration en fin de sous-section VI-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 VI-43 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 $\{U_{x_1}, \ldots, U_{x_N}\}$. 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 $C_1, \ldots, C_N$ deux à deux disjoints dont l'union est K. Puisque le théorème est vrai dans chaque U_{x_i}, on a

$$\forall j \in \{1, \ldots, N\}, \qquad \lambda_n[\varphi(C_j)] = \int_{C_j} |\det d_x\varphi|\, dx.$$

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. □

Seconde preuve du Théorème VI-38. 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 VI-44. *Soient U et V deux ouverts de $\mathbb{R}^n$ et $\varphi : U \to V$ 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 $y \in B_r(x)$,

$$(1-\delta)\,|\det d_y\varphi| \le |\det d_x\varphi| \le (1+\delta)\,|\det d_y\varphi|;$$

(b) pour tous $y, z \in B_r(x)$,

$$(1-\delta)\,|T_x\varphi(y) - T_x\varphi(z)| \le |\varphi(y) - \varphi(z)| \le (1+\delta)\,|T_x\varphi(y) - T_x\varphi(z)|.$$

Preuve du Lemme VI-44. 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

$$|x - y| \le \varepsilon \Longrightarrow \quad |\det d_y\varphi - \det d_x\varphi| \le \delta\,|\det d_x\varphi|,$$

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 $\|d_y\varphi - d_x\varphi\| \le \eta$ pour $|x - y| \le r$. 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

$$\Big|\big(\varphi(y)-\varphi(z)\big)-\big(T_x\varphi(y)-T_x\varphi(z)\big)\Big| \leq \left(\int_0^1 \|d_{x(t)}\varphi - d_x\varphi\|\,dt\right)|y-z| \leq \eta\,|y-z|.$$

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

$$\Big|\big|\varphi(y)-\varphi(z)\big|-\big|T_x\varphi(y)-T_x\varphi(z)\big|\Big| \leq (\varepsilon K^{-1})|T_x\varphi(y)-T_x\varphi(z)|.$$

La conclusion découle du choix $\eta=K\delta$. □

Revenons maintenant à la preuve du Théorème VI-38. Soit $\delta\in]0,1[$; soit $x\in U$ et soit $r>0$ tels que les énoncés (a) et (b) du Lemme VI-44 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 VI-22 (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 VI-22.)

Un raisonnement identique à partir de l'inégalité de gauche dans (b) mène finalement à

$$(1-\delta)^n\,\lambda_n[T_x\varphi(B)]\leq\lambda_n[\varphi(B)]\leq(1+\delta)^n\,\lambda_n[T_x\varphi(B)]; \tag{63}$$

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 VI-38, on peut partitionner K en boréliens disjoints $C_1,\ldots,C_N$ tels que chaque C_i est inclus dans un U_{x_i}. On peut alors appliquer (63) à chaque C_i :

$$(1-\delta)^n\,\lambda_n[T_{x_i}\varphi(C_i)]\leq\lambda_n[\varphi(C_i)]\leq(1+\delta)^n\,\lambda_n[T_{x_i}\varphi(C_i)]. \tag{64}$$

Pour chaque i, on peut appliquer le Théorème VI-27 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 (64) 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

$$\frac{(1-\delta)^n}{1+\delta}\int_K |\det d_y\varphi|\,dy \le \lambda_n[\varphi(K)] \le \frac{(1+\delta)^n}{1-\delta}\int_K |\det d_y\varphi|\,dy.$$

On conclut la preuve en faisant tendre δ vers 0. □

VI-5*Mesurabilité, non-mesurabilité, et paradoxes de Banach–Tarski

Cette section est l'occasion de développer les éléments abordés dans la Mise au point axiomatique en début de cours. On y trouvera peu de démonstrations, mais plutôt un survol de notions fondamentales et parfois très subtiles.

VI-5.1. Ensembles boréliens et Lebesgue-mesurables. On a déjà noté que la tribu borélienne, intuitivement, est obtenue à partir des boules en appliquant une infinité dénombrable de fois les opérations d'union dénombrable et d'intersection dénombrable. Pour passer de la tribu borélienne dans $\mathbb{R}$ à la tribu des ensembles Lebesgue-mesurables, il a suffi 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

$$A \subset E \subset B, \qquad |B \setminus A| = 0.$$

De manière équivalente, les ensembles Lebesgue-mesurables sont les ensembles E 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. En particulier, tout sous-ensemble d'un ensemble négligeable est Lebesgue-mesurable.

La régularité de la mesure de Lebesgue permet de donner une autre caractérisation, en apparence un peu plus précise.

Proposition VI-45 (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 II-57. 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. □

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.

VI-5.2. Fonctions boréliennes et Lebesgue-mesurables. 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 III-69) implique que toute fonction borélienne $f : \mathbb{R}^n \to \mathbb{R}$, 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 III-70 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

$$\int |f_k(x) - f(x)|\, dx \xrightarrow[k\to\infty]{} 0.$$

Le théorème de Vitali-Carathéodory (Théorème III-72) 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.

Enfin on sait (Théorème III-29) qu'une fonction Lebesgue-mesurable est une fonction qui coïncide avec une fonction borélienne presque partout.

Pour résumer informellement : une fonction borélienne réelle est donc "pas loin d'être continue" et elle est "presque limite de fonctions continues". Quant à une fonction Lebesgue-mesurable, c'est une fonction "presque borélienne", et aussi "presque limite de fonctions continues".

VI-5.3. Existe-t-il des ensembles non mesurables ? Cette question d'apparence anodine va nous entraîner à l'assaut de questions très subtiles, dont certaines touchent à rien moins que les fondations logiques du raisonnement mathématique.

Commençons par nous demander s'il existe des parties non boréliennes. La réponse est affirmative : il n'est pas facile d'exhiber une partie non borélienne, mais un argument de cardinalité permettra de prouver que l'immense majorité des parties de $\mathbb{R}$ est non borélienne. On pourra objecter qu'il serait plus satisfaisant de construire explicitement des ensembles non boréliens ; c'est vrai, mais au moins l'argument non constructif des cardinaux apportera un premier jalon. On peut comparer cette situation au problème des nombres algébriques (c'est à dire les nombres réels qui sont racines d'un polynôme à coefficients entiers, comme $\sqrt{2}$) : il est très facile de montrer que les nombres algébriques forment un ensemble dénombrable, et donc l'immense majorité des nombres réels sont transcendants, c'est à dire non algébriques ; mais il est bien plus difficile de construire explicitement un nombre transcendant (comme le nombre de Liouville, $\sum 10^{-k!}$) ; et il est encore bien plus difficile de montrer que certains nombres bien connus, comme e ou π, sont transcendants (théorèmes de Lindemann).

Théorème VI-46 (Rareté des ensembles boréliens). *L'ensemble $\mathcal{B}(\mathbb{R})$ des parties boréliennes de $\mathbb{R}$ a le même cardinal que $\mathbb{R}$, soit $c = 2^{\aleph_0}$; alors que l'ensemble des parties non boréliennes de $\mathbb{R}$ a pour cardinal 2^c. Le même résultat vaut pour $\mathbb{R}^n$, $n \geq 2$.*

Pour éviter une digression trop importante dans la théorie des cardinaux, je ne donnerai ici qu'une esquisse de preuve ; on trouvera une preuve complète dans [Rudin, p. 53].

Esquisse de preuve du Théorème VI-46. Comment "compter" les boréliens ? On commence par se donner une base dénombrable de voisinages faits d'intervalles ouverts : par exemple, tous les intervalles ouverts de longueur $2^{-\ell}$ centrés en les rationnels $(q_m)_{m\in\mathbb{N}}$. Cette énumération $(I_k)_{k\in\mathbb{N}}$ est fixée une fois pour toute. Une union dénombrable de ces intervalles ouverts peut se représenter comme une suite de 0 et de 1 : pour chaque $k \in \mathbb{N}$ on indique 1 si I_k en fait partie, et 0 sinon. Une intersection dénombrable d'unions dénombrables de I_k se représente alors comme un tableau de 0 et de 1, indexé par $\mathbb{N}^2$; mais cela peut aussi se réindexer par $\mathbb{N}$ (exercice). Pour obtenir les boréliens, on applique une infinité dénombrable de fois l'opération "intersection dénombrable d'unions dénombrables" ; cela peut donc se représenter comme un tableau infini d'entiers, ou encore une fonction de $\mathbb{N}$ dans $\mathbb{N}$. Tout se ramène donc au lemme suivant : *L'ensemble des fonctions de* $\mathbb{N}$ *dans* $\mathbb{N}$ *est en bijection avec* $\mathbb{R}$.

Prouvons ce lemme. On se souvient d'abord que $\mathbb{R}$ est en bijection avec $[0,1]$, qui est lui-même en bijection avec $\{0,1\}^{\mathbb{N}}$, c'est à dire les fonctions de $\mathbb{N}$ dans $\{0,1\}$: il suffit pour cela d'écrire $x \in [0,1]$ en écriture binaire, en traitant à part les nombres dyadiques, qui sont en quantité dénombrable. Reste à montrer que les fonctions de $\mathbb{N}$ dans $\mathbb{N}$ sont en bijection avec les fonctions de $\mathbb{N}$ dans $\{0,1\}$. L'inclusion de $\{0,1\}^{\mathbb{N}}$ dans $\mathbb{N}^{\mathbb{N}}$ est évidente. Réciproquement, soit f une fonction de $\mathbb{N}$ dans $\mathbb{N}$, on va lui associer une suite à valeurs dans $\{0,1\}$: pour cela on inscrit $f(1)$ fois le chiffre 1, puis 0, puis $f(2)$ fois le chiffre 1, puis 0, puis $f(3)$ fois le chiffre 1, etc. (Exercice : vérifier que c'est une injection.) Par théorème de Cantor–Bernstein, les ensembles $\{0,1\}^{\mathbb{N}}$ et $\mathbb{N}^{\mathbb{N}}$ sont bien en bijection.

L'énoncé sur les parties non boréliennes découle de ce premier résultat, et d'un peu de théorie des cardinaux : en retirant un ensemble de cardinal c à un ensemble de cardinal 2^c, on obtient un ensemble dont le cardinal est toujours 2^c et cela est strictement supérieur à c.

Enfin la généralisation à $\mathbb{R}^n$ est facile, quitte à remplacer les intervalles par des boules, par exemple la famille des boules dont le centre a toutes ses coordonnées rationnelles et dont le rayon est de la forme $2^{-\ell}$. $\square$

Cet argument ne s'applique plus à 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 se rappeler que l'ensemble triadique de Cantor C sur $[0,1]$ a même cardinal que $\mathbb{R}$, et que toutes ses parties sont Lebesgue-mesurables puiqu'il est de mesure nulle. La cardinalité de l'ensemble des parties Lebesgue-mesurables est donc au moins la même que $\mathcal{P}(C)$, soit 2^c.

On peut se demander à quoi ressemblerait un ensemble Lebesgue-mesurable non borélien. Lebesgue avait déjà identifié de tels ensembles. Le mathématicien russe Nikolaï Nikolaïevitch Lusin (Louzine) a construit dans les années 1920 des exemples assez explicites, au moyen de fractions continues : on rappelle que tout nombre $x \in [0,1]$ peut s'écrire uniquement sous la forme

$$x = \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}};$$

où les a_k sont des entiers naturels ; on appelle $(a_k)_{k\in\mathbb{N}}$ le développement en fraction continue de x.

Théorème VI-47 (Ensemble non borélien de Lusin). *Soit L l'ensemble des x dont le développement en fraction continue admet une sous-suite infinie dont chaque terme divise le terme suivant. Alors L est Lebesgue-mesurable, mais non borélien.*

Une intuition derrière cet ensemble, et qu'il est défini par des propriétés qui font intervenir une infinité potentielle de suites infinies. Il est traité dans [Nikolaï Luzin, *Sur les ensembles analytiques*, Fundamenta Mathematica, vol. 10, 1927, p. 1-95, p. 77]. Ces thèmes ont déjà été abordés dans le Chapitre d'approfondissement V.

On en vient maintenant à formuler une question bien plus délicate : *Existe-t-il des parties de $\mathbb{R}$ qui ne sont 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 grand logicien Robert Solovay [*A model of set-theory in which every set of reals is Lebesgue measurable*, Annals of Mathematics, vol. 92 (1970), pp. 1-56]

Solovay s'appuyait sur les techniques introduites par Paul 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, le théorème d'indécidabilité de Cohen et le théorème d'indécidabilité de Solovay sont peut-être les trois résultats de logique les plus marquants du vingtième siècle.

Comment construire un ensemble non mesurable ? On se souvient que d'après la Proposition VI-5, 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 VI-48 (Paradoxe de Vitali). *Sous l'hypothèse de l'axiome du choix, il existe une partie V de $\mathbb{T}$ telle que $\mathbb{T}$ puisse s'écrire comme réunion dénombrable disjointe de translatés de V. En particulier, V n'est pas Lebesgue-mesurable.*

Démonstration. La seconde partie du théorème découle de la première : en effet, si V 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 V serait nulle, par σ-additivité.

Pour définir V, on introduit une relation d'équivalence $\mathcal{R}$ dans $\mathbb{T}$ comme suit :

$$x\mathcal{R}y \Longleftrightarrow x - y \in \mathbb{Q}.$$

L'ensemble $\mathbb{T}$ est alors partagé en une infinité de classes d'équivalence, et on choisit un représentant dans chaque classe ; on note V 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 $q_k + V$, où les q_k sont les rationnels de $[0, 1[$ et l'addition est considérée modulo 1. □

Dans le raisonnement précédent, on a utilisé l'axiome du choix pour *choisir* arbitrairement un représentant dans chaque classe d'équivalence. Au vu du théorème de Solovay, cela est inévitable. Le contre-exemple de Sierpiński, mentionné dans la Remarque IV-60, 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 :

— Sergei Bernstein a défini 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]

— Wacław Sierpiński a construit un ensemble $S \subset \mathbb{R}^2$ tel que (i) S intersecte tout ensemble fermé mesurable de $\mathbb{R}^2$ de mesure positive, et (ii) on ne peut pas trouver trois points de S alignés. Il s’ensuit que S n’est pas mesurable, sinon le théorème de Fubini impliquerait que $|S| = 0$ (puisque l’intersection de S avec toute droite verticale est réduite à au plus deux points), et on pourrait trouver un ensemble de mesure positive dans $\mathbb{R}^2 \setminus S$... 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 $\mathbb{R}$... mais 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, même ce dernier espoir est ruiné par le théorème suivant, dû à Banach et Kuratowski, généralisé par Ulam [Dudley, pp.526–527 ; Billingsley, p.37] ; il utilise l’axiome du choix et l’hypothèse du continu.

Théorème VI-49 (Obstruction de Banach–Kuratowski). *Sous hypothèse de l’axiome du choix et de l’hypothèse du continu, 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.*

Ici l’obstruction a trait à la théorie des cardinaux, comme le montre une généralisation abstraite due à Ulam (voir [Oxtoby, p. 25–26]).

Corollaire VI-50. *La seule mesure finie et sans atome que l’on puisse définir sur toutes les parties de $[0,1]$ est la mesure nulle.*

VI-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. 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. L’obstruction de Banach–Kuratowski ne laisse guère le choix : si l’on veut mesurer tous les ensembles, il faut abandonner l’axiomatique même de la mesure, et le seul coupable envisageable est le fameux axiome de σ-additivité, imposé avant tout 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, et 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, et bien sûr qu'elle soit non triviale (non identiquement nulle). 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, j'emploierai parfois informellement le terme de "mesure" pour des fonctions additives qui ne sont pas forcément σ-additives, aucune confusion n'étant possible.

Définition VI-51 (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 $A_1, \ldots, A_k$, et des déplacements (i.e. des isométries affines de déterminant 1) $g_1, \ldots, g_k$ tels que les morceaux $g_1(A_1), \ldots, g_k(A_k)$ 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 VI-52 (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 Felix Hausdorff [Wagon, p. 18], répond à cette question.

Théorème VI-53 (paradoxe de Hausdorff). *Sous hypothèse de l'axiome du choix, il existe un sous-ensemble dénombrable D de la sphère S^2 tel que $S^2 \setminus D$ est paradoxal.*

Ce paradoxe 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 VI-54 (non-existence de mesures finiment additives). *Sous l'hypothèse de l'axiome du choix, soit μ une mesure finiment additive définie sur toutes les parties de $\mathbb{R}^3$, finie sur les compacts et invariante par déplacement. Alors $\mu[K] = 0$ pour tout compact de $\mathbb{R}^3$.*

Corollaire VI-55 (non-existence de mesures finiment additives, autre formulation). *Il est impossible de prouver l'existence d'une mesure finiment additive, non nulle, définie sur toutes les parties de $\mathbb{R}^3$, finie sur les compacts et invariante par déplacement.*

Ce dernier énoncé apporte un point final à notre quête : il est impossible de construire une notion raisonnable de volume sur toutes les parties de $\mathbb{R}^3$. La conclusion est la même pour toute dimension $n \geq 4$: il suffit pour le voir de prendre le produit cartésien des contre-exemples dans $\mathbb{R}^3$, avec $[0,1]^{n-3}$.

Preuve du Théorème VI-53. Soit μ une mesure vérifiant les hypothèses, non nulle sur les compacts ; il existe donc $R > 0$ tel que $\mu[R\,B^3] > 0$, 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 $x \longmapsto x/|x|$ 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 $\mu[D] = 0$. $\square$

VI-5.5. Paradoxe de Banach–Tarski. Du paradoxe de Hausdorff on déduit 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 approfondissent la discussion dans une autre direction, et mènent à s'interroger sur l'axiomatique mathématique, en fournissant des conclusions qui violent le bon sens.

Théorème VI-56 (paradoxes de Banach–Tarski). *(i) Sous l'axiome du choix, 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 toute la mathématique", 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 $A_1, \ldots, A_5$ et de trouver des isométries $g_1, \ldots, g_5$ telles que l'union des $g_i(A_i)$ forme deux boules disjointes, chacune étant de rayon 1.*

Les paradoxes de Banach–Tarski mènent à la même conclusion que celui de Hausdorff : *Au moins pour $n \geq 3$, il est impossible de définir une mesure finiment additive sur l'ensemble de toutes les parties de $\mathbb{R}^n$.*

Or il se trouve que la conclusion est différente pour $n \leq 2$!

Théorème VI-57 (existence de mesure finiment additive dans $\mathbb{R}^2$). *Sous hypothèse de l'axiome du choix, 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 VI-57 ne prétend pas que la mesure μ coïncide avec la mesure de Lebesgue sur tous les boréliens, ce qui suggère que son maniement est délicat. Quoi qu'il en soit, la différence de comportement entre les dimensions inférieures ou égales à 2 d'une part, et supérieures ou égalés 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.

Cela dit, 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.

VI-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, Sierpiński ou Banach–Tarski. En outre il n'est pas vraiment intuitif, car dans le cas où la famille d'ensembles est infinie on ne pourra jamais construire explicitement leurs représentants, ni même esquisser une méthode.

D'un autre côté, si l'on supprime complètement l'axiome du choix, on tombe vite sur des paradoxes qui heurtent tout autant 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, Sierpiński 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.

VI-5.7. Quelle attitude adopter ? Au vu de la discussion précédente, il y a trois attitudes possibles :

- *l'attitude classique :* accepter l'axiome du choix, se résoudre à ce que certains ensembles ne soient pas mesurables, à ce que certains paradoxes existent (après tout, ce ne sont pas les seuls) 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 :* n'accepter que l'axiome du choix dépendant, s'interdire l'usage de l'axiome du choix, mais sans supposer non plus la mesurabilité de toutes les parties de $\mathbb{R}^n$; vérifier la mesurabilité dans les démonstrations, et se résoudre à ce que certains énoncés, comme la non-mesurabilité de certaines parties, soient indémontrables.

L'attitude classique est le choix le plus fréquent. L'axiome du choix mène parfois à des démonstrations formellement élégantes, parfois via ses divers avatars tels que le Lemme de Zorn, ou le Théorème de Tychonov dans sa version la plus générale (un produit quelconque de compacts est compact). Il mène parfois à des énoncés très synthétiques ; on le verra par exemple au sujet de la mesure de Haar au Chapitre **??**.

L'attitude iconoclaste est quelque peu dangereuse, car cela réclame une certaine discipline que de savoir quels sont les résultats qui nécessitent l'axiome du choix et quels sont ceux qui ne le réclament pas. Mais surtout, l'axiome que l'on adopte (la non-existence de parties non mesurables) est un axiome très fort.

C'est finalement l'attitude sceptique que je recommande sans hésitation : c'est la plus économe en axiomes, et l'expérience montre qu'elle suffit à couvrir tous les énoncés classiques d'analyse réelle. Après tout c'est un devoir, en mathématique, de se passer des hypothèses superflues ! Cette attitude n'exclut pas, bien sûr, de recourir à l'axiome du choix dans une phase prospective de recherche de preuve.

VI-5.8. Justification pratique de la mesurabilité. Au vu de la discussion précédente, 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 mesurable 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 VI-58. Soit $f : \mathbb{R}^n \to \mathbb{R}$ 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 $q \le r \le q'$, on a

$$|B_q(x)| \le |B_r(x)| \le |B_{q'(x)}|; \qquad \int_{B_q(x)} f \le \int_{B_r(x)} f \le \int_{B_{q'}(x)} f.$$

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 !

Remarque VI-59. Une situation du même type, un peu plus subtile, survient quand on considère des sommes de Minkowski :

$$A + B := \big\{x \in \mathbb{R}^n;\ \exists\, a \in A;\ \exists\, b \in B;\ x = a + b\big\}.$$

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 une mesure extérieures pour mesurer $A + B$. Ou encore on se rappellera que les images continues de boréliens sont des ensembles analytiques et en particulier Lebesgue-mesurables (c'est la théorie défrichée dans le Chapitre V) ; et il est clair que la somme de Minkowski de deux ensembles boréliens fait partie de cette catégorie. Donc : Si A et B sont boréliens, $A + B$ n'est pas forcément borélien, mais à tout le moins Lebesgue-mesurable.

VI-5.9. Subtilités liées au produit. Plus traîtres que les exemples précédents sont les subtilités liées à la combinaison de négligeabilité et de 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 de $\mathbb{R}$, alors

$$\mathcal{L}(\mathbb{R}^m) \otimes \mathcal{L}(\mathbb{R}^n) \ne \mathcal{L}(\mathbb{R}^{m+n}) \quad !$$

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 $\phi : x \longmapsto (x, x)$, ou sur l'axe horizontal ($y = 0$) dans $\mathbb{R}^2$, via l'application $\phi : x \longmapsto (x, 0)$. 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. 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]).

Ce contre-exemple utilisait l'axiome du choix ; sans cet axiome, ce qu'il en reste est que l'identité $\mathcal{L}(\mathbb{R}^m) \otimes \mathcal{L}(\mathbb{R}^n) = \mathcal{L}(\mathbb{R}^{m+n})$ est tout simplement indémontrable.

VI-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 ; en fait elle risque de compliquer les justifications. C'est une des raisons pour lesquelles je recommande vivement 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 basé sur la tribu borélienne ; et dans ma propre carrière de recherche, je n'ai jamais rencontré un problème où il était vraiment utile de compléter la tribu borélienne.

VI-5.11. En conclusion. La discussion de toute cette section appelle à des recommandations d'humilité : ne pas chercher à mesurer toutes les parties, ne pas chercher à rendre mesurables les parties négligeables, mais se contenter des ensembles boréliens (ou de leur généralisation, les ensembles analytiques et coanalytiques) ; ne pas imposer l'axiome du choix mais se contenter de l'axiome du choix dépendant ; accepter des zones d'ombre et de non-démontrabilité.

CHAPITRE VII

Les mesures de Hausdorff

Ce chapitre est consacré aux mesures de Hausdorff, qui généralisent la mesure de Lebesgue. Même si elles n'ont pas la même importance pratique et universelle que la mesure de Lebesgue, elles jouent un rôle majeur dans de nombreux domaines de la mathématique et des sciences naturelles.

VII-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.

VII-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.

VII-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 $\mathbb{R}^m \to \mathbb{R}^n$, 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 $z \in \mathbb{R}^{m+n} \to (x, y) \in \mathbb{R}^m \times \mathbb{R}^n$ 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 $2\,dr$ et la section a (par homogénéité) une "surface" $r^{n-1}\,d\sigma$. On en déduit la formule de changement de variable*

$$dx = r^{n-1}\,dr\,d\sigma.$$

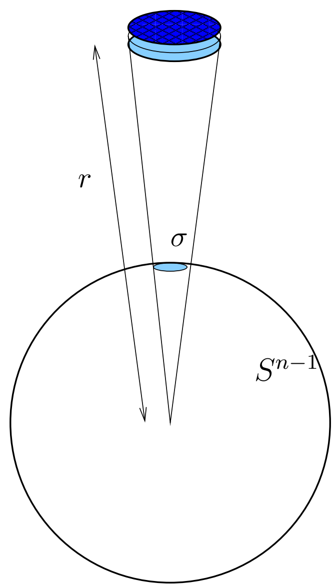

FIGURE 1. Elément de volume autour de x dans les variables sphériques

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 raport 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.

VII-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 $[0,1] \to [0,1]^2$, 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.

Il existe aussi une autre notion populaire de dimension, antérieure à celle de dimension de Hausdorff, et 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**. Mais 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 du mathématicien polono-français Benoît 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].

VII-1.4. Mesures de référence 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 $N > n$.

VII-1.5. Ensembles de Besicovitch et problème de Kakeya. En 1919 le mathématicien japonais Sōichi Kakeya posa le problème suivant, vaguement motivé par des applications mécaniques : *quelle est l'aire minimale qu'une épingle doit balayer pour pouvoir s'orienter dans toutes les directions du plan ?* Il proposa un ensemble de surface fort réduite, mais Besicovitch montra que l'infimum des aires de surfaces balayées était en fait... zéro ! La solution de ce problème était moins intéressante que l'outil technique qu'il introduisit pour cela, : *un ensemble de mesure nulle, compact, comprenant des segments orientés dans toutes les directions.* La construction de Besicovitch, qui avait un parfum de géométrie fractale, lui permettait aussi de construire des contre-exemples en théorie de l'intégration : si f est une fonction définie dans le plan, valant 1 quand (x, y) est dans ce fameux ensemble et a au moins une coordonnée rationnelle, et 0 sinon, alors quelle que soit la direction que l'on choisit, il existe un axe parallèle à cette direction selon lequel f n'est pas Riemann-intégrable ; cette fonction f est alors Riemann-intégrable en dimension 2, sans qu'aucun choix de coordonnées ne permette de la considérer comme Riemann-intégrable dans une direction, puis dans l'autre.

On appelle en son honneur **ensemble de Besicovitch** tout ensemble $B \subset \mathbb{R}^n$, qui pour tout angle $\sigma \in \mathbb{S}^{n-1}$ contient au moins un segment de longueur unité orienté selon σ. Ces ensembles sont reliés à de nombreux problèmes d'analyse, en particulier harmonique, ce qui motiva des recherches approfondies pour déterminer leur taille minimale. Le **problème de Kakeya** consiste à montrer que tout ensemble de Besicovitch dans $\mathbb{R}^n$ est de dimension (de Hausdorff et de Minkoswki) au moins n. Après sa résolution en dimension $n = 2$ par le mathématicien britannique Roy Davies en 1971, il a suscité une quantité considérable de travaux par des analystes de premier plan, établissant avec ténacité des bornes partielles ou étudiant des cas particuliers. Il fallut attendre 2025 pour que la Chinoise Hong Wang et l'Américain Joshua Zahl résolvent le cas $n = 3$, causant l'une des grandes sensations mathématiques de l'époque. Ainsi les questions originelles de Kakeya et Besicovitch ont mené à des développements profonds dans la théorie des dimensions fractionnaires.

VII-2. Construction des mesures de Hausdorff

VII-2.1. Définition. Comme on l'a déjà vu, 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 :

$$|A| := \inf\left\{\sum \ell(I_k);\ I_k \text{ intervalle};\ A \subset \cup I_k\right\}.$$

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

$$|B_r|_d = \alpha(d)\, r^d, \qquad \alpha(d) := \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)},$$

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 analyticité la fonction "volume d'une boule de rayon r en dimension d".

Voici maintenant une première tentative de construction de la "mesure d-dimensionnelle", copiant la définition de la mesure de Lebesgue :

$$\mu^d[A] := \inf\left\{\sum_{k=1}^{\infty}\alpha(d)r_k^d;\quad A\subset\bigcup_{k=1}^{\infty}B_{r_k}(x_k)\right\}\qquad ?$$

On se rend compte tout de suite que cette définition est absurde : le volume 1-dimensionnel d'une boule de $\mathbb{R}^2$ serait fini ! Le problème vient de ce que la notion de dimension doit dépendre uniquement de la structure locale d'un objet, et que donc on doit forcer le recouvrement par des boules à "épouser les détails" de l'ensemble A ; autrement dit, il faut définir la mesure d-dimensionnelle en fonction de *recouvrements par des petites boules.* Avec la mesure de Lebesgue sur $\mathbb{R}$, cette propriété était 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 équivalents en termes de mesure.

Voici donc une deuxième tentative de définition de "mesure d-dimensionnelle" :

$$\mu^d[A] := \lim_{\varepsilon\to 0}\inf\left\{\sum_{k=1}^{\infty}\alpha(d)\,r_k^d;\quad A\subset\bigcup_{k=1}^{\infty}B_{r_k}(x_k);\quad r_k\le\varepsilon\right\}.$$

La mesure ainsi définie est dite **mesure de Hausdorff sphérique** [Falconer1, p. 7]. Elle a le défaut 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 forcément une boule dans A (il suffit que le centre de la boule n'appartienne pas à A...). Ce qui est vrai en revanche, c'est que le diamètre de $A\cap B$ est inférieur ou égal à $2r$.

Pour avoir une notion aussi intrinsèque que possible, et être sûr que la mesure d'un objet ne dépend pas de la taille de l'espace dans lequel on le plonge, on souhaiterait donc définir les mesures de Hausdorff en fonction des diamètres, sans référence à la notion de boule. Cette troisième tentative est la bonne, elle mène à la définition finalement retenue pour les mesures de Hausdorff :

Définition VII-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*

$$\mathcal{H}^d[A] = \lim_{\varepsilon\to 0}\inf\left\{\sum_{k=1}^{\infty}\alpha(d)\,r(C_k)^d;\quad A\subset\bigcup_{k=1}^{\infty}C_k\quad \operatorname{diam}(C_k)\le\varepsilon\right\},\tag{65}$$

où les C_k sont des parties arbitraires de $\mathbb{R}^n$, $r(C_k) := \operatorname{diam}(C_k)/2$ 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 VII-2. (i) Posons

$$\mathcal{H}^d_\varepsilon[A] = \inf\left\{\sum_{k=1}^{\infty}\alpha(d)\,r(C_k)^d;\quad A\subset\bigcup_{k=1}^{\infty}C_k\quad \operatorname{diam}(C_k)\le\varepsilon\right\}.$$

Comme $\mathcal{H}^d_\varepsilon[A]$ est clairement une fonction décroissante de ε, l'existence de $\lim_{\varepsilon\to 0}\mathcal{H}^d_\varepsilon[A]$ 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 $\mathcal{H}^d_\varepsilon[A] < +\infty$, et pour tout $\delta > 0$ on peut trouver une famille dénombrable (C_k) d'ensembles de diamètre au plus ε, recouvrant A, telle que

$$\sum_{k=1}^{\infty} \alpha(d)\, r(C_k)^d \le \mathcal{H}^d_\varepsilon[A] + \delta.$$

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

$$\sum_{k=1}^{\infty} \alpha(d) r(C'_k)^d \le \mathcal{H}^d_\varepsilon[A] + \delta.$$

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 VII-3 (la mesure de Hausdorff est une mesure de Borel). *Pour tout $d \ge 0$, la fonction $A \longmapsto \mathcal{H}^d[A]$ 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

$$\mathcal{H}^d_\varepsilon[\cup A_k] \le \sum_{k\in\mathbb{N}} \mathcal{H}^d_\varepsilon[A_k].$$

En passant à la limite $\varepsilon \to 0$ dans le terme de gauche, et en utilisant l'inégalité $\mathcal{H}^d_\varepsilon \le \mathcal{H}^d$ dans le terme de droite, on trouve

$$\mathcal{H}^d[\cup A_k] \le \sum_{k\in\mathbb{N}} \mathcal{H}^d[A_k].$$

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 II-82 ; 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 II-92. 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 $\mathcal{H}^d_\varepsilon[A \cup B] = \mathcal{H}^d_\varepsilon[A] + \mathcal{H}^d_\varepsilon[B]$, et la conclusion en découle par passage à la limite. □

Exemples VII-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 le 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 $\mu = \delta_0 \otimes \mathcal{H}^1\lfloor_{[0,1]}$, où le symbole $\lfloor$ 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 $d > 1$.

VII-2.2. Propriétés élémentaires. Commençons par un critère de négligeabilité, conséquence immédiate de la structure de mesure extérieure :

Proposition VII-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 $\sum_{k=1}^\infty \operatorname{diam}(B_k)^d$ 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 VII-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*

$$\mathcal{H}^d[f(A)] \le k^d\, \mathcal{H}^d[A].$$

(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

$$|f(x) - f(y)| \le |g(x) - g(y)|,$$

alors pour tout borélien $B \subset A$ et pour tout $d \in [0, n]$ on a

$$\mathcal{H}^d[f(A)] \le \mathcal{H}^d[g(A)].$$

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 $C \subset A$, $\operatorname{diam}(f(C)) \leq k \operatorname{diam}(C)$. L'énoncé (i) en découle immédiatement.

La démonstration de (ii) est similaire : l'hypothèse implique $\operatorname{diam}(f(C)) \leq \operatorname{diam}(g(C))$. □

Passons maintenant à des propriétés d'invariances, qui elles aussi découlent directement de la définition :

Proposition VII-7 (invariance par isométrie-multiplication). *Soit T une application affine de la forme*

$$T(x) = \alpha A x + b,$$

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 VII-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 VII-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 VII-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 VII-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 VII-5 : les petits cubes sont de diamètre $\sqrt{n}2^{-k}$, et la somme de leurs diamètres à la puissance d vaut donc

$$2^{nk} n^{d/2} 2^{-dk} \xrightarrow[k\to\infty]{} 0.$$

Il s'ensuite 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. $\square$

VII-2.3. Régularité. Il résulte de la Proposition VII-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 II-56. Cependant, la propriété II-57 (que de nombreux auteurs appellent aussi régularité) reste vraie :

Théorème VII-10 (régularité faible 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,

$$\mathcal{H}^d[A] = \sup\big\{\mathcal{H}^d[K];\ K \textit{ compact};\ K \subset A\big\}.$$

Remarque VII-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 VII-10. (i) Sans perte de généralité, on suppose que $\mathcal{H}^d[A] < +\infty$. Pour tout k, on a $\mathcal{H}^d_{1/k}[A] \leq \mathcal{H}^d[A] < +\infty$, et par la remarque VII-2(ii), on peut trouver un recouvrement de A par des ouverts $C'_{k,j}$ de diamètre au plus $2/k$, tel que

$$\sum_j \alpha(d)\, r(C'_{k,j})^d \leq \mathcal{H}^d_{1/k}[A] + \frac{1}{k}.$$

On pose alors

$$O_k := \bigcup_{j\in\mathbb{N}} C'_{k,j}, \qquad G := \bigcap_{k\geq 1} O_k.$$

Il est clair que G contient A, et d'autre part pour tout k on a

$$\mathcal{H}^d_{2/k}[G] \leq \sum_j \alpha(d)\, r(C'_{k,j})^d \leq \mathcal{H}^d_{1/k}[A] + \frac{1}{k}.$$

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,

$$\lim_{j\to\infty} \mathcal{H}^d[A \cap F_{k,j}] = \mathcal{H}^d[A \cap O_k] = \mathcal{H}^d[A].$$

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 $\cap O_k = G$, 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 $F := \bigcap_{k\geq 1} F_{1/k}$. $\square$

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 VII-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 $x \in A$,*

$$2^{-d} \leq \limsup_{r\to 0} \frac{\mathcal{H}^d[B_r(x) \cap A]}{\alpha(d)\, r^d} \leq 1$$

et pour $\mathcal{H}^d$-presque tout $x \in \mathbb{R}^n \setminus A$,

$$\lim_{r\to 0} \frac{\mathcal{H}^d[B_r(x) \cap A]}{\alpha(d)\, r^d} = 0.$$

Remarque VII-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

$$\lim_{r\to 0} \frac{\mathcal{H}^d[B(x,r) \cap A]}{\alpha(d)\, r^d} = 1.$$

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.

VII-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 (65) 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.

VII-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, on a 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. On verra en outre que la mesure de Hausdorff 1-dimensionnelle prolonge une définition courante de la longueur.

VII-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 VII-14 (inégalité isodiamétrique). *Soit $A \subset \mathbb{R}^n$ un ensemble Lebesgue-mesurable, et r son demi-diamètre. Alors*

$$\lambda^n[A] \le \alpha(n)r^n.$$

En d'autres termes, à diamètre fixé, les boules maximisent le volume.

Remarques VII-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.

La preuve du Théorème VII-14 sera l'occasion d'utiliser pour la première fois la technique puissante de **symétrisation de Steiner**.

Définition VII-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.*

J'admettrai le lemme suivant [Evans-Gariepy pp. 67-68], et suggère comme exercice de le démontrer informellement.

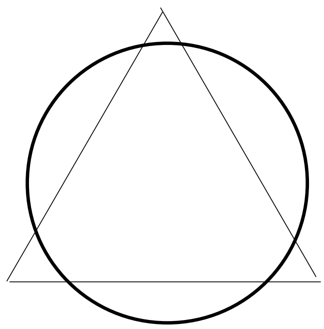

FIGURE 2. Le triangle équilatéral ne rentre pas dans le disque de même diamètre.

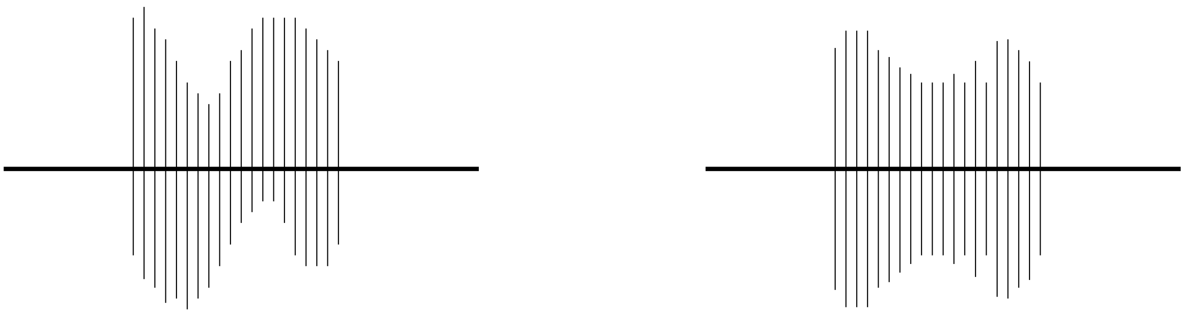

FIGURE 3. Représentation schématique de la symétrisation de Steiner

LEMME VII-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 $(e_1, \ldots, e_n)$ une base euclidienne de $\mathbb{R}^n$. On note S_a la symétrisation de Steiner par rapport à P_a, et $A^* := S_{e_n} S_{e_{n-1}} \ldots S_{e_1} A$. 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 $P_{e_1}, \ldots, P_{e_n}$, et donc symétrique par rapport à l'origine. Il s'ensuit que A^* est contenu dans une boule de centre 0 et de rayon $\operatorname{diam}(A^*)/2$. Le résultat en découle. □

VII-3.2. Dimension n : le volume. La mesure de Hausdorff n-dimensionnelle en dimension n coïncide avec la mesure de Lebesgue λ_n :

THÉORÈME VII-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 VII-18 est plus subtile et utilisera l'inégalité isodiamétrique.

Démonstration partielle du Théorème VII-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 $\mathcal{H}^b[C_n] \in (0, +\infty)$, où $C_n = [0, 1]^n$.

Soit $\varepsilon > 0$, et $k \in \mathbb{N}$ tel que $2^{-k} \le \varepsilon/\sqrt{n} \le 2^{-k+1}$. On peut recouvrir C_n par 2^{nk} cubes de côté 2^{-k}, dont chacun aura un diamètre $\sqrt{n}2^{-k} \le \varepsilon$. Il s'ensuit que $\mathcal{H}^n_\varepsilon[C_n] \le C(n)\,2^{nk}2^{-nk} = C(n)$, où $C(n)$ est une constante ne dépendant que de n. En prenant la limite quand $\varepsilon \to 0$ 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 $C'(n)\mathrm{diam}\,(A)^n$, 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 $\sum C'(n)(\mathrm{diam}\,(A))^n \ge 1$. On en déduit que $\mathcal{H}^n_\varepsilon[C_n]$ est minoré par une constante positive indépendante de ε, et en faisant tendre ε vers 0 on conclut que

$$\mathcal{H}^n[C_n] > 0.$$

□

Démonstration complète du Théorème VII-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 $(C_k)_{k\ge1}$ un recouvrement de A par des ensembles de diamètre inférieur ou égal à ε. Grâce à l'inégalité isodiamétrique, on a

$$\lambda_n[A] \le \sum_k \lambda_n[C_k] \le \sum_k \alpha(n) r(C_k)^n.$$

En passant à l'infimum, on voit que $\lambda_n[A] \le \mathcal{H}^n_\varepsilon[A]$, et donc $\lambda_n[A] \le \mathcal{H}^n[A]$. Il nous reste à montrer l'inégalité inverse.

Il est facile de montrer, en utilisant des cubes dyadiques, que

$$\lambda_n[A] = \inf\left\{\sum_{k=1}^{\infty} \lambda_n[Q_k]; \quad A \subset \bigcup Q_k,\ r(Q_k) \le \varepsilon\right\},$$

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 $\mathcal{H}^n \le c_n\lambda_n$.

Pour conclure, on utilise le résultat suivant, conséquence du Lemme de recouvrement de Vitali, et plus précisément de son Corollaire II-103 (et de la Propriété VI-7 : Étant donné un cube Q et $\varepsilon > 0$, on peut écrire

$$Q = \bigcup_{j\le1} B_j \cup N,$$

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

$$\mathcal{H}^n_\varepsilon[A] \leq \sum_{k\geq 1} \mathcal{H}^n[Q_k] = \sum_{k\geq 1}(\sum_{j\geq 1} \mathcal{H}[B_{k,j}] + \mathcal{H}(N_k))$$

$$\leq \sum_{k\geq 1}\sum_{j\geq 1} \lambda_n[B_{k,j}] = \sum_{k\geq 1} \lambda_n[\cup B_{k,j}] = \sum_{k\geq 1} \lambda_n[Q_k] \leq \lambda_n[A] + \delta.$$

Ceci conclut l'argument. □

VII-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 VII-19 (rectifiabilité). *Soient I un intervalle de $\mathbb{R}$ et $\gamma : I \to \mathbb{R}^n$ une courbe continue injective. On dit que γ est rectifiable sur I si pour tout intervalle compact $[a, b] \subset I$,*

$$L_{[a,b]}(\gamma) :=$$

$$\sup \left\{ \sum_{k=0}^{N} |\gamma(t_{k+1}) - \gamma(t_k)|; \qquad a = t_0 \leq t_1 \leq \ldots \leq t_N \leq t_{N+1} = b, \quad N \in \mathbb{N} \right\} < +\infty$$

où le supremum est pris sur toutes les subdivisions finies $(a = t_0, t_1, \ldots, t_N, t_{N+1} = b)$ de $[a, b]$. On appelle alors

$$L(\gamma) := \sup_{[a,b]\subset I} L_{[a,b]}(\gamma) \tag{66}$$

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. Ce procédé de calcul de longueur est la **rectification** de la courbe.

REMARQUES VII-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'ensuivraient si l'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 $\gamma([a, b])$. Mais par souci de simpliciité, dans cette partie je ne travaillerai qu'avec des courbes injectives.

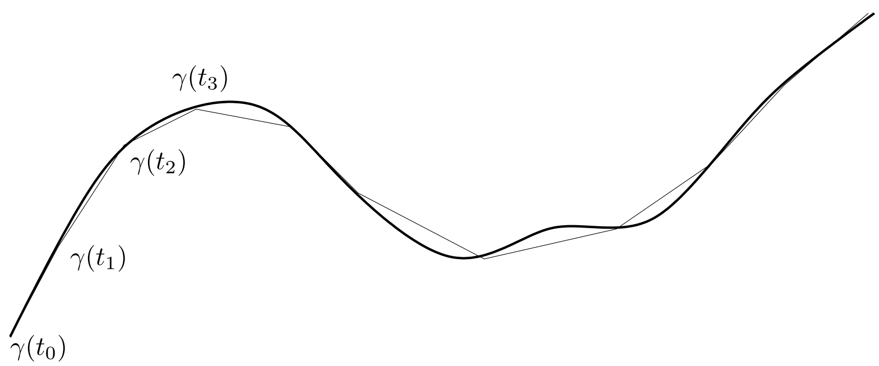

FIGURE 4. Approximation polygonale d'une courbe

Une courbe γ étant donnée, on appelle **reparamétrage** de γ toute courbe (injective) $\widetilde{\gamma}$, 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

$$\ell_{x_0}(x) = \begin{cases} L_{[x_0,x]}(\gamma) & (x \geq x_0) \\ -L_{[x,x_0]}(\gamma) & (x < x_0) \end{cases} \quad ;$$

on vérifie que la fonction ℓ_{x_0} est continue et strictement croissante, en particulier inversible sur son image. On définit alors le reparamétrage $\widetilde{\gamma}$ par

$$\widetilde{\gamma}(x_0 + \ell_{x_0}(x)) = \gamma(x), \qquad \ell(x) = L_{[x_0,x]}(\gamma).$$

Les propriétés suivantes découlent presque immédiatement de la définition.

PROPOSITION VII-21 (propriétés du paramétrage par longueur d'arc). *Soit $\gamma : I \to \mathbb{R}$ une courbe paramétrée par longueur d'arc. Alors pour tout $[a,b] \subset I$,*

$$L_{[a,b]}(\gamma) = b - a;$$

En particulier,

$$|\gamma(b) - \gamma(a)| \leq b - a, \tag{67}$$

et

$$L(\gamma) = |I|.$$

Le théorème suivant montre que la dimension de Hausdorff de dimension 1 est une généralisation du concept de rectifiabilité.

THÉORÈME VII-22 ($L = \mathcal{H}^1$). *Soient I un intervalle de $\mathbb{R}$, et $\gamma : I \to \mathbb{R}$ 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 (67) implique alors que

$\operatorname{diam}(B_k) \geq \operatorname{diam}(A_k)$. 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'identié $\mathcal{H}^1 = \lambda_1$ 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 $a = t_0 \leq t_1 \leq \ldots \leq t_N \leq t_{N+1} = b$ une subdivision de $[a,b]$; cette subdivision découpe l'intervalle I en sous-intervalles ouverts $I_0, I_1, \ldots, I_{N+1}, I_{N+2}$. 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. □

VII-3.4. Autres dimensions entières. On vient de constater 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 seconde est une définition possible du “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) $S(A) := \liminf_{\varepsilon \to 0} \dfrac{\lambda_n[A_\varepsilon] - \lambda_n[A]}{\varepsilon}$;

(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 l'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 σ... la mesure de surface sur ∂A.

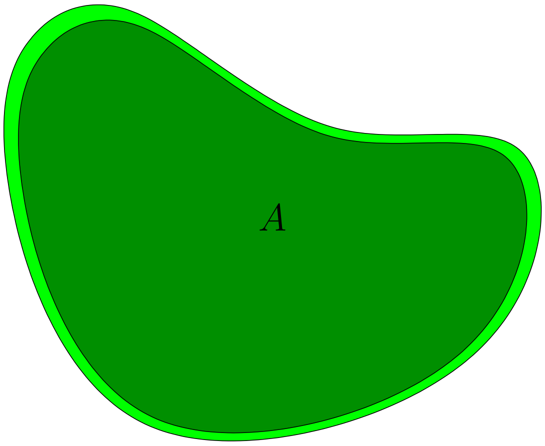

Figure 5. Surface au sens de Minkowski : l'accroissement infinitésimal du volume est donné par le produit de la surface par la largeur d'épaississement

Les trois définitions précédentes de la surface de A peuvent donner des résultats différents pour des ensembles A "pathologiques". Attention donc, dans un contexte peu régulier, à préciser la notion de "surface k-dimensionnelle" employée.

VII-4. Dimension

VII-4.1. Échelle 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 VII-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 VII-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 $\mathcal{H}^d_\varepsilon[A] < +\infty$, 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 $\sum_k r_k^{d'} \leq C\varepsilon^{d'-d} \longrightarrow 0$ dès que $d' > d$. Cela prouve que $\mathcal{H}^{d'}_\varepsilon[A] = O(\varepsilon^{d'-d})$, et en particulier $\mathcal{H}^{d'}[A] = 0$.

Si d'autre part $\mathcal{H}^d[A] > 0$ pour un certain $d > 0$, alors pour tout $\varepsilon > 0$ assez petit on a $\mathcal{H}^d_\varepsilon[A] \geq \delta > 0$; 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 $d' < d$,
$$\sum_k \alpha(d') r_k^{d'} \geq \varepsilon^{d-d'} \delta \, \frac{\alpha(d')}{\alpha(d)} \xrightarrow[\varepsilon \to 0]{} +\infty,$$
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). $\square$

VII-4.2. Dimension de Hausdorff. Au vu de la Proposition VII-23, la fonction $d \longmapsto \mathcal{H}^d[A]$ 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.

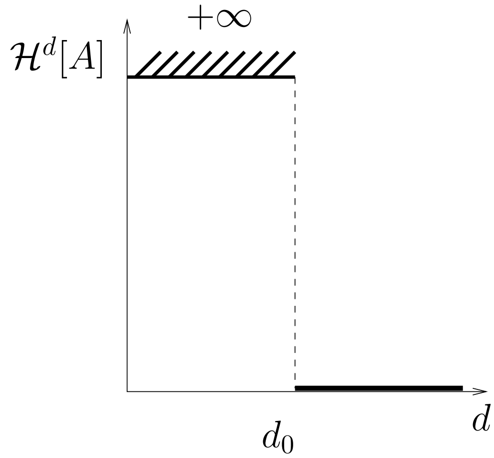

FIGURE 6. Graphe de $\mathcal{H}^d[A]$; $\mathcal{H}^{d_0}[A]$ peut se situer n'importe où sur la ligne pointillée.

Ceci mène naturellement à la définition de la dimension de Hausdorff.

DÉFINITION VII-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*

$$\dim(A) := \inf\{d \geq 0;\ \mathcal{H}^d[A] = 0\} \in [0, n]$$

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 VII-6 :

THÉORÈME VII-25 (dimension des graphes et images). *Soit* $f : \mathbb{R}^n \to \mathbb{R}^m$ *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 VII-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 lipschitziennes.

(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 lipschitziennes, 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 lipschitzienne est toujours de dimension inférieure ou égale à 1. Si on considère une fonction lipschitzienne 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** (bilipschitzienne = bijective lipschitzienne de réciproque lipschitzienne). C'est une des raisons pour lesquelles les applications bilipschitziennes constituent une notion naturelle d'"isomorphisme" dans l'étude des objets fractals.

VII-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

$$\dim_{\mathcal{M}}(A) = \lim_{\varepsilon\to 0} \frac{\log N_\varepsilon(A)}{|\log \varepsilon|},$$

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 $\varepsilon \to 0$ 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 :

$$\dim_{\mathcal{M}}(A) := n + \lim_{\varepsilon\to 0} \frac{\log \lambda_n[A_\varepsilon]}{|\log \varepsilon|}$$

où A_ε est le ε-voisinage de A, i.e.

$$A_\varepsilon := \{x \in \mathbb{R}^n;\ d(x,A) \le \varepsilon\}.$$

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.

VII-4.4. Comparaisons. Nous voici avec deux notions de dimension fractionnaire : Hausdorff et Minkowski, qui ne coïncident pas forcément.

De manière générale, la dimension de Minkowski est toujours supérieure ou égale à la dimension de Hausdorff ; mais 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.

VII-4.5. 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.

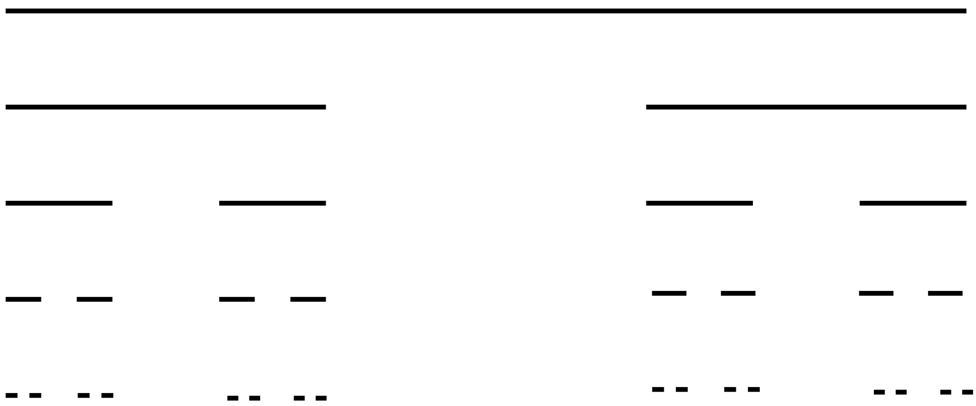

Figure 7. Premières étapes de la construction de l'ensemble triadique de Cantor

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

$$\sum_k |I_k|^d \geq \frac{1}{2}, \tag{68}$$

et il s'ensuit que $\mathcal{H}^d[C] > 0$.

Pour établir l'inégalité (68), 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 $\alpha(d)$ par le coefficient dominant de N_ε quand $\varepsilon \to 0$, 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 avec $\mathcal{H}^d[C]$. Mais 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 d'une 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

$c(k) \to 1$ pour $k \to \infty$, 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 $d > \log 2/\log c(k)$.

VII-4.6. 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 $\ell/3$).

Figure 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

$$\iint_{A\times A} \frac{\mu(dx)\,\mu(dy)}{|x-y|^s} < +\infty,$$

alors $\dim_{\mathcal{H}}(A) \geq s$.

Malgré ces méthodes, le calcul de la dimension de Hausdorff est parfois un casse-tête insoluble, ou presque. 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 a longtemps conjecturé que la dimension de Hausdorff a la même valeur, et cela a été affirmé par Benoît Mandelbrot ; mais pendant longtemps cela restait une conjecture, on savait seulement que $\lim_{\lambda\to\infty} \dim_{\mathcal{H}}(G(f_{s,\lambda},[0,1])) = s$; ce n’est qu’en 2015 que le mathématicien chinois Weixiao Shen fournit une preuve complète [Mathematische Zeitschrift. 289 (1–2) : 223–266 (2018)].

Enfin, comme je l’ai mentionné en début de chapitre, l’estimation de la dimension minimale des ensembles de Besicovitch, aussi appelée problème de Kakeya, a tenu en haleine des générations de spécialistes, à l’interface de la théorie géométrique de la mesure et de l’analyse harmonique ; et on sait aujourd’hui résoudre ce problème seulement en dimensions $n \leq 3$.

VII-5*Changements de variables : aire et co-aire

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 VII-27 (formule de l’aire). *Soit $T : \mathbb{R}^n \to \mathbb{R}^m$ une application lipschitzienne, avec $m \geq n$, et soit $A \subset \mathbb{R}^n$ un ensemble Lebesgue-mesurable. Alors*

$$\int_A |\det \nabla T|\, d\lambda_n = \int_{\mathbb{R}^m} \mathcal{H}^0[A \cap T^{-1}\{y\}]\, \mathcal{H}^n(dy).$$

Théorème VII-28 (formule de la co-aire). *Soit $T : \mathbb{R}^n \to \mathbb{R}^m$ une application lipschitzienne, avec $m \leq n$, et soit $A \subset \mathbb{R}^n$ un ensemble Lebesgue-mesurable. Alors*

$$\int_A |\det \nabla T|\, d\lambda_n = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}[A \cap T^{-1}\{y\}]\, \lambda_m(dy).$$

Exemple VII-29. Soit $f : [0,1] \to \mathbb{R}^n$ une courbe lipschitzienne simple ; alors par la formule de l’aire,

$$\int_0^1 |f'(s)|\, ds = \int_{\mathbb{R}^n} 1_{f([0,1])} \mathcal{H}^1(dy) = \mathcal{H}^1[f([0,1])], \tag{69}$$

ce qui identifie $\mathcal{H}^1$ avec une notion naturelle de plus de longueur d’une courbe. Finalement, pour une courbe lipschitzienne simple à valeurs dans $\mathbb{R}^n$, on peut calculer la longueur de trois façons équivalentes : par mesure de Hausdorff, par rectification (Définition VII-19) ou par intégration de la vitesse. Cela reste vrai si la courbe est absolument continue.

CHAPITRE VIII

Espaces de Lebesgue et mesures signées

Jusqu'ici, on a considéré des fonctions "individuellement". Dans ce chapitre et le suivant, l'attention portera sur des familles entières de fonctions : des "espaces de fonctions", ou **espaces fonctionnels**. On munira ces espaces de structures géométriques et topologiques : par exemple un produit scalaire pour définir l'orthogonalité et plus généralement les angles, une norme pour mesurer la taille des fonctions et leur éloignement, une description de leurs formes linéaires, qui sont autant de façons de les cartographier par des coordonnées. En première approximation, on peut dire que c'est cette étude des propriétés géométriques et topologiques des espaces de fonctions qui 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 faire remonter ce point de vue à Fourier lui-même, avec des motivations issues de la physique mathématique.

Dans ce chapitre, j'introduirai deux types d'espaces fonctionnels. Dans un premier temps, je fixerai une mesure, et construirai des espaces de **fonctions** mesurables, définis par leur "degré d'intégrabilité" : intégrabilité de la puissance p pour les espaces L^p de Lebesgue. L'exploration de ces espaces fut, sous l'impulsion des mathématiciens polonais du Café écossais, le premier grand projet de l'analyse fonctionnelle. Leur étude nous mènera à quelques développements sophistiqués, en particulier les puissantes techniques d'interpolation entre espaces de Lebesgue. Par extension, on considèrera également l'espace de toutes les fonctions mesurables.

Le deuxième type d'espace fonctionnel ne sera pas constitué de fonctions à proprement parler, mais des fonctions d'ensembles : ce sera l'espace des mesures, ou plus précisément des **mesures signées**, que l'on peut considérer comme des "fonctions généralisées".

Les questions prioritaires que l'on se pose sur les espaces fonctionnels sont : les normes, la complétude, la séparabilité, la réflexivité, l'uniforme convexité, l'existence de systèmes de coordonnées commodes. L'étude de ces questions commencera dans le présent chapitre, et se poursuivra dans le chapitre suivant.

La section la plus importante de ce chapitre est la Section VIII-1, qui introduit les espaces de Lebesgue et leurs propriétés élémentaires, approfondies ensuite dans la Section VIII-2. La Section VIII-2.4 explore les fonctions mesurables en général. La Section VIII-4 est consacrée aux mesures signées.

VIII-1. Espaces L^p de Lebesgue

VIII-1.1. Définitions.

Définition VIII-1 (espaces L^p). *Soit $(X, \mathcal{A}, \mu)$ 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 est noté $L^p(X, \mathcal{A}, \mu)$, ou simplement $L^p(\mu)$ (ou encore $L^p(X, \mu)$ ou $L^p(X)$ ou $L^p(d\mu)$ ou $\mathcal{L}^p$, etc).

Remarques VIII-2. (i) L'espace $L^p(\mu)$ dépend de μ ; c'est évident pour $p < \infty$, mais on se laisse plus facilement piéger dans le cas $p = \infty$. Ainsi la fonction $1/x$ appartient à $L^\infty([0,1], \delta_{1/2})$.

(ii) On rencontre aussi parfois l'exposant de Lebesgue en bas : $L_p(X)$; mais je recommande de garder la place d'indice en bas pour le comportement à l'infini, comme dans $C_0(X)$, $C_c(X)$; dans le cadre des espaces L^p, cela pourra se traduire par des poids à l'infini.

Exemples VIII-3. (i) Si $X = \mathbb{N}$, muni de la tribu triviale de toutes les parties, 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 sur la tribu borélienne de $\mathbb{R}^n$. Alors, la fonction $f_\alpha : x \longmapsto |x|^{-\alpha}$ appartient à $L^p(B_1, \lambda_n)$ si et seulement si $p < n/\alpha$, et 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 VIII-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

$$h_{\alpha,\beta} : x \longmapsto \frac{[\log(1/|x|)]^\beta}{|x|^\alpha}$$

pour des valeurs différentes de β. D'autres espaces fonctionnels plus "fins" permettent de distinguer ces fonctions : par exemple, les espaces de Lorentz

$L^{p,q}$, introduits par l'analyste russo-américain George G. Lorentz dans les années 1950. 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, du nom de Józef Marcinkiewicz (brillant représentant de l'illustre école d'analyse harmonique polonaise, tué en 1940 dans les massacres des élites polonaises par l'armée soviétique). Les définitions, données ci-après, peuvent être omises en première lecture.

Définition VIII-5. *Soit* $(X, \mathcal{A}, \mu)$ *un espace mesuré, et soient* $p, q \in]0, +\infty[$*. On appelle espace de Lorentz* $L^{p,q}(X, \mu)$ *l'ensemble des fonctions mesurables* $f : X \to \overline{\mathbb{R}}$ *telles que*

$$\|f\|_{L^{p,q}(X,\mu)} = p^{\frac{1}{q}} \Big\| t\,\mu\big[|f| \geq t\big]^{1/p} \Big\|_{L^q\left((0,+\infty),\frac{dt}{t}\right)} = \left(\int_0^\infty p\, t^{q-1}\, \mu\big[|f| \geq t\big]^{\frac{q}{p}}\, dt \right)^{\frac{1}{q}}$$

soit fini (attention, malgré la notation, il ne s'agit pas d'une norme).

Définition VIII-6. *Soit* $(X, \mathcal{A}, \mu)$ *un espace mesuré, et soit* $p \in]0, +\infty[$*. On appelle espace de Marcinkiewicz* $M^p(X, \mu)$ *l'ensemble des fonctions mesurables* $f : X \to \overline{\mathbb{R}}$ *telles qu'il existe une constante* $C \geq 0$ *telle que*

$$\forall t \geq 0, \quad \mu\big[\{x;\ |f(x)| \geq t\}\big] \leq \left(\frac{C}{t}\right)^p.$$

On notera alors $\|f\|_{M^p(X,\mu)}$ *ou* $\|f\|_{L^{p,\infty}(X,\mu)}$ *l'infimum des constantes* C *admissibles (ce n'est pas une norme non plus).*

Exercice VIII-7. En utilisant l'inégalité de Tchebychev, montrer que pour tout $p \geq 1$, $L^p(X, \mu) \subset L^{p,\infty}(X, \mu)$, 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.

Exercice VIII-8. En utilisant le principe de sommation par tranches, montrer que pour tout $p \geq 1$, $L^{p,p}(X, \mu) = L^p(X, \mu)$.

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 VIII-9. *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 VIII-10 (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* $f : X \to E$ *telles que* $d(0, f) \in L^p(X, \mu)$*.*

VIII-1.2. Inégalité de Minkowski. Le point de départ de l'analyse fonctionnelle des espaces de Lebesgue est l'inégalité suivante, dont le cœur remonte aux travaux de Hermann Minkowski sur les volumes à la fin du dix-neuvième siècle.

THÉORÈME VIII-11 (inégalité de Minkowski). *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et soit $p \in [1, +\infty[$. Alors, pour toutes fonctions f, g mesurables $X \to \overline{\mathbb{R}}$,*

$$\left(\int |f+g|^p \, d\mu\right)^{1/p} \leq \left(\int |f|^p \, d\mu\right)^{1/p} + \left(\int |g|^p \, d\mu\right)^{1/p},$$

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.

La preuve la plus populaire de cette inégalité découle de celle de Hölder [Rudin p. 64, Lieb–Loss p. 48], voir Exercice VIII-14 ci-après. On va donner un autre argument ci-dessous, et en même temps démontrer quelques variantes de l'inégalité de Minkowski, selon un plan parallèle à la présentation de l'inégalité de Hölder dans la Section IV-4.4.

THÉORÈME VIII-12 (variantes de l'inégalité de Minkowski). *Soit $p \in [1, +\infty[$.*

(i) Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, et 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 $(X, \mathcal{A}, \mu)$ un espace mesuré, $k \in \mathbb{N}$, $f_1, \ldots, f_k$ des fonctions mesurables sur X, à valeurs dans $\overline{\mathbb{R}}$. Alors

$$\left(\int \Big|\sum_i f_i\Big|^p \, d\mu\right)^{1/p} \leq \sum_i \left(\int |f_i|^p \, d\mu\right)^{1/p}.$$

(iii) Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \pi)$ deux espaces mesurés σ-finis. Alors, pour toute fonction F mesurable de $X \times Y$ dans $\mathbb{R}_+ \cup \{+\infty\}$, on a

$$\left(\int_X \left(\int_Y F(x,y)\, \pi(dy)\right)^p \mu(dx)\right)^{1/p} \leq \int_Y \left(\int_X F(x,y)^p \, \mu(dx)\right)^{1/p} \pi(dy).$$

(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 $(X, \mathcal{A}, \mu)$ un espace mesuré, $(E, \|\cdot\|)$ un espace vectoriel normé, et $f, g : X \to E$ 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 $(X, \mathcal{A}, \mu)$ un espace mesuré, $f, g : X \to \mathbb{C}$ 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. On va se contenter ici de démontrer l'inégalité (i) du Théorème VIII-12 et d'en déduire l'inégalité de Minkowski du Théorème VIII-11. Le reste (discussion des cas d'égalité dans le Théorème VIII-11, énoncés (ii) à (vi) du Théorème VIII-12) est laissé en exercice.

Pour démontrer (i), on écrit d'abord, par convexité de la fonction $t \longmapsto |t|^p$,

$$|f(x)+g(x)|^p = |\lambda(f(x)/\lambda)+(1-\lambda)(g(x)/(1-\lambda))|^p \leq \lambda|f(x)/\lambda|^p+(1-\lambda)|g(x)/(1-\lambda)|^p.$$

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

$$\inf_{0\leq\lambda\leq 1} \left(\frac{a}{\lambda^{p-1}} + \frac{b}{(1-\lambda)^{p-1}}\right) = \left(a^{\frac{1}{p}} + b^{\frac{1}{p}}\right)^p.$$

On peut être plus explicite : cela correspond à choisir

$$\lambda = \frac{\left(\int |f|^p\right)^{\frac{1}{p}}}{\left(\int |f|^p\right)^{\frac{1}{p}} + \left(\int |g|^p\right)^{\frac{1}{p}}},$$

qui donne l'inégalité souhaitée. □

Remarque VIII-13. Cette méthode de preuve (démonstration d'une inégalité auxiliaire dépendant d'un paramètre, puis optimisation sur ce paramètre) est très répandue en analyse.

Exercice VIII-14. Écrire $(f+g)^p = f(f+g)^{p-1} + g(f+g)^{p-1}$ et appliquer l'inégalité de Hölder pour contrôler $\int f(f+g)^{p-1}$ et $\int g(f+g)^{p-1}$ séparément. Retrouver ainsi l'inégalité de Minkowski.

VIII-1.3. Distances L^p. À ce stade on a seulement défini l'ensemble des fonctions L^p ; on va maintenant munir cet ensemble d'une structure qui, selon les cas, sera soit une “semi-distance”, soit une semi-norme.

Théorème VIII-15 (semi-distances L^p). *Soit $(X, \mathcal{A}, \mu)$ 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*

$$N_p(f) = \left(\int |f|^p \, d\mu\right)^{\min(1,1/p)} \qquad (0 < p < \infty);$$

$$N_\infty(f) = \inf\Big\{C;\ \mu\big[\{x;\ |f(x)| > C\}\big] = 0\Big\};$$

$$N_0(f) = \mu\big[\{x;\ f(x) \neq 0\}\big].$$

La quantité $N_\infty(f)$ *est appelée supremum essentiel de* $|f|$*, ce que l'on note* esssup $|f|$*. Les quantités* $N_p(f)$ *sont également notées* $\|f\|_{L^p(X,\mu)}$ *ou* $\|f\|_{L^p(X)}$ *ou* $\|f\|_{L^p(\mu)}$*, ou* $\|f\|_{L^p}$*, voire* $\|f\|_p$*.*

L'application N_p *définit alors sur* L^p

(i) pour $1 \leq p \leq \infty$ *: une semi-norme, i.e. pour toutes fonctions* f, g *et tout* $\lambda \in \mathbb{R}$*,*

$$N_p(f) \geq 0; \qquad N_p(f+g) \leq N_p(f) + N_p(g); \qquad N_p(\lambda f) = |\lambda| N_p(f);$$

(ii) pour $0 \leq p < 1$ *: une application positive, homogène de degré p, vérifiant l'inégalité triangulaire, i.e. pour toutes fonctions* f, g *et tout* $\lambda \in \mathbb{R}$*,*

$$N_p(f) \geq 0; \qquad N_p(f) \leq N_p(f) + N_p(g); \qquad N_p(\lambda f) = |\lambda|^p N_p(f).$$

En outre, pour tout $p \in [0, +\infty]$*, une fonction* f *dans* $L^p(X, \mu)$ *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 VIII-16 (espaces de Lebesgue). *Soit* $(X, \mathcal{A}, \mu)$ *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* $L^p(X)$ *ou* $L^p(\mu)$ *ou* $L^p(d\mu)$*, ou simplement* L^p*), l'espace vectoriel de toutes les classes d'équivalence de fonctions dans* $L^p(X, \mu)$*, 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* $\|f\|_{L^p(X,\mu)}$ *ou* $\|f\|_{L^p}$*, ou simplement* $\|f\|_p$*.*

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 VIII-17. (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 ; dès qu'il existe des ensembles négligeables non vide, ce n'est pas un espace normé. Le deuxième est obtenu à partir du premier en identifiant des fonctions qui coïncident presque partout, et c'est un espace normé. Cette identification nous mène dans un univers a priori 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. Mais sans cette identification, on ne peut aller bien loin dans l'analyse fonctionnelle. Certains auteurs distinguent les deux espaces, par exemple en utilisant la notation $\mathcal{L}^p$ pour l'espace non quotienté, mais la plupart du temps on tolère la confusion entre les deux, et ce sera le cas dans ce chapitre.

(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 du Théorème VIII-11 ne s'applique plus, et l'inégalité triangulaire non plus. 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 du Théorème VIII-11 (de même que l'inégalité de Hölder). On peut aller plus loin et montrer que L^p n'est pas normable pour $0 \leq p < 1$. 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. On pourra démontrer l'inégalité de Minkowski renversée en exercice.

Pour conclure ce paragraphe, introduisons les semi-distances L^p sur les espaces de Lebesgue à valeurs vectorielles $L^p(X;E)$.

Proposition VIII-18 (espaces de Lebesgue à valeurs vectorielles). *Soit $(X,\mathcal{A},\mu)$ 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 ; j'en admettrai quelques-unes sans démonstration. Le seul point un tant soit peu délicat dans le maniement des espaces de Lebesgue à valeurs vectorielles ne concerne pas les opérations dans les espaces L^p, mais la construction de l'intégrale.

VIII-1.4. Théorème de convergence dominée L^p. Avant d'aller plus loin, voici une variante simple et utile du théorème de convergence dominée, adaptée aux espaces de Lebesgue.

Théorème VIII-19 (convergence dominée dans les L^p). *Soient $(X,\mathcal{A},\mu)$ 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*

$$\int |f_k - f|^p \, d\mu \xrightarrow[k\to\infty]{} 0,$$

c'est-à-dire que f_k converge vers f dans L^p.
Le même énoncé reste vrai si l'on suppose que $|f_k| \leq g_k$ presque partout, avec $g_k \to g$ presque partout et $\int g_k^p \to \int g^p$.

Démonstration. Il est clair que $|f(x)| \leq g(x)$ pour presque tout x, et donc que $f \in L^p$. On applique alors 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. □

VIII-1.5. Théorème de Riesz–Fischer. L'analyse étant basée pour une grande part sur des procédés de limite et d'approximation, on n'utilise le plus souvent les espaces vectoriels normés que s'ils sont **complets**, c'est à dire que toute suite de Cauchy converge. Le théorème suivant assure la complétude des espaces de Lebesgue et ouvre donc la voie à leur usage dans toutes sortes de problèmes. Il est issu des travaux simultanés du mathématicien juif hongrois Frigyes Riesz, déjà rencontré dans le Chapitre III, et du mathématicien juif autrichien Ernst Sigismund Fischer, qui fut élève entre autres de Minkowski. Trois ans à peine après l'intégrale de Lebesgue, ils réalisaient ainsi un pont entre cette nouvelle branche de l'analyse, et des idées issues de la géométrie et de l'algèbre, et offraient à l'analyse fonctionnelle naissante l'un de ses premiers succès.

Théorème VIII-20 (théorème de complétude de Riesz–Fischer). *Soit $(X,\mathcal{A},\mu)$ 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 $f_k \longrightarrow f$ 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

$$|f_{k'}| \leq g \quad \mu\text{-presque partout};$$

$$f_{k'}(x) \xrightarrow[k'\to\infty]{} f(x) \quad \text{pour } \mu\text{-presque tout } x.$$

Corollaire VIII-21 (statut des espaces L^p). *Soit $(X,\mathcal{A},\mu)$ 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, et complet.*

Remarque VIII-22. 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 développera dans le chapitre suivant les résultats les plus fondamentaux, et on les appliquera aux espaces de Lebesgue. L'analyse dans les espaces de Fréchet est plus délicate et d'usage moins répandu.

Démonstration du théorème de Riesz–Fischer. Soit $(f_k)_{k\in\mathbb{N}}$ une suite de Cauchy dans $L^p(X,\mu)$. Si l'on démontre l'existence d'une sous-suite convergente, alors toute la suite convergera (c'est une propriété générale des suites de Cauchy dans les espaces métriques).

Par récurrence, on construit une suite extraite, disons $(f_{k_\ell})_{\ell\in\mathbb{N}}$, telle que

$$N_p(f_{k_{\ell+1}} - f_{k_\ell}) \leq 2^{-\ell}.$$

Le problème est de construire une limite à cette suite. Dans la suite, pour alléger la notation, j'écrirai $(f_k)_{k\in\mathbb{N}}$ au lieu de $(f_{k_\ell})_{\ell\in\mathbb{N}}$.

Pour construire la limite, distinguons plusieurs cas.

1. Supposons d'abord $1 \leq p < \infty$. On pose $f_0 = 0$, et, pour tout $x \in X$,

$$g_k(x) = \sum_{j=1}^{k} |f_j(x) - f_{j-1}(x)|, \qquad g(x) = \sum_{j=1}^{\infty} |f_j(x) - f_{j-1}(x)|$$

Par convergence monotone,

$$\int g(x)^p \, d\mu = \lim_{k\to\infty} \int g_k(x)^p \, d\mu;$$

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 $g \in L^p(X,\mu)$. En particulier, il existe un ensemble négligeable N tel que $g(x) < +\infty$ pour tout $x \notin N$.

Pour x hors de N, la série $\sum |f_j(x) - f_{j-1}(x)|$ converge ; la série $\sum (f_j(x) - f_{j-1}(x))$ est donc aussi (absolument) convergente, par complétude de $\mathbb{R}$. On pose

$$f(x) = \sum_{j=1}^{\infty} (f_j(x) - f_{j-1}(x)) = \lim_{n\to\infty} f_n(x).$$

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,

$$|f_n(x) - f(x)| \leq \sum_{k=n}^{\infty} \|f_k - f_{k+1}\|_{L^\infty} \xrightarrow[n\to\infty]{} 0,$$

ce qui prouve la convergence de f_k vers f dans L^∞.

3. Dans le cas où $0 < p < 1$, on pose

$$g_k(x) = \Big(|f_0(x)|^p + \sum_{j=1}^{k} |f_j(x) - f_{j-1}(x)|^p\Big)^{\frac{1}{p}}, \qquad g(x) = \Big(|f_0(x)|^p + \sum_{j=1}^{\infty} |f_j(x) - f_{j-1}(x)|^p\Big)^{\frac{1}{p}}.$$

Par convergence monotone, on a toujours

$$\int |g(x)|^p \, d\mu = \lim_{k\to\infty} \int |g_k(x)|^p \, d\mu;$$

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

$$\lim_{k_0\to\infty} \sum_{k\geq k_0} \int_X 1_{f_k \neq f_{k-1}} \, d\mu = 0,$$

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 complète (c'est le cas de le dire). $\square$

Remarque VIII-23. Dans le cas où $p = 1$, on a retrouvé une variante de la réciproque du théorème de convergence dominée (Théorème IV-24).

VIII-1.6. Produit tensoriel d'espaces de Lebesgue. Soient $(X_1, \mathcal{A}_1, \mu_1)$ et $(X_2, \mathcal{A}_2, \mu_2)$ 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 ;
- ou 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, dont la démonstration est repoussée 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 VIII-24. *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 une démonstration en p. **??**.

VIII-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 VIII-25 (espaces de Lebesgue locaux). *Soit (X, d) un espace métrique. On note $L^p_{\mathrm{loc}}(X)$ l'ensemble des fonctions mesurables $X \to \mathbb{R}$ 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.

VIII-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.

VIII-2.1. Inégalité de Hölder L^p et dualité des normes L^p.

Théorème VIII-26 (inégalité de Hölder dans les espaces L^p). *Soit $(X, \mathcal{A}, \mu)$ 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, soit $k \in \mathbb{N}$, soient $f_1, \ldots, f_k$ des fonctions mesurables, et $p_1, \ldots, p_k$ des exposants dans $[1, +\infty]$, tels que $\sum p_k^{-1} \geq 1$. Alors

$$\|\prod_j f_j\|_{L^r} \leq \prod_j \|f_j\|_{L^{p_j}}, \qquad \text{où} \quad \frac{1}{r} = \sum_{j=1}^k \frac{1}{p_j}.$$

Preuve du Théorème VIII-26. 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 (Théorème IV-83) aux fonctions $|f|^r$ et $|g|^r$, avec les exposants conjugués p/r et q/r (en effet, $(r/p) + (r/q) = 1$). □

Corollaire VIII-27 (convergence de produits). *Soient $(X, \mathcal{A}, \mu)$ 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 $f_n \longrightarrow f$ dans $L^p(X, \mu)$ et $g_n \longrightarrow g$ dans $L^{p'}(X, \mu)$. Alors $f_n g_n$ converge vers fg dans L^1, et en particulier*

$$\int f_n g_n\, d\mu \longrightarrow \int fg\, d\mu.$$

Plus généralement, si l'on se donne k suites de fonctions mesurables $(f_{1,n}), \ldots, (f_{k,n})$ telles que

$$\forall j, \quad f_{j,n} \longrightarrow f_j \qquad \text{dans } L^{p_j}(\mu),$$

avec $\sum p_j^{-1} \leq 1$, alors

$$\prod_j f_{j,n} \xrightarrow[L^r(\mu)]{} \prod_j f_j, \qquad \text{où} \quad \frac{1}{r} = \sum_{j=1}^k \frac{1}{p_j}.$$

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 $f_n \longrightarrow f$ dans L^p, $g_n \longrightarrow g$ dans L^q, et on cherche à montrer que $f_n g_n \longrightarrow fg$ 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 $n \to \infty$. □

On verra 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 VIII-28 (représentation duale des normes L^p). *Soient* (X,μ) *un espace mesuré, et* $p \in [1,\infty[$. *Alors, pour tout* $f \in L^p(\mu)$,

$$\|f\|_{L^p(\mu)} = \sup\left\{\int fg\,d\mu; \qquad \|g\|_{L^{p'}(\mu)} = 1\right\} = \sup_{\|g\|_{L^{p'}(\mu)}\neq 0} \frac{\int fg\,d\mu}{\|g\|_{L^{p'}(\mu)}},$$

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 $(X,\mathcal{A},\mu)$ *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

$$\frac{\int fg\,d\mu}{\|g\|_{L^{p'}(\mu)}} \leq \|f\|_{L^p(\mu)},$$

pour tout $g \in L^{p'}(\mu)$, ce qui implique

$$\sup_{\|g\|_{L^{p'}(\mu)}\neq 0} \frac{\int fg\,d\mu}{\|g\|_{L^{p'}(\mu)}} \leq \|f\|_{L^p(\mu)}.$$

Il suffit donc de montrer que

$$\|f\|_{L^p(\mu)} \leq \sup_{\|g\|_{L^{p'}(\mu)}\neq 0} \frac{\int fg\,d\mu}{\|g\|_{L^{p'}(\mu)}}. \tag{70}$$

Commençons par le cas où $p<\infty$; pour montrer (70) 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 $\|f_k\|_{L^p} = \int f_k g_k\,d\mu$. On écrit, en utilisant l'inégalité de Hölder,

$$\int fg_k\,d\mu = \int f_k g_k\,d\mu + \int (f-f_k)\,g_k\,d\mu \geq \|f_k\|_{L^p} - \|f-f_k\|_{L^p}\|g_k\|_{L^{p'}}.$$

Le premier terme du membre de droite tend vers $\|f\|_{L^p}$, tandis que le second tend vers 0 puisque $f_k \longrightarrow f$ dans L^p. On en déduit que

$$\liminf \int fg_k\,d\mu \geq \|f\|_{L^p},$$

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 (c'est là d'une manière très faible d'utiliser la σ-finitude). On pose alors

$$g := \frac{1_{Z_\varepsilon}\, sign(f)}{\mu[Z_\varepsilon]};$$

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. □

VIII-2.2. Relations d'inclusion. Il faut garder en tête les relations d'inclusion entre espaces de Lebesgue. Limitons 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 VIII-29 (emboîtement décroissant des espaces de Lebesgue). *Soit $(X, \mathcal{A}, \mu)$ 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 :

$$q \geq p \geq 1 \Longrightarrow L^q \subset L^p,$$

et cette injection est continue.

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 l'on vient de voir.

Théorème VIII-30 (emboîtement croissant des espaces de Lebesgue). *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré tel que*

$$\exists \varepsilon > 0; \qquad \forall x \in X, \quad \mu[\{x\}] \geq \varepsilon.$$

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 :

$$q \geq p \geq 1 \Longrightarrow L^p(X, \mu) \subset L^q(X, \mu),$$

et l'injection est continue.

Exemples VIII-31. 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. Quant aux espaces $L^p(\mathbb{R})$, ils ne sont emboîtés ni dans le sens croissant, ni dans le sens décroissant.

Preuve du Théorème VIII-29. 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 $1/q$. □

Preuve du Théorème VIII-30. 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

$$\int f^q \leq \frac{1}{\varepsilon^{\frac{q}{p}-1}} \|f\|_{L^p}^p \|f\|_{L^p}^{q-p},$$

d'où l'on déduit facilement le résultat. □

Même quand ils ne sont pas emboîtés, les espaces de Lebesgue sont "*en interpolation*", l'espace de Lebesgue L^p est "entre" L^{p_0} et L^{p_1} dès que p est entre p_0 et p_1 :

Théorème VIII-32 (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 $f : X \to \overline{\mathbb{R}}$,*

$$\|f\|_{L^q} \leq \|f\|_{L^p}^\theta \|f\|_{L^r}^{1-\theta}, \tag{71}$$

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 (71) 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 à la lectrice. 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 VIII-19, on vérifie facilement que $\|f_k - f\|_{L^q} \longrightarrow 0$ quand $k \to \infty$. □

Remarque VIII-33. Je présenterai 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 VIII-34 (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.*

VIII-2.3. Continuité de la norme en p. On a 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.

Théorème VIII-35 (continuité de la norme L^p en p). *Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, et $f : X \to \mathbb{R}$ une fonction mesurable. Soit*

$$J := \Big\{p \in [0,\infty];\ N_p(f) < +\infty\Big\}.$$

Alors J est un intervalle (éventuellement vide) et $p \longmapsto N_p(f)$ est continue sur l'adhérence de J (à valeurs dans $[0,+\infty]$).

Remarques VIII-36. (i) Remarquons que $p \longmapsto N_p(f)$ n'est en général pas continue sur $[0,+\infty]$ tout entier ; c'est pourquoi on se restreint 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 $p \to \infty$. 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 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

$$N_p(f) \xrightarrow[p\to\infty]{} \|f\|_{L^\infty}.$$

Cet énoncé est 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 (71) 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, on va 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 VIII-32. 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 $p \longmapsto X^{\min(1,1/p)}$ pour $p \in]0,+\infty[$, il suffit de prouver que pour toute suite p_k convergeant vers p, $f \in L^{p_k}$,

$$\int |f|^{p_k}\, d\mu \xrightarrow[k\to\infty]{} \int |f|^p\, d\mu.$$

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

$$\int_{|f|\geq 1} |f|^{p_k} \longrightarrow \int_{|f|\geq 1} |f|^p; \qquad \int_{|f|<1} |f|^{p_k} \longrightarrow \int_{|f|<1} |f|^p.$$

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

$$\|f\|_{L^q} = \||f|^q\|_{L^1}^{1/q}; \qquad \||f|^q\|_{L^\infty} = \|f\|_{L^\infty}^q,$$

on voit que l'on peut remplacer le problème sur f par le problème sur $|f|^q$, et que l'on peut donc supposer sans perte de généralité

$$f \in L^1; \qquad f \geq 0.$$

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 Tchebychev,

$$\|f\|_{L^p} \geq [a(K)M^p]^{1/p} = Ma(K)^{1/p} \xrightarrow[p\to\infty]{} M.$$

En faisant tendre M vers $\|f\|_{L^\infty}$, on en déduit (que $\|f\|_{L^\infty}$ soit fini ou non)

$$\liminf_{p\to\infty} \|f\|_{L^p} \geq \|f\|_{L^\infty}.$$

Si $\|f\|_{L^\infty} = +\infty$, ceci achève la preuve.

6. Supposons maintenant que $\|f\|_{L^\infty} < +\infty$. Comme $f \in L^1$, on peut 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

$$\limsup_{p\to\infty} \|f\|_{L^p} \leq \|f\|_{L^\infty},$$

ce qui conclut la preuve. □

VIII-2.4. Interpolation entre espaces de Lebesgue. Le Théorème VIII-32 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

$$\|T\|_{E\to F} := \sup_{\|x\|_E \neq 0} \frac{\|Tx\|_F}{\|x\|_E}.$$

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 VIII-37 (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 $L^{q_0}(X)$, 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*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad \frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

En outre, si l'on pose $M_\theta = \|T\|_{L^{p_\theta}\to L^{q_\theta}}$, 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 $\log \|T\|_{L^p\to L^q}$ 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 VIII-38 (Théorème d'interpolation de Marcinkiewicz). *Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ 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*

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

En outre, si l'on note $M_0 = \|T\|_{L^{p_0}\to L^{q_0,\infty}}$, $M_1 = \|T\|_{L^{p_1}\to L^{q_1,\infty}}$, $M_\theta = \|T\|_{L^{p_\theta}\to L^{q_\theta}}$, 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 $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ 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 $p \in [1, +\infty]$,*

$$\|T\|_{L^p \to L^p} \leq \frac{Cp}{p-1} \|T\|^{1/p}_{L^1 \to L^{1,\infty}} \|T\|^{1-1/p}_{L^\infty \to L^\infty}.$$

Parlons maintenant des démonstrations de ces théorèmes. C'est Marcel Riesz, le petit frère de Frigyes 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, dans une forme un peu plus restrictive. Vers la fin des années 1930, le mathématicien suédois Olof Thorin mit au point la preuve esquissée ci-après, basée sur l'analyse complexe et devenue très populaire. À peu près au même moment, Józef Marcinkiewicz démontrait le théorème qui porte son nom par des méthodes très différentes.

Un outil-clé dans la preuve 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 VIII-39 (Lemme des trois droites). *Soit $S := \{x + iy;\ x \in [0, 1];\ y \in \mathbb{R}\} \subset \mathbb{C}$ une bande du plan complexe, et soit $f : S \to \mathbb{C}$ une fonction continue bornée, holomorphe dans l'intérieur de S. Alors,*

(i) $\sup_S |f| = \sup_{\partial S} |f|$ *;*

(ii) soit $M_\theta := \sup_{y \in \mathbb{R}} |f(\theta + iy)|$; 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 $M_j = \|T\|_{L^{p_j}\to L^{q_j}}$. On va utiliser le Théorème VIII-28, sous la forme

$$\|f\|_{L^q(\mu)} = \sup_{\|g\|_{L^{q'}}\neq 0} \frac{\int fg\,d\mu}{\|g\|_{L^{q'}(\mu)}},$$

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 $L^p \to L^q$ avec norme au plus $M_1^\theta M_0^{1-\theta}$ revient à prouver que

$$\|Tf\|_{L^q} \le M_1^\theta M_0^{1-\theta} \|f\|_{L^p} \tag{72}$$

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 (72) dans le cas où f est une fonction simple. Notre but est donc

$$\left|\int (Tf)g\right| \le M_1^\theta M_0^{1-\theta}\|f\|_{L^p}\|g\|_{L^{q'}}. \tag{73}$$

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

$$\varphi : z \longmapsto \int (Tf_z)g_z$$

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

$$\varphi(z) = \sum_{1\le k\le K} a_k^{\lambda_k z+\mu_k}, \qquad \lambda \in \mathbb{R},\ \mu \in \mathbb{R};$$

en particulier φ est holomorphe et bornée dans S, et on peut appliquer le lemme des trois droites :

$$|\varphi(\theta)| \le \left(\sup_{t\in\mathbb{R}} |\varphi(it)|^{1-\theta}\right)\left(\sup_{t\in\mathbb{R}} |\varphi(1+it)|^{1-\theta}\right).$$

Mais $\varphi(\theta)$ n'est autre que $\int Tfg$. Par ailleurs,

$$\left|\int Tf_{it}\,g_{it}\right| \le \|Tf_{it}\|_{L^{q_0}}\|g_{it}\|_{L^{q_0'}} \le \|T\|_{L^{p_0}\to L^{q_0}}\|f\|_{L^p}^{p/p_0}\|g\|_{L^{q'}}^{q'/q_0'},$$

et l'on peut faire une majoration similaire pour les $z = 1 + it$. La conclusion en découle facilement. $\square$

Démonstration du théorème de Marcinkiewicz. On se contentera de démontrer le "cas particulier", qui est utile dans de nombreuses situations. La lectrice 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 on va 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*.

Écrivons donc
$$\|Tf\|_{L^\infty} \le M_1 \|f\|_{L^\infty}, \qquad \|Tf\|_{L^{1,\infty}} \le M_0 \|f\|_{L^1}.$$
La deuxième inégalité se réécrit
$$\forall t > 0, \quad t\,\mu[\{|Tf| > t\}] \le M_0 \|f\|_{L^1}.$$
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 (25) :
$$\int |f|^p = p \int_0^{+\infty} \mu[\{|f| > t\}] t^{p-1}\, dt.$$
De même,
$$\int |Tf|^p = p \int_0^{+\infty} \mu[\{|Tf| > t\}] t^{p-1}\, dt.$$
Pour tout $t \ge 0$ on écrit alors
$$f = f_1^{(t)} + f_2^{(t)}, \qquad f_1^{(t)} = f 1_{|f| \le At}, \quad f_2^{(t)} = f_2 1_{|f| > At}.$$
La borne $L^\infty \to L^\infty$ entraîne que pour tout $t \ge 0$,
$$|Tf_1^{(t)}| \le M_1 A t.$$
En particulier,
$$\mu[\{|Tf| > t\}] \le \mu[\{|Tf_2^{(t)}| > (1 - M_1 A)t\}].$$
En reportant cette inégalité dans la représentation de $\int |Tf|^p$, on trouve
$$\int |Tf|^p \le p \int_0^{+\infty} \mu[\{|Tf_2^{(t)} > (1 - M_1 A)t\}] t^{p-1}\, dt$$
$$= (1 - M_1 A)^{-1} p \int_0^{+\infty} \Big((1 - M_1 A) t \mu[\{|Tf_2^{(t)} > (1 - M_1 A)t\}] \Big) t^{p-2}\, dt$$
$$\le (1 - M_1 A)^{-1} p M_0 \int_0^{+\infty} \|f_2^{(t)}\|_{L^1} t^{p-2}\, dt = (1 - M_1 A)^{-1} p M_0 \int_0^{+\infty} \int |f| 1_{|f| > At}\, t^{p-2}\, dt.$$
On applique alors Fubini et un changement de variable évident pour réécrire le dernier terme sous la forme
$$(1 - M_1 A)^{-1} p M_0 \int |f| \left(\int_0^{|f|/A} t^{p-2}\, dt \right)$$

$$= \frac{pM_0}{(p-1)(1-M_1A)A^{p-1}} \int |f|^p.$$

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. □

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 $z \longmapsto T_z f$ est holomorphe pour tout f simple. On peut alors changer, dans l'énoncé du Théorème de Riesz–Thorin, l'opérateur T en une famille holomorphes d'opérateurs T_z ; l'hypothèse de bornes $L^{p_0} \to L^{q_0}$ et $L^{p_1} \to L^{q_1}$ sur T est alors remplacée par une hypothèse similaire sur T_0 et T_1 respectivement.

Théorème VIII-40 (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 $L^{p_0}(X) \to L^{p_1}(Y)$, et $L^{q_0}(X) \to L^{q_1}(Y)$. Alors,*

$$T_\theta \text{ est borné } L^p(X) \to L^q(Y),$$

En outre, si on pose $M_\theta = \|T_z\|_{L^{p_\theta}\to L^{q_\theta}}$, alors

$$M_\theta \leq M_0^{1-\theta} M_1^\theta.$$

La démonstration de cet énoncé dû à Elias Stein (élève d'Antoni Zygmund, comme Marcienkiewicz) est similaire à celle du théorème de Riesz–Thorin.

Exemple VIII-41. Soit μ une mesure et w une fonction positive ; la famille d'opérateurs

$$T_z : f \longmapsto w^z f$$

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

$$\|f\|_{L^p_\kappa} = \|f v^\kappa\|_{L^p},$$

alors on a, pour tout opérateur linéaire S,

$$\|S\|_{L^{p_\theta}_{\kappa_\theta}\to L^{q_\theta}_{\lambda_\theta}} \leq \|S\|^{1-\theta}_{L^{p_0}_{\kappa_0}\to L^{q_0}_{\lambda_0}} \|S\|^{\theta}_{L^{p_1}_{\kappa_1}\to L^{q_1}_{\lambda_1}}.$$

VIII-3*Espace des fonctions mesurables

On va maintenant introduire une notion naturelle de convergence des fonctions mesurables, ne présupposant aucune intégrabilité, et étudier ses liens avec la convergence L^p. Sans être particulièrement difficile, cette section contient des notions d'usage beaucoup moins fréquent que les autres de ce chapitre, et pourra être omise en première lecture.

VIII-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, 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 (exercice). Or on a vu (Exemple IV-23(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 VIII-42 (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 $f_n : X \to Y$ 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, on l'appellera "convergence dans L".

Proposition VIII-43 (convergence dans L et convergence en mesure). *Soient $(X, \mathcal{A}, \mu)$ un espace mesuré $\sigma-fini$, (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

$$\forall \varepsilon > 0, \quad \mu\big[\{x \in A;\ d(f_n(x), f(x)) \geq \varepsilon\}\big] \xrightarrow[n\to\infty]{} 0.$$

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.

Remarque VIII-44. 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

$$\varphi_k = \sum_{k\geq 1} \frac{1_{A_k}}{k^2\mu[A_k]}.$$

Définition VIII-45 (convergence en probabilité). *Dans $(X, \mathcal{A}, \mu)$ un espace de probabilité, la convergence en mesure est aussi appelée convergence en probabilité.*

En mots, la convergence en probabilité dit que la probabilité de dévier sensiblement de la limite tend vers 0.

Preuve de la Proposition VIII-43. Je donnerai 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 $\Delta(f_{n''}, f) \longrightarrow 0$. Comme la sous-suite extraite $f_{n'}$ était arbitraire, et que Δ définit une métrique, on en déduit que $\Delta(f_n, f) \longrightarrow 0$.

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

$$\int_{B_{n,\varepsilon}} \varphi\, d\mu \xrightarrow[n\to\infty]{} 0.$$

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 $n \to \infty$; comme η est arbitrairement petit, on conclut que

$$\mu[B_{n,\varepsilon} \cap A] \xrightarrow[n\to\infty]{} 0,$$

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 $\varepsilon \to 0$, et le deuxième tend vers 0 quand $n \to \infty$, ε étant fixé. On en déduit que $\Delta(f_n, f) \longrightarrow 0$.

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 $(\mathbb{R}, \delta)$). Par convergence dominée, on montre alors que

$$\int_{A_k} \delta(f_m(x), f_n(x)) \, d\mu(x) \xrightarrow[n\to\infty]{} \int_{A_k} \delta(f_m(x), f(x)) \, d\mu(x).$$

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

$$\int_{A_k} \delta(f_m(x), f(x)) \, d\mu(x) \xrightarrow[m\to\infty]{} 0,$$

ce qui est bien sûr équivalent à

$$\int_{A_k} \varphi \, \delta(f_m(x), f(x)) \, d\mu(x) \xrightarrow[m\to\infty]{} 0.$$

On a donc, pour tout k_0,

$$\sum_{k\leq k_0} \int \varphi \, \delta(f_m(x), f(x)) \, d\mu(x) \xrightarrow[m\to\infty]{} 0;$$

et d'autre part, puisque $\varphi \in L^1(d\mu)$,

$$\sum_{k>k_0} \int \varphi \, \delta(f_m(x), f(x)) \, d\mu(x) \leq \sum_{k>k_0} \int \varphi \xrightarrow[k_0\to\infty]{} 0.$$

On conclut que

$$\int \varphi \, \delta(f_m(x), f(x)) \, d\mu(x) \xrightarrow[m\to\infty]{} 0.$$

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

$$A_k := \big\{x \in X;\ |f(x)| \leq k \text{ et } \varphi(x) \geq k^{-1}\big\}.$$

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 $\nu[X \setminus A_k] \longrightarrow 0$. 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. $\square$

VIII-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, que je limiterai aux fonctions à valeurs réelles.

Théorème VIII-46 (convergence dans L et dans L^p). *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables à valeurs rélles sur un espace mesuré (X,μ) ; soit également f une fonction mesurable à valeurs réelles. Alors*

(i) Si $f_n \longrightarrow f$ dans $L^p(X,\mu)$ ($0 \leq p \leq \infty$), alors $f_n \longrightarrow f$ dans $L(X,\,\mu)$.

(ii) Si $f_n \longrightarrow f$ dans $L(X,\mu)$ et qu'il existe $g \in L^p(\mu)$ ($0 < p < \infty$) tel que $|f_n| \leq g$ pour tout n, alors $f_n \longrightarrow f$ dans $L^p(X,\,\mu)$.

Démonstration. L'assertion (i) est facile : pour toute sous-suite extraite n', on a $f_{n'} \longrightarrow f$ 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 IV-4.5.

Théorème VIII-47. *Soient $(X,\mathcal{A},\mu)$ 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 VIII-48. Dans le chapitre suivant, on retrouvera le cas particulier $p = 1$ de ce théorème comme une conséquence du théorème de Schur.

Démonstration du Théorème VIII-47. 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

$$\int |f|^p \, d\mu \leq \liminf_{n\to\infty} \int |f_n|^p < +\infty.$$

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 $\Delta(f_n, f) \longrightarrow 0$, 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. □

VIII-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.

VIII-4.1. Mesures signées. Commençons par un rappel (Proposition IV-7). Si $(X, \mathcal{A}, \mu)$ 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 VIII-49 (mesures étrangères). *Soit $(X, \mathcal{A})$ 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 $A \cap B = \emptyset$, $\mu[X \setminus A] = 0$, $\nu[X \setminus B] = 0$.*

On peut maintenant définir la notion de mesure signée, comme une généralisation du concept de fonction mesurable :

Définition VIII-50 (mesure signée). *Soit $(X, \mathcal{A})$ 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 VIII-51. 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 VIII-52. 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 $\mu[\mathbb{R}]$.

VIII-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 $\mathcal{A} \to [-\infty, +\infty]$, 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 VIII-53 (σ-additivité à valeurs réelles). *Soit $\mathcal{A}$ une σ-algèbre de parties d'un ensemble X, et soit $\mu : \mathcal{A} \to \mathbb{R}$ une fonction d'ensembles. 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*

$$\mu[\cup A_k] = \sum_{k\in\mathbb{N}} \mu[A_k], \tag{74}$$

où le membre de droite est défini comme la limite des sommes partielles $\sum_{1\leq k\leq \ell} \mu[A_k]$.

Remarque VIII-54. Le membre de gauche de (74) 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 VIII-55 (théorème de décomposition de Hahn). *Soit $(X, \mathcal{A})$ un espace mesurable ; alors on peut identifier*

- d'une part, les fonctions $\mu : \mathcal{A} \to \mathbb{R}$, 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 $A \in \mathcal{A}$,

$$|\mu|[A] = \mu_+[A] + \mu_-[A] = \sup\left\{\sum_{j\in\mathbb{N}} |\mu[A_j]|; \quad A_j \in \mathcal{A},\ (A_j)_{j\in\mathbb{N}} \text{ partition de } A\right\}. \tag{75}$$

Remarque VIII-56. 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”. Bien noter que l'énoncé contient l'*unicité* de cette décomposition. Ce résultat est dû au mathématicien autrichien Hans Hahn, grand spécialiste d'analyse fonctionnelle, très actif dans l'entre-deux guerres.

Remarque VIII-57. La conclusion du Théorème VIII-55 est bien sûr en défaut pour des fonctions σ-additives $\mathcal{A} \to \overline{\mathbb{R}}$ ou même $\mathcal{A} \to [0, +\infty]$ (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 $\nu_\pm$. 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 (75) : $|\mu|[A] = \sup \sum |\mu[A_i]|$, 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 $(A^k_j)_{j\in\mathbb{N}}$ de chaque A^k, on définit automatiquement une partition $(A^k_j)_{k,j\in\mathbb{N}}$ 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 $A^k_j = A^k \cap A_j$, de sorte que $(A^k_j)_{j\in\mathbb{N}}$ constitue une partition de A^k. Par σ-additivité de μ,

$$\sum_{j\in\mathbb{N}} |\mu[A_j]| = \sum_{j\in\mathbb{N}} \Big|\sum_{k\in\mathbb{N}} \mu[A^k_j]\Big| \leq \sum_{j\in\mathbb{N}}\sum_{k\in\mathbb{N}} |\mu[A^k_j]| = \sum_{k\in\mathbb{N}}\sum_{j\in\mathbb{N}} |\mu[A^k_j]| \leq \sum_{k\in\mathbb{N}} |\mu|[A_k];$$

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

$$\Big|\sum_{j\in\mathcal{J}} \mu[A_j]\Big| \geq |\mu[A]| + 1.$$

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 $|\mu|[A_1] = +\infty$ 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
$$\mu[\bigcup A_k] = \lim_{k\to\infty} \mu[A_k].$$
Enfin, en passant au complémentaire et en utilisant $|\mu[X]| < +\infty$, on voit que pour toute famille (A_k) décroissante,
$$\mu[\bigcap A_k] = \lim_{k\to\infty} \mu[A_k].$$

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 supposerons 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 $\mu[A] = \lim_{\ell\to\infty} \mu[C_\ell]$. 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 $m \geq \ell$,
$$\mu[A_\ell \cup \ldots \cup A_m] \geq \mu[A_\ell] - \sum_{k=\ell}^{m} 2^{-k}M \geq \mu[A_\ell] - 2^{-(\ell-1)}M.$$
En passant à la limite quand $m \to \infty$, 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 (75), 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

$$\mu_+[A] + \mu_-[A] = |\mu[A \cap S_+]| + |\mu[A \cap S_-]| \leq |\mu|[A].$$

$\square$

Remarque VIII-58. La fin de la preuve montre a posteriori que dans le membre de droite de la formule (75), 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 VIII-59 (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

$$|\mu[A] - \mu[K]| \leq \varepsilon, \qquad |\mu[O] - \mu[A]| \leq \varepsilon. \tag{76}$$

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 $|\mu|[A \setminus K_n] \longrightarrow 0$; et une suite décroissante d'ouverts $(O_n)_{n\in\mathbb{N}}$ contenant A, telle que $|\mu|[O_n \setminus A] \longrightarrow 0$. Alors

$$\big|\mu[O_n] - \mu[A]\big| = |\mu[O_n \setminus A]| \leq |\mu|[O_n \setminus A] \longrightarrow 0,$$

donc $\mu[O_n] \to \mu[A]$, et de même $\mu[K_n] \to \mu[A]$. La propriété (iii) est donc vérifiée.

Supposons maintenant (iii), et prouvons par exemple que μ_+ est régulière. Soit A un ensemble mesurable quelconque, et $A' = A \cap S_+$. Par (76) on peut trouver une suite de compacts $(K_n)_{n\in\mathbb{N}}$, inclus dans A' et donc dans A, tels que

$$\mu_+[K_n] = \mu[K_n] \longrightarrow \mu[A'] = \mu_+[A'] = \mu_+[A];$$

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 $\mu[L_n] = \mu_+[L_n] \longrightarrow \mu_+[B] = \mu[B]$. L'ouvert $O_n = X \setminus L_n$ contient alors A, et on a

$$\mu_+[O_n] = \mu_+[X] \setminus \mu_+[L_n] \longrightarrow \mu_+[X] \setminus \mu_+[B] = \mu_+[X \setminus B] = \mu_+[A \cup S_-] = \mu_+[A].$$

□

Le corollaire suivant est une conséquence immédiate du Théorème VIII-55 et de la Proposition VIII-59 :

Corollaire VIII-60 (Théorème de Hahn pour les mesures régulières). *Soit $(X, \mathcal{A})$ un espace mesurable ; alors on peut identifier*

- d'une part, les fonctions $\mu : \mathcal{A} \to \mathbb{R}$, 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

$$\lim_{n\to\infty} \mu[K_n] = \lim_{n\to\infty} \mu[O_n] = \mu[A];$$

- 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.

VIII-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 VIII-60 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 VIII-61 (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*

$$\mu \le \nu \Longrightarrow \quad \mu_+ \le \nu_+,\ \mu_- \ge \nu_-;$$
$$\forall \alpha \ge 0,\ (\alpha\mu)_\pm = \alpha\mu_\pm; \qquad \forall \alpha < 0,\ (\alpha\mu)_\pm = |\alpha|\mu_\mp;$$
$$(-\mu)_+ = \mu_-; \qquad |-\mu| = |\mu|;$$
$$(\mu+\nu)_\pm \le \mu_\pm + \nu_\pm; \qquad |\mu+\nu| \le |\mu| + |\nu|.$$

Pour mesurer la taille d'une mesure signée, un concept naturel est fourni par la **variation totale** :

Définition VIII-62 (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 $\|\mu\|_{\mathrm{VT}(X)}$ ou simplement $\|\mu\|_{\mathrm{VT}}$, la quantité positive*

$$|\mu|[X] = \mu_+[X] + \mu_-[X].$$

Si A est une partie mesurable de X, on notera $\|\mu\|_{\mathrm{VT}(A)} = |\mu|[A]$.

Proposition VIII-63 (propriétés de la variation totale). *Soient X un espace mesurable et μ une mesure signée sur X. Alors*

(i) $A \longmapsto \|\mu\|_{VT(A)}$ est une fonction σ-additive d'ensembles, qui coïncide avec $|\mu|$;

(ii) Pour toute partie mesurable A de X,

$$|\mu[A]| \leq \|\mu\|_{\mathrm{VT}}.$$

Plus généralement, pour toutes parties disjointes $(A_k)_{k\in\mathbb{N}}$,

$$\sum_{k\in\mathbb{N}} |\mu[A_k]| \leq \|\mu\|_{\mathrm{VT}}.$$

(iii) $\|\mu\|_{\mathrm{VT}} = \sup_{|h|\leq 1} \int h\, d\mu$, 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) $\|\mu\|_{\mathrm{VT}} = \inf\Big\{\nu_+[X] + \nu_-[X];\quad \mu = \nu^+ - \nu^-\Big\}$, 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 $\mu_\pm$ 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 $\int h\, d\mu \leq \|\mu\|_{\mathrm{VT}}$. 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

$$\mu = \nu^+ - \nu^- \Longrightarrow \nu^+[S_\pm] + \nu^-[S_\pm] \geq \mu_\pm[S_\pm].$$

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 difficuté. □

Décrivons maintenant de manière un peu plus précise l'espace des mesures signées finies :

Théorème VIII-64 (espace des mesures signées). *Soit $(X, \mathcal{A})$ 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 $f \longmapsto f\nu$: en particulier,*

$$\|f\|_{L^1(\nu)} = \|f\nu\|_{\mathrm{VT}}.$$

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 $M_{\mathrm{reg}}(X)$.

Remarque VIII-65. Si X est polonais, $M_{\mathrm{reg}}(X) = M(X)$ en vertu du Théorème II-62.

Démonstration. 1. Il est facile de vérifier que la variation totale définit bien une norme, en utilisant la Proposition VIII-61.

2. Montrons maintenant que $M(X)$ est complet : soit $(\mu_k)_{k\in\mathbb{N}}$ une famille de mesures signées finies telles que

$$\|\mu_k - \mu_\ell\|_{\mathrm{VT}} \xrightarrow[k,\ell\to\infty]{} 0.$$

Pour toute partie A mesurable, on a, d'après la Proposition VIII-63(i),

$$|\mu_k[A] - \mu_\ell[A]| \leq \|\mu_k - \mu_\ell\|_{\mathrm{VT}} \xrightarrow[k,\ell\to\infty]{} 0.$$

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 :

$$\mu_k[\bigcup A_j] = \sum_j \mu_k[A_j].$$

On peut passer à la limite quand $\ell \to \infty$ dans le premier terme ; pour passer à la limite dans le deuxième, et donc prouver la σ-additivité de μ, il suffit d'établir

$$\sum_j |\mu_\ell[A_j] - \mu[A_j]| \xrightarrow[\ell\to\infty]{} 0.$$

Mais, les A_j étant disjoints, on a, pour tout $k \geq \ell$, grâce à la Proposition VIII-63(ii),

$$\sum_j |\mu_\ell[A_j] - \mu_k[A_j]| \leq \|\mu_\ell - \mu_k\|_{\mathrm{VT}},$$

et le membre de droite converge vers 0 quand $\ell \to \infty$, 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 $k \to \infty$. Pour prouver la complétude, il reste à montrer que $\|\mu_k - \mu\|_{\mathrm{VT}}$ tend vers 0. Soit h une fonction mesurable valant ± 1 sur X, et $\varepsilon(k) := \sup_{\ell \geq k} \|\mu_k - \mu_\ell\|_{\mathrm{VT}}$. On a, d'après la Proposition VIII-63(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 $\ell \to \infty$, 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 VIII-63(ii) encore, on conclut que $\|\mu_k - \mu\|_{\mathrm{VT}} \leq \varepsilon(k)$, ce qui conclut l'argument.

3. Vérifions maintenant l'identité

$$\|f\|_{L^1(d\nu)} = \|f\nu\|_{\mathrm{VT}}$$

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

$$\|f\nu\|_{\mathrm{VT}} = f_+\nu[X] + f_-\nu[X] = \int f_+\,d\nu + \int f_-\,d\nu = \int (f_+ + f_-)\,d\nu = \int |f|\,d\nu.$$

4. Comme on l'a déjà remarqué, le Corollaire VIII-60 montre que $M_{\mathrm{reg}}(X)$ 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

$$|\mu[O \setminus A]| \leq \|\mu_k - \mu\|_{\mathrm{VT}} + |\mu_k[O \setminus A]| \leq \varepsilon,$$

et de même

$$|\mu[A \setminus K]| \leq \|\mu_k - \mu\|_{\mathrm{VT}} - |\mu_k[A \setminus K]| \leq \varepsilon.$$

La propriété (iii) de la Proposition VIII-59 est donc satisfaite, ce qui prouve la régularité de μ. □

VIII-4.4. Théorème de Riesz pour les mesures signées. Comme nous l'avons vu au chapitre III, 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 VIII-66 (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 les formules, valables pour toute fonction $f \in C_0(X)$ et tout ouvert O,

$$\Lambda f = \int f\,d\mu := \int f\,d\mu_+ - \int f\,d\mu_-$$

et

$$\mu[O] = \sup\big\{\Lambda f;\ f \geq 0,\ f \in C_0(X),\ \mathrm{Spt}\,(f) \subset O\big\} \\ + \inf\big\{\Lambda f;\ f \leq 0,\ f \in C_0(X),\ \mathrm{Spt}\,(f) \subset O\big\}.$$

En bref,

$$C_0(X)^* = M_{\mathrm{reg}}(X),$$

où $C_0(X)$ est muni de la norme uniforme et $M_{\mathrm{reg}}(X)$ de la norme de la variation totale. En particulier

$$\|\mu\|_{\mathrm{VT}} = \sup\left\{\int_X f\,d\mu; \quad |f| \leq 1,\ f \in C_c(X)\right\} \\ = \sup\left\{\int_X f\,d\mu; \quad |f| \leq 1,\ f \in C_0(X)\right\}.$$

Remarque VIII-67. 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 $C_b(X)$, et il est en fait impossible d'identifier $C_b(X)^*$ à $M(X)$. En effet, sous hypothèse de l'axiome du choix, on peut identifier $C_b(X)^*$ à l'espace des fonctions d'ensemble *finiment additives* régulières, et montrer si X est non compact que cet espace est *strictement plus grand* que $M(X)$ [Dunford–Schwarz, IV.6].

Exemple VIII-68. Fixons un espace localement compact non compact, vérifiant aux hypothèses du Théorème II-64 (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 $\lim_\infty$ (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 II-60, en fait L est “concentrée à l'infini”.

Preuve du Théorème VIII-66. 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 $f \longmapsto \int f\,d\mu$ 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

$$\Phi(f) := \sup\Big\{\langle\Lambda, h\rangle;\ h \in C_c(X);\ 0 \leq h \leq f\Big\}.$$

La fonctionnelle Φ, définie sur l'ensemble des fonctions continues positives à support compact, est positive et croissante ($f \leq g \Longrightarrow \Phi(f) \leq \Phi(g)$) ; 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 $i = 1, 2$,

$$0 \le h_i \le f_i, \qquad \Lambda h_i \ge \Phi(f_i) - \varepsilon.$$

Alors $h := h_1 + h_2$ est une fonction continue à support compact telle que $0 \le h \le f_1 + f_2$, et on a

$$\Lambda h = \Lambda h_1 + \Lambda h_2 \ge \Phi(f_1) + \Phi(f_2) - 2\varepsilon.$$

En passant au supremum sur tous les h admissibles, on obtient

$$\Phi(f_1 + f_2) \ge \Phi(f_1) + \Phi(f_2) - 2\varepsilon.$$

En faisant finalement tendre ε vers 0, on conclut à la sur-additivité de Φ. On peut alors appliquer la Remarque III-64 (v) suivant l'énoncé du Théorème de Riesz III-63 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 $0 \le h \le f$, alors bien sûr $\|h\|_\infty \le \|f\|_\infty$, et par continuité de Λ il existe $C > 0$, indépendant de f et h, tel que $\Lambda h \le C\|f\|_\infty$. En passant au supremum, on obtient $\Phi(f) \le C\|f\|_\infty$, soit

$$\int f\, d\mu_+ \le C\|f\|_\infty.$$

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

$$\mu_+[K] \le C.$$

Par ailleurs, X étant ouvert et μ_+ étant pré-régulière, on a

$$\mu_+[X] = \sup \Big\{ \mu_+[K];\ K \text{ compact} \Big\};$$

ce qui prouve $\mu_+[X] \le C$. On conclut que μ_+ est finie. Par la Remarque III-64 (iii) suivant l'énoncé du Théorème III-63 (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

$$\Lambda \le \mu_+.$$

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 pré-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

$$\begin{aligned} \|\mu\|_{\mathrm{VT}} &\ge \sup \left\{ \int_X f\, d\mu; \quad |f| \le 1,\ f \in C_0(X) \right\} \\ &\ge \sup \left\{ \int_X f\, d\mu; \quad |f| \le 1,\ f \in C_c(X) \right\}. \end{aligned}$$

Pour conclure la démonstration, il suffit donc d'établir

$$\|\mu\|_{\mathrm{VT}} \le \sup \left\{ \int_X f\, d\mu; \quad |f| \le 1,\ f \in C_c(X) \right\}.$$

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 II-3.3). 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

$$\begin{aligned}\int f\,d\mu &\geq \mu_+[K_+] + \mu_-[K_-] - \|\mu\|_{\mathrm{VT}}(X \setminus (K_+ \cup K_-))\\ &\geq (\mu_+[X] - \varepsilon) + (\mu_-[X] - \varepsilon) - 2\varepsilon = \|\mu\|_{\mathrm{VT}} - 4\varepsilon.\end{aligned}$$

On conclut en faisant tendre ε vers 0. □

VIII-4.5. Représentation duale de la variation totale. Comme corollaire du Théorème VIII-66 (Théorème de Riesz pour les mesures signées), nous avons obtenu une représentation duale de la variation totale :

$$\|\mu\|_{\mathrm{VT}} = \sup\left\{\int_X f\,d\mu; \quad |f| \leq 1,\ f \in C_0(X)\right\} \tag{77}$$

$$= \sup\left\{\int_X f\,d\mu; \quad |f| \leq 1,\ f \in C_c(X)\right\}. \tag{78}$$

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 VIII-69 (représentation faible de la variation totale). *Soit X un espace métrique et μ une mesure (de Borel) signée sur X, régulière. Alors*

$$\|\mu\|_{\mathrm{VT}} = \sup\left\{\int_X f\,d\mu; \quad |f| \leq 1,\ f \in C_b(X)\right\}.$$

En particulier, cette formule est automatiquement vérifiée si μ est finie et X est un espace polonais.

Remarque VIII-70. En dépit de cette proposition, le dual de $C_b(X)$ est a priori plus gros que $M_{\mathrm{reg}}(X)$.

Démonstration. La première partie de l'énoncé implique la deuxième puisque toute mesure finie sur un espace polonais est régulière (Théorème II-62). D'autre part il est clair que $\int \varphi\,d\mu \leq \|\mu\|_{\mathrm{VT}}$ pour tout φ continu à valeurs dans $[-1,1]$; il suffit donc de prouver que $\|\mu\|_{\mathrm{VT}} \leq \sup\{\int \varphi\,d\mu\}$, où le supremum est pris sur les fonctions continues à valeurs dans $[-1,1]$.

Soient S_+ et S_- des ensembles disjoints tels que $\|\mu\|_{\mathrm{VT}} = \mu_+[S_+] + \mu_-[S_-]$. 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

$$\|\mu\|_{\mathrm{VT}} \leq \mu_+[K_+] + \mu_-[K_-] + \varepsilon;$$

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

$$\mu_+[K_+] = \|\mu_+\|_{\mathrm{VT}(K_+)} = \sup\Big\{\int \varphi\, d\mu_+;\ \varphi \in C(K_+),\ \|\varphi\| \le 1\Big\}.$$

On peut donc trouver $\varphi_\pm$ (continue sur $K_\pm$ et à valeurs dans $[-1,1]$) tels que

$$\mu_\pm[K_\pm] \le \int_{K_\pm} \varphi_\pm\, d\mu_\pm + \varepsilon.$$

(Quitte à remplacer $\varphi_\pm$ 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

$$\|\mu\|_{\mathrm{VT}} \le \int_{K_+\cup K_-} \varphi\, d\mu + 3\varepsilon.$$

Par le Théorème d'extension de Tietze–Urysohn (rappelé dans la sous-section II-3.3), on peut prolonger φ en une fonction continue sur X, toujours notée φ, à valeurs dans $[-1,1]$. On a alors

$$\int_{K_+\cup K_-} \varphi\, d\mu \le \int_X \varphi\, d\mu + \|\mu\|_{\mathrm{VT}(X\setminus(K_+\cup K_-))} \le \int_X \varphi\, d\mu + \varepsilon.$$

On conclut que

$$\|\mu\|_{\mathrm{VT}} \le \int_X \varphi\, d\mu + 4\varepsilon,$$

et on achève l'argument en faisant tendre ε vers 0. □

VIII-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. Elles sont nommées en hommage au mathématicien autrichien Johannes Radon, pionnier de la théorie de la mesure abstraite (et découvreur de la "transformée de Radon" très populaire dans les technologies d'imagerie). Avant d'introduire cette classe de mesures, notons que sa définition même varie de manière assez importante d'un auteur à l'autre.

Définition VIII-71 (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 $M_{\mathrm{loc}}(\Omega)$ 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, en englobant tous les supports compacts possibles par une suite croissante d'ouverts, alors $C_c(\Omega)$ est un espace complet (contrairement à ce qui se passe pour la topologie uniforme), et il s'avère que $(C_c(\Omega))^* = M_{\mathrm{loc}}(\Omega)$ (c'est bien sûr un avatar du théorème de Riesz). En d'autres termes, les mesures de Radon s'identifient alors au dual de l'espace des fonctions continues à support compact. C'est ce que traduit l'énoncé suivant [Schwartz] (non démontré dans ce cours) :

Théorème VIII-72 (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 $K \subset X$.*

VIII-4.7. Convergence dans $M(X)$. On reviendra sur ce sujet dans le chapitre suivant, mais dressons dès maintenant la liste des trois notions de convergence couramment utilisées dans $M(X)$: on distingue

• la **convergence en variation totale** :

$$\mu_k \xrightarrow[\mathrm{VT}]{} \mu \qquad \text{si} \qquad \|\mu_k - \mu\|_{\mathrm{VT}} \xrightarrow[k\to\infty]{} 0.$$

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** :

$$\mu_k \xrightarrow[w-*]{} \mu \qquad \text{si} \qquad \forall \varphi \in C_c(X), \int \varphi\, d\mu_k \xrightarrow[k\to\infty]{} \int \varphi\, d\mu.$$

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 $C_c(X) = \{0\}$, par exemple si X est un espace de Banach de dimension infinie, auquel cas la définition de convergence faible-$*$ devient vide...).

• la **convergence faible**, ou convergence étroite :

$$\mu_k \xrightarrow[w]{} \mu \qquad \text{si} \qquad \forall \varphi \in C_b(X), \int \varphi\, d\mu_k \xrightarrow[k\to\infty]{} \int \varphi\, d\mu.$$

Cette notion est bien plus faible de la convergence en variation totale, mais plus forte (un peu plus forte en dimension finie, beaucoup plus forte en dimension infinie) que la convergence faible-étoile. Elle est très populaire parmi les probabilistes, quand les μ_k sont des mesures de probabilité.

Bibliographie

Les textes suivants sont 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).

Une introduction à la théorie descriptive des ensembles se trouve dans les notes de cours de **J. Melleray**, *Some aspects of topological dynamics of Polish groups.* Une introduction à l'histoire et aux bases de la théorie des sélections mesurables se trouve dans l'exposé de **C. Dellacherie**, *Ensembles analytiques, théorèmes de séparation et applications* (Séminaire de probabilités de Strasbourg, tome 9 (1975), pp. 336-372).

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. Les domaines non lisses sont traités dans **P. Grisvard**, *Elliptic problems in nonsmooth domains* (Pitman, Boston, 1985).

L’ouvrage de référence en théorie descriptive des ensembles est celui de **A. Kechris**, *Classical descriptive set theory* (Springer-Verlag, 1995) ; on y trouvera aussi une introduction à la technique du forçage de Paul Cohen. D’autres introductions à la théorie descriptive des ensembles, plus modernes et bien plus concises, se trouvent dans les notes de cours d’**A. Tserunyan**, *Introduction to descriptive set theory* (très populaire dans le milieu même de la théorie descriptive des ensembles, disponible sur sa page Web avec d’autres notes de cours) ; ou dans celles de **J. Melleray** (*Some aspects of topological dynamics of Polish groups*, cours donné à Turin en 2024) ; ou encore dans celles de **B. Miller** (*An introduction to classical descriptive set theory*, Vienne 2015, aussi disponible sur sa page Web).

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) ; celui de **K.R. Parthasarathy**, *Probability Measures on Metric Spaces* (Academic Press, New York, 1967) ; ou celui, riche en énoncés et en références, de **V. Bogachev**, *Weak Convergence of Measures* (Vol. 234 de la collection Mathematical Surveys and Monographs de l’American Mathematical Society, Providence, 2018). On peut aussi parcourir la synthèse claire et concise en 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. Il y aura quelques exceptions, en particulier 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).