Cours au S5 des parcours Mathématiques-Informatique et Mathématiques-économie : Topologie et Théorie de la mesure.

Yoann Dabrowski

27 septembre 2025

Ce polycopié de cours est issu d'un cours donné en 2019 puis aussi de 2023 à 2025 à l'Université Claude Bernard Lyon 1. Il a été rendu plus accessible aux lecteurs dyslexiques en utilisant le travail décrit dans *Making an Accessible Open Logic Textbook (for Dyslexics)* par Richard Zach [1].

Le code latex pour la présentation du livre *forallx: Calgary (Accessible)* par P.D. Magnus, Tim Button, Robert Trueman et Richard Zach, a été utilisé sous licence CC BY 4.0, disponible sur le site https ://creativecommons.org/licenses/by/4.0/. Ce cours et disponible au lien :
https ://forallx.openlogicproject.org/forallxyyc-accessible.pdf . Vu notre utilisation systématique d'environnements pour énoncer des définitions, Théorèmes, Lemmes, etc. comme il est d'usage en mathématiques, nous avons en plus veillé à appliquer les mêmes usages typographiques recommandés dans ces environnements dans les versions accessibles de ce cours.

Certaines sections indiquées au cours du texte sont tirées d'un polycopié du même cours de 2018-2019 de Thomas Blossier, Maria Carrizosa et Julien Melleray avec permission.

L'auteur ne prétend bien sûr à aucune originalité mathématiques sur des sujets si classiques. Il espère cependant, après quinze ans d'enseignements de l'analyse et des probabilités en parcours mathématiques et économie, qu'il a atteint son objectif pédagogique de permettre plusieurs niveaux de lecture à un publique qui a principalment besoin des applications du sujet en probabilité et modélisation. Au niveau minimum, il suffit d'apprendre les définitions et résultats principaux avec ★ et de bien comprendre les exemples qui seront la source d'exercices types incontournables. A un deuxième niveau, les étudiants hésitant avec des études de mathématiques appliquées devraient comprendre les résultats du corps du texte et leurs preuves. C'est l'enseignement que l'auteur donne en pratique au tableau pendant les 50 heures de ce cours. Enfin, les étudiants à l'aise qui se destinent à la recherche mathématique, malgré leur parcours inhabituel, auront tout intérêt à faire des excursions dans les compléments en appendices, qui rassemblent des preuves supplémentaires et des prolongements immédiats, le plus souvent nécessaires pour les preuves supplémentaires de d'autres sections de l'appendices. Ce sont des

1. Voir aussi du même auteur *Accessible Open Textbooks in Math-Heavy Disciplines The challenge*

matériaux soit enseignées à d'autres niveaux, soit enseignées dans des versions précédentes de ce cours et qui ce sont révélées trop ambitieuses pour le public visé.

Table des matières

CHAPITRE 1
Ensembles dénombrables et Familles sommables

Un espace de probabilité discret (disons dénombrable) associe des nombres, les probabilités aux évènements de base $\{\omega_i\}$, correspondant aux éléments ω_i de l'espace des réalisations et en sommant à des évènements plus compliqués. Comme ces nombres vont être associés à des ensembles, l'ordre de sommation de ces nombres ne doit pas importer. On va donc étudier une notion de sommation de série où l'ordre de sommation n'importe pas. Le but est donc pour une famille de nombres $(u_i)_{i\in I}$, indicée par un ensemble infini I (le plus souvent dénombrable) de définir la somme :

$$\sum_{i\in I} u_i,$$

en conservant les propriétés de commutativité et d'associativité des sommes finies.

Même dans le cas $I = \mathbb{N}$, le but est d'obtenir une notion de sommation qui ne privilégie pas les sous-ensembles finis $[\![0, n]\!]$ comme la notion de somme de série usuelle. On verra que dans ce cas, cette notion de sommation coïncide avec la convergence absolue que vous connaissez déjà.

Le but de la Théorie de la mesure sera d'étendre cette construction à des espaces dits mesurés (de probabilité ou de masse totale différente de 1), incluant les espaces probabilités continues. Le principe de la construction sera le même et généralisera le cas plus simple de ce chapitre.

1 Ensembles (au plus) dénombrables

Rappels sur les ensembles

Définition 1.1. La **fonction indicatrice** d'une partie A est l'application $1_A : \Omega \to \{0; 1\}$ définie par

$$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

On a admis en L1 l'existence de l'ensemble $\mathbb{N}$ des entiers naturels et d'un ensemble constitué des parties de Ω (ce sont des axiomes de base de la théorie des ensembles).

Définition 1.2. L'ensemble des parties de Ω est noté $\mathcal{P}(\Omega)$. Une **famille** $\mathcal{F}$ de parties de Ω est une partie de $\mathcal{P}(\Omega)$ (soit $\mathcal{F} \subset \mathcal{P}(\Omega)$ ou $\mathcal{F} \in \mathcal{P}(\mathcal{P}(\Omega))$). Les éléments de $\mathcal{F}$ sont des parties de Ω.

Lemme 1.1. La fonction indicatrice $A \mapsto 1_A$ réalise une bijection entre $\mathcal{P}(\Omega)$ et $\{0, 1\}^\Omega$ (l'ensemble des applications de Ω dans $\{0, 1\}$).

Démonstration. L'inverse est $h \mapsto h^{-1}(\{1\})$. La vérification que c'est bien un inverse est facile, et laissée en exercice. □

Rappel 1.1. Si A et B sont deux parties de Ω (i.e. deux éléments de $\mathcal{P}(\Omega)$).

1. On a les relations $A \subset B$ ou $B \subset A$ ou ($A \not\subset B$ et $B \not\subset A$). $A \subset B$ s'écrit aussi $B \supset A$.
2. On a défini en $L1$: $A \times B$ l'ensemble des couples (a, b) $a \in A$, $b \in B$, l'intersection $A \cap B$ (ensemble des éléments à la fois dans A **et** dans B), l'union $A \cup B$ (ensemble des éléments à la fois dans A **ou** dans B), le complémentaire de B dans A :$A - B = A \cap B^c = \{x \in A : x \notin B\}$ et la différence symétrique $A \Delta B = (A - B) \cup (B - A)$. On remarquera la relation de ces opérations avec les connecteurs logiques de base.
3. Plus généralement on définit l'union d'une famille $A_i \in \mathcal{P}(\Omega)$, $i \in I$:
$$\bigcup_{i \in I} A_i = \{x \in \Omega : \exists i \in I : x \in A_i\},$$
et de l'intersection d'une même famille :
$$\bigcap_{i \in I} A_i = \{x \in \Omega : \forall i \in I : x \in A_i\}.$$

qui vérifie les relations de distributivités :

$$\Big(\bigcup_{i\in I} A_i\Big)\cap C = \bigcup_{i\in I}(A_i\cap C)$$

$$\Big(\bigcap_{i\in I} A_i\Big)\cup C = \bigcap_{i\in I}(A_i\cup C)$$

et plus généralement

$$\Big(\bigcup_{i\in I} A_i\Big)\cap \Big(\bigcup_{j\in J} C_j\Big) = \bigcup_{i\in I, j\in J}(A_i\cap C_j).$$

$$\Big(\bigcap_{i\in I} A_i\Big)\cup \Big(\bigcap_{j\in J} C_j\Big) = \bigcap_{i\in I, j\in J}(A_i\cup C_j).$$

4. A et B sont disjoints si $A\cap B=\emptyset$.
5. On a les relations fondamentales du complémentaire $(A^c)^c = A$ et pour le complémentaire des unions

$$\Big(\bigcup_{i\in I} A_i\Big)^c = \bigcap_{i\in I} A_i^c$$

et (de façon équivalente) des intersections :

$$\Big(\bigcap_{i\in I} A_i\Big)^c = \bigcup_{i\in I} A_i^c.$$

$*$ *Rappel* 1.2. Soit $A\subset E$ et $f:\Omega\to E$, on rappelle que l'image réciproque $f^{-1}(A)$ est définie par :

$$f^{-1}(A) = \{\omega\in\Omega : f(\omega)\in A\}.$$

On a vu en L1 les relations

$$f^{-1}(A\cup B) = f^{-1}(A)\cup f^{-1}(B),$$

$$f^{-1}(A\cap B) = f^{-1}(A)\cap f^{-1}(B),$$

$$\begin{aligned} f^{-1}(A^c) &= [f^{-1}(A)]^c,\\ f^{-1}\Big(\bigcup_{i\in I} A_i\Big) &= \bigcup_{i\in I} f^{-1}(A_i),\\ f^{-1}\Big(\bigcap_{i\in I} A_i\Big) &= \bigcap_{i\in I} f^{-1}(A_i). \end{aligned} \tag{1.1}$$

Un ensemble A qui n'est pas fini est dit **infini**.

Ensembles infinis dénombrables

★ **Définition 1.3.** Un ensemble infini A est **dénombrable** s'il existe une bijection $f : A \to \mathbb{N}$.
Un ensemble A est **au plus dénombrable** s'il existe une injection $f : A \to \mathbb{N}$.

Remarque 1.3. Certains auteurs disent dénombrable pour ce que nous appelons au plus dénombrable et infini dénombrable avec le sens de dénombrable ci-dessus.

On va utiliser librement le lemme suivant :

Lemme 1.2.
1. Toute partie non-vide de $\mathbb{N}$ a un minimum.
2. Une application strictement croissante $f : \mathbb{N} \to \mathbb{N}$ (resp. $f : [\![0, n]\!] \to \mathbb{N}$) vérifie $f(p) \geq p$ pour tout p dans son domaine.

Démonstration. 1. Si P est non-vide et donc, disons, contient n, alors $[\![0, n]\!] \cap P$ est aussi non-vide et FINI, donc a clairement un minimum. 2. Il suffit de voir le deuxième cas (en restreignant aux segments initiaux), on le montre par récurrence sur n. Si $n = 0$, $f(0) \in \mathbb{N}$ donc c'est évident. En supposant l'hypothèse vraie au rang n, on considère $f : [\![0, n+1]\!] \to \mathbb{N}$, la restriction à $[\![0, n]\!]$ vérifie l'hypothèse de récurrence, donc $f(p) \geq p$ pour $p \leq n$ et $f(n+1) > f(n) \geq n$ mais dans $\mathbb{N}$ cela implique $f(n+1) \geq n+1$ et conclut l'étape d'induction. □

On peut représenter les éléments d'un ensemble dénombrable A à l'aide d'une suite infinie en écrivant $A = \{x_n ; n \geq 1\}$ (x est l'inverse de la bijection f).

★ **Proposition 1.3.** Les ensembles au plus dénombrables sont soit finis, soit dénombrables. De plus, pour une partie infinie $P \subset \mathbb{N}$, il existe une bijection strictement croissante et une seule de $\mathbb{N} \to P$.

Démonstration. Les ensembles au plus dénombrables sont par définition en bijection avec les parties de $\mathbb{N}$. Dans le cas infini, il suffit de voir le second point pour obtenir la bijection avec $\mathbb{N}$. On définit par récurrence la bijection $f : \mathbb{N}^* \to P$. Plus précisément, on construit par récurrence sur n une application strictement croissante $f_n : [\![1, n]\!] \to P$ telle que pour tout $x \in Im(f_n)$, $y \in P - Im(f_n)$, $x < y$ et $f_n|_{[\![1,k]\!]} = f_k$. Comme P, infini, il est non-vide donc admet un élément $a_0 = \min(P)$ On pose $f_0(0) = a_0$ d'où l'initialisation.

On suppose construit f_n, et on prend $a_{n+1} = min(P - Im(f_n))$ qui existe car cette partie est infinie de $\mathbb{N}$ donc non vide (si elle n'était pas infinie, P serait finie comme union finie de parties finies). On pose $f_{n+1}(k) = f_n(k)$, $k \le n$, $f_{n+1}(n+1) = a_{n+1}$ de sorte que par l'hyp de rec sur f_n, $a_{n+1} > f_n(k)$, $k \le n$ ce qui donne la stricte croissance de f_{n+1} en combinant avec celle de f_n. Enfin, si $y \in P - Im(f_{n+1}) \subset P - Im(f_n)$ on a par hyp de rec $y > f_n(k) k \le n$ et $y > a_{n+1}$ car c'est le min donc $\ge$ et on a $y \neq a_{n+1}$ par construction. Donc la relation demandée à l'étape suivante est vérifiée.

On obtient f strictement croissante donc injective en rassemblant les valeurs des f_n qui s'accordent ($f(n) = f_n(n) = f_m(n)$, $m \ge n$).

Pour voir que f bijective, par l'absurde, sinon il existe $b \in P - Im(f)$ mais par stricte croissance d'entiers $f(n) \to \infty$ donc il existe n minimal tel que $b < f(n) = f_n(n)$ ce qui impose par minimalité $b > f(n-1)$ et contredit $f_n(n) = Min(P - Im(f_{n-1}))$ vu $b \in P - Im(f_{n-1})$.

Pour l'unicité, si g est une autre telle bijection $g^{-1} \circ f$ est une bijection strictement croissante de $\mathbb{N} \to \mathbb{N}$ ainsi que sa réciproque et le lemme 1.2 donne donc
$g^{-1} \circ f(n) \ge n$, $f^{-1} \circ g(n) \ge n$ et donc, d'où par croissance de g, f appliquée encore à ces relations : $f = g$. □

★ **Proposition 1.4.** Un ensemble P est au plus dénombrable si et seulement si il existe une surjection $f : \mathbb{N} \to P$.

Démonstration. Pour l'implication directe, si P est dénombrable, la bijection de la définition convient, si P est fini, en bijection avec $[[0, n-1]]$ alors le reste modulo n donne la surjection $\mathbb{N} \to [[0, n-1]]$ qui composée à la bijection donne la surjection cherchée. Réciproquement, l'ensemble $f^{-1}(p)$, $p \in P$ est une partie de $\mathbb{N}$ qui a un plus petit element a_p : $a : P \to \mathbb{N}$ est l'injection cherchée. □

On va obtenir des exemples d'ensembles dénombrables les plus courants. Pour cela on a besoin de quelques méthodes de constructions.

Lemme 1.5.
1. La réunion d'une suite $(X_n)_{n \ge 0}$ d'ensembles finis 2 à 2 disjoints est au plus dénombrable.
2. Un ensemble X est au plus dénombrable si et seulement si il admet une **suite exhaustive** de parties finies, c'est à dire une suite croissante de parties finies dont l'union est X.

3. Le produit cartésien d'un nombre fini d'ensembles au plus dénombrables est au plus dénombrable.

Démonstration. 1. Soit $a_n = Card(X_n)$ et $A_n = \sum_{k=0}^{n} a_k$ ($A_{-1} = 0$). On a des bijections $h_n : [\![A_{n-1}+1, A_n]\!] \to [\![1, a_n]\!] \to X_n$ qui induisent une application $h : \mathbb{N}^* \to \cup_n X_n$ dès qu'un nombre infini de X_i n'est pas vide, ou $h : [\![1, A_p]\!] \to \cup_n X_n$ qui est par construction surjective. L'injectivité des h_n et le fait que les X_n sont disjoints donne l'injectivité de h. 2. Si X est fini on prend la suite constante, sinon, pour une bijection $h : \mathbb{N} \to X$ on prend $X_n = h([\![0, n]\!])$ comme suite croissante cherchée. Réciproquement, la suite croissante X_n donne une suite disjointe X_0, $X_{n+1} - X_n$ de parties finies, donc 1 donne que l'union est au plus dénombrable.

3. Une récurrence triviale ramène au cas du produit de 2 ensembles A, B. Soit $h : \mathbb{N} \to A$, $g : \mathbb{N} \to B$ des surjections données par la proposition 1.4. $f = h \times g : \mathbb{N}^2 \to A \times B$ est une surjection qui ramène au cas $\mathbb{N}^2$ qui admet pour suite exhaustive d'ensembles finis $[\![0, n]\!]^2$. □

★ **Proposition 1.6.** Les ensembles $\mathbb{N}^k$, $k \in \mathbb{N}^*$; $\mathbb{Z}$ et $\mathbb{Q}$ sont infinis dénombrables.

Démonstration. On a vu le cas du produit $\mathbb{N}^k$ au lemme précédent. $[\![-n, n]\!]$ est une suite exhaustive d'ensemble fini pour $\mathbb{Z}$ qui est donc au plus dénombrable par la proposition précédente, il est infini car il contient $\mathbb{N}$. Enfin $(p, q) \mapsto p/q$ est une surjection de $\mathbb{Z} \times \mathbb{N}^* \to \mathbb{Q}$, donc, par la proposition 1.4, $\mathbb{Q}$ est au plus dénombrable, et infini car il contient $\mathbb{N}$. □

Enfin, on améliore le lemme précédent.

Proposition 1.7. Une réunion au plus dénombrable d'ensembles au plus dénombrables est au plus dénombrable.

Démonstration. Soit $(X_n)_{n \geq 0}$ une suite d'ensembles dénombrables (si la suite est finie, on peut la prolonger en une suite infinie.). Soit $f_n : \mathbb{N} \to X_n$ une surjection donnée par la proposition 1.4. Petite subtilité, on a besoin de former une suite $(f_n)_{n \in \mathbb{N}}$, c'est à dire une application de $\mathbb{N} \to \left(\bigcup_{n \in \mathbb{N}} X_n\right)^{\mathbb{N}}$, ce qui n'est pas complètement anodin et utilise l'axiome du choix dénombrable). On pose $f : \mathbb{N}^2 \to \bigcup_{n \in \mathbb{N}} X_n$ défini par $f(n, p) = f_n(p)$ et en composant avec une surjection $\mathbb{N} \to \mathbb{N}^2$, on obtient le résultat par la réciproque dans la proposition juste citée. □

Les ensembles au plus dénombrables serviront de base aux probabilités discrètes.

Ensembles infinis non dénombrables

Les ensembles qui n'appartiennent pas aux catégories précédentes (finis ou infinis dénombrables) sont dits infinis non dénombrables. On va voir que par exemple, $\mathbb{R}$ et $\mathbb{C}$, $[a, b]$, $a < b$ sont infinis non dénombrables.

Le résultat clef est toujours un argument diagonal :

★ **Lemme 1.8** (Théorème de Cantor)**.** Il n'existe pas de surjection $h : E \to \mathcal{P}(E)$ entre un ensemble E et l'ensemble de ses parties.

Démonstration. En effet une application $h : E \to \mathcal{P}(E)$ permet de considérer l'ensemble $A = \{x \in E : x \notin h(x)\}$. Il n'existe pas de y tel que $h(y) = A$ car par l'absurde, si il existait, soit $y \in A$ et alors $y \notin h(y) = A$ une contradiction, soit $y \notin A$ et alors $y \in h(y) = A$ encore une contradiction. □

Remarque 1.4. En conséquence de ce lemme et de la proposition 1.4, $\mathcal{P}(\mathbb{N})$ n'est pas dénombrable (il est infini à cause de l'injection $x \mapsto \{x\}$ défini sur $\mathbb{N}$), car sinon on aurait une surjection de $\mathbb{N} \to \mathcal{P}(\mathbb{N})$. En conséquence $\{0, 1\}^{\mathbb{N}}$, en bijection par la fonction indicatrice n'est pas non-plus dénombrable.

★ **Théorème 1.9.** $[0, 1]$ et $\mathbb{R}$ ne sont pas dénombrables.

En conséquence un intervalle quelconque $[a, b]$, pour $a < b$, en bijection avec $[0, 1]$ ne l'est pas non plus. et un intervalle quelconque contenant au moins deux points (qui contient donc aussi un $[a, b]$) est aussi non-dénombrable.

Démonstration. On construit une injection $\varphi : \{0, 1\}^{\mathbb{N}} \to [0, 1]$ (le cas $\mathbb{R}$ s'en déduit. (l'image de cette injection va être l'ensemble triadique de Cantor). On fixe $a = (a_n) \in \{0, 1\}^{\mathbb{N}}$ on définit une suite de segments emboîtés, on pose $J_0 = [0, 1]$ et si $J_n = [x_n, y_n]$ alors on découpe l'intervalle en trois en posant $u_n = (2x_n + y_n)/3$ et $v_n = (x_n + 2y_n)/3$. Si $a_n = 0$, on pose

$J_{n+1} = [x_n, u_n]$, et si $a_n = 1$, on pose $J_{n+1} = [v_n, y_n]$. On obtient par construction une suite de segments emboîtés, x_n, y_n sont des suites adjacentes et $y_n - x_n \leq 1/3^n$ (récurrence facile) donc l'intersection est un singleton $\cap_n J_n = \{\varphi(a)\}$.

Pour voir que φ est injective on note que si $a \neq a'$ sont deux suites et n le premier indice avec $a_n \neq a'_n$, alors $J_n \cap J'_n = \emptyset$ et les images sont donc distinctes. $\square$

Remarque 1.5. L'ensemble triadique de Cantor a plein de propriétés intéressantes. Topologiquement, il est fermé, totalement disconnecté (les composantes connexes sont les singletons). Il est de longueur nulle (car inclus dans l'union sur tous les cas possibles des J_n dont la longueur perd un facteur 2/3 à chaque n). Le sens de cette longueur sera vu au chapitre 3 (c'est la mesure de Lebesgue). Il est en fait fractal de dimension de Hausdorff $\ln(2)/\ln(3) < 1$ (ce qui réexplique la longueur nulle, mais c'est un sujet beaucoup plus avancé des mesures intermédiaires entre discret et continue).

Exemple 1.1. L'ensemble des nombres irrationnels $\mathbb{R} - \mathbb{Q}$ est non-dénombrable, car sinon son union avec $\mathbb{Q}$ à savoir $\mathbb{R}$ serait dénombrable, ce qui n'est pas le cas.

2 Familles sommables à termes positifs

Rappels

Rappel 1.6. La somme $x + y$ avec $x, y \in \overline{\mathbb{R}}$, est définie à l'exception du cas où $x = \pm\infty$ et $y = -x$.
Contrairement au cas des limites, on pose $0. + \infty = 0$, $t. + \infty = +\infty$ pour $t > 0$.

Pour un ensemble A non-vide (non-nécessairement borné), on utilise $\sup A$ pour le plus petit majorant $M \in \overline{\mathbb{R}}$ de A et $\inf A$ pour le plus grand minorant $m \in \overline{\mathbb{R}}$ de A.

On utilisera aussi $\inf \emptyset = +\infty$, $\sup \emptyset = -\infty$.

Si $(a_i)_{i=1,\dots,n}$ est une suite finie (disons de nombres complexes) et $\sigma : [\![1, n]\!] \to [\![1, n]\!]$ une bijection.

La propriété de commutativité de la somme donne :

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_{\sigma(i)}.$$

Démonstration. En voyant σ comme produit de transpositions, il suffit de montrer le résultat pour $\sigma = (jk)$ une transposition avec $j < k$.

Mais par commutativité ($a+b=b+a$) et associativité ($(a+b)+c=a+(b+c)$) de la somme :

$$\begin{aligned}\sum_{i=1}^{n} a_{\sigma(i)} &= \sum_{i=1}^{j-1} a_{\sigma(i)} + a_{\sigma(j)} + \sum_{i=j+1}^{k-1} a_{\sigma(i)} + a_{\sigma(k)} + \sum_{i=k+1}^{n} a_{\sigma(i)} \\ &= \sum_{i=1}^{j-1} a_i + a_k + \sum_{i=j+1}^{k-1} a_i + a_j + \sum_{i=k+1}^{n} a_i = \sum_{i=1}^{n} a_i.\end{aligned}$$

□

Corollaire 1.10. Si E est fini et $e : [\![1,n]\!] \to E$ une bijection, $f : E \to \mathbb{C}$ alors $\sum_{i=1}^{n} f(e_i)$ ne dépend pas de la bijection e. On note

$$\sum_{e\in E} f(e) = \sum_{i=1}^{n} f(e_i).$$

Démonstration. Si on prend une autre bijection e' on considère la bijection $\sigma = e^{-1}\circ e'$ de sorte que $e\circ\sigma = e'$. La formule de commutativité de la somme conclut :

$$\sum_{i=1}^{n} f(e_i) = \sum_{i=1}^{n} f(e_{\sigma(i)}) = \sum_{i=1}^{n} f(e'_i).$$

□

Le résultat suivant résume les propriétés de manipulation de ces sommes :

Proposition 1.11.

1. Si E fini, on a

$$Card(E) = \sum_{e\in E} 1.$$

2. (Sommation par paquet) Si E fini est une union disjointe finie $E = \bigcup_{i\in I} E_i$ (c'est à dire I fini et $E_i \cap E_j = \emptyset$ si $i \neq j$) et $f : E \to \mathbb{C}$ alors :

$$\sum_{e\in E} f(e) = \sum_{i\in I} \sum_{e\in E_i} f(e).$$

En particulier, on a $Card(E) = \sum_{i\in I} Card(E_i)$.

3. (interversion de sommes finies) Si E, F sont finis et $a : E \times F \to \mathbb{C}$, alors :

$$\sum_{e\in E}\sum_{f\in F} a_{e,f} = \sum_{(e,f)\in E\times F} a_{e,f} = \sum_{f\in F}\sum_{e\in E} a_{e,f}.$$

En particulier, on a $Card(E \times F) = Card(E)Card(F)$.

Démonstration. 1. Si $Card(E) = n$, $E = \{e_1, \dots, e_n\}$ pour une bijection $e : [\![1, n]\!] \to E$, on a donc $\sum_{e\in E} 1 = \sum_{i=1}^{n} 1 = n$ par définition.

2. On pose $j : [\![1, m]\!] \to I$ une bijection et $n_i = Card(E_{j(i)})$ On note $N_0 = 0$, $N_i = \sum_{l=1}^{i} n_l$.

On a $N_i - N_{i-1} = n_i$, $i \geq 1$ donc on a une bijection (en composant la soustraction de N_{i-1} : $[\![N_{i-1}+1, N_i]\!] \to [\![1, n_i]\!]$ avec la bijection donnée par la définition du cardinal $[\![1, n_i]\!] \to E_{j(i)}$, $g_i : [\![N_{i-1}+1, N_i]\!] \to E_{j(i)}$. On pose $g(k) = g_i(k)$, si $k \in [\![N_{i-1}+1, N_i]\!]$. Montrons que g réalise une bijection de $[\![1, N_m]\!] \to E$. En effet, par hypothèse, E est l'union des $E_{j(i)}$, dont tous les éléments sont atteints par g_i, donc par g qui est donc surjective. De plus, si $g(k) = g(l) \in E_i$, comme l'union décrivant E est disjointe, on a $k, l \in [\![N_{i-1}+1, N_i]\!]$ et $g_i(k) = g_i(l)$ et comme g_i est injective, on déduit $k = l$ et donc comme k, l sont arbitraires, on déduit que g est aussi injective.

Donc par définition de la somme sur un ensemble (au début et aux deux dernières lignes) :

$$\begin{aligned}
\sum_{e\in E} f(e) &= \sum_{k=1}^{N_m} f(g(k)) \\
&= \sum_{k=1}^{N_1} f(g(k)) + \sum_{k=N_1+1}^{N_2} f(g(k)) + \cdots + \sum_{k=N_{m-1}+1}^{N_m} f(g(k)) \\
&= \sum_{l=1}^{m} \sum_{k=N_{l-1}+1}^{N_l} f(g(k)) \\
&= \sum_{l=1}^{m} \sum_{k=N_{l-1}+1}^{N_l} f(g_l(k)) \\
&= \sum_{l=1}^{m} \sum_{e\in E_{j(l)}} f(e) \\
&= \sum_{i\in I} \sum_{e\in E_i} f(e)
\end{aligned}$$

Le résultat sur le cardinal est une application du 1. et de la sommation par paquet pour la fonction $f = 1$ constante :

$$Card(E) = \sum_{e\in E} 1 = \sum_{i\in I}\sum_{e\in E_i} 1 = \sum_{i\in I} Card(E_i).$$

3. Il suffit d'appliquer la sommation par paquet aux unions disjointes

$$E \times F = \cup_{e\in E}\{e\} \times F = \cup_{f\in F} E \times \{f\}.$$

Pour le cardinal on a par le 1 et la distributivité de la multiplication par rapport à l'addition :

$$\begin{aligned} Card(E\times F) &= \sum_{(e,f)\in E\times F} 1 = \sum_{e\in E}\sum_{f\in F} 1 \\ &= \sum_{e\in E} Card(F) = Card(F)\sum_{e\in E} 1 \\ &= Card(E)Card(F). \end{aligned}$$

□

Définition et premières propriétés

Définition 1.4. Une famille $(a_i)_{i\in I}$ de nombres réels positifs est dite **sommable** si

$$\sup\left\{\sum_{j\in J} a_j : J \subset I, \text{ fini}\right\} < \infty$$

et alors on note

$$\sum_{i\in I} a_i = \sup\left\{\sum_{j\in J} a_j : J \subset I, \text{ fini}\right\}.$$

Tout d'abord, le résultat simple suivant ramène au cas I dénombrable, ce que l'on supposera par la suite :

Lemme 1.12. Si $(a_i)_{i\in I}$ est une famille sommable, alors le support $I_0 = \{i \in I : a_i \neq 0\}$ est au plus dénombrable.

Démonstration. Si $S = \sum_{i\in I} a_i = 0$, alors $I_0 = \emptyset$. Sinon si $S = \sum_{i\in I} a_i > 0$ et si $I_n = \{i \in I : a_i \geq S/n\}$, alors $I_0 = \cup_{n\geq 1} I_n$ est au plus dénombrable comme union d'une suite d'ensembles finis car $Card(I_n) \leq n$. En effet, si $j \in I_n$, $a_j \geq S/n$ donc si $J \subset I_n$ fini $S \geq \sum_{j\in J_n} a_j \geq SCard(J)/n$ donc $Card(J) \leq n$ et donc $Card(I_n) \leq n$. □

On résume les propriétés générales dans l'énoncé suivant :

Proposition 1.13. 1. (critère des suites exhaustives) Si $(J_n)_{n\geq 0}$ est une suite exhaustive de parties finies de I, alors la famille $(a_i)_{i\in I}$ est sommable si et seulement si la suite $(\sum_{i\in J_n} a_i)_{n\geq 0}$ est bornée et alors on a

$$\sum_{i\in I} a_i = \sup_{n\in\mathbb{N}} \sum_{i\in J_n} a_i = \lim_{n\to\infty} \sum_{i\in J_n} a_i.$$

2. (lemme de domination) Si $a_i \leq b_i$ pour tout i et (b_i) sommable, alors $(a_i)_{i\in I}$ est sommable et alors $\sum_{i\in I} a_i \leq \sum_{i\in I} b_i$.
3. (lemme de permutation) Si $(a_i)_{i\in I}$ est sommable et $\sigma : I \to I$ est une bijection, alors $(a_{\sigma(i)})_{i\in I}$ est sommable de même somme.

Démonstration. 1/ La famille $\sum_{i\in J_n} a_i$ étant inclus dans la famille des sommes finies, il est clair qu'elle est majorée si la famille est sommable (et on a en passant au sup la partie $\geq$ de l'égalité énoncée). Mais réciproquement toute famille finie est inclus dans un certain J_n, par définition d'une suite exhaustive, d'où la borne inverse et la réciproque.

2/ Il suffit de borner les sommes partielles finies $\sum_{i\in J} a_i \leq \sum_{i\in J} b_i$ et passer au sup.

3/ Pour tout J fini, $\sigma(J)$ est fini donc $\sum_{i\in J} a_{\sigma(i)} = \sum_{i\in\sigma(J)} a_i \leq \sum_{i\in I} a_i$. D'où la sommabilité et la première inégalité en passant au sup. En considérant la bijection réciproque σ^{-1} on obtient de même l'autre inégalité. $\square$

Le dernier résultat généralise la commutativité des sommes.

Corollaire 1.14. Une famille à termes positifs $(a_n)_{n\in\mathbb{N}}$ est sommable si et seulement si la série $\sum_{n=0}^{\infty} a_n$ est convergente.

Sommation par paquet et applications

On conclut avec les deux résultats importants, le premier généralise l'associativité des sommes finies. On rappelle qu'une partition $(I_\lambda)_{\lambda\in\Lambda}$ de I est une famille d'ensembles 2 à 2 disjoints d'union égale à I.

★ **Théorème 1.15** (de sommation par paquets – Cas Positif)**.** Soit $(I_\lambda)_{\lambda\in\Lambda}$ une partition de I. Une famille $(a_i)_{i\in I}$ est sommable si et seulement si on a à la fois les deux propriétés suivantes :

1. pour chaque $\lambda \in \Lambda$, $(a_i)_{i\in I_\lambda}$ est sommable, disons de somme σ_λ
2. et $(\sigma_\lambda)_{\lambda\in\Lambda}$ est sommable.

Dans tous les cas (même en l'absence de sommabilité), on a l'égalité :

$$\sum_{i\in I} a_i = \sum_{\lambda\in\Lambda} \sigma_\lambda \equiv \sum_{\lambda\in\Lambda}\left(\sum_{i\in I_\lambda} a_i\right).$$

Démonstration. Commençons par la condition nécessaire. Si $(a_i)_{i\in I}$ est sommable alors les sommes finies d'une sous famille $(a_i)_{i\in I_\lambda}$ sont bornées par les sommes de la famille totale donc on a la première condition de sommabilité et $\sigma_\lambda \leq \sum_{i\in I} a_i$. Plus si on a des sous ensembles finis $J_1 \subset I_{\lambda_1}, \dots, J_n \subset I_{\lambda_n}$ pour des λ_j distincts, ils sont disjoints et leur union $J = \bigcup_{k=1}^{n} J_k$ est un sous-ensemble fini de I donc

$$\sum_{k=1}^{n}\sum_{i\in J_k} a_i = \sum_{i\in J} a_i \leq \sum_{i\in I} a_i$$

Donc en passant successivement au sup sur les J_k fini, on obtient :

$$\sum_{k=1}^{n}\sigma_{\lambda_k} \leq \sum_{i\in I} a_i.$$

Donc la famille $(\sigma_\lambda)_{\lambda\in\Lambda}$ est sommable et on obtient la première inégalité $\geq$ en passant au sup.

Réciproquement, pour tout J partie finie de I on définit $J_\lambda = J \cap I_\lambda$ et on obtient un nombre fini de λ tel que $J = \bigcup_{k=1}^{n} J_{\lambda_k}$. On déduit

$$\sum_{i\in J} a_i = \sum_{k=1}^{n}\sum_{i\in J_k} a_i \leq \sum_{k=1}^{n}\sigma_{\lambda_k} \leq \sum_{\lambda\in\Lambda}\sigma_\lambda.$$

D'où la bornitude sur J qui donne la sommabilité, et l'autre inégalité en passant au sup. □

Un cas particulier est la "version famille sommable" du théorème de Fubini (qui se généralise à un théorème d'intégration). Le cas positif est nommé théorème de Fubini-Tonelli. Il correspond à la décomposition

$$I \times J = \cup_{i\in I}\{i\}\times J = \cup_{j\in J} I \times \{j\}.$$

Il donne un résultat d'interversion des sommes.

★ **Théorème 1.16** (de Fubini–Tonelli)**.** Une famille double $(a_{i,j})_{i\in I, j\in J}$ à termes positifs est sommable si et seulement si on a l'une des deux propriétés équivalentes suivantes :

1. pour tout $i \in I$, $(a_{i,j})_{j\in J}$ est sommable et la famille des sommes $(\sum_{j\in J} a_{i,j})_{i\in I}$ est sommable
2. pour tout $j \in J$, $(a_{i,j})_{i\in I}$ est sommable et la famille des sommes $(\sum_{i\in I} a_{i,j})_{j\in J}$ est sommable

Dans tous les cas (même en l'absence de sommabilité), on a l'égalité :

$$\sum_{(i,j)\in I\times J} a_{i,j} = \sum_{i\in I}\left(\sum_{j\in J} a_{i,j}\right) = \sum_{j\in J}\left(\sum_{i\in I} a_{i,j}\right).$$

Démonstration. C'est une application directe du résultat de sommation par paquets avec les partitions ci-dessus. □

Exemple 1.2*.* Calculons la somme $I = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{1}{(i+j+1)^2}$.
Comme c'est une série à coefficient positifs, chaque somme est somme d'une famille sommable, donc par Fubini–Tonelli, on obtient une somme sur le produit :

$$I = \sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}} \frac{1}{(i+j+1)^2} = \sum_{(i,j)\in\mathbb{N}^2} \frac{1}{(i+j+1)^2}.$$

Comme chaque terme de la somme ne dépend que de $n = i+j+1$, on a envie de considérer la partition de $\mathbb{N}^2 = \cup_{n\in\mathbb{N}^*}\Lambda_n$ avec $\Lambda_n = \{(i,j)\in\mathbb{N}^2 : i+j+1 = n\}$. Par le théorème de sommation par paquet, on a :

$$I = \sum_{n=1}^{\infty}\sum_{(i,j)\in\Lambda_n} \frac{1}{(i+j+1)^2}.$$

Il suffit donc de calculer $\sum_{(i,j)\in\Lambda_n} \frac{1}{(i+j+1)^2}$ Mais Λ_n est fini de taille n vu $\Lambda_n = \{(i, n-1-i) : 0 \le i \le n-1\} \simeq [\![0, n-1]\!]$, donc $\sum_{(i,j)\in\Lambda_n} \frac{1}{(i+j+1)^2} = \frac{\mathrm{Card}(\Lambda_n)}{n^2} = \frac{1}{n}$.
C'est le terme d'une série de Riemann divergente, donc $I = +\infty$ et les familles ne sont pas sommables.

3 Familles sommables à termes scalaires

Comme pour les séries, on se ramène au cas à valeur positif en prenant le module. On pourrait traiter de façon semblable le cas à valeurs vectorielles (par exemple dans $\mathbb{R}^n$ ou dans ce qu'on appelera au chapitre suivant un e.v.n. où toute suite de Cauchy converge, un e.v.n dit complet) en prenant la norme à la place du module. On note $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$ le corps de référence.

★ **Définition 1.5.** Une famille $(z_i)_{i\in I}$ de nombres complexes ou réels est dite **sommable** si la famille $(|z_i|)_{i\in I}$ est sommable. On note $\ell^1(I,\mathbb{K})$ l'ensemble des familles sommables d'éléments de $\mathbb{K}$ indexées par I.

On note

$$||z||_1 = \sum_{i\in I} |z_i|.$$

Lemme 1.17. $\ell^1(I,\mathbb{K})$ est un espace vectoriel et de plus on a pour $u, v \in \ell^1(I,\mathbb{K}), \mu, \nu \in \mathbb{K}$:

$$||\lambda u + \mu v||_1 \leq |\lambda| ||u||_1 + |\mu| ||v||_1.$$

Démonstration. On voit que c'est un sous-espace vectoriel de l'ensemble des fonctions $\mathbb{K}^I$. D'abord, la famille nulle est sommable et de plus si $\lambda, \mu \in \mathbb{K}$, (a_i), (b_i) des familles sommables, pour J fini, on a par l'inégalité triangulaire (des nombres) :

$$\begin{aligned}\sum_{i\in J} |\lambda a_i + \mu b_i| &\leq \sum_{i\in J} |\lambda||a_i| + |\mu||b_i| \\ &= |\lambda| \sum_{i\in J} |a_i| + |\mu| \sum_{i\in J} |b_i| \\ &\leq |\lambda| ||a||_1 + |\mu| ||b||_1\end{aligned}$$

donc comme la valeur est bornée, on obtient, la sommabilité de la famille $(\lambda a_i + \mu b_i)$, donc $\ell^1(I,\mathbb{K})$ est stable par combinaison linéaire et est donc un sous-espace vectoriel de $\mathbb{K}^I$, puisqu'il contient aussi la famille nulle (0).

De plus en passant au sup sur J on obtient $||\lambda a + \mu b||_1 \leq |\lambda| ||a||_1 + |\mu| ||b||_1$. □

Comme d'habitude pour définir l'intégrale (ici on va définir de même la somme), on sépare les parties positives, négatives des parties réelles et imaginaires, pour définir la somme. On note donc $(a_i)_+ = \max(a_i, 0)$, $(a_i)_- = \max(-a_i, 0)$ de sorte que

$$z_j = (\Re z_j)_+ - (\Re z_j)_- + i(\Im z_j)_+ - i(\Im z_j)_-$$

Comme $(\Re z_j)_+ + (\Re z_j)_-$, $(\Im z_j)_+ + (\Im z_j)_- \leq |z_j|$ on déduit que si (z_j) est sommable, alors $((\Re z_j)_+)$, $((\Re z_j)_-$, $((\Im z_j)_+)$, $((\Im z_j)_-)$ le sont aussi par domination.

Définition 1.6. La somme d'une famille sommable $(z_i)_{i\in I}$ est la valeur :

$$\sum_{j\in I} z_j := \sum_{j\in I}(\Re z_j)_+ - \sum_{j\in I}(\Re z_j)_- + i\sum_{j\in I}(\Im z_j)_+ - i\sum_{j\in I}(\Im z_j)_-.$$

Exercice 1.1. Vérifier que la somme d'une famille sommable est une application linéaire. (indication : considérer une suite exhaustive de parties finies pour se ramener au cas des sommes finies).

On a le résultat qui résume les propriétés élémentaires :

Proposition 1.18.
1. Une famille $(z_i)_{i\in I}$ est sommable si et seulement si $(\Re z_i)_{i\in I}$ et $(\Im z_i)_{i\in I}$ sont sommables.
2. $(z_i)_{i\in I}$ est sommable si et seulement si $(\overline{z_i})_{i\in I}$ est sommable et on a :
$$\overline{\sum_{j\in I} z_j} = \sum_{j\in I}\overline{z_j},$$
3. Pour $(z_i)_{i\in I}$ sommable, on a l'inégalité triangulaire généralisée :
$$\left|\sum_{j\in I} z_j\right| \leq \sum_{j\in I}|z_j|.$$
4. (lemme de permutation) Si $(z_i)_{i\in I}$ est sommable et $\sigma : I \to I$ est une bijection, alors $(z_{\sigma(i)})_{i\in I}$ est sommable de même somme. En particulier, si $\sum a_n$ est une série absolument convergente et σ une permutation de $\mathbb{N}$ alors $\sum a_{\sigma(n)}$ est absolument convergente de même somme.

Démonstration. 1/ Les bornes $|\Re z_i| \leq |z_i|$ et $|\Im z_i| \leq |z_i|$ donnent la condition nécessaire par domination. Réciproquement $|z_i| = \sqrt{|\Re z_i|^2 + |\Im z_i|^2} \leq |\Re z_i| + |\Im z_i|$ et comme ℓ^1 est un e.v, on a vu que l'hypothèse implique $(|\Re z_i| + |\Im z_i|)_{i\in I}$ sommable d'où le résultat à nouveau par domination.

2/ l'équivalence est évidente en utilisant 2 fois le 1. L'égalité vient directement de la définition.

3/ On fixe une suite exhaustive J_n de I. D'après le critère des suites exhaustives pour les quatre séries à termes positives intervenant dans la somme,

$\sum_{j\in I} z_j = \lim_{n\to\infty} \sum_{j\in J_n} z_j$, $\sum_{j\in I} |z_j| = \lim_{n\to\infty} \sum_{j\in J_n} |z_j|$ donc par l'inégalité triangulaire pour les sommes finies (et continuité du module)

$$\left|\sum_{j\in I} z_j\right| = \left|\lim_{n\to\infty} \sum_{j\in J_n} z_j\right| = \lim_{n\to\infty} \left|\sum_{j\in J_n} z_j\right| \leq \lim_{n\to\infty} \sum_{j\in J_n} |z_j| = \sum_{j\in I} |z_j|.$$

4/ Tout vient du cas positif, soit par la définition de sommabilité soit par la définition de la somme en terme de somme de familles à termes positifs. Le cas particulier vient du fait que si la famille est indicée par $\mathbb{N}$, le critère des suites exhaustives (appliqué à la suite $[\![0, n]\!]$) implique qu'être sommable équivaut à être absolument convergente. □

Remarque 1.7. Une série $\sum a_n$ telle que pour tout σ permutation de $\mathbb{N}$ on ait $\sum a_{\sigma(n)}$ convergeant est dite inconditionnellement convergente. Un résultat classique qu'on trouve par exemple dans Bourbaki Topologie Générale III.44 dit qu'une série numérique inconditionnellement convergente est absolument convergente. Il n'y a donc pas d'extension possible du dernier énoncé.

On finit avec les résultats de sommation par paquets et de Fubini. Dans les deux cas, on n'a plus d'équivalence comme dans le cas à terme positif. On utilise alors souvent/toujours le cas à terme positif pour montrer la sommabilité nécessaire à appliquer le cas avec signe/complexe.

★ **Théorème 1.19** (de sommation par paquets – Cas Général)**.** Soit $(I_\lambda)_{\lambda\in\Lambda}$ une partition de I. Si une famille $(z_i)_{i\in I}$ est sommable alors on a les deux propriétés suivantes :

1. pour chaque $\lambda \in \Lambda$, $(z_i)_{i\in I_\lambda}$ est sommable, disons de somme σ_λ
2. et $(\sigma_\lambda)_{\lambda\in\Lambda}$ est sommable.

De plus, on a l'égalité :

$$\sum_{i\in I} z_i = \sum_{\lambda\in\Lambda} \sigma_\lambda \equiv \sum_{\lambda\in\Lambda}\left(\sum_{i\in I_\lambda} z_i\right).$$

Démonstration. Comme $(|z_i|)_{i\in I}$, la sommabilité de $(|z_i|)_{i\in I_\lambda}$ vient du cas positif. De plus, par l'inégalité triangulaire des familles sommables (proposition 1.18), $|\sum_{i\in I_\lambda} z_i| \leq \sum_{i\in I_\lambda} |z_i|$ et le théorème de sommation par paquets assure la sommabilité du membre de droite, donc par comparaison, celle de $(\sigma_\lambda)_{\lambda\in\Lambda}$ comme voulu. L'égalité vient du cas positif appliqué aux parties positives et négatives des parties réelle et imaginaire. □

En appliquant la sommation par paquets à la même partition que dans le cas positif, on obtient :

★ **Théorème 1.20** (de Fubini)**.** Si une famille double $(z_{i,j})_{i\in I, j\in J}$ est sommable alors on a les deux propriétés suivantes :

1. pour tout $i \in I$, $(z_{i,j})_{j\in J}$ est sommable et la famille des sommes $(\sum_{j\in J} z_{i,j})_{i\in I}$ est sommable
2. pour tout $j \in J$, $(z_{i,j})_{i\in I}$ est sommable et la famille des sommes $(\sum_{i\in I} z_{i,j})_{j\in J}$ est sommable

De plus, on a l'égalité :

$$\sum_{(i,j)\in I\times J} z_{i,j} = \sum_{i\in I}\left(\sum_{j\in J} z_{i,j}\right) = \sum_{j\in J}\left(\sum_{i\in I} z_{i,j}\right).$$

CHAPITRE 2

Introduction à la Topologie des Espaces Métriques

Dans tout le cours, $\mathbb{K} = \mathbb{R}$, (le corps des nombres réels) ou $\mathbb{C}$ (le corps des nombres complexes). $|\lambda|$ est la valeur absolue ou le module de $\lambda \in \mathbb{K}$.

1 Distance et Norme sur un espace vectoriel

★ **Définition 2.1.** Soit X un ensemble (en général supposé non-vide). Une **distance** sur X est une application $d : X \times X \to [0, +\infty[$ telle que :

i $\forall x, y \in X,\ d(x, y) = d(y, x)$ (symétrie)

ii $\forall x, y, z \in X\ d(x, z) \leq d(x, y) + d(y, z)$ (inégalité triangulaire ou sous-additivité)

iii $\forall x, y \in X\ d(x, y) = 0 \Longleftrightarrow x = y$ (séparation)

Un couple (X, d) est appelé espace métrique (em).

★ **Définition 2.2.** Soit E un $\mathbb{K}$-e.v. Une **norme** sur E est une application $n : E \to [0, +\infty[$ telle que :

i $\forall x \in E,\ \lambda \in \mathbb{K}\ n(\lambda x) = |\lambda| n(x)$ (homogénéité)

ii $\forall x, y \in E\ n(x + y) \leq n(x) + n(y)$ (inégalité triangulaire ou sous-additivité)

iii $\forall x \in E\ n(x) = 0 \Longleftrightarrow x = 0$ (séparation)

Souvent on note $n(x) = ||x||$, sauf dans l'exemple $E = \mathbb{K}$, $n(x) = |x|$. Un couple $(E, ||.||)$ est appelé espace vectoriel normé (evn).

Exemple 2.1. Soit $X \subset E$ une partie (non-vide) avec $d(x, y) = ||x - y||$, alors (X, d) est un espace métrique et tout espace métrique est de cette forme.

Exemple 2.2. Si $E=\mathbb{R}^n$ on a trois normes classiques, si $X=(x_1,\dots,x_n)$:

$$||X||_1 = \sum_{i=1}^{n} |x_i|$$

$$||X||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \text{ (norme euclidienne)}$$

$$||X||_\infty = \max_{i=1\dots n} |x_i|$$

Exercice 2.1. Montrer que ce sont des normes (cf. TD de L2).

★ *Exemple* 2.3. Si $E=C^0([a,b],\mathbb{R})$ l'ensemble des fonctions continues sur $[a,b]$, on a trois normes :

$$||f||_1 = \int_a^b |f(t)|dt$$

$$||f||_2 = \sqrt{\int_a^b |f(t)|^2\,dt}$$

$$||f||_\infty = \sup_{t\in[a,b]} |f(t)|$$

Cette dernière norme est la norme de la convergence uniforme (la convergence pour $||.||_\infty$ coïncidera avec la convergence uniforme)

Le lemme 1.17 se reformule en disant :

Lemme 2.1. $(\ell^1(I,\mathbb{K}),||\cdot||_1)$ est un espace vectoriel normé.

Démonstration. $||\cdot||_1$ vérifie l'inégalité triangulaire (cas $\lambda=\mu=1$ du lemme 1.17). De plus $||\cdot||_1$ est positif. Comme $|a_i|\le||a||_1$, $a_i=0$ si $||a||_1=0$, pour tout i donc $a=0$ ce qui donne l'axiome de séparation. Enfin $\sum_{i\in J}|\lambda\, a_i| = |\lambda|\sum_{i\in J}|a_i|$ donc en passant au sup : $|\lambda|\;||a||_1 = ||\lambda\, a||_1$ (d'où l'homogénéité). □

Exemple 2.4. Si $Z=X\times Y$ avec (X,d_X), (Y,d_Y) des espaces métriques. On définit :

$$d_Z((x,y),(x',y')) = \max(d_X(x,x'),d_Y(y,y')).$$

C'est une distance sur Z (exo) que l'on utilisera dans cette situation ultérieurement (distance produit).

Exemple 2.5. $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ est un espace métrique avec la distance

$$d_{\overline{\mathbb{R}}}(x,y) = \begin{cases} \min(1, |x-y|) & \text{si} \quad x, y \in \mathbb{R} \\ 0 & \text{si} \quad x = y \in \{-\infty, +\infty\} \\ 1 & \text{sinon} \end{cases}$$

Proposition 2.2. (Inégalité triangulaire inverse) Soit (X, d) un espace métrique.

$$\forall x, y, z \in X \left| \, d(x,z) - d(y,z) \, \right| \leq d(x,y).$$

Démonstration. Cas $d(x,z) \geq d(y,z)$: Comme $d(x,z) \leq d(x,y) + d(y,z)$ par l'inégalité triangulaire, on en déduit $\left| \, d(x,z) - d(y,z) \, \right| = d(x,z) - d(y,z) \leq d(x,y)$.

Dans le cas $d(y,z) \geq d(x,z)$, on échange x et y par symétrie. □

2 Métriques équivalentes

★ **Définition 2.3.** Soit X un ensemble. Deux distances d_1 et d_2 sur X sont dites **équivalentes** si

$$\exists c, C > 0, \forall x, y \in X, \quad c d_1(x,y) \leq d_2(x,y) \leq C d_1(x,y).$$

On note alors $d_1 \sim d_2$. Des normes sont **équivalentes** si les distances induites le sont.

Remarque 2.1. L'équivalence des distances est une relation d'équivalence, c'est à dire qu'elle est réflexive ($d_1 \sim d_1$), symétrique ($d_1 \sim d_2 \Rightarrow d_2 \sim d_1$) et transitive ($d_1 \sim d_2, d_2 \sim d_3 \Rightarrow d_1 \sim d_3$). Si deux normes sont équivalentes les notions d'analyses (limite, continuité, ...) sont les mêmes pour les deux normes.

Exemple 2.6. Dans $\mathbb{R}^n$, $||.||_1, ||.||_2, ||.||_\infty$ sont équivalentes (cf. TD de L2). On verra plus tard qu'en dimension finie toutes les normes sont équivalentes.

3 Boules dans un espace métrique

★ **Définition 2.4.** Soient $a \in X$ et $r \in [0, \infty[$.
On appelle **boule ouverte de centre a et de rayon r** de X la partie :

$$B(a, r) = \{x \in E, \mid d(x, a) < r\}.$$

et **boule fermée de centre a et de rayon r** de E la partie :

$$B_F(a, r) = \{x \in E, \mid d(x, a) \leq r\}.$$

On appelle **sphère de centre a et de rayon r** de E la partie :

$$S(a, r) = \{x \in E, \mid d(x, a) = r\}.$$

Dans le cas $r = 0$, $B(x, 0) = \emptyset$, $B_F(x, 0) = \{x\}$.

Exercice 2.2. Dessiner les boules de $\mathbb{R}^2$ pour $||.||_1, ||.||_2, ||.||_\infty$

Parties bornées

Définition 2.5. Un ensemble $A \subset X$ est dit **borné** si
$\exists M \in [0, \infty[, a \in X \forall x \in A, d(x, a) \leq M$, c'est à dire s'il est contenu dans une boule.

4 Suites dans un espace métrique

On rappelle qu'une suite de E est une application $u : \mathbb{N} \to E$ notée $(u_n)_{n \geq 0}$.

Convergence

Définition 2.6 (Convergence)**.** Soit (u_n) une suite d'un espace métrique (X, d). On dit que u_n **converge vers** $l \in X$ (et on note $l = \lim_{n\to\infty} u_n$ ou $u_n \to_{n\to\infty} l$) si la suite numérique $d(u_n, l)$ converge vers 0, c'est-à-dire :

$$\forall \epsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, \quad d(u_n, l) \leq \epsilon.$$

Remarque 2.2. Ceci équivaut à $\forall\epsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, \quad u_n \in B(l, \epsilon)$. Comme dans $\mathbb{R}$ on a unicité de la limite (justifiant la notation). En effet si on a deux limites l_1, l_2 pour n grand $u_n \in B(l_1, \epsilon) \cap B(l_2, \epsilon)$ donc par inégalité triangulaire
$d(l_1, l_2) \leq d(l_1, u_n) + d(u_n, l_2) \leq 2\epsilon$ Comme $\epsilon > 0$ arbitraire $d(l_1, l_2) = 0$, soit par l'axiome de séparation $l_1 = l_2$.

Proposition 2.3. (i) Si $u_n \to u$, alors pour tout x, $d(u_n, x) \to d(u, x)$.

(ii) Toute suite convergente est bornée (réciproque fausse).

(iii) Si E est un evn $u_n \to u$, $v_n \to v$ alors pour toute suite $\lambda_n \in \mathbb{K}$, tel que $\lambda_n \to \lambda$ on a $\lambda_n u_n + v_n \to \lambda u + v$.

Démonstration. (i) Par l'inégalité triangulaire inverse $|d(u_n, x) - d(u, x)| \leq d(u_n, u)$

(ii) Par (i) et le cas réel.

(iii) Vu $\lambda_n u_n + v_n - (\lambda u + v) = \lambda_n(u_n - u) + (v_n - v) + (\lambda_n - \lambda)u$, homogénéité et inégalité triangulaire implique :

$$||\lambda_n u_n + v_n - (\lambda u + v)|| \\ \leq |\lambda_n|||u_n - u|| + ||v_n - v|| + |\lambda_n - \lambda|||u|| \to 0.$$

$$||\lambda_n u_n + v_n - (\lambda u + v)|| \leq |\lambda_n|||u_n - u|| + ||v_n - v|| + |\lambda_n - \lambda|||u|| \to 0.$$

□

Suite extraite, valeur d'adhérence

Définition 2.7. Soit (u_n) une suite de X on appelle **suite extraite** ou **sous-suite** une suite de la forme $v_n = u_{\phi(n)}$, pour $\phi : \mathbb{N} \to \mathbb{N}$ une application strictement croissante

Définition 2.8. On appelle **valeur d'adhérence** d'une suite (u_n) toute limite d'une suite extraite convergente.

Proposition 2.4. Toute suite extraite d'une suite convergente converge vers la même limite. (Autrement dit, toute suite convergente n'a qu'une seule valeur d'adhérence, sa limite.)

Démonstration. Supposons $u_n \to l$ et si v_n une suite extraite, $d(v_n, l)$ est extraite de $d(u_n, l)$ (le résultat est donc une conséquence du cas réel). □

5 Suite de Cauchy, Complétude

Définition 2.9. Une suite (u_n) de X est dite **de Cauchy** si :

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall (p,q) \in \mathbb{N}^2, p \geq N \text{ et } q \geq N$$
$$\Rightarrow d(u_p, u_q) \leq \epsilon.$$

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall (p,q) \in \mathbb{N}^2, p \geq N \text{ et } q \geq N \Rightarrow d(u_p, u_q) \leq \epsilon.$$

La proposition suivante est similaire au cas réel (cf. cours de L2).

Proposition 2.5. Toute suite convergente est de Cauchy. Toute suite de Cauchy est bornée. Toute suite de Cauchy possédant une valeur d'adhérence est convergente.

Définition 2.10. Un espace métrique X est dit **complet** si toute suite de Cauchy de X converge dans X. Si un evn E est complet on dit que c'est un **espace de Banach**.

On a vu en première année que $\mathbb{K}$ est complet (mais pas $\mathbb{Q}$). Vous avez vu en L2 que $(\mathbb{R}^n, ||\cdot||_2)$ est complet. On verra que tout evn de dimension finie est complet.

★ **Proposition 2.6.** Un evn E est complet si et seulement si toute série absolument convergente est convergente.

Démonstration. Si E est complet et (x_i) est absolument convergente, la suite des sommes partielles $S_p = \sum_{i=1}^{p} x_i$ vérifie, pour $q > p$, $||S_p - S_q|| \leq \sum_{k=p}^{q-1} ||x_i||$ donc comme $\sum_{k=1}^{q} ||x_i||$ est convergente donc de Cauchy, on déduit que (S_p) est de Cauchy donc converge.

Réciproquement, si toute suite absolument convergente converge, soit (x_i) une suite de Cauchy. Il suffit de montrer qu'elle admet une sous-suite convergente pour voir qu'elle converge. Par la propriété de Cauchy, on trouve par induction $||x_{n_{k+1}}||$ avec $||x_{n_{k+1}} - x_{n_k}|| \leq \frac{1}{2^k}$ de sorte que la série télescopique $\sum x_{n_{k+1}} - x_{n_k}$ est absolument convergente donc converge, et donc la sous-suite (x_{n_k}) converge. □

Exemple 2.7. Dans le cadre de l'exemple 2.3, vous avez vu en L2 que toute série normalement convergente de $(C^0([a,b],\mathbb{R}), ||.||_\infty)$ converge uniformément. D'après le résultat précédent, c'est équivalent à dire que $(C^0([a,b],\mathbb{R}), ||.||_\infty)$ est un espace de

Banach (aussi vu directement en L2 en analyse 2 Prop 7.6). Par contre ce n'est pas le cas de $(C^0([a,b],\mathbb{R}),||.||_i)$, $i=1,2$. On verra qu'ils sont denses dans les espaces de Lebesgue $L^i([a,b],\mathbb{R})$ qui seront eux complets, et sont les constructions de base de la théorie de l'intégration de Lebesgue.

Proposition 2.7. Si X, Y sont des espaces métriques complets. Alors $X\times Y$ (munie de la distance produit de l'exemple 2.4) est complet.

Démonstration. Si (u_n,v_n) est de Cauchy dans $X\times Y$, de même, (u_n) est de Cauchy dans X, et (v_n) dans Y, donc par complétude (u_n) converge vers u et (v_n) vers v. En conséquence (u_n,v_n) converge vers (u,v) vu $d((u_n,v_n),(u,v))=\max(d(u_n,u),d(v_n,v))\to 0$. □

Théorème de Point fixe

★ **Théorème 2.8** (du point fixe de Banach)**.** Soit (X,d) un espace métrique complet, et $f:X\to X$ une application telle que

$$\exists k<1\ \forall x\neq y\in X\quad d(f(x),f(y))\leq kd(x,y)\ .$$

Alors f admet un unique point fixe.

Démonstration. Soit $x_0\in X$ on définit par récurrence $x_n=f(x_{n-1})=f^{\circ n}(x_0)$. Donc

$$d(x_{n+1},x_n)=d(f(x_n),f(x_{n-1}))\leq kd(x_n,x_{n-1})\leq k^n d(x_1,x_0). \tag{2.1}$$

Montrons que x_n **est bornée** en voyant par récurrence que $d(x_n,x_0)\leq\sum_{i=0}^{n-1}k^i d(x_1,x_0)$. C'est évident pour $n=1$. Et par l'inégalité triangulaire et (2.1) :

$$\begin{aligned} d(x_{n+1},x_0) &\leq d(x_{n+1},x_n)+d(x_n,x_0)\\ &\leq k^n d(x_1,x_0)+\sum_{i=0}^{n-1}k^i d(x_1,x_0)\\ &=\sum_{i=0}^{n}k^i d(x_1,x_0)\end{aligned}$$

Or on reconnaît une série géométrique convergente, d'où la borne : $d(x_{n+1},x_0)\leq\frac{1}{1-k}d(x_1,x_0)$.

Montrons que x_n **est de Cauchy.** En effet, pour $m > n$,

$$d(x_n, x_m) = d(f^{\circ n}(x_0), f^{\circ n}(x_{m-n})) \leq k^n d(x_0, x_{m-n}) \leq k^n \frac{1}{1-k} d(x_1, x_0)$$

Comme $k^n \frac{1}{1-k} \to 0$, on déduit que pour N grand et $m > n \geq N$ $d(x_n, x_m)$ est arbitrairement petit, donc x_n est de Cauchy. Par complétude de X, on obtient donc que x_n converge, disons vers x. Maintenant, en passant à la limite dans (2.1), on obtient
$d(f(x), x) = \lim_n d(f(x_n), x_n) \leq \limsup_n d(f(x_n), x_n) \leq \limsup_n k^n d(x_1, x_0) = 0$ donc par séparation $f(x) = x$ et x est le point fixe cherché.

□

6 Ouverts dans un espace métrique

Soit (X, d) un espace métrique.

★ **Définition 2.11.** Une partie $O \subset X$ est un **ouvert** (ou une partie ouverte) si

$$\forall x \in O, \exists r > 0, \quad B(x, r) \subset O.$$

Exemples d'ouverts et propriétés

$X, \emptyset$ sont des ouverts de X. $[a, b]$, $[a, b[$ ne sont pas ouverts dans $\mathbb{R}$ mais $]a, b[$ l'est.

Proposition 2.9. Les boules ouvertes sont ouvertes.

On remarquera que le mot ouvert a deux sens dans "boules ouvertes" et "parties ouvertes" mais qu'ils sont cohérents grâce à la proposition (les boules fermées ne sont pas des ouverts, cf. TD).

Démonstration. Soit $a \in X$, $r > 0$ montrons que $B(a, r)$ est un ouvert ($B(a, 0)$ est vide donc ouvert). Soit $x \in B(a, r)$, $r - d(x, a) > 0$, il suffit donc de montrer que :

$$B(x, r - d(x, a)) \subset B(a, r).$$

C'est une conséquence de l'inégalité triangulaire. En effet, si $y \in B(x, r - d(x, a))$, alors $d(y, a) \leq d(y, x) + d(x, a) < (r - d(x, a)) + d(x, a) = r$, donc $y \in B(a, r)$. □

★ **Proposition 2.10.** 1. La partie vide $\emptyset$ et X sont des ouverts.
2. la réunion d'une famille d'ouverts est ouverte.
3. l'intersection d'une famille finie d'ouverts est ouverte.

Remarque 2.3. On appelle **topologie** une famille de parties d'un ensemble, qui, comme la famille des ouverts d'un espace métrique, vérifie ces trois propriétés. La famille des ouverts de X est donc appelée topologie (métrique) de X. $\cap_{n\in\mathbb{N}} B(a, 1/n) = \{a\}$ qui n'est pas ouvert dans X montre que l'hypothèse "finie" est cruciale dans 3.

Démonstration. 1. évident.
2. Soit $(O_i)_{i\in I}$ une famille d'ouverts. On peut supposer I non vide (sinon l'union vide étant vide on est ramené à 1). Soit $x \in O = \cup_{i\in I} O_i$, donc il existe $j \in I$, $x \in O_j$. Comme O_j est ouvert il existe $r > 0$, $B(x, r) \subset O_j \subset O$. Donc O est ouvert.
3. Soit $O_1, \dots, O_n$ une famille finie d'ouverts. Soit $x \in O = O_1 \cap \cdots \cap O_n$. Comme $x \in O_i$, et O_i ouvert, il existe $r_i > 0$, $B(x, r_i) \subset O_i$. Soit $r = \min_{i=1\dots n} r_i > 0$. On déduit de la définition que $B(x, r) \subset B(x, r_i) \subset O_i$ donc $B(x, r) \subset O$, ce qui montre que O est ouvert.

□

Exemple 2.8. Soit $O = \{(x, y), x > 0\}$. Montrons que c'est un ouvert de $\mathbb{R}^2$ pour la norme $||.||_\infty$. En effet

$$\begin{aligned} O &= \bigcup_{(x,y)\in O}]0, 2x[\times]y - x, y + x[\\ &= \bigcup_{(x,y)\in O} B_{||.||_\infty}((x, y), x), \end{aligned}$$

est ouvert comme union d'ouverts.

★ **Proposition 2.11** (Ouverts pour la métrique induite). Soit $A \subset (X, d)$ avec la métrique induite, O est un ouvert de A, si et seulement si il existe un ouvert U de X tel que $O = U \cap A$.

Démonstration. On suppose O ouvert de A. Pour chaque $x \in O$, on fixe $r_x > 0$ tel que $B_A(x, r_x) \subset O$. On pose alors

$$U = \bigcup_{x \in O} B_X(x, r_x)$$

qui est un ouvert de X par union de boules ouvertes. Or $O \subset U \cap A$ car $r_x > 0$ donc pour tout $x \in O$, $x \in B_X(x, r_x) \subset U$. Et $U \cap A = \bigcup_{x \in O} B_X(x, r_x) \cap A = \bigcup_{x \in O} B_A(x, r_x) \subset O$. Donc $U \cap A = O$.

Réciproquement, comme U est ouvert soit $x \in O \subset U$, il existe $r > 0$, $B_X(x, r) \subset U$ donc $B_A(x, r) = B_X(x, r) \cap A \subset U \cap A = O$ donc O est ouvert dans A. □

Intérieur

Définition 2.12. Soit $A \subset X$, on dit que x est **intérieur** à A (ou A est un voisinage de x) si $\exists r > 0, B(x, r) \subset A$.
On note $\mathrm{Int}(A)$ ou $\mathring{A}$ l'ensemble des points intérieurs à A.

★ **Proposition 2.12.** $\mathrm{Int}(A)$ est le plus grand ouvert contenu dans A.

Démonstration.

1. $\mathrm{Int}(A)$ **contient tous les ouverts inclus dans A.**
   Soit U un ouvert contenu dans A. Soit $x \in U$, alors comme U est ouvert,
   $\exists r > 0, B(x, r) \subset U \subset A$, donc x est intérieur à A. Ainsi $U \subset \mathrm{Int}(A)$
2. $\mathrm{Int}(A)$ **est un ouvert.** Soit $x \in \mathrm{Int}(A)$. Soit donc $r > 0$ tel que $B(x, r) \subset A$. Comme $B(x, r)$ est ouvert, tout $y \in B(x, r)$ est intérieur à $B(x, r)$ donc intérieur à A. En bilan, $\forall x \in \mathrm{Int}(A), \exists r > 0, \ B(x, r) \subset \mathrm{Int}(A)$, ce qui conclut.

□

Corollaire 2.13 (exo, cf TD).

1. A ouvert si et seulement si $A = \mathrm{Int}(A)$.
2. $A \subset B \Rightarrow \mathrm{Int}(A) \subset \mathrm{Int}(B)$
3. $\mathrm{Int}(A) \cup \mathrm{Int}(B) \subset \mathrm{Int}(A \cup B)$
4. $\mathrm{Int}(A) \cap \mathrm{Int}(B) = \mathrm{Int}(A \cap B)$

Exemple 2.9. Soit $F = \{(x, y), x \geq 0\}$. Montrons que $Int(F) = O := \{(x, y), x > 0\}$. On a vu à l'exemple 2.8 que O est ouvert, donc comme $O \subset F$, on a $O \subset Int(F)$. Il reste à voir que $Int(F) \cap \{(x, y), x = 0\} = \emptyset$ (car alors $Int(F) \subset F - \{(x, y), x = 0\} = O$). Mais soit $(-\epsilon, y) \in B_{||.||_\infty}((0, y), \epsilon) \cap F^c$ pour tout $\epsilon > 0$, donc $B_{||.||_\infty}((0, y), \epsilon) \not\subset F$ donc $(0, y)$ n'est pas intérieur à F, ce qu'il fallait démontrer.

7 Fermés dans un espace métrique.

Soit (X, d) un espace métrique.

Rappel 2.4. Soit $A \subset X$, on note $A^c = \{x \in X \mid x \notin A\}$ le complémentaire de A. On rappelle que $\emptyset^c = X$, $X^c = \emptyset$, $(A^c)^c = A$, $A \cup A^c = X$, $A \cap A^c = \emptyset$. Les lois de De Morgan impliquent que pour une famille $(A_i)_{i \in I}$

$$\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c,$$

$$\left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c.$$

Définition 2.13. Soit $F \subset X$. On dit que F est un **fermé** de X si F^c est un ouvert de X.

Le résultat suivant est obtenu en passant au complémentaire le résultat sur les ouverts.

★ **Proposition 2.14.**
1. La partie vide $\emptyset$ et X sont des fermés.
2. l'intersection d'une famille de fermés est fermée.
3. l'union d'une famille finie de fermés est fermée.

★ **Proposition 2.15** (Caractérisation séquentielle des fermés)**.** Une partie F d'un espace métrique X est fermée si et seulement si toute suite convergente (x_n) d'éléments de F a sa limite dans F.

Démonstration. Supposons F fermé. Soit (x_n) une suite d'éléments de F, convergente vers x. Soit $y \in F^c$, comme F^c est ouvert il existe $\epsilon > 0$ $B(y, \epsilon) \subset F^c$, d'où $x_n \notin B(y, \epsilon)$ Donc $d(x_n, y) \geq \epsilon$. En passant à la limite on déduit

$$d(x, y) \geq | \ d(x_n, x) - d(x_n, y) \ | \geq \epsilon - d(x_n, x) \to_{n \to \infty} \epsilon > 0,$$

Donc $d(x, y) \geq \epsilon$ donc $x \neq y$. Comme y était arbitraire dans F^c, $x \in F$.

Réciproquement, supposons que F n'est pas fermé et montrons que la seconde caractérisation est fausse. Soit $x \in F^c$ montrant que F^c n'est pas ouvert, donc pour tout $n \in \mathbb{N}$, $B(x, 1/n) \cap F \neq \emptyset$. Soit $x_n \in B(x, 1/n) \cap F$ $d(x_n, x) \leq 1/n \to_{n \to \infty} 0$, donc (x_n) est une suite d'éléments de F qui converge vers $x \in F^c$. □

Exemple 2.10. Montrons avec la caractérisation séquentielle que $A = \{(x,y), x > 0, y > 0\}$ n'est pas fermé pour la norme $||.||_\infty$. En effet $A \ni (1/n, 1/n) \to (0,0) \notin A$, ce qui contredirait l'hypothèse que A fermé. Montrons de même que $B = \{(x,y), x \geq 0, y \geq 0\}$ est fermé. En effet, Soit $(x_n, y_n) \in B$ tel que $(x_n, y_n) \to (x,y)$ on a $x_n \to x$, $y_n \to y$ donc comme $x_n \geq 0$, on déduit $x \geq 0$, et de même $y \geq 0$ donc $(x,y) \in B$. Ainsi, comme toute limite de suite de B est dans B, on déduit que B est fermé.

Vous avez vu en L2 le résultat suivant :

Proposition 2.16 (Relations Fermé–Complet)**.** Soit E un espace métrique.

1. Si $C \subset E$ est complet alors il est fermé.
2. Si $C \subset E$ est complet et $F \subset C$ est un fermé de E, alors F est complet.

Démonstration.

1. Si $C \subset E$ est complet alors si on considère une suite (x_n) convergente vers x dans E, elle est de Cauchy, donc converge dans C, donc $x \in C$ par unicité de la limite.
2. Si $C \subset E$ est complet et $F \subset C$. Soit x_n une suite de Cauchy de F, elle converge dans C, donc comme F est fermé, la limite est dans F, donc toute suite de Cauchy de F converge dans F.

□

En passant au complémentaire la proposition 2.11, on obtient :

Proposition 2.17 (Fermés pour la métrique induite)**.** Soit $A \subset (X, d)$ avec la métrique induite, F est un fermé de A, si et seulement si il existe un fermé C de X tel que $F = C \cap A$.

Adhérence

Définition 2.14. Soit $A \subset X$. Un point $x \in X$ est dit **adhérent** à A si $\forall \epsilon > 0 B(x, \epsilon) \cap A \neq \emptyset$. On note $\overline{A}$ (ou $\mathrm{Adh}(A)$) l'ensemble des points adhérents à A.

Exemple 2.11. $\overline{X} = X$, $\overline{\emptyset} = \emptyset$, $A \subset \overline{A}$. Si $r > 0$, $\overline{B(a,r)} = B_F(a,r)$. Si $A = \{x_n\}_{n\in\mathbb{N}}$ les valeurs d'adhérence de la suite (x_n) sont dans $\overline{A}$ qui est l'union de l'ensemble des valeurs d'adhérence et de A (exo).

Proposition 2.18.

$$(\mathrm{Adh}(A))^c = \mathrm{Int}(A^c).$$

$$(\mathrm{Int}(B))^c = \mathrm{Adh}(B^c).$$

Démonstration. Un point $x \in X$ n'appartient pas à $\mathrm{Adh}(A)$ si et seulement si $\exists\epsilon > 0, B(x,\epsilon) \cap A = \emptyset \iff \exists\epsilon > 0, B(x,\epsilon) \subset A^c$. C'est par définition équivalent à dire que x est un point adhérent à A^c.En appliquant le premier résultat à $A = B^c$, on en déduit le second. □

On en déduit toutes les propriétés en passant au complémentaire celles de l'intérieur.

Corollaire 2.19.

1. $\overline{A}$ est le plus petit fermé contenant A.
2. A fermé si et seulement si $A = \overline{A}$.
3. $A \subset B \Rightarrow \overline{A} \subset \overline{B}$
4. $\overline{A} \cap \overline{B} \supset \overline{A \cap B}$
5. $\overline{A} \cup \overline{B} = \overline{A \cup B}$

Démonstration. 1. $\overline{A}$ est fermé vu que son complémentaire est l'ouvert $\mathrm{Int}(A^c)$. Si F est un fermé contenant A, F^c est un ouvert contenu dans A^c donc dans $\mathrm{Int}(A^c)$ le plus grand ouvert contenant A^c. En passant au complémentaire, $F \supset \overline{A}$. Les résultats 2.3.4.5 sont analogues, par passage au complémentaire, de résultats sur l'intérieur. □

★ **Proposition 2.20** (Caractérisation séquentielle de l'adhérence)**.** $x \in \overline{A}$ si et seulement si il existe une suite (a_n) d'éléments de A vérifiant $a_n \to x$.

Démonstration. Si x est adhérent à A pour tout entier n $B(x,1/n) \cap A$ est non vide donc contient un élément a_n. La suite $(a_n) \in A^{\mathbb{N}}$ converge vers x vu $||a_n - x|| \leq 1/n \to 0$. La réciproque vient de la caractérisation séquentielle des fermés vu $\overline{A}$ fermé. □

★ *Exemple* 2.12. Montrons que si $A = \{(x, y), x > 0, y > 0\}$ alors $\overline{A} = B = \{(x, y), x \geq 0, y \geq 0\}$. On a vu à l'exemple 2.10 que B est fermé, donc comme $A \subset B$, on en déduit $\overline{A} \subset B$
Il reste à montrer que $B - A = \{(x, y), x = 0, y \geq 0 \text{ ou } y = 0, x \geq 0\} \subset \overline{A}$. Or $(0, y) = \lim_{n\to\infty}(1/n, y + 1/n)$ et si $y \geq 0$, $(1/n, y + 1/n) \in A$, donc $(0, y) \in \overline{A}$. De même $(x, 0) = \lim_{n\to\infty}(x + 1/n, 1/n) \in \overline{A}$ si $x \geq 0$.

Densité, Frontière

Définition 2.15. Une partie A est dite **dense** dans X si $\overline{A} = X$.

Exemple 2.13. $\mathbb{Q}$ et $\mathbb{Q}^c$ sont denses dans $\mathbb{R}$.

Définition 2.16. Un point $x \in X$ est dit **point frontière** d'une partie A si pour tout $r > 0$, $B(x, r)$ est d'intersection non vide avec A et A^c. On note $\mathrm{Fr}(A)$ l'ensemble des points frontières de A.

Remarque 2.5. D'après la définition, $\mathrm{Fr}(A) = \mathrm{Fr}(A^c) = \overline{A} \cap \overline{A^c}$ est un fermé.

Exercice 2.3. Montrer que $\mathrm{Int}(A^c)$, $\mathrm{Fr}(A)$, $\mathrm{Int}(A)$ forment une partition de X (i.e. sont disjoints deux à deux et leur union est X).

8 Fonctions continues

Définitions équivalentes

On considère $(X, d_X = d)$ et $(Y, d_Y = d)$ deux espaces métriques.

★ **Définition 2.17.** Soient $A \subset X$, Y des espaces métriques et $f : A \to Y$.

1. Soit $a \in A$, f est dit **continue en a** si $\lim_{x\to a} f(x) = f(a)$, soit

$$\forall \epsilon > 0, \exists \delta > 0, \forall x \in A d_X(x, a) < \delta \Rightarrow d_Y(f(x), f(a)) < \epsilon.$$

$$\forall \epsilon > 0, \exists \delta > 0, \forall x \in A d_X(x, a) < \delta \Rightarrow d_Y(f(x), f(a)) < \epsilon.$$

2. f est **continue sur** A si f est continue en tout point de A. Autrement dit,

$$\forall a \in A, \forall \epsilon > 0, \exists \delta > 0, \forall x \in A d_X(x, a) < \delta \Rightarrow d_Y(f(x), f(a)) < \epsilon.$$

$$\forall a \in A, \forall \epsilon > 0, \exists \delta > 0, \forall x \in A d_X(x, a) < \delta \Rightarrow d_Y(f(x), f(a)) < \epsilon.$$

Remarque : $\delta = \delta(a, \epsilon)$ dépend à la fois de ϵ et de a. Vous avez vu en L2, le résultat suivant.

★ **Proposition 2.21** (Caractérisation séquentielle de la continuité)**.** Soit $f : X \to Y$. L'application f est continue en $x \in X$ si et seulement si pour toute suite (x_n) d'éléments de X : si x_n converge vers x, alors $f(x_n)$ converge vers $f(x)$.

Démonstration. Supposons que f tend vers $l = f(x)$ en x. Soit $\epsilon > 0$ il existe $\eta > 0$ tel que $f(B(x, \eta)) \subset B(l, \epsilon)$. Vu que $x_n \to a$ il existe N, tel que $\forall n \geq N$, $d(x_n, a) \leq \eta$ donc $\forall n \geq N$, $d(f(x_n), l) \leq \epsilon$. Ceci indique que $f(x_n) \to l$.

Réciproquement, supposons par contraposition, qu'il existe $\epsilon > 0$ tel que pour tout $\eta > 0$ $f(B(x, \eta)) \cap B(l, \epsilon)^c \neq \emptyset$. Donc, en prenant, $\eta = 1/n$, on obtient $x_n \in B(x, 1/n)$, tel que $d(f(x_n), l) \geq \epsilon$. Pour tout n, donc $x_n \to a$ et $f(x_n)$ ne converge pas vers l comme voulu. □

★ **Proposition 2.22** (Caractérisation topologique de la continuité)**.** Soit $f : X \to Y$. Les assertions suivantes sont équivalentes :

1. f est continue sur X.
2. Pour tout ouvert O de Y, l'image inverse $f^{-1}(O)$ est ouverte dans X.
3. Pour tout fermé F de Y, l'image inverse $f^{-1}(F)$ est fermée dans X.

Démonstration. 2. $\Longleftrightarrow$ 3. vient de $(f^{-1}(B))^c = (f^{-1}(B^c))$ et de la relation fermés/ouverts.

1. $\Rightarrow$ 2. Soit O un ouvert de Y et $x \in O$, il existe et on choisit $\epsilon(x) > 0$ tel que $B(x, \epsilon(x)) \subset O$. Par continuité de f, soit $y \in f^{-1}(O)$, $f(y) = x \in O$, il existe $\delta(y) > 0$ tel que $f(B(y, \delta(y))) \subset B(x, \epsilon(f(y))) \subset O$. Donc $B(y, \delta(y)) \subset f^{-1}(O)$ et comme y est arbitraire, $f^{-1}(O)$ est ouvert.

2. $\Rightarrow$ 1. Soit $a \in A$. Montrons que $\lim_{x \to a} f(x) = f(a)$. Soit $\epsilon > 0$. Par 1. $V = f^{-1}(B(f(a), \epsilon))$ est un ouvert X. Or $a \in V$ donc $\exists \delta > 0$ tel que $B(a, \delta) \subset V$. En conséquence

$$f(B(a, \delta)) \subset f(V) = f(f^{-1}(B(f(a), \epsilon))) \subset B(f(a), \epsilon),$$

ce qui conclut. □

Corollaire 2.23 (Stabilité par composition de la continuité)**.** Si $f : X \to Y$ et $g : Y \to Z$ sont continues, alors $g \circ f : X \to Z$ est continue.

Démonstration. Pour tout ouvert U de Z, $g^{-1}(U)$ est ouvert de Y par coninuité de g, puis $f^{-1}(g^{-1}(U))$ est ouvert par coninuité de f, mais $f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U)$. Comme c'est vrai pour tout ouvert U, on déduit de nouveau du théorème précédent que $g \circ f$ est continue. □

Exemple 2.14.
1. $f : X \to \mathbb{R}$ définit par $f(x) = d(x, z)$ est continue sur E car $|d(x, z) - d(x_0, z)| \leq d(x, x_0)$ (inégalité triangulaire inverse).
2. Soit $0 \leq p \leq n = r + s$, $p : R^n \to \mathbb{R}^s$ définie par si $x = (y, z) \in \mathbb{R}^n = \mathbb{R}^r \times \mathbb{R}^s$, $p(x) = z$. On munit $\mathbb{R}^n$ et $\mathbb{R}^s$ des normes $||.||_1$, on voit $||p(x)||_1 \leq ||x||_1$, donc comme p est linéaire, p est continue car $||p(x) - p(y)||_1 = ||p(x - y)||_1 \leq ||x - y||_1$.

Remarque 2.6. Il résulte des théorèmes sur les limites que les opérations algébriques usuelles (somme, produit, composition) préservent la continuité. En particulier si P est une fonction polynomiale $P : \mathbb{R}^n \to \mathbb{R}$ c'est à dire de la forme $P(x) = \sum_{\text{finie}} a_{i_1,\ldots,i_n} x_1^{i_1} \ldots x_n^{i_n}$ est continue comme somme et produits des projections $(x_1, \ldots, x_n) \mapsto x_i$.

★ **Théorème 2.24** (de prolongement des identités)**.** Si $f, g : (X, d) \to (Y, d)$ sont deux applications continues et $D \subset X$ est dense. Si f et g sont égales sur D, alors elles sont égales (sur tout X).

Démonstration. Soit $x \in X$, on sait par caractérisation séquentielle de l'adhérence qu'il existe $a_n \in D$ avec $a_n \to x$. Par continuité de f, g en x, et caractérisation séquentielle de la continuité : $f(a_n) \to f(x)$, $g(a_n) \to g(x)$. Mais on sait que $f(a_n) = g(a_n)$ par hypothèse, donc par unicité de la limite dans Y, $f(x) = g(x)$. Comme x est arbitraire, on a $f = g$. □

Homéomorphismes, Continuité uniforme, Lipschitzianité

Définition 2.18. Une application $f : X \to Y$ est dite un **homéomorphisme** (ou une application bicontinue) si elle est bijective et si $f : X \to Y$ et $f^{-1} : Y \to X$ sont continues.

★ **Définition 2.19.** Une application $f : X \to Y$ est **uniformément continue** si :

$$\forall \epsilon > 0, \exists \delta > 0 : \forall (x, x') \in X^2, d(x, x') \leq \delta \Rightarrow d(f(x), f(y)) \leq \epsilon.$$

Une application $f : X \to Y$ est **K-lipschitzienne** avec $K \in [0, +\infty[$ si :

$$\forall (x, y) \in X^2, d(f(x), f(y)) \leq K d(x, y).$$

Remarque : dans la continuité uniforme, $\delta = \delta(\epsilon)$ ne dépend PAS de x, contrairement au cas de la continuité.

Proposition 2.25. Une application uniformément continue est continue.

Proposition 2.26. Un application K-lipschitzienne est uniformément continue.

Démonstration. Pour $\epsilon > 0$ dans la définition il suffit de prendre $\delta = \epsilon / K$. □

Exemple 2.15. $f : \mathbb{R}_+ \to \mathbb{R}$ $f(x) = \sqrt{x}$ est uniformément continue mais pas lipschitzienne (cf TD.). Toute application uniformément continue est continue mais la réciproque est fausse : $g : \mathbb{R} \to \mathbb{R}$ $g(x) = x^2$ n'est pas uniformément continue sur $\mathbb{R}$ (cf TD.).
$x \mapsto d(x, z)$ est 1-lipschitzienne $X \to \mathbb{R}$, $(x, y) \mapsto x + y$ est 2-lipschitzienne $E \times E \to E$.

Le résultat suivant ne doit pas être confondu avec le Théorème 2.24qui ne donne que l'unicité d'un prolongement mais pas son existence.

★ **Théorème 2.27** (de prolongement des applications uniformément continues)**.** Si $f : (D, d) \to (Y, d)$ est une application **uniformément continue**, $D \subset (X, d)$ est dense et (Y, d) est **complet**. Alors f admet un unique prolongement continue $g : (X, d) \to (Y, d)$ et celui-ci est uniformément continue.

Démonstration. L'unicité vient du Théorème 2.24.

Soit $x \in X$, et par densité $x_n \in D$, $x_n \to x$. Comme f est uniformément continue soit $\epsilon > 0$ et $\delta > 0$ tel que $d_X(x, y) < \delta \Rightarrow d_Y(f(x), f(y) \leq \epsilon$. Si on prend N tel que $d(x_n, x_m) < \delta$, pour $n, m \geq N$, on voit que $d_Y(f(x_n), f(x_m)) \leq \epsilon$, donc comme ϵ est arbitraire, $(f(x_n))$ est de Cauchy. Donc $(f(x_n))$ converge vers $z \in Y$ par complétude.

Soit $y_n \to x$ une autre telle suite, alors $d(f(y_n), z) \leq d(f(x_n), f(y_n)) + d(f(x_n), z) \to 0$, car $d(f(x_n), f(y_n)) \leq \epsilon$ dès que $d(x_n, y_n) \leq \delta$et on voit donc que $d(x_n, y_n) \to 0$ implique que $d(f(x_n), f(y_n)) \to 0$. Donc la limite z ne dépend pas de la suite choisie. On pose $g(x) = z$.

En particulier, g étend f (en considérant la suite constante). Soit $z \in X$ avec $d(x, z) < \delta$ et $z_n \to z$ alors pour n assez grand $d(x_n, z_n) < \delta$ donc $d_Y(f(x_n), f(z_n) \leq \epsilon$ et on déduit en passant à la limite $d_Y(g(x), g(z)) \leq \epsilon$. Donc g est uniformément continue (avec même constantes que f). □

Fonctions continues bornées

Exemple 2.16. Soit X un espace métrique, F un e.v.n. et $C_b(X, F)$ l'ensemble des fonctions continues bornées sur X à valeur dans F, on a la norme uniforme (exo : vérifier que c'est bien une norme) :

$$||f||_\infty = \sup_{x \in X} ||f(x)||_F$$

Le résultat suivant a été vu en L2 pour $F = \mathbb{R}$.

★ **Théorème 2.28.** Les espaces $(C_b(X, F), ||.||_\infty)$, pour X espace métrique et F espace de Banach est un espace de Banach.

Démonstration. On a vu que ce sont des espaces normés. Montrons qu'ils sont complets. Soit f_n une suite de Cauchy, donc comme $||f_p(x) - f_q(x)||_F \leq ||f_p - f_q||_\infty$, pour tout $x \in X$, $(f_p(x))$ est de Cauchy, donc par complétude de F, converge vers une valeur $f(x)$. Soient p, q tels que pour tout x $||f_p(x) - f_q(x)|| \leq \epsilon$ en prenant la limite $q \to \infty$, on déduit $||f_p(x) - f(x)|| \leq \epsilon$ donc $||f_p - f|| \leq \epsilon$. Donc f_p converge uniformément vers f, donc f est continue (résultat de L2 ou exo). De plus, $||f_p||_\infty$ est convergente, donc de Cauchy, donc bornée, disons par M. En passant à la limite dans l'inégalité $||f_p(x)||_F \leq M$, on obtient $||f(x)||_F \leq M$ et donc f est aussi bornée par M. Donc la limite f est continue bornée et f_p converge vers f dans $C_b(X, F)$. Ce qui donne la complétude. □

9 Applications linéaires continues

On considère $(E, ||.||)$ et $(F, ||.||)$ deux evn.

Rappel 2.7. Une application $u : E \to F$ est dite **linéaire** si :

(i) $\forall x, y \in E$, $u(x+y) = u(x) + u(y)$

(ii) $\forall x \in E, \lambda \in \mathbb{K}$, $u(\lambda x) = \lambda u(x)$.

Proposition 2.29. Si $u : E \to F$ est une application linéaire, les assertions suivantes sont équivalentes :

1. u est lipschitzienne.
2. u est continue.
3. u est continue en 0.
4. u est continue en un point.
5. Il existe $a \in E$,$\eta > 0$ tel que $u(B(a,\eta)) \subset B(u(a), 1)$.
6. u est bornée sur la boule unité fermée $\overline{B_E(0,1)}$

Démonstration. (Preuve facultative) 1. $\Rightarrow$ 2., 2. $\Rightarrow$ 3.,3. $\Rightarrow$ 4.,4. $\Rightarrow$ 5. sont évidentes (et n'utilisent pas la linéarité). Si on suppose 5., il existe $\eta > 0$ tel que si $||x - a|| \le \eta$ alors $||u(x) - u(a)|| \le 1$. Soit $h \in E$, $h \neq 0$, $x = a + h\eta/||h||$ de sorte que $||x - a|| \le \eta$, on déduit donc $||u(h)||\eta/||h|| = ||u(x - a)|| \le 1$ c'est-à-dire $||u(h)|| \le ||h||/\eta$ (ce qui est aussi vrai pour $h = 0$). En particulier, si $||h|| \le 1$, on obtient donc 6.

Si on suppose 6., on montre finalement 1, on pose $C = \sup_{||h|| \le 1} ||u(h)|| < \infty$ et on obtient de même pour $h \neq 0$, $||u(h/||h||)|| \le C$ donc $||u(h)|| \le C||h||$ (ce qui est aussi vrai pour $h = 0$). Donc pour tout x, y en utilisant encore la linéarité $u(x - y) = u(x) - u(y)$, on obtient :

$$||u(x) - u(y)|| \le C||x - y||,$$

donc u est C-lipschitzienne. □

Proposition 2.30. Si $\phi : E \to \mathbb{K}$ est une application linéaire (forme linéaire), ϕ est continue si et seulement si son noyau $H = \mathrm{Ker}\,\phi = \phi^{-1}(\{0\})$ est fermé.

Démonstration. Si ϕ est continue, $\phi^{-1}(\{0\})$ est fermé comme image inverse d'un singleton, qui est fermé. Réciproquement, supposons ϕ non nulle, soit e tel que $\phi(e) = 1$. Comme le complémentaire de H est ouvert soit $r > 0$ tel que $B(e, r) \subset H^c$.

Montrons par l'absurde que pour tout $x \in B(e,r)$, $\phi(x) \in B(1,1)$. En effet, sinon soit x avec $|\phi(x)-1| \geq 1$. Si $t = -\phi(x)/(1-\phi(x))$, on $\phi(te+(1-t)x) = t1+(1-t)\phi(x) = t(1-\phi(x))+\phi(x) = 0$. Or $||te+(1-t)x-e|| = |1-t|||x-e|| = ||x-e||/|\phi(x)-1| \leq r$ une contradiction car alors $y = te+(1-t)x \in B(e,r) \cap H$.

On a donc vu $\phi(B(e,r)) \subset B(\phi(e),1)$ d'où ϕ continue par la proposition précédente.

□

Définition 2.20. L'espace $E' := L(E,\mathbb{K})$ des formes linéaires continues sur un e.v.n. E est munie de la **norme duale**

$$||f||_{E'} := \sup_{x \in E, ||x||_E \leq 1} |f(x)|.$$

★ **Définition 2.21.** L'espace $L(E,F)$ des applications linéaires continues d'un e.v.n. E vers un e.v.n. F est munie de la **norme subordonnée** (ou **norme d'opérateur**) :

$$|||f||| := \sup_{x \in E, ||x||_E \leq 1} ||f(x)||_F.$$

Remarque 2.8. La preuve de 6. implique 5. dans la proposition 2.29 montre en fait que si $f \in L(E,F)$ alors f est $|||f|||$−lipschitzienne.

Un espace dual est toujours complet par le résultat suivant :

Théorème 2.31. Si E est un e.v.n. et F un espace de Banach, alors $(L(E,F), |||.|||)$ est un espace de Banach.

Démonstration. Soit B la boule fermée de E de centre 0 et de rayon 1 et $i : L(E,F) \to C_b(B,F)$ la restriction à la boule. Par définition des normes, c'est une isométrie qui identifie donc $L(E,F)$ à un sous espace de $C_b(B,F)$. Montrons que ce sous espace est fermé (il sera donc complet par complétude de $C_b(B,F)$ par théorème 2.28).

Montrons que

$$i(L(E,F)) = \{u \in C_b(B,F) : \forall \lambda, \mu \in K |\lambda| + |\mu| \leq 1,$$
$$\forall x, y \in B,$$
$$u(\lambda x + \mu y) = \lambda u(x) + \mu u(y)\}.$$

Cela suffit car cela décrit $i(L(E,F))$ comme une intersection de fermé vu que $u \mapsto u(y)$ est une application continue sur $C_b(B,F)$. L'inclusion $\subset$ est évidente. Réciproquement si u est continue sur B donc en 0 et dans l'ensemble indiqué, pour $x \in E \setminus \{0\}$, on pose $u_E(x) = ||x||_E u(\frac{x}{||x||_E})$ et $u_E(0) = 0$. D'abord, si $||x|| \leq 1$ on remarque que u_E étend la précédente valeure de u sur B (en prenant $y = 0$ dans la relation). De même, u_E est positivement homogène. Donc, si $(x,y) \neq 0$, on pose $x' = x/\max(||x||, ||y||)$, $y' = y/\max(||x||, ||y||)$, $\lambda' = \lambda/(|\lambda| + |\mu|)$, $\mu' = \mu/(|\lambda| + |\mu|)$ pour obtenir par homogénéité et la relation appliquée à x', y', λ', μ' :

$$\begin{aligned} u_E(\lambda x + \mu y) &= (|\lambda| + |\mu|) \max(||x||, ||y||) u(\lambda' x' + \mu' y') \\ &= (|\lambda| + |\mu|) \max(||x||, ||y||) [\lambda' u(x') + \mu' u(y')] \\ &= \lambda u_E(x) + \mu u_E(y) \end{aligned}$$

Donc u_E est linéaire continue en 0, donc linéaire continue et $u = i(u_E)$ comme souhaité. $\square$

Définition 2.22. Une application linéaire $u : E \to F$ est une **isométrie** (linéaire) si :

$$\forall x \in E, ||u(x)|| = ||x||.$$

Proposition 2.32. Une isométrie (linéaire) est toujours injective.

Une isométrie $u : E \to F$ identifie donc E au sous-espace vectoriel $u(E) \subset F$ avec la norme induite.

Démonstration. Si $u(x) = 0$ alors $0 = ||u(x)|| = ||x||$ donc par séparation $x = 0$. $\square$

10 Propriétés particulières des evn de dimension finie.

Complétude

★ **Théorème 2.33.** Tout evn de dimension finie est complet.

Démonstration. C'est bien connu en dimension 1. On montre donc le résultat par récurrence sur la dimension. On suppose donc le résultat acquis en dimension strictement inférieure à n, soit $(E, ||.||)$ de dimension n. Soit ϕ une forme linéaire non nulle sur E, son noyau F est de dimension $(n-1)$, donc par hypothèse de récurrence $(F, ||.||)$ (muni de la restriction de la norme de E) est complet. Par conséquent F est fermé dans E, donc ϕ est continue.

Soit $e \in E$ avec $\phi(e) = 1$. L'isomorphisme linéaire $u : (\lambda, f) \to \lambda e + f$ de $\mathbb{K} \times F$ (avec la norme produit donc complet par la proposition 6) sur E est continue ($(1 + ||e||)$–lipschitzien). Son isomorphisme réciproque est donné par :

$$\forall x \in E, \quad u^{-1}(x) = (\phi(x), x - \phi(x)e).$$

u^{-1} est donc aussi continue comme ϕ . u^{-1} étant lipschitzienne (car linéaire continue et par la proposition 2.29), si (x_n), suite de E, est de Cauchy $u^{-1}(x_n) \in \mathbb{K} \times F$ l'est aussi donc converge par complétude de $\mathbb{K} \times F$, d'où $x_n = u(u^{-1}(x_n))$ converge aussi par continuité de u^{-1}. □

Applications linéaires

Rappel 2.9. Si E de dimension n et F de dimension p. Soit $(e_1, \ldots, e_n)$ une base de E, $(f_1, \ldots, f_p)$ une base de F. Une application linéaire u est décrite par sa matrice $A = (a_{ij})_{i \in [1,p], j \in [1,n]}$ dans ces bases. Alors, si $x = \sum_{j=1}^{n} x_j e_j$ et $y = u(x) = \sum_{i=1}^{p} y_i f_i$, on rappelle que :

$$y_i = \sum_{j=1}^{n} a_{ij} x_j.$$

On définit aussi la base duale $(e_1^*, \ldots, e_n^*)$ de l'ev des formes linéaires sur E caractérisés par $e_j^*(e_k) = 1$ si $j = k$ et 0 sinon. En conséquence, pour tout $x \in E$:

$$u(x) = \sum_{j=1}^{n} x_j u(e_j) = \sum_{j=1}^{n} e_j^*(x) u(e_j).$$

★ **Théorème 2.34.** Toute application linéaire entre evn de dimensions finies est continue (et même lipschitzienne).

Démonstration. En utilisant la représentation du rappel

$$u = \sum_{i=1}^{n} u(e_i) e_i^*,$$

il suffit de montrer que les formes linéaires e_i^* sont continues. Mais Ker e_i^* est un sous-espace vectoriel de dimension fini donc complet (Théorème 2.33), donc fermé (proposition 2.16) dans E, d'où la continuité voulue (proposition 2.30). La lipschitzianité vient de la proposition 2.29. □

Équivalence des normes et conséquences.

★ **Théorème 2.35.** Toutes les normes d'un espace vectoriel normé de dimension finie sont équivalentes.

Démonstration. Si $||.||_1$ et $||.||_2$ sont deux normes sur E, l'application linéaire identité $u = Id_E$ vu de $(E, ||.||_1)$ vers $(E, ||.||_2)$ est continue ainsi que son inverse u^{-1} (théorème 2.34), donc elles sont C et $1/c$–lipschitzienne respectivement (proposition 2.29). On en déduit, pour tout $x \in E$:

$$||x||_2 = ||u(x) - u(0)||_2 \leq C||x||_1,$$
$$||x||_1 = ||u^{-1}(x) - u^{-1}(0)||_1 \leq \frac{1}{c}||x||_2,$$

d'où l'équivalence des normes souhaitée. □

Remarque 2.10. Sur R^n on peut donc parler de continuité, limite etc. sans préciser la norme.

Proposition 2.36. Soient E un evn, $A \subset E$, $f : A \to \mathbb{R}^n$. Si $x \in A$, on note $f(x) = (f_1(x), \dots, f_n(x))$ où les f_i sont les fonctions composantes de f : $f_i : A \to \mathbb{R}$. Soit $x \in \overline{A}$ et $b = (b_1, \dots, b_n) \in \mathbb{R}^n$, alors on a l'équivalence :

$$\lim_{x \to a} f(x) = b \Longleftrightarrow \forall i = 1 \dots n \lim_{x \to a} f_i(x) = b_i.$$

Démonstration. On a $f_i = p_i \circ f$, où p_i est i−ème projection $p_i : \mathbb{R}^n \to \mathbb{R}$ définie par $p_i(x_1, \dots, x_n) = x_i$. p_i est continue d'après l'exemple 2.14.2.

Si $\lim_{x \to a} f(x) = b$, on déduit $\lim_{x \to a} f_i(x) = b_i$ d'après le Théorème de composition des limites.

Réciproquement, on munit $\mathbb{R}^n$ de la norme $||.||_\infty$. Si pour tout i $\lim_{x \to a} f_i(x) = b_i$ on a donc pour $\epsilon > 0$, l'existence de $\delta_i > 0$ tel que si $||x - a|| \leq \delta_i$, $||f_i(x) - b_i|| \leq \epsilon$. On pose $\delta = \min_{i=1 \dots n}(\delta_i) > 0$. Donc si $||x - a|| \leq \delta$, pour tout i $||f_i(x) - b_i|| \leq \epsilon$ donc $||f(x) - b||_\infty = \max ||f_i(x) - b_i|| \leq \epsilon$.

□

Corollaire 2.37. Soient E un evn, $A \subset E$, $f : A \to \mathbb{R}^n$. Si $x \in A$, on note $f(x) = (f_1(x), \dots, f_n(x))$ où les f_i sont les fonctions composantes de f : $f_i : A \to \mathbb{R}$. f est continue sur A (resp. en $a \in A$) si et seulement si les f_i sont continues sur A (en resp. $a \in A$).

La preuve du résultat suivant est semblable et omise.

Proposition 2.38. Soit $X_n = (x_n^{(1)}, \dots, x_n^{(p)})$ une suite de $\mathbb{R}^p$ et soit $L = (\ell_1, \dots, \ell_p)$. Alors X_n converge vers L si et seulement si pour tout $i = 1 \dots p$ $x_n^{(i)} \to \ell_i$.

Proposition 2.39. Soient $A \subset \mathbb{R}^n$, $p_i : \mathbb{R}^n \to \mathbb{R}$ la i−ème projection définie par $p_i(x_1, \dots, x_n) = x_i$. Alors A est bornée dans $\mathbb{R}^n$ si et seulement si pour tout i, $p_i(A)$ est bornée dans $\mathbb{R}$.

11 Compacité dans les espaces métriques

★ **Définition 2.23.** Soit K une partie de (X, d) espace métrique K est dite **(séquentiellement) compacte** si elle possède la propriété suivante (dite de Bolzano–Weierstrass) : De toute suite de K, on peut extraire une suite convergente dans K.

Rappel 2.11. Dans $\mathbb{R}$ le théorème de Bolzano–Weierstrass indique que toute suite bornée admet une sous-suite convergente et donc que tout fermé borné est compact.

Proposition 2.40. Un compact K d'un espace métrique X est un fermé borné de X. Un sous-ensemble fermé d'un compact est compact. Le produit de 2 espaces compacts est compact.

Démonstration.

1. Un compact K est fermé, car si une suite (u_n) converge vers l dans E, elle admet une sous-suite convergeant vers $k \in K$, dont la limite est nécessairement $l = k$ (proposition 2.4), donc $l \in K$.
2. On montre par contraposée qu'un ensemble non borné A ne peut pas être compact. Si A non-borné, soit $x_n \in A$ tel que $d(x_n, y) \geq n$, si une suite extraite $x_{\phi(n)} \to x$ convergeait, elle serait bornée, ce qui n'est pas le cas car $d(x_{\phi(n)}, y) \geq \phi(n) \to_{n\to\infty} \infty$.
3. Si $F \subset K$ avec K compact, F fermé, une suite de F admet une sous suite convergeant dans K par compacité, donc sa limite est dans F par fermeture, d'où F compacte.
4. Si K, L sont compacts, pour une suite $(x_n, y_n) \in K \times L$, on extrait une suite $(x_{\phi(n)})$ convergente dans K, puis on réextrait $(y_{\phi(\psi(n))})$ convergente dans L (et a fortiori $(x_{\phi(\psi(n))})$ est aussi convergente) donc $(x_{\phi(\psi(n))}, y_{\phi(\psi(n))})$ converge dans $K \times L$.

□

Exemple 2.17. Soit $F = \{(x, y) \in \mathbb{R}^2, xy = 1\}$ est fermé mais pas compact. En effet, si $f(x, y) = xy$ est polynomiale donc continue $\mathbb{R}^2 \to \mathbb{R}$ donc $F = f^{-1}(\{1\})$ est fermé comme image réciproque d'un fermé par une application continue. Mais F n'est pas compact car pas borné. $x_n = (1/n, n) \in F$ et $||x_n||_\infty = n \to \infty$.

☡ *Remarque* 2.12. En général dans un evn un **fermé borné n'est PAS toujours compact**. Dans $C^0([0, 1], \mathbb{R})$, montrons que la boulde unité fermée n'est pas compacte. $f_n(x) = x^n$ vérifie $||f_n||_\infty = 1$, mais comme $f_n(x) \to f(x)$ (on dit converge simplement vers

f) avec $f(x) = 0$ si $x < 1$, $f(1) = 1$, donc f non continue. Toute suite extraite de f devrait converger vers cette limite qui n'est pas continue, donc elle ne peut pas converger dans $C^0([0,1],\mathbb{R})$ vers cette limite qui n'est pas dans $C^0([0,1],\mathbb{R})$. En général, on peut montrer que les boules fermées d'evn sont compactes si et seulement si l'evn est de dimension finie, on montre une implication ci-dessous.

★ **Théorème 2.41.** Si $u : E \to F$ est continue et $K \subset E$ est compacte alors $u(K)$ est compacte.

Démonstration. Soit y_n une suite de $u(K)$ donc $y_n = u(x_n)$,avec (x_n) suite de K, on extrait donc une suite $x_{\phi(n)}$ convergeant vers $x \in K$. Par continuité, la suite extraite $y_{\phi(n)} = u(x_{\phi(n)}) \to u(x) \in u(K)$. □

★ **Corollaire 2.42** (Thm. de Weierstrass)**.** Si $K \subset X$ espaces métriques est compacte et $f : K \to \mathbb{R}$ est continue, alors la fonction f est bornée et atteint ses bornes : $\exists x_0, x_1 \in K, \forall x \in K f(x_0) \leq f(x) \leq f(x_1)$.

Démonstration. $f(K)$ est compacte donc fermée et bornée. Donc f est bornée, et le $f(K)$ contient son *sup* et son *inf* (par fermeture) c'est-à-dire, il existe $y_0, y_1 \in f(K)$ $y_0 = \inf_{x\in K} f(x)$, $y_1 = \sup_{x\in K} f(x)$. Finalement $y_i = f(x_i)$ avec $x_i \in K$. □

Corollaire 2.43. Soit X, K deux espaces métriques avec K compact et $f : K \to X$ une bijection continue, alors f est une homéomorphisme (c'est-à-dire f^{-1} est continue et X est aussi compacte).

Démonstration. Comme f bijective, pour un fermé $F \subset K$, donc un compact, $(f^{-1})^{-1}(F) = f(F)$ est l'image directe du compact F dans X, donc est compact donc fermé. f^{-1} envoie donc un fermé sur un fermé, donc est continue par caractérisation topologique de la continuité (Proposition 2.22). □

★ **Théorème 2.44.** Dans un evn de dimension finie, les compacts sont exactement les fermés bornés.

Démonstration. Il reste à montrer que les fermés bornés sont compacts. D'après le théorème 2.34 un isomorphisme linéaire u de E sur $\mathbb{K}^n$ est continu de $(E, ||.||)$ sur $(K^n, ||.||_\infty)$, et u^{-1} également. $u(K)$ est fermé comme image réciproque d'un fermé par u^{-1} continue, $u(K)$ est borné comme image d'un borné par une application lipschitzienne. Donc $L = u(K)$ est un fermé borné de $(K^n, ||.||_\infty)$. Il suffit de voir que c'est un compact, car alors $K = u^{-1}(L)$ est compact comme image continue d'un compact (theorème 2.41). Soit $(x_p) = (x_p^{(1)}, \dots, x_p^{(n)})$ une suite de L, par définition de la norme $(x_p^{(i)})$ sont bornés, elles admettent donc, par le théorème de Bolzano–Weierstrass dans $\mathbb{K}$, une sous-suite simultanément convergente. $x_{\phi(p)}^{(i)} \to x^{(i)}$ Donc si $x = (x^{(1)}, \dots, x^{(n)})$, on a $||x_{\phi(p)} - x|| = \max_{i=1\dots n} |x_{\phi(p)}^{(i)} - x^{(i)}| \to 0$ et comme L est fermé ; $x \in L$ ce qui conclut. □

Exemple 2.18. Soit $K = \{(x, y) \in \mathbb{R}^2 x^2 + y^2/2 = 1\}$ est compact. En effet, si $f(x, y) = x^2 + y^2/2$ est polynomiale donc continue $\mathbb{R}^2 \to \mathbb{R}$ donc $F = f^{-1}(\{1\})$ est fermé comme image réciproque d'un fermé par une application continue. De plus $K \subset B_{||.||_\infty, F}(0, \sqrt{2})$ donc K est borné, donc fermé borné dans $\mathbb{R}^2$ de dimendion finie, donc K est compact.

Exemple 2.19. Soit $g : K \to \mathbb{R}$ définie par $g(x, y) = x^2 + y^2$ g est continue donc atteint ses bornes sur K compact. En effet g est la distance euclidienne à l'origine, il est facile de voir qu'elle atteint son maximum 2 en $(0, \pm\sqrt{2})$ sur K et son minimum 1 en $(\pm 1, 0)$ sur K. Le théorème des extremas liés permettra de retrouver ce résultat pour des g et des K plus généraux.

★ **Théorème 2.45** (de Heine)**.** Toute fonction continue f sur un compact $K \subset X$ est uniformément continue.

Démonstration. Soit $g : (x, y) \to d(f(x), f(y))$ de K^2 dans $\mathbb{R}$ elle est continue (pour la distance produit sur X^2 par composition) donc $g(K^2)$ est compact. Soit $\epsilon > 0$ reste à trouver un δ de continuité uniforme.

$$A = \{(x, y) \in K^2 \mid d(f(x), f(y)) \geq \epsilon\} = g^{-1}([\epsilon, +\infty[)$$

est fermé dans K^2 donc compact. Donc l'application continue $(x, y) \mapsto d(x, y)$ atteint sa borne inférieure m. On a $m \neq 0$ car sinon on aurait un $(x, x) \in A$, ce qui n'est pas possible vu $\epsilon > 0$.

Finalement si $\delta > 0$ est tel que $\delta < m$, si $d(x, y) \leq \delta$, on a $(x, y) \notin A$, donc $d(f(x), f(y)) < \epsilon$. □

Complément : un résultat reliant complétude et compacité (facultatif)

Proposition 2.46. Tout espace métrique compact X est complet.

Démonstration. Soit (x_n) une suite de Cauchy de X, elle admet par compacité une suite extraite convergente, donc elle converge (proposition 2.5). □

Définition 2.24. Un espace métrique (X, d) est **précompact** si pour tout $\epsilon > 0$, X peut être couvert par un nombre fini de boules ouvertes de rayon ϵ.

On rappelle le résultat suivant (cf. e.g. Zuily-Quéffelec Th II.1 p135 ou Gourdon d'Analyse p 32) ou la proposition A.7.

Proposition 2.47. Un espace métrique X est compact si et seulement si il est précompact et complet.

Complément : Compacité topologique (facultatif)

On rappelle le résultat suivant (cf. e.g. Gourdon d'Analyse Thm 1 p 28)

Théorème 2.48 (Propriété de Borel-Lebesgue)**.** Pour un ensemble K d'un espace métrique X est compact, si et seulement si, pour tout $(U_i)_{i\in I}$ est un recouvrement de K par des ouverts U_i de X, au sens où $K \subset \bigcup_{i\in I} U_i$ alors K admet un sous-recouvrement fini : il existe $I_0 \subset I$ fini tel que $K \subset \bigcup_{i\in I_0} U_i$.

En passant au complémentaire et à la contraposée, on obtient aussi la version équivalente :

Théorème 2.49. Pour un ensemble K d'un espace métrique X est compact, si et seulement si, pour tout $(F_i)_{i\in I}$ est un fermé de K, si pour toute intersection finie (i.e. avec I_0 fini) est non-vide $\bigcap_{i\in I_0} F_i \neq \emptyset$ alors l'intersection complète est aussi non-vide $\bigcap_{i\in I} F_i \neq \emptyset$.

12 Intégrale de Riemann à valeur Espace de Banach

Nous référons par exemple au Gourdon d'Analyse (chapitre 3 secion 1) pour cette section. Soit F un evn complet. Soit $I = [a, b] \subset \mathbb{R}$ un segment. On rappelle les définitions :

Définition 2.25. Une **subdivision** de $[a, b]$ est suite finie $(a_i)_{i=0,\cdots,n}$ de la forme $a = a_0 < a_1 < \cdots < a_n = b$. Une fonction **continue par morceaux** sur I est une fonction $f : I \to F$ telle qu'il existe une subdivision $(a_i)_{i=0,\cdots,n}$, telle que pour $i \in [[0, n-1]]$, chaque restriction $f_{]a_i,a_{i+1}[}$ est continue et admette des limites en a_i, a_{i+1}. Une fonction $f : I \to F$ est dite **en escalier** si il existe une subdivision $(a_i)_{i=0,\cdots,n}$, telle que pour $i \in [[0, n-1]]$, $f_{]a_i,a_{i+1}[}$ est constante.

On définit $E = \mathcal{C}M(I, F)$ l'ensemble des fonctions continues par morceaux sur I à valeur F. Comme chaque prolongement par continuité de $f_{]a_i,a_{i+1}[}$ est continue sur un compact $[a_i, a_{i+1}]$, donc bornée, les fonctions continues par morceaux sont bornées. On note $D \subset E$ l'ensemble des fonctions en escaliers.

E est donc un Evn (PAS complet) pour la norme de la convergence uniforme, si $f \in E$:

$$||f||_E = \sup_{t \in I} ||f(t)||_F.$$

On va utiliser le théorème suivant de prolongement des applications linéaires continues pour définir l'intégrale à valeur dans F. C'est une application immédiate du Théorème 2.27 :

Proposition 2.50. Toute application linéaire continue u d'un sous-espace vectoriel dense D d'un evn E vers un *evn* complet F se prolonge en une unique application linéaire continue $v : E \to F$, ayant la même constante de lipschitzianité que u.

Démonstration. Comme u est continue donc K−lipschitzienne (par proposition 2.29) donc uniformément continue, l'unique prolongement est donné par le Théorème 2.27.

Si $x_n \to x$, $y_n \to y$ en passant à la limite dans la relation $u(\alpha x_n + \beta y_n) = \alpha u(x_n) + \beta u(y_n)$, on déduit que v est linéaire et avec $||u(x_n - y_n)|| \leq K||x_n - y_n||$, on déduit que v est K−lipschitzienne. □

Pour une fonction en escalier $\phi : [a, b] \to F$ de subdivision $(a_i)_{i=0,\cdots,n}$. On définit

$$I(\phi) = \int_{[a,b]} \phi(t)dt = \sum_{i=1}^{n} (a_i - a_{i-1})\phi\Big(\frac{a_{i-1} + a_i}{2}\Big).$$

I est une application linéaire continue, car par l'inégalité triangulaire

$$||I(\phi)|| \leq \sum_{i=1}^{n} |a_i - a_{i-1}| \Big\|\phi\Big(\frac{a_{i-1} + a_i}{2}\Big)\Big\|_F \leq ||\phi||_E |b - a|.$$

Comme les fonctions en escalier sont denses dans les fonctions continues par morceaux (exo. TD), la proposition précédente permet d'étendre l'intégrale comme quand $F = \mathbb{R}$ et on a :

Définition 2.26. L'**intégrale** des fonctions continues par morceaux $\mathcal{CM}(A, F)$ est l'unique prolongement linéaire continu de l'intégrale des fonctions en escaliers, noté $\int_a^b dt f(t) = \int_a^b f(t)dt$.

Proposition 2.51. (Inégalité triangulaire) $|| \int_a^b dt f(t)||_F \leq \int_a^b dt ||f(t)||_F$.

Démonstration.

$$||I(\phi)||_F \leq \sum_{i=1}^{n} |a_i - a_{i-1}| \Big\|\phi\Big(\frac{a_{i-1} + a_i}{2}\Big)\Big\|_F = \int_a^b ||\phi(t)||_F dt$$

pour ϕ en escalier et on prolonge par continuité. □

On a toutes les propriétés usuelles, Chasles, linéarité, en particulier si $F = \mathbb{R}^n$ et $f = (f_1, \ldots, f_n)$ $\int_a^b f(t)dt = (\int_a^b f_1(t)dt, \ldots, \int_a^b f_n(t)dt)$.

Rappel sur les Intégrales impropres

Définition 2.27. Pour une fonction f continue sur un intervalle $I \to \mathbb{R}$ qui n'inclut pas toutes ses bornes ou qui n'est pas borné, on définit l'**intégrale impropre** de la manière suivante :

1. Dans le cas $I = [a, b[$ avec $a < b$, $b \in \mathbb{R} \cup \{+\infty\}$
$$\int_a^b f(x)dx = \lim_{c \nearrow b} \int_a^c f(x)dx$$
2. Dans le cas $I =]a, b]$ avec $a < b$, $a \in \mathbb{R} \cup \{-\infty\}$
$$\int_a^b f(x)dx = \lim_{c \searrow a} \int_c^b f(x)dx$$

3. Dans le cas $I =]a, b[$ avec $a < b$, $a \in \mathbb{R} \cup \{-\infty\}$, $b \in \mathbb{R} \cup \{+\infty\}$ on prend $a < c < b$ et on pose
$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Dans tous ces cas, on dit que l'intégrale est **convergente** si la limite existe et est finie.

Dans tous les cas, on s'occupera surtout du cas $I = [a, b[$ puisque le cas $I =]a, b]$ est similaire en remplaçant f par $x \mapsto f(-x)$

Le cas le plus important est le cas suivant (car on va disposer de théorèmes de comparaison avec des fonctions positives de références) :

Définition 2.28. Pour une fonction f continue sur un intervalle I (comme dans la définition précédente) est dite **intégrable** sur I si $\int_a^b |f(x)|dx$ converge. Dans ce cas on dit aussi que $\int_a^b f(x)dx$ est **absolument convergente**.

Exercice 2.4. Convergence et valeur de
$$\int_0^1 \frac{1}{\sqrt{x}} dx.$$

La limite infinie est en 0. Donc Soit $t > 0$ on Calcule $\int_t^1 \frac{1}{\sqrt{x}} dx = [2\sqrt{x}]_t^1 = 2 - 2\sqrt{t}$. La limite en $t \to 0$ est finie donc l'intégrale converge et vaut 2.

Exemples de référence (à connaître TRES BIEN)

1. $\int_0^\infty e^{-x} dx$ converge et vaut 1. En effet,$\int_0^A e^{-x} dx = 1 - e^{-A} \to_{A\to\infty} 1$.
   Plus généralement, $\int_0^\infty e^{-ax} dx$ **converge si et seulement si a > 0, et vaut alors 1/a.**
2. $\int_1^\infty \frac{1}{t^\alpha} dt$ **converge si et seulement si** $\alpha > 1$ **(intégrale de Riemann) et vaut**
$$\int_1^\infty \frac{1}{t^\alpha} dt = \frac{1}{\alpha - 1}, \alpha > 1,$$
$$\int_1^\infty \frac{1}{t^\alpha} dt = +\infty, \alpha \leq 1.$$
   En effet, si $\alpha \neq 0$, $\int_1^A \frac{1}{t^\alpha} dt = \frac{A^{-\alpha+1}-1}{-\alpha+1}$ et pour $\alpha > 1$, $A^{-\alpha+1} \to_{A\to+\infty} 0$, tandis que pour $\alpha < 1$ $A^{-\alpha+1} \to_{A\to+\infty} +\infty$
   Si $\alpha = 1$, $\int_1^A \frac{1}{t} dt = \ln(A) \to_{A\to+\infty} +\infty$
3. $\int_0^1 \frac{1}{t^\alpha} dt$ **converge si et seulement si** $\alpha < 1$ **(intégrale de Riemann) et vaut**
$$\int_0^1 \frac{1}{t^\alpha} dt = \frac{1}{1 - \alpha}, \alpha < 1$$

$$\int_0^1 \frac{1}{t^\alpha} dt = +\infty, \alpha \geq 1$$

.

En effet si $\alpha \neq 0$, $\int_a^1 \frac{1}{t^\alpha} dt = \frac{1-a^{-\alpha+1}}{-\alpha+1}$ et pour $\alpha > 1$, $a^{-\alpha+1} \to_{a\to 0} +\infty$, tandis que pour $\alpha < 1$ $a^{-\alpha+1} \to_{a\to 0} 0$

Si $\alpha = 1$, $\int_a^1 \frac{1}{t} dx = |\ln(a)| \to_{a\to\infty} \infty$.

4. $\int_0^\infty \frac{1}{t^\alpha} dt = +\infty$ diverge toujours pour tout $\alpha \in \mathbb{R}$(en combinant les 2 points précédents).

Théorèmes de comparaison

Le contexte est le suivant : on se donne une fonction continue $f : I = [a, b[\to \mathbb{R}$ et on étudie la nature de l'intégrale impropre $\int_a^b f(x)dx$

La méthode la plus simple consiste à chercher une fonction convenable continue et positive $g : I = [a, b[\to [0, \infty[$ et de comparer f à g. Les trois résultats de base à utiliser sont les suivants (avec C>0 une constante).

Théorème 2.52. Théorème de comparaison .

5. Si $|f(x)| \leq Cg(x)$, $\forall x \in [a, b[$, et si $\int_a^b g(x)dx$ converge, alors $\int_a^b f(x)dx$ converge (absolument).
6. Si $f(x) \geq Cg(x)$, $\forall x \in [a, b[$ et si $\int_a^b g(x)dx = +\infty$ alors $\int_a^b f(x)dx = +\infty$.

13 Espaces métriques séparables

Définition 2.29. Une partie A est dite **dense** dans E si $\overline{A} = E$. Un ensemble est dit **séparable** si il admet un sous-ensemble au plus dénombrable dense (ou autrement dit une suite dense).

Lemme 2.53. Un sous-ensemble F d'un espace métrique séparable est séparable.

Démonstration. On peut supposer F non-vide, sinon, c'est évident (la partie vide donc finie est dense). On fixe donc $x_0 \in F$

Soit u_n une suite dénombrable dense. Soit $a_{m,n} \in B(u_m, 1/n) \cap F$ si cet ensemble est non-vide, et sinon on pose $a_{m,n} = x_0$. La famille $\{a_{m,n}, m, n \in \mathbb{N}\}$ est finie ou dénombrable et dense car si $x \in F$ il existe $d(u_m, x) < 1/2n$ donc $a_{m,2n}$ existe car $B(u_m, 1/2n) \cap F$ est non vide et par inégalité triangulaire $d(u_m, a_{m,2n}) < 1/n$. □

Proposition 2.54. $(\mathbb{R}^n, ||.||_\infty)$ est séparable.

Démonstration. On a vu que $\mathbb{Q}^n$ est dénombrable comme produit d'ensembles dénombrables. Montrons qu'il est dense dans $\mathbb{R}^n$. En effet si $x = (x_1, \ldots, x_n)$ on pose $x_p = (\frac{\lfloor px_1 \rfloor}{p}, \ldots, \frac{\lfloor px_n \rfloor}{p})$ avec $\lfloor x \rfloor$ la partie entière de x. Donc $\lfloor px_i \rfloor \leq px_i \leq \lfloor px_i \rfloor + 1$ et

$$\left| \frac{\lfloor px_i \rfloor}{p} - x_i \right| \leq \frac{1}{p}$$

donc $||x_p - x||_\infty \leq 1/p \to_{p\to\infty} 0$. Donc vu $x_p \in \mathbb{Q}^n$, $x \in \overline{\mathbb{Q}^n}$. Comme x est arbitraire. $\mathbb{R}^n \subset \overline{\mathbb{Q}^n}$ CQFD. □

Exercice 2.5. Montrer que $\mathbb{Q}^c$ est dense dans $\mathbb{R}$.

CHAPITRE 3
Ensembles et fonctions convexes, Introduction à l'optimisation

Vous avez vu en L2 qu'une fonction C^1 f qui atteint un minimum sur un ouvert en x satisfait une condition nécessaire du première ordre $\nabla f(x) = 0$ et si f est C^2 on peut garantir que c'est un minimum local si sa hessienne est définie positive.

Il reste les questions : comment avoir un minimum global ? comment avoir unicité du minimum ? Une réponse va être obtenue en étudiant une notion, qui, dans le cas des fonctions C^2, sera équivalente à une positivité globale de la hessienne. L'avantage est qu'on peut trouver une définition : la notion de fonction convexe, sans hypothèse de dérivabilité et qui va être robuste et permettre d'obtenir aussi des critères d'optimisation du premier ordre, sur des ensembles eux aussi convexes (pas forcément ouverts).

On suppose donc que E est un espace vectoriel (e.v.) sur $\mathbb{R}$.

1 Ensembles Convexes

Soit $x, y \in E$, on appelle **segment d'extrémité** x **et** y la partie

$$[x, y] = \{\lambda x + (1 - \lambda) y, \lambda \in [0, 1]\}.$$

On retrouve bien sûr la définition usuelle du segment dans $\mathbb{R}$. (avec la notation inhabituelle $[-1, -2] = [-2, -1]$)

★ **Définition 3.1.** Un ensemble $C \subset E$ est dit **convexe** si $\forall x, y \in C, [x, y] \subset C$.

Par convention , $C = \emptyset$ est convexe même si les convexes intéressants sont les convexes non-vides...

Proposition 3.1. Si E est un e.v.n., les boules (ouvertes et fermés) sont des convexes.

Démonstration. Considérons le cas des boules ouvertes. Soient $x, y \in B(a, r)$, $z = \lambda x + (1-\lambda)y$, $\lambda \in [0, 1]$.

Par l'inégalité triangulaire et homogénéité, on a :

$$\begin{aligned} ||z-a|| &= ||\lambda(x-a) + (1-\lambda)(y-a)|| \\ &\le |\lambda| ||x-a|| + |1-\lambda| ||y-a|| < |\lambda| r + |1-\lambda| r = r. \end{aligned}$$

Donc $z \in B(a, r)$. Le cas des boules fermées est similaire. □

Exemple 3.1. On pose $||(x, y)||_{1/2} = (|x|^{1/2} + |y|^{1/2})^2$. On note $B = \{(x, y) \in \mathbb{R}^2 : ||(x, y)||_{1/2} \le 1\}$. On remarque que $(1, 0), (0, 1) \in B$, $(1/4, 1/4) \in B$ mais $(1/2, 1/2) \notin B$ donc B n'est pas convexe et $|| \cdot ||_{1/2}$ n'est PAS une norme sur $\mathbb{R}^2$.

Exercice 3.1. Montrer que les ensembles convexes de $\mathbb{R}$ sont exactement les intervalles.

Le résultat suivant est laissé en exercice.

Proposition 3.2. Si C est convexe, alors son adhérence $\overline{C}$ et son intérieur $Int(C)$ sont convexes. Une intersection (finie ou infinie) d'ensembles convexes est convexe. Si $C_1 \subset E$, $C_2 \subset F$ sont convexes, alors $C_1 \times C_2$ est convexe dans $E \times F$.

Cônes tangents et normaux dans $\mathbb{R}^n$

On suppose $E = \mathbb{R}^n$ (ou un espace préhilbertien comme au dernier chapitre pour avoir un produit scalaire). On rappelle $\langle f, x\rangle = \sum_{i=1}^{n} f_i x_i$, pour $f, x \in E$.

Les deux ensembles suivant seront importants pour formuler des conditions pour des problèmes de minimisation sous contrainte. On rappelle que pour $A, B \subset E$, $C \subset \mathbb{R}$, $x \in E$, $A + B = \{a + b : a \in A, b \in B\}$, $CA = \{ca, c \in C, a \in A\}$, $A - x = \{a - x : a \in A\}$, $x + A = \{a + x : a \in A\}$.

★ **Définition 3.2.** Le **cône tangent** (au sens de l'analyse convexe) du convexe $S \subset E$ e.v.n. au point $x \in S$ est

$$T_S(x) := \overline{\{\frac{u-x}{s}, u \in S, s > 0\}} = \overline{\mathbb{R}_+^*(S-x)},$$

Le **cône normal** est son polaire, c'est à dire le cône convexe fermé :

$$\begin{aligned} N_S(x) &:= \{f \in E : \forall u \in S, \langle f, u - x\rangle \leq 0\} \\ &= \{f \in E : \forall v \in T_S(x)\langle f, v\rangle \leq 0\}. \end{aligned}$$

Exercice 3.2. Si L est un s.e.v de E (de dimension finie), $a \in L$. Montrer que $T_L(a) = L$ et $N_L(a) = L^{\perp} = \{y \in E : \langle y, a\rangle = 0 \forall y \in L\}$, est l'orthogonal de L.

Exercice 3.3. Si S convexe et $a \in Int(S)$. Montrer que $T_S(a) = E$ et $N_L(a) = \{0\}$.

2 Fonctions convexes

Il est pratique de considérer des fonctions $f : E \to]-\infty, +\infty] = \mathbb{R} \cup \{+\infty\}$. Dans ce cas on parle de **domaine de** f :

$$D(f) = \{x \in E : f(x) < \infty\}.$$

Les propriétés que l'on considère dans cette section vont être déterminées par l'ensemble des valeurs au dessus du graphe de f, que l'on appelle **épigraphe de f** :

$$\text{Epi}(f) = \{(x, \lambda) \in E \times \mathbb{R} : f(x) \leq \lambda\}.$$

On utilise les conventions $\infty + \infty = \infty$ et $\lambda . \infty = \infty$ si $\lambda > 0$, $0 . \infty = 0$.

★ **Définition 3.3.** Soit C un ensemble convexe.

1. Une fonction $f : C \to]-\infty, +\infty]$ est dite **convexe** si pour tout $\lambda \in]0, 1[, x, y, \in C$,

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y).$$

2. Une fonction $f : C \to]-\infty, +\infty]$ est dite **strictement convexe** si pour tout $\lambda \in]0, 1[, x, y, \in C$, avec $x \neq y$

$$f(\lambda x + (1 - \lambda) y) < \lambda f(x) + (1 - \lambda) f(y).$$

3. Une fonction $f : C \to [-\infty, +\infty[$ est dite **concave** si $-f$ est convexe.

Exemple 3.2. Une fonction affine $f(x_1, \dots, x_n) = \sum_{i=1}^{n} a_i x_i + b$ est convexe et concave mais pas strictement convexe ! Une norme sur E est convexe.

Remarque 3.1. Si f est convexe, alors $C \cap D(f)$ est convexe car si $f(x) < +\infty$, $f(y) < \infty$ alors

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) < \infty.$$

On peut donc toujours remplacer soit C par E soit C par $C \cap D(f)$ selon votre goût (pour les fonctions infinies ou les ensembles convexes).

Proposition 3.3. Soit E un e.v. et $f : C \to]-\infty, \infty]$.

1. f est convexe si et seulement si $Epi(f)$ est convexe

1'. Si f est convexe alors pour tout $t \in \mathbb{R}$, $f^{-1}(]-\infty, t])$ est convexe. La réciproque est fausse.

2. Si $\mu > 0$, f, g convexes alors $\mu f + g$ est convexe. De plus, elle est aussi strictement convexe si f ou g l'est.
3. Si f_i, $i \in I$ sont convexes alors l'enveloppe supérieure $f(x) = \sup_{i \in I} f_i(x)$ est convexe.
4. (facultatif) f est convexe ssi $g : E \to]-\infty, +\infty]$, définie par $g(x) = f(x)$ si $x \in C$ et $g(x) = +\infty$ sinon, est convexe.
5. Si f est strictement convexe, alors f a au plus un minimum sur C.

Le dernier point donne la première relation simple des fonctions convexes à l'optimisation.

Démonstration. Pour (1), l'énoncé est vide si $f(x)$ ou $f(y) = \infty$. Soit donc $(x, t_1), (y, t_2) \in Epi(f)$ (comme on veut $t_i < \infty$ cela utilise la réduction précédente). On remarque que $(\lambda x + (1-\lambda)y, \lambda t_1 + (1-\lambda)t_2) \in Epi(f)$ ssi $f(\lambda x + (1-\lambda)y) \leq \lambda t_1 + (1-\lambda)t_2$.

Si les épigraphes sont convexes, cette propriété est vérifiée et donc en prenant l'infimum sur t_1, t_2 (qui donne $f(x), f(y)$) on a le résultat. Si f vérifie l'inégalité, on utilise $f(x) \leq t_1, f(y) \leq t_2$ pour conclure :

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \leq \lambda t_1 + (1-\lambda)t_2.$$

(1)' On montre la convexité de $D = \{x : f(x) \leq t\}$ comme ci-dessus. Soit $x, y \in D$ alors pour $\lambda \in [0,1]$: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \leq \lambda t + (1-\lambda)t = t$. Donc $\lambda x + (1-\lambda)y \in D$.

Par contre si $g = 1_{[0,\infty[}$ alors si $t < 0$, $g^{-1}(]-\infty, t]) = \emptyset$, si $0 \le t < 1$, $g^{-1}(]-\infty, t]) =]-\infty, 0[$ et sinon pour $t \ge 1$, $g^{-1}(]-\infty, t]) = \mathbb{R}$ et ce sont 3 intervalles donc 3 ensembles convexes. Mais g n'est pas convexe $g(0) = 1 > 1/2g(-1) + 1/2g(1) = 1/2$.

(2) est évident en utilisant l'inégalité :

$$\begin{aligned}
&\mu f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y) \\
&\quad \le \mu(\lambda f(x) + (1-\lambda)f(y)) + (\lambda g(x) + (1-\lambda)g(y)) \\
&\quad = (\lambda(\mu f + g)(x) + (1-\lambda)(\mu f + g)(y)).
\end{aligned}$$

(3) vient de la stabilité des convexes par intersection et de $Epi(f) = \cap_{i\in I} Epi(f_i)$.

(4) est évident car $Epi(f) = Epi(g)$.

(5) si $x \neq y$ sont deux points atteignant le minima, $f((x+y)/2) < (f(x) + f(y))/2$ contredisant la minimalité. □

Une propriété importante des fonctions convexes est le fait qu'on peut les caractériser en terme d'accroissements :

Proposition 3.4. Soit $f : E \to]-\infty, +\infty]$ une fonction. f est convexe si et seulement si pour tout $x, h \in E$ la fonction $\Delta_{x,h}f(t) := \frac{f(x+th)-f(x)}{t}$ est croissante sur $\mathbb{R}_+^*$.

Démonstration. Il suffit de noter que $g(t) = \Delta_{x,h}f(t) = \frac{f(x+th)-f(x)}{t}$ est croissante si et seulement si $g(t) \le g(s)$ pour $0 < t < s$ si et seulement si on a l'inégalité de convexité :

$$f(x+th) = f(\frac{t}{s}(x+sh) + x(1-\frac{t}{s})) \le f(x+sh)\frac{t}{s} + f(x)(1-\frac{t}{s}).$$

Donc la convexité de f implique la croissance énoncée et réciproquement en prenant $s = 1$ on écrit toute paire x, y sous la forme $y = x + h$ et l'inégalité ci-dessus se réécrit en l'inégalité définissant la convexité de f :

$$\begin{aligned}
f((1-t)x + ty) = f(x+th) &\le f(x+h)t + f(x)(1-t) \\
&= f(y)t + f(x)(1-t).
\end{aligned}$$

□

Cela implique une régularité minimale des fonctions convexes :

Corollaire 3.5. Si $f : E \to]-\infty, \infty]$ est convexe, pour tout $x \in D(f)$ et tout $h \in E$, la dérivée directionnelle $D'_h f(x)$ existe dans $[-\infty, \infty]$ au sens où la limite suivante existe et vaut :

$$D'_h f(x) := \lim_{t \to 0^+} \frac{f(x+th) - f(x)}{t} = \inf_{t>0} \frac{f(x+th) - f(x)}{t}.$$

Démonstration. Par la proposition précédente $g(t) = \frac{f(x+th)-f(x)}{t}$ est croissante donc admet une limite pour $t \to 0^+$ qui coïncide avec l'infimum. □

Calcul des cônes normaux courants

Soient $g_1, \ldots, g_n$ des fonctions convexes C^1 définies $U \to \mathbb{R}$ avec U **ouvert convexe** tel qu'il existe $x_0 \in U$ avec $g_i(x_0) < 0$ pour tout i.

Soit la contrainte :

$$C = \{x \in U : \forall i \in \{1, \ldots, n\}, g_i(x) \leq 0\}.$$

On sait que chaque $g_i^{-1}(]-\infty, 0])$ est convexe comme image réciproque d'un intervalle borné supérieurement par une application convexe. Par intersection, on sait donc que $C = \cap_{i=1}^n g_i^{-1}(]-\infty, 0]) \subset U$ est aussi convexe.

★ **Théorème 3.6** (admis, cf Section B.1)**.** Soit $x \in C$ tel que :

1. les l premières contraintes sont actives, c'est à dire : $g_1(x) = \ldots = g_l(x) = 0$
2. les autres contraintes ne sont pas actives, c'est à dire $g_{l+1}(x) < 0, \ldots g_n(x) < 0$

Si $l = 0$, on a $N_C(x) = \{0\}$ et sinon, le cône normal à C en x est donné par

$$N_C(x) = \left\{ \sum_{i=1}^{l} \lambda_i \nabla g_i(x), \lambda_i \geq 0 \right\}.$$

Exemple 3.3. Soit $A = \{(x, y) \in \mathbb{R}^2 : x \geq y \geq 0, \}$. Si on pose $g_1(x, y) = y - x$, $g_2(x, y) = -y$ qui sont linéaires donc convexes et C^1, on a :

$$A = \{(x, y) \in \mathbb{R}^2 : g_1(x, y) \leq 0, g_2(x, y) \leq 0\}$$

Calculons $N_A(0)$ le cône normal en $0 = (0, 0)$.
On a $g_1(0, 0) = 0 = g_2(0, 0)$ donc toutes les contraintes sont actives.
On calcule donc $\nabla g_1(0, 0) = (-1, 1)$, $\nabla g_2(0, 0) = (0, -1)$. D'après le théorème, on a :

$$N_A(0) = \mathbb{R}_+(-1, 1) + \mathbb{R}_+(0, -1).$$

Exercice 3.4. 1. Pour A de l'exemple précédent, si $a=(x,x)$ pour $x>0$. Montrer que $N_A(a)=\mathbb{R}_+(-1,1)$.
2. Pour $b=(x,0)$, $x>0$. Montrer que $N_A(b)=\mathbb{R}_+(0,-1)$.
3. Y-a-t-il d'autres valeurs de $N_A(c)$ et si oui, pour quels points $c\in A$?

Fonctions convexes sur $\mathbb{R}$

Soit I un intervalle de $\mathbb{R}$. Pour une fonction $f:I\to\mathbb{R}$ et $a\in I$, on considère la fonction (taux d'accroissement de f en a) $\Delta_a f$ définie par $\Delta_a f(x)=\dfrac{f(x)-f(a)}{x-a}$ pour tout $x\in I\setminus\{a\}$. La proposition 3.4 se reformule sous la forme :

Proposition 3.7. Une fonction $f:I\to\mathbb{R}$ est convexe si et seulement si pour tout $a\in I$, la fonction $\Delta_a f$ est croissante sur $I\setminus\{a\}$.

On en déduit les inégalités suivantes (inégalité des pentes, cf dessin en cours) sur une fonction f :

★ **Proposition 3.8.** Une fonction convexe $f:I\to\mathbb{R}$ vérifie l'**inégalité des pentes** :

$$\forall a,\ b,\ c\in I,\ \ a<b<c\Rightarrow \frac{f(b)-f(a)}{b-a}\leq\frac{f(c)-f(a)}{c-a}\leq\frac{f(c)-f(b)}{c-b}.$$

★ **Théorème 3.9.** Soit I un intervalle ouvert de $\mathbb{R}$, et $f: I\to\mathbb{R}$ une fonction convexe. Alors pour tout $a\in I$, f admet des dérivées à droite et à gauche en a. On a pour tout $x\in I$: $f(x)\geq f_d'(a)(x-a)+f(a)$ et $f(x)\geq f_g'(a)(x-a)+f(a)$. En particulier, il existe une fonction affine g telle que $g(a)=f(a)$ et $g(x)\leq f(x)$ pour tout $x\in I$. De plus, si $a<b$ sont dans I, on a $f_g'(a)\leq f_d'(a)\leq f_g'(b)$.

Démonstration. Soit $a\in I$. Dans le cas d'une fonction à une variable, le corollaire 3.5 implique l'existence de dérivées à droites et à gauches (pour l'instant peut-être infinies). Dans l'inégalité des pentes en faisant $c\to b^+$ ou $a\to b^-$, on obtient :

$$-\infty<\frac{f(b)-f(a)}{b-a}\leq f_d'(b),$$
$$f_g'(b)\leq\frac{f(c)-f(b)}{c-b}<+\infty.$$

Pour $a<b$, $0<\epsilon_i<(b-a)/2$, l'inégalité des pentes appliquée aux points $a\leq a+\epsilon_1<b-\epsilon_2<b$ donne :

$$\frac{f(a+\epsilon_1)-f(a)}{\epsilon_1}\leq\frac{f(b-\epsilon_2)-f(a+\epsilon_1)}{(b-a-\epsilon_1-\epsilon_2)}\leq\frac{f(b-\epsilon_2)-f(b)}{-\epsilon_2}$$

et en passant à la limite $\epsilon_1 \to 0^+$ ou $\epsilon_2 \to 0^+$ puis les deux, on obtient :

$$f'_d(a) \leq \frac{f(b-\epsilon_2) - f(b)}{-\epsilon_2},$$

$$\frac{f(a+\epsilon_1) - f(a)}{\epsilon_1} \leq f'_g(b),$$

$$f'_d(a) \leq f'_g(b).$$

Donc $f'_d(a) < +\infty$, $f'_g(a) > -\infty$, ce qui termine la preuve des dérivabilités à droite et à gauche, et on a l'inégalité attendue.

De plus, la formulation comme infimum, dans le corollaire 3.5, montre que pour tout $x > a$ que $\dfrac{f(x) - f(a)}{x-a} \geq f'_d(a)$ et donc $f(x) \geq f'_d(a)(x-a) + f(a)$. De même, pour tout $x < a$ on a $\dfrac{f(x) - f(a)}{x-a} \leq f'_g(a)$; en multipliant par $x - a$ (qui est négatif!) on a donc que pour tout $x < a$ $f(x) \geq f(a) + f'_g(a)(x-a)$.

De plus, $f'_g(b) \leq f'_d(b)$ (en passant aux limites $a \to b^-$, $c \to b^+$ dans l'inégalité des pentes) ; par conséquent, pour $x < a$ $f'_g(a)(x-a) \geq f'_d(a)(x-a)$, et on voit finalement que l'inégalité $f(x) \geq f'_d(a)(x-a) + f(a)$ est valide pour tout $x \in \mathbb{R}$. Le même raisonnement s'applique pour montrer que l'autre inégalité est vraie pour tout $x \in \mathbb{R}$. □

Corollaire 3.10. Soit I un intervalle ouvert de $\mathbb{R}$, alors une fonction convexe $f : I \to \mathbb{R}$ est continue.

Exercice 3.5. Trouver une fonction convexe $f : [0, 1[\to \mathbb{R}$ qui n'est pas continue en $\{0\}$.

Proposition 3.11. Si $E = \mathbb{R}$ et f est dérivable sur un ouvert convexe $U \subset E$ (donc un intervalle ouvert) alors f est convexe si et seulement si f' est croissante.

Démonstration. $\Rightarrow$) Supposons f convexe, l'inégalité qu'on a montrée au (2) du théorème précédent s'écrit $(f'(u) - f'(v))(u-v) \geq 0$ donc $(f'(u) - f'(v)), (u-v)$ ont même signe et f' est croissante. On peut alternativement utilisé pour $a < b$, $f'(a) = f'_d(a) \leq f'_g(b) = f'(b)$ grâce à l'inégalité vue au théorème 3.9.

$\Leftarrow$) Réciproquement si f' croissante, montrons que f convexe, on veut voir $f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda)f(b)$ pour $a < b, \lambda \in]0, 1[$. Par l'égalité des accroissements finis, la pente $\frac{f(\lambda a+(1-\lambda)b)-f(a)}{(1-\lambda)(b-a)}$ est atteinte par f' en un point de $]a, \lambda a + (1-\lambda)b[$, et de même $\frac{f(b)-f(\lambda a+(1-\lambda)b)}{(\lambda(b-a)}$ est atteinte par f' en un point de $]\lambda a + (1-\lambda)b, b[$ donc par croissance de la dérivée :

$$\frac{f(\lambda a+(1-\lambda)b)-f(a)}{(1-\lambda)(b-a)} \leq \frac{f(b)-f(\lambda a+(1-\lambda)b)}{\lambda(b-a)}$$
$$\Longleftrightarrow f(\lambda a+(1-\lambda)b)(\frac{1}{(1-\lambda)(b-a)}+\frac{1}{\lambda(b-a)}) \leq \frac{f(a)}{(1-\lambda)(b-a)}+\frac{f(b)}{\lambda(b-a)}$$
$$\Longleftrightarrow f(\lambda a+(1-\lambda)b)(\frac{1}{\lambda(1-\lambda)}) \leq \frac{f(a)}{(1-\lambda)}+\frac{f(b)}{\lambda}.$$

Ceci conclut. □

3 Propriétés différentielles des fonctions convexes.

Rappel sur la différentiabilité (au sens de Fréchet)

On rappelle que pour E, F des e.v.n. l'ensemble des applications linéaires continues $L(E,F)$ est un e.v.n. avec la norme d'opérateur (dite aussi norme subordonnée) $|||f||| = \sup_{||x||_E \leq 1} ||f(x)||_F$.

Définition 3.4. Soit E, F des e.v.n., $U \subset E$ un ouvert, $f : U \to F$ est différentiable (au sens de Fréchet) en x si il existe $T \in L(E,F)$ notée $df(x)$ telle que

$$||f(x+h) - f(x) - df(x)(h)|| = o(||h||), \quad si \quad ||h|| \to 0.$$

f est C^1 (ou continuement différentiable) sur U si f est différentiable en tout $x \in U$ et $df : U \to L(E,F)$ est continue. On note aussi $D_h f(x) = df(x)(h)$
f est C^2 si f est C^1 et df est aussi C^1. On note $d^2 f(x)(h,k) = D_k(D_h f)(x)$.

On rappelle que si $g : U \to V \subset F$, $f : V \to Z$ sont différentiables, alors $f \circ g$ aussi et $d(f \circ g)(x) = df(g(x)) \circ dg(x)$. De plus si $Z = \mathbb{R}$ et f a un minimum local en $x \in V$ avec V ouvert, alors $df(x) = 0$.

Remarque 3.2. Il est important de noter que $df(x)$ est une application linéaire, donc $df(x)(h)$ est linéaire en h, mais pas forcément en x. Pour insister sur ce point, on note parfois de façon équivalente :

$$df(x)(h) \equiv df(x).h \equiv df(x).[h]$$

Dans le cas le plus fréquent pour nous où $E = \mathbb{R}^n$, $F = \mathbb{R}$, si f est différentiable, alors elle admet des dérivées partielles, le gradient de f en a est noté $\nabla f(a) = (\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a))$.

Alors, on a :

$$df(a)(h) = \langle \nabla f(a), h \rangle = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) h_j.$$

Caractérisations différentielles des fonctions convexes

Le théorème suivant résume les 3 caractérisations principales de la convexité en terme de différentiabilité, par la position relative des plans tangents et du graphe, par la monotonie de la dérivée première ou par la positivité de la dérivée seconde (le résultat n'est pas optimal, il suffit en fait d'une dérivabilité directionnelle appelée dérivée au sens de Gâteaux) :

★ **Théorème 3.12.** Soit E un e.v.n. et U un ouvert convexe, $f : U \to \mathbb{R}$ une fonction différentiable en tout point de U.

1. f est convexe ssi pour tout $u, v \in U$:

$$f(u) - f(v) \geq df(v).[u - v]$$

2. f est convexe ssi pour tout $u, v \in U$:

$$[df(u) - df(v)].[u - v] \geq 0$$

3. Si f est en plus C^2, f est convexe ssi $d^2 f(x)$ est positive pour tout $x \in U$ au sens où $d^2 f(x)(h, h) \geq 0$ pour tout $x \in U$, $h \in E$. De plus, si $E = \mathbb{R}^n$ avec la norme euclidienne, ou plus généralement si E est préhilbertien (cf. chapitre 5), si $d^2 f(x)$ est définie positive, pour tout $x \in U$ (c'est-à-dire pour tout $h \neq 0$, $d^2 f(x)(h, h) > 0$) alors f est strictement convexe.

Remarque 3.3. (Rappel d'algèbre linéaire) Si $E = \mathbb{R}^n$, alors $d^2 f(x)$ est positive si et seulement si la matrice hessienne $Hf(x)$ est positive (rappel $(Hf(x))_{ij} = (\frac{\partial^2 f}{\partial x_i \partial x_j}(x))$). Comme elle est toujours symétrique et donc diagonalisable en base orthonormale, cela équivaut à ce que ces valeurs propres soient toutes positives. Dans le cas $n = 2$ $H(f)(x) = \begin{pmatrix} r & s \\ s & t \end{pmatrix}$ (c'est à dire on prend les notations de Monge $r = \frac{\partial^2 f}{\partial x^2}(x)$, $s = \frac{\partial^2 f}{\partial x \partial y}(x)$, $t = \frac{\partial^2 f}{\partial y^2}(x)$) alors $H(f)(x)$ est positive si et seulement si $rt - s^2 \geq 0$ et $r \geq 0$. [a]

---

a. En effet $D^2 f(x)((h_1, h_2), (h_1, h_2)) = rh_1^2 + 2sh_1h_2 + th_2^2 = (h_1^2)P(h_2/h_1)$ si $h_1 \neq 0$, avec $P(\lambda) = r + 2s\lambda + t\lambda^2$ le polynôme de second degré de discriminant $\Delta = 4s^2 - 4rt$. Si $\Delta < 0$ pas de racine et selon le signe de r, P est soit toujours positif (cas $D^2 f(a)$ définie positive) soit toujours négative ($D^2f(a)$ définie négative). Si $\Delta = 0$, il y a une racine double et on a la même conclusion sur la positivité. Si $h_1 = 0$, alors $D^2f(x)((h_1, h_2), (h_1, h_2))) = 2sh_1h_2$ n'est positive que si $s = 0$ car sinon en $(h_1, h_2) = (s, -1)$, on a la valeur strictement négative $-2s^2$ et c'est aussi le seule cas ou le déterminant $rt - s^2$ est positif pour $r = 0$). Si $\Delta > 0$ on a 2 racines réelles et P prend à la fois des valeurs positives et négatives.

Remarque 3.4. Un cas particulier du (3) est le cas où il existe $c > 0$ telle que $d^2f(x)(h, h) \geq c||h||^2$ pour tout $x \in U$, $h \in E = \mathbb{R}^n$. Le cas de stricte convexité se déduit donc en décomposant $f = g + \frac{c}{2}||x||^2$. L'inégalité donne que $d^2g = d^2f - c$ est positive donc g convexe et on verra au dernier chapitre que l'identité du parallélogramme implique que $\frac{c}{2}||x||^2$ est strictement convexe, donc par somme f est strictement convexe (de façon très uniforme). C'est une situation intéressante pour les problèmes de minimisation qui permet d'obtenir la convergence de suites minimisantes et des stratégies algorithmiques de minimisation (cf. cours de recherche opérationnelle au S6).

Démonstration. (1) Si f convexe, l'inégalité vient du corollaire 3.5 en comparant l'infimum à la valeur en $t = 1$ pour $h = u - v$:

$$df(v).[u - v] = \inf_{t>0} \frac{f(v + th) - f(v)}{t} \leq f(u + h) - f(u) = f(u) - f(v).$$

Réciproquement on applique l'inégalité en $z = tx + (1 - t)y \in U$ par convexité de U pour $x, y \in U$ d'où :

$$(A)\ f(x) - f(z) \geq df(z)[x - z], \quad (B)\ f(y) - f(z) \geq df(z)[y - z],$$

et $t(A) + (1 - t)(B)$ donne

$$tf(x) + (1 - t)f(y) - f(z) \geq df(z)[t(x - z) + (1 - t)(y - z)] = df(z)(0) = 0$$

ce qui donne l'inégalité de convexité.

(2) Si f convexe, on utilise de même les inégalités du corollaire 3.5 :

$$df(u)(v - u) \leq f(v) - f(u), \quad df(v)(u - v) \leq f(u) - f(v)$$

En sommant, on obtient l'inégalité $(df(u) - df(v))(v - u) \leq 0$. Réciproquement, on utilise $\phi(t) = f(tx + (1 - t)y)$ qui par composition est dérivable de dérivée

$\phi'(t) = df(tx + (1-t)y)(x-y)$. Or si $t < s$

$$\begin{aligned}
&\phi'(s) - \phi'(t) \\
&= [df(y + s(x-y)) - df(y + t(x-y))](x-y) \\
&= \frac{1}{s-t}[df(y + s(x-y)) - df(y + t(x-y))] \\
&\qquad (y + s(x-y) - (y + t(x-y))) \geq 0
\end{aligned}$$

Donc ϕ' est croissante et par un résultat à 1 variable (proposition 3.12) ϕ est convexe.

(3)Si f est C^2, on dérive en t la relation du (2) avec $v = x$, $u = x + th$ une fois divisée par t^2 et on obtient $d^2f(x)(h,h) \geq 0$. Réciproquement, en dérivant en t, la fonction g définie par $g(t) = df(v + t(u-v))(u-v)$ (qui est C^1 car df est C^1) et en appliquant le théorème fondamental du calcul :

$$[df(u) - df(v)][u-v] = g(1) - g(0) = \int_0^1 dt df(v + t(u-v))(u-v, u-v) \geq 0$$

et on retrouve le critère du (2).

Pour la stricte convexité, commençons par le cas $E = \mathbb{R}$, donc $U = I$ un intervalle ouvert. Soit $[a,b] \subset I$ il suffit de voir f strictement convexe sur $[a,b]$. On fixe $[a,b] \subset]a',b'[\subset [a',b'] \subset I$

On suppose dans ce cas $f''(x) > 0$ pour tout $x \in I$ et f'' continue (vue f de classe C^2). Donc f'' atteint son minimum sur $[a',b']$ en x_0 de sorte que $f''(x) \geq c = f''(x_0) > 0$ pour tout $x \in]a',b'[\subset [a',b']$. Donc comme à la remarque 3.4 implique $f = g + c\frac{x^2}{2}$ avec $g'' \geq 0$ donc g convexe et donc f strictement convexe sur $]a',b'[$.

On pose $g_{a,b}(t) = ta + (1-t)b$. Soit maintenant le cas général $E = \mathbb{R}^n$. Par définition, f est strictement convexe si et seulement si pour tout segment $[a,b] \subset U$, $a \neq b$, $h_{a,b} = f \circ g_{a,b}$ est strictement convexe sur $[0,1]$ (ou sur $]0,1[$ en élargissant les intervalles comme avant). Or $h''_{a,b}(t) = df^2(g_{a,b}(t))(a-b, a-b) > 0$ pour tout $t \in]0,1[$. On déduit donc du premier cas que $h_{a,b}$ est strictement convexe sur $]0,1[$ et donc aussi f. Comme U ouvert, on peut trouver $a', b' \in U$ avec $[a,b] \subset [a',b'] - \{a',b'\}$, $[a',b'] \subset U$.

Pour montrer Comme $g_{a',b'}$ est continue bijective de $[0,1] \to [a',b']$ si $a' \neq b'$, $[a',b']$ est compact comme image direct du compact $[0,1]$ par une application continue.

$t \mapsto h''_{a,b}(at + (1-t)(b-a)) = d^2f(at + (1-t)(b-a))(b-a, b-a)$ est continue sur $[a',b']$ donc atteint son minimum en $x_0 \in [a',b']$ qui est donc $h''_{a,b}(x) = d^2f(x_0)(b-a, b-a) \geq c_{x_0}(b-a, b-a)$. En appliquant à l'intervalle ouvert $]a',b'[$ le premier cas, on déduit que $h_{a',b'}$ est strictement convexe sur $]a', b'[$, donc aussi par restriction $h_{a,b}$. Comme $a \neq b \in U$ arbitraires, f est aussi strictement convexe. □

Exercice 3.6. Montrer que $f(x) = x^4$ est strictement convexe sur $\mathbb{R}$ mais que sa dérivée seconde n'est pas bornée inférieurement par $c > 0$.

Convexité, Critère d'extremum global

On retrouve d'abord un critère d'optimisation du premier ordre

Proposition 3.13. Si f est de classe $\mathcal{C}^1$ sur un ouvert convexe U et f est convexe, alors tout point $a \in U$ est un minimum global de f si et seulement si c'est un point critique de f (c'est à dire un point a tel que $df(a) = 0$).

Démonstration. On sait déjà par le cours de L2 que si f a un minimum local en a alors $df(a) = 0$. En effet, rappelons la preuve, pour tout $h \in E$, il existe $\epsilon > 0 : B(a, \epsilon||h||) \subset U$ (car U ouvert) et $f(a \pm th) \geq f(a)$ pour tout $t \in]-\epsilon, \epsilon[$. Donc, en divisant par $t > 0$ on obtient :

$$\frac{f(a+th) - f(a)}{t} \to_{t \to 0^+} df(a)(h) \geq 0$$

$$\frac{f(a-th) - f(a)}{-t} \to_{t \to 0^+} df(a)(h) \leq 0$$

donc $df(a)(h) = 0$ pour tout h ce qui veut dire $df(a) = 0$.

La nouveauté est la réciproque, on suppose f convexe. Il suffit de noter par le théorème 3.12 que pour $c \in C$, $f(c) - f(a) \geq df(a)(c-a) = 0$ donc $f(a) = \inf_{c \in C} f(c)$ et a atteint l'infimum de f sur C. □

On a un critère d'optimisation plus général sur un convexe $C \subset \mathbb{R}^n$. On rappelle que $\nabla f(a) = (\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a))$.

★ **Théorème 3.14.** Soit C un convexe de $\mathbb{R}^n$ avec $C \subset U$ un ouvert et $f : U \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$, convexe sur C. Alors a est un minimum global de f sur C si et seulement si $-\nabla f(a) \in N_C(a)$ c'est à dire si et seulement si

$$\forall c \in C, \langle \nabla f(a), c - a \rangle \geq 0.$$

Démonstration. On rappelle la définition $N_C(a) = \{f \in E : \forall c \in S, \langle f, c-a\rangle \leq 0\}$ ce qui donne la dernière reformulation. Si a est un minimum global $f(a) \leq f(tc + (1-t)a)$ pour $c \in C$, $t \in]0,1[$ vu que par convexité $tc + (1-t)a \in C$. En prenant la limite, on obtient

$$\langle \nabla f(a), c-a\rangle = \lim_{t\to 0^+} \frac{f(t(c-a)+a) - f(a)}{t} \geq 0$$

Réciproquement, si l'inégalité est vérifiée donc on peut utiliser le théorème 3.12 (dont la preuve du 1 s'applique même si C n'est pas ouvert) et on obtient :

$$0 \leq \langle \nabla f(a), c-a\rangle = df(a)(c-a) \leq f(c) - f(a).$$

donc $f(c) \geq f(a)$ pour tout $c \in C$ et donc a est un minimum de f sur C. □

Exemple 3.4. On prend $g(c) = ||f-c||_2^2$ le carré de la distance euclidienne à $f \in E$. Alors $\nabla g(a) = -2(f-a)$ et donc on obtient que $a \in C$ minimise la distance de x à C si et seulement si :

$$\forall c \in C, \langle x-a, c-a\rangle \leq 0.$$

Ce sera le critère du théorème de projection sur un convexe fermé C où l'on verra l'existence d'un tel point a au dernier chapitre. Dans $\mathbb{R}^n$ on peut aussi voir l'existence par compacité de $C \cap B$ pour une boule fermée B assez grande pour qu'une inégalité grossière permette d'assurer que tout minimum doive s'y trouver. On obtient ainsi le résultat suivant.

★ **Théorème 3.15** (théorème de projection sur un convexe fermé de $\mathbb{R}^n$)**.** Soit $C \subset \mathbb{R}^n = E$ un convexe fermé non–vide et $||.||_2$ la norme euclidienne. Pour tout $f \in \mathbb{R}^n$, il existe un unique $u = P_C(f)$ tel que

$$||f-u||_2 = \inf_{v\in C} ||f-v||_2.$$

De plus, c'est l'unique vecteur $u \in C$ tel que :

$$\forall v \in C, \langle f-u, v-u\rangle \leq 0.$$

De plus, pour tout $c \in C$, $c + N_C(c) = P_C^{-1}(\{c\})$ et forment une partition de $\mathbb{R}^n$.

La preuve suivante par compacité ne fonctionnera pas en dimension infinie, mais le résultat sera encore vrai dans un espace de Hilbert (cf. chapitre 5).

Démonstration. Comme C non vide $r = \inf_{v\in C} ||f-v||_2 < \infty$. Soit $D = C \cap \overline{B(f, r+1)}$. Comme la boule fermé est un convexe fermé, D est un convexe fermé comme intersection de convexes fermés, et il est aussi borné par définition, donc c'est un compact de $\mathbb{R}^n$. De plus, $D \subset C$, donc $\inf_{v\in C} \|f-v\|_2 \leq \inf_{v\in D} \|f-v\|_2$ par définition de l'infimum. Mais soit $1 > \epsilon > 0$ et $v \in C$ tel que $\|f-v\|_2 \leq r+\epsilon$ alors par définition $v \in D$ et donc $\inf_{d\in D} \|f-d\|_2 \leq \|f-v\|_2 \leq r+\epsilon$. Donc en passant à la limite $\epsilon \to 0$, on a obtenu :

$$\inf_{v\in D} \|f-v\|_2 \leq r = \inf_{v\in C} \|f-v\|_2 \leq \inf_{v\in D} \|f-v\|_2.$$

Or $v \mapsto \|f-v\|_2$ est continue sur le compact D, donc atteint son infimum en $u \in D \subset C$. Par croissance du carré, c'est aussi le point où $\|f-v\|_2^2$ atteint son infimum. La hessienne de $v \mapsto \|f-v\|_2^2$ est l'identité, donc cette application est strictement convexe, elle a donc un unique minimum $P_C(f)$. La caractérisation du minimum a été vue à l'exemple précédent. Enfin cette caractérisation donne (en retraduisant avec la définition de $N_C(c)$)

$$\begin{aligned} P_C^{-1}(\{c\}) &= \{f \in E : \forall v \in C, \langle f-c, v-c\rangle \leq 0\} \\ &= \{f \in E : f-c \in N_C(c)\} = c + N_C(c). \end{aligned}$$

Le fait que $P_C : E \to C$ est une application surjective (vu que $P_C(c) = c$ pour $c \in C$) implique le résultat sur la partition. □

4 Premières Inégalités de convexité

Citons un exemple important et simple.

Exercice 3.7. Soit $f\colon [0,+\infty[\to [0,+\infty[$ une fonction concave. Montrer que pour tout $x, y \geq 0$ on a $f(x+y) \leq f(x)+f(y)$.

Démonstration. Fixons $y \geq 0$ et considérons la fonction $g\colon [0,+\infty[\to \mathbb{R}$ définie par $g(x) = f(x)+f(y)-f(x+y)$.

Alors, pour tout $a < b \in [0,+\infty[$, on a

$$\frac{g(b)-g(a)}{b-a} = \frac{f(b)-f(a)}{b-a} - \frac{f(b+y)-f(a+y)}{b-a}.$$

Puisque $\dfrac{f(b+y)-f(a+y)}{b-a} = \dfrac{f(b+y)-f(a+y)}{(b+y)-(a+y)}$ est le taux d'accroissement de f entre $(b+y)$ et $(a+y)$, l'inégalité des pentes nous donne donc que $\dfrac{g(b)-g(a)}{b-a} \geq 0$, autrement dit g est croissante.

Par conséquent, on a pour tout x que $g(x) \geq g(0) = f(0)$, et donc $f(x)+f(y)-f(x+y) \geq f(0) \geq 0$, ce qu'on voulait démontrer. □

On verra au chapitre intégration section 5.2 l'inégalité la plus importante, l'inégalité de Jensen, qu'on appliquera ensuite au chapitre Espace L^p.

Voici en exercice un cas (très) particulier de l'inégalité de Jensen (cf. Corollaire 5.6 pour une preuve).

Exercice 3.8. Soit I un intervalle de $\mathbb{R}$, $\alpha_1, \ldots, \alpha_n$ des réels positifs tels que $\sum_{i=1}^{n} \alpha_i = 1$, et φ une fonction convexe sur I. Alors, pour tout $x_1, \ldots, x_n \in I$ on a

$$\varphi\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i \varphi(x_i) \ .$$

CHAPITRE 4

Intégration de Lesbesgue : Construction de l'intégrale et grands théorèmes

Le but de ce chapitre " Construction de l'intégrale et grands théorèmes" est de donner le cadre pour votre cours de probabilité du second semestre, en pensant l'espérance comme une intégrale, tout en généralisant l'intégrale de Riemann et la somme de séries vues en L1 ou en L2. Ce seront aussi les 2 exemples importants unifiés dans ce chapitre (qui donnent les exemples des variables aléatoires continues et discrètes).

On va se concentrer dans ce chapitre sur la construction de l'intégrale et les grands théorèmes qu'il faut apprendre à utiliser. On verra le minimum des définitions requises pour formuler cette construction. Pour cela, on va s'appuyer sur les similarités avec vos cours de probabilités et avec le chapitre 1. Ce sont des constructions importantes dont la démarche sera reprise par exemple au semestre prochain pour la construction de l'espérance conditionnelle. On reporte au deux chapitres suivants les résultats plus techniques dont il est moins important de retenir une idée des preuves.

Dans ce chapitre, le corps est $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$. Pour l'intégration, on a aussi besoin de la droite réelle étendue : $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ avec les mêmes conventions qu'au chapitre précédent : $\infty + \infty = \infty$ et $\lambda.\infty = \infty$ si $\lambda > 0$, $0.\infty = 0$.

Rappels

Droite réelle étendue

Rappel 4.1. La somme $x + y$ avec $x, y \in \overline{\mathbb{R}}$, est définie à l'exception du cas où $x = \pm\infty$ et $y = -x$.
Contrairement au cas des limites, on pose $0. + \infty = 0$, $t. + \infty = +\infty$ pour $t > 0$.

Pour un ensemble A non-vide (non-nécessairement borné), on utilise $\sup A$ pour le plus petit majorant $M \in \overline{\mathbb{R}}$ de A et $\inf A$ pour le plus grand minorant $m \in \overline{\mathbb{R}}$ de A.

On utilisera aussi $\inf \emptyset = +\infty$, $\sup \emptyset = -\infty$.

Exercice 4.1. Soient A, B parties non vides de R. Montrer que :

1. $M = \sup A$ ssi M est un majorant de A et il existe une suite (x_n), avec $x_n \in A$ telle que $x_n \to M$. Caractérisation analogue de inf A .
2. Tout A (non-vide) admet une borne supérieur $\sup A \in]-\infty, \infty]$ et une borne inférieur $\inf A \in [-\infty, \infty[$.
3. sup A et inf A sont uniques.
4. $\sup(-tA) = -t \inf A$, $\forall t \in]0, \infty[$. Formules analogues pour $\sup(tA)$, $\inf(tA)$, $\inf(-tA)$.
5. $\sup(A+B) = \sup A + \sup B$ et $\inf(A+B) = \inf A + \inf B$ (avec la somme usuelle d'ensemble $A+B = \{a+b : a \in A, b \in B\}$.
6. Si $A \subset B$, alors $\inf B \leq \inf A \leq \sup A \leq \sup B$.
7. Si $(x_n)_{n\geq 0}$ est une suite croissante de réels, alors $\lim x_n = \sup\{x_n; n \geq 0\} = \sup x_n$. Énoncé analogue pour une suite décroissante.
8. Si $\sup A > x \in \mathbb{R}$, alors il existe un $y \in A$ tel que $y > x$.

Limites inférieures et supérieures

★ **Définition 4.1.** Pour une suite $x_n \in \mathbb{R}$, sa **limite supérieure** est le nombre :

$$\limsup_n x_n = \inf_{n\geq 1} \sup_{k\geq n} x_k = \lim_{n\to\infty} \sup_{k\geq n} x_k$$

(L'égalité vient de la décroissance de la suite des sup, et c'est aussi la plus grande valeur d'adhérence :exo), sa **limite inférieure** est le nombre :

$$\liminf_n x_n = \sup_{n\geq 1} \inf_{k\geq n} x_k = \lim_{n\to\infty} \inf_{k\geq n} x_k.$$

(c'est aussi la plus petite valeur d'adhérence exo)

Lemme 4.1. On a les formules suivantes (pour $t > 0$) :

$$\limsup -x_n = -\liminf x_n, \liminf -x_n = -\limsup x_n$$

$$\limsup tx_n = t \limsup x_n, \liminf tx_n = t \liminf x_n$$

$$\limsup x_n + y_n \leq \limsup x_n + \limsup y_n$$

$$\liminf x_n + y_n \geq \liminf x_n + \liminf y_n$$

Enfin, $\limsup x_n = \liminf x_n = \ell \in \overline{\mathbb{R}}$ si et seulement si $x_n \to \ell$.

Démonstration. Toutes les (in)égalités sont des conséquences des propriétés des sup, inf puis un passage à la limite :

$$\sup_{k\geq n} -x_n = -\inf_{k\geq n} x_n,\ \inf_{k\geq n} -x_n = -\sup_{k\geq n} x_n$$
$$\sup_{k\geq n} tx_n = t\sup_{k\geq n} x_n,\ \inf_{k\geq n} tx_n = t\inf_{k\geq n} x_n$$
$$\sup_{k\geq n} x_n + y_n \leq \sup_{k\geq n} x_n + \sup_{k\geq n} y_n$$
$$\inf_{k\geq n} x_n + y_n \geq \inf_{k\geq n} x_n + \inf_{k\geq n} y_n$$

Enfin, le sens intéressant est celui où $\limsup x_n = \liminf x_n = \ell \in \mathbb{R}$ et alors $X_k = \inf_{k\geq n} x_k \leq x_k \leq \sup_{k\geq n} x_n = Y_k$ et le théorème des gendarmes permet de conclure que la limite commune de X_k, Y_k est aussi la limite ℓ de x_k. Réciproquement, si $x_n \to \ell$, alors pour tout $\epsilon > 0$, pour n grand, $\ell - \epsilon \leq x_n \leq \ell + \epsilon$ d'où on déduit $\ell - \epsilon \leq \liminf x_n \leq \limsup x_n \leq \ell + \epsilon$ et $\epsilon \to 0$ conclut. □

1 Tribus, fonctions mesurables et mesures

Tribus

Vous avez l'habitude de parler d'évènement d'un espace de probabilité et de considérer la famille $\mathcal{T} \subset \mathcal{P}(\Omega)$ des évènements d'un tel espace. Souvent (pour les probabilités discrètes), on peut prendre $\mathcal{T} = \mathcal{P}(\Omega)$, l'ensemble de toutes les parties de Ω, mais cela ne sera pas possible pour généraliser l'intégrale de Riemann, on ne pourra pas définir l'intégrale de n'importe quel ensemble. La définition suivante retient donc les propriétés essentielles de la famille des évènements que l'on veut pour définir une probabilité sur une telle famille.

★ **Définition 4.2.** Une **tribu (ou σ–algèbre)** sur Ω est une famille $\mathcal{T}$ de partie de Ω, soit $\mathcal{T} \subset \mathcal{P}(\Omega)$ telle que :

1. $\emptyset \in \mathcal{T}$
2. Si $A \in \mathcal{T}$ alors $A^c \in \mathcal{T}$.
3. Pour toute suite infinie (dénombrable) $(A_n)_{n\geq 1}$ de parties de $\mathcal{T}$, alors leur union est aussi dans la tribu $\bigcup_{n\geq 1} A_n \in \mathcal{T}$.

Un ensemble $A \in \mathcal{T}$ est appelée partie $\mathcal{T}$**−mesurable** ou simplement **mesurable**.
Un **espace mesurable** est une paire $(\Omega, \mathcal{T})$ formée d'un ensemble Ω et d'une tribu $\mathcal{T}$ sur Ω. Les enembles $A \in \mathcal{T}$ sont appelés **ensembles mesurables (pour la tribu $\mathcal{T}$ ou $\mathcal{T}$−mesurables).**

Le résultat suivant est assez évident

Lemme 4.2. Pour toute suite finie $A_1, \cdots, A_n$ de $\mathcal{T}$, alors $A_1 \cup \cdots \cup A_n \in \mathcal{T}$.
Pour toute suite infinie (dénombrable, resp. finie) $(A_n)_{n \geq 1}$ (resp. $A_1, \cdots, A_n$) de parties de $\mathcal{T}$, alors leur intersection $\bigcap_{n \geq 1} A_n \in \mathcal{T}$ (resp. $\bigcap_{i=1}^{n} A_i \in \mathcal{T}$).

Démonstration. Pour le premier, il suffit de prolonger la suite en $A_k = \emptyset \in \mathcal{T}$ pour $k \geq n+1$ et alors $A_1 \cup \cdots \cup A_n = \bigcup_{n \geq 1} A_n \in \mathcal{T}$

Pour l'intersection, il suffit de combiner union et complémentaire, par exemple dans le cas dénombrable : $\bigcap_{n \geq 1} A_n = \left(\bigcup_{n \geq 1} A_n^c \right)^c \in \mathcal{T}$. □

Remarque 4.2. On verra au chapitre suivant la notion plus élémentaire d'algèbre de parties (ou clan) où l'on demande seulement la stabilité par union finie, mais elle ne suffira pas pour la construction de l'intégrale. Il faut comparer la notion de tribu à celle de topologie de la remarque 2.3, qui était l'axiomatisation des parties ouvertes d'un espace métrique. Comme une topologie, une tribu est stable par intersection finie, mais même plus elle est stable par intersection dénombrable. Mais par contre, elle n'est pas stable par union quelconque, mais seulement par union dénombrable. Donc aucune des notions n'est plus générale que l'autre. Enfin, la nouveauté est la stabilité par complémentaire, ou autrement dit par toutes les opérations logiques de bases sur les ensembles (complémentaire, intersection et union binaires), et c'est la clef pour son application en probabilité (on veut aussi que les évènements soient stables par toutes les opérations logiques). On va cependant traiter dans beaucoup d'aspect la notion de tribu comme la famille des ouverts d'un espace métrique (ou plus généralement topologique).

Mesure et Probabilité sur une tribu

L'intégation va dépendre d'un objet de base qui permet la "mesure du volume" (ou en physique la "mesure de la masse" ou d'autres grandeurs extensives) et qui va généraliser la notion de probabilité.

★ **Définition 4.3** (Définition d'une mesure). Soit $(\Omega, \mathcal{T})$ un **espace mesurable**.
On appelle **mesure (positive)** est une application $\mu : \mathcal{T} \to [0, +\infty]$ ayant les propriétés suivantes :

1. $\mu(\emptyset) = 0$
2. (σ−additivité) Pour toute suite au plus dénombrable $(A_i)_{i\in I}$ [a] d'éléments de $\mathcal{T}$ deux à deux disjoints,
$$\mu(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mu(A_i).$$

Une **mesure de probabilité** P est une mesure positive P vérifiant en plus $P(\Omega) = 1$. Un **espace mesuré (resp. de probabilité)** est un triplet $(\Omega, \mathcal{T}, \mu)$ (resp. $(\Omega, \mathcal{T}, P)$) formée d'une mesure positive μ (resp. une mesure de probabilité P) sur un espace mesurable $(\Omega, \mathcal{T})$.

---

a. c'est à dire soit $I = [[0, n]]$ et dans ce cas $\sum_{i\in I} \mu(A_i) = \sum_{i=0}^{n} \mu(A_i)$, soit soit $I = \mathbb{N}$ et dans ce cas $\sum_{i\in I} \mu(A_i) = \sum_{i=0}^{\infty} \mu(A_i)$ est la somme de la série, finie ou $+\infty$

Une mesure a des propriétés très similaires à celle d'une probabilité dont vous avez l'habitude (exo).

★ **Proposition 4.3.** i) Si $A \subset B$ alors $\mu(A) \leq \mu(B)$ (μ est croissante).

ii) Pour toute suite au plus dénombrable $(A_i)_{i\in I}$, $\mu(\bigcup_{i\in I} A_i) \leq \sum_{i\in I} \mu(A_i)$ (μ est sous−additive).

iii) Si $(A_n)_{n\geq 1}$ est une suite croissante,
$$\mu(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} \mu(A_n) = \sup_{n\geq 1} \mu(A_n).$$

iv) Si $(A_n)_{n\geq 1}$ est une suite décroissante avec $\mu(A_1) < \infty$,
$$\mu(\bigcap_{n\geq 1} A_n) = \lim_{n\to\infty} \mu(A_n) = \inf_{n\geq 1} \mu(A_n).$$

v) Si $\mu(\Omega)$ est finie : $\mu(A^c) = \mu(\Omega) - \mu(A)$.

Ensembles μ−négligeables

★ **Définition 4.4.** Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, un ensemble $A \subset \Omega$ est μ**−négligeable** si il existe $B \in \mathcal{T}$ contenant $A \subset B$ et avec $\mu(B) = 0$.

Attention, A n'est par forcément mesurable donc on ne peut PAS déduire $\mu(A) = 0$. Mais la seule extension possible, si A devenait mesurable, serait la valeur 0.

Lemme 4.4. Une union au plus dénombrable d'ensembles μ−négligeables est μ−négligeable.

Démonstration. Si $(A_n)_{n\geq 0}$ est μ−négligeable, alors il existe une suite $B_n \in \mathcal{T}$ avec $\mu(B_n) = 0$ et $A_n \subset B_n$, donc

$$\bigcup_{n\geq 0} A_n \subset \bigcup_{n\geq 0} B_n \in \mathcal{T}, \mu\Big(\bigcup_{n\geq 0} B_n\Big) \leq \sum_{n\geq 0} \mu(B_n) = 0.$$

□

Exercice 4.2. Montrer que le seul ensemble ν−négligeable pour la mesure de comptage ν est l'ensemble vide.

Définition 4.5. Une propriété $P(\omega)$ des points $\omega \in \Omega$ est dite vraie **presque partout** (par rapport à μ, ou μ−presque partout, ou μ−.p.p) si $\{\omega \in \Omega : \neg P(\omega)\}$ est μ−négligeable. Autrement dit, si il existe $B \in \mathcal{T}$ avec $\mu(B) = 0$ telle que P est vraie sur B^c.

Exercice 4.3. Montrer que l'indicatrice de $\mathbb{Q}$ $1_{\mathbb{Q}}$ est nulle λ−p.p.

Un ensemble peut donc être dense et λ−négligeable.

Exemples de tribus

Exemple 4.1. $\mathcal{T} = \mathcal{P}(\Omega)$ est une tribu (appelée **tribu discrète**) et $\mathcal{T} = \{\emptyset, \Omega\}$ est aussi une tribu (appelée **tribu grossière**).

Tribus engendrés par une famille d'ensembles

En pratique, on n'a pas besoin de connaître en détail, tous les éléments contenus dans une tribu, il suffit de savoir qu'on a assez d'élements voulus (les générateurs de la tribu). Ceci est permis par le lemme suivant.

Lemme 4.5. Si $(\mathcal{T}_i)_{i\in I}$ est une famille de tribus, alors $\bigcap_{i\in I}\mathcal{T}_i$ est une tribu. On peut donc parler de la plus petite tribu contenant une famille $\mathcal{A}\subset\mathcal{P}(\Omega)$, qui est l'intersection de toutes les tribus contenant $\mathcal{A}$, elle est notée $\sigma(\mathcal{A})$ et appelée la **tribu engendrée par** $\mathcal{A}$.

Démonstration. C'est une conséquence directe de la forme de la définition. $\emptyset\in\mathcal{T}_i$ pour tout i, donc $\emptyset\in\bigcap_{i\in I}\mathcal{T}_i$.

De plus, si $A\in\bigcap_{i\in I}\mathcal{T}_i$, alors $A\in\mathcal{T}_i$ pour tout i, donc comme $\mathcal{T}_i$ est une tribu, $A^c\in\mathcal{T}_i$ pour chaque i et donc $A^c\in\bigcap_{i\in I}\mathcal{T}_i$.

Enfin, si pour chaque $n\geq 1$, $A_n\in\bigcap_{i\in I}\mathcal{T}_i$, alors $A_n\in\mathcal{T}_i$ pour tout i, donc comme $\mathcal{T}_i$ est une tribu, $\bigcup\limits_{n\geq 0}A_n\in\mathcal{T}_i$ pour chaque i et donc $\bigcup\limits_{n\geq 1}A_n\in\bigcap\limits_{i\in I}\mathcal{T}_i$. □

Exemple 4.2 ((cf. TD)). Si $A\subset\Omega$, la tribu engendrée par A est $\sigma(\{A\})=\{A,A^c,\emptyset,\Omega\}$.

Exemple 4.3 ((cf. TD)). Si $A_1,\cdots,A_n\subset\Omega$ forment une partition (c'est à dire sont 2 à 2 disjoints et d'union Ω), la tribu engendrée $\sigma(\{A_1,\cdots,A_n\})=\{\cup_{i\in I}A_i : I\subset[\![1,n]\!]\}$.

★ **Définition 4.6.** Pour (X,d) un espace métrique dont $\mathcal{T}$ est la topologie des ouverts, on appelle **tribu borélienne** sur X, notée $\mathcal{B}(X)=\sigma(\mathcal{T})$ la tribu engendrée par les ouverts de X.

Le résultat suivant est montré en annexe C en section 2.

★ **Lemme 4.6.** Sur $\mathbb{R}^n$, la tribu borélienne a le système de générateurs :

$$\mathcal{B}(\mathbb{R}^n)=\sigma\Big(\prod_{i=1}^{n}]a_i,b_i[,\ a_i<b_i\in\mathbb{R}\Big)$$

A partir de là, on obtiendra en TD les autres générateurs usuels.

★ **Lemme 4.7** (cf. TD). Sur $\mathbb{R}^n$, la tribu borélienne a les différents systèmes de générateurs :

$$\begin{aligned}\mathcal{B}(\mathbb{R}^n) &= \sigma\Big(\prod_{i=1}^{n}]-\infty, b_i],\, b_i \in \mathbb{R}\Big)\\ &= \sigma\Big(\prod_{i=1}^{n}[a_i, +\infty[,\, a_i \in \mathbb{R}\Big)\\ &= \sigma\Big(\prod_{i=1}^{n}[a_i, b_i],\, a_i < b_i \in \mathbb{R}\Big)\\ &= \sigma\Big(F : F \text{ fermé de } \mathbb{R}^n\Big)\end{aligned}$$

Tribu engendrée par une fonction

Lemme 4.8. Soit $f : \Omega \to (E, \mathcal{B})$ une fonction, $\sigma(f) := f^{-1}(\mathcal{B}) = \{f^{-1}(B), B \in \mathcal{B}\} \subset \mathcal{T}$ est une tribu sur Ω. On l'appelle **tribu engendrée par** f.

Démonstration. C'est essentiellement une application des rappels sur l'image réciproque de fonctions (1.1). D'abord $f^{-1}(\emptyset) = \emptyset \in \sigma(f)$, $f^{-1}(E) = \Omega \in \sigma(f)$. Pour $A \in \mathcal{B}$ (resp, $A_n \in \mathcal{B}$, $n \geq 1$) :

$$\begin{aligned}[f^{-1}(A)]^c = f^{-1}(A^c) \in \sigma(f) \text{ car } A^c \in \mathcal{B},\\ \bigcup_{n\geq 1} f^{-1}(A_n) = f^{-1}\Big(\bigcup_{n\geq 1} A_n\Big) \in \sigma(f) \text{ car } \bigcup_{n\geq 1} A_n \in \mathcal{B}\end{aligned}$$

□

Exemples de mesures

Exemple 4.4 ((Mesure de comptage)). Sur tout ensemble Ω, on définit sur $\mathcal{P}(\Omega)$, la mesure suivante (dite de comptage)

$$\nu(A) = \begin{cases} \mathrm{Card}(A) & \text{si} \quad \text{A fini} \\ +\infty & \text{sinon} \end{cases}$$

Exemple 4.5 ((Mesure discrète sur Ω fini)). Sur tout ensemble dénombrable $\Omega = \{\omega_n, n \in [\![1, n]\!]\}$, pour $(\mu_i) \in [0, +\infty[^n$ on définit sur $\mathcal{P}(\Omega)$:

$$\mu(A) = \sum_{\omega_i \in A} \mu_i.$$

C'est une mesure sur $\mathcal{P}(\Omega)$. Une fois connue l'intégration pour la mesure de comptage (ou de façon équivalente si on connaît la notion de **famille sommable**, on pourra généraliser cet exemple au cas Ω dénombrable)

Enfin, l'exemple fondamental est le théorème donnant l'existence de la mesure de Lebesgue (admis)

★ **Théorème 4.9** (définissant l'intégrale de Lebesgue). (admis) Il existe une unique mesure λ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ invariante par translation [a] telle que

$$\lambda\Big([0, 1]^n\Big) = 1.$$

Cette mesure est appelée **mesure de Lebesgue** sur $\mathbb{R}^d$ et notée $\lambda \equiv \lambda_d$ et elle vérifie pour $a_i < b_i$:

$$\lambda\Big(\prod_{i=1}^{n}[a_i, b_i]\Big) = \lambda\Big(\prod_{i=1}^{n}]a_i, b_i[\Big) = \prod_{i=1}^{n}(b_i - a_i).$$

a. au sens ou pour tout $a \in \mathbb{R}^d$, $B \in \mathcal{B}(\mathbb{R}^d)$, si on note $a + B = \{a + b, b \in B\}$, alors $\lambda(a + B) = \Lambda(B)$

Proposition 4.10 (définissant la mesure image). Soit $f : \Omega \to (E, \mathcal{B})$une fonction et $(\Omega, \sigma(f), \mu)$ un espace mesuré alors la formule $\mu^f(B) = \mu(f^{-1}(B))$ for $B \in \mathcal{B}$ est une mesure sur $\mathcal{T}$, appelée **mesure image de μ par f.**

Démonstration. Pour voir que μ^f est une mesure sur $\mathcal{B}$, il faut noter $\mu^f(\emptyset) = \mu(\emptyset) = 0$. Puis pour la σ−additivité, pour $A_i \in \mathcal{B}$, $i \in I$ deux à deux disjoints avec I au plus dénombrable, on a :

$$\begin{aligned}\mu^f\Big(\bigcup_{i \in I} A_i\Big) &= \mu\Big(f^{-1}\Big(\bigcup_{i \in I} A_i\Big)\Big) = \mu\Big(\bigcup_{i \in I} f^{-1}(A_i)\Big) \\ &= \sum_{i \in I} \mu(f^{-1}(A_i)) = \sum_{i \in I} \mu^f(A_i),\end{aligned}$$

vu que les $f^{-1}(A_i) \in \sigma(f)$ sont aussi deux à deux disjoints par (1.1), on a pu utilisé à l'avant−dernière égalité la σ−additivité de μ.

□

Fonctions mesurables

Il nous reste à spécifier les fonctions qu'on va pouvoir intégrer. Il faut lire la définition suivante comme l'analogue de la définition topologique de la continuité (proposition 2.22)

Définition 4.7. Une fonction $f : (\Omega, \mathcal{T}) \to (E, \mathcal{B})$ entre espaces mesurables est **mesurable** si $f^{-1}(\mathcal{B}) \subset \mathcal{T}$ c'est à dire si pour tout $B \in \mathcal{B}$, $f^{-1}(B) \in \mathcal{T}$. Si $(\Omega, \mathcal{T}) = (X, \mathcal{B}(X)), (E, \mathcal{B}) = (Y, \mathcal{B}(Y))$, on appelle **fonction borélienne** une fonction mesurable $f : (X, \mathcal{B}(X)) \to (Y, \mathcal{B}(Y))$.

On déduit immédiatement de la définition comme le corollaire 4.11 :

★ **Lemme 4.11** (Stabilité par composition de la mesurabilité)**.** Si $f : (D, \mathcal{A}) \to (E, \mathcal{B})$ et $g : (E, \mathcal{B}) \to (F, \mathcal{C})$ sont mesurables, alors, $g \circ f : D \to F$ est mesurable.

Démonstration. Pour tout ensemble mesurable $U \in \mathcal{C}$, $g^{-1}(U) \in \mathcal{B}$ est mesurable de Y par mesurabilité de g, puis $f^{-1}(g^{-1}(U)) \in \mathcal{T}$ est mesurable de X par mesurabilité de f, mais $f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U) \in \mathcal{A}$. Comme c'est vrai pour tout ensemble mesurable U, on déduit de la définition précédente $g \circ f$ est mesurable. □

Comme en probabilité, l'intérêt principal de la notion de mesurabilité est de permettre de définir la notion de mesure image (analogue de la loi d'une variable aléatoire).

Proposition 4.12. Soit $f : \Omega \to (E, \mathcal{B})$ une fonction, la tribu engendrée par f du lemme 4.8 $\sigma(f) := f^{-1}(\mathcal{B}) = \{f^{-1}(B), B \in \mathcal{B}\}$ est la plus petite tribu rendant f mesurable. Autrement dit : Si $\mathcal{T} \subset \mathcal{P}(\Omega)$ est une tribu, $f : (\Omega, \mathcal{T}) \to (E, \mathcal{B})$ est mesurable si et seulement si $\sigma(f) \subset \mathcal{T}$.

Démonstration. On a vu au lemme 4.8 que $\sigma(f)$ est une tribu. $f : (\Omega, \sigma(f)) \to (E, \mathcal{B})$ est mesurable par définition, car pour tout $B \in \mathcal{B}$, on a $f^{-1}(B) \in \sigma(f)$ par définition de $\sigma(f)$, et cela veut dire $f : (\Omega, \sigma(f)) \to (E, \mathcal{B})$ est mesurable par définition de la mesurabilité. L'équivalence $f : (\Omega, \mathcal{T}) \to (E, \mathcal{B})$ est mesurable si et seulement si $\sigma(f) \subset \mathcal{T}$ vient aussi directement des deux mêmes définitions. L'inclusion $\sigma(f) \subset \mathcal{T}$ dit justement que $\sigma(f)$ est plus petite (pour l'inclusion) que toute tribu rendant f mesurable. □

Exemple 4.6. Si $A \in \mathcal{T}$, la fonction indicatrice $1_A : (\Omega, \mathcal{T}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable, car $\sigma(1_A) = \sigma(\{A\}) = \{A, A^c, \emptyset, \Omega\}$ par l'exemple 4.2 et donc $\sigma(1_A) \subset \mathcal{T}$ par la définition d'une tribu.

En pratique, on a besoin d'une description en terme de parties génératrices :

Lemme 4.13. Une fonction $f : (\Omega, \mathcal{T}) \to (E, \sigma(\mathcal{A}))$, vers un espace mesurable engendré par une famille $\mathcal{A}$, est mesurable si et seulement si $f^{-1}(\mathcal{A}) \subset \mathcal{T}$ c'est à dire si pour tout $A \in \mathcal{A}$, $f^{-1}(A) \in \mathcal{T}$.

Démonstration. Si f mesurable, vu que $\mathcal{A} \subset \sigma(\mathcal{A})$, le fait que $f^{-1}(A) \in \mathcal{T}$ est une conséquence directe de la définition. Le contenu du lemme est donc la réciproque.

On introduit une fammille $\mathcal{B}$ (qui va se révéler être la plus grande tribu de E rendant f mesurable, la preuve est donc très similaire à celle sur $\sigma(f)$) :

$$\mathcal{B} = \{B \in \mathcal{P}(E) : f^{-1}(B) \in \mathcal{T}\}.$$

Par hypothèse $\mathcal{A} \subset \mathcal{B}$. Vérifions que $\mathcal{B}$ est une tribu (par la définition) :

- $\emptyset \in \mathcal{B}$ car $f^{-1}(\emptyset) = \emptyset \in \mathcal{T}$
- Si $B \in \mathcal{B}$, alors $f^{-1}(B^c) = f^{-1}(B)^c \in \mathcal{T}$ car $\mathcal{T}$ est une tribu donc $B^c \in \mathcal{B}$
- Si $A_n \in \mathcal{B}$, $n \geq 1$, alors $f^{-1}(\bigcup_{n\geq 1} A_n) = \bigcup_{n\geq 1} f^{-1}(A_n) \in \mathcal{T}$ car $\mathcal{T}$ est une tribu donc $\bigcup_{n\geq 1} A_n \in \mathcal{B}$

En conséquence, $\mathcal{B}$ est une tribu qui contient $\mathcal{A}$, donc $\sigma(\mathcal{A}) \subset \mathcal{B}$ ce qui dit exactement : $\forall B \in \sigma(\mathcal{A}) : f^{-1}(B) \in \mathcal{T}$ soit la définition de f mesurable. □

Corollaire 4.14. Une fonction $f : (\Omega, \mathcal{T}) \to (Y, \mathcal{B}(Y))$ vers la tribu borélienne d'un espace métrique est mesurable, si et seulement si pour tout ouvert U (resp. tout fermé F) on a $f^{-1}(U) \in \mathcal{T}$ (resp. $f^{-1}(F) \in \mathcal{T}$). En particulier, si $(\Omega, \mathcal{T}) = (X, \mathcal{B}(X))$ pour un espace métrique X, alors, toute fonction continue f est borélienne.

Démonstration. Le premier résultat est une conséquence directe du lemme vu que $\mathcal{B}(Y) = \sigma\Big(\{U \subset Y : U \text{ ouvert}\}\Big) = \sigma\Big(\{F \subset Y : F \text{ fermé}\}\Big)$. Par la proposition 2.22, $f^{-1}(U)$ est ouvert (resp. $f^{-1}(F)$ est fermé) donc dans $\mathcal{B}(X)$ pour tout ouvert U de Y, on déduit que la continuité implique la mesurabilité. □

En composant, avec les produits et sommes qui sont des applications continues, on obtient les mêmes stabilités algébriques que pour les fonctions continues :

Corollaire 4.15. Les fonctions mesurables $(\Omega, \mathcal{T}) \to \mathbb{R}$ sont stables par combinaisons linéaires, produits, fractions rationnelles à dénominateur non nulle, passage à l'exponentielle (etc.)

On tire de même immédiatement des lemmes 4.6 et 4.7 :

Corollaire 4.16. Une fonction $f = (f_1, \cdots, f_n) : (\Omega, \mathcal{T}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ est mesurable si et seulement si l'une des assertions suivantes est vérifiée :

1. Pour tout $b_1, \cdots, b_n \in \mathbb{R}$, $f^{-1}\Big(\prod_{i=1}^{n}]-\infty, b_i]\Big) \in \mathcal{T}$
2. Pour tout $a_1, \cdots, a_n \in \mathbb{R}$, $f^{-1}\Big(\prod_{i=1}^{n} [a_i, +\infty[\Big) \in \mathcal{T}$
3. Pour tout $a_1 < b_1, \cdots, a_n < b_n \in \mathbb{R}$, $f^{-1}\Big(\prod_{i=1}^{n} [a_i, b_i]\Big) \in \mathcal{T}$
4. Pour tout $a_1 < b_1, \cdots, a_n < b_n \in \mathbb{R}$, $f^{-1}\Big(\prod_{i=1}^{n}]a_i, b_i[\Big) \in \mathcal{T}$.
5. Pour tout $i = 1, \cdots, n$, $f_1, \cdots, f_n : (\Omega, \mathcal{T}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sont toutes mesurables.

Corollaire 4.17. Une fonction $f : (\Omega, \mathcal{T}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ (à valeur dans l'espace métrique $(\overline{\mathbb{R}}, d_{\overline{\mathbb{R}}})$ de l'exemple 2.5) est mesurable si et seulement si les trois assertions suivantes sont vérifiées :

1. $f^{-1}(\{\infty\}) \in \mathcal{T}$
2. $f^{-1}(\{-\infty\}) \in \mathcal{T}$
3. Pour tout $a < b \in \mathbb{R}$, $f^{-1}([a, b]) \in \mathcal{T}$

On renvoie aussi à l'annexe section 3 pour le résultat important suivant

★ **Théorème 4.18.** Les constructions suivantes sont mesurables :

1. Un supremum d'une suite $f_n : (\Omega, \mathcal{T}) \to \overline{\mathbb{R}}$ de fonctions mesurables
2. La $\limsup$, $\liminf$ d'une suite $f_n : (\Omega, \mathcal{T}) \to \overline{\mathbb{R}}$ de fonctions mesurables
3. Une limite simple d'une suite $f_n : (\Omega, \mathcal{T}) \to \overline{\mathbb{R}}$ de fonctions mesurables

Unicité des mesures σ–finies

Définition 4.8. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. On dit que $(X, \mathcal{A}, \mu)$ est σ**–fini** s'il existe une suite de parties mesurables $(A_n)_{n\in\mathbb{N}}$ telle que $\mu(A_n) < +\infty$ pour tout n, et $\Omega = \bigcup_n A_n$.

Cette hypothèse est par exemple vérifiée quand $\mu(\Omega) < +\infty$ (donc en particulier quand μ est une mesure de probabilité), quand $\Omega = \mathbb{N}$ muni de la mesure de comptage, ou quand $\Omega = \mathbb{R}^n$ muni de la mesure de Lebesgue.

On renvoie à l'annexe C en section 1 pour une preuve d'un corollaire très classique au lemme de classe monotone pour les mesures dans le cas des mesures σ–finies.

Corollaire 4.19 (au lemme de classe monotone)**.** Soient μ et ν des mesures sur un espace mesurable $(\Omega, \mathcal{T})$. Soit $\mathcal{E}$ une famille stable par intersection finie qui engendre $\mathcal{T}$. Si μ et ν coïncident sur $\mathcal{E}$ (i.e. $\mu(E) = \nu(E), \forall E \in \mathcal{E}$) et si il existe une suite de parties $A_n \in \mathcal{E}$ telle que $\Omega = \cup_n A_n$ et $\mu(A_n) = \nu(A_n) < +\infty$ alors μ et ν sont égales (i.e. $\mu(B) = \nu(B), \forall B \in \mathcal{T}$).

2 Les fonctions étagées (mesurables) et leur intégrale

Comme les fonctions en escalier sont la base pour l'intégrale de Riemann, on considère ici la classe des fonctions étagées (mesurables) qui sont la base de l'intégrale de Lebesgue. Les fonctions en escaliers sont les combinaisons linéaires des indicatrices d'intervalles $1_{]a,b]}$. On les prend pour base de l'intégrale de Riemann car on sait définit $\int_{\mathbb{R}} 1_{]a,b]}(x)\,dx = (b - a)$.

On fixe à partir de maintenant un espace mesuré $(\Omega, \mathcal{T}, \mu)$.

Maintenant, qu'on dispose d'une mesure μ, on veut définir de même pour $A \in \mathcal{T}$:

$$\int_\Omega 1_A\, d\mu \equiv \int_\Omega 1_A(\omega)\, d\mu(\omega) = \mu(A).$$

Plus généralement, on définit :

Définition 4.9. Pour $A, B \in \mathcal{T}$, **l'intégrale de** 1_A **sur** B **par rapport à** μ est notée et définie par :
$$\int_B f d\mu \equiv \int_B 1_A(\omega) d\mu(\omega) = \mu(A \cap B).$$

Les combinaisons linéaires de fonctions indicatrices (mesurables) vont donc être de même la base de l'intégrale de Lebesgue :

Définition 4.10. Soit $(\Omega, \mathcal{T})$ un espace mesurable, on appelle **fonction étagée** $f : (\Omega, \mathcal{T}) \to \mathbb{R}^d$ une fonction de la forme
$$f(\omega) = \sum_{i=1}^{n} a_i 1_{A_i}(\omega)$$
pour $a_i \in \mathbb{R}^d$ et $A_i \in \mathcal{T}$. Pour $d = 1$, la représentation est dite **canonique** si $a_1 < \cdots < a_n$, tous non nuls ($\forall i$, $a_i \neq 0$) et les $A_1, \cdots, A_n$ sont deux à deux disjoints et non vides.

Exercice 4.4. Les fonctions étagées sur $(\Omega, \mathcal{T})$ forment un sous espace vectoriel des fonctions $\Omega \to \mathbb{R}^d$.

Comme on veut que l'intégrale soit linéaire, on est conduit à la définition suivante :

Définition 4.11. Soit f une fonction étagée positive $f(\omega) = \sum_{i=1}^{n} a_i 1_{A_i}(\omega)$ avec $A_i \in \mathcal{T}$ des ensembles mesurables **deux à deux disjoints** ($a_i > 0$), on définit **l'intégrale de** f **sur** $B \in \mathcal{T}$ par rapport à μ par :
$$\int_B f d\mu \equiv \int_B f(\omega) d\mu(\omega) = \sum_{i=1}^{n} a_i \mu(A_i \cap B).$$

On reporte à l'annexe C section 4 la preuve facile mais fastidieuse du lemme suivant :

Lemme 4.20. Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, et $f, h : (\Omega, \mathcal{T}) \to [0, +\infty]$ étagées positives, $B \in \mathcal{T}$:

1. Si $f \geq 0$, alors $\int_B f d\mu = \int_\Omega 1_B f d\mu$.
2. Si $f \geq 0$, $c > 0$, alors $\int_B c f d\mu = c \int_B f d\mu$.
3. (additivité) $\int_B f + h d\mu = \int_B f d\mu + \int_B h d\mu$.
4. (monotonie) Si $0 \leq f \leq h$ alors $0 \leq \int_B f d\mu \leq \int_B h d\mu$.

Le résultat crucial qui va permettre l'extension de l'intégrale est le résultat suivant :

★ **Lemme 4.21.** Soit $(\Omega, \mathcal{T})$ un espace mesurable. Toute fonction mesurable positive $f : (\Omega, \mathcal{T}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ est limite simple d'une suite croissante de fonctions étagées positives.

Démonstration. On prend

$$\begin{aligned} f_n(x) &= 2^n 1_{\{x : f(x) = +\infty\}} + \sum_{k=0}^{4^n - 1} \frac{k}{2^n} 1_{f^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}[)}(x) \\ &= \begin{cases} \frac{k}{2^n} & \text{si} \quad \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}, 0 \leq k < 4^n \\ 0 & \text{si} \quad \frac{4^n - 1 + 1}{2^n} = 2^n \leq f(x) < +\infty \\ 2^n & \text{si} \quad f(x) = +\infty \end{cases} \leq f(x). \end{aligned}$$

1. Comme f mesurable, chacun des $f^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}[) \in \mathcal{T}$ et $f^{-1}(\{+\infty\}) \in \mathcal{T}$ et donc f_n est étagée (comme combinaison linéaire de fonctions indicatrices mesurables).
2. La suite est croissante $0 \leq f_n \leq f_m$ pour $n \leq m$. Sur $f^{-1}([0, 2^n])$, on découpe chaque intervalle de définition de f_n en 2^{m-n} ensembles dans la définitions de f_m. Si $f_m(x) = \frac{k}{2^m} \leq f(x) < \frac{k+1}{2^m}, 0 \leq k < 2^{m+n}$, on trouve $k = \kappa 2^{m-n} + l$ pour $0 \leq l < 2^{m-n}, 0 \leq \kappa < 4^n$ par division euclidienne et

$$\begin{aligned} f_n(x) = \frac{\kappa}{2^n} \leq f_m(x) = \frac{k}{2^m} &= \frac{\kappa}{2^n} + \frac{l}{2^m} \\ &\leq f(x) < \frac{\kappa}{2^n} + \frac{l+1}{2^m} \leq f_n(x) + \frac{l+1}{2^m} \end{aligned}$$

   Sur $f^{-1}(]2^n, +\infty[)$ on a $f_n(x) = 0 \leq f_m(x)$. Vu $f_n(x) \leq f(x) \leq f_n(x) + \frac{1}{2^n}$ on en déduit $f(x) - \frac{1}{2^n} \leq f_n(x) \leq f(x)$ si $f(x) \leq 2^n$, on déduit la convergence simple.

□

3 Intégrale des fonctions mesurables positives

On peut maintenant définir l'intégrale des fonctions mesurables positives :

★ **Définition 4.12.** Soit $f : \Omega \to [0, +\infty]$ une fonction mesurable positive sur un espace mesuré $(\Omega, \mathcal{T}, \mu)$, on définit **l'intégrale de f sur $B \in \mathcal{T}$ par rapport à μ** par :

$$\int_B f d\mu \equiv \int_B f(\omega) d\mu(\omega) = \sup \left\{ \int_B g d\mu : g \text{ étagée}, 0 \leq g \leq f \right\} \in [0, +\infty].$$

Remarque 4.3. Pour la mesure de comptage ν sur I, toute suite $a : I \to [0,+\infty]$ est mesurable positive et l'intégrale correspond à la définition de la somme d'une famille sommable :

$$\int_I f d\nu = \sum_{i\in I} a_i = \sup\left\{\sum_{j\in J} a_j : J \subset I,\ \text{fini}\right\}.$$

Remarque 4.4. Si f est étagée positive, pour chaque $g \leq f$ étagée positive, on a vu au lemme 4.20, $\int_B g d\mu \leq \int_B f d\mu$ donc

$$\int_B f d\mu \geq \sup\left\{\int_B g d\mu : g \text{ étagée, } 0 \leq g \leq f\right\}.$$

Et comme f fait parti des g du sup, on a en fait égalité, et la valeur de la définition du cas étagé positif coïncide avec la nouvelle valeure.

Premières propriétés

On reporte à l'annexe C section 4 la preuve facile mais fastidieuse du lemme suivant :

Lemme 4.22. Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, et $f, h : (\Omega, \mathcal{T}) \to [0,+\infty]$ mesurable positive, $A, B \in \mathcal{T}$:

1. (monotonie) Si $0 \leq f \leq h$ alors $0 \leq \int_B f d\mu \leq \int_B h d\mu$.
2. Si $f \geq 0$, alors $\int_B f d\mu = \int_\Omega 1_B f d\mu$. En particulier, pour $A \subset B$, $0 \leq \int_A f d\mu \leq \int_B f d\mu$.
3. Si $f \geq 0$, $c \geq 0$, alors $\int_B c f d\mu = c \int_B f d\mu$.
4. Si $f = 0$ ou $\mu(B) = 0$, alors $\int_B f d\mu = 0$.
5. (sur-additivité) $\int_B f + h d\mu \geq \int_B f d\mu + \int_B h d\mu$.

La dernière propriété n'est pas optimale, nous verrons l'additivité en utilisant le théorème de convergence monotone. Nous la mentionnons ici pour signaler que l'additivité n'est pas évidente à partir de la définition.

Théorème de convergence monotone de Beppo Levi

★ **Théorème 4.23** (Théorème de convergence monotone ou TCM)**.** Soit $Z_n : (\Omega, \mathcal{T}) \to [0,+\infty]$, une suite croissante de fonctions mesurables positives qui tend simplement vers Z. Alors Z est mesurable et pour tout $B \in \mathcal{T}$:

$$\lim_{n\to\infty} \int_B Z_n d\mu = \int_B Z d\mu \equiv \int_B \lim_{n\to\infty} Z_n d\mu.$$

Démonstration. La mesurabilité de Z vient du théorème 4.18. Posons $\alpha = \sup_n \int_B Z_n d\mu$.

Comme $Z_n \leq Z_m \leq Z$ pour $n \leq m$, la monotonie de l'intégrale (du lemme 4.22) montre que

$$\int_B Z_n d\mu \leq \int_B Z_m d\mu \leq \int_B Z d\mu$$

Donc, comme la suite $\int_B Z_n d\mu$ est croissante, elle converge vers son sup et :

$$\lim_{n\to\infty} \int_B Z_n d\mu = \alpha \leq \int_B Z d\mu.$$

Pour la réciproque, soit $1 > \epsilon > 0$ et une fonction étagée $g(\omega) = \sum_{i=1}^{m} b_i 1_{B_i}(\omega) \leq Z(\omega)$. On pose $A_n = \{\omega \in \Omega : Z_n(\omega) \geq Z(\omega) - \epsilon Z(\omega)\}$. Par la monotonie de l'intégrale et la formule pour les fonctions étagées :

$$\begin{aligned}\int_B Z_n d\mu &\geq \int_B Z_n 1_{A_n} d\mu \\ &\geq (1-\epsilon) \int_B g 1_{A_n} d\mu \\ &= (1-\epsilon) \sum_{i=1}^{m} b_i \mu(B_i \cap A_n \cap B).\end{aligned} \tag{4.1}$$

Remarquons finalement que $\bigcup_{n\geq 0} A_n = \Omega$ vu que pour tout $\omega \in \Omega$, $Z_n(\omega) \to Z(\Omega) > Z(\omega) - \epsilon Z(\omega)$. Comme Z_n est croissante, A_n est aussi croissante donc par la proposition 4.3,

$$\mu(B_i \cap A_n \cap B) \to \mu(\bigcup_n B_i \cap A_n \cap B) = \mu(B_i \cap B).$$

En passant à la limite dans (4.1), on obtient :

$$\alpha \geq (1-\epsilon) \sum_{i=1}^{m} b_i \mu(B_i \cap B) = (1-\epsilon) \int_B g d\mu$$

soit en passant au sup sur $g \leq Z$ puis à la limite $\epsilon \to 0$, on obtient l'inégalité voulue $\alpha \geq \int_B Z d\mu$.

□

On obtient un résultat concret d'approximation pour $\int_B f d\mu$.

Corollaire 4.24. Soit f mesurable positive. Pour toute suite croissante de fonctions étagées telle que $f_n \to f$, on a $\int_B f_n d\mu \to \int_B f d\mu$.

Corollaire 4.25 (Linéarité de l'intégrale : cas positif)**.** Soient f, g mesurables positives et $\alpha, \beta > 0$, on a :

$$\int_B \alpha f + \beta g d\mu = \alpha \int_B f d\mu + \beta \int_B g d\mu.$$

Démonstration. Par le lemme 4.21, on a des suites croissantes de fonctions étagées $f_n \to f$, $g_n \to g$ donc $\alpha f_n + \beta g_n$ est une suite croissante de fonctions étagées et $\alpha f_n + \beta g_n \to \alpha f + \beta g$. Par le TCM ou le corollaire précédent, en passant à la limite dans l'égalité du lemme 4.20 :

$$\int_B \alpha f_n + \beta g_n d\mu = \alpha \int_B f_n d\mu + \beta \int_B g_n d\mu$$
$$\to \int_B \alpha f + \beta g d\mu = \alpha \int_B f d\mu + \beta \int_B g d\mu.$$

□

★ **Corollaire 4.26** (Interversion Série–intégrale : cas positif)**.** Soient $f_n : \Omega \to [0, +\infty]$ une suite de fonctions mesurables positives alors la somme $\sum_{n \geq 0} f_n : \Omega \to [0, +\infty]$ est mesurable et on a pour tout $B \in \mathcal{T}$:

$$\int_B \sum_{n \geq 0} f_n d\mu = \sum_{n \geq 0} \int_B f_n d\mu.$$

Démonstration. La suite des sommes partielles $S_n = \sum_{k=0}^{n} f_k$ est croissante mesurable (par somme finie). Le résultat est donc une application du TCM. □

Lemme de Fatou

★ **Théorème 4.27** (Lemme de Fatou)**.** Soient $B \in \mathcal{T}$ et $X_n : (\Omega, \mathcal{T}) \to [0, +\infty]$, une suite de fonctions mesurables positives alors $\liminf_{n\to\infty} X_n$ est mesurable et

$$\int_B \liminf_{n\to\infty} X_n d\mu \leq \liminf_{n\to\infty} \int_B X_n d\mu.$$

Démonstration. La mesurabilité de $\liminf_{n\to\infty} X_n$ vient du théorème 4.18.

Par définition, $\liminf_{n\to\infty} X_n = \sup_m Z_m$ pour la suite croissante $Z_m = \inf_{n\geq m} X_n \leq X_m$. En particulier, par monotonie de l'intégrale, $\int_B Z_m \, d\mu \leq \int_B X_n \, d\mu$ pour $n \geq m$, donc en passant à l'infimum : $\int_B Z_m \, d\mu \leq \inf_{n\geq m} \int_B X_n \, d\mu$.

Par le théorème de convergence monotone, on obtient (en combinant à l'inégalité ci-dessus) :

$$\begin{aligned}\int_B \liminf_{n\to\infty} X_n \, d\mu &= \lim_{m\to\infty} \int_B Z_m \, d\mu = \sup_m \int_B Z_m \, d\mu \\ &\leq \sup_m \inf_{n\geq m} \int_B X_n \, d\mu \equiv \liminf_{n\to\infty} \int_B X_n \, d\mu.\end{aligned}$$

□

4 Intégrale des fonctions intégrables

Comme pour les séries et les intégrales impropres en L2, le deuxième cas après le cas positif est celui qu'on appelle “absoluement convergent" pour les séries ou “intégrable" pour les intégrales. Ils ont en commun de considérer la même opération (somme de série ou intégrale) pour la valeur absolue, et si la grandeur obtenue est finie, on peut alors définir l'opération sans valeur absolue. On suit la même stratégie pour l'intégrale de Lebesgue.

On aura besoin de la :

Remarque 4.5. Soit $f : (\Omega, \mathcal{T}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une fonction mesurable, sa **partie positive** est $f_+ = \max(f, 0)$ et sa **partie négative** est $f_- = \max(-f, 0)$. f_+, f_- et la valeur absolue $|f|$ sont mesurables par composée de f avec des applications continues. Elles vérifient $f = f_+ - f_-$ et $|f| = f_+ + f_-$.
De même, pour $f : (\Omega, \mathcal{T}) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ une fonction mesurable, son module $|f|$, et ses parties réelles et imaginaires $\Re(f), \Im(f)$ sont mesurables et

$$f = \Re(f) + i\Im(f) = \Re(f)_+ - \Re(f)_- + i\Im(f)_+ - i\Im(f)_-.$$

★ **Définition 4.13.** Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, une fonction mesurable $f : (\Omega, \mathcal{T}) \to \overline{\mathbb{R}})$ est **intégrale par rapport à μ sur** $B \in \mathcal{T}$ si son module $|f| : (\Omega, \mathcal{T}) \to [0, +\infty]$ est d'intégrale finie sur B, i.e. $\displaystyle\int_B |f| d\mu < +\infty$. On note $\mathcal{L}^1(\Omega, \mathcal{T}, \mu)$ l'ensemble des fonctions intégrables à valeur $\mathbb{R}$.

Si $f : (\Omega, \mathcal{T}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ est intégrable sur B, on a donc $\int_B f_+ d\mu, \int_B f_- d\mu \leq \int_B |f| d\mu < +\infty$ et on peut définir **l'intégrale de f par rapport à μ sur B** :

$$\int_B f d\mu = \int_B f_+ d\mu - \int_B f_- d\mu.$$

Si on dit f est intégrable, c'est qu'on veut implicitement dire sur Ω, son ensemble de définition. Dans ce cas, on écrit aussi : $\int f d\mu = \int_\Omega f d\mu$.

Définition 4.14. Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, une fonction mesurable $f : (\Omega, \mathcal{T}) \to \mathbb{C}$ (resp. $f = (f_1, \cdots, f_n) : (\Omega, \mathcal{T}) \to \mathbb{R}^n$)) est **intégrale par rapport à μ sur $B \in \mathcal{T}$** si ses parties réelles et imaginaire $\Re f, \Im f : (\Omega, \mathcal{T}) \to \mathbb{R}$ (resp. ses coordonnées f_i) sont intégrables sur B, i.e. de façon équivalente si $\int_B |f| d\mu < +\infty$. On note $\mathcal{L}^1(\Omega, \mathcal{T}, \mu; \mathbb{C})$ l'ensemble des fonctions intégrables à valeur $\mathbb{C}$.
On pose alors :

$$\int_B f d\mu = \int_B \Re f d\mu + i \int_B \Im f d\mu \in \mathbb{C},$$
$$(resp. \int_B f d\mu = \left(\int_B f_1 d\mu, \cdots, \int_B f_n d\mu \right) \in \mathbb{R}^n)$$

L'équivalence vient de $\int_B |\Re f| d\mu, \int_B |\Im f| d\mu \leq \int_B |f| d\mu \leq \int_B |\Re f| d\mu + \int_B |\Im f| d\mu$.

Premières propriétés

Lemme 4.28. Si $f : (\Omega, \mathcal{T}, \mu) \to \overline{\mathbb{R}}$ est intégrable (sur Ω), alors $\mu(\{\omega : |f|(\omega) = +\infty\}) = 0$

Démonstration. En effet, si $A = \{\omega : |f|(\omega) = +\infty\}$, on a $(+\infty)1_A \leq |f|$ et donc $+\infty\mu(A) \leq \int_B |f| d\mu < +\infty$ ce qui n'est possible que pour $\mu(A) = 0$. □

★ **Lemme 4.29.** Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, et $f, g : (\Omega, \mathcal{T}) \to \mathbb{K}$ des fonctions intégrables sur $B \in \mathcal{T}$, alors

1. $1_B f$ est intégrable sur Ω et $\int_B f d\mu = \int_\Omega 1_B f d\mu$.
2. (linéarité) Si $\alpha, \beta \in \mathbb{K}$ alors $\alpha f + \beta g$ est intégrable sur B et
$$\int_B \alpha f + \beta g d\mu = \alpha \int_B f d\mu + \beta \int_B g d\mu.$$
3. (domination) Si $h : (\Omega, \mathcal{T}) \to \mathbb{K}$ est mesurable et dominée par $|f|$ au sens $|h| \leq |f|$ alors h est intégrable sur B.

4. (inégalité triangulaire) Si $\mathbb{K}=\mathbb{R}$, on a :

$$\left|\int_B f d\mu\right| \leq \int_B |f| d\mu.$$

On verra le cas complexe de l'inégalité triangulaire un peu plus loin.

Démonstration. 1. Vu $|1_B f| = 1_B |f|$, en utilisant le cas positif du lemme 4.22, on a $\int_\Omega |1_B f| d\mu = \int_B |f| d\mu < +\infty$ d'où l'intégrabilité. Le calcul de l'intégral se déduit alors du même résultat en prenant partie positive et négative des parties réelles et imaginaires.

2. Par l'inégalité triangulaire $|\alpha f + \beta g| \leq |\alpha||f| + |\beta||g|$, donc en passant à l'intégrale et utilisant le cas positif de la linéarité de l'intégrale (Corollaire 4.25) :

$$\int_B |\alpha f + \beta g| d\mu \leq \int_B |\alpha||f| + |\beta||g| d\mu = |\alpha| \int_B |f| d\mu + |\beta| \int_B |g| d\mu < +\infty.$$

De même, l'égalité des intégrales vient en prenant partie positive et négative des parties réelles et imaginaires.

3. Il suffit d'utiliser la monotonie de l'intégrale $\int_B |h| d\mu \leq \int_B |f| d\mu < +\infty$.

4. Dans le cas réel, on a utilise juste l'inégalité triangulaire :

$$\left|\int_B f d\mu\right| = \left|\int_B f_+ d\mu - \int_B f_- d\mu\right| \leq \int_B f_+ d\mu + \int_B f_- d\mu = \int_B |f| d\mu.$$

□

Théorème de Convergence dominée de Lebesgue

★ **Théorème 4.30** (Théorème de Convergence dominée ou TCD). Soient $Z_n, Z : (\Omega, \mathcal{T}, \mu) \to \mathbb{K}$ des fonctions mesurables et $A \in \mathcal{T}$ avec $\mu(A^c) = 0$ satisfaisant :

1. (Condition de domination) il existe une fonction Y intégrable (positive) telle que $|Z_n| \leq Y$,
2. pour tout $\omega \in A$, $Z_n(\omega) \to Z(\omega)$

alors on a :

3. Z est intégrable
4. $\int_\Omega |Z_n - Z| d\mu \to 0$
5. on peut intervertir limite et intégrale

$$\lim_{n\to\infty} \int_\Omega Z_n d\mu = \int_\Omega Z d\mu = \int_A \lim_{n\to\infty} Z_n d\mu.$$

Définition 4.15. Si une propriété est vraie sur un ensemble $A \in \mathcal{T}$ avec $\mu(A^c) = 0$, on dit que A est vraie **presque partout**.

L'hypothèse 2. se formule en disant que Z_n converge vers Z presque partout. On étudiera cette notion avec plus de détail au chapitre suivant.

Démonstration. En appliquant aux parties réelles et imaginaires, il suffit de montrer le cas $\mathbb{K} = \mathbb{R}$.

1. L'inégalité $|Z_n| \leq Y$ implique en passant à la limite $|Z| \leq Y$ sur A, ou autrement dit par domination, Z est intégrable sur A. Comme $\mu(A^c) = 0$, on a aussi $|Z| \leq Y + \infty 1_{A^c}$ et $Y + \infty 1_{A^c}$ est aussi intégrable, donc Z est même intégrable.

3. L'inégalité $|Z_n| \leq Y$ se traduit aussi par $Y - Z_n$, $Z_n + Y \geq 0$ et on peut appliquer le lemme de Fatou 4.27 :

$$\begin{aligned}\int_A (Y - Z)\,d\mu &= \int_A \liminf_n (Y - Z_n)\,d\mu \\ &\leq \liminf_n \int_A (Y - Z_n)\,d\mu \\ &= \int_A Y d\mu - \limsup_n \int_A Z_n\,d\mu,\end{aligned}$$

$$\begin{aligned}\int_A (Y + Z)\,d\mu &= \int_A \liminf_n (Y + Z_n)\,d\mu \\ &\leq \liminf_n \int_A (Y + Z_n)\,d\mu \\ &= \int_A Y d\mu + \liminf_n \int_A Z_n\,d\mu,\end{aligned}$$

donc en soustrayant le terme en Y,

$$\int_A Z d\mu \leq \liminf_n \int Z_n\,d\mu \leq \limsup_n \int Z_n\,d\mu \leq \int_A Z d\mu$$

et on en déduit donc l'égalité et la dernière convergence.

2. Enfin, par l'inégalité triangulaire, on déduit $|Z_n - Z| \leq |Z_n| + |Z| \leq 2Y$ sur A et il satisfait la même condition de domination et pour tout $\omega \in A$, $|Z_n - Z|(\omega) \to 0$. En appliquant le reste du résultat, on obtient donc $\int_\Omega |Z_n - Z|\,d\mu \to \int_\Omega 0\,d\mu = 0$ □

★ **Corollaire 4.31** (Interversion Série–intégrale : cas général)**.** Soient $f_n : \Omega \to \mathbb{K}$ une suite de fonctions mesurables telle que $\sum_{n\geq 0}\int_B |f_n| d\mu < \infty$ pour $B \in \mathcal{T}$, alors la somme $\sum_{n\geq 0} f_n : \Omega \to \mathbb{K}$ converge (absolument) pour presque tout ω dans B et est intégrable sur B et on a :

$$\int_B \sum_{n\geq 0} f_n d\mu = \sum_{n\geq 0}\int_B f_n d\mu.$$

Démonstration. On considère la suite des sommes partielles $S_n = \sum_{k=0}^{n} f_k$ qui vérifie, grâce à l'inégalité triangulaire, la condition de domination $|S_k| \leq \sum_{k=0}^{n} |f_k| \leq \sum_{k=0}^{\infty} |f_k| =: Z$. Or par le cas positif de l'interversion, $\int_B Z d\mu = \sum_{n\geq 0}\int_B |f_n| d\mu < \infty$ donc Z est intégrable sur B. Soit $A = \{\omega \in B : Z(\omega) < \infty\}$, de sorte que $\sum_k f_k$ converge absolument sur A donc S_n converge simplement vers la somme (qui est donc mesurable par le théorème 4.18). Par le lemme 4.28 on a $\mu(A^c) = 0$ donc le TCD s'applique (sur B à la place de Ω) et donne le résultat. □

5 Théorème de transfert

★ **Théorème 4.32** (Théorème de transfert)**.** Soit $f : (\Omega, \mathcal{T}, \mu) \to (E, \mathcal{E})$ une fonction mesurable de mesure image μ_f et $h : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une autre fonction mesurable. Alors, si h est à valeur positive :

$$\int (h \circ f)\ d\mu = \int_E h(x)\ d\mu_f(x).$$

De plus, si h n'est pas à valeur positive $h \circ f \in L^1(\Omega, \mathcal{T}, \mu)$ si et seulement si $h \in L^1(E, \mathcal{E}, \mu_f)$ et on a encore $\int (h \circ f) d\mu = \int h(x) d\mu_f(x)$.

Autrement dit, on ramène une intégrale sur Ω à une intégrale sur $\mathbb{R}$:

$$\int_\Omega h(f(\omega)) d\mu(\omega) = \int_{\mathbb{R}} h(x) d\mu_f(x).$$

Démonstration. On procède comme pour la construction de l'intégrale. Si $h = 1_B$ avec $B \in \mathcal{E}$, $h \circ f = 1_{f^{-1}(B)}$ et donc

$$\int h \circ f d\mu = \mu(f^{-1}(B)) = \mu_f(B) = \int h(x) d\mu_f(x).$$

Par linéarité, on obtient le cas de h étagé. Si h positive, h est la limite croissante d'une suite de fonctions étagées h_n (du lemme 4.21). Comme $h_n(x) \to h(x)$ par construction, on applique le théorème de convergence monotone aux deux mesures :

$$\begin{aligned}\int h \circ f \, d\mu &= \lim_{n\to\infty} \int (h_n \circ f) d\mu \\ &= \lim_{n\to\infty} \int h_n(x) d\mu_f(x) = \int h(x) d\mu_f(x).\end{aligned}$$

Le dernier résultat du cas intégrable est évident par le cas positif pour l'équivalence et par linéarité pour l'égalité. □

Le résultat similaire suivant est important en probabilité. Nous avons vu la tribu engendrée par f : $\sigma(f)$ au lemme 4.8. Le résultat suivant donne une interprétation concrète des fonctions $\sigma(f)$−mesurables.

Proposition 4.33 (Lemme de Doob–Dynkin). Soit f une fonction mesurable, $f : (\Omega, \mathcal{T}, \mu) \to (E, \mathcal{E})$, et soit $\sigma(f) = \{A = f^{-1}(B), B \in \mathcal{E}\}$ la tribu engendrée par f. Alors $g : \Omega \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ est $\sigma(f)$−mesurable si et seulement si il existe $h : (E, \mathcal{E}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ mesurable telle que $g = h \circ f$.

Démonstration. La condition suffisante est évidente car pour un borélien A, $(h \circ f)^{-1}(A) = f^{-1}(h^{-1}(A))$ qui est mesurable car $h^{-1}(A) \in \mathcal{E}$ car h borélienne et l'image inverse par f est alors par définition un élément de $\sigma(f)$.

Réciproquement, on raisonne comme pour le transfert par le cas étagé $g = \sum_i \lambda_i 1_{A_i}$ et $A_i = f^{-1}(B_i)$ et alors $h = \sum_i \lambda_i 1_{B_i}$ convient. Sinon, si g positive, on la prend pour limite simple de g_n étagée de la forme $h_n \circ f$ par le cas étagé, et on pose

$$h(x) = \liminf_{n\to\infty} h_n(x).$$

h convient car mesurable positive (comme lim inf de fonctions mesurables) et car $g(\omega) = \lim_n h_n(f(\omega)) = h(f(\omega))$ vu qu'en $f(\omega)$ la suite (h_n) converge d'après le choix de g_n. Le cas général se montre par linéarité à partir du cas positif. □

6 Comparaison aux constructions de L2

Intégrale de Riemann des fonctions continues par morceau

Comme on a vu au chapitre 2, la base de l'intégrale de Riemann est la notion de fonction en escalier. Ce sont des combinaisons linéaires d'indicatrices d'intervalles de forme $1_{]a,b[}$ et $1_{\{c\}}$. Or les intervalles sont des boréliens, donc les fonctions en escalier sont boréliennes étagées. On a

$$\int 1_{]a,b[}d\lambda = (b-a) = \int 1_{]a,b[}(x)dx, \int 1_{\{c\}}d\lambda = 0 = \int 1_{\{x\}}(x)dx,$$

donc par combinaison linéaire, intégrale de Riemann et de Lebesgue par rapport à la mesure de Lebesgue coïncident.

Soit f continue par morceau sur $[a,b]$, l'intégrale de Riemann est construite en choisissant f_n en escalier convergent uniformément vers f et donc simplement, donc f est borélienne comme limite simple de fonctions boréliennes (cf. le théorème 4.18). De plus, elle est bornée donc intégrable sur $[a,b]$.

Quitte à décomposé en partie réelle et imaginaire, on suppose f réelle. DOnc pour tout $x \in [a,b]$ on a $|f(x) - f_n(x)| \leq ||f_n - f||_\infty$ soit

$$f_n(x) - ||f_n - f||_\infty \leq f(x) \leq f_n(x) + ||f_n - f||_\infty.$$

En intégrant au sens de Lebesgue, et en utilisant que les deux côtés coïncident avec celle de Riemann, on obtient l'inégalité :

$$\begin{aligned}&\int_a^b f_n(x)dx - ||f_n - f||_\infty(b-a)\\ &\leq \int_{[a,b]} fd\lambda \leq \int_a^b f_n(x)dx + ||f_n - f||_\infty(b-a).\end{aligned}$$

En passant à la limite $n \to \infty$, on a $||f_n - f||_\infty \to 0$ et $\int_a^b f_n(x)dx \to \int_a^b f(x)dx$ par définition de l'intégrale de Riemann. On a donc obtenu :

★ **Théorème 4.34.** 1. Toute fonction continue par morceau sur un segment $[a,b]$ est intégrable par rapport à la mesure de Lebesgue λ et son intégrale de

Riemann coïncide avec celle pour la mesure de Lebesgue :

$$\int_a^b f(x)dx = \int_{[a,b]} fd\lambda.$$

2. Toute fonction continue par morceau **intégrable** sur un intervalle I ($]a,b]$, $]a,b[$ ou $[a,b[$) est intégrable par rapport à la mesure de Lebesgue λ et son intégrale de Riemann coïncide avec celle pour la mesure de Lebesgue : $\int_a^b f(x)dx = \int_I fd\lambda$.

On pourra donc appliquer les théorèmes précédents aux intégrales (de Riemann) usuelles vues en L2.

Remarque 4.6. Pour les fonctions $f : [a,b] \to \mathbb{R}$, on peut définir une notion plus générale de fonction "Riemann intégrable", elle même plus générale que l'intégrale des fonctions continues par morceaux. L'intégrale de Lebesgue généralise aussi cette version plus générale, cf. e.g. http://math.univ-lyon1.fr/homes-www/mironescu/resources/cours_mesure_integration.pdf section 6.8.1

Mesures à densité

Le résultat suivant est laissé en exercice

★ **Proposition 4.35** (Mesures à densité (ou absolument continue)). Soit $f : X \to [0,+\infty]$ une fonction mesurable. On définit une application $\nu : \mathcal{A} \to [0,+\infty]$ par

$$\nu(A) = \int_A fd\mu \ .$$

Alors, ν est une mesure sur X, appelée **mesure de densité** f par rapport à μ. De plus h est intégrable par rapport à ν si et seulement si fh est intégrable par rapport à μ et :

$$\int_X hd\nu = \int_X fhd\mu \ .$$

Pour une mesure à densité ν par rapport à μ, si $\mu(A) = 0$ alors $\nu(A) = 0$. En fait, cette propriété caractérise les mesures à densité (c'est un théorème beaucoup plus dur, le théorème de Radon-Nikodym cf. section 5)

Exemple 4.7. On peut définir une mesure de probabilité sur les boréliens de $\mathbb{R}$ en posant

$$\mu(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-\frac{x^2}{2}} \, d\lambda(x) \ .$$

Cette mesure s'appelle la **mesure gaussienne**. C'est un exemple de probabilité à densité par rapport à la mesure de Lebesgue. Pour vérifier qu'il s'agit bien d'une probabilité, il faut vérifier que :

$$\mu(\mathbb{R}) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} \, d\lambda(x) = 1.$$

On le vérifiera plus loin par changement de variable à la fin du chapitre 5 à la formule (5.1)

Lien avec les Séries

Soit Ω un ensemble. On considère l'espace mesuré $(\Omega, \mathcal{P}(\Omega), \nu)$. Tout fonction $f : \Omega \to \overline{\mathbb{R}}$ est $\mathcal{P}(\Omega)$–mesurable. On peut donc ignorer la mesurabilité pour le cas des séries.

Cas $\Omega = \{\omega_1, \cdots, \omega_n\}$ fini

Toute fonction s'écrit $f = \sum_{k=1}^{n} f(\omega_k) 1_{\{\omega_k\}}$ et est donc étagée. On déduit que $\int_\Omega f d\nu = \sum_{k=1}^{n} f(\omega_k)$, d'abord pour les fonctions étagées, puis positives, puis quelconques (on peut prendre toutes les limites constantes).

Cas $\Omega = \mathbb{N}$

Lemme 4.36.
1. Si $f \geq 0$ alors $\int_\Omega f d\nu = \sum_{n=0}^{\infty} f(n)$
2. f est intégrable si et seulement si $\sum f(n)$ est absolument convergente et encore

$$\int_\Omega f d\nu = \sum_{n=0}^{\infty} f(n).$$

Démonstration. 1) Soit $f_n = \sum_{k=1}^{n} f(k) 1_{\{k\}}$ est une suite croissante de fonctions donc par le TCM $\int_\Omega f d\nu = \lim_n \int_\Omega f_n d\nu = \lim_n \sum_{k=0}^{n} f(k) = \sum_{n=0}^{\infty} f(n)$

2) L'équivalence vient du 1) f est intégrable ssi $|f|$ a une intégrale fini, donc ssi $\sum_{n=0}^{\infty} |f(n)|$ c'est à dire ssi $\sum f(n)$ est absolument convergente. La définition de l'intégrale et de la somme coïncident alors

$$\int_{\Omega} f d\nu = \int_{\Omega} f_+ d\nu - \int_{\Omega} f_- d\nu = \sum_{n=0}^{\infty} f(n)_+ - \sum_{n=0}^{\infty} f(n)_- = \sum_{n=0}^{\infty} f(n).$$

□

Cas $\Omega = \{\omega_n, n \in \mathbb{N}\}$ dénombrable

On a $\omega : \mathbb{N} \to \Omega$ une bijection, donc la mesure image $\nu_\omega(\{i\}) = \nu(\{\omega^{-1}(i)\} = 1 = \nu(\{i\})$ est encore la mesure de comptage, le théorème de transfert donne donc :

Lemme 4.37. Pour tout $f : \Omega \to [0, +\infty]$,

$$\int_{\Omega} f d\nu = \int_{\mathbb{N}} f(\omega) d\nu = \sum_{n=0}^{\infty} f(\omega_n)$$

En particulier, si $\sigma : \mathbb{N} \to \mathbb{N}$ est une bijection $\sum_{n=0}^{\infty} f(\sigma(n)) = \sum_{n=0}^{\infty} f(n)$ et le même résultat est valide pour les séries absolument convergentes (on dit qu'elles sont **commutativement convergentes**.) Aussi $L^1(\Omega, \nu) = \ell^1(\Omega)$ est l'ensemble des familles sommables sur Ω avec la norme 1.

Probabilité discrète sur $\Omega = \{\omega_n, n \in \mathbb{N}\}$ dénombrable

C'est une densité $f : \Omega \to [0, +\infty]$ par rapport à la mesure de comptage telle que $\int_\Omega f d\nu = \sum_{n=0}^{\infty} f(\omega_n) = 1$.

7 Intégrales dépendant d'un paramètre

Soient $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, E un evn. Soit finalement A une partie de E.

Définition 4.16. Soit $f : A \times \Omega \to \mathbb{K}$. On suppose que pour tout $x \in A$, $t \mapsto f(x, t)$ est intégrable (soit dans $L^1(\Omega, \mathcal{T}, \mu)$). Dans ce cas, on peut poser :
$F(x) = \int_\Omega f(x, t) d\mu(t)$. On définit ainsi une **intégrale dépendant d'un paramètre** la fonction $F : A \to \mathbb{K}$.

★ **Théorème 4.38** (Théorème de continuité avec hypothèse de domination)**.** Soit $f : A \times \Omega \to \mathbb{K}$. On suppose :

1. Pour tout $x \in A$, $t \mapsto f(x,t)$, est mesurable sur Ω.
2. Pour tout presque tout $t \in \Omega$, $x \mapsto f(x,t)$ est continue en $x_0 \in A$.
3. (Hypothèse de domination) Il existe une fonction intégrable $g : \Omega \to \mathbb{R}_+$, $g \in L^1(\Omega, \mathcal{T}, \mu)$ telle que

$$\forall t \in \Omega, \forall x \in A, \quad |f(x,t)| \leq g(t).$$

Alors la fonction $x \mapsto F(x) = \int_\Omega f(x,t)d\mu(t)$ est continue en x_0.

On remarquera que dans l'hypothèse de domination, la fonction g ne dépend PAS de x.

Démonstration. L'hypothèse de domination garantit que $t \mapsto f(x,t)$ est intégrable. Soit $x_n \in A$ tel que $x_n \to x_0$. Par continuité de $x \mapsto f(x,t)$, pour chaque t, $f(x_n,t) \to f(x_0,t)$. On peut donc appliquer le théorème de convergence dominée (avec domination par g) pour conclure

$$\lim_{n\to\infty} \int_\Omega f(x_n,t)d\mu(t) = \int_\Omega f(x_0,t)d\mu(t).$$

□

Exemple 4.8 ((cf TD.))*.* Soit $f : \mathbb{R} \to \mathbb{C}$ intégrable sur $\mathbb{R}$ (par rapport à la mesure de Lebesgue λ). Sa **transformée de Fourier** est définie par :

$$\hat{f}(x) = \int_\mathbb{R} f(t)e^{itx}dt.$$

Elle est continue sur $\mathbb{R}$ en utilisant une domination par $|f|$.

Théorème 4.39 (Théorème de dérivabilité avec hypothèse de domination)**.** Soit $f : U \times \Omega \to \mathbb{K}$ avec $U \subset \mathbb{R}^n$ un ouvert.
On suppose :

1. Pour tout $x \in U$, $t \mapsto f(x,t)$, est intégrable sur Ω.
2. Il existe N avec $\mu(N^c) = 0$, tel que pour tout $t \in N$, la fonction $x \mapsto f(x,t)$ admet une i-ème dérivée partielle sur U.

3. (Hypothèse de domination) Pour tout compact $K \subset U$, il existe une fonction intégrable $g_K \in L^1(\Omega)$ telle que
$$\forall t \in N, \forall x \in K, \quad \left|\frac{\partial f}{\partial x_i}(x,t)\right| \leq g_K(t).$$

Alors la fonction $x \mapsto F(x) = \int_\Omega f(x,t)d\mu(t)$ admet une i-ème dérivée partielle sur U, $\frac{\partial f}{\partial x_i} \in L^1(\Omega)$ et :
$$\frac{\partial F}{\partial x_i}(x) = \int_\Omega \frac{\partial f}{\partial x_i}(x,t)d\mu(t).$$

Remarque 4.7. Soit $f = (f_1, \dots, f_m) : U \times \Omega \to \mathbb{R}^m$ avec $U \subset \mathbb{R}^n$ un ouvert. Si chaque $f_i(x,.)$ est intégrable sur Ω pour tout $x \in U$, on peut définir l'intégrale coordonnée par coordonnée :
$$\int_\Omega f(x,t)d\mu(t) = (\int_\Omega f_1(x,t)d\mu(t), \cdots, \int_\Omega f_n(x,t)d\mu(t)).$$
Alors le théorème s'applique en remplaçant la valeur absolue par la norme dans la domination (et en appliquant le résultat coordonnée par coordonnée.)

Démonstration. On peut supposer $n = m = 1$ (car les dérivées partielles se calculent coordonnée par coordonnée). On fixe x_0 et montre la dérivabilité en x_0. On pose $h(x,t) = 0$ si $t \in N^c$ et pour $t \in N$
$$h(x,t) = \begin{cases} \frac{f(x,t)-f(x_0,t)}{x-x_0}, & \text{si } \ x \neq x_0 \\ frac\partial f \partial x(x_0,t) & \text{sinon} \end{cases}.$$

Pour $x \neq x_0$,
$$\frac{F(x) - F(x_0)}{x - x_0} = \int_\Omega h(x,t)d\mu(t).$$

Il suffit donc de prouver que $x \mapsto \int_\Omega h(x,t)d\mu(t)$ est continue en x_0. Par hypothèse, $t \mapsto h(x,t)$ est mesurable pour $x \neq x_0$ et par exemple en tant que lim inf (sur N) aussi ex x_0 et $x \mapsto h(x,t)$ est continue pour $t \in N$ (par continuité d'une fonction dérivable d'une variable). Enfin l'inégalité des accroissements finis à $x \mapsto f(x,t)$ donne, pour $x \neq x_0$, $x \in K = [x_0 - \epsilon, x_0 + \epsilon] \subset U$ (un compact car fermé borné de $\mathbb{R}$ contenu dans U pour ϵ assez petit) :
$$||h(x,t)|| \leq \sup_{u \in [x_0,x]} \left|\frac{\partial f}{\partial x_i}(u,t)\right| \leq g_K(t).$$
La même inégalité étant évidente en x_0, on a la condition de domination et le théorème de continuité appliqué à K conclut. □

★ **Corollaire 4.40** (Théorème de dérivation successive). Soit $f : U \times V \to \mathbb{R}^l$ avec $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ des ouverts, une fonction C^k ($k \in \mathbb{N} \cup \{\infty\}$). Soit μ une mesure sur une tribu $\mathcal{T} \supset \mathcal{B}(V)$.
On suppose qu'il existe $\phi_0, \phi_1, \ldots, \phi_k$ μ−intégrables sur V telles que $||f(x,t)|| \leq \phi_0(t)$ et
Alors la fonction $x \mapsto F(x) = \int_V f(x,t)d\mu(t)$ est de classe C^k sur U et pour $p = i_1 + \ldots + i_n \leq k$:

$$\frac{\partial^p F}{\partial x_1^{i_1} \ldots \partial x_n^{i_n}}(x) = \int_V \frac{\partial^p f}{\partial x_1^{i_1} \ldots \partial x_n^{i_n}}(x,t)d\mu(t).$$

Démonstration. Il suffit d'appliquer le théorème de dérivation avec condition de domination par récurrence simple (coordonnées f_i par coordonnée $f = (f_1, \cdots, f_n)$) . La mesurabilité de f vient de sa continuité vu que $\mathcal{T}$ contient les boréliens. Son intégrabilité vient de la domination $|f(x,t)| \leq \phi_0(t)$ et sur les autres dérivées successives des autres dominations. On peut prendre $N = \emptyset$. □

Un exemple : la transformée de Fourier d'une mesure avec moments d'ordre 2.

Soit μ une mesure (de masse) finie sur $\mathcal{B}(\mathbb{R}^n)$ (par exemple une probabilité à densité par rapport à λ) tel que $x_i, x_i x_j, i, j = 1, \cdots, n$ sont intégrables c'est à dire :

$$\int_{\mathbb{R}^n} |x_i| d\mu(x) < +\infty, \qquad \int_{\mathbb{R}^n} |x_i x_j| d\mu(x) < +\infty.$$

On verra plus tard grâce à l'inégalité de Cauchy−Schwarz qu'il suffit de supposer x_i^2 intégrable. On reprend la transformée de Fourier vu en TD et à l'exemple 4.8 qui est définie par :

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^n} e^{i\langle \xi, x\rangle} d\mu(x) = \int_{\mathbb{R}} f(\xi, x) d\mu(x), \qquad f(\xi, x) = e^{i\langle \xi, x\rangle}.$$

f est C^2 (même C^∞) sur $\mathbb{R}^{2n}$ et vérifie les dominations :

$$\begin{aligned} & & |f(\xi,x)| \leq 1 \\ \frac{\partial}{\partial \xi_i} f(\xi,x) &= i x_i e^{i\langle \xi, x\rangle}, & \left|\frac{\partial}{\partial \xi_i} f(\xi,x)\right| \leq |x_i| \\ \frac{\partial^2}{\partial \xi_i \partial \xi_j} f(\xi,x) &= -x_i x_j e^{i\langle \xi, x\rangle}, & \left|\frac{\partial^2}{\partial \xi_i \partial \xi_j} f(\xi,x)\right| \leq |x_i x_j| \end{aligned}$$

et par l'hypothèse μ de masse finie, 1 est intégrable et par les hypothèses d'intégrabilité, les autres membres de droite des dominations sont intégrables aussi par rapport à μ. Par le théorème de dérivation avec condition de domination, on déduit donc que $\hat{\mu}$ est C^2 et :

$$\frac{\partial}{\partial \xi_i}\hat{\mu}(\xi) = i \int_{\mathbb{R}^n} x_i e^{i\langle \xi, x\rangle} d\mu(x)$$
$$\frac{\partial^2}{\partial \xi_i \partial \xi_j}\hat{\mu}(\xi) = -\int_{\mathbb{R}^n} x_i x_j e^{i\langle \xi, x\rangle} d\mu(x).$$

Cet exemple sera utilisé au S6 pour montrer le Théorème centrale limite dans $\mathbb{R}^n$.

CHAPITRE 5

Intégration avancée : Théorème de Fubini, Changements de variables

1 Mesure produit et théorèmes de Fubini

Tribus produits

La méthode de base pour calculer une intégrale d'une fonction de 2 variables est de se ramener à des intégrales de fonctions de 1 variable. Pour cela il nous faut d'abord expliquer comment on peut munir $X \times Y$ d'une structure d'espace mesuré quand X, Y sont tous les deux munis d'une telle structure.

★ **Définition 5.1.** Soient $(X, \mathcal{A}, \mu_1)$ et $(Y, \mathcal{B}, \mu_2)$ deux espaces mesurés σ−finis. On note $\mathcal{A} \otimes \mathcal{B}$ la tribu engendrée par les parties de la forme $A \times B$, où $A \in \mathcal{A}$, $B \in \mathcal{B}$; on l'appelle **tribu produit** des tribus $\mathcal{A}$ et $\mathcal{B}$.

Lemme 5.1. Si $\mathcal{A} = \sigma(\mathcal{E})$ et $\mathcal{B} = \sigma(\mathcal{F})$, on a $\mathcal{A} \otimes \mathcal{B} = \sigma\Big(\{E \times F,\, E \in \mathcal{E},\, F \in \mathcal{F}\}\Big)$.
En particulier, $\mathcal{B}(\mathbb{R}^{n+m}) = \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m)$. De plus, si $f : (X, \mathcal{A}) \to (Z, \mathcal{C})$ et $g : (Y, \mathcal{B}) \to (Z, \mathcal{D})$ sont mesurables, l'application $(f, g) : (X \times Y, \mathcal{A} \otimes \mathcal{B}) \to (Z \times T, \mathcal{C} \otimes \mathcal{D})$ définie par $(f, g)(x, y) = (f(x), g(y))$ est mesurable.

Démonstration. Vu $\{E \times F,\, E \in \mathcal{E},\, F \in \mathcal{F}\} \subset \mathcal{A} \otimes \mathcal{B}$, on obtient en passant à la tribu engendrée $\mathcal{G} := \sigma\Big(\{E \times F,\, E \in \mathcal{E},\, F \in \mathcal{F}\}\Big) \subset \mathcal{A} \otimes \mathcal{B}$.

Réciproquement, on pose $\mathcal{A}' = \{A \in \mathcal{A} : \forall F \in \mathcal{F},\, A \times F \in \mathcal{G}\}$. On a clairement que $\mathcal{A}'$ contient $\mathcal{E}$ et on vérifie facilement que c'est une tribu (vu que $A^c \times F = (\Omega \times F) - (A \times F) \in \mathcal{G}$ pour $F \in \mathcal{F}$.) D'où $\mathcal{A}' = \sigma(\mathcal{E}) = \mathcal{A}$. De même, on pose ensuite, $\mathcal{B}' = \{B \in \mathcal{B} : \forall A \in \mathcal{A},\, A \times B \in \mathcal{G}\}$ et on déduit du point précédent que $\mathcal{F} \subset \mathcal{B}' \subset \mathcal{B}$ et comme avant que $\mathcal{B}'$ est une tribu d'où $\mathcal{B} = \mathcal{B}'$. Finalement, on a donc $\mathcal{A} \times \mathcal{B} \subset \mathcal{G}$ d'où l'inclusion complémentaire de tribus.

Le cas particulier $\mathcal{B}(\mathbb{R}^{n+m}) = \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m)$ est une conséquence immédiate du Corollaire 4.16.

Pour le dernier point, comme $\mathcal{C} \otimes \mathcal{D} = \sigma\Big(\{E \times F, E \in \mathcal{C}, F \in \mathcal{D}\}\Big)$ il suffit de noter que $(f,g)^{-1}(E \times F) = f^{-1}(E) \times g^{-1}(F) \in \mathcal{A} \times \mathcal{B} \subset \mathcal{A} \otimes \mathcal{B}$ et le lemme 4.13 conclut. □

Mesure produit

Théorème 5.2 (définissant la mesure produit)**.** Soient $(\Omega_1, \mathcal{T}_1, \mu_1)$ et $(\Omega_2, \mathcal{T}_2, \mu_2)$ deux espaces mesurés σ−finis. Alors il existe une unique mesure ν sur $\mathcal{T}_1 \otimes \mathcal{T}_2$ vérifiant

$$\nu(A \times B) = \mu_1(A)\mu_2(B)$$

pour tout $A \in \mathcal{T}_1$ et tout $B \in \mathcal{T}_2$ (avec la convention usuelle $0.(+\infty) = 0$). Cette mesure est notée $\mu_1 \otimes \mu_2 = \nu$, et est σ−finie.

Exemple 5.1. Si λ_n désigne la mesure de Lebesgue sur $\mathbb{R}^n$, alors on a toujours $\lambda_{n+m} = \lambda_n \otimes \lambda_m$. On applique le corollaire 4.19 au lemme de classe monotone à l'ensemble des pavés $\mathcal{E}$. Par définition, $\lambda_{n+m}, \lambda_n \otimes \lambda_m$ coïncident sur les pavés. Or $\cup_{M \in \mathbb{N}}[-M, M]^{n+m} = R^{n+m}$ et $\lambda_{n+m}([-M, M]^{n+m}) = (2M)^{n+m} = (\lambda_n \otimes \lambda_m)([-M, M]^{n+m}) < +\infty$ donc on conclut à l'égalité voulue.

La preuve va être basée sur le fait de montrer un cas particulier du théorème de Fubini suivant pour les fonctions indicatrices.

Démonstration. **Unicité** On applique le même corollaire 4.19 au lemme de classe monotone. ON prend $\mathcal{E} = \{A \times B, A \in \mathcal{T}_1, B \in \mathcal{T}_2\}$ qui engendre $\mathcal{T}_1 \otimes \mathcal{T}_2$ par définition. Deux mesures ν_1, ν_2 vérifiant le théorème coïncident sur $\mathcal{E}$. Or comme μ_1, μ_2 sont σ−finies, on obtient $\Omega_i = \cup_n A_{i,n}$ avec $A_{i,n} \in \mathcal{T}_i$ et $\mu_i(A_{i,n}) < +\infty$. Alors, on a $A_{1,n} \times A_{2,n} \in \mathcal{E}$ et est de mesure $\mu_1(A_{1,n})\mu_2(A_{2,n}) < +\infty$ pour ν_1, ν_2. Ceci donne la dernière hypothèse du corollaire 4.19 qui conclut à $\mu_1 = \mu_2$.

Existence Pour $C \in \mathcal{T}_1 \otimes \mathcal{T}_2$, on pose $C_x = \{y \in \Omega_2 : (x,y) \in C\}$. On cherche à voir que $C_x \in \mathcal{T}_2$. Supposons d'abord μ_2 finie. On considère

$$\mathcal{C} = \{C \in \mathcal{T}_1 \otimes \mathcal{T}_2 : \forall x C_x \in \mathcal{T}_2 \text{ et } x \mapsto \mu_2(C_x) \text{ est } \mathcal{T}_1 - \text{mesurable}\}.$$

Alors on a

- $\mathcal{C}$ contient les pavés mesurables $C = A \times B$ avec $A \in \mathcal{T}_1$, $B \in \mathcal{T}_2$ car $(A \times B)_x \in \{\emptyset, B\}$ en distinguant le cas $x \in A$, $x \notin A$ donc $\mu_2(C_x) = 1_A(x)\mu_2(B)$.
- $\mathcal{C}$ est une classe monotone car si $C' \subset C$, $C' \in \mathcal{C}$ $(C \setminus C')_x = C_x \setminus C'_x$ d'où la mesurabilité et $\mu_2(C \setminus C')_x = \mu_2(C_x) - \mu_2(C'_x)$ par finitude de μ_2 qui est mesurable par différence donc $C \setminus C' \in \mathcal{C}$. De même si C_n est une suite croissante $(\cup_n C_n)_x = \cup_n (C_n)_x$ qui est dans $\mathcal{T}_2$ et $\mu_2((\cup_n C_n)_x) = \sup_n \mu_2((C_n)_x)$ est bien mesurable.

Donc $\mathcal{C}$ contient la classe monotone engendrée par les pavés, donc (par le lemme de classe monotone) est égale à $\mathcal{T}_1 \otimes \mathcal{T}_2$.

Si μ_2 est σ-finie, on regarde les mesures induites et déduit le même résultat de mesurabilité de $\mu_2(C_x)$ par limite croissante.

On peut donc poser

$$\nu(C) = \int_{\Omega_1} \mu_2(C_x) d\mu_1(x).$$

Il faut voir que c'est une mesure en montrant la σ-additivité : Soient C^n des ensembles mesurables disjoints, (en utilisant qu'alors les C_x^n sont disjoints), il suffit d'utiliser l'interversion série intégrale :

$$\begin{aligned}
\nu(\bigcup_n C^n) &= \int_{\Omega_1} \mu_2(\bigcup_n C_x^n) d\mu_1(x) \\
&= \int_{\Omega_1} \sum_n \mu_2(C_x^n) d\mu_1(x) \\
&= \sum_n \int_{\Omega_1} \mu_2(C_x^n) d\mu_1(x) = \sum_n \nu(C^n).
\end{aligned}$$

Enfin, ν convient par le calcul précédent de $\mu_2((A \times B)_x)$:

$$\nu(A \times B) = \int_{\Omega_1} 1_A(x)\mu_2(B) d\mu_1(x) = \mu_1(A)\mu_2(B).$$

□

Théorème de Fubini-Tonelli et Fubini (admis)

La mesure produit $\mu_1 \otimes \mu_2$ étant définie à partir de μ_1 et μ_2, on s'attend à ce qu'il en soit de même de l'intégrale d'une fonction mesurable relativement à $\mu_1 \otimes \mu_2$.[1] Et c'est effectivement le contenu des théorèmes de Fubini. On commence par le cas positif.

1. Cette sous-section reprend le cours de 2018-2019 de T. Blossier, M. Carrizosa et J. Melleray.

★ **Théorème 5.3** (Fubini-Tonelli)**.** Soient $(\Omega_1, \mathcal{T}_1, \mu_1)$ et $(\Omega_2, \mathcal{T}_2, \mu_2)$ deux espaces mesurés σ-finis. Soit $f : \Omega_1 \times \Omega_2 \to [0, +\infty]$ une fonction $\mathcal{T}_1 \otimes \mathcal{T}_2$-mesurable. Alors :

1. $y \mapsto f(x, y)$ est une fonction mesurable (sur $(\Omega_2, \mathcal{T}_2)$ dans $[0, +\infty]$) pour tout $x \in \Omega_1$, et $x \mapsto \int_{\Omega_2} f(x, y) d\mu_2(y)$ est une fonction mesurable (sur $(\Omega_1, \mathcal{T}_1)$).
2. $x \mapsto f(x, y)$ est une fonction mesurable (sur $(\Omega_1, \mathcal{T}_1)$ dans $[0, +\infty]$) pour tout $y \in \Omega_2$, et $y \mapsto \int_{\Omega_1} f(x, y) d\mu_1(x)$ est une fonction mesurable (sur $(\Omega_2, \mathcal{T}_2)$).
3. On a

$$\begin{aligned}
&\int_{\Omega_1 \times \Omega_2} f(x, y) d\mu_1 \otimes \mu_2(x, y) \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) d\mu_2(y) \right) d\mu_1(x) \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) d\mu_1(x) \right) d\mu_2(y) \ .
\end{aligned}$$

Exercice 5.1*.* Calculer l'aire du disque unité $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$.

Comme dans le cas des fonctions définies sur $\mathbb{R}^n$, on en déduit facilement un théorème qui s'applique à toutes les fonctions intégrables (et pour vérifier qu'une fonction est intégrable, on peut commencer par appliquer le théorème de Fubini-Tonelli à $|f|$).

★ **Théorème 5.4** (Fubini)**.** Soient $(\Omega_1, \mathcal{T}_1, \mu_1)$ et $(\Omega_2, \mathcal{T}_2, \mu_2)$ deux espaces mesurés σ-finis. Soit $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ une fonction **intégrable**. Alors :

1. $y \mapsto f(x, y)$ est une fonction intégrable (sur Ω_2) pour presque tout $x \in \Omega_1$, et $x \mapsto \int_{\Omega_2} f(x, y) d\mu_2(y)$ est une fonction intégrable (sur Ω_1).
2. $x \mapsto f(x, y)$ est une fonction intégrable (sur Ω_1) pour presque tout $y \in \Omega_2$, et $y \mapsto \int_{\Omega_1} f(x, y) d\mu_1(x)$ est une fonction intégrable (sur Ω_2)
3. On a

$$\begin{aligned}
&\int_{\Omega_1 \times \Omega_2} f(x, y) d\mu_1 \otimes \mu_2(x, y) \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) d\mu_2(y) \right) d\mu_1(x) \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) d\mu_1(x) \right) d\mu_2(y) \ .
\end{aligned}$$

Exercice 5.2. Soit f, g des fonctions mesurables positives sur $\mathbb{R}$, on définit la convolution de f, g par :

$$f * g(x) = \int_{\mathbb{R}} f(x-y)g(y)d\lambda(y) \in [0, \infty].$$

On rappelle que

$$||f||_1 = \int_{\mathbb{R}} |f(x)|d\lambda(x).$$

1. Montrer que $f * g$ est mesurable et que

$$||f * g||_1 = ||f||_1 ||g||_1.$$

2. Montrer que la définition de $f * g$ s'étend pour presque tout x au $f, g \in L^1(\mathbb{R}, d\lambda)$ et que $f * g \in L^1(\mathbb{R}, d\lambda)$.
3. Montrer que pour f, g, h toutes mesurables positives ou toutes intégrables, alors

$$f * (g * h) = (f * g) * h.$$

2 Une Inégalité de convexité : l'Inégalité de Jensen

La convexité (ou la concavité) est souvent utilisée pour établir des inégalités.[2]

Voyons maintenant l'inégalité de convexité la plus importante de notre cours.

★ **Théorème 5.5** (Inégalité de Jensen)**.** Soit $(X, \mathcal{A}, \mu)$ un espace de probabilité, g une fonction μ–intégrable à valeurs dans un intervalle I, et $\varphi \colon I \to \mathbb{R}$ une fonction convexe. Alors on a

$$\varphi\left(\int_X g d\mu\right) \leq \int_X \varphi \circ g d\mu$$

(l'intégrale de droite peut être égale à $+\infty$!).

Démonstration. D'abord, par le théorème 3.9, φ est dérivable à droite et à gauche, donc continue sur l'intérieur de I, donc borélienne sur I (exo) donc la composée $\varphi \circ g$ est bien mesurable. Posons $m = \int_X g d\mu$. Notons que $m \in I$. En effet I est définie par une ou deux inégalités, $I = I_1 \cap I_2$ avec ($I_1 = \{x : x \geq a\}$ ou $I_1 = \{x : x > a\}$ ou $I_1 = \mathbb{R}$) et de même ($I_2 = \{x : x \leq b\}$ ou $I_2 = \{x : x < b\}$ ou $I_2 = \mathbb{R}$). Expliquons d'abord que si g est à valeur dans

2. Cette partie reprend le cours de 2018–2019 de T. Blossier, M. Carrizosa et J. Melleray.

$I_1 = \{x : x \geq a\}$, alors comme l'intégrale préserve les inégalités larges $\int_X g d\mu \geq \int_X a d\mu = a$ car $\mu(X) = 1$ et donc $m \in I_1$. De même si $I_1 = \{x : x > a\}$ si on n'avait pas $\int_X g d\mu > a$, on aurait donc $\int_X g d\mu = a = \int_X a d\mu$ donc $\int_X (g - a) d\mu = 0$ mais alors $g - a$ serait nulle μ−presque partout, donc $\{x \in X : g(x) > a\} = X$ serait de mesure nulle, contredisant l'hypothèse que X est un espace de probabilité. On conclut donc aussi dans ce cas $\int_X g d\mu \in I_1$. On raisonne pareil pour I_2 (ou on applique le premier cas à $-g$ pour changer le sens des inégalités).

Maintenant qu'on a vu que $m \in I$, on distingue 3 cas. Si jamais m est le minimum de I (s'il existe !) alors on a $\int_X (g - m) d\mu = 0$ et $g - m \geq 0$, donc $g - m$ est nulle presque partout, par conséquent on a

$$\int_X \varphi \circ g d\mu = \int_X \varphi(m) d\mu = \varphi(m) = \varphi\left(\int_X g d\mu\right) .$$

On traite de même le cas où m est le maximum de I ; finalement, le cas qui nous reste est celui où m appartient à l'intérieur de I.

Alors, on sait que $\varphi'_g(m)$ existe et en posant $\alpha = \varphi'_g(m)$, le théorème 3.9 donne que

$$\forall t \in I \quad \varphi(t) - \varphi(m) \geq \alpha(t - m) .$$

En particulier, pour tout $x \in X$ on a $\varphi(g(x)) \geq \varphi(m) + \alpha(g(x) - m)$. Comme g est intégrable et les fonctions constantes sont intégrables (car μ est finie), donc la borne inférieure est intégrable, et on en déduit que la partie négative de $\varphi \circ g$ est d'intégrale finie ; et en intégrant cette inégalité, on obtient aussi que

$$\begin{aligned}\int_X \varphi \circ g d\mu &\geq \int_X \varphi(m) d\mu + \alpha \int_X (g - m) d\mu \\ &= \varphi(m) + \alpha(\int_X g d\mu - m) = \varphi(m) .\end{aligned}$$

□

Le corollaire suivant est un cas (très) particulier de l'inégalité de Jensen, qui peut se montrer élémentairement, sans théorie de la mesure.

Corollaire 5.6. Soit I un intervalle de $\mathbb{R}$, $\alpha_1, \ldots, \alpha_n$ des réels positifs tels que $\sum_{i=1}^{n} \alpha_i = 1$, et φ une fonction convexe sur I. Alors, pour tout $x_1, \ldots, x_n \in I$ on a

$$\varphi\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i \varphi(x_i) .$$

Démonstration. On fixe $x_1, \dots, x_n \in I$ et on considère l'espace mesuré d'ensemble sous-jacent $X = \{x_1, \dots, x_n\}$, où toutes les parties sont mesurables et $\mu = \sum_{i=1}^{n} \alpha_i \delta_{x_i}$, où δ_{x_i} désigne la mesure de Dirac en x_i. Alors μ est une mesure de probabilité ; de plus pour toute fonction $g\colon X \to \mathbb{R}$ on a

$$\int_X g d\mu = \sum_{i=1}^{n} \alpha_i g(x_i) \ .$$

En considérant pour g la fonction identité, on a donc $\int_X \varphi \circ g d\mu = \sum_{i=1}^{n} \alpha_i \varphi(x_i)$, et $\int_X g d\mu = \sum_{i=1}^{n} \alpha_i x_i$. L'inégalité de Jensen nous donne donc comme attendu

$$\varphi\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i \varphi(x_i) \ .$$

□

Remarque 5.1. Dans le corollaire ci-dessus, le cas $n = 2$ correspond exactement à la définition de la convexité. En particulier, une application φ qui satisfait l'inégalité de Jensen pour toute fonction intégrable sur un espace de probabilité, est nécessairement convexe.

3 Théorème de changement de variables

En pratique, pour calculer une intégrale multiple, on est souvent amené à faire un changement de variables pour se ramener à un domaine plus simple sur lequel appliquer le théorème de Fubini. On énonce le théorème dans le cadre le plus courant où les fonctions que l'on peut utiliser pour faire un changement de variables sont les **difféomorphismes de classe** $\mathcal{C}^1$.

Cas affine

On commence par montrer le cas des fonctions affines. Nous allons baser la preuve sur une caractérisation de la mesure de Lebesgue :

Théorème 5.7. (admis) La mesure de Lebesgue sur R^n est invariante par translation, au sens où pour tout $A \in \mathcal{B}(R^n)$ et tout $x \in R^n$, on a $\lambda_n(x+A) = \lambda_n(A)$ avec $x + A := \{x + a, a \in A\}$.
Inversement, si μ est une mesure sur $(R^n, \mathcal{B}(\mathbb{R}^n))$ finie sur les parties bornées et invariante par translation, alors il existe une constante $c \geq 0$ telle que $\mu = c\lambda_n$.

Exercice 5.3. On cherche à montrer l'unicité. On pose $c = \mu([0,1[^n)$. Montrer en utilisant des recouvrements par des translations d'un ensemble fixé que

1. $\mu([0, \frac{1}{m}[^n) = c\frac{1}{m^n}$
2. pour $a_1, \ldots, a_n \geq 0$, on a

$$\mu(\prod_{i=1}^n [0, \frac{\lfloor ma_i \rfloor}{m}[) = c\frac{\prod_{i=1}^n \lfloor ma_i \rfloor}{m^n}$$

En déduire que $\mu(\prod_{i=1}^n [a_i, b_i[) = c\prod_{i=1}^n (b_i - a_i)$ et conclure (en utilisant un corollaire du lemme de classe monotone).

Lemme 5.8. Soit $b \in \mathbb{R}^n$ et $A \in M_n(\mathbb{R})$ une matrice inversible. On pose $f(x) = Ax + b$ avec $f : \mathbb{R}^n \to \mathbb{R}^n$, alors pour tout borélien B de $\mathbb{R}^n$, on a :

$$\lambda_n(f(B)) = |det(A)|\lambda_n(B).$$

Exercice 5.4. Si A n'est pas inversible montrer que $\lambda(f(B)) = 0$. (Indication : on pourra montrer que $f(B)$ est inclus dans un hyperplan affine, i.e. un sous-espace affine de dimension $n-1$, dans le cas $b = 0$ dans un s.e.v. de dimension $n-1$).

Démonstration. $f(B) = (f^{-1})^{-1}(B)$ est bien borélien car f^{-1} est linéaire (en dimension finie donc) continue donc borélienne. De même $\lambda(f(\cdot)) = f^{-1}.\lambda$ est la mesure image par f^{-1} donc c'est bien une mesure finie sur les parties bornées (car $f(B)$ est borné pour tout borné B, cf chapitre 3 $f(B(0,M)) \subset B(0, ||b|| + M|||f|||)$ avec $|||f|||$ la norme subordonnée de f). Montrons qu'elle est invariante par translation.

On a pour $a \in \mathbb{R}^n$ $\lambda_n(f(a+B)) = \lambda_n(b + A(a+B)) = \lambda_n(Aa + f(B)) = \lambda_n(f(B))$ par invariance par translation de la mesure de Lebesgue. Le théorème précédent montre donc que $\lambda_n(f(B)) = c\lambda_n(B)$ pour tout borélien B. Il suffit donc de bien choisir le borélien pour chaque A pour montrer que $c = |det(A)|$.

Par décomposition polaire, une matrice réelle s'écrit $A = OS$ avec O orthogonale et S symétrique. Cette matrice S peut se diagonaliser en base orthogonale $S = O_2^t D O_2$ donc, ensemble, cela donne une décomposition $A = O_1 D O_2$ où $O_1 = OO_2^t$, O_2 sont orthogonales et D est diagonale réelle.

Comme λ_n est invariante par translation, on est donc ramené au cas $b = 0$.

On est donc ramener au deux cas A orthogonale et A diagonale inversible.

Si A orthogonale, alors on choisit la boule unité euclidienne $B = B_n$ car une matrice orthogonale laisse invariante cette boule (c'est par définition une isométrie pour la norme euclidienne) donc $\lambda_n(f(B_n)) = \lambda_n(B_n)$ et $c = 1 = |det(A)|$ (vu $AA^t = I$, $det(A)^2 = det(A)det(A^t) = det(I) = 1$).

Si $A = diag(d_1, \ldots, d_n)$ alors on prend $B = [0, 1]^n$ car $A(B) = \prod_{i=1}^n [0, d_i]$ avec $[0, d_i] = [d_i, 0]$ si $d_i < 0$. Dans tous les cas $\lambda_n(A(B)) = \prod_{i=1}^n |d_i| = |det(A)|\lambda(B)$ comme voulu.

Dans le cas général, $A = O_1 S O_2$, par composition, on obtient :

$$\lambda(A(B)) = |det(O_1)||det(D)| \det(O_2)|\lambda(B) = |\det(A)|\lambda(B).$$

□

Rappel (de L2) sur les difféomorphismes

Définition 5.2. Soient $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^p$. Une application $f : U \to V$ une fonction différentiable. f est un **difféomorphisme** si f est bijective et que f^{-1} est différentiable. On dit que f est un C^k**–difféomorphisme** ($k \in \mathbb{N}^* \cup \infty$) si de plus f et f^{-1} sont de classe C^k.

Proposition 5.9. Soit $f : U \to V$ un difféomorphisme, alors $\forall x \in U$, $df(x) : \mathbb{R}^n \to \mathbb{R}^p$ est un isomorphisme linéaire (en particulier nécessairement $n = p$) et on a :

$$(df(x))^{-1} = df^{-1}(f(x)).$$

Remarque 5.2. 1. Le résultat précédent montre que la dimension est invariante par difféomorphisme. De même des ouverts de $\mathbb{R}^n$ et $\mathbb{R}^p$ ne peuvent être homéomorphes que si $n = p$ mais c'est beaucoup plus dur (Théorème d'invariance du domaine de Brouwer). Par contre, il existe des applications continues surjectives de $[0, 1]$ dans $[0, 1]^2$.

2. Le théorème d'inversion locale va donner des conditions pour la réciproque de la proposition précédente

Démonstration. Comme $f^{-1} \circ f(y) = y$, en différenciant $f^{-1} \circ f$ par le théorème des fonctions composées en x, on obtient : $df^{-1}(f(x)) \circ df(x) = id$.

De même en différenciant $f \circ f^{-1}(y) = y$ en $z = f(x)$ on obtient : $df(f^{-1}(z)) \circ df^{-1}(z) = Id$. Donc $df(x)$ et $df^{-1}(f(x))$ sont inverses l'une de l'autre, ce qui conclut. □

Définition 5.3. Soit $f : U \to \mathbb{R}^p$ une application différentiable sur un ouvert $U \subset R^n$. $f(x) = (f_1(x), \dots, f_p(x))$. La matrice de l'application linéaire $df(x)$ dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^p$ est appelée, **matrice jacobienne** de f et notée $J(f)(x)$:

$$(J(f)(x))_{ij} = (\frac{\partial f_i}{\partial x_j}(x)).$$

Remarque 5.3. Le théorème de dérivation des fonctions composées donne donc :

$$J(g \circ f)(x_0) = J(g)(f(x_0))J(f)(x_0),$$

et le résultat pour les inverses de la proposition précédente s'écrit :

$$J(f^{-1})(y_0) = [J(f)(f^{-1}(y_0))]^{-1}.$$

Le théorème suivant avec $k = 1$ permettra de vérifier l'hypothèse du théorème de changement de variable.

Théorème 5.10 (d'inversion globale)**.** Soit $f : U \to \mathbb{R}^n$ une application de classe C^k (avec $k \geq 1$) injective et telle que pour tout $x \in U$, $df(x) : \mathbb{R}^n \to \mathbb{R}^n$ est un isomorphisme linéaire, alors $f(U)$ est un ouvert de $\mathbb{R}^n$ et $f : U \to f(U)$ est un C^k−difféomorphisme.

Remarque 5.4. $df(x)$ est un isomorphisme si et seulement si $det(Jf(x)) \neq 0$.

Cas général (admis)

Nous pouvons maintenant énoncer le théorème de changement de variables. [3]

★ **Théorème 5.11** (Théorème de changement de variables)**.** Soient U, V deux ouverts de $\mathbb{R}^n$, et $\varphi\colon U \to V$ un difféomorphisme de classe $\mathcal{C}^1$. Rappelons qu'on note λ_n la mesure de Lebesgue sur $\mathbb{R}^n$. Alors on a :

1. Pour toute partie B borélienne de U, $\lambda_n(\varphi(B)) = \int_B |\det(J\varphi(x))| d\lambda_n(x)$.
2. Si $f\colon V \to [0, +\infty]$ est borélienne, alors
$$\int_V f(x)d\lambda_n(x) = \int_U f\circ\varphi(y)|\det(J\varphi(y))|d\lambda_n(y)\ .$$
3. Si $f\colon V \to \mathbb{R}$ est intégrable, alors $y \mapsto f\circ\varphi(y)|\det(J\varphi(y))|$ est intégrable sur U et on a
$$\int_V f(x)d\lambda_n(x) = \int_U f\circ\varphi(y)|\det(J\varphi(y))|d\lambda_n(y)\ .$$

Remarque 5.5*.* Le cas affine est une conséquence du lemme 5.8 et du théorème de transfert appliqué $f = \varphi^{-1} : (V, \mathcal{B}(V), \lambda_n) \to (U, \mathcal{B}(U))$. Le 1 du théorème ou le lemme 5.8 ci-dessus, s'interprète comme le calcul de la mesure image de la mesure de Lebesgue induite sur V : $(\lambda_{n,V})_X$ ayant une densité $f_X(x) = |\det(J\varphi(x))|1_U(x)$ par rapport à λ_n. Le résultat correspond à $h = f\circ\varphi$ de sorte que :

$$\begin{aligned}\int_V fd\lambda_n &= \int_V h(X)d\lambda_n\\ &= \int_{\mathbb{R}^n} h(y)f_X(y)d\lambda_n(y)\\ &= \int_U f\circ\varphi(y)|\det(J\varphi(y))|d\lambda_n(y).\end{aligned}$$

Exemple 5.2 ((changement de variables en coordonnées polaires))*.* On considère l'application $\phi : U =]0, +\infty[\times]0, 2\pi[\to \mathbb{R}^2$ définie par $\phi(r, \theta) = (r\cos\theta, r\sin\theta)$.
Alors, la matrice jacobienne de ϕ est $\begin{pmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{pmatrix}$, de déterminant r.
De plus, ϕ est injective et $\phi(U) = \mathbb{R}^2 \setminus ([0, +\infty[\times\{0\}) = V$.
Ainsi, ϕ est un $\mathcal{C}^1$-difféomorphisme de U sur V. Comme $\lambda_2(\mathbb{R}^2 \setminus V) = 0$, c'est-à-dire $\mathbb{R}^2 \setminus V$ est négligeable, il n'est pas gênant que ϕ ne soit pas un difféomorphisme de U sur $\mathbb{R}^2$ tout entier.

3. Cette sous-section reprend le cours de 2018-2019 de T. Blossier, M. Carrizosa et J. Melleray.

Par exemple, calculons

$$I = \int_D (x+y)^2\,dxdy, \text{ où } D = \{(x,y) : x^2+y^2 < 1\}.$$

En utilisant le théorème de changement de variables avec les coordonnées polaires (et le théorème de Fubini), on obtient $\phi^{-1}(D \cap V) =]0,1[\times]0,2\pi[$ et

$$\begin{aligned}
I &= \int_{D\cap V} (x+y)^2\,dxdy \\
&= \int_{\phi^{-1}(D\cap V)} (r\cos\theta + r\sin\theta)^2\,rdrd\theta \\
&= \int_0^1 \left(\int_0^{2\pi} r^3(\cos^2\theta + \sin^2\theta + 2\cos\theta\sin\theta)d\theta\right) dr \\
&= \int_0^1 r^3 \left(\int_0^{2\pi} (1+\sin 2\theta)d\theta\right) dr \\
&= \int_0^1 2\pi r^3\,dr \\
&= \frac{\pi}{2}.
\end{aligned}$$

Exemple 5.3. Calculons $\Gamma(\frac{1}{2}) = \int_0^{+\infty} t^{-1/2} e^{-t}\,dt$.
On commence par le changement de variable (pour les intégrales à une variable) $u^2 = t$, $dt = 2udu$:

$$\Gamma(\frac{1}{2}) = \int_0^{+\infty} t^{-1/2} e^{-t}\,dt = 2\int_0^{+\infty} e^{-u^2}\,du = \int_{-\infty}^{+\infty} e^{-u^2}\,du$$

avec la dernière égalité venant de la parité de la fonction $u \mapsto e^{-u^2}$.
Enfin, on calcule le carré de cette intégrale en utilisant d'abord Fubini–Tonelli pour obtenir une intégrale double (on utilise $\mathbb{R}^2 \setminus (\{0\}\times[0,+\infty[) = V$ vérifiant $\lambda_2(V^c) = 0$ comme à l'exemple précédent).

$$\begin{aligned}
(\Gamma(\frac{1}{2}))^2 &= \left(\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\ e^{-x^2-y^2}\right) \\
&= \int_{\mathbb{R}^2} dxdy\ e^{-x^2-y^2} = \int_V dxdy\ e^{-x^2-y^2}
\end{aligned}$$

d'où par changement de variable en coordonnée polaire (comme à l'exemple précédent on utilise $\phi^{-1}(V) = U$ pour le domaine d'intégration) :

$$\begin{aligned}(\Gamma(\frac{1}{2}))^2 &= \left(\int_0^{2\pi} d\theta \int_0^{+\infty} dre^{-r^2} 2r/2\right)\\ &= \left(\int_0^{2\pi} d\theta 1\right)\left[-e^{-r^2}/2\right]_0^{+\infty}\\ &= (2\pi).\frac{1}{2} = \pi.\end{aligned}$$

On a aussi vérifier que

$$\int_{-\infty}^{+\infty} e^{-u^2} du = \sqrt{\pi}.$$

En faisant, le changement de variable linéaire $u = x/\sqrt{2}$, on obtient :

$$\frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{\pi}. \tag{5.1}$$

CHAPITRE 6
Introduction aux espaces L^p

Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré ($\mathcal{T}$ la tribu, μ la mesure). On va travailler en identifiant les fonctions si elles coïncident μ−presque partout. Autrement dit, on écrira $f = g$ quand $\mu(\{x : f(x) \neq g(x)\}) = 0$; en particulier, $f = 0$ signifiera que f vaut 0 presque partout. Par exemple, si f est la fonction caractéristique de $\mathbb{Q}$, on pourra écrire $f = 0$. Ainsi, dit en mots, on va en fait travailler avec les "classes d'équivalence de fonctions à égalité μ−presque partout près". $\mathbb{K}$ sera égale à $\mathbb{R}$ ou $\mathbb{C}$.

1 L'espace $L^\infty(\Omega, \mu)$

★ **Définition 6.1.** Soit $f : \Omega \to \mathbb{K}$ une fonction mesurable. On dit que $M \in [0, +\infty[$ est une **borne essentielle** de f ou que f est **essentiellement bornée** par M si $\mu(\{x : |f(x)| > M\}) = 0$, autrement dit, si $f \leq M$ μ−presque partout.

On définit leur ensemble :

$$\mathrm{L}^\infty(\Omega, \mathcal{T}, \mu; \mathbb{K}) = \\ \{\dot{f}; f : \Omega \to \mathbb{K}, \text{ mesurable et } \exists C < \infty : |f| \leq C\mu - p.p.\}$$

et la fonction (qui est une norme selon le lemme suivant) :

$$||f||_\infty = \inf\{C : |f| \leq C\mu - p.p.\} =: \text{ess supp}_{x\in\Omega}|f(x)|.$$

On note aussi plus brièvement $\mathrm{L}^\infty(\Omega; \mathbb{K}) = \mathrm{L}^\infty(\Omega, \mu; \mathbb{K}) = \mathrm{L}^\infty(\Omega, \mathcal{T}, \mu; \mathbb{K})$ et $\mathrm{L}^\infty(\Omega) = \mathrm{L}^\infty(\Omega; \mathbb{R})$, si il n'y a pas de confusion possible.

Exercice 6.1. (cf TD) Montrer que $|f| \leq ||f||_\infty p.p.$

Lemme 6.1. $(\mathrm{L}^\infty(\Omega, \mathcal{T}, \mu; \mathbb{K}), || \cdot ||_\infty)$ est un espace vectoriel normé.

Démonstration. On montre qu'il s'agit d'un sous-espace vectoriel de l'espace des classes d'équivalences de fonctions mesurables. Bien sûr 0 est bornée donc essentiellement bornée.

Soient $f, g \in \mathrm{L}^\infty(\Omega, \mathcal{T}, \mu; \mathbb{K})$, $\lambda \in \mathbb{K}$. Par l'exo

$$\mu(\{\omega : |f(\omega)| > ||f||_\infty\}) = 0, \qquad \mu(\{\omega : |g(\omega)| > ||g||_\infty\}) = 0.$$

Or par l'inégalité triangulaire des nombres on a :$|(\lambda f + g)(\omega)| \leq |\lambda||f(\omega)| + |g(\omega)|$ donc

$$\begin{aligned}&\{\omega :|f(\omega)| \leq ||f||_\infty\} \cap \{\omega : |g(\omega)| \leq ||g||_\infty\}\\ &\subset \{\omega : |(\lambda f + g)(\omega)| \leq |\lambda|||f||_\infty + ||g||_\infty\}\end{aligned}$$

et en passant au complémentaire

$$\begin{aligned}&\mu(\{\omega : |(\lambda f + g)(\omega)| > |\lambda|||f||_\infty + ||g||_\infty\})\\ &\leq \mu(\{\omega : |f(\omega)| > ||f||_\infty\}) + \mu(\{\omega : |g(\omega)| > ||g||_\infty\}) = 0\end{aligned}$$

Donc, par définition, $\lambda f + g$ est essentiellement bornée et $||\lambda f + g||_\infty \leq |\lambda|||f||_\infty + ||g||_\infty$. On déduit que $\mathrm{L}^\infty(\Omega; \mathbb{K})$ est bien un espace vectoriel et l'inégalité triangulaire. En fait $\mu(\{\omega : |f(\omega)| > C\}) = \mu(\{\omega : |\lambda f(\omega)| > |\lambda|C\})$ donc en comparant les infima, $||\lambda f||_\infty = |\lambda|\ ||f||_\infty$ ce qui donne la positive homogénéité. Enfin par définition, si $||f||_\infty = 0$ alors $f = 0$ presque partout donc sa classe d'équivalence est nulle. □

Théorème 6.2. $(\mathrm{L}^\infty(\Omega, \mathcal{T}, \mu; \mathbb{K}), ||\cdot||_\infty)$ est un espace de Banach.

Démonstration. Il reste à montrer la complétude : Soit f_n une suite de Cauchy de fonctions mesurables essentiellement bornées. Montrons que que f_n converge vers $f(\omega) = \limsup_{n\to\infty} f_n(\omega)$ qui est une fonction mesurable comme lim sup de fonctions mesurables et dont on va voir qu'elle est essentiellement bornée. Donc, par l'hypothèse d'avoir une suite de Cauchy, pour $n > 0$, $\epsilon = 1/n$ il existe N_n tel que $\forall p, q \geq N_n$, $||f_p - f_q||_\infty \leq \frac{1}{n}$. Par définition de la norme, on peut donc fixer $A_{n,p,q}$ (pour $p, q \geq N_n$) avec $\mu(A^c_{n,p,q}) = 0$ tel que

$$\sup_{\omega\in A_{n,p,q}} |f_p(\omega) - f_q(\omega)| \leq \frac{1}{n}.$$

On va intersecter tous ces ensembles (une intersection dénombrable) pour avoir μ-p.p. une suite de Cauchy. On prend donc $A = \cap_{n>0} \cap_{p,q\geq N_n} A_{n,p,q}$. On a $\mu(A^c) \leq \sum_{n>0} \sum_{p,q\geq N_n} \mu(A^c_{n,p,q}) = 0$ (vu que A^c est une union dénombrable).

De plus pour $\omega \in A^c$, on a

$$\forall n, \forall p, q \geq N_n, |f_p(\omega) - f_q(\omega)| \leq \frac{1}{n}$$

donc $(f_n(\omega))$ est de Cauchy dans $\mathbb{K}$ donc converge. Sa limite est forcément $f(\omega)$ et en passant à la limite $q \to \infty$ ci dessus, pour tout $\omega \in A$:

$$\forall n, \forall p \geq N_n, |f_p(\omega) - f(\omega)| \leq \frac{1}{n}.$$

Comme $\mu(A^c) = 0$ on déduit

$$\forall n, \forall p \geq N_n, ||f_p - f||_\infty \leq \frac{1}{n}.$$

Ceci implique $||f||_\infty \leq ||f_p||_\infty + ||f_p - f||_\infty$ donc f est dans $\mathrm{L}^\infty(\Omega, \mathcal{T}, \mu; \mathbb{K})$ et la convergence de f_n vers f dans cet espace. Comme toute suite de Cauchy converge, on a obtenu la complétude voulue. □

2 Définitions et propriétés élémentaires des espaces $L^p(\Omega, \mu)$

On définit les espaces :

$$\mathcal{L}^p(\Omega, \mathcal{T}, \mu; \mathbb{K}) = \{f : \Omega \to \mathbb{K} \text{ mesurable} \mid \int |f|^p d\mu < \infty\},$$

pour $p \in [1, \infty[$. Alors

$$||f||_p = (\int d\mu |f|^p)^{1/p}.$$

n'est pas une norme (mais une seminorme sur $\mathcal{L}^p(\Omega, \mathcal{T}, \mu)$ car si $||f||_p = 0$ alors f est seulement nulle presque partout. On considère donc l'espace des classes d'équivalences à égalité presque partout près de fonctions $\dot{f}$ et l'espace de Lebesgue :

★ Définition 6.2.

$$\mathrm{L}^p(\Omega, \mathcal{T}, \mu; \mathbb{K}) = \{\dot{f}; f : \Omega \to \mathbb{K} \text{ mesurable et} \int |f|^p d\mu < \infty\},$$

pour $p \in [1, \infty[$.

Comme pour le cas $p = \infty$, on on note aussi plus brièvement

$$\mathrm{L}^p(\Omega; \mathbb{K}) = \mathrm{L}^p(\Omega, \mu; \mathbb{K}) = \mathrm{L}^p(\Omega, \mathcal{T}, \mu; \mathbb{K})$$

et $\mathrm{L}^p(\Omega) = \mathrm{L}^p(\Omega; \mathbb{R})$, si il n'y a pas de confusion possible.

Par la suite, on identifie f à $\dot{f}$ dans ce contexte, on répète que les égalités sont des égalités $\mu - p.p.$.

Montrons que $||.||_p$ est une norme sur $L^p(\Omega, \mathcal{T}, \mu)$. La séparation et l'homogénéité sont maintenant évidentes. On rappelle l'inégalité de Hölder d'abord dans le cas le plus simple

Proposition 6.3. Si f, g sont mesurables, $\|f\|_p < +\infty$ et $\|g\|_\infty < +\infty$, alors $fg \in L^p(\Omega, \mathcal{T}, \mu; \mathbb{K})$ et $\|fg\|_p \leq \|f\|_p \|g\|_\infty$.

Démonstration. Il suffit de noter que, μ–presque partout, on a $|g(x)| \leq \|g\|_\infty$, et donc $|f(x)g(x)|^p \leq |f(x)|^p \|g\|_\infty^p$. En intégrant cette inégalité, on obtient bien

$$\|fg\|_p^p = \int_\Omega |f(x)g(x)|^p d\mu \leq \int_\Omega |f(x)|^p \|g\|_\infty^p d\mu = \|f\|_p^p \|g\|_\infty^p \ .$$

□

La version générale est la suivante

★ **Lemme 6.4** (inégalité de Hölder)**.** Si $p, q \in [1, \infty[$ tels que $1/p + 1/q = 1/r \leq 1$, $f \in L^p(\Omega, \mathcal{T}, \mu; \mathbb{K})$, $g \in L^q(\Omega, \mathcal{T}, \mu; \mathbb{K})$ alors $fg \in L^r(\Omega, \mathcal{T}, \mu; \mathbb{K})$ et

$$||fg||_r \leq ||f||_p ||g||_q.$$

Démonstration. En remplaçant f, g par $|f|^r, |g|^r$ on se ramène au cas $r = 1$.

Par hypothèse dans le cas $r = 1$, $1 < p < \infty$, on remarque que par concavité du logarithme, on a pour a, $b > 0$

$$\begin{aligned} \log\left(a^p/p + b^q/q\right) &\geq \log\left(a^p\right)/p + \log\left(b^q\right)/q \\ &= \log\left(ab\right) . \end{aligned}$$

Donc on obtient en exponentiant (et en vérifiant directement les cas d'annulations), l'inégalité d'Young :

$$|f(x)g(x)| \leq \frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q}.$$

Donc en intégrant, on obtient $fg \in L^1$ et appliquant à λf, $\lambda > 0$:

$$||fg||_1 \leq \frac{\lambda^{p-1}}{p} ||f||_p^p + \frac{\lambda^{-1}}{q} ||g||_q^q.$$

Comme le cas d'annulation $||f||_p = 0$ ou $||g||_q = 0$ sont évidents (car alors $fg = 0$ $\mu - p.p.$), on conclut en supposant $||f||_p \neq 0, ||g||_q \neq 0$ et en prenant la valeur de λ donnant le minimum $\lambda = ||f||_p^{-1} ||g||_q^{q/p}$. □

Une conséquence importante est l'exercice suivant :

Exercice 6.2. Si μ est une mesure finie pour $1 \leq p \leq q \leq \infty$, montrer que :

$$\begin{aligned} L^\infty(\Omega, \mathcal{T}, \mu; \mathbb{K}) &\subset L^q(\Omega, \mathcal{T}, \mu; \mathbb{K}) \\ &\subset L^p(\Omega, \mathcal{T}, \mu; \mathbb{K}) \subset L^1(\Omega, \mathcal{T}, \mu; \mathbb{K}). \end{aligned}$$

On en déduit l'inégalité triangulaire :

★ **Théorème 6.5** (Inégalité de Minkowski)**.** Soient $p \in [1, +\infty]$ et $f, g \in L^p(\Omega)$. Alors $f + g \in L^p(\Omega)$ et $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

Démonstration. On a déjà traité le cas $p = +\infty$, et le cas $p = 1$ est simplement l'inégalité triangulaire habituelle. Supposons donc $p \in]1, +\infty[$ et $f, g \in L^p(\Omega)$.

Commençons par montrer que $\|f + g\|_p < +\infty$. Comme $x \mapsto x^p$ est convexe et croissante, on a pour tout x que

$$\left(\left|\frac{1}{2}f(x) + \frac{1}{2}g(x)\right|\right)^p \leq \left(\left|\frac{1}{2}f(x)\right| + \left|\frac{1}{2}g(x)\right|\right)^p \leq \frac{1}{2}|f(x)|^p + \frac{1}{2}|g(x)|^p \ .$$

En intégrant cette inégalité, on obtient que

$$\frac{1}{2^p}\|f + g\|_p^p \leq \frac{1}{2}(\|f\|_p^p + \|g\|_p^p) \ .$$

Ceci nous prouve que $\|f + g\|_p < +\infty$.

Maintenant, notons $q = \dfrac{p}{p-1}$ l'exposant conjugué de p. Ci-dessous, on va utiliser l'inégalité de Hölder, et le fait que

$$\begin{aligned} \left\||f + g|^{p-1}\right\|_q &= \left(\int_\Omega |f + g|^{(p-1)q} d\mu\right)^{\frac{1}{q}} \\ &= \left(\int_\Omega |f + g|^p\right)^{1-\frac{1}{p}} = \|f + g\|_p^{p-1} \ . \end{aligned}$$

Alors on a

$$\begin{aligned} \|f + g\|_p^p &= \int_\Omega |f + g|^p d\mu \\ &\leq \int_\Omega (|f| + |g|)|f + g|^{p-1} d\mu \\ &= \int_\Omega |f||f + g|^{p-1} d\mu + \int_\Omega |g||f + g|^{p-1} d\mu \\ &\leq \|f\|_p \left\||f + g|^{p-1}\right\|_q + \|g\|_p \left\||f + g|^{p-1}\right\|_q \\ &= (\|f\|_p + \|g\|_p) \left\||f + g|^{p-1}\right\|_q \\ &= (\|f\|_p + \|g\|_p)\|f + g\|_p^{p-1} \end{aligned}$$

Si jamais $\|f+g\|_p = 0$ on n'a rien à démontrer ; sinon, en divisant des deux côtés par $\|f+g\|_p^{p-1}$ on obtient finalement $\|f+g\|_p \leq \|f\|_p + \|g\|_p$. □

Exercice 6.3. Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesure σ−fini. Soit $f \geq 0$ une fonction mesurable positive, alors pour $p \in]0, \infty[$

$$\int f^p d\mu = \int_0^\infty dtpt^{p-1}\mu(\{\omega : f(\omega) > t\}).$$

On rappelle d'abord la version L^p du théorème de convergence dominée.

★ **Théorème 6.6** (Théorème de convergence dominée L^p). Soit $p \in [1, +\infty[$. Soit (Ω, μ) un espace mesuré, et f_n une suite de fonctions mesurables convergeant μ−presque partout vers f, et vérifiant la domination $|f_n| \leq g$ avec $g \in L^p(\Omega, \mu)$. Alors, $f_n, f \in L^p(\Omega, \mu)$ et f_n converge vers f dans $L^p(\Omega, \mu)$, c'est à dire.

$$\lim_{n\to\infty} ||f_n - f||_p = 0.$$

Démonstration. On a $|f_n - f|^p \to 0$ μ−presque partout. De $|f_n| \leq g$ on déduit que $f_n, inL^p(\Omega, \mu; \mathbb{K})$ en passant à la limite on obtient $|f| \leq g$ et donc $f \in L^p(\Omega, \mu; \mathbb{K})$. De plus, on a la domination :

$$|f_n - f|^p \leq (|f_n| + |f|)^p \leq (2g)^p = 2^p g^p$$

et comme $g \in L^p(\Omega, \mu)$ et positive, on déduit que $g^p = |g|^p$ est μ−intégrable et sert donc de domination pour appliquer le théorème de convergence dominée usuelle qui donne le résultat :

$$||f_n - f||_p^p = \int_\Omega |f_n - f|^p d\mu \to_{n\to\infty} \int_\Omega 0\, d\mu = 0.$$

□

★ **Théorème 6.7** (de Riesz−Fischer). Soit (Ω, μ) un espace mesuré, les espaces $L^p(\Omega, \mu, \mathbb{K})$ pour $p \in [1, \infty]$ sont des espaces de Banach.

Démonstration. On vient de voir que $L^p(\Omega, \mu, \mathbb{K})$ est un espace vectoriel normé, et même la complétude dans le cas $p = \infty$.

Il reste le cas $p < \infty$. En décomposant en partie réelle et imaginaire, on peut supposer et donc on suppose $\mathbb{K} = \mathbb{R}$.

Pour la complétude, on utilise la proposition 2.6. Soit $\sum u_n$ qui est absolument convergente, il faut montrer qu'elle converge dans L^p. Soit $g_k = \sum_{n=1}^{k} |u_n|$, $||g_k||_p \leq \sum ||u_n||_p$ et $|g_k|^p$ est croissante, donc par convergence monotone converge vers g avec $||g||_p \leq \sum ||u_n||_p$. Donc $|g|^p \in L^1$ qui donne une domination pour $|\sum u_n|^p$ et $\sum u_n$ est p.p. absolument convergente, donc a p.p. une limite et par convergence dominée, converge donc dans L^p. . □

Résultats de convergences

En suivant le même raisonnement on obtient le résultat suivant :

★ **Théorème 6.8.** Soient $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, $p \in [1, +\infty[$, et (f_n) une suite d'éléments de $L^p(\Omega)$ qui converge vers f dans $(L^p(\Omega), \|\cdot\|_p)$. Alors il existe une suite extraite (f_{n_k}) telle que (f_{n_k}) tend vers f, μ–presque partout et dans $L^p(\Omega)$.

Démonstration. On extrait (f_{n_k}) telle que $||f_{n_{k+1}} - f_{n_k}||_p \leq 1/2^k$. (c'est possible car la suite est de Cauchy dans L^p donc on prend n_k telle que $||f_q - f_{n_k}||_p \leq 1/2^k$ pour $q \geq n_k$.)

Donc on pose $g_n = \sum_{k=1}^{n} |f_{n_{k+1}} - f_{n_k}|$ qui est une suite croissante avec

$$||g_k||_p \leq \sum_k ||f_{n_{k+1}} - f_{n_k}||_p \leq \sum_{k=1}^{\infty} 1/2^k = 1.$$

On déduit donc en appliquant le théorème de convergence monotone que g_k a une limite $g = \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}|$ telle que $||g||_p \leq 1$. On l'utilise maintenant comme condition de domination. Donc $\sum_k (f_{n_{k+1}} - f_{n_k})$ est absolument convergente sur $A = \{\omega : g(\omega) < \infty\}$ et on a $\mu(A^c) = 0$, vu $||g||_p < \infty$. Donc par série télescopique $(f_{n_k}(\omega))$ converge pour $\omega \in A$. (et comme suite extraite elle converge aussi dans L^p mais en fait elle est dominée par $|f_{n_0}| + g \in L^p$ et converge aussi par convergence dominée). □

Proposition 6.9. Soient $(\Omega, \mathcal{T}, \mu)$ un espace de probabilité et $f : \Omega \to [0, +\infty]$ une fonction mesurable. Alors on a

$$\|f\|_\infty = \lim_{p \to +\infty} \|f\|_p \ .$$

Démonstration. Commençons par remarquer que l'on a toujours

$$\|f\|_p = \left(\int_\Omega |f|^p \, d\mu\right)^{\frac{1}{p}} \leq \left(\|f\|_\infty^p \mu(\Omega)\right)^{\frac{1}{p}} = \|f\|_\infty \ .$$

Par conséquent, si $\|f\|_p \to +\infty$ quand $p \to +\infty$ alors $\|f\|_\infty = +\infty$. Pour voir la réciproque, notons que pour $t < \|f\|_\infty$ fixé, l'ensemble $A_t = \{x \in \Omega : |f(x)| > t\}$ est de mesure strictement positive, par conséquent

$$\|f\|_p \geq (t^p \mu(A_t))^{\frac{1}{p}} = t\mu(A_t)^{\frac{1}{p}} \to t \text{ quand } p \to +\infty \ .$$

Ceci montre que si $\|f\|_\infty = +\infty$ alors $\|f\|_p$ tend vers $+\infty$; mais aussi que, si $\|f\|_\infty < +\infty$ on a pour tout $\varepsilon > 0$ que pour p suffisamment grand $\|f\|_\infty - \varepsilon \leq \|f\|_p \leq \|f\|_\infty$. □

Résultats de densité

On rappelle le résultat suivant qui se déduit de la construction de l'intégrale (cf. lemme 4.21)

Lemme 6.10. Soit $(\Omega, \mu, \mathcal{T})$ un espace σ−fini. L'ensemble S des fonctions étagées intégrables est dense dans tous les $L^p(\Omega, \mu, \mathcal{T})$, $1 \leq p < \infty$. En particulier, $L^1(\Omega, \mu, \mathcal{T}) \cap L^\infty(\Omega, \mu, \mathcal{T})$ est dense dans $L^p(\Omega, \mu, \mathcal{T})$ pour $1 \leq p < \infty$.

Lemme 6.11. Soit $(\Omega, \mu, \mathcal{T})$ un espace σ−fini avec $\mathcal{T} = \sigma(\mathcal{E})$ pour $\mathcal{E}$ une famille stable par intersection finie et de mesure finie pour μ, et contenant une suite A_n avec $\mu(A_n) < \infty$ et $\Omega = \cup_n A_n$. Alors l'espace vectoriel $E = Vect\{1_A, A \in \mathcal{E}\}$ est dense dans tous les $L^p(\Omega, \mu, \mathcal{T})$, $1 \leq p < \infty$. En particulier, si $\mathcal{E}$ est dénombrable, alors $L^p(\Omega, \mu, \mathcal{T})$, $1 \leq p < \infty$ est séparable.

En général $L^\infty(\Omega, \mu, \mathcal{T})$ n'est PAS séparable, sauf si Ω est un ensemble fini, par exemple $\ell^\infty(\mathbb{N})$ n'est pas séparable (c'est un exercice plus dur de niveau M1).

Démonstration. Soit $A_n \in \mathcal{E}$ avec $\mu(A_n) < \infty$ et $\Omega = \cup_n A_n$.

Soit $\mathcal{M} := \{A \in \mathcal{T} : \forall n, 1_{A \cap A_n} \in \overline{E}^{L^p}\}$. Clairement $\mathcal{E} \subset \mathcal{M}$. On va montrer que $\mathcal{M}$ est une classe monotone :

- $\Omega \in \mathcal{M}$ car $1_{A_n} \in E$
- Si $A \subset B$ et $A, B \in \mathcal{M}$, on a $1_{(B \setminus A) \cap A_n} = 1_{B \cap A_n} - 1_{A \cap A_n}$ par le TD 1 donc dans l'espace vectoriel $\overline{E}^{L^p}$.

- Si $B_m \in \mathcal{M}$ suite croissante d'union B alors $1_{B_m \cap A_n} \to 1_{B\cap A_n}$ partout par le TD 1, Or on a domination par $1_{A_n} \in L^p(\Omega, \mu, \mathcal{T})$ donc par convergence dominée $1_{B_m\cap A_n} \to 1_{B\cap A_n}$ dans $L^p(\Omega, \mu, \mathcal{T})$ et donc $1_{B\cap A_n} \in \overline{E}^{L^p}$

Le lemme de classe monotone implique $\mathcal{M} \supset \mathcal{T}(\mathcal{E})$. Donc si $B \in \mathcal{T}(\mathcal{E})$ est de mesure finie, on a $1_{B\cap A_n} \in \overline{E}^{L^p}$ et par la même application du théorème de convergence dominée (par 1_B cette fois) on déduit $1_B \in \overline{E}^{L^p}$. Donc $\overline{E}^{L^p}$ contient toute fonction étagée intégrable et le résultat précédent conclut. La séparabilité vient de la densité de l'ensemble dénombrable $Vect_{\mathbb{Q}}(1_A, A \in \mathcal{E})$. □

Le support d'une fonction continue f est le fermé $\mathrm{supp}(f) = \overline{f^{-1}(\{0\})^c}$. Un fonction sur $\mathbb{R}^n$ est donc à support compact quand elle est nulle en dehors d'un ensemble borné. On note $C^0_c(\Omega)$ est l'ensemble des fonctions à support compact sur un ouvert Ω.

★ **Théorème 6.12.** Soit $\Omega \subset \mathbb{R}^n$ un ouvert et λ la mesure de Lebesgue sur la tribu borélienne $\mathcal{B}(\Omega) = \mathcal{B}(\mathbb{R}^n)_\Omega$ (tribu induite sur Ω). Alors l'ensemble des fonctions continues à support compact $C^0_c(\Omega)$ est dense dans $L^p(\Omega, \mathcal{B}(\Omega), \lambda)$ pour $1 \leq p < \infty$, qui est séparable.

Démonstration. Par le lemme précédent avec $\mathcal{E} = \{A = \prod_{i=1}^n [a_i, b_i], a_i \leq b_i\}$ l'ensemble des pavés, il suffit de voir que les 1_A sont approchés par des fonctions continues à support compact pour $A = \prod_{i=1}^n [a_i, b_i]$. Par produit de fonctions (de variables différentes), cela se ramène au cas $n = 1$. Soit $f = 1_{[a,b]}$ et $f_n(t) = 1$ si $t \in [a, b]$, $f_n(t) = 1 - max(n(t - b), 1)$ si $t > b$, $f_n(t) = 1 - max(n(a - t), 1)$ si $t < a$. Alors il est facile de voir que $(f_n)_{n\geq 1}$ est une suite dans $C^0_c(\Omega)$ qui converge ponctuellement vers f (exo). Elle est dominée par $1_{[a-1,b+1]}$ qui est dans $L^p(\Omega, \mathcal{B}(\Omega), \lambda)$ pour $1 \leq p < \infty$ donc par convergence dominée, $||f_n - f||_p \to 0$. Donc on peut appliquer le lemme précédent et conclure. □

3 Cas discret : espaces $\ell^p(I)$, $p \in [1, \infty[$ (cf. TD)

Définition 6.3. Soit $p \in [1, \infty[$. Une famille $(z_i)_{i\in I}$ de nombres complexes ou réels est dite **p–sommable** si la famille $(|z_i|^p)_{i\in I}$ est sommable. On note $\ell^p(I, \mathbb{K})$ l'ensemble des familles d'éléments de $\mathbb{K}$ p–sommable.

Un examen de la définition indique que $\ell^p(I, \mathbb{K}) = L^p(I, \mathcal{P}(I), \nu)$ avec ν la mesure de comptage, c'est donc un espace de Banach. On a aussi par définition (dans le cas positif puis le cas quelconque) :

$$\sum_{i \in I} a_i = \int_I a d\nu.$$

On note

$$||z||_p = \left(\sum_{i \in I} |z_i|^p \right)^{1/p}.$$

L'inégalité de Hölder s'écrit donc pour $x \in \ell^q(I)$, $y \in \ell^p(I)$: avec $1/p + 1/q = 1$, $p, q \in]1, \infty[$:

$$\left| \sum_{i \in I} x_i y_i \right| \leq \left(\sum_{i \in I} |x_i|^q \right)^{1/q} \left(\sum_{i \in I} |y_i|^p \right)^{1/p}$$

CHAPITRE 7

Espaces de Hilbert ; bases hilbertiennes

1 Généralités

Soit H un espace vectoriel sur $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$

★ **Définition 7.1.** Un produit scalaire sur H est une application

$$\langle ., . \rangle : H \times H \to \mathbb{K}$$

telle que :

1. pour tout $y \in H$, $\langle y, . \rangle : H \to \mathbb{K}$ est linéaire
2. – Si $\mathbb{K} = \mathbb{R}$ $\forall x, y \in H$, $\langle x, y \rangle = \langle y, x \rangle$ (symétrie)
   – Si $\mathbb{K} = \mathbb{C}$ $\forall x, y \in H$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (symétrie hermitienne)
3. pour $x \in H$, $\langle x, x \rangle \in \mathbb{R}^+$
4. pour $x \in H$, $\langle x, x \rangle = 0$ si et seulement si $x = 0$.

Un espace H avec un tel produit scalaire est un espace préhilbertien réel (si $\mathbb{K} = \mathbb{R}$) et complexe (si $\mathbb{K} = \mathbb{C}$).

On remarque que dans le cas complexe, $\langle ., y \rangle$ est antilinéaire, c'est-à-dire avec $\overline{\lambda}$ le conjugué complexe,

$$\forall x, y, z \in H, \lambda \in \mathbb{C}, \quad \langle \lambda x + z, y \rangle = \overline{\lambda} \langle x, y \rangle + \langle z, y \rangle.$$

Exemple 7.1. Sur $H = \ell^2(\mathbb{N}, \mathbb{C}) := L^2(\mathbb{N}, \nu; \mathbb{C})$ (espace L^2 avec la mesure de comptage ν) on a le produit scalaire (hermitien canonique) :

$$\langle x, y \rangle = \sum_{i \in I} \overline{x_i} y_i$$

Dans le cas réel, la même formule sans conjugaison complexe fonctionne.

Exemple 7.2. Sur $H = L^2(\Omega, \mu; \mathbb{C})$ avec (Ω, μ) un espace mesuré σ−fini, on a le produit scalaire (hermitien canonique) :

$$\langle f, g\rangle = \int_\Omega \overline{f(x)}g(x)d\mu(x).$$

Exemple 7.3. Sur $H = C^0([a, b], \mathbb{C})$ on a le produit scalaire :

$$\langle f, g\rangle = \int_a^b \overline{f(x)}g(x)dx.$$

Proposition 7.1. Si H est muni d'un produit scalaire on a l'inégalité de Cauchy−Schwarz :

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle$$

avec égalité si et seulement si x, y sont liés. De plus $||x|| = \sqrt{\langle x, x\rangle}$ est une norme sur H vérifiant l'identité du parallélogramme :

$$\left\|\frac{x+y}{2}\right\|^2 + \left\|\frac{x-y}{2}\right\|^2 = \frac{1}{2}(||x||^2 + ||y||^2).$$

Démonstration. On a

$$\langle x + ty, x + ty\rangle = ||x||^2 + t^2||y||^2 + 2t\Re(\langle x, y\rangle) \geq 0$$

c'est un polynôme de degré 2 qui est toujours positif ou nul, donc son discriminant $\Delta = 4\Re(\langle x, y\rangle)^2 - 4||x||^2||y||^2 \leq 0$. En remplaçant y par uy avec $u = \frac{\overline{\langle x,y\rangle}}{|\langle x,y\rangle|}$ si $\langle x, y\rangle \neq 0$ on obtient

$$\Re(\langle x, y\rangle u) = |\langle x, y\rangle| \leq ||x||^2\langle uy, uy\rangle = ||x||^2||y||^2\overline{u}u = ||x||^2||y||^2.$$

Le même calcul donne pour u de module 1 la norme de

$$\Big\| \, ||y||x - u||x||y \, \Big\|^2 = 2||y||^2||x||^2 - 2||x||||y||\Re(\langle x, uy\rangle)$$

qui vaut 0 si on choisit u tel que $\langle x, y\rangle u = |\langle x, y\rangle|$ et que l'on est dans le cas d'égalité de C−S, ce qui donne la relation de dépendance linéaire cherchée $||y||x - u||x||y = 0$. (La réciproque, c'est à dire l'égalité en cas de dépendance linéaire, est évidente).

Pour vérifier que l'on a une norme, la positivité vient de l'axiome 3, la séparation vient du dernier axiome, l'homogénéité vient de

$$\langle \lambda y, \lambda y\rangle = \overline{\lambda}\lambda\langle y, y\rangle = |\lambda|^2\langle y, y\rangle$$

et l'inégalité triangulaire vient d'une application de C–S :

$$\langle x+y, x+y\rangle = ||x||^2 + ||y||^2 + 2\Re\langle x, y\rangle \leq ||x||^2 + ||y||^2 + 2||x|| ||y|| = (||x|| + ||y||)^2.$$

Enfin, on a aussi la relation :

$$\langle x-y, x-y\rangle = ||x||^2 + ||y||^2 - 2\Re\langle x, y\rangle$$

soit en faisant la somme (avec l'égalité débutant le calcul pour l'inégalité triangulaire), on obtient l'identité du parallélogramme. □

Remarque 7.1. L'identité du parallélogramme implique que $\left\|\frac{x+y}{2}\right\|^2 \geq \frac{1}{2}(||x||^2 + ||y||^2)$ avec égalité si et seulement si $x = y$ ce qui donne un résultat de convexité (en faite stricte car l'inégalité est stricte si $x \neq y$). (On a vu en TD que par continuité la convexité à mi point implique la convexité).
Une autre identité importante s'établit en prenant la différence des égalités donnant la preuve de l'identité du parallélogramme ci–dessus, c'est l'identité de polarisation :

$$\Re\langle x, y\rangle = \frac{||x+y||^2 - ||x-y||^2}{4}$$

On retrouve aussi

$$\Im\langle y, x\rangle = \Re\langle iy, x\rangle = \frac{||x+iy||^2 - ||x-iy||^2}{4}$$

d'où la formule de polarisation complexe :

$$\langle y, x\rangle = \frac{||x+y||^2 - ||x-y||^2 + i||x+iy||^2 - i||x-iy||^2}{4}$$

ou encore en bref

$$\langle y, x\rangle = \frac{1}{4}\sum_{i=0}^{3} i^k ||x + i^k y||^2 \tag{7.1}$$

★ **Définition 7.2.** Un espace pré–hilbertien complet est appelé **espace de Hilbert**.

★ **Théorème 7.2.** Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré. Alors $H = L^2(\Omega, \mathcal{T}, \mu; \mathbb{K})$ est un espace de Hilbert sur $\mathbb{K}$ avec le produit scalaire défini pour $f, g \in H$ par :

$$\langle f, g\rangle = \int_\Omega \overline{f} g d\mu.$$

Démonstration. On ne traite que le cas $\mathbb{K} = \mathbb{C}$. Si $f, g \in H$, l'inégalité de Hölder avec $p = q = 2$ donne $\overline{f} g \in L^1(\Omega, \mathcal{T}, \mu; \mathbb{K})$ et donc l'intégrale définissant le produit scalaire est bien définie. On vérifie les axiomes des produits scalaires : 1/ $\langle f, g\rangle$ est linéaire en la deuxième variable g par linéarité de l'intégrale.

2/ la symétrie hermitienne vient du calcul suivant :

$$\langle f, g\rangle = \int_\Omega \overline{f} g d\mu = \int_\Omega \overline{f\overline{g}} d\mu = \overline{\int_\Omega f\overline{g} d\mu} = \overline{\langle g, f\rangle}.$$

3/

$$\langle f, f\rangle = \int_\Omega |f|^2 \, d\mu = ||f||_2^2 \in [0, +\infty[$$

4/ Comme on sait déjà que $||.||_2$ la séparation de la norme implique que si $||f||_2 = 0$ alors $f = 0$ (μ–presque partout c'est à dire) dans $H = L^2(\Omega, \mathcal{T}, \mu; \mathbb{K})$.

On a donc bien un espace pré–hilbertien, et le Théorème de Riesz–Fischer 6.7 dit que $L^2(\Omega, \mathcal{T}, \mu; \mathbb{K})$ est complet, donc un espace de Hilbert. □

Exemple 7.4. $\ell^2(\mathbb{N}; \mathbb{C})$ sont des espaces de Hilbert (cf. chapitre 6 pour la complétude), mais pas $C^0([a, b], \mathbb{C})$ dont la complétion est l'espace de Hilbert $L^2([a, b], \lambda; \mathbb{C})$. La complétion d'un espace préhilbertien en tant qu'e.v.n. (cf. annexe A section 3) est toujours un espace de Hilbert.

2 Projection sur un convexe fermé

On va généraliser l'existence de projection orthogonale sur un sous–espace d'un espace euclidien d'abord au cas des convexes fermés et en dimension infinie.

★ **Théorème 7.3.** Soit H un espace de Hilbert et $C \subset H$ un convexe fermé non–vide. Pour tout $f \in H$ il existe un unique $u = P_C(f) \in C$ tel que

$$||f - u|| = \inf_{v \in C} ||f - v||.$$

De plus c'est l'unique vecteur $u \in C$ vérifiant la propriété caractéristique :

$$\forall v \in C, \quad \Re(\langle f - u, v - u\rangle) \leq 0$$

Enfin, P_C est une application 1–lipschitzienne appelée **projection sur** C.

Remarque 7.2. Un théorème de projection similaire sur un convexe fermé est valide dans $L^p(\Omega, \mathcal{T}, \mu)$ pour tout $1 < p < \infty$ (et pas seulement $p = 2$), mais il n'y a pas de caractérisation aussi simple de la projection P_C (en l'absence de produit scalaire) et la projection P_C est seulement uniformément continue (et plus nécessairement Lipschitz). Mais ce résultat est beaucoup plus dur (un exercice difficile de M1 Math).

Démonstration. On fait une preuve directe, utilisant l'identité du parallélogramme.

Soit $v_n \in C$ tel que $||f - v_n|| \to d = \inf_{v\in C} ||f - v||$

En appliquant l'identité à $a = f - v_n$, $b = f - v_m$, on trouve :

$$\left\|f - \frac{v_n + v_m}{2}\right\|^2 + \left\|\frac{v_n - v_m}{2}\right\|^2 = \frac{1}{2}(||f - v_n||^2 + ||f - v_m||^2) \to d^2.$$

Or par convexité $\frac{v_n+v_m}{2} \in C$ donc $\left\|f - \frac{v_n+v_m}{2}\right\|^2 \geq d^2$ donc

$$\left\|\frac{v_n - v_m}{2}\right\|^2 \leq \frac{1}{2}(||f - v_n||^2 + ||f - v_m||^2) - d^2 \to 0.$$

On déduit donc que v_n est de Cauchy, donc converge vers u et par continuité de la norme $d = ||f - u||$.

Soit $g : v \mapsto ||f - v||_2^2$. On peut calculer la différentielle $dg(u) = \Re(\langle f - u, .\rangle)$. Or si g atteint son minimum en u, pour $v \in C$, $t \in [0, 1]$,

$$||f - tv - (1 - t)u||_2^2 = ||f - u||_2^2 + t^2||v - u||_2^2 - 2t\Re(\langle f - u, v - u\rangle) \geq ||f - u||_2^2$$

donc $2\Re(\langle f - u, t - u\rangle) \leq t||v - u||_2^2$ et la limite $t \to 0$ donne l'inégalité caractéristique. Réciproquement, on a en $t = 1$, l'inégalité qui conclut :

$$||f - u||_2^2 - ||f - v||_2^2 = 2\Re(\langle f - u, v - u\rangle) - ||v - u||_2^2 \leq 0.$$

Pour voir l'unicité, si $u_1, u_2 \in C$, on peut utiliser la convexité stricte sous la forme de l'identité du parallélogramme, on a

$$\left\|f - \frac{u_1 + u_2}{2}\right\|^2 + \left\|\frac{u_1 - u_2}{2}\right\|^2 = \frac{1}{2}(||f - u_1||^2 + ||f - u_2||^2) = d^2$$

soit comme $\left\|f - \frac{u_1+u_2}{2}\right\|^2 \geq d^2$ on déduit $\left\|\frac{u_1-u_2}{2}\right\|^2 \leq 0$ donc $u_1 = u_2$.

Par l'unicité, P_C est bien définie et il ne reste qu'à voir la lipschitizianité. En appliquant la propriété caractéristique pour f_1, f_2 :

$$\Re(\langle f_1 - P_C(f_1), P_C(f_2) - P_C(f_1)\rangle) \leq 0,$$

$$\Re(\langle f_2 - P_C(f_2), P_C(f_1) - P_C(f_2)\rangle) \leq 0,$$

soit en additionnant :

$$\Re(\langle f_1 - f_2 + P_C(f_2) - P_C(f_1), P_C(f_2) - P_C(f_1)\rangle) \leq 0$$

soit en utilisant Cauchy-Schwarz :

$$\begin{aligned} ||P_C(f_2) - P_C(f_1)||^2 &\leq \Re(\langle f_1 - f_2, P_C(f_2) - P_C(f_1)\rangle) \\ &\leq ||f_1 - f_2||\, ||P_C(f_2) - P_C(f_1)||. \end{aligned}$$

□

★ **Théorème 7.4.** Soit H un espace de Hilbert et $K \subset H$ un sous espace vectoriel fermé. Pour tout $f \in H$, il existe un unique $u = P_K(f) \in K$ tel que

$$||f - u||_2 = \inf_{v \in K} ||f - g||_2.$$

De plus c'est l'unique vecteur $u \in K$ tel que

$$\forall v \in K, \quad \langle v, f - u\rangle = 0$$

Enfin, P_K est une application linéaire bornée appelée **projection orthogonale sur** K.

Démonstration. Il reste à voir la nouvelle caractérisation équivalente car celle-ci étant une relation linéaire, elle impose la linéarité de P_K ($\lambda P_K(f) + P_K(g)$ vérifie la relation pour $\lambda f + g$ et doit donc être par unicité $P_K(\lambda f + g)$). La nouvelle caractérisation est plus forte.
Réciproquement, si $\Re(\langle f - u, v - u\rangle) \leq 0$, en prenant $v = 2u$ et $v = 0$, on trouve $\Re(\langle f - u, u\rangle) = 0$ donc $\Re(\langle f - u, v\rangle) \leq 0$ pour tout v dans K donc aussi pour $-v$ par linéarité d'où l'égalité à 0. □

Exemple 7.5. Si $H = L^2(\Omega, \mu, \mathbb{R})$

$$C = \{f \geq 0 \ p.p.\}.$$

Alors $P_C(f) = f 1_{\{f \geq 0\}}$. (exo) Trouver aussi de même la projection sur l'ensemble de $f : \Omega \to [0, 1]$.

3 Applications : Orthogonalité et Dualité

Orthogonalité

On peut définir dans un espace de Hilbert une notion d'orthogonal comme en dimension finie.

★ **Définition 7.3.** Si $F \subset H$ est un sous-espace, alors l'orthogonal de F est

$$F^\perp = \{x \in H, \forall y \in F, \langle x, y\rangle = 0\}$$

On dit que x est orthogonal à F si $x \in F^\perp$. On remarque que

$$F^\perp = \bigcap_{y \in F} (\langle y, \cdot\rangle)^{-1}(\{0\})$$

est toujours un sous-espace fermé comme intersection de sous-espaces fermé, comme image inverse d'un sous-espace fermé par une application linéaire continue (le produit scalaire). La proposition suivante décrit la décomposition en somme directe orthogonale. Tout se passe comme en dimension finie pour les sous-espaces fermés, et sinon, il faut ajouter une adhérence.

★ **Proposition 7.5.** Si F est un sous-espace de l'espace de Hilbert H alors $F^{\perp\perp} = \overline{F}$, et on a la somme directe orthogonale

$$H = \overline{F} \oplus F^\perp$$

et alors p_F et $p_{F^\perp} = 1 - p_{\overline{F}}$ sont les projections associées à cette décomposition.

Ici $F^{\perp\perp} = (F^\perp)^\perp$ est l'orthogonal de l'orthogonal.

Démonstration. 1. On remarque d'abord que $F \subset F^{\perp\perp}$. En effet par définition de $F^\perp$ si $x \in F$, $y \in F^\perp$, $\langle x, y\rangle = 0$ et donc comme c'est pour tout $y \in F^\perp$ la définition du biorthogonal donne $x \in F^{\perp\perp}$.

2. On remarque ensuite que $F^{\perp\perp} \cap F^\perp = \{0\}$. En effet, si $x \in F^{\perp\perp} \cap F^\perp$ alors $\langle x, x\rangle = 0$ donc $x = 0$ (par l'axiome de séparation).
3. Montrons ensuite que $p_{F^\perp} = 1 - p_{\overline{F}}$ (les projections sont bien définies car on a des sous-espaces fermés l'espace de Hilbert H donc on peut utiliser le théorème de projection). En effet, si $y \in H$ la relation caractéristique de la projection othogonale dit que $y - p_{\overline{F}}(y)$ est orthogonal à $\overline{F}$ donc dans $F^\perp$ et comme $y - (y - p_{\overline{F}}(y)) = p_{\overline{F}}(y)$ est orthogonal à $F^\perp$, on doit avoir $y - p_{\overline{F}}(y) = p_{F^\perp}(y)$ par caractérisation de la projection.

4. On en déduit la somme $H = \overline{F} + F^{\perp}$ (par l'inclusion du 1 et l'intersection du 2, on sait que cette somme doit être directe). Le point précédent donne la relation

$$y = p_{F^{\perp}}(y) + p_{\overline{F}}(y)$$

ce qui montre que tout vecteur H se décompose comme somme d'un vecteur de $\overline{F}$ et d'un vecteur de $F^{\perp}$. L'énoncé sur les projections associées à la décomposition est évident à partir de là.
5. Il reste à voir que $F^{\perp\perp} \subset F$ ce qui donne l'égalité avec le point 1. Mais si $y \in F^{\perp\perp}$, $y - P_{\overline{F}}(y) \in F^{\perp\perp}$ par 1 et le fait fait que $F^{\perp\perp}$ est un sous-espace vectoriel. Mais on vient de voir au 3 que $y - P_{\overline{F}}(y) = p_{F^{\perp}}(y) \in F^{\perp}$. Donc $y - P_{\overline{F}}(y) \in F^{\perp\perp} \cap F^{\perp} = \{0\}$ par le 2. donc $y = P_{\overline{F}}(y) \in \overline{F}$, ce qui conclut.

□

Dualité : le théorème de représentation de Riesz

On en déduit maintenant le calcul du dual de H (voir sous-section 9 pour des rappels).

★ **Théorème 7.6** (théorème de représentation de Riesz). Soit ϕ une forme linéaire continue sur un espace de Hilbert H alors il existe un unique $f \in H$ tel que

$$\forall v \in H, \phi(v) = \langle f, v \rangle.$$

De plus, on a l'expression duale pour la norme :

$$||f|| = \sup_{||v|| \leq 1} |\langle f, v \rangle|.$$

Remarque 7.3. (facultative) Dans le cas complexe, $f \mapsto \langle f, . \rangle$ est une isométrie antilinéaire identifiant H et H' (et donc identifiant linéairement H' au conjugué $\overline{H}$ ayant la même structure normique et de groupe mais $\lambda.\overline{v} = \overline{\overline{\lambda} v}$ si $v \mapsto \overline{v}$ est la bijection/identité de $H \to \overline{H}$ notée $\bar{.}$ pour le caractère suggestif de la relation à la conjugaison complexe). Dans la cas complexe on a donc $H' \simeq \overline{H}$ et dans le cas réel $H' \simeq H$.

Démonstration. Soit $K = \phi^{-1}(\{0\})$ le noyau de ϕ. Si $K = H$ alors $f = 0$ convient. On suppose donc $K \neq H$. Soit donc $g_0 \notin K$ et $g = \frac{g_0 - P_K(g_0)}{||g_0 - P_K(g_0)||_2}$ un vecteur de norme 1 et orthogonal à K. Comme ϕ est une forme linéaire, on s'attend à ce que K et g engendrent L^2, sorte de généralisation du théorème du rang (on va voir cela plus loin en utilisant l'orthogonalité). En effet, soit $v \in H$, $w = v - \frac{\phi(v)}{\phi(g)} g$ vérifie $\phi(w) = \phi(v) - \frac{\phi(v)}{\phi(g)}\phi(g) = 0$ donc $w \in K = Ker\phi$ et $v = \lambda g + w$ avec $\lambda = \frac{\phi(v)}{\phi(g)}$.

On montre donc que $f = \overline{\phi(g)} g$ convient, en montrant l'égalité sur un v quelconque en utilisant la forme précédente :

$$\begin{aligned}\langle f, v\rangle &= \phi(g)\langle g, v\rangle = \phi(g)\langle g, \lambda g + w\rangle \\ &= \phi(g)\lambda ||g||_2^2 = \phi(g)\lambda = \phi(v).\end{aligned}$$

L'égalité des normes vient de Cauchy Schwarz qui implique que $\geq$ avec égalité en prenant $v = f/||f||$ si $f \neq 0$. □

Remarque 7.4. (facultative) Il n'est parfois pas judicieux d'identifier un espace de Hilbert à son dual, notamment quand plusieurs espaces de Hilbert sont considérés et que les identifications sont incompatibles à des relations de sous-espaces. Soit $H = \ell^2(\mathbb{N})$ et $K = \{u \in H, \sum_{n\in\mathbb{N}} n^2 |u_n|^2 < \infty\}$ Si on considère l'ensemble des suites telles que $L = \{(u_n) \sum_{n\in\mathbb{N}} \frac{1}{n^2}|u_n|^2 < \infty\}$. Il est facile de voir que $K \subset H \subset L$ et que La transposé de l'inclusion $K \subset H$ s'identifie à $H \simeq H' \subset K' \simeq L$. Il vaut alors mieux identifier K' à L (et pas K) en ayant une identification compatible avec les inclusions avec H.

4 Bases Hilbertiennes

★ **Définition 7.4.** Soit H un espace préhilbertien. Une famille $(x_i)_{i\in I}$ est dite **orthogonale** si pour tout $i \neq j$, $\langle x_i, x_j\rangle = 0$.
Si de plus $||x_i|| = 1$, elle est dite **orthonormale**.
Une **base hilbertienne** (ou base orthonormale) de H est une famille orthonormale $(e_i)_{i\in I}$ telle que $Vect(e_i, i \in I)$ est dense dans H.

Exemple 7.6. e_i la suite dont la seule coordonnée non-nulle est la i-ème égale à 1 donne une base hilbertienne de $\ell^2(I)$. (par construction de $\ell^2(I)$) Les bases hilbertiennes vont permettre d'identifier tout espace de Hilbert à cet exemple.

Procédé d'orthonormalisation de Gram–Schmidt

Notons[1] tout d'abord que la projection d'un point sur un sous-espace vectoriel de dimension finie se calcule facilement à l'aide d'une base (de préférence orthonormale) de F :

Proposition 7.7. Soit H un espace de Hilbert et F un sous-espace vectoriel de dimension finie avec $(x_1, \dots, x_n)$ une base de F (non nécessairement orthonormale). Soit $B_{i,j} = \langle x_i, x_j\rangle$. Alors B est inversible et pour tout $x \in E$, on a

$$p_F(x) = \sum_{i,j=1}^{n} (B^{-1})_{j,i}\langle x_i, x\rangle x_j.$$

Démonstration. Pour voir que B est inversible, il suffit de montrer que les vecteurs de ces lignes $(\langle x_i, x_j\rangle)_{j=1,\dots,n}$ sont linéairement indépendants. Si on a $\sum_{i=1}^{n} \lambda_i(\langle x_i, x_j\rangle)_{j=1,\dots,n} = 0$, on a $\langle \sum_{i=1}^{n} \overline{\lambda_i} x_i, x_j\rangle = 0$ pour tout j. En prenant une combinaison linéaire

$$0 = \sum_{j=1}^{n} \overline{\lambda_j} \langle \sum_{i=1}^{n} \overline{\lambda_i} x_i, x_j\rangle = || \sum_{i=1}^{n} \overline{\lambda_i} x_i ||^2,$$

donc $\sum_{i=1}^{n} \overline{\lambda_i} x_i = 0$ donc comme $x_1, \dots, x_n$ était une base, on obtient $\overline{\lambda_i} = 0$ pour tout i, ce qui donne la liberté voulue.

Pour $x \in H$, on a

$$\begin{aligned}
&\langle x_k, x - \sum_{i,j=1}^{n} (B^{-1})_{j,i}\langle x_i, x\rangle x_j\rangle \\
&\quad = \langle x_k, x\rangle - \sum_{i,j=1}^{n} (B^{-1})_{j,i}\langle x_i, x\rangle\langle x_k, x_j\rangle \\
&\quad = \langle x_k, x\rangle - \sum_{i,j=1}^{n} (B^{-1})_{j,i}\langle x_i, x\rangle B_{k,j} = 0
\end{aligned}$$

donc $x - \sum_{i,j=1}^{n} (B^{-1})_{j,i}\langle x_i, x\rangle x_j \in F^{\perp}$ donc par caractérisation de la projection orthogonale

$$p_F(x) = \sum_{i,j=1}^{n} (B^{-1})_{j,i}\langle x_i, x\rangle x_j.$$

□

1. Cette sous-section reprend le cours de 2018-2019 de T. Blossier, M. Carrizosa et J. Melleray.

Remarque 7.5. Voici un cas particulier important du résultat précédent. Soit E un espace de Hilbert et F un sous-espace vectoriel de dimension finie avec $(e_1, \dots, e_n)$ une base orthonormale de F. Alors pour tout $x \in E$, on a

$$p_F(x) = \sum_{i=1}^{n} \langle e_i, x \rangle e_i.$$

Exemple 7.7. Soit $H = L^2(\Omega, \mathcal{T}, \mu)$ et $A \in \mathcal{T}$, on a vu en TD que $\mathcal{T}(A) = \{\emptyset, A, A^c, \Omega\}$. $F = L^2(\Omega, \mathcal{T}(A), \mu)$ et un espace de dimension au plus 2 engendrée par $e_1 = 1_A$, $e_2 = 1_{A^c}$ (du moins si A, Ω ont mesures finis). Cette famille est orthogonale mais pas orthonormale. $||e_1||^2 = \int 1_A d\mu = \mu(A), ||e_2||^2 = \mu(A^c)$. Supposons ces deux nombres non nuls et finis de sorte que F a exactement dimension 2. Alors la matrice de la proposition précédente est $B = diag(\mu(A), \mu(A^c))$ et $B^{-1} = diag(1/\mu(A), 1/\mu(A^c))$, la formule de projection donne donc pour $f \in L^2(\Omega, \mathcal{T}, \mu)$:

$$\begin{aligned} &p_{L^2(\Omega, \mathcal{T}(A), \mu)}(f) \\ &= \left(\frac{1}{\mu(A)} \int_A f d\mu\right) 1_A + \left(\frac{1}{\mu(A^c)} \int_{A^c} f d\mu\right) 1_{A^c}. \end{aligned} \tag{7.2}$$

Rappelons que le procédé de Gram-Schmidt permet de calculer une base orthonormale d'un espace euclidien à partir d'une base donnée :

Proposition 7.8 (Procédé de Gram-Schmidt). Soit E un espace euclidien et $(e_1, \dots, e_n)$ une base (resp. une famille libre) de E. Pour chaque $0 < i \le n$, notons F_i le sous-espace vectoriel $\mathrm{Vec}(e_1, \dots, e_i)$ engendré par $e_1, \dots, e_i$. Alors, la famille $(e'_1, \dots, e'_n)$ définie de la manière suivante est une base orthonormale (resp. une famille orthonormale) de E :

$$\begin{aligned} e'_1 &= \frac{e_1}{\|e_1\|} \\ e'_i &= \frac{e_i - p_{F_{i-1}}(e_i)}{\|e_i - p_{F_{i-1}}(e_i)\|} = \frac{e_i - \sum_{k=1}^{i-1} \langle e'_k, e_i \rangle e'_k}{\|e_i - \sum_{k=1}^{i-1} \langle e'_k, e_i \rangle e'_k\|} \end{aligned}$$

pour $1 < i \le n$.

Exercice 7.1. Vérifier que les vecteurs $e_1 = (1, 1, 1)$, $e_2 = (1, 1, -1)$ et $e_3 = (0, 1, 1)$ forment une base de $\mathbb{R}^3$. Utiliser le procédé de Gram-Schmidt sur cette base pour obtenir une base orthonormale.

Théorème des bases

Exemple 7.8. $e_n(x) = \exp(inx)$, $n \in \mathbb{Z}$ définit une base hilbertienne de l'espace pré–hilbertien $C^0_{2\pi}(\mathbb{R}, \mathbb{C})$ l'ensemble des fonctions continues 2π périodiques, muni du produit scalaire :

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} \overline{f(t)} g(t) dt.$$

C'est la base des décompositions en série de Fourier (on montrera cela plus en détail dans la section suivante). Le but est de décomposer de façon similaire tout vecteur de H comme somme d'une série en fonction d'une base.

★ **Théorème 7.9.** Soit H un espace préhilbertien et I un ensemble dénombrable.

1. Une famille orthonormale $(x_i)_{i \in I}$ est libre et vérifie l'inégalité de Bessel, pour tout $x \in H$:
$$\sum_{i \in I} |\langle x, x_i \rangle|^2 \le ||x||^2$$
2. De plus une famille orthonormale $(e_i)_{i \in I}$ est une base hilbertienne si et seulement si on a l'égalité de Bessel–Parseval, pour tout $x \in H$:
$$\sum_{i \in I} |\langle x, e_i \rangle|^2 = ||x||^2$$
De plus, dans ce cas, pour tout $x \in H$, la série suivante converge (dans H mais pas absolument)
$$x = \sum_{i \in I} e_i \langle e_i, x \rangle.$$
3. Si H est un espace de Hilbert séparable, toute famille orthonormale peut être complétée en une base hilbertienne au plus dénombrable $(e_i)_{i \in I}$ de H et $J : x \mapsto (\langle e_i, x \rangle)_{i \in I}$ établit alors une isométrie surjective $J : H \simeq \ell^2(I)$.

Remarque 7.6. De la formule pour x, on tire par continuité la formule pour le produit scalaire (qui est une série absolument convergente par Cauchy–Schwarz) :

$$\langle y, x \rangle = \sum_{i \in I} \langle y, e_i \rangle \langle e_i, x \rangle.$$

Démonstration. Comme I est dénombrable, on peut supposer et on suppose $I = \mathbb{N}$.

(1) Si $\sum \lambda_i x_i = 0$, on calcule $\lambda_j = \langle x_j, \sum \lambda_i x_i \rangle = 0$ donc x_i est bien libre. Soit $V_n = Vect(e_i, i \in [\![0, n]\!])$, on a déjà vu la formule pour la projection orthogonale sur V_n :

$$p_n(x) = \sum_{i=0}^{n} e_i \langle e_i, x \rangle.$$

Donc par la propriété de contraction de p_n et l'orthogonalité En passant à la limite $n \to \infty$ on obtient l'inégalité de Bessel pour la somme et on trouve en particulier $(\langle x, e_i \rangle)_{i \in \mathbb{N}} \in \ell^2(\mathbb{N})$.

(2) Si $(e_i)_{i \in \mathbb{N}}$ est une base soit $x_n \in Vect(e_i, i \in I)$ convergeant vers x.

De plus, pour n assez grand $|||x||^2 - ||x_n||^2| \leq \epsilon/2$ et pour tout m,

$$\begin{aligned}\left| ||p_m(x)||^2 - ||p_m(x_n)||^2 \right| &\leq ||p_m(x_n - x)||(||x_n|| + ||x||) \\ &\leq ||(x_n - x)||(||x_n|| + ||x||) \leq \epsilon/2\end{aligned}$$

(avec la dernière inégalité pour n assez grand) d'où en prenant m tel que $p_m(x_n) = x_n$ (car x_n est dans un certain V_m comme combinaison linéaire finie des e_i), on obtient

$$\left| \sum_{i=0}^{m} |\langle e_j, x \rangle|^2 - ||x||^2 \right| \leq \epsilon$$

et donc la somme de la série est $||x||$ d'où l'égalité de Parseval.

Réciproquement, Si on a égalité, on a la limite

$$\sum_{j=0}^{n} |\langle e_j, x \rangle|^2 = ||p_n(x)||^2 \to_{n \to \infty} ||x||^2$$

et ceci implique par le théorème de Pythagore :

$$||p_n(x) - x||_2^2 = ||x||_2^2 - ||p_n(x)||_2^2 \to_{n \to \infty} 0$$

donc tout élément de H est limite d'éléments de $Vect(e_i, i \in I)$ d'où la propriété de densité manquante pour obtenir une base hilbertienne.

De plus un calcul donne la formule pour x :

$$||x - \sum_{i=0}^{n} e_i \langle e_i, x \rangle||^2 = \sum_{i=n+1}^{\infty} |\langle e_i, x \rangle|^2 \to 0.$$

(3) Soit O la famille othonormale de départ. Soit $K = \overline{Vect(O)}$, on cherche une base orthonormale de $K^\perp$ pour compléter O, il est bien séparable comme sous espace de H. Soit

$(x_n)_{n\in\mathbb{N}}$ une famille dénombrable dense de $K^\perp$. Quitte à extraire une sous-suite, on peut supposer que $x_n \notin Vect(x_0, \ldots, x_{n-1})$ de sorte que $(x_n)_{n\in\mathbb{N}}$ est une famille libre.

On peut donc orthonormaliser $(x_0, \ldots, x_n)$ et obtenir $(e_0, \ldots, e_n)$ tel que $Vect(x_0, \ldots., x_n) = Vect(e_0, \ldots., e_n)$. Par la construction, on remarque que l'orthonormalisation pour $(x_0, \ldots, x_{n+1})$ on commence par les mêmes vecteurs et on obtient donc une famille orthonormale $(f_n)_{n\in\mathbb{N}}$. Comme

$$\begin{aligned} Vect(x_n, n\in\mathbb{N}) &= \cup_{n=0}^{\infty} Vect(x_0, \ldots., x_n) \\ &= \cup_{n=0}^{\infty} Vect(f_0, \ldots., f_n) \\ &= Vect(f_n, n\in\mathbb{N}), \end{aligned}$$

ces deux ensembles sont denses et donc $(f_n)_{n\in\mathbb{N}}$ est une base de $K^\perp$. Maintenant, O et $(f_n)_{n\in\mathbb{N}}$ forment une famille orthonormale de H et tout O est une base de K par définition de K, donc la décomposition orthogonale $x = P_K(x) + P_{K^\perp}(x)$ permet d'approcher $P_K(x)$ par un élément $y_n \in Vect(O)$, $P_{K^\perp}(x)$ par un élément $z_n \in Vect(f_n, n\in\mathbb{N})$ et $y_n + z_n \in Vect(O, f_n, n\in\mathbb{N})$ tend vers x, d'où la densité voulue pour que $\{e_n, n\in\mathbb{N}\} = O \cup \{f_n, n\in\mathbb{N}\}$ forme une base de H.

Une fois l'existence d'une base, l'isométrie est évidente par le (2), et si on a une suite $(\lambda_i)_{i\in I}$ dans $\ell^2(I)$, on voit que $\sum \lambda_i e_i$ converge par complétude comme ci-dessus et on obtient ainsi la surjectivité.

On vient de voir (en prolongeant la famille vide) qu'un espace de Hilbert séparable a une base dénombrable. Réciproquement, un espace de Hilbert à base dénombrable est isométrique à $\ell^2(\mathbb{N})$ pour lequel $\mathrm{Vect}_{\mathbb{Q}}(e_n, n\in\mathbb{N}\}$ donne une famille dénombrable dense. □

Exemples de base 1 : Séries de Fourier

On va obtenir un premier exemple de base en utilisant le théorème d'approximation de Weierstrass.

Vous pouvez voir dans la section de compléments le corollaire A.10 pour une preuve probabiliste basée sur la loi faible des grands nombres.

★ **Théorème 7.10** (d'approximation de Weierstrass)**.** Soit K un compact de $\mathbb{R}^n$, les fonctions polynômiales (à coefficients réels ou même rationnels) sont denses dans $C^0(K, \mathbb{R})$.

En conséquence, $(C^0(K,\mathbb{R}), ||.||_\infty)$ est séparable et sa tribu borélienne $\mathcal{B}(C^0(K,\mathbb{R}))$ est dénombrablement engendrée (c'est à dire admet une partie génératrice au plus dénombrable).

Remarque 7.7. Le mouvement brownien sur $[0,1]$, un objet probabiliste important (vu en M1) peut être défini comme une probabilité sur la tribu borélienne de $(C^0([0,1],\mathbb{R}), ||.||_\infty)$.

Exemple 7.9. Montrons que $e_n(x) = \exp(inx)$, $n \in \mathbb{Z}$ forme une base hilbertienne de $L^2([0,2\pi],\mathbb{C})$:

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} \overline{f(t)}g(t)dt.$$

D'abord, on sait que $C_b^0(]0,2\pi[,\mathbb{C})$ est dense car il contient $C_c^0(]0,2\pi[)$ qui est dense par le Théorème 6.12. Il s'agit donc presque de la complétion de l'exemple précédent.
Ensuite on vérifie l'orthonormalité :

$$\langle e_n, e_m\rangle = \frac{1}{2\pi}\int_0^{2\pi} exp(i(m-n)t)dt = 1_{\{m=n\}}.$$

Enfin, il reste à voir que $Vect(e_n)$ est dense. Or, on a $Vect(e_n) = \{P(e^{ix}, e^{-ix}), P \in \mathbb{C}[X,Y]\} = \{P(\cos(x), \sin(x)), P \in \mathbb{C}[X,Y]\}$. Soit $D = \{(x,y) \in \mathbb{R}^2, x^2+y^2 = 1\}$, soit $f \in C_{2\pi}^0(\mathbb{R},\mathbb{C})$ On définit $g : D \to \mathbb{C}$ par $g(\cos(x), \sin(x)) = f(x)$. Il est facile de voir que g est continue sur D (utiliser tan, cot selon le point comme carte coordonnée) donc par le théorème d'approximation de Weierstrass 7.10, il existe un polynôme P tel que $||P-g||_\infty \leq \epsilon$ donc, si $Q = P(\cos(.), \sin(.)) \in Vect(e_n)$, on a $||Q-f||_2 \leq ||Q-f||_\infty \leq ||P-g||_\infty \leq \epsilon$. D'où la densité voulue.
C'est la base des décompositions en série de Fourier.

Exemple de base 2 : Polynômes d'Hermite

L'exercice suivant est corrigé à l'annexe E en section 3. Vérifier qu'une famille est orthonormée est toujours un exercice calculatoire.

Exercice 7.2. Soit $H = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$ l'espace de Hilbert réel des fonctions de carrés intégrables pour la mesure gaussienne standard définie pour un borélien B par $\gamma(B) = \int_B \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx$. H muni de la norme usuelle :

$$||f||_2 = \sqrt{\int_{\mathbb{R}} |f(x)|^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx}.$$

Soit

$$H_n(x) = (-1)^n \frac{e^{x^2/2}}{\sqrt{n!}} \left(\frac{d}{dx}\right)^n (e^{-x^2/2})$$

(et donc $H_0(x) = 1$). On appelle les H_n les **polynômes d'Hermite**.

1. Montrer que pour $n \geq 1$, H_n est un polynôme de la forme :

$$H_n(x) = \frac{x^n}{\sqrt{n!}} + \sum_{k=0}^{n-1} a_k x^k.$$

2. Montrer que $(H_n)_{n\geq 0}$ est une famille orthonormale de H.

Montrer le résultat de densité sous–jacent pour obtenir une base est souvent plus dur. Quand on ne peut pas utiliser un résultat connu, on utilise souvent la méthode qui consiste à montrer que l'orthogonale est $\{0\}$ en utilisant la proposition 7.5. On va donc déduire le résultat suivant de cela et du théorème d'inversion de Fourier :

Théorème 7.11. Soit γ la mesure gaussienne standard sur $\mathbb{R}$. Alors la famille des polynômes d'Hermite $(H_n)_{n\geq 0}$ est une base orthonormale de $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$. En particulier, les polynômes sont denses dans $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$ qui est séparable.

Démonstration. Montrons d'abord que la série $\exp(-t^2/2) \sum_{n=0}^{\infty} \frac{(it)^n}{\sqrt{n!}} H_n$ converge dans $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$.

On calcule la norme du terme général de la suite $S_N = \exp(-t^2/2) \sum_{n=0}^{N} \frac{(it)^n}{\sqrt{n!}} H_n$ par orthonormalité de s (H_n) :

$$\begin{aligned} ||S_N||_2^2 &= \exp(-t^2) \sum_{n=0}^{N} \frac{|(it)^n|^2}{n!} \\ &= \exp(-t^2) \sum_{n=0}^{N} \frac{(t^2)^n}{n!} \leq \exp(t^2 - t^2) = 1 \end{aligned}$$

Donc pour $p \geq q \geq N$, $||S_{p+1} - S_q||_2^2 \leq \exp(-t^2) \sum_{n=N}^{\infty} \frac{(t^2)^n}{n!} \to_{N\to\infty} 0$. Donc S_n est de Cauchy et donc converge dans L^2. Quitte à extraire on sait qu'elle converge presque partout, donc sa limite ponctuelle sera aussi sa limite dans L^2. Concluons que F_t, définie par $F_t(x) = exp(itx)$, est la limite. Il suffit donc de voir que pour tout $x \in \mathbb{R}$:

$$F_t(x) = \exp(-t^2/2) \sum_{n=0}^{\infty} \frac{(-it)^n}{\sqrt{n!}} H_n(x).$$

Ceci équivaut, vu la définition de H_n à

$$F_t(x) \exp(t^2/2 - x^2/2) = exp(-(it - x)^2/2) = \sum_{n=0}^{\infty} \frac{(-it)^n}{n!} \left(\frac{d}{dx}\right)^n (e^{-x^2/2})$$

ce qui est la somme de la série de Taylor en x évaluée en $a = it$ de $f(x) = exp(-x^2/2)$ (pour f somme de série entière sur $\mathbb{C}$ $f(x+a) = \sum_{n=0}^{\infty} \frac{a^n}{n!} f^{(n)}(x)$. Ceci est bien vérifié car la fonction du milieu est analytique par composée de fonctions analytiques sur $\mathbb{C}$ (un polynôme et exp sont sommes de séries entières sur $\mathbb{C}$ donc aussi leur composée).

Conclusion : on a $F_t \in \overline{Vect(H_n, n \in \mathbb{N})}$.

On montre maintenant que toute fonction $f \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$, orthogonale à $K := Vect(H_n, n \in \mathbb{N})$ est nulle. On peut supposer f réelle en prenant partie réelle et imaginaire. Si f orthogonale à tout H_n on a $\langle f, F_t \rangle = 0$ et donc

$$u(t) = \int f(x) exp(itx - x^2/2) = 0.$$

Or si $g(x) = f(x)exp(-x^2/2)$ $g \in L^1(\mathbb{R}, \lambda)$ est équivalent à $f \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$ ce qui est le cas car γ est une mesure de probabilité et donc $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma) \subset L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$. Donc on a $\hat{g}(t) = 0$ et par le théorème d'inversion de Fourier, $g(x) = 0$ presque partout, soit $f = 0$ dans $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$.

Bilan pour $K = Vect(H_n, n \in \mathbb{N})$ $K^\perp = \{0\}$ donc $\overline{K} = K^{\perp\perp} = \{0\}^\perp = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \gamma)$, d'où la densité voulue. □

On a utilisé le théorème suivant (peut-être vu en cours de probabilité, cf. annexe E section 4 pour la variante sur les mesures de probabilité, cf. aussi le livre de Rudin d'analyse réelle et complexe Thm 9.11 et 9.12 pour $n = 1$)

Définition 7.5. Soit $f \in L^1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$ la transformée de Fourier de f est la fonction de $t \in \mathbb{R}^n$:

$$\hat{f}(t) = \int_{\mathbb{R}^n} e^{i\langle x,t\rangle} f(x) \lambda(dx).$$

On renvoie à la section E.4 pour une preuve du résultat fondamental suivant.

★ **Théorème 7.12** (Théorème d'injectivité de la transformation de Fourier (admis))**.** Soient deux fonctions $f_1, f_2 \in L^1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$ On suppose que pour tout $t \in \mathbb{R}^n$ les transformées de Fourier sont égales :

$$\hat{f}_1(t) = \hat{f}_2(t), \qquad \forall t \in \mathbb{R}^n.$$

Alors $f_1 = f_2$ presque partout.
De plus, si $\hat{f}_1 \in L^1(\mathbb{R}^n, \lambda)$ alors f_1 est (égale presque partout à) une fonction continue :

$$f_1(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}_1(t) exp(-i\langle x, t\rangle) dt.$$

5 Une Application : Le théorème de convergence des martingales bornées dans $L^2(\Omega, \mathcal{T}, P)$ (facultatif)

Dans cette section, on conclut par une application en probabilité. On prend $(\Omega, \mathcal{T}, P)$ un espace de probabilité. Une filtration est une suite croissante de sous–tribu $(\mathcal{T}_n)_{n\geq 0}$. Un exemple de telle suite est $\mathcal{T}_n = \mathcal{T}((X_0, \ldots, X_n))$ de la tribu engendrée par un vecteur aléatoire. On peut considérer les espaces de Hilbert $H_n = L^2(\Omega, \mathcal{T}_n, P) \subset L^2(\Omega, \mathcal{T}, P)$. C'est un sous–espace fermé car si $H_n \ni X_m \to_{m\to\infty} X$ on a vu au chapitre précédent, que quitte à extraire X_{m_k} converge p.p. vers X et donc X est aussi $\mathcal{T}_n$–mesurable et donc est dans H_n. Par caractérisation séquentielle cela dit H_n fermé. On dispose donc de la projection orthogonale P_{H_n}. EN probabilité, vous noterez $P_{H_n}(X) = E(X|\mathcal{T}_n)$ et vous interpréterez cette projection comme une espérance conditionnelle.

Définition 7.6. Une suite $(X_n)_{n\in\mathbb{N}}$ est une **martingale dans** L^2 (pour la filtration $(\mathcal{T}_n)_{n\geq 0}$ si pour tout $m \geq n$ $P_{H_n}(X_m) = X_n$.

Cette condition dit que la moyenne de la future variable X_m, conditionnellement au présent H_n, est égale à X_n (si X_n est la valeur d'un gain au temps n, en moyenne on n'a rien gagné à attendre le temps $m > n$). Une somme de v.a. i.i.d. dans L^2 d'espérance nulle est une telle martingale. Par exemple, la somme des n premiers termes d'une suite de variables gaussiennes centrées indépendantes donne une martingale dans L^2. On va montrer un théorème de convergence pour les martingales bornées dans L^2.

Théorème 7.13. Soit $(\mathcal{T}_n)_{n\geq 0}$.Soit $(X_n)_{n\in\mathbb{N}}$ est une martingale dans $H = L^2(\Omega,\mathcal{T},P)$ qui est une suite bornée, c'est-à-dire, qu'il existe $M > 0$ telle que $\sup_n ||X_n||_2 \leq M$. Alors X_n converge dans $L^2(\Omega,\mathcal{T},P)$ vers une variable X et $X_n = P_{H_n}(X)$.

Ce théorème se généralise à un théorème de convergence des martingales bornées dans L^p, $1 < p < \infty$. Il y a aussi une version pour les martingales L^1 mais il faut une hypothèse technique plus compliquée (dite d'uniforme intégrabilité). (On dit que X_n est une martingale fermée quand $X_n = P_{H_n}(X)$ comme ci-dessus).

Démonstration. On considère la décomposition orthogonale $H_{n+1} = K_n \oplus H_n$ avec $H_0 = K_0$ On voudrait dire que $L^2(\Omega,\mathcal{T}(\cup_{n\geq 0}\mathcal{T}_n),P) = \oplus_{n\geq 0} K_n$ est une somme orthogonale infinie, mais comme on n'a pas introduit la notion,on va donc faire une preuve directe.

Remarquez déjà que $X_{n+1} - X_n = X_{n+1} - P_{H_n}(X_{n+1}) \in K_n$ par la condition de martingale. Donc par le théorème de Pythagore et une récurrence triviale, on obtient :

$$||X_{n+1}||_2^2 = ||X_{n+1} - X_n||_2^2 + ||X_n||_2^2 = ||X_0||_2^2 + \sum_{k=0}^{n} ||X_{k+1} - X_k||_2^2.$$

On déduit donc de la bornitude en prenant la limite $||X_0||_2^2 + \sum_{k=0}^{\infty} ||X_{k+1} - X_k||_2^2 \leq M^2$ et donc la série est convergente. On déduit aussi que pour $p \geq q \geq N$

$$||X_{p+1} - X_q||_2^2 = \sum_{k=q}^{p} ||X_{k+1} - X_k||_2^2 \leq \sum_{k=N}^{\infty} ||X_{k+1} - X_k||_2^2 \to_{N\to\infty} 0.$$

Donc (X_n) est de Cauchy dans un espace de Hilbert donc converge vers X. Comme P_{H_n} est 1-lipschitz donc continue, on déduit en passant à la limite dans la relation
$X_n = P_{H_n}(X_m) \to_{m\to\infty} P_{H_n}(X) = X_n$

□

ANNEXE A
Compléments facultatifs au chapitre 2 : Topologie des espaces métriques

1 Théorème de Tietze (niveau L3-M1)

Comme jolie application de la complétude, on va donner en exercice (corrigé), la preuve du théorème de Tietze

Exercice A.1. **Extension de Tietze-Urysohn**
Soit F un fermé de X espace métrique. Soit $E = C_b^0(X,\mathbb{R})$ l'espace des fonctions continues bornées et $p : E \to C_b^0(F,\mathbb{R})$ l'application de restriction (pour $f : X \to \mathbb{R}$, $p(f) = f|_F$ est la restriction de f à F. On va montrer que p est surjective.

1. Est-ce que E est complet ?
2. Soit $g \in C_b^0(F,\mathbb{R})$ avec $||g||_\infty \leq 1$. Soient $K_1 := g^{-1}([1/3, 1])$ et $K_2 := g^{-1}([-1, -1/3])$. Soit :
$$f(x) = \frac{1}{3}\frac{d(x, K_2) - d(x, K_1)}{d(x, K_2) + d(x, K_1)},$$
$$d(x, K_i) := \inf\{d(x, y), y \in K_i\}.$$
(On comprend la valeur comme 0 si K_1 et K_2 vides et sinon, $-1/3$ si K_1 vide, $1/3$ si K_2 vide). Vérifier que $f \in E$
3. Montrer que $||f||_\infty \leq 1/3$ et $||p(f) - g||_\infty \leq \alpha = 2/3$..
4. Construire une suite f_n par récurrence à partir du résultat précédent telle que $f_n = F_0 + \dots + F_n$ et
$$\sum_{k=0}^{n} ||F_k||_\infty \leq \frac{1}{3}(1 + \dots + \frac{2^n}{3^n})$$
et
$$||p(f_n) - g||_\infty \leq \frac{2^{n+1}}{3^{n+1}}.$$
5. Montrer que f_n converge. En déduire, qu'il existe $F \in E$, $||F||_\infty \leq 1$ telle que $p(F) = g$.

Extension de Tietze-Urysohn (Correction)

Soit F un fermé de X espace métrique. Soit $E = C_b^0(X, \mathbb{R})$ et $p : E \to C_b^0(F, \mathbb{R})$ l'application de restriction. On va montrer que p est surjective (et un peu mieux).

1. Soit $g \in C^0(K)$ avec $||g||_\infty \leq 1$. Soient $K_1 := g^{-1}([1/3, 1])$ et $K_2 := g^{-1}([-1, -1/3])$. Soit :
$$f(x) = \frac{1}{3}\frac{d(x, K_2) - d(x, K_1)}{d(x, K_2) + d(x, K_1)}, .$$
Vérifions que $f \in E, ||f||_\infty \leq 1/3$ et $||p(f) - g||_\infty \leq \alpha = 2/3$. (on dit que p est presque surjective)
f est continue car $d(., K_i)$ est continue et le dénominateur est non nul car $K_1 \cap K_2 = \emptyset$ et $d(., K_i) > 0$ sur K_i^c.
2. Or par l'inégalité triangulaire :
$$|f(x)| \leq \frac{1}{3}\frac{d(x, K_2) + d(x, K_1)}{d(x, K_2) + d(x, K_1)} = \frac{1}{3}$$
donc f est bornée et $||f||_\infty \leq 1/3$.
$$\begin{aligned}|p(f) - g| &= 1_{K_1}|\frac{1}{3} - g| + 1_{K_2}| - \frac{1}{3} - g| + (1_K - 1_{K_1} - 1_{K_2})|f - g| \\ &\leq 1_{K_1}||1_{K_1}(\frac{1}{3} - g)||_\infty + 1_{K_2}||1_{K_1}(-\frac{1}{3} - g)||_\infty + (1_K - 1_{K_1} - 1_{K_2})(||f||_\infty + ||g||_\infty)\end{aligned}$$
et tous les termes sont inférieurs à $2/3$ par définition.
3. On construit construire une suite f_n par récurrence à partir du résultat précédent telle que $f_n = F_0 + \ldots + F_n$
$$\sum_{k=0}^{n} ||F_k||_\infty \leq \frac{1}{3}(1 + \ldots + \frac{2^n}{3^n})$$
et
$$||p(f_n) - g||_\infty \leq \frac{2^{n+1}}{3^{n+1}}.$$
On prend $f_0 = F_0 = f$ donné par 1 à partir de g. On prend $F_n/||p(f_{n-1}) - g||_\infty$ donné par 1 à partir de $-[p(f_{n-1}) - g]/||p(f_{n-1}) - g||_\infty$ (si le dénominateur est 0 on s'arrête et on prend la suite constante).
Donc on a les deux inégalités
$$||F_n||_\infty \leq \frac{1}{3}||p(f_{n-1}) - g||_\infty \leq \frac{1}{3}\frac{2^n}{3^n}$$
et
$$||p(F_n) + p(f_{n-1}) - g||_\infty \leq \frac{2}{3}||p(f_{n-1}) - g||_\infty \leq \frac{2^{n+1}}{3^{n+1}}$$
La deuxième inégalité donne $||p(f_n) - g||_\infty \leq \frac{2^{n+1}}{3^{n+1}}$. La première inégalité suit par l'hypothèse de récurrence.

4. Déduisons qu'il existe $F \in E$, $||F||_\infty \le 1$ telle que $p(F) = g$. $\sum F_n$ est donc absolument convergente dans E, donc par complétude convergente, donc soit $F = \sum_{n=0}^{\infty} F_n = \lim f_n$. En passant à la limite on obtient (par la somme d'une série géométrique)

$$||F||_\infty \le \sum_{n=0}^{\infty} ||F_n||_\infty \le \frac{1}{3}\sum_{n=0}^{\infty} \frac{2^{n+1}}{3^{n+1}} \le \frac{1}{3}\frac{1}{1-2/3} = 1$$

et $||p(F) - g||_\infty = 0$ donc $p(F) = g$ par séparation.

2 Complément sur l'Espace dual (niveau début de M1)

Définition A.1. L'espace $E' := L(E, \mathbb{K})$ des formes linéaires continues sur un e.v.n. E est munie de la norme duale

$$||f||_{E'} := \sup_{x\in E, ||x||\le 1} |f(x)|.$$

On a vu dans la section précédente que c'est toujours un **espace de Banach**. Il sera très utile dans ce cours pour étudier E lui-même.

Le résultat suivant, conséquence de Hahn-Banach permet de décrire réciproquement la norme de E en terme de celle de E' (cela ressemble à la définition de $||f||_{E'}$ mais c'est un théorème difficile ! que l'on exploitera pour relier E au dual du dual dans la section suivante) :

Proposition A.1. Soit $(E, ||.||_E)$ un e.v.n., alors

$$\|x\|_E = \sup_{f\in E', ||f||_{E'}\le 1} |f(x)| = \max_{f\in E', ||f||_{E'}\le 1} |f(x)|.$$

Démonstration. Par définition, on a

$$\sup_{f\in E', ||f||_{E'}\le 1} |f(x)| \le \sup_{f\in E', ||f||_{E'}\le 1} ||f||_{E'}\|x\|_E = \|x\|_E.$$

Inversement, on applique le Théorème de Hahn-Banach B.9 à $G = \mathbb{R}x$ en posant $g(tx) = t||x||_E$ de sorte que $g(tx)|| \le ||tx||_E$. Donc, il existe $f \in E'$ tel que $f(x) = g(x) = ||x||_E$ et $f(y) \le ||y||_E$ c'est-à-dire $||f||_{E'} \le 1$. En particulier, le sup est atteint en f et est donc un maximum. □

On rappelle deux exemples d'espaces classiques.

Exemple A.1. $c_0(I)$ est l'ensemble des suites $(x_i)_{i\in I}$ qui tendent vers 0 dans le sens où si $\epsilon > 0$, il existe une partie F finie telle que $|x_i| \leq \epsilon$ pour tout $i \notin F$. On munit $c_0(I)$ de la norme sup :

$$||x||_\infty = \sup_{i\in I} |x_i| < \infty.$$

$\ell^\infty(I)$ est l'ensemble des suites bornée $(x_i)_{i\in I}$ avec la même norme $||x||_\infty$.

Exemple A.2. $\ell^1(I)$ est l'ensemble des suites $(x_i)_{i\in I}$ sommables, tel qu'il existe une constante C, tel que pour toute partie F finie telle que $\sum_{i\in F} |x_i| \leq C$. On munit $\ell^1(I)$ de la norme :

$$||x||_1 = \sup_F \sum_{i\in F} |x_i| =: \sum_{i\in I} |x_i| < \infty.$$

On étudiera la dualité des espaces L^p dans un chapitre ultérieur. Le résultat suivant donne un exemple de calcul de dual :

Proposition A.2. Le dual de $c_0(I)$ est isométrique à

$$\ell^1(I) \simeq (c_0(I))'.$$

Démonstration. On définit $T : \ell^1(I) \to (c_0(I))'$ par :

$$T((u_i))[(v_i)] = \sum_{i\in I} u_i v_i.$$

Bien sûr, on a l'inégalité montrant que T est bien défini et contractant :

$$|T((u_i))[(v_i)]| \leq \sum_{i\in I} |u_i|\,|v_i| \leq ||c||_\infty \sum_{i\in I} |u_i|.$$

Montrons que T est isométrique. Comme les suites à support fini sont denses dans $\ell^1(I)$ il suffit de montrer l'égalité dans ce cas, et cela vient en posant $(v_i) = 1_{\{v_i\neq 0\}} \frac{\overline{v_i}}{|v_i|} \in c_0(I)$ si (u_i) à support fini de $T((u_i))(v_i) = ||(u_i)||_{\ell^1}$. Donc comme $||(v_i)||_{c_0} \leq 1$ on a l'inégalité manquante :

$$||T((u_i))||_{(c_0)'} \geq ||(u_i)||_{\ell^1}.$$

Montrons que T est surjectif. Soit $f \in (c_0(I))'$ et e_i la suite valant 1 en i et 0 ailleurs. Soit $u_i = f(e_i)$, montrons que $(u_i) \in \ell^1(\mathbb{N})$. Or par l'isométrie

$$||(u_i 1_{i\in F})||_{\ell^1} \leq ||T((u_i 1_{i\in F}))||_{(c_0)'} = ||T((u_i)) \circ v_F||_{(c_0)'} = ||f \circ v_F||_{(c_0)'} \leq ||f||_{(c_0)'}$$

car $v_F((x_i)) = (1_{i\in F} x_i)$ est une contraction sur c_0 pour F fini (et par le calcul à support fini qui suit qui implique $f \circ v_F = T((u_i)) \circ v_F$). Donc pour tout F fini :

$$\sum_{i\in F} |u_i| \leq ||f||_{(c_0)'}$$

ce qui donne la sommabilité $u \in \ell^1(I)$.

Montrons enfin que $f = T((u_i))$.

En effet, si v est à support fini, $f(v) = T((u_i))(v)$ par linéarité mais comme les deux côtés sont continus en v et que (par définition) les suites à support fini sont denses dans $c_0(I)$, on obtient $f = T((u_i))$.

□

Un autre résultat de base permet d'associer à une application continue $u : E \to F$ une application (dite transposée ou adjoint) entre les duaux $u^t : F' \to E'$.

Proposition A.3. Si $u : E \to F$ est une application linéaire continue $u^t(f) = f \circ u$ définie une application linéaire continue $u^t : F' \to E'$ et on a

$$|||u^t||| = |||u|||.$$

Démonstration. Par composition, si $f \in F'$, u linéaire continue, $f \circ u$ est linéaire continue donc appartient à E'. La linéarité en f est évidente. de plus $||u^t(f)(x)|| \leq ||f||_{F'}|||u|||||x||_E$ donc

$$||u^t(f)||_{E'} \leq ||f||_{F'}|||u|||.$$

Ceci donne $|||u^t||| \leq |||u|||$.

Réciproquement on utilise la proposition précédente pour obtenir :

$$||u(x)||_F = \sup_{||f||_{F'}\leq 1} |(u^t(f)(x))| \leq \sup_{||f||_{F'}\leq 1} ||(u^t(f)||_{E'}||x||_E \leq |||u^t|||||x||_E.$$

Ceci donne par définition de la norme subordonnée, l'autre inégalité : $|||u||| \leq |||u^t|||$.

□

3 Bidual, Complété (niveau début de M1)

Le dual du dual $E'' = (E')'$ est appelé **bidual de** E.

Définition A.2. L'application $J : E \to E''$ qui envoie $J(x)(f) = f(x)$ pour $f \in E'$ est appelée **injection canonique** de E dans E''.

Proposition A.4. L'injection canonique $J : E \to E''$ est une isométrie (c'est pour cela que c'est une injection).

Démonstration. En appliquant la définition de la norme du dual puis la conséquence de Hahn–Banach de la section précédente (proposition A.1), on obtient :

$$||J(x)||_{E''} = \sup_{||f||_{E'} \leq 1} |J(x)(f)| = \sup_{||f||_{E'} \leq 1} |f(x)| = ||x||_E.$$

□

On donne un exemple :

Proposition A.5.

$$(c_0(I))'' \simeq (\ell^1(I))' \simeq \ell^\infty(I).$$

Démonstration. On définit $T : \ell^\infty(I) \to (\ell^1(I))'$ par :

$$T((u_i))[(v_i)] = \sum_{i \in I} u_i v_i.$$

Bien sûr, on a l'inégalité montrant que T est bien défini et contractant :

$$|T((u_i))[(v_i)]| \leq \sum_{i \in I} |u_i|\ |v_i| \leq ||c||_\infty \sum_{i \in I} |u_i|.$$

Montrons que T est surjectif. Soit $f \in (\ell^1(I))'$ et e_i la suite valant 1 en i et 0 ailleurs. Soit $u_i = f(e_i)$, alors $|u_i| \leq ||f||_{\ell^1}$ donc $(u_i) \in \ell^\infty(I)$, montrons que $f = T((u_i))$.

En effet, si v est à support fini, $f(v) = T((u_i))(v)$ par linéarité mais comme les deux côtés sont continus en v et que (par définition) les suites à support fini sont denses dans $\ell^1(I)$, on obtient $f = T((u_i))$.

Montrons que T est isométrique. Mais $||T(u_i)|| \geq |T(u_i)(e_i)| = |u_i|$ donc $||T(u_i)|| \geq ||(u_i)||_{\ell^\infty(I)}$ et on obtient donc l'égalité.

□

Définition A.3. L'adhérence $\widehat{E} := \overline{J(E)}^{E''}$ E dans E'' est appelée **complété de** E.

Comme c'est un espace fermé d'un espace complet, c'est un espace de Banach muni d'une injection $i : E \to \widehat{E}$ (qui est id si E est déjà n espace de Banach). Il est caractérisé par la propriété universelle suivante. Contrairement à la compacité qui est dure à trouver en dimension infinie, la complétude est simple grâce à cette construction, car il suffit de passer au complété (mais, dans des espaces de fonctions, il faut travailler pour décrire plus explicitement ce complété, comme espace de fonctions concrètes).

Proposition A.6. Soit F un espace de Banach et $u : E \to F$ une application linéaire continue, il existe une unique extension $\widehat{u} : \widehat{E} \to F$ telle que $\hat{u} \circ i = u$. De plus, on a $|||\hat{u}||| = |||u|||$.

Démonstration. pour l'existence on considère $(u^t)^t : E'' \to F''$ et on regarde sa restriction $\widehat{u}$ à $\widehat{E}$. Sur E, $\widehat{u}$ coincide avec u donc est à valeur dans F. Par densité de E, l existe une suite $u_n \to u \in \widehat{E}$ et donc $\widehat{u}(\widehat{E}) \subset \widehat{F}$. Or comme F est complet il est fermé dans son bidual donc $\widehat{F} = F$. Cela donne l'existence. L'unicité vient de la densité de E dans $\widehat{E}$. Par la construction on a $|||\hat{u}||| \leq |||u|||$. L'autre inégalité vient par densité. □

4 Compléments sur la compacité et complétude (niveau L2-L3)

Définition A.4. Un espace métrique (X, d) est précompact si pour tout $\epsilon > 0$, X peut être couvert par un nombre fini de boules ouvertes de rayon ϵ.

On rappelle le résultat suivant (cf. e.g. Zuily-Quéffelec Th II.1 p135 ou Gourdon d'Analyse p 32) :

Proposition A.7. Un espace métrique X est compact si et seulement si il est précompact et complet.

Démonstration. L'implication, compact implique précompact vient de la définition. L'implication compact implique complet vient de Bolzano-Weierstrass (vu qu'une suite de Cauchy ayant une sous-suite convergente converge).

Réciproquement, on utilise aussi Bolzano-Weierstrass. On va construire une suite extraite de Cauchy par extraction diagonale. Soit (x_n) suite de $X.X$ est recouvert par un nombre fini de boules $B(a,1)$ donc par principe des tiroirs, il existe une sous-suite $(x_{\phi_0(n)})$ de (x_n) contenu dans une de ces boules $B(a_0,1)$. Par récurrence, on obtient une suite extraite $(x_{\phi_0\circ\ldots\circ\phi_p(n)})$ contenu dans $B(a_p,1/2^p)$ en ayant choisi un recouvrement fini $B(a,1/2^p)$ de $B(a_{p-1},1/2^{p-1})$ et un terme de ce recouvrement contenant une sous-suite de la suite-extraite précédente $(x_{\phi_0\circ\ldots\circ\phi_{p-1}(n)})$. On considère l'extraction diagonale $y_n = x_{\phi_0\circ\ldots\circ\phi_n(n)}$. Vu que $\phi_i(n)\geq n$ car les ϕ_i sont strictement croissantes,
$\psi(n)=\phi_0\circ\ldots\circ\phi_n(n)\geq\phi_0\circ\ldots\circ\phi_{n-1}(n)>\phi_0\circ\ldots\circ\phi_{n-1}(n-1)=\psi(n-1)$ donc $y_n=x_{\psi(n)}$ est bien une suite extraite telle que à partir du rang n, $(y_k)_{k\geq n}$ extraite de $(x_{\phi_0\circ\ldots\circ\phi_n(k)})$ est dans la boule $B(a_n,1/2^n)$. Donc y_k est de Cauchy donc converge par complétude. □

Théorème A.8. (de Tychonov)Un produit $\prod_{i\in I}X_i$ d'espaces topologiques compacts est compact.

Comme le cas non-métrique, non-dénombrable utilise l'axiome du choix sous la forme du lemme de Zorn, on reverra cela plus loin.

Exercice A.2. Si I dénombrable, X_i métriques, montrer que $\prod_{i\in I}X_i$ est un espace métrique compact. (Indication utiliser le résultat précédent.)

5 Théorème d'approximation de Weierstrass (niveau L3-M1)

Théorème A.9 (de Bernstein)**.** Soit $f:[0,1]^n\to\mathbb{C}$ continue et définissons le polynôme de Bernstein :

$$B_N(f)(x_1,\ldots,x_n)=\sum_{k_1=0}^{N}\cdots\sum_{k_n=0}^{N}$$
$$C_N^{k_1}\ldots C_N^{k_n}f(\frac{k_1}{N},\ldots,\frac{k_n}{N})x_1^{k_1}(1-x_1)^{N-k_1}\ldots x_n^{k_n}(1-x_n)^{N-k_n}$$

Alors $B_N(f)$ converge uniformément sur $[0,1]^n$ vers f

Démonstration. On interprète de façon probabiliste $B_N(f)$. Soit $\Omega = \{0,1\}^{Nn}$ avec la mesure de probabilité

$$P(\omega_1 = i_1, \ldots, \omega_{Nn} = i_n) = x_1^{k_1}(1-x_1)^{N-k_1} \ldots x_n^{k_n}(1-x_n)^{N-k_n}$$

avec k_i le nombre de 1 parmi $i_{N(i-1)+1}, \ldots, i_{Ni}$. On note $S_1(\omega) = \frac{\omega_1 + \ldots + \omega_N}{N}, \ldots, S_n(\omega) = \frac{\omega_{N(n-1)+1} + \ldots + \omega_{Nn}}{N}$, $S = (S_1, \ldots, S_n)$ qui sont des variables de loi binomiales indépendantes du point de vue probabiliste. Alors $\int dPf(S_1, \ldots, S_n) = B_N(f)(x_1, \ldots, x_n)$, donc si $\omega(h) = \sup\{|f(x) - f(y)| : |x-y| \leq h\}$ est le module d'uniforme continuité de f, on a :

$$\begin{aligned}|f(x_1, \ldots, x_n) - B_N(f)(x_1, \ldots, x_n)| &\leq ||f(x_1, \ldots, x_n) - f(S)||_1 \\ &\leq \omega(\delta) + 2||f||_\infty P\left(|(x_1, \ldots, x_n) - S| \geq \delta\right)\end{aligned}$$

Or par union disjointe et l'inégalité de Markov :

$$P\left(|(x_1, \ldots, x_n) - S| \geq \delta\right) \leq \sum_{i=1}^n P\left(|x_i - S_i| \geq \delta\right) \leq \sum_{i=1}^n \frac{E(|x_i - S_i|^2)}{\delta^2}$$

Or un calcul simple donne $E(|x_i - S_i|^2) = Var(S_i) = \frac{x_i(1-x_i)}{N} \leq \frac{1}{4N}$ donc

$$\begin{aligned}&\limsup_{N\to\infty} \sup_{(x_1,\ldots,x_n)\in[0,1]^n} |f(x_1, \ldots, x_n) - B_N(f)(x_1, \ldots, x_n)| \\ &\leq \limsup_{N\to\infty} \omega(\delta) + \frac{2n||f||_\infty}{4N\delta^2} = \omega(\delta) \to_{\delta\to 0} 0.\end{aligned}$$

□

Corollaire A.10 (Théorème d'approximation de Weierstrass)**.** Soit K un compact de $\mathbb{R}^n$ les polynômes (à coefficients complexes) sont denses dans $C^0(K, \mathbb{C})$. En conséquence, $C^0(K, \mathbb{C})$ est séparable.

Démonstration. Comme K est fermé borné, $K \subset [-N, N]^n$ et par le théorème de Tietze D.3, f continue sur K se prolonge en une fonction continue sur $[-N, N]^n$, il suffit donc du cas $K = [-N, N]^n$ que l'on obtient par translation et dilatation (qui conservent les polynômes) du résultat précédent. Comme $\mathbb{Q}[i] := \mathbb{Q} + i\mathbb{Q}$ est dense dans $\mathbb{C}$, on voit facilement que les polynômes à coefficients dans $\mathbb{Q}[i]$ sont aussi denses, et forment un ensemble dénombrable, comme union dénombrable des polynomes de degré au plus m en chaque variable (c'est plus simple à décrire qu'en terme de degré total) qui s'écrivent sous la forme $\sum_{i_1,\ldots,i_n=0}^m \lambda_i x_1^{i_1} \ldots x_n^{i_n}$ et qui s'identifient donc au produit $\mathbb{Q}[i]^{m^n} \simeq \mathbb{Q}^{2m^n}$, qui est dénombrable comme produit fini d'ensembles dénombrables. □

Remarque A.1. Plus généralement, le théorème de Stone Weierstrass indique que toute sous-algèbre A (stable par conjugaison complexe) de $C^0(K, \mathbb{C})$ avec K compact qui contient les fonctions constantes et sépare les points (au sens pour $x \neq y$ il existe $P \in A$ avec $P(x) \neq P(y)$) est dense pour la norme uniforme :$\overline{A} = C^0(K, \mathbb{C})$.

6 Un résultat de compacité : le Théorème d'Ascoli (niveau L3 Math)

Les compacts sont difficiles à trouver en dimension infinie, et la moitié viennent (ou sont des variantes) du résultat suivant (l'autre moitié sont des conséquences du Théorème de Tychonov), que l'on va déduire de la relation entre complétude et compacité.

Remarque A.2. Soit (Y, d) un espace métrique borné, $d_y \in (C_b^0(Y, \mathbb{R}))$, $d_y(x) = d(y, x)$ la distance à y. $||d_y - d_z|| = \sup_{x \in Y} |d(y, x) - d(z, x)| = d(y, z)$ (car $\leq$ par l'inégalité triangulaire inverse et $\geq$ en prenant $x = y$ ou $x = z$) Donc $d : Y \to C_b^0(Y, \mathbb{R})$ est une isométrie.

Définition A.5. Soient X, Y des espaces métriques, une partie $F \subset C^0(X, Y)$ est **équicontinue** si pour tout $\epsilon > 0$, il existe $\delta = \delta(\epsilon) > 0$, tel que $\forall x, y \in X, \forall f \in F$, si $d(x, y) \leq \delta$ alors $d(f(x), f(y)) \leq \epsilon$.

Par exemple une famille d'application K-lipschitziennes (comme une famille de la boule unité fermé de rayon K des applications linéaires continues entre espaces de Banach) forme une famille équicontinue.

Théorème A.11 (d'Ascoli)**.** Soient X, Y des espaces métriques compacts, si une partie F est équicontinue alors $\overline{F}$ est compacte (pour la topologie de la convergence uniforme donnée par la distance $d(f, g) = \sup_{x \in X} d(f(x), g(x))$).

Exercice A.3. Montrer la réciproque facile.

Démonstration. Comme Y compact il est complet borné donc $d : Y \to C_b^0(Y, \mathbb{R})$ est une isométrie et $d(Y)$ est complet donc fermé. Elle induit une isométrie de $C^0(X, Y) \to C^0(X, C_b^0(Y, \mathbb{R}))$ qui est un espace de Banach. Les équations $f(x) \in d(Y), x \in X$ montrent que l'image de l'isométrie est fermé (comme intersection de fermés $\cap_{x \in X} ev_x^{-1}(d(Y))$,

$ev_x(f) = f(x))$ donc complet. Donc $C^0(X, Y)$ est aussi complet (on aurait aussi pu reprendre la preuve du cas Y Banach) et $\overline{F}$ aussi.

Il reste à voir que $\overline{F}$ est précompact. Or en recouvrant F par des boules de rayon $\epsilon/2$, $\overline{F}$ est recouvert par les boules de même centre et rayon ϵ, donc il suffit de voir F précompact. Soit $\epsilon > 0$, on fixe $\delta(\epsilon) > 0$ donné par l'équicontinuité et R les centres d'un recouvrement de X par des boules de rayons $\delta(\epsilon)$ donné par sa précompacité.

Remarquons que si $d(f(r), g(r)) \leq \epsilon$ pour tout $r \in R$, en prenant r avec $d(x, r) \leq \delta(\epsilon)$, on a par l'équicontinuité et l'inégalité triangulaire :

$$\begin{aligned}
& d(f(x), g(x)) \leq d(f(x), f(r)) + d(f(r), g(r)) + d(g(r), g(x)) \leq 3\epsilon \\
& \Rightarrow d(f, g) \leq 3\epsilon.
\end{aligned}$$

Soit enfin S les centres des boules de rayon $\epsilon/2$ recouvrant Y. Nous allons indicer les boules d'un 4ϵ recouvrement par les applications S^R de R vers S en nombre fini. Pour $\phi \in S^R$, soit

$$F_\phi = \{f \in F, \forall r \in R, d(\phi(r), f(r)) \leq \epsilon/2\}$$

Si $f, g \in F_\phi$ alors l'inégalité triangulaire donne, $d(g(r), f(r)) \leq \epsilon$ pour tout r donc $d(f, g) \leq 3\epsilon$ et si F_ϕ est non-vide il est inclus dans $B(b_\phi, 4\epsilon)$.

Enfin, il suffit donc de voir que $F \subset \cup_{\phi \in S^R} F_\phi$. Or chaque valeur possible de $f(r)$ est à distance inférieure à $\epsilon/2$ d'un $s = \phi(r) \in S$ pour un certain ϕ, ce qui conclut. □

Théorème A.12 (d'Ascoli)**.** Soient X un espace métrique compact et E un e.v.n. de dimension finie, si une partie F est équicontinue et bornée de $C^0(X, E)$ alors $\overline{F}$ est compacte (pour la topologie de la convergence uniforme donnée par la norme $||.||_\infty$).

Démonstration. Si $M = \sup\{||f||_\infty, f \in F\}$, $F \subset C^0(X, B_F(0, M))$ et $Y = B_F(0, M)$ est fermé borné donc compact comme E est de dimension finie. Le théorème précédent conclut. □

ANNEXE B
Compléments facultatifs et hors programme au chapitre 3 :Convexité

1 Propriétés des Cônes tangents et normaux dans $\mathbb{R}^n$

En pratique, on peut utiliser le résultat suivant pour se ramener à des cas plus simples :

Proposition B.1. Soient A, B des convexes de E.

1. Si $A \subset B$ alors pour tout $x \in A$, $T_A(x) \subset T_B(x)$ et $N_A(x) \supset N_B(x)$.
2. Si $a \in Int(A)$, $T_A(a) = E$ et $N_A(a) = \{0\}$.
3. Si $u_1, \dots, u_n \in N_A(x)$ alors $\{\sum_{i=1}^{n} \lambda_i u_i, \lambda_i \geq 0\} \subset N_A(x)$.
4. Si $a \neq b$ alors $N_{[a,b]}(a) = (\mathbb{R}(b-a))^{\perp} + \mathbb{R}_+(a-b)$
   et pour $u \in [a,b] - \{a,b\}$ $N_{[a,b]}(u) = (\mathbb{R}(b-a))^{\perp}$.
5. Pour $x \in A$, $A \subset x + T_A(x)$ et $T_{x+T_A(x)} = T_A(x)$ et donc $N_{x+T_A(x)} = N_A(x)$.

Démonstration. (1) $T_A(x) = \overline{\mathbb{R}_+^*(A-x)} \subset T_B(x)$ est par monotonie de l'adhérence. Si $f \in N_B(x)$ alors pour tout $y \in T_B(x)$ (en particulier $y \in T_A(x)$ on a $\langle f, x\rangle \leq 0$ et donc $f \in N_A(x)$. Donc on a l'inclusion $N_A(x) \supset N_B(x)$.

(2) $a \in Int(A)$ il existe une boule donc un convexe $B(a,r) \subset A$ $r > 0$ et donc par (1) $T_A(a) \supset T_{B(a,r)}(a) \supset \mathbb{R}_+(B(a,r) - a) = \mathbb{R}_+ B(0,r) = E$ par la définition. Vu $E^{\perp} = \{0\}$ le résultat sur le cône normal s'en déduit.

(3) C'est la propriété de cône. Par hypothèse pour $x \in T_A(x)$ on a $\langle u_i, x\rangle \leq 0$ donc pour $\lambda_i \geq 0$ $\langle \sum_{i=1}^{n} \lambda_i u_i, x\rangle = \sum_{i=1}^{n} \lambda_i \langle u_i, x\rangle \leq 0$ et donc $\sum_{i=1}^{n} \lambda_i u_i \in N_A(x)$.

(4) Comme $[a, b]$ est convexe, on obtient $T_u([a, b]) = \mathbb{R}_+[a - u, b - u]$ et $u = \lambda a + (1 - \lambda) b$ donc $(a - u) = (1 - \lambda)(a - b)$, $b - u = \lambda(b - a)$ donc $T_u([a, b]) = \mathbb{R}_+[a - u, b - u] = \mathbb{R}(b - a)$ d'où le calcul du cône normal par l'exo 3.2. De même $T_a([a, b]) = \mathbb{R}_+(b - a)$ donc clairement $f \in N_a([a, b])$ se décompose selon la somme directe orthogonale $\mathbb{R}(b - a) \oplus (\mathbb{R}(b - a))^\perp$ $f = \lambda(b - a) + v$ et on $\langle f, b - a \rangle = \lambda ||b - a||^2$ qui est négatif si et seulement si $\lambda \leq 0$. Donc si et seulement si $f \in (\mathbb{R}(b - a))^\perp + \mathbb{R}_+(a - b)$ comme annoncé.

(5) Par la formule $x + T_A(x) = x + \overline{\mathbb{R}_+^*(A - x)} \supset x + (A - x) = A$. Par l'inclusion $T_{x+T_A(x)} \supset T_A(x)$. Mais $x + T_A(x) - x = T_A(x)$ donc $T_{x+T_A(x)} = \overline{\mathbb{R}_+^* T_A(x)} = T_A(x)$ car $T_A(x)$ est un cône fermé. On déduit directement le cas des cônes normaux. □

Exemple B.1. Soit $A = \{(x, y) \in \mathbb{R}^2 : x \geq y \geq 0, \}$. Calculons $N_A(0)$ le cône normal en $0 = (0, 0)$.
D'abord on essaye de borner supérieurement l'ensemble. En prenant $[(0, 0), (1, 1)] \subset A$, on a

$$\begin{aligned} N_A(0) &\subset N_{[(0,0),(1,1)]}(0) \\ &= (\mathbb{R}(1, 1))^\perp + \mathbb{R}_+(-1, -1) \\ &= \{\lambda(1, -1) + \mu(-1, -1), \lambda \in \mathbb{R}, \mu \geq 0\} \end{aligned}$$

De même
Donc $N_A(0)$ est inclus dans l'intersection, résolvons le système $(-\mu', \lambda') = (\lambda - \mu, -\lambda - \mu)$ avec les conditions ci-dessus ,$\lambda, \lambda' \in \mathbb{R}, \mu, \mu' \geq 0$ Il faut donc $-\lambda - \mu = -\lambda + \mu - 2\mu = \mu' - 2\mu$ donc

$$N_{(0,0)}(A) \subset \{\mu'(-1, 1) + \mu(0, -2), \mu, \mu' \geq 0\}.$$

Montrons qu'il y a égalité en montrant que $(-1, 1) \in N_A(0)$ et $(0, -1) \in N_A(0)$ (car on a alors l'autre inclusion par le 3 de la précédente proposition).
La formule du cas convexe donne $T_A(0) = A$ donc soit $(x, y) \in A$, on calcule $\langle (x, y), (-1, 1) \rangle = y - x \leq 0$ d'après l'équation de A donc $(-1, 1) \in N_A(0)$.
Enfin $\langle (x, y), (0, -1) \rangle = -y \leq 0$ donc $(0, -1) \in N_A(0)$ comme voulu.
On a donc

$$N_A(0) = \mathbb{R}_+(-1, 1) + \mathbb{R}_+(0, -2).$$

On est maintenant prêt pour la :

Preuve du Théorème 3.6. On rappelle que

$$C = \{x \in U : \forall i \in \{1, \dots, n\}, g_i(x) \leq 0\}.$$

On a supposé $x_0 \in Int(C) \subset U$ existe. Soit $x \in C$ tel que :

1. les l premières contraintes sont actives, c'est à dire : $g_1(x) = \ldots = g_l(x) = 0$
2. les autres contraintes ne sont pas actives, c'est à dire $g_{l+1}(x) < 0, \ldots g_n(x) < 0$

Si $l = 0$, on a

$$x \in Int(C) = \{x \in U : \forall i \in \{1, \ldots, n\}, g_i(x) < 0\}$$

donc $N_C(x) = \{0\}$ par la proposition B.1.2. Sinon, le but est de voir :

$$N_C(x) = \left\{ \sum_{i=1}^{l} \lambda_i \nabla g_i(x), \lambda_i \geq 0 \right\}.$$

Etape 1 : inclusion $\supset$.

Par la proposition B.1.3. il suffit de voir que $\nabla g_i(x) \in N_C(x)$ pour $1 \leq i \leq l$, soit autrement dit par définition de $N_C(x)$, il faut voir :

$$\langle \nabla g_i(x), u - x \rangle \leq 0, \forall u \in C$$

Or par le théorème 3.12, on a $\forall u, x \in U$

$$\langle \nabla g_i(x), u - x \rangle \leq g_i(u) - g_i(x) = g_i(u) \leq 0,$$

car $u \in C$.

Etape 2 : inclusion $\subset$.

Soit $f \in N_C(x)$.

On remarque d'abord que si on prend $h_0 = x_0 - x$ on a
$dg_i(x)(h_0) \leq g_i(x_0) - g_i(x) = g_i(x_0) < 0$ pour tout $i = 1, \ldots, l$.

Soit donc maintenant h tel que $dg_i(x)(h) < 0$, $i = 1, \ldots, l$ (il en existe par la remarque), alors $g_i(x + th) - g_i(x) = tdg_i(x)(h) + o(t)$ donc $g_i(x + th) < 0$ pour $t > 0$ petit, et $i = 1, \ldots l$ De plus pour t assez petit comme $g_{l+1}(x) < 0, \ldots g_n(x) < 0$, on déduit par continuité $g_{l+1}(x + th) < 0, \ldots g_n(x + th) < 0$ d'où $x + th \in A$ pour tout t assez petit.

Par définition de $N_C(x)$, on a donc $\langle f, x + th - x \rangle \leq 0$ donc en particulier $\langle -f, h \rangle \geq 0$ et on ne peut pas avoir $-\langle f, h \rangle < 0$. Donc $-f$, $dg_1(x), \ldots, dg_l(x)$ vérifient la première condition de la Proposition B.15 (avec $E = \mathbb{R}^n$) donc aussi la seconde et sont donc positivement linéairement dépendants. On a donc des λ_i positifs non tous nuls tel que $-\lambda_0 f + \sum_{i=1}^{l} \lambda_i \nabla g_i(x) = 0$.

Montrons enfin que $\lambda_0 \neq 0$. Si on avait $\sum_{i=1}^{l} \lambda_i \nabla g_i(x) = 0$, il n'y aurait pas de h tel que $dg_i(x)(h) < 0$ pour tout $i = 1, \ldots, l$ ce qui contredit $dg_i(x)(h_0) < 0$.

On conclut à l'égalité voulu :

$$f = \sum_{i=1}^{l} \frac{\lambda_i}{\lambda_0} \nabla g_i(x) \in \left\{ \sum_{i=1}^{l} \lambda_i \nabla g_i(x), \lambda_i \geq 0 \right\}$$

□

2 Enveloppe convexe, cônes tangents et cônes normaux pour tout e.v.n. E (Niveau L3)

Comme pour les adhérences, la stabilité par intersection garantit l'existence d'un plus petit convexe contenant A.

Définition B.1. L'enveloppe convexe d'un ensemble A, notée $Conv(A)$ est le plus petit convexe contenant A.

Lemme B.2.
$$Conv(A) = \bigcup_{n \in \mathbb{N}^*} \{ \sum_{i=1}^{n} t_i x_i,\, x_i \in A,\, avec \sum_{i=1}^{n} t_i = 1,\, t_i \geq 0 \}$$

Démonstration. Soit $Conv'(A)$ le membre de droite.
$Conv'_n(A) = \{ \sum_{i=1}^{n} t_i x_i,\, x_i \in A,\, avec \sum t_i = 1,\, t_i \geq 0 \}$ Le cas $n = 1$ dans l'union est A donc $A \subset Conv'(A)$. Si $y_1 = \sum_{i=1}^{n} t_i x_i \in Conv'_n(A)$, $y_2 = \sum_{j=1}^{m} s_j z_j \in Conv'_m(A)$ sont deux points quelconques, alors pour $\lambda \in [0, 1]$

$$\lambda y_1 + (1 - \lambda) y_2 = \sum_{i=1}^{n} \lambda t_i x_i + \sum_{j=1}^{m} (1 - \lambda) s_j z_j.$$

Comme $\sum_{i=1}^{n} \lambda t_i + \sum_{j=1}^{m} (1 - \lambda) s_j = \lambda + (1 - \lambda)$ on déduit $\lambda y_1 + (1 - \lambda) y_2 \in Conv'_{n+m}(A)$. Ceci montre que $Conv'(A)$ est un convexe qui contient A.

Il est facile de voir que tout ensemble convexe est stable par combinaison convexe $\sum_{i=1}^{n} t_i x_i$ avec $\sum t_i = 1$, $t_i \geq 0$ par récurrence sur n et ainsi $Con'_n(A) \subset Conv(A)$. Si $t_n = 1$, les autres sont nuls et rien n'est à montrer. En écrivant $\sum_{i=1}^{n} t_i x_i = (1-t_n)(\frac{1}{1-t_n}\sum_{i=1}^{n} t_i x_i) + t_n x_n$ on a par l'hypothèse de récurrence $\frac{1}{1-t_n}\sum_{i=1}^{n} t_i x_i \in Conv(A)$ car $y_n := \frac{1}{1-t_n}\sum_{i=1}^{n} t_i = (1-t_n)/(1-t_n) = 1$ (et les coefficients sont positifs). Donc on a aussi la combinaison convexe $\sum_{i=1}^{n} t_i x_i = (1-t_n)y_n + t_n x_n \in Conv(A)$. □

Dans $\mathbb{R}^n$ il ne suffit que du barycentre de $n+1$ points.

Théorème B.3 (de Carathéodory)**.** (admis) Si $A \subset \mathbb{R}^n$, on a

$$Conv(A) = \{\sum_{i=1}^{n+1} t_i x_i,\, x_i \in A,\, avec \sum_{i=1}^{n+1} t_i = 1,\, t_i \geq 0\}.$$

Les deux ensembles suivant seront importants pour formuler des conditions pour des problèmes de minimisation sous contrainte.

Définition B.2. Le cône tangent de l'ensemble $A \subset E$ e.v.n. au point $a \in A$ est

$$T_A(a) := \{b \in E : \exists a_i \to a,\, a_i \in A,\, t_i > 0,\, t_i \to 0 : b = \lim \frac{a_i - a}{t_i}\}$$

Le cône normal est son polaire, c'est à dire le cône convexe fermé :

$$N_A(a) := \{f \in E^* : \forall x \in T_A(a),\, f(x) \leq 0\}.$$

Exemple B.2. $T_A(a)$ est toujours fermé. Si L est un s.e.v de E $a \in L$, $T_L(a) = \overline{L}$ et $N_L(a) = L^{\perp}$. Si $a \in Int(A)$, $T_A(a) = E$ et $\mathbb{N}_A(a) = \{0\}$.

Le résultat montrer l'accord avec la définition du cas $E = \mathbb{R}^n$ dans le cas convexe (avec l'identification usuelle de E' à E comme pour tout espace de Hilbert.)

Proposition B.4. Si S est convexe et $x \in S$, alors $T_x(S)$ est convexe et $S \subset x + T_x(S)$. De plus, on a

$$T_x(S) = \overline{\{\frac{u-x}{s},\, u \in S,\, s > 0\}},$$
$$N_x(S) = \{f \in E' : \forall u \in S,\, f(u-x) \leq 0\}$$

Démonstration. $\mathbb{R}_+^*(S-x)$ est convexe comme $S-x$ donc en prenant l'adhérence, aussi l'ensemble $W=\overline{\mathbb{R}_+^*(S-x)}$ que l'on veut montrer être $T_S(x)$. Si on a une suite $(x_n-x)/t_n \to u \in T_S(x)$ comme tous les éléments sont dan W, on obtient par fermeture aussi la limite, donc $T_S(x)\subset W$. Réciproquement, pour $t>0$, $x_n := \frac{t}{n}(u-x)+x=\frac{t}{n}u+(1-\frac{t}{n})x\in S$ pour n assez grand par convexité et $(x_n-x)/t_n = t(u-x)$ si $t_n = 1/n \to 0$ donc $t(u-x)\in T_S(x)$ comme voulu. Les autres relations sont alors évidentes, car $S-x\subset T_S(x)$ (car $s=1$) et par la définition de $N_S(x)$ comme polaire. □

3 Points selles (Niveau L2-L3)

Les points critiques a qui ne sont pas des extrema peuvent être de différents types. L'absence d'extrema peut être visible sur une droite passant par a s'il y a un point d'inflexion (comme pour $x\mapsto x^3$ dans $\mathbb{R}$) et il peut y avoir des points critiques qui sont des maxima dans certaines directions et des minima dans d'autres. Ces points ont un certain intérêt et seront nommés points selles.

Définition B.3. Soit $U\subset\mathbb{R}^n$ et $f:U\to\mathbb{R}$ et $a\in U$.

1. Soient deux sous-espaces vectoriels F et G supplémentaires $\mathbb{R}^n = F\oplus G$ (c'est à dire $F\cap G=\{0\}$ et $\mathbb{R}^n=F+G$) On dit que a est un **point selle (resp. point selle local)** de f selon la décomposition $\mathbb{R}^n=F\oplus G$ si a est un minimum (resp. minimum local) pour la restriction $f_{|a+F}$ de f au sous espace affine $a+F$, et si a est un maximum (resp. maximum local) pour la restriction $f_{|a+G}$ de f au sous espace affine $a+G$. On parle de point selle si il existe une telle décomposition.
2. Si f de classe $\mathcal{C}^1$. Soit a un point critique de f, un sous espace vectoriel $H\subset\mathbb{R}^n$ est un plan d'inflexion si pour toute droite Δ passant par a inclus dans $a+H$, $f_{|\Delta}$ n'a pas d'extrema local en a.

Remarque B.1. La décomposition $F\oplus G$ d'un point selle n'est pas forcément unique et on ne demande rien en dehors $(a+G)\cup(F+a)$, en particulier, il peut y avoir des plans d'inflexion en un point selle (ex $f(x,y)=x^2-y^2+(x-y)^3$, $(0,0)$ est un point selle local dans la décomposition $(\mathbb{R},0)\oplus(0,\mathbb{R})$ car x^2+x^3 a un minimum local en 0 et $-y^2-y^3$ un maximum local, de même $(0,0)$ est un point selle dans la décomposition $\mathbb{R}(1,1/2)\oplus\mathbb{R}(1/2,1)$ mais $\mathbb{R}(1,-1)$ est une droite d'inflexion)

Proposition B.5. Soit $f:U\to\mathbb{R}$ de classe $\mathcal{C}^1$

1. Si a est un point selle de f, c'est un point critique de f.

2. Si f est $\mathcal{C}^2$ et a est un point critique de f. Si $D^2 f(a)$ est non-dégénérée, ni positive ni négative, alors a est un point selle local de f.
3. Si a est un point critique de f H est un plan d'inflexion en a de dimension $dim(H) > n/2$ alors a n'est pas un point selle local. De plus si f est $\mathcal{C}^2$ pour tout $h \in H$, $D^2 f(a)(H,H) = 0$.

Démonstration. Pour (1) on remarque qu'il suffit de montrer $df(a) = 0$ ce qui ne dépend pas de la base de $\mathbb{R}^n$ on peut donc supposer a point selle pour la décomposition $F = \mathbb{R}^k \times \{0\}$, $G = \{0\} \times \mathbb{R}^{n-k}$. Comme f restreint à $a + F$ à un minimum local, les k premières dérivées partielles s'annulent, les n-k dernières s'annulent à cause du maximum sur $a + G$, d'où $df(a) = 0$.

La preuve de (2) nécessite quelques bases d'algèbre linéaire. Pour (2), comme $D^2 f(a)$ est non dégénérée, les valeurs propres de $H(f)(a)$ (les racines du polynôme $X \mapsto det(H(f)(a) - Xid))$ sont non nulles. Comme la matrice $D^2 f(a)$ n'est ni positive ni négative, il y a à la fois des valeurs propres λ positives et négatives. Soit F l'espace vectoriel engendré par les vecteurs propres u (les $u \in \mathbb{R}^n$ tels que $H(f)(a)u = \lambda u$ qui existent car si $det(H(f)(a) - \lambda id) = 0$, $H(f)(a) - \lambda id$ n'est pas injective donc a un noyau) des valeurs propres λ strictement positives, et de même G avec les négatives. $D^2 f(a)$ restreint à F est positive donc $f|_{a+F}$ admet un minimum local et de même pour G.

Pour (3), si $dim(H) > n/2$ et supposons par l'absurde a point selle, on a $dim(F) + dim(G) = n$, on a soit $dim(F) \geq n/2$, soit $dim(G) \geq n/2$, disons qu'on se trouve dans le premier cas, alors $n \geq dim(H+F) = dim(F) + dim(H) - dim(F \cap H)$ implique $dim(F \cap H) \geq dim(F) + dim(H) - n > n/2 + n/2 - n = 0$ donc $F \cap H \neq \{0\}$ une contradiction car la restriction de f à toute droite dans $a + F \cap H$ devrait avoir un minimum local en a et un point d'inflexion à la fois. Si $D^2 f(a)(H,H) \neq 0$, on a vu que cela suffit à ce que f ait un extremum local sur la droite $a + \mathbb{R}H$, vu si $\phi(\lambda) = f(a + \lambda H)$, $\phi''(0) = D^2 f(a)(H,H)$. □

Théorème B.6. Soient $A \subset \mathbb{R}^{n-k}$, $B \subset \mathbb{R}^k$ des compacts convexes et $K : C = A \times B \to \mathbb{R}$ continue. Si pour tout $(a,b) \in C$, $a \in \mathbb{R}^{n-k}$, $b \in \mathbb{R}^k$, $x \mapsto K(x,b)$ est convexe et $y \mapsto K(a,y)$ est concave, alors il existe un point de C qui soit un point selle (x_0, y_0) selon la décomposition $\mathbb{R}^{n-k} \times \{0\} \oplus \{0\} \times \mathbb{R}^k$ autrement dit :

$$\forall x \in A, y \in B \quad K(x_0, y) \leq K(x_0, y_0) \leq K(x, y_0). \tag{B.1}$$

De plus, (B.1) est équivalente à l'égalité :

$$\mathrm{Min}_{x \in A} \mathrm{Max}_{y \in B} K(x,y) = \mathrm{Max}_{y \in B} \mathrm{Min}_{x \in A} K(x,y). \tag{B.2}$$

Remarque B.2. On a des *Min* et *Max* au lieu d'inf et sup car des fonctions continues sur des compacts atteignent leurs bornes (cf. la preuve pour la continuité de $x \mapsto \mathrm{Max}_{y\in B} K(x,y)$ et de façon similaire de $y \mapsto \mathrm{Max}_{x\in A} K(x,y)$.
Dans le cas où f est bilinéaire, ce résultat s'appelle le théorème du min−max de von Neumann. Il a une signification en théorie des jeux. Si f donne la valeur que gagne un joueur A en position $x \in U$ si $f(x) \geq 0$ et $-f(x)$ la valeur que gagne le joueur B (et perd le joueur A) si $f(x) \leq 0$. Si A ne peut influencer que la direction $\{0\}\times\mathbb{R}^k$ et B seulement la direction $\mathbb{R}^{n-k}\times\{0\}$. Alors un point selle est un "équilibre de Nash" c'est−à−dire un point où ni A ni B n'ont intérêt à changer leur stratégie, car si A change sa stratégie celle de B étant constante, étant donné que le point selle est un maximum, A va perdre en gain, et de même si B change sa position avec celle de A constante, le caractère de minimum dans la direction du changement de B montre que B ne peut que perdre plus.

Démonstration.

- $\mathrm{Max}_{y\in B}\mathrm{Min}_{x\in A} K(x,y) \leq \mathrm{Min}_{x\in A}\mathrm{Max}_{y\in B} K(x,y)$ est toujours vrai.
  Comme pour tout $x \in A$, $y \in B$, $\mathrm{Min}_{x\in A} K(x,y) \leq K(x,y) \leq \mathrm{Max}_{y\in A} K(x,y)$, on déduit en prenant le max : $\mathrm{Max}_{y\in B}\mathrm{Min}_{x\in A} K(x,y) \leq \mathrm{Max}_{y\in B} K(x,y)$ soit en prenant un *Min* en x :

$$\mathrm{Max}_{y\in B}\mathrm{Min}_{x\in A} K(x,y) \leq \mathrm{Min}_{x\in A}\mathrm{Max}_{y\in B} K(x,y).$$

- $(B.1) \Rightarrow (B.2)$
  De plus, en considérant (x_0, y_0) de $(B.1)$, on a :

$$\begin{aligned} K(x_0,y_0) &\leq \mathrm{Min}_{x\in A} K(x,y_0) \\ &\leq \mathrm{Max}_{y\in B}\mathrm{Min}_{x\in A} K(x,y), \\ K(x_0,y_0) &\geq \mathrm{Max}_{y\in B} K(x_0,y) \\ &\geq \mathrm{Min}_{x\in A}\mathrm{Max}_{y\in B} K(x,y), \end{aligned}$$

  d'où l'égalité complète en rassemblant les 3 dernières inégalités.
- $g : x \mapsto \mathrm{Max}_{y\in B} K(x,y)$ est continue.
  Soit $x, x_n \in A$, $x_n \to x$, soit y_n (resp t) atteignant le *max* pour x_n (resp x) c'est à dire : $\mathrm{Max}_{y\in B} K(x_n,y) = K(x_n,y_n)$. Supposons que $g(x_n) = K(x_n,y_n)$ ne converge pas vers $g(x)$. Par compacité, on peut extraire une suite telle que $y_{\phi(n)} \to Y$. Par continuité de K :
  Or $K(x_{\phi(n)},t) \leq K(x_{\phi(n)},y_{\phi(n)})$ donc en passant à la limite par continuité de K, $K(x,t) \leq K(x,Y) < K(x,t)$, une contradiction.
- $(B.1) \Leftarrow (B.2)$ On prend $x_0 \in A$ réalisant le minimum c'est à dire tel que :

$$\alpha = \mathrm{Min}_{x\in A}\mathrm{Max}_{y\in B} K(x,y) = \mathrm{Max}_{y\in B} K(x_0,y)$$

Il existe par la continuité du point précédent et par compacité. De même, il existe $y_0 \in B$ réalisant le maximum :

$$\mathrm{Min}_{x\in A} K(x, y_0) = \mathrm{Max}_{y\in B}\mathrm{Min}_{x\in A} K(x, y) = \alpha.$$

Donc pour tout $x \in A$, $y \in B$, en utilisant (B.2) pour l'égalité du milieu, on obtient :

$$\begin{aligned} K(x_0, y) &\leq \mathrm{Max}_{Y\in B} K(x_0, Y) \\ &= \alpha = \mathrm{Min}_{X\in A} K(X, y_0) \leq K(x, y_0). \end{aligned}$$

En prenant $x = x_0$, $y = y_0$, on voit $\alpha = K(x_0, y_0)$, ce qui dit donc que (x_0, y_0) est un point selle.

- Montrons (B.2). Considérons, pour $\epsilon > 0$,

$$K_\epsilon(x, y) = K(x, y) + \epsilon||x||_2^2.$$

Comme $x \mapsto \epsilon||x||_2^2$ est strictement convexe, il en est de même de $K_\epsilon(., y)$ pour tout $y \in B$ (convexe plus strictement convexe donne strictement convexe).
Montrons que pour tout y, la fonction $K_\epsilon(., y)$ a un unique minimum. En effet, si $x_1 \neq x_2$ sont deux minima, par stricte convexité :
$K_\epsilon((x + y)/2, y) < K_\epsilon(x_1, y)/2 + K_\epsilon(x_2, y)/2 = K(x_i, y)$ en contradiction avec le caractère de minimum. Donc on a un unique $E(y)$ atteignant le minimum de $K_\epsilon(., y)$ Par le deuxième point (appliqué à $-K_\epsilon(y, x)$) $f_\epsilon(y) = K_\epsilon(E(y), y)$ est continue, donc atteint son maximum en y^*. En conséquence, par la définition de f_ϵ et le choix de y^*

$$f_\epsilon(y^*) = \mathrm{Max}_{y\in B}\mathrm{Min}_{x\in A} K_\epsilon(x, y) = K_\epsilon(E(y^*), y^*) = \mathrm{Min}_{x\in A} K_\epsilon(x, y^*).$$

Soit $x \in A$, $y \in B$, $t \in]0, 1[$, on a par concavité :

$$K_\epsilon(x, (1 - t)y^* + ty) \geq (1 - t)K_\epsilon(x, y^*) + tK_\epsilon(x, y) \geq (1 - t)f_\epsilon(y^*) + tK_\epsilon(x, y).$$

En prenant $x = E((1 - t)y^* + ty)$, on obtient
$f_\epsilon((1 - t)y^* + ty) \geq (1 - t)f_\epsilon(y^*) + tK_\epsilon(E((1 - t)y^* + ty), y)$.
Vu que y^* maximise f_ϵ, en soustrayant et divisant par t, on a :

$$f_\epsilon(y^*) \geq K_\epsilon(E((1 - t)y^* + ty), y) \quad (*).$$

On veut prendre $t \to 0$, voyons que $y \mapsto E(y)$ est continue. Supposons $y_n \to y$, et supposons $E(y_n) \not\to E(y)$ par compacité, on a une suite extraite $y_{\phi(n)}$ telle que $E(y_{\phi(n)}) \to Z \neq E(y)$. Par continuité $K_\epsilon(E(y_{\phi(n)})), y_{\phi(n)}) \to K_\epsilon(Z, y) > K_\epsilon(E(y), y)$, l'inégalité stricte venant de l'unicité du minimum d'une fonction strictement convexe. Or par définition $K_\epsilon(E(y)), y_{\phi(n)}) \geq K_\epsilon(E(y_{\phi(n)})), y_{\phi(n)})$ donc en passant à la limite $K_\epsilon(E(y), y) \geq K_\epsilon(Z, y) > K_\epsilon(E(y), y)$, une contradiction.

On a donc montré la continuité de $y \mapsto E(y)$.
Donc en passant à la limite dans l'inégalité $(*)$, on obtient : $f_\epsilon(y^*) \geq K_\epsilon(E(y^*), y)$ et ce pour tout $y \in B$ Par ailleurs par définition de f_ϵ, $f_\epsilon(y^*) \leq K_\epsilon(x, y^*)$. Autrement dit $(E(y^*), y^*)$ est un point selle de K_ϵ. Par l'implication $(B.1) \Rightarrow (B.2)$, on déduit, vu $K(x, y) \leq K_\epsilon(x, y) \leq K(x, y) + \epsilon D$ (avec $D = Max_{x\in A}||x||_2^2 < \infty$ par compacité) :

$$\begin{aligned}\text{Min}_{x\in A}\text{Max}_{y\in B}K(x, y) &\leq \text{Min}_{x\in A}\text{Max}_{y\in B}K_\epsilon(x, y)\\ &= \text{Max}_{y\in B}\text{Min}_{x\in A}K_\epsilon(x, y)\\ &\leq \epsilon C + \text{Max}_{y\in B}\text{Min}_{x\in A}K(x, y).\end{aligned}$$

En prenant $\epsilon \to 0$, on obtient l'inégalité qui manque pour avoir $(B.2)$ pour K.

□

4 Jauge de Minkowski d'un ensemble convexe (Niveau M1)

L'un des objectifs principaux de ce chapitre est d'utiliser le théorème de Hahn–Banach pour séparer des convexes par des hyperplans fermés, lieu d'annulation d'une forme linéaire continue. Pour cela, nous devons associer à un convexe une fonction (qui sera souvent une semi–norme) et que l'on pourra utiliser comme domination dans le théorème d'Hahn–Banach.

Définition B.4. Soit E un $\mathbb{R}$–e.v., un convexe $C \subset E$ est dit **absorbant** si pour tout $x \in E$, $x \in \lambda C$ pour un $\lambda > 0$.

Définition B.5. Soit E un $\mathbb{R}$–e.v. et C un convexe absorbant. La **jauge de Minkowski** de C est la fonction :

$$\mu_C(x) := \inf\{\lambda > 0 : \lambda^{-1}x \in C\} \in [0, \infty)$$

Théorème B.7. Soit E un $\mathbb{R}$–e.v. et C un convexe absorbant. Alors

1. $\mu_C(x + y) \leq \mu_C(x) + \mu_C(y)$.
2. $\mu_C(tx) = t\mu_C(x)$ si $t \geq 0$.
3. Si $-C = C$, μ_C est une seminorme.
4. Si $A = \{x : \mu_C(x) < 1\}$, $B = \{x : \mu_C(x) \leq 1\}$ alors $A \subset C \subset B$ sont des convexes et $\mu_A = \mu_B = \mu_C$

5. Si E est un e.v.n. et $0 \in Int(C)$ (ce qui implique C absorbant), μ_C est continue et de plus
$$A = Int(C),\ B = \overline{C}.$$

Démonstration. Soit $t = \mu_C(x) + \epsilon > 0$, $s = \mu_C(y) + \epsilon > 0$ de sorte que $x/t, y/s \in C$. Or on peut écrire la combinaison convexe suivante $\frac{x+y}{s+t} = \frac{t}{s+t}\frac{x}{t} + \frac{s}{s+t}\frac{y}{s} \in C$ et donc $\mu_C(x+y) \leq s+t$. Comme $\epsilon > 0$ est arbitraire, on déduit (1).

(2) est une conséquence directe de la définition. Si $-C = C$ $\mu_C(x) = \mu_C(-x)$ d'où on déduit $\mu_C(tx) = |t|\mu_C(x)$, la seule relation manquante pour (3).

Les inclusions entre A, B, C viennent de la définition : $x \in C$ donne $x/1 \in C$ et donc $\mu_C(x) \leq 1$ et si $\mu_C(x) < 1$, alors $x/1 \in C$. Elles impliquent $\mu_B \leq \mu_C \leq \mu_A$. Si $\mu_B(x) < s < t$ alors $x/s \in B$ donc $\mu_C(x/s) \leq 1$ donc $\mu_C(x/t) \leq s/t < 1$ d'où $x/t \in A$ donc $\mu_A(x) \leq t$ soit en passant à l'infimum des t, $\mu_A(x) \leq mu_B(x)$ ce qui donne la dernière égalité de (4). A, B convexes sont semblables à la convexité des boules en utilisant (1) et (2).

Pour (5), on remarque qu'il existe $B(0,\epsilon) \subset C$ donc $\mu_C(\epsilon x/||x||) \leq 1$ soit $\mu_C(x) \leq ||x||/\epsilon$.

De plus par l'inégalité triangulaire $\mu_C(x) \leq |\mu_C(x-y)| + \mu_C(y)$ et de même en inversant x, y donc
$$|\mu_C(x) - \mu_C(y)| \leq |\mu_C(x-y)| \leq ||x-y||/\epsilon$$
donc μ_C est $1/\epsilon$−lipschitzienne donc continue. On déduit que A est ouvert, B fermé et donc $A \subset Int(C), \overline{C} \subset B$. Or, soit ϵ, si $x \in B$ $x(1-1/n) \in C$ et converge vers $x \in \overline{C}$ donc $B \subset \overline{C}$. De même si $x \in A^c$, $(1+\epsilon)x \notin C$ donc $x \in \overline{C^c}$ donc $A^c \subset \overline{C^c}$ d'où en prenant le complémentaire $Int(C) \subset A$. □

Vous pouvez aussi en exercice essayer de montrer le résultat suivant directement.

Corollaire B.8. Soit C un convexe d'intérieur non vide d'un e.v.n., $Int(\overline{C}) = Int(C)$ et $\overline{Int(C)} = \overline{C}$.

Démonstration. En translatant, on peut supposer $0 \in Int(C)$,, Alors comme $\mu_C = \mu_{Int(C)} = \mu_{\overline{C}}$, par le (5) ci-dessus, le calcul de l'intérieur/adhérence en terme de la jauge donne que ces trois ensembles ont même intérieur et même adhérence. □

5 Séparation des convexes (Niveau M1)

Un élément $f \in E'$ tel que $f \neq 0$ permet de construire un hyperplan fermé (translation de $Ker(\phi)$, voir lemma 2.30) : $\{x \in E, f(x) = c\}$. Les deux ensembles $\{x \in E, f(x) \leq c\}$ et $\{x \in E, f(x) \geq c\}$ sont des demi-espaces. On dit que deux ensembles sont séparés (par l'hyperplan) si chaque ensemble est dans un des demi-espaces. On parle de séparation stricte si $C_1 \subset \{x \in E, f(x) < c\}$ et $C_2 \subset \{x \in E, f(x) > d\}$ pour $d > c$.

On va obtenir un résultat de séparation en utilisant un résultat abstrait de prolongement :

Théorème B.9 (de prolongement de Hahn-Banach) (admis)**.** Soient E un espace vectoriel, $p : E \to \mathbb{R}$ une application positivement homogène et sous-additive, c'est-a-dire vérifiant :

- $p(tx) = tp(x) x \in E, t > 0$
- $p(x+y) \leq p(x) + p(y), x, y \in E$.

Soient $G \subset E$ un sous-espace vectoriel et $g : G \to \mathbb{R}$ une application linéaire dominée par p :

$$\forall x \in G, g(x) \leq p(x).$$

Alors il existe une forme linéaire f sur E qui prolonge g (c'est-à-dire $\forall x \in G, g(x) = f(x)$) et encore dominée par p, c'est-à-dire telle que

$$\forall x \in E, f(x) \leq p(x).$$

La version suivante du théorème de Hahn-Banach permet de séparer des ensembles convexes bien choisis.

Théorème B.10 (de séparation de Hahn-Banach)**.** Soient A, B deux convexes non-vides disjoints d'un e.v.n. E, ils sont séparés par un hyperplan dans les deux cas suivants :

1. Si A est ouvert, alors il existe $f \in E'$ et $c \in \mathbb{R}$ telle que
$$\forall x \in A, y \in B : f(x) < c \leq f(y).$$
2. Si A est compact et B est fermé, alors il existe $f \in E'$ et $c < d \in \mathbb{R}$ telle que
$$\forall x \in A, y \in B : f(x) < c < d < f(y).$$

Démonstration. 1) **Premier cas :** $B = \{x_0\}$.

On peut supposer que $0 \in A$ pour utiliser la fonctionnelle μ_A comme fonctionnelle sous-additive et positivement homogène p du théorème de Hahn Banach. Soit $G = \mathbb{R}x_0$ et $g(tx_0) = t$.

On remarque que $\mu_A(x_0) \geq 1$ car $A = Int(A) = \{x : \mu_A(x) < 1\}$ par le théorème B.7 et $x_0 \notin A$.

donc pour t>0 $g(tx_0) = t \leq t\mu_A(x_0) = \mu_A(tx_0)$ et pour $t \leq 0$ $g(tx_0) \leq 0 \leq \mu_A(tx_0)$. Donc on obtient la domination hypothèse de Hahn-Banach :

$$\forall x \in G, g(x) \leq \mu_A(x).$$

En appliquant le théorème, on obtient donc f linéaire étendant g et telle que (en réutilisant la lipshitzianité obtenue dans la preuve du théorème B.7 (5))

$$\forall x \in E, f(x) \leq \mu_A(x) \leq M||x||.$$

Ceci implique en particulier $f \in E^*$, $f(x) < 1$ pour $x \in A$ et $f(x) = 1$ sur B. Ce qui donne la séparation.

Second cas : B quelconque.

On pose $C = A - B$ qui est convexe, ouvert (comme union $\cup_{y \in B} A - y$) et $0 \notin C$. Donc d'après le premier cas il existe $f \in E'$ telle que $f(z) < 0$ pour $z = a - b \in A - B$ soit $f(a) < f(b)$ pour $a \in A$, $b \in B$. En passant au sup on obtient :

$$Sup_{x \in A} f(x) \leq Inf_{y \in B} f(y) := c.$$

De plus, comme A ouvert on obtient $A \subset Int(\{x : f(x) \leq c\}) = \{x : f(x) < c\}$.

2)Vérifions qu'il existe $\epsilon > 0$ tel que $A + B(0, \epsilon)$ et $B + B(0, \epsilon)$ soient disjoints (ce sont aussi des convexes ouverts comme au 1). Sinon, on trouve $x_n \in A + B(0, 1/n) \cap B + B(0, 1/n)$ donc $y_n \in A$, $z_n \in B$ avec $||y_n - x_n||, ||z_n - x_n|| \leq 1/n$. En extrayant par compacité une sous-suite $y_{n_k} \to y \in A$ on obtient $z_{n_k} \to y \in B$, une contradiction.

Donc on peut appliquer le cas 1) à $A + B(0, \epsilon)$ et $B + B(0, \epsilon)$. On obtient $f \in E'$ non-nulle telle que :

$$\forall a \in A, \forall z \in B(0, \epsilon), \forall b \in B : \quad f(a) + f(z) \leq \alpha \leq f(b) + f(z)$$

En prenant des sup sur la boule unité :

$$\forall a \in A, \forall b \in B : \quad f(a) + ||f||\epsilon \leq \alpha \leq f(b) - ||f||\epsilon.$$

Comme $||f|| \neq 0$, il suffit de prendre $c = \alpha - ||f||\epsilon/2 < d = \alpha + ||f||\epsilon/2$. □

Applications

Il vient de l'application directe au cas $A = \{x\}$, $B = \{y\}$ qui sont des compacts.

Proposition B.11 (separation des points). E' sépare les points de E : Pour $x \neq y \in E$ il existe $f \in E'$ telle que $f(x) \neq f(y)$.

Le deuxième cas particulier permet de séparer un point et un espace fermé $\overline{F}$

Proposition B.12. Si $F \subset E$ un sous-espace vectoriel de l'e.v.n. E. Si $x \notin \overline{F}$ alors il existe $f \in E'$ telle que $f(x) = 1$ et $F \subset Ker(f)$.
En particulier, $F^{\perp} = 0$ ssi F est dense dans E.

La proposition précédente a des conséquences intéressantes pour comprendre l'injectivité et la surjectivité (ou plutôt la densité de l'image) des applications linéaires en dimension infinie.

On commence par un préliminaire algébrique sur l'orthogonalité dans les espaces de Banach.

Définition B.6. Soit E un e.v.n. et F un sous-espace de E et N un sous-espace de E'. Les orthogonaux de F et N sont les sous-espaces fermés :

$$F^{\perp} := \{f \in E', f(x) = 0 x \in F\},$$

$$^{\perp}N := \{x \in E, f(x) = 0 \forall f \in N\}.$$

Proposition B.13. Soient X, Y des e.v.n et $T \in L(X, Y)$. Alors

$$Ker(T^t) = [Im(T)]^{\perp} \quad Ker(T) = {}^{\perp}[Im(T^t)].$$

Démonstration. En effet, $y \in Ker(T^t)$ ssi pour tout $x \in E$, $0 = [T^t(y)](x) = y(T(x))$ ssi $y \in [Im(T)]^{\perp}$.

De même, $y \in Ker(T)$ ssi pour tout $x \in E^*$, $0 = x[T(y)] = [T^t(x)(y)$ ssi $y \in {}^{\perp}[Im(T^t)]$. □

Proposition B.14. Soient X, Y des e.v.n et $T \in L(X, Y)$.

1. $Im(T)$ est dense dans Y si et seulement si T^t est injectif.
2. Si $X \subset Y$, ${}^{\perp}(X^{\perp}) = \overline{X}$ est la fermeture normique de X dans Y.

Démonstration. Pour 1, T^t est injectif si et seulement si $Im(T)^{\perp} = Ker(T^t) = 0$ (proposition B.13) ssi $Im(T)$ est dense par la proposition précédente.

Pour 2, $X \subset {}^{\perp}(X^{\perp})$ donc comme le terme de droite est fermé, l'adhérence est inclus. Réciproquement, soit $x \notin \overline{X}$ par la conséquence de Hahn–Banach ci–dessus, soit $f \in E'$ telle que $f(x) = 1$, et $f \in X^{\perp}$, on déduit que $x \notin {}^{\perp}(X^{\perp})$. □

Le résultat suivant qu'on a utilisé pour les calculs de cônes normaux est un exercice typique d'application de Hahn–Banach.

Proposition B.15. Soit $\{f_i : i = 1, 2, \cdots, k\}$ un ensemble fini dans E' (pour un e.v.n. E). Les affirmations suivantes sont équivalentes :

(1) Il n'y a aucun $v \in E$ tel que $f_i(v) < 0$ pour tout $i \in [1, n]$;

(2) L'ensemble $\{f_i : i = 1, 2, \ldots, k\}$ est positivement linéairement dépendant : il existe un vecteur non–nul $\lambda = (\lambda_1, \cdots, \lambda_k) \neq 0$ avec $\lambda_k \geq 0$ tel que $\sum_{i=1}^{k} \lambda_i f_i = 0$.

Démonstration. Montrons premièrement le sens facile : $(2) \Rightarrow (1)$. A partir de $\lambda_i > 0$, on obtient en appliquant à v, $\sum_{i=1}^{k} \lambda_i f_i(v) = 0$, Or $f_i(v) \leq 0$ pour tout i implique $\sum_{i=1}^{k} \lambda_i f_i(v) < 0$, donc cela implique (1) par contraposée.

Dans l'autre sens $(1) \Rightarrow (2)$, on utilise le théorème de séparation de Hahn–Banach pour

$$K_1 = \{y \in \mathbb{R}^k : y_i < 0, \forall i \in \{1, 2, \ldots, k\}\},$$
$$K_2 = \{(f_1(v), f_2(v), \ldots, f_k(v)) : v \in E\}.$$

Vu que $p_i(y) = y_i$ est linéaire sur $\mathbb{R}^k$ de dimension finie, donc convexe continue, on obtient que $K_1 = \cap_{i=1}^{k} p_i^{-1}(]-\infty, 0[)$ est une intersection finie de convexes ouverts, donc un convexe ouvert.

$K_2 = Im(f_1, \cdots, f_k) \ni 0$ est un s.e.v de $\mathbb{R}^k$, donc un convexe non–vide. (1) indique qu'ils sont disjoints. Par conséquent le cas 1 du théorème B.10 s'applique et donne

$\lambda = (\lambda_1, \cdots, \lambda_k) \in E' = \mathbb{R}^k$ et c tels que :

$$\forall x \in K_1, y \in K_2, \qquad \langle \lambda, x \rangle < c \leq \langle \lambda, y \rangle$$

.

Comme K_2 est un s.e.v., pour $t \to 0$ on a $c \leq t\langle \lambda, y \rangle \to 0$, donc $c \leq 0$. De plus $c \leq \pm n\langle \lambda, y \rangle$ et donc $\pm n\langle \lambda, y \rangle \leq -c = |c|$ force $|\langle \lambda, y \rangle| \leq \frac{|c|}{n} \to 0$ donc $\langle \lambda, y \rangle = 0$.

De plus $(-\frac{1}{n}, \cdots, -1, \cdots, -\frac{1}{n}) \in K_1$ so $-\lambda_i - \frac{1}{n}\sum_{j \neq i} \lambda_j < c \leq 0$. Donc en passant à la limite, $n \to \infty$, on obtient $-\lambda_i \leq 0$, donc $\lambda_i \geq 0$. Et $\lambda \neq 0$ vient de $\langle \lambda, (1, \cdots, 1) \rangle < 0$.

□

ANNEXE C
Compléments facultatifs au chapitre 4 : Espaces mesurés.

1 Lemme de classe monotone

Définitions

Au lemme 4.3 iii), on a vu comment on remplace les unions dénombrables par des unions croissantes d'une suite d'unions finies. Cela suggère que la notion d'union croissante pourrait remplacer utilement (pour la théorie) celle d'union dénombrable et suggère la définition suivante de classe monotone. [1]

Définition C.1. Une **classe monotone** sur Ω est une famille $\mathcal{M}$ de partie de Ω contenant Ω et stable par différence et unions croissantes, c'est-à-dire $\mathcal{M} \subset \mathcal{P}(\Omega)$ telle que :

1. $\Omega \in \mathcal{M}$
2. Si $A, B \in \mathcal{M}$ avec $B \subset A$ alors $A - B \in \mathcal{M}$.
3. Si $\{A_n, n \geq 0\} \subset \mathcal{M}$ est une suite croissante (i.e. $A_n \subset A_{n+1}$, alors $\bigcup_{n \geq 0} A_n \in \mathcal{M}$.

Lemme C.1 (cf. TD).
1. Une tribu est une classe monotone.
2. Une classe monotone stable par intersection finie est une tribu.
3. Si $(\mathcal{M}_i)_{i \in I}$ sont des classes monotones, alors leur intersection $\bigcap_{i \in I} \mathcal{M}_i$ est une classe monotone.

On peut donc parler de la plus petite classe monotone contenant une famille $\mathcal{A} \subset \mathcal{P}(\Omega)$, qui est l'intersection de toutes les classes contenant $\mathcal{A}$, elle est notée $\mathcal{M}(\mathcal{A})$ et appelée la **classe monotone engendrée par** $\mathcal{A}$.

1. Comme dans le livre de Barbe-Ledoux, on suit la tradition française pour cette définition (différente de la tradition anglo-saxone venant du livre de Durett de Probabilités). Attention, ce n'est pas la même définition dans le cours du MGA.

Le résultat principal

Théorème C.2 (Lemme de classe monotone)**.** Soit $\mathcal{E}$ une famille de partie de Ω stable par intersection finie, alors la classe monotone et la tribu engendrée par $\mathcal{E}$ coïncident : $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.

Démonstration. Par le lemme C.1 1), $\sigma(\mathcal{E})$ est une classe monotone contenant $\mathcal{E}$, donc comme $\mathcal{M}(\mathcal{E})$ est la plus petite telle classe, on a $\mathcal{M}(\mathcal{E}) \subset \sigma(\mathcal{E})$.

$\mathcal{M}(\mathcal{E})$ **est une tribu.** Par le lemme C.1 2), il suffit de voir que $\mathcal{M} := \mathcal{M}(\mathcal{E})$ est stable par intersection binaire. On pose

$$\mathcal{K} = \{A \in \mathcal{M} : \forall B \in \mathcal{E}, A \cap B \in \mathcal{M}\}.$$

Comme E est stable par intersection finie, $\mathcal{E} \subset \mathcal{K}$. On a $\Omega \in \mathcal{M}$ et si $A \subset C$ avec $A, C \in \mathcal{K}$, $B \in \mathcal{E}$, alors $(C - A) \cap B = (C \cap B) - (A \cap B) \in \mathcal{M}$ par différence d'ensembles de $\mathcal{M}$. Enfin, de même comme intersection distribue sur les unions croissantes, $\mathcal{K}$ est stable par intersection croissante et donc une classe monotone. Or elle contient $\mathcal{E}$, comme ona vu, donc $\mathcal{M}(\mathcal{E}) \subset \mathcal{K}$ et comme par définition $\mathcal{K} \subset \mathcal{M}(\mathcal{E})$. on a égalité.

On est maintenant prêt à définir la classe qui va vérifier la stabilité voulue par intersection :

$$\mathcal{L} = \{A \in \mathcal{M} : \forall C \in \mathcal{M}, A \cap C \in \mathcal{M}\}.$$

On montre comme avant que $\mathcal{L}$ est une classe monotone (exo). Montrons que $\mathcal{E} \subset \mathcal{L}$. Soit donc $B \in \mathcal{E}$, alors $C \in \mathcal{M} \subset \mathcal{K}$ donc, par définition de $\mathcal{K}$, pour $B \cap C \in \mathcal{M}$. Et comme c'est vrai pour tout $C \in \mathcal{M}$, on en déduit par définition de $\mathcal{L}$ que $B \in \mathcal{L}$, comme voulu.

Finalement, $\mathcal{L}$ est une classe monotone telle que $\mathcal{E} \subset \mathcal{L} \subset \mathcal{M}(\mathcal{E})$ donc, par définition de la classe monotone engendrée, $\mathcal{L} = \mathcal{M}(\mathcal{E})$.

Inclusion réciproque. Comme $\mathcal{M}(\mathcal{E})$ est une tribu contenant $\mathcal{E}$ et que $\sigma(\mathcal{E})$ est la plus petite telle tribu, on obtient $\mathcal{M}(\mathcal{E}) \supset \sigma(\mathcal{E})$. □

Corollaire C.3 (au lemme de classe monotone)**.** Soient μ et ν des mesures finies de mêmes masses (i.e. $\mu(\Omega) = \nu(\Omega)$) sur un espace mesurable $(\Omega, \mathcal{T})$. Soit $\mathcal{E}$ une famille stable par intersection finie qui engendre $\mathcal{T}$. Si μ et ν coïncident sur $\mathcal{E}$ (i.e. $\mu(E) = \nu(E), \forall E \in \mathcal{E}$) alors μ et ν sont égales (i.e. $\mu(B) = \nu(B), \forall B \in \mathcal{T}$).

Démonstration. Soit $\mathcal{M} = \{B \subset E : \mu(B) = \nu(B)\}$. Par l'hypothèse, $\mathcal{M}$ contient $\mathcal{E}$. Vérifions que c'est une classe monotone :

- $\Omega \in \mathcal{M}$ car μ et ν ont même masse.
- Si $A, B \in \mathcal{M}$, $A \subset B$, alors par la proposition 4.3 v) on a $\mu(B - A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B - A)$.
- Si $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$, $A_n \in \mathcal{M}$, est une suite croissante, par la proposition 4.3 iii),

$$\mu\Big(\bigcup_{n\geq 1} A_n\Big) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \nu(A_n) = \nu\Big(\bigcup_{n\geq 1} A_n\Big).$$

Bilan, $\mathcal{M}$ est une classe monotone qui contient $\mathcal{E}$, donc $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}$. Or par le lemme de classe monotone $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{T}$ d'où le résultat. □

Preuve du corollaire 4.19 au lemme de classe monotone sur l'unicité des mesures sigma-finie

On commence par le cas où la suite de parties $A_n \in \mathcal{E}$ est croissante.

Notons μ_n, ν_n les mesures induites par μ, ν sur A_n respectivement. On a deux mesures finies avec $\mu_n(E) = \mu(E \cap A_n) = \nu(E \cap A_n) = \nu_n(E)$ pour tout $E \in \mathcal{E}$ donc par le corollaire au lemme de classe monotone pour les mesures finies, on déduit $\mu_n = \nu_n$. Pour tout $B \in \mathcal{T}$, on a $B = B \cap (\bigcup_n A_n) = \bigcup_n (B \cap A_n)$ donc par union croissante :

$$\mu(B) = \lim_{n\to\infty} \mu_n(B) = \lim_{n\to\infty} \nu_n(B) = \nu(B).$$

Dans le cas où la suite A_n n'est pas croissante, on utilise $B_n = \cup_{i=1}^n A_i$ qui est une suite croissante, mais pas forcément dans $\mathcal{E}$, donc il faut travailler plus pour vérifier l'hypothèse pour la mesure induite sur B_n. D'abord, par la formule de Poincaré :

$$\mu(\cup_{k=1}^n A_k) = \sum_{k=1}^n (-1)^{k-1} \sum_{1\leq i_1<\cdots<i_k\leq n} \mu(A_{i_1} \cap \cdots \cap A_{i_k}) < +\infty.$$

Et comme toutes les intersections sont dans $\mathcal{E}$ tous les termes de la formule sont égaux aux termes correspondants pour ν donc $\mu(B_n) = \nu(B_n)$. On considère les mesures induites pour $B \in \mathcal{T}(E)$,$\mu_n(B) = \mu(B \cap B_n)$, $\nu(B \cap B_n) = \nu_n(B)$. On vient de voir que μ_n, ν_n sont finies. Montrons que pour $E \in \mathcal{E}$ $\mu_n(E) = \nu_n(E)$ En effet $E \cap B_n = \cup_{k=1}^n (E \cap A_k)$ et en appliquant la formule de Poincaré encore (en remarquant que les intersections sont celles d'éléments de E.

$$\begin{aligned}\mu(\cup_{k=1}^n (E \cap A_k)) &= \sum_{k=1}^n (-1)^{k-1} \sum_{1\leq i_1<\cdots<i_k\leq n} \mu(E \cap A_{i_1} \cap \cdots \cap A_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k-1} \sum_{1\leq i_1<\cdots<i_k\leq n} \nu(E \cap A_{i_1} \cap \cdots \cap A_{i_k}) = \nu(\cup_{k=1}^n (E \cap A_k)).\end{aligned}$$

On conclut comme avant du corollaire au lemme de classe monotone pour les mesures finies, que $\mu_n = \nu_n$. Puis pour tout $B \in \mathcal{T}$, on a $B = B \cap (\bigcup_n B_n) = \bigcup_n (B \cap B_n)$ donc par union croissante :

$$\mu(B) = \lim_{n\to\infty} \mu_n(B) = \lim_{n\to\infty} \nu_n(B) = \nu(B).$$

2 Compléments sur les Boréliens

On rappelle que la tribu des boréliens d'un espace métrique (X, d) est la tribu engendrée par l'ensemble des ouverts $\mathcal{T}$. (cf. définition 4.6). En pratique, il est difficile de décrire tous les boréliens (les ouverts sont déjà difficiles à décrire), mais on n'a pas besoin de description explicite (juste de familles génératrices simples, et stables par intersections finies).

Remarque C.1. Il existe des ensembles qui ne sont pas boréliens sur $\mathbb{R}$, et donc des fonctions non-boréliennes. Ils ne sont pas si faciles à définir, donc en pratique, tous les ensembles qu'on rencontrera seront boréliens.

Espaces métriques séparables et leurs boréliens

Définition C.2. Une partie A est dite **dense** dans E si $\bar{A} = E$. Un ensemble est dit **séparable** si il admet un sous-ensemble au plus dénombrable dense (ou autrement dit une suite dense).

Lemme C.4. Un sous-ensemble F d'un espace métrique séparable est séparable.

Démonstration. On peut supposer F non-vide, sinon, c'est évident (la partie vide donc finie est dense). On fixe donc $x_0 \in F$

Soit u_n une suite dénombrable dense. Soit $a_{m,n} \in B(u_m, 1/n) \cap F$ si cet ensemble est non-vide, et sinon on pose $a_{m,n} = x_0$. La famille $\{a_{m,n}, m, n \in \mathbb{N}\}$ est finie ou dénombrable et dense car si $x \in F$ il existe $d(u_m, x) < 1/2n$ donc $a_{m,2n}$ existe car $B(u_m, 1/2n) \cap F$ est non vide et par inégalité triangulaire $d(u_m, a_{m,2n}) < 1/n$. □

Proposition C.5. $(\mathbb{R}^n, ||.||_\infty)$ est séparable.

Démonstration. On a vu que $\mathbb{Q}^n$ est dénombrable comme produit d'ensembles dénombrables. Montrons qu'il est dense dans $\mathbb{R}^n$. En effet si $x = (x_1, \ldots, x_n)$ on pose $x_p = (\frac{\lfloor px_1 \rfloor}{p}, \ldots, \frac{\lfloor px_n \rfloor}{p})$ avec $\lfloor x \rfloor$ la partie entière de x. Donc $\lfloor px_i \rfloor \leq px_i \leq \lfloor px_i \rfloor + 1$ et

$$\left|\frac{\lfloor px_i \rfloor}{p} - x_i\right| \leq \frac{1}{p}$$

donc $||x_p - x||_\infty \leq 1/p \to_{p\to\infty} 0$. Donc vu $x_p \in \mathbb{Q}^n$, $x \in \overline{\mathbb{Q}^n}$. Comme x est arbitraire. $\mathbb{R}^n \subset \overline{\mathbb{Q}^n}$ CQFD. □

Exercice C.1. Montrer que $\mathbb{Q}^c$ est dense dans $\mathbb{R}$.

Proposition C.6. Soit (X, d) un espace métrique séparable, alors la tribu borélienne est engendrée par les boules ouvertes $\mathcal{B}(X) = \sigma\Big(B : \text{Boule ouverte de X}\Big)$.

Démonstration. Toute boule ouverte est un ouvert donc $\{B : \text{Boule ouverte de X}\} \subset \mathcal{B}(X)$ et donc en passant à la tribu engendrée : $\sigma\Big(B : \text{Boule ouverte de X}\Big) \subset \mathcal{B}(X)$.

Le contenu de la proposition est la réciproque. Il suffit de montrer que $\mathcal{T} \subset \sigma\Big(B : \text{Boule ouverte de X}\Big)$ car alors, en passant à la tribu engendré, on obtient $\mathcal{B}(X) \subset \sigma\Big(B : \text{Boule ouverte de X}\Big)$.

Montrons qu'en fait, tout ouvert U est union au pus dénombrable de boules ouvertes. Comme X est séparable, c'est aussi le cas de U. Soit $D = \{x_n : n \in \mathbb{N}\} \subset U$ une suite dense. Comme U est ouvert, il existe $r_n \in \mathbb{Q}\cap]0, +\infty[$ tel que $B(x_n, r_n) \subset U$. Soit donc $A = \{(x_n, r_n) : r_n \in \mathbb{Q}\cap]0, +\infty[, B(x_n, r_n) \subset U\}$ est donc a.p.d comme sous-ensemble d'un produit d'ensembles dénombrables. Donc en passant à l'union on a : $\bigcup_{(x_n, r_n)\in A} B(x_n, r_n) \subset U$. Montrons que

$$U = \bigcup_{(x_n, r_n)\in A} B(x_n, r_n) \in \sigma\Big(B : \text{Boule ouverte de X}\Big)$$

Soit $x \in U$, il existe $r > 0$ avec $B(x, r) \subset U$. Puis il existe n tel que $d(x, x_n) < \frac{r}{3}$ et soit $r_n \in \mathbb{Q}$ avec $r/3 < r_n < r/2$ (par densité de $\mathbb{Q}$ dans $\mathbb{R}$= donc $x \in B(x_n, r/3) \subset B(x_n, r_n) \subset B(x_n, r/2) \subset B(x, r) \subset U$ donc $(x_n, r_n) \in A$ et $x \in \bigcup_{(x_n, r_n)\in A} B(x_n, r_n)$. Comme x est arbitraire, on a l'insclusion réciproque qui conclut : $U \subset \bigcup_{(x_n, r_n)\in A} B(x_n, r_n)$. □

Preuve du Corollaire 4.17. Il suffit de remarquer que $\mathcal{B}(\overline{\mathbb{R}}) = \sigma\Big(\{\{+\infty\}, \{-\infty\}\} \cup \{]a, b[: a < b < a + 2\}\Big)$, car $\overline{\mathbb{R}}$ est séparable ($\{+\infty, -\infty\} \cup \mathbb{Q}$ est dense car la densité sur $\mathbb{R}$ coïncide avec la densité usuelle vu qu'on a les mêms ouverts, cf TD 1) et que les ensembles de la partie génératrice sont les boules ouvertes pour $d_{\overline{\mathbb{R}}}$.

Il suffit de noter que $]a, b[= \bigcap_{n\geq 1} \left[a - \frac{1}{n}, b + \frac{1}{n}\right] \in \sigma\Big(\{\{+\infty\}, \{-\infty\}\} \cup \{[a, b] : a < b\}\Big)$ pour déduire que

$$\mathcal{B}(\overline{\mathbb{R}}) = \sigma\Big(\{\{+\infty\}, \{-\infty\}\} \cup \{[a, b] : a < b\}\Big)$$

Par le lemme 4.13, on a alors que f est mesurable si et seulement si

1. $f^{-1}(\{\infty\}) \in \mathcal{T}$
2. $f^{-1}(\{-\infty\}) \in \mathcal{T}$
3. Pour tout $a < b \in \mathbb{R}$, $f^{-1}([a, b]) \in \mathcal{T}$

C'est exactement le résultat voulu (et on a vu que le dernier point équivaut à la mesurabilité de la restrition de f à $\mathbb{R}$.

□

Preuve du lemme 4.6

Pour rappel, on veut montrer que

$$\mathcal{B}(\mathbb{R}^n) = \sigma\Big(\prod_{i=1}^{n}]a_i, b_i[, a_i < b_i \in \mathbb{R}\Big).$$

Comme les produits d'intervalles ouverts sont des ouverts, et que les boules ouvertes pour la norme infini sont des boules ouvertes, on a

$$\begin{aligned}&\{B : \text{Boule ouverte de } \mathbb{R}^n, \|\cdot\|_\infty\}\\ &\quad\subset \left\{\prod_{i=1}^{n}]a_i, b_i[, a_i < b_i \in \mathbb{R}\right\} \subset \mathcal{T}.\end{aligned}$$

Donc en prenant la tribu engendrée et en appliquant la proposition C.6 sachant que $\mathbb{R}^n$ est séparable par la proposition C.5, on obtient :

$$\begin{aligned}\mathcal{B}(\mathbb{R}^n) &= \sigma\Big(\{B : \text{Boule ouverte de } \mathbb{R}^n, \|\cdot\|_\infty\}\Big)\\ &\subset \sigma\Big(\prod_{i=1}^{n}]a_i, b_i[, a_i < b_i \in \mathbb{R}\Big) \subset \sigma\Big(\mathcal{T}\Big) = \mathcal{B}(\mathbb{R}^n).\end{aligned}$$

3 Stabilité des fonctions mesurables

Lemme C.7. Un supremum d'une suite $f_n : (\Omega, \mathcal{T}) \to \overline{\mathbb{R}}$ de fonctions mesurables est mesurable.

Démonstration. On note $f = \sup_{n\geq 1} f_n$ et on remarque que

$$f^{-1}([-\infty, a]) = \{\omega \in \Omega : \sup_{n\geq 1} f_n(\omega) \leq a\} = \cap_{n\geq 1} f_n^{-1}([-\infty, a]).$$

Or par le corollaire 4.17, on sait que $f_n^{-1}(\{-\infty\}$ est dans $\mathcal{T}$ et aussi $f_n^{-1}(]-\infty, a]) = \cup_{n\geq 1} f_n^{-1}([-n, a]) \in \mathcal{T}$ par union dénombrable. Donc chaque $f_n^{-1}([-\infty, a]) \in \mathcal{T}$ et donc par intersection dénombrable, on a $f^{-1}([-\infty, a]) \in \mathcal{T}$. Par le corollaire 4.16, on déduit que la restriction de f à $\mathbb{R}$ est mesurable et donc pour tout $a < b$, on a $f^{-1}([a, b]) \in \mathcal{T}$.

Enfin, $f^{-1}(\{-\infty\}) = \cap_{n\geq 1} f_n^{-1}(\{-\infty\}) \in \mathcal{T}$ et $f^{-1}(\{+\infty\}) = \bigcap_{n\geq 1} f^{-1}(]n, +\infty]) \in \mathcal{T}$. Or $f^{-1}(]n, +\infty]) = f^{-1}([-\infty, n])^c \in \mathcal{T}$ donc par intersection dénombrable, on a bien $f^{-1}(\{+\infty\}) \in \mathcal{T}$. Par la réciproque du corollaire 4.17, on déduit que f est mesurable. □

Lemme C.8. La lim sup, lim inf d'une suite $f_n : (\Omega, \mathcal{T}) \to \overline{\mathbb{R}}$ de fonctions mesurables est mesurable.

Démonstration. Comme $\inf_n f_n = -\sup_n -f_n$, on déduit qu'un infimum d'une suite de fonctions mesurables est mesurable. Or, comme rappelé au chapitre précédent,

$$\limsup_n f_n = \inf_{n\geq 0} \sup_{k\geq n} f_k, \liminf_n f_n = \sup_{n\geq 0} \inf_{k\geq n} f_k$$

est donc mesurable en utilisant le résultat du lemme précédent sur le suprémum (ou l'infinimum) de fonctions mesurables. □

Proposition C.9. Une limite simple d'une suite $f_n : (\Omega, \mathcal{T}) \to \overline{\mathbb{R}}$ de fonctions mesurables est mesurable.

Démonstration. Si une suite converge simplement, on a $\lim_{n\to\infty} f_n = \limsup_n f_n$ qui est donc mesurable par le lemme précédent. □

4 Compléments sur la construction de l'intégrale

Intégrale des fonctions étagées

La définition 4.11 est motivée par le résultat suivant :

Lemme C.10. L'intégrale $\int_B f d\mu$ ne dépend pas de la décomposition $f(\omega) = \sum_{i=1}^{n} a_i 1_{A_i}(\omega)$ en somme d'indicatrice d'ensembles deux à deux disjoints mais seulement de f.

Démonstration. Pour $f = \sum_{i=1}^{n} a_i 1_{A_i}$, il existe toujours une (unique) représentation canonique de f en voyant $b_1 < \cdots < b_m$ tel que l'image $f(\Omega) - \{0\} = \{b_1, \cdots, b_m\}$ et en prenant $B_i = f^{-1}(\{b_i\}) \in \mathcal{T}$ car f est mesurable. Alors, on a $f(\omega) = \sum_{i=1}^{n} b_i 1_{B_i}(\omega)$. Comme les A_i sont 2 à deux disjoints, on voit B_j comme union disjointe de A_i et donc $\mu(B_i \cap B) = \sum_{\{j : a_j = b_i\}} \mu(A_j \cap B)$ donc, en regroupant par paquet :

$$\begin{aligned}
\int_B f d\mu &= \sum_{j=1}^{n} a_i \mu(A_j \cap B) \\
&= \sum_{i=1}^{m} \sum_{\{j : a_j = b_i\}} b_i \mu(A_i \cap B) \\
&= \sum_{i=1}^{m} b_i \mu(B_i \cap B).
\end{aligned}$$

C'est la formule qui ne dépend que de f (comme sa représentation canonique). □

Preuve du lemme 4.20

1. Si $f(\omega) = \sum_{i=1}^{n} a_i 1_{A_i}(\omega)$ avec les A_i deux à deux disjoints, alors $1_B f(\omega) = \sum_{i=1}^{n} a_i 1_{A_i \cap B}(\omega)$. Donc $\int_B f d\mu = \sum_{i=1}^{n} a_i \mu(A_i \cap B) = \int_\Omega 1_B f d\mu$.

2. De même, $cf(\omega) = \sum_{i=1}^{n} c a_i 1_{A_i}(\omega)$, alors $\int_B cf d\mu = \sum_{i=1}^{n} c a_i \mu(A_i \cap B) = c \int_B f d\mu$.

3.Si de plus $h = \sum_{j=1}^{m} b_j 1_{B_j}(\omega)$ avec les B_j deux à deux disjoints, et soit $B_0 = \Omega - \bigcup_{j=1}^{m} B_j$, $A_0 = \Omega - \bigcup_{i=1}^{n} A_i$, alors les $A_i \cap B_j$ deux à deux disjoints $i = 0, \cdots, n, j = 0, \cdots, m$. De même, avec $a_0 = b_0 = 0$, $f(\omega) = \sum_{j=0}^{m} \sum_{i=0}^{n} a_i 1_{A_i \cap B_j}(\omega)$, $h(\omega) = \sum_{j=0}^{m} \sum_{i=0}^{n} b_j 1_{A_i \cap B_j}(\omega)$. Donc

$$f(\omega) + h(\omega) = \sum_{j=0}^{m} \sum_{i=0}^{n} (a_i + b_j) 1_{A_i \cap B_j}(\omega)$$

On obtient donc :

$$\begin{aligned}
\int_B f + h d\mu &= \sum_{j=0}^{m} \sum_{i=0}^{n} (a_i + b_j) \mu(A_i \cap B_j) \\
&= \sum_{j=0}^{m} \sum_{i=0}^{n} a_i \mu(A_i \cap B_j) + \sum_{j=0}^{m} \sum_{i=0}^{n} b_j \mu(A_i \cap B_j) \\
&= \int_B f d\mu + \int_B h d\mu.
\end{aligned}$$

4. Si $0 \leq f \leq h$ alors $h = (h - f) + f$ est la somme de deux fonctions étagées positives et par le 3, $\int_B f d\mu \leq \int_B f d\mu + \int_B (h - f) d\mu = \int_B h d\mu$.

Preuve du lemme 4.22

On va utiliser que toutes les propriétés sont vraies si f, h sont étagées par définition de l'intégrale dans ce cas.

1. Soit g étagée avec $g \leq f$ alors $g \leq h$, donc par définition $0 \leq \int_B g d\mu \leq \int_B h d\mu$. En passant au sup sur les g, on obtient $0 \leq \int_B f d\mu \leq \int_B h d\mu$.

2. Soit g étagée avec $g \leq f$, alors $1_B g \leq 1_B f$. Or par le cas étagé du lemme 4.20, on a $\int_B g d\mu = \int_\Omega 1_B g d\mu$ et donc par définition : $\int_B g d\mu = \int_\Omega 1_B g d\mu \leq \int_\Omega 1_B f d\mu$. En passant au sup, on obtient $\int_B f d\mu \leq \int_\Omega 1_B f d\mu$.

En sens inverse, g étagée positive avec $g \leq 1_B f \leq f$ vérifie donc $g 1_B = g$ et par définition $\int_\Omega g d\mu = \int_\Omega g 1_B d\mu = \int_B g d\mu \leq \int_B f d\mu$ soit en passant au sup $\int_\Omega 1_B f d\mu \leq \int_B f d\mu$.

Le cas particulier vient du 1. appliqué à l'inéaglité $1_A f \leq 1_B f$ sous la forme : $0 \leq \int_A f d\mu = \int_\Omega 1_1 f d\mu \leq \int_\Omega 1_B f d\mu = \int_B f d\mu$.

3. Si $c = 0$ c'est évident, on suppose donc $c > 0$. Alors pour $g \leq f$ avec g étagée positive, on a $cg \leq cf$ donc par le cas étagé du lemme 4.20, on a $c \int_B g d\mu = \int_B cg d\mu \leq \int_B cf d\mu$. En

passant au sup, on a obtenu :

$$c \int_B f d\mu \leq \int_B c f d\mu$$

mais en appliquant) cf à la place de f et $f = \frac{1}{c} cf$, on obtient :

$$\frac{1}{c} \int_B c f d\mu \leq \int_B c \frac{1}{c} f d\mu = \int_B f d\mu$$

d'où l'inégalité dans l'autre sens $\int_B c f d\mu \leq c \int_B f d\mu$ et donc l'égalité.

4. Si $f = 0$ $0 \leq g \leq f$ implique $g = 0$ et en passant au sup de 0, on obtient le résultat.

Si $\mu(B) = 0$, $\int_B g d\mu = \int_\Omega 1_B g d\mu$ et si $g(\omega) = \sum_{i=1}^{n} a_i 1_{A_i}(\omega)$, on a $\int_\Omega 1_B g d\mu = \sum_{i=1}^{n} a_i \mu(B \cap A_i) = 0$ car chaque $\mu(B \cap A_i) \leq \mu(B) = 0$.

5. Si on a $0 \leq g \leq f$, $0 \leq k \leq h$ avec g, k mesurable positive, alors $g + k \leq f + h$ est mesurable positive, donc $\int_B f + h d\mu \geq \int_B g + k d\mu = \int_B g d\mu + \int_B k d\mu$. En passant au sup, on obtient le résultat.

ANNEXE D

Compléments facultatifs et hors programme au chapitre 6 : Espaces L^p

1 Formule alternative de la norme (niveau L3)

On va en déduire l'expression alternative suivante dont l'inégalité triangulaire se déduit facilement. Cette méthode a l'avantage d'être utile pour le calcul du dual.

Proposition D.1. Soit μ σ−finie, $p \in [1,\infty]$, q tel que $1/p + 1/q = 1$ le coefficient conjugué, alors pour tout g mesurable

$$||g||_q = \sup\left\{\left|\int fgd\mu\right| ; ||f||_p \leq 1,\right.$$
$$\left. fg \in L^1(\Omega,\mu), f \in L^1(\Omega,\mu) \cap L^\infty(\Omega,\mu)\right\}.$$

Démonstration. Soit A_n croissant telle que $\cup A_n = \Omega$, $\mu(A_n) < \infty$. On commence par le cas $g \in L^q(\Omega,\mu)$.

Par Hölder, $fg \in L^1$ donc l'intégrale est définie (avec la condition $||f||_p \leq 1$ seule) et

$$\left|\int fgd\mu\right| \leq ||fg||_1 \leq ||f||_p||g||_q$$

d'où $||g||_q$ est plus grand que le sup de l'énoncé. Mais, pour $p \in]1,\infty[$, si on prend $f = \overline{g}|g|^{q-2}/||g||_q^{q-1}$ on a $|f|^p = |g|^{p(q-1)}/||g||_q^{p(q-1)} = |g|^q/||g||_q^q$ car $p(q-1) = qp(1-1/q) = q$, donc $f \in L^p$ et $||f||_p^p = E(|f|^p) = ||g||_q^q/||g||_q^q = 1$. Donc $||f1_{A_n}||_p^p \leq ||f||_p^p \leq 1$ donc comme $L^p(A_n,\mu) \subset L^1(A_n,\mu)$ car $\mu(A_n) < \infty$ on a $f1_{A_n} \in L^1(\Omega,\mu)$ et donc

$$g_{n,m}(f) = 1_{\{f1_{A_n}\neq 0\}}\frac{f1_{A_n}}{|f1_{A_n}|}min(m,|f1_{A_n}|) \in L^\infty(\Omega,\mu) \cap L^1(\Omega,\mu)$$

d'où le sup est supérieur à

$$\left|\int g_{n,m}(f)gd\mu\right| \to_{m\to\infty} \left|\int f1_{A_n}gd\mu\right| \to_{n\to\infty} \left|\int fgd\mu\right|$$

(par convergence dominée par $|g_{n,m}(f)g| \leq |fg|$) et le sup est supérieur à $\left|\int fgd\mu\right| = \int |g|^q d\mu/||g||_q^{q-1} = ||g||_q$. On déduit donc l'égalité énoncée.

Si $p = 1$, $q = \infty$, soit

$$C > \sup\left\{\left|\int fgd\mu\right|; ||f||_1 \leq 1 fg \in L^1(\Omega,\mu),\right.$$
$$\left. f \in L^1(\Omega,\mu) \cap L^\infty(\Omega,\mu)\right\}$$

et $A = \{x : |g(x)| > C\}$. Supposons par l'absurde que $\mu(A) > 0$ soit $B \subset A$ avec $\mu(B) \in]0,\infty[$. Alors $f = 1_B \frac{\overline{g}}{|g|\mu(B)}$ est dans L^1 et $||f||_1 = 1$ (et borné par $1/\mu(B)$ donc dans L^∞) mais $\left|\int fgd\mu\right| = \int 1_B \frac{|g|}{\mu(B)} \geq C$ en contradiction avec le choix de C donc $\mu(A) = 0$ ce qui implique $||g||_\infty \leq C$ ce qui donne le résultat en prenant l'inf des C.

Si $p = \infty$, $q = 1$, il suffit de prendre $f = 1_{g\neq 0}\frac{\overline{g}}{|g|} \in L^\infty(\Omega)$ et $f1_{A_n} \in L^\infty(\Omega) \cap L^1(\Omega)$ de sorte que $f1_{A_n}g = |f|1_{A_n}$ et la norme $||f1_{A_n}||_\infty \leq 1$. Donc le supremum, est supérieur à $\int |f|1_{A_n}d\mu \to ||f||_1$ par convergence monotone.

Si on ne suppose plus $g \in L^q(\Omega,\mu)$ mais $||g||_q = \infty$. Soit alors $g_{n,m} = 1_{\{g\neq 0\}}\frac{g}{|g|}min(m,|g|)1_{A_n} \in L^q(\Omega,\mu)$ on obtient $f_{n,m,k} \in L^1 \cap L^\infty$ de norme ≤ 1 dans L^p tel que

$$|\int f_{n,m,k}g_{n,m}| \to_{k\to\infty} ||g_{n,m}||_q.$$

Comme on a l'inégalité par Hölder,

$$\left|\int f_{n,m,k}(g_{n,m} - g1_{A_n})\right| \leq ||f_{n,m,k}||_p||(g_{n,m} - g1_{A_n})||_q$$
$$\leq ||(g_{n,m} - g1_{A_n})||_q \to_{m\to\infty} 0$$

par convergence monotone car $|\min(|g|,m) - |g||^q$ décroit vers 0, on trouve une suite m_k tel que

$$|\int f_{n,m_k,k}g1_{A_n}| \to_{k\to\infty} ||g1_{A_n}||_q$$

(fini ou infini). Enfin comme par convergence monotone $||g1_{A_n}||_q \to ||g||_q$, on trouve une suite

$$|\int f_{n_k,m_k,k}g1_{A_n}| \to_{k\to\infty} ||g||_q = \infty.$$

Comme $||f_{n_k,m_k,k}1_{A_n}||_p \leq 1$, et $f_{n_k,m_k,k}1_{A_n} \in L^1 \cap L^\infty$ et $f_{n_k,m_k,k}g1_{A_n} \in L^1$ cela donne la solution :

$$\sup\{\left|\int fgd\mu\right|; ||f||_p \leq 1, fg \in L^1(\Omega), f \in L^1 \cap L^\infty\} = \infty = ||g||_q.$$

□

Exemple D.1. Dans le cas où μ est la mesure de comptage sur I (σ−finie si I dénombrable), $\mu(A) = Card(A)$, on obtient l'espace $\ell^p(I, \mathbb{K})$ des suites indicées par I de puissance p sommable, i.e. telles que

$$\sum_{i \in I} |x_i|^p < \infty$$

pour $p \in [1, \infty[$ et l'ensemble des suites bornées, c'est-à-dire telles que

$$||x||_\infty = \sup_{i \in I} |x_i| < \infty$$

pour $p = \infty$.

2 Premiers résultats de densité (niveau M1)

On rappelle qu'une fonction étagée intégrable sur $(\Omega, \mu, \mathcal{T})$ est une combinaison linéaire (finie) de fonctions indicatrices 1_A avec $\mu(A) < \infty$.

Lemme D.2. Soit $(\Omega, \mu, \mathcal{T})$ un espace σ−fini. L'ensemble S des fonctions étagées intégrables est dense dans tous les $L^p(\Omega, \mu, \mathcal{T})$, $1 \leq p < \infty$. En particulier, $L^1(\Omega, \mu, \mathcal{T}) \cap L^\infty(\Omega, \mu, \mathcal{T})$ est dense dans $L^p(\Omega, \mu, \mathcal{T})$ pour $1 \leq p < \infty$.

Démonstration. Cela vient de la construction de l'intégrale, et du fait que les fonctions étagées sont dans $L^1(\Omega, \mu, \mathcal{T}) \cap L^\infty(\Omega, \mu, \mathcal{T})$, mais rappelons une preuve. En décomposant en parties réelle et imaginaire puis parties positive et négative, on se ramène à approcher $f \in L^p$ avec $f \geq 0$. Si $\Omega = \cup A_n \mu(A_n) < \infty$, on a $||f 1_{A_n} - f||_p \to 0$ par convergence dominée, donc on prend $h = f 1_{A_m}$.

On prend

$$h_n(x) = \sum_{k=0}^{4^n} \frac{k}{2^n} 1_{[\frac{k}{2^n}, \frac{k+1}{2^n}[}(h(x)) = \sum_{k=0}^{4^n} \frac{k}{2^n} 1_{h^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}[)}(x) \leq h(x)$$

Comme h mesurable, il est facile de voir que $h \in S$,

$$||h - h_n||_p \leq ||h 1_{h(x) \geq 2^n}||_p + ||1_{h(x) \leq 2^n} 1_{A_m}||_p \frac{1}{2^n}$$

et le premier terme tend vers 0 par convergence dominée (par $|h|^p$), le second car $\mu(A_m)^{1/p} < \infty$. Donc h puis f sont dans l'adhérence.

□

Pour obtenir un résultat de densité des fonctions continues, on a besoin d'un résultat de continuité sur un grand ensemble pour les fonctions mesurables. On a besoin d'une compatibilité entre théorie de la mesure et topologie qui fait l'objet de la définition suivante. L'essentiel est que la mesure de Lebesgue sur $\mathbb{R}^n$ est un exemple de mesure de Radon, ainsi que toutes les mesures à densité par rapport à la mesure de Lebesgue (et aussi les mesures discrètes).

Définition D.1. Une **mesure de Radon positive** sur X localement compact est une mesure positive définie sur une tribu $\mathcal{T}$ contenant la tribu borélienne $\mathcal{B}$ et telle que :

1. $\mu(K) < \infty$ pour K compact (on parle de mesure de Borel).
2. μ est extérieurement régulière au sens où pour tout $E \in \mathcal{T}$, on a :
$$\mu(E) = \inf\{\mu(V) | E \subset V, \ V \text{ ouvert } \}$$
3. μ vérifie pour tout E ouvert et $E \in \mathcal{T}$ avec $\mu(E) < \infty$, on a :
$$\mu(E) = \sup\{\mu(K) | E \supset K, \ K \text{ compact } \}$$
4. $\mathcal{T}$ est complète pour μ au sens où si $E \in \mathcal{T}$, $A \subset E$ et $\mu(E) = 0$ alors $A \in \mathcal{T}$.

On va utiliser deux lemmes topologiques (en fait reliés) :

Théorème D.3 (de prolongement de Tietze)**.** (exo en section A) Soit X un espace métrique, F un fermé de X et $f : F \to \mathbb{R}$ une fonction continue bornée par C, alors il existe une fonction $g : X \to \mathbb{R}$ bornée par C et prolongeant f.

On rappelle qu'un espace topologique est dit **localement compact** si tout point a un voisinage (d'adhérence) compact. [Rmq : pour nous, un voisinage d'un point n'est pas forcément ouvert, c'est seulement un ensemble contenant un ouvert contenant le point] Par exemple c'est le cas de $X = \mathbb{R}^n$!

Lemme D.4 (d'Urysohn)**.** Si X est un espace métrique localement compact, V un ouvert contenant un compact K, alors il existe f continue à support compact tel que $1_K \leq f \leq 1_V$.

Démonstration. Pour tout $x \in K$, soit U_x voisinage ouvert d'adhérence compact inclus dans V (pour voir que l'adhérence peut être inclus dans V il suffit d'intersecter le voisinage avec $\{y : d(y, V^c) > \epsilon/2\}$ pour $\epsilon = d(x, V^c)$). On recouvre K par un nombre fini de U_x, $K \subset U := \cup_{i=1}^n U_{x_i}$ et $\overline{U} = \cup_{i=1}^n \overline{U_{x_i}}$ est compact et on trouve un ouvert d'adhérence compact W, $V \supset W \supset \overline{U}$ et on pose $F = W^c \cup K$. On définit $g : F \to \mathbb{R}$ par $g = 1_K$. Si $x_n \in F$, $x_n \to x \in K$

nécessairement pour n grand $x_n \in U$ donc $x_n \in K$ donc $g(x_n) = g(x) = 1$. De même si $x \in W^c$, pour n grand, $x_n \in (\overline{U})^c$, donc $x_n \in W^c$ et $g(x_n) = g(x) = 0$. Donc g est continue sur F et s'étend en une fonction $f : X \to \mathbb{R}$ continue par le théorème précédent et en centrant on a même, $0 \le f \le 1$ ($|f - 1/2| \le 1/2$). Donc le support de f est dans $\overline{W}$ compact et $1_K \le f \le 1_W \le 1_V$ ce qui conclut. □

Théorème D.5 (de Lusin)**.** Soit X un espace métrique localement compact. μ une mesure de Radon positive. Soit f une fonction complexe mesurable sur X s'annulant en dehors de A avec $\mu(A) < \infty$. Alors, pour tout $\epsilon > 0$, il existe g continue à support compact avec $\sup_{x \in X} |g(x)| \le \sup_{x \in X} |f(x)|$ et telle que :

$$\mu(\{x : f(x) \neq g(x)\}) \le \epsilon.$$

Démonstration. **Cas A compact,** $0 \le f \le 1$**.** On pose

$$f_n(x) = \sum_{k=0}^{2^n} \frac{k}{2^n} 1_{[\frac{k}{2^n}, \frac{k+1}{2^n}[}(f(x)) \le f_{n+1}(x) \le f(x).$$

Remarquons que $t_n := f_{n+1}(x) - f_n(x) = \frac{1}{2^{n+1}} \sum_{k=0}^{2^{n+1}} 1_{[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}[}(f(x)) = \dfrac{1}{2^{n+1}} 1_{T_n}$, $(f_{-1} := 0)$ avec $T_n \subset A$ de sorte que :

$$f(x) = \sum_{n=-1}^{\infty} t_n(x).$$

Comme dans la preuve du lemme d'Urysohn, il existe un ouvert $A \subset V$ avec $\overline{V}$ compact, puis par régularité extérieure, on trouve V_n ouvert avec $T_n \subset V_n \subset V$ et enfin par intérieure régularité sur les ensembles de mesures finies $K_n \subset T_n$ avec $\mu(V_n - K_n) \le 2^{-n-2}\epsilon$. Par le lemme d'Urysohn, on trouve h_n continue à support compact avec $1_{K_n} \le h_n \le 1_{V_n}$. On pose

$$g(x) = \sum_{k=-1}^{\infty} 2^{-n-1} h_n(x).$$

Par convergence uniforme (car normale) de la série, g est continue, à support compact car inclus dans $\overline{V}$. Enfin $2^{-n-1}h_n(x) = t_n(x)$ sauf sur $V_n - K_n$ donc $f = g$ sauf sur $\cup_n(V_n - K_n)$ qui est de mesure au plus ϵ

Cas A quelconque, $0 \le f \le 1$**.** Par régularité, on prend $A \subset V$ ouvert, $K \subset V$ compact avec $\mu(A \cap K^c) \le \mu(V \cap K^c) \le \epsilon/2$ et on applique à $f 1_K$ (vu $\{f 1_K \neq f\} \subset A \cap K^c$) le cas précédent en remplaçant ϵ par $\epsilon/2$.

Cas général Soit $B_n = \{x | f(x)| > n\}$ de sorte que $\cap B_n = \emptyset$, comme $\mu(B_1) < \infty$ en utilisant le TCM sur $1_{B_1} - 1_{B_n}$, $\mu(B_n) \to 0$, on applique à $(1 - 1_{B_n})f$ en décomposant la fonction en somme de 4n fonctions à valeur $[0,1]$ (4 pour décompositions en parties positives, négatives des parties imaginaires et réelles, et ces fonctions sont dans $[0,n]$ d'où la décomposition en somme de n fonctions à valeurs $[0,1]$). Enfin pour avoir l'inégalité on remplace g par $\phi \circ g$ avec $\phi(x) = x, |x| \leq R = \sup_{x\in X} |f(x)|$, $\phi(x) = Rx/|x|, |x| > R$. On a $g(x) = \phi \circ g(x)$ pour tout x tel que $f(x) = g(x)$, donc on n'augmente pas l'ensemble sur lequel f et g diffèrent. □

Corollaire D.6. Soit $(X,\mu,\mathcal{T})$ un espace métrique localement compact avec μ mesure de Radon σ−finie. L'ensemble $C_c(X)$ des fonctions continues à support compact est dense dans tous les $L^p(X,\mu,\mathcal{T})$, $1 \leq p < \infty$. De plus si $f \in L^p(X,\mu,\mathcal{T})$ et $\int f\phi = 0$, pour tout $\phi \in C_c(X)$ alors $f = 0$ $p.p$.

Démonstration. Par le lemme précédent, il suffit d'approcher les éléments de S. Par le théorème de Lusin D.5, pour chaque $f \in S$, $\epsilon > 0$, on a $g \in C_c(X)$ avec $\mu(g \neq f) \leq \epsilon$ et $\sup |g| \leq sup|f| = C$ donc $||f - g||_p \leq 2C\mu(g \neq f)^{1/p}$ et cette quantité est arbitrairement petite. Pour le résultat d'annulation, si $p > 1$, On utilise la densité dans L^q, q exposant conjugué, pour obtenir $\int fg = 0$ pour $g \in L^q$, d'où on déduit $||f||_p = 0$ par la proposition D.1. Si $p = 1$, on remplace f par $f|_V$ avec V ouvert $\overline{V}$ compact, qui couvrent X par locale compacité de sorte qu'on peut supposer $\mu(X) < \infty$. On peut supposer f réelle. Soit $f_1 \in C_c(X)$ avec $||f - f_1||_1 \leq \epsilon$, $K_1 = f_1^{-1}([\epsilon,\infty[)$ et $K_{-1} = f_1^{-1}(]-\infty,\epsilon])$ sont compacts, on prolonge par le Théorème de Tietze D.3, $u \in C_c(X)$ valant ϵ sur K_ϵ et soit $K = K_1 \cup K_{-1}$. Donc

$$||f_1||_1 = \int_K f_1 u + \int_{X-K} |f_1| \leq \int_X f_1 u + 2\int_{X-K} |f_1| \leq \epsilon + \int_X fu + 2\mu(X-K)\epsilon \leq \epsilon + 2\mu(X)\epsilon$$

car $|f_1| \leq \epsilon$ sur $X - K$. Donc $||f||_1 \leq 2\epsilon + 2\mu(X)\epsilon$ pour tout $\epsilon > 0$ ce qui donne $f = 0$. □

Donnons une application.

Proposition D.7. Soit $1 \leq p < \infty$ et soit $\tau_h f(x) := f(x+h)$ pour $h, x \in \mathbb{R}^d$, $f \in L^p(\mathbb{R}^d)$. La translation $\tau_h : L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ est isométrique et pour tout $f \in L^p(\mathbb{R}^d)$ $h \mapsto \tau_h(f)$ est continue de $\mathbb{R}^d \to L^p(\mathbb{R}^d)$.

Démonstration. L'isométrie est évidente par invariance de la mesure de Lebesgue par translation. Montrons que $||\tau_h f - f||_p \to_{h\to 0} 0$. En effet pour $\epsilon > 0$, par densité du lemme D.6, on trouve $f_1 \in C_c(\mathbb{R}^d)$ avec $||f_1 - f||_p \leq \epsilon/3$ donc comme τ_h est une isométrie : on obtient :

$$\begin{aligned}||\tau_h f - f||_p &\leq ||\tau_h f_1 - \tau_h f||_p + ||\tau_h f_1 - f_1||_p + ||f_1 - f||_p \\ &\leq 2\epsilon/3 + Leb(B(0,||h||) + \mathrm{Supp}(f_1))^{1/p} ||\tau_h f_1 - f_1||_\infty\end{aligned}$$

Pour h assez petit, comme f_1 est uniformément continue (car continue à support compact et par le Théorème de Heine), on peut trouver $1 \geq \delta > 0$ de sorte que si $||h|| \leq \delta$, $||\tau_h f_1 - f_1||_\infty = \sup_x |f_1(x+h) - f_1(x)| \leq \epsilon/[3Leb(B(0,1) + \mathrm{Supp}(f_1))^{1/p}]$ ce qui conclut. □

3 Dualité des espaces de Lebesgue $L^p(\Omega)$ (Niveau M1)

On rappelle que (Ω, μ) est un espace mesuré σ−fini. On se souvient que pour $p \in [1, \infty]$, q tel que $1/p + 1/q = 1$ la proposition D.1 donne pour g mesurable :

On a même le théorème suivant (on notera que $p < \infty$ contrairement au cas de la formule pour la norme) :

Théorème D.8 (de représentation de Riesz L^p)**.** Soit l'application définie grâce à l'inégalité de Hölder :

$$I : f \in L^q(\Omega, \mu) \mapsto (g \in L^p(\Omega, \mu) \mapsto \int fgd\mu)$$

Alors $I : L^q(\Omega, \mu) \to (L^p(\Omega, \mu))'$, réalise une isométrie SURJECTIVE pour $p \in [1, \infty[$ et q exposant conjugué c'est−à−dire tel que $1/p + 1/q = 1$.

Attention le cas $p = \infty$ est EXCLU... $L^\infty(\Omega)'$ est un espace très gros de mesures sur un espace stonien compact X tel que $L^\infty(\Omega) = C^0(X)$.

Démonstration. On a déjà montré l'isométrie, il reste à voir la surjectivité.

On fixe A_n avec $\mu(A_n) < \infty$ et $\bigcup_{n\in\mathbb{N}} A_n = \Omega$, A_n croissant.

Le **cas** $p = 2$ **a été traité par le théorème de représentation de Riesz.**

(1) **cas** $p = 1$Soit $\phi \in (L^1(\Omega, \mu))'$ avec $||\phi|| \leq 1$. D'abord on définit T application linéaire continue sur $L^2(\Omega)$ (en fait à valeur dans son dual identifié à lui même) par :

$$\langle Tx, y\rangle = \phi(\overline{x}y)$$

vu que $\overline{x}y \in L^1(\Omega)$ par Hölder et on a

$$\begin{aligned} ||T|| &:= \sup\{||Tx||_2, ||x||_2 \leq 1\} \\ &= \sup\{|\langle Tx, y\rangle|, ||x||_2 \leq 1, ||y||_2 \leq 1\} \leq ||\phi||_{L^1(\Omega)'}. \end{aligned}$$

La première égalité est la définition de la norme des applications linéaires bornées, la deuxième est le résultat de dualité du cas $p = 2$, la troisième utilise Hölder et la définition de la norme du dual. Notons que si $z \in L^\infty(\Omega)$,

$$\langle Tzx, y\rangle = \phi(\overline{zx}y) = \langle Tx, \overline{z}y\rangle = \langle zTx, y\rangle$$

la deuxième relation en utilisant la commutativité des espaces de fonctions soit la relation $\overline{zx}y = \overline{xz}y$ et la seconde la définition du produit scalaire $\langle Tx, \overline{z}y\rangle = \int \overline{Tx}\overline{z}yd\mu$. donc on déduit si m_z est la multiplication par $z \in L^\infty$, $Tm_z = m_zT$. Montrons que $T = m_g$ pour $g \in L^\infty$. (on dit que cette algèbre est son propre commutant dans $B(L^2(\Omega))$, ou qu'elle est maximale commutative).

En effet, soit $x_n = T(1_{A_n}) \in L^2$. On a $||T|| \leq 1$ car $||\phi|| \leq 1$.

Pour $g \in L^\infty$ avec $||g||_1 \leq 1$,

$$\begin{aligned}\left|\int T(1)gd\mu\right| &= \left|\int (|g|^{1/2}T)(1)g|g|^{-1/2}d\mu\right| \\ &= \left|\int T(|g|^{1/2})g|g|^{-1/2}d\mu\right| \\ &\leq |||g|^{1/2}||_2||g|g|^{-1/2}||_2 = ||g||_1 \leq 1\end{aligned}$$

où on a utilisé à la deuxième égalité la commutation avec $m_{|g|^{1/2}}$. On voit donc par la formule de la proposition D.1 que $||T(1_{A_n})||_\infty \leq 1$. Comme $T(1_{A_m}) = T(1_{A_m}1_{A_n}) = 1_{A_m}T(1_{A_n})$ donc on définit $g(x) = T(1_{A_n})(x)$ pour $x \in A_n$ de façon cohérente de sorte que $g1_{A_n} = T(1_{A_n})$ d'où $||g||_\infty = \sup_n ||g1_{A_n}||_\infty \leq 1$.

Et pour $z \in\in L^\infty \cap L^1 \subset L^2$ $T(z1_{A_n}) = m_g(z1_{A_n})$ donc par densité dans L^2 $T = m_z$. Enfin pour $f \in L^1(\Omega)$ $f = |f|^{1/2}g$ avec $g \in L^2$, on obtient

$$\begin{aligned}\phi(f) &= \phi(|f|^{1/2}g) = \langle T(|f|^{1/2}), g\rangle \\ &= \langle z(|f|^{1/2}), g\rangle = I(\overline{z})(|f|^{1/2}g) = I(\overline{z})(f).\end{aligned}$$

donc $\phi = I(\overline{z})$ d'où la surjectivité de I.

(2) **cas** $p > 1$ $\mu(\Omega) < \infty$ **utilisant les cas** $p = 1, 2$**.** (On l'appliquera ensuite à $\Omega = A_n$.) Après normalisation, on peut supposer $\mu(\Omega) = 1$.

On commence par montrer que via I, $L^p(\Omega)' \subset L^1(\Omega)$. Si $p \leq 2$, c'est évident par l'inclusion $L^2(\Omega) \subset [L^p(\Omega)]$ et par restriction et théorème de representation de Riesz, on obtient $g \in L^2(\Omega) \subset L^1(\Omega)$ tel que

$$\phi|_{L^2(\Omega)}(f) = \langle\overline{g}, f\rangle$$

Si $p > 2$ pour $x \in L^\infty$, et $\phi \in (L^p)'$,

$$|\phi(x)|^p \leq \int |x|^p d\mu \leq \int |x|^2 ||x||_\infty^{p-2} d\mu \leq ||x||_2^2 ||x||_\infty^{p-2}.$$

Par l'inégalité d'Young (cas particulier d'Holder utilisé dans sa preuve) $|ab| \leq a^P/P + b^Q/Q$ utilisé avec $1/P + 1/Q = 1$, $P = p/2$, $Q = p/(p-2)$, $a = ||x||_2^{1/P}/\epsilon^{1/Q}$, $b = (\epsilon ||x||_\infty)^{1/Q}$, on obtient :

$$|\phi(x)| \leq \frac{\epsilon}{Q}||x||_\infty + \frac{1}{P\epsilon^{P/Q}}||x||_2.$$

En incluant $\{(x,x), x \in (L^\infty(\Omega))\} \subset L^\infty(\Omega) \times L^2(\Omega)$ avec norme $||(x,y)|| = \frac{\epsilon}{Q}||x||_\infty + \frac{1}{P\epsilon^{P/Q}}||y||_2$ on étend par Hahn Banach ϕ à $L^\infty(\Omega) \times L^2(\Omega)$ donnant un élément de $(\phi_1, \phi_2) \in L^\infty(\Omega)' \times L^2(\Omega)$ avec $||\phi_1|| \leq \epsilon/Q, ||\phi_2|| \leq \frac{1}{P\epsilon^{P/Q}}$ (car en calculant la norme duale on a $\max(Q||\phi_1||/\epsilon, P\epsilon^{P/Q}||\phi_2||) \leq 1$) Donc $||\phi|_{L^\infty(\Omega)} - J(\phi_2)||_{(L^\infty(\Omega))'} = ||\phi_1||_{(L^\infty(\Omega))'} \leq \epsilon/Q$ et $\phi_2 \in L^1(\Omega)$. Or par le cas $p = 1$, $(L^1(\Omega))'' = L^\infty(\Omega)'$ et il contient $L^1(\Omega)$ comme espace fermé isométriquement via J (comme tout espace de Banach est inclus isométriquement comme espace fermé dans son bidual). Comme le résultat précédent indique $\phi \in \overline{L^2(\Omega)}^{(L^1(\Omega))''}$, on déduit $\phi \in J(L^1(\Omega))$ comme voulu. On a donc une fonction g telle que pour tout $f \in L^\infty(\Omega)$

$$\phi(f) = \int_\Omega g f d\mu$$

Soit donc g l'image dans L^1 de ϕ (on revient au cas général $p \in]1, \infty[$). Or dans le cas d'un espace avec mesure finie, l'équation de la proposition D.1 donne :

$$\begin{aligned} ||\phi||_{(L^p)'} &= \sup\{|\phi(x)|, ||x||_p \leq 1, x \in L^\infty\} \\ &= \sup\{|\int gx d\mu|, ||x||_p \leq 1, x \in L^\infty\} = ||g||_q \end{aligned}$$

On déduit donc $g \in L^q$ comme on voulait et $\phi = T(g)$ (en étendant la relation depuis$L^\infty(\Omega)$ par densité dans $L^p(\Omega)$.

(3)**cas** $1 < p < \infty$ **et** μ σ**−fini.** Soit $\phi \in (L^p(\Omega, \mu))'$, il faut montrer qu'elle vient d'un élément de $L^q(\Omega, \mu)$. On pose $\phi_n(f) = \phi(f 1_{A_n})$ pour $f \in L^p(A_n, \mu) \subset L^p(\Omega, \mu)$. Par le cas précédent, il existe $g_n \in L^q(A_n, \mu)$ telle que

$$\forall f \in L^p(A_n, \mu), \quad \int g_n f d\mu = \phi(f 1_{A_n}).$$

et

$$||g_n||_q = \sup\{\left|\phi(f 1_{A_n})\right|; ||f||_p \leq 1, f \in L^\infty(A_n, \mu)\} \leq ||\phi||_{(L^p)'} < \infty.$$

Or par unicité dans le cas (2) et vu les A_n croissant pour $n > m$, $g_n 1_{A_m} = g_m$ et donc $|g_n|$ est croissant et $g = \sup |g_n|$ vérifie par convergence monotone $||g||_q \leq ||\phi||_{(L^p)'}$, vu $|g_n| \leq |g|$ et comme $g_n \to g$ p.s., on déduit par convergence dominée $||g_n - g||_q \to 0$ et en passant à la limite $g_n = g1_{A_n}$.

Or $f1_{A_n} \to f$ dans L^p et donc par continuité la relation $\phi(f1_{A_n}) = T(g)(f1_{A_n})$ devient $\phi(f) = T(g)(f)$ pour tout $f \in L^p$ donc $\phi = T(g)$.

□

4 Convolution

Dans cette section, on considère l'espace mesuré $(\Omega, \mu, \mathcal{T}) = (\mathbb{R}^d, Leb, \mathcal{B})$ muni de la tribu borélienne et de la mesure de Lebesgue. On note alors $L^p(\mathbb{R}^d) = L^p(\mathbb{R}^d, Leb, \mathcal{B})$. Vu l'accord avec l'intégrale de Riemann, on note aussi $dy = d\lambda(y)$.

Théorème D.9 (définissant la Convolution)**.** Soient $f \in L^1(\mathbb{R}^d)$, $g \in L^p(\mathbb{R}^d)$, $1 \leq p \leq \infty$. Pour presque tout $x \in \mathbb{R}^d$, $y \mapsto f(x-y)g(y)$ est dans $L^1(\mathbb{R}^d)$. La **convolution** de f et g est la fonction $f * g$ définie par :

$$(f * g)(x) = \int_{\mathbb{R}^d} f(x-y)g(y)dy.$$

Alors $f * g \in L^p(\mathbb{R}^d)$ et :

$$||f * g||_p \leq ||f||_1 ||g||_p.$$

Démonstration. Si $p = \infty$, comme $|g| \leq ||g||_\infty p.p.$, $f(x-y)g(y) \leq ||g||_\infty |f(x-y)|$ d'où l'intégrabilité et la borne souhaitée en intégrant (comme la mesure de Lebesgue est invariante par translation).

On suppose d'abord $p = 1$ et on utilise le Théorème de Fubini Tonelli pour calculer :

$$\begin{aligned}\int dx |f| * |g|(x) &= \int dx \int dy |f(x-y)||g(y)| \\ &= \int dy \int dx |f(x-y)||g(y)| \\ &= ||f||_1 \int dy |g(y)| = ||f||_1 \; ||g||_1 < \infty\end{aligned}$$

On déduit du théorème de Fubini que pour presque tout x, $y \mapsto f(x-y)g(y)$ est intégrable et on obtient la borne souhaitée

$$||f * g||_1 \leq ||f||_1 ||g||_1.$$

Pour $1 < p < \infty$, soit q l'exposant conjugué. Du cas $p = 1$ on déduit $y \mapsto |f(x-y)||g(y)|^p$ est dans L^1 donc $y \mapsto |f(x-y)|^{1/p}|g(y)|$ est dans L^p pour presque tout x. Or $y \mapsto |f(x-y)|^{1/q} \in L^q$ donc par Hölder, $y \mapsto |f(x-y)||g(y)| = |f(x-y)|^{1/p}|g(y)|.|f(x-y)|^{1/q}$ est dans L^1 et

$$\begin{aligned} |(f * g)(x)|^p &\leq \left(\int |f(x-y)||g(y)|dy\right)^p \\ &\leq \left(\int |f(x-y)||g(y)|^p dy\right) ||f||_1^{p/q}. \end{aligned}$$

Par l'inégalité précédente du cas $p = 1$, on obtient donc en intégrant :

$$||f * g||_p^p \leq ||f||_1^{p/q} |||f| * |g|^p||_1 \leq ||f||_1^{p/q} ||g||_p^p ||f||_1 = ||f||_1^p ||g||_p^p.$$

□

Exercice D.1. (cf TD) Soit $f \in L^1$, $g \in L^p$, $h \in L^q$, $\check{f}(x) = \overline{f}(-x)$ Montrer que :

$$\int \overline{(f * g)} h = \int \overline{g} (\check{f} * h).$$

5 Support de la convolution

Si f continue, $\mathrm{Supp}(f) = \overline{\{x : f(x) \neq 0\}}$. Le résultat suivant permet d'étendre la définition aux fonctions mesurables.

Lemme D.10. Pour $f : \mathbb{R}^d \to \mathbb{K}$ mesurable, soit $(\omega_i)_{i \in I}$ la famille de tous les ouverts tels que, pour chaque i, $f = 0$ p.p sur ω_i. Si $\omega = \cup_{i \in I} \omega_i$ alors $f = 0$ p.p. sur ω. De sorte que ω est le plus grand ouvert sur lequel $f = 0$ p.p.

Démonstration. Il faut écrire ω comme union dénombrable car I n'est pas forcément dénombrable. Soit $K_n = \{x \in \omega : ||x|| \leq n, d(x, \omega^c) \geq 1/n\}$ comme la distance à un fermé est continue, on voit que K_n fermé borné de $\mathbb{R}^n$ (e.v.n de dimension finie) donc est compact et $\omega = \cup_{n \in \mathbb{N}} K_n$. Par compacité, K_n, recouvert par une union finie $K_n \subset \omega_{i_{n,1}} \cup \ldots \cup \omega_{i_{n,r_n}}$. donc $\omega = \cup_{n \in \mathbb{N}, j \leq r_n} \omega_{i,j}$ est union dénombrable d'ouvert sur lesquels $f = 0$ p.p. d'où le résultat. □

Définition D.2. Soit $f : \mathbb{R}^d \to \mathbb{K}$ mesurable, On pose $\mathrm{Supp}(f) = \mathbb{R}^d - \omega$ où ω est le plus grand ouvert sur lequel $f = 0$ p.p. Si $f \in L^p(\mathbb{R}^d)$, on pose $\mathrm{Supp}(f) = \mathrm{Supp}(f_1)$ pour n'importe quel représentant $f_1 \in f$ de la classe d'égalité presque partout.

Proposition D.11. Si $f \in L^1(\mathbb{R}^d)$, $g \in L^p(\mathbb{R}^d)$ alors :

$$\mathrm{Supp}(f * g) \subset \overline{\mathrm{Supp}(f) + \mathrm{Supp}(g)}.$$

Démonstration. On fixe $x \in \mathbb{R}^d$ avec $y \mapsto f(x-y)g(y) \in L^1$. Si $x \notin \mathrm{Supp}(f) + \mathrm{Supp}(g)$, on a $(x - \mathrm{Supp}(f)) \cap \mathrm{Supp}(g) = \emptyset$ donc en intégrant $f * g(x) = 0$ sur $Int((\mathrm{Supp}(f) + \mathrm{Supp}(g))^c) = (\overline{\mathrm{Supp}(f) + \mathrm{Supp}(g)})^c$. Donc $f * g$ est 0, p.p. sur cet ouvert de sorte qu'il est inclus dans $\mathrm{Supp}(f * g)^c$. □

6 Régularisation par convolution

On étudiera plus systématiquement au chapitre suivant certaines classes importantes de fonctions continues. Pour $\Omega \subset \mathbb{R}^d$ un ouvert. On note $C^k(\Omega)$ l'ensemble des fonctions k−fois différentiables avec leurs dérivées continues et $C^k_c(\Omega)$ les fonctions à support compact de $C^k(\Omega)$. Pour simplifier si $\alpha \in \mathbb{N}^d$, on note

$$D^\alpha f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}} f.$$

On note $|\alpha| = |\alpha_1| + \ldots + |\alpha_d|$. On note

$$C^\infty(\Omega) = \cap_{k \in \mathbb{N}} C^k(\Omega), \quad C^\infty_c(\Omega) = \cap_{k \in \mathbb{N}} C^k_c(\Omega).$$

Proposition D.12. Soit $1 \leq p \leq \infty$. Si $f \in C^k_c(\mathbb{R}^d)$, $g \in L^p(\mathbb{R}^d)$, $k \in \mathbb{N} \cup \{\infty\}$ alors $f * g \in C^k(\mathbb{R}^d)$ et si $|\alpha| \leq k$:

$$D^\alpha(f * g) = D^\alpha(f) * g.$$

De plus, si $p < \infty$, on a aussi la formule comprise comme intégrale de Riemann à valeur $L^p(\mathbb{R}^d)$, si Supp(f) $\subset [-C, C]^d$:

$$f * g = \int_{[-C,C]^d} dy f(y) \tau_{-y} g.$$

Démonstration. Par récurrence il suffit du cas $k = 1$. On applique le théorème de dérivation avec condition de domination. $\frac{\partial}{\partial x_i} f(x-y)g(y) = (\frac{\partial}{\partial x_i} f)(x-y)g(y)$.

Comme $(\frac{\partial}{\partial x_i} f)$ est à support compact et continue, il est borné par $||(\frac{\partial}{\partial x_i} f)||_\infty$ et

$$|\frac{\partial}{\partial x_i} f(x-y)g(y)| \leq ||\frac{\partial}{\partial x_i} f||_\infty 1_K(x-y)g(y),$$

avec K le compact support de f. Or par Hölder $\int 1_{B-K}(y)|g|(y)dy \leq Leb(B-K)^{1/q}||g||_p$, donc on a une domination par une fonction intégrable $c1_{B-K}g$ si $x \in B$ avec B compact. Le théorème de dérivation 4.39 conclut donc. De plus, par changement de variables linéaire si Supp(f) $\subset [-C, C]^d$, on a

$$\begin{aligned} f * g(x) &= \int_{\mathbb{R}^d} f(x-y)g(y)dy \\ &= \int_{\mathbb{R}^d} f(y)g(x-y)dy \\ &= \int_{[-C,C]^d} f(y)(\tau_{-y}g)(x)dy \end{aligned}$$

avec $\tau_h(g)(x) = g(x+h)$. On a vu à la proposition D.7 que $y \mapsto f(y)(\tau_{-y}g)$ est continue à valeur $L^p(\mathbb{R}^d)$ on peut donc parler de son intégrale de Riemann, sur $[-C, C]^d$ (calculée successivement variable par variable). On obtient une suite (de sommes de Riemann) qui converge dans $L^p(\mathbb{R}^d)$, donc quitte à extraire une suite qui converge p.p. et donc p.p. la limite $\int_{[-C,C]^d} dyf(y)(\tau_{-y}g)$ coïncide avec l'intégrale de Riemann $\int_{[-C,C]^d} dyf(y)(\tau_{-y}g)(x)$ par exemple si g est continue à support compact et cette intégrale vaut l'intégrale de Lebesgue donc $f * g(x)$. On en déduit l'égalité voulue dans L^p si g continue à support compact. Or par densité, on a une suite de fonctions g_n continues à support compact convergeant dans L^p vers g. Et comme $\sup_{\mathbb{R}^d} ||\tau_{-y}g_n - \tau_{-y}g||_p \to 0$, $f(.)(\tau_{-.}g_n)$ converge uniformément vers $f(.)(\tau_{-.}g)$ et comme l'intégrale de Riemann est continue pour la convergence uniforme $\int_{[-C,C]^d} dyf(y)(\tau_{-y}g)$ est la limite de $\int_{[-C,C]^d} dyf(y)(\tau_{-y}g_n)$ dans L^p qu'on a déjà vu valoir $f * g_n$, qui a pour limite $f * g$ donc $\int_{[-C,C]^d} dyf(y)(\tau_{-y}g) = f * g$. □

7 Suites régularisantes et densité par convolution

Définition D.3. Une **suite régularisante** est une suite de fonctions $\rho_n \in C_c^\infty(\mathbb{R}^d)$ avec $\int_{\mathbb{R}^d} \rho_n = 1$, $\rho_n \geq 0$ et $\text{Supp}(\rho_n) \subset B_{||.||_2}(0, 1/n)$.

Exercice D.2. Montrer que si $\rho_n(x) = Cn^d\rho(nx)$ avec $C\int \rho = 1$ et $\rho(x) = 1_{\{||x||_2<1\}} \exp(\frac{1}{||x||_2^2 - 1})$ alors ρ_n est une suite régularisante sur $\mathbb{R}^d$.

Lemme D.13. Soit ρ_n suite régularisante et $f \in L^p(\mathbb{R}^d)$ pour $1 \leq p < \infty$. Alors $||\rho_n * f - f||_p \to 0$.

Démonstration. On a comme $||.||_p$ est une norme on a par l'inégalité triangulaire (de l'intégrale de Riemann et la proposition D.12) :

$$||\rho_n * f - f||_p = ||\int dy \rho_n(y)(\tau_{-y} f - f)||_p \leq \int_{B(0,1/n)} dy \rho_n(y)||\tau_{-y} f - f)||_p$$

Or si n assez grand, on a vu à la proposition D.7 que $||\tau_{-y} f - f)||_p \leq \epsilon$ pour $y \in B(0, 1/n)$ de sorte que la dernière intégrale est bornée par $\epsilon \int_{B(0,1/n)} dy \rho_n(y) = \epsilon$. □

Proposition D.14. Soit $\Omega \subset \mathbb{R}^d$ un ouvert, alors $C_c^\infty(\Omega)$ est dense dans $L^p(\Omega)$ pour $1 \leq p < \infty$.

Démonstration. Soit $f \in L^p(\Omega)$ et $K_n = \{x \in \Omega : ||x||_2 \leq n, d(x, \Omega^c) \geq 1/n\}$. On a déjà remarqué que K_n compact et $\cup K_n = \Omega$ donc $f 1_{K_n} \to f$ p.p. et par la domination $|f 1_{K_n} - f| \leq |f|$ on conclut par le TCD à $||f 1_{K_n} - f||_p \to 0$. Soit $m > n$, si on considère $\rho_m * (f 1_{K_n}) \in C^\infty(\mathbb{R}^d)$, on a par la relation sur les supports des convolution,

$$\begin{aligned} \mathrm{Supp}(\rho_m * f 1_{K_n}) &\subset \mathrm{Supp}(\rho_m) + \mathrm{Supp}(f 1_{K_n}) \\ &\subset B(0, 1/m) + K_n \subset \Omega \end{aligned}$$

(vu que pour K, F compacts $K + F$ est compact et en comparant les distances pour la dernière inclusion). Donc $\rho_m * (f 1_{K_n}) \in C_c^\infty(\mathbb{R}^d)$. Mais on a vu $||\rho_m * (f 1_{K_n}) - f 1_{K_n}||_{L^p(\Omega)} = ||\rho_m * (f 1_{K_n}) - f 1_{K_n}||_p \to_{m \to \infty} 0$. Donc $f 1_{K_n}$ puis f sont dans l'adhérence de $C_c^\infty(\Omega)$. □

ANNEXE E
Compléments facultatifs et hors programme au chapitre 7

Le théorème des bases ne nécessite pas l'hypothèse I dénombrable ou H séparable, voici la version générale.

Comme l'existence de base algébrique d'un espace vectoriel de dimension infinie, elle requière un lemme général de théorie des ensembles :

1 Rappel sur le lemme de Zorn

Si on était en dimension finie, on voudrait faire une récurrence sur le cardinal d'une famille orthonormale en ajoutant un vecteur de plus pris dans un ensemble dense. Une façon de rédiger la preuve est de considérer un sous-espace de dimension maximale et d'obtenir une contradiction en construisant une famille libre de cardinal 1 de plus.

Dans le cas de la dimension infinie on pourrait faire une récurrence transfinie en complétant une base de G en une base de E et mettant un "bon ordre" sur la base. En analyse (ou en algèbre), on préfère souvent utiliser la conséquence suivante de l'axiome du choix, le lemme de Zorn, qui utilise une notion de maximalité pour obtenir une contradiction comme dans la preuve par induction.

Soit P muni d'un ordre partiel $\leq$. $Q \subset P$ est dit totalement ordonné si tout $a, b \in Q$ on a soit $a \leq b$, soit $b \leq a$. $c \in P$ est un majorant de Q si $\forall a \in Q, a \leq c$.

$m \in P$ est un élément maximal de P si tout $x \in P$ tel que $m \leq x$ on a $x = m$.

Enfin P est dit inductif si tout ensemble totalement ordonné de P admet un majorant.

Lemme E.1 (de Zorn)**.** Tout ensemble ordonné, inductif, non vide admet un élément maximal.

2 Théorème des bases dans le cas général

Théorème E.2. Soit H un espace préhilbertien.

1. Une famille orthonormale $(x_i)_{i\in I}$ est libre et vérifie l'inégalité de Bessel, pour tout $x \in H$:
$$\sum_{i\in I} |\langle x, x_i\rangle|^2 \leq ||x||^2$$
2. De plus une famille orthonormale $(e_i)_{i\in I}$ est une base hilbertienne si et seulement si on a l'égalité de Bessel–Parseval :
$$\sum_{i\in I} |\langle x, x_i\rangle|^2 = ||x||^2$$
De plus, dans ce cas, pour tout $x \in H$, la série suivante converge (dans H mais pas absolument)
$$x = \sum_{i\in I} e_i\langle e_i, x\rangle.$$
3. Si H est un espace de Hilbert, toute famille orthonormale peut être complétée en une base hilbertienne de H et $J : x \mapsto (\langle x, e_i\rangle)_{i\in I}$ établit alors une isométrie surjective $J : H \simeq \ell^2(I)$.

Remarque E.1. De la formule pour x, on tire par continuité la formule pour le produit scalaire (qui est une série absolument convergente par Cauchy–Schwarz) :
$$\langle y, x\rangle = \sum_{i\in I}\langle y, e_i\rangle\langle e_i, x\rangle.$$

Démonstration. (1) Si $\sum_{i\in I} \lambda_i x_i = 0$, on calcule $\lambda_j = \langle x_j, \sum_{i\in I} \lambda_i x_i\rangle = 0$ donc x_i est bien libre. Si F est une partie finie de I, et $V = V_F = Vect(e_i, i \in F)$, on a déjà vu la formule pour la projection orthogonale sur V_F :
$$p_V(x) = \sum_{i\in F} e_i\langle e_i, x\rangle.$$

Donc par la propriété de contraction de p_F et l'ortogonalité
$$||p_F(x)||^2 = \Big\langle \sum_{i\in F} e_i\langle e_i, x\rangle, \sum_{j\in F} e_j\langle e_j, x\rangle\Big\rangle = \sum_{i\in F} |\langle e_i, x\rangle|^2 \leq ||x||^2$$
la famille est donc sommable et on a l'inégalité de Bessel pour la somme (en passant au supremum) et on trouve en particulier $(\langle x, e_i\rangle)_{i\in I} \in \ell^2(I)$.

(2) Si $(e_i)_{i\in I}$ est une base soit $x_n \in Vect(e_i, i \in I)$ convergeant vers x.

De plus, pour n assez grand $|||x||^2 - ||x_n||^2| \leq \epsilon/2$ et pour tout J,

$$\begin{aligned}|||p_{V_J}(x)||^2 - ||p_{V_J}(x_n)||^2| &\leq ||p_{V_J}(x_n - x)||(||x_n|| + ||x||)\\ &\leq ||(x_n - x)||(||x_n|| + ||x||) \leq \epsilon/2\end{aligned}$$

d'où en prenant J tel que $p_{V_J}(x_n) = x_n$ on obtient

$$\left|\sum_{i\in J} |\langle e_j, x\rangle|^2 - ||x||^2\right| \leq \epsilon$$

et donc la somme de la série est $||x||$ d'où l'égalité de Parseval.

Réciproquement, Si on a égalité, on trouve J_n tel que

$$\sum_{j\in J_n} |\langle e_j, x\rangle|^2 = ||p_{V_{J_n}}(x)||^2 \to ||x||^2$$

et ceci implique par le théorème de Pythagore :

$$||p_{V_{J_n}}(x) - x||_2^2 = ||x||_2^2 - ||p_{V_{J_n}}(x)||_2^2 \to 0$$

donc tout élément de H est limite d'éléments de $Vect(e_i, i \in I)$ d'où la propriété de base hilbertienne.

De plus un calcul donne la formule pour x :

$$||x - \sum_{i\in F} e_i\langle e_i, x\rangle||^2 = \sum_{i\notin F} |\langle e_i, x\rangle|^2 \to 0.$$

(3) Considérons l'ensemble des familles orthonormales contenant une famille orthonormale donnée, et ordonné par inclusion. C'est un ensemble non-vide. Si on a une famille totalement ordonnée de familles orthonormales, l'union est un majorant, donc l'ensemble ordonné est inductif, il admet donc par le lemme de Zorn un élément maximal $(e_i)_{i\in I}$. Si ce n'était pas une base (complétant la famille orthonormale de départ), on aurait un x avec

$$\sum_{i\in I} |\langle x, e_i\rangle|^2 < ||x||^2.$$

Comme H est complet la somme $y = \sum_{i\in I} e_i\langle e_i, x\rangle$ converge car si (I_n) croissante telle que $\sum_{i\in I_n} |\langle e_i, x\rangle|^2 \to \sum_{i\in I} |\langle e_i, x\rangle|^2$ la suite $y_n = \sum_{i\in I_n} e_i\langle e_i, x\rangle$ est de Cauchy car pour $q > p$

$$||y_p - y_q||_2^2 = \sum_{i\in I_q - I_p} |\langle e_i, x\rangle|^2 \leq \sum_{i\notin I_p} |\langle e_i, x\rangle|^2 \to 0.$$

On déduit que $y-x$ est orthogonal à tout e_i car tout i tel que $\langle e_i, x\rangle \neq 0$ est dans un I_n et que $\langle y_n - x, e_i\rangle = 0$ pour n assez grand pour un tel i. Donc par orthogonalité

$$||y-x||_2^2 = ||x||^2 - \sum_{i\in I} |\langle x, x_i\rangle|^2 > 0$$

donc ajouter $(y-x)/||y-x||$ à la famille orthonormale contredit la maximalité et conclut.

Une fois l'existence d'une base, l'isométrie est évidente par le (2), et si on a une suite $(\lambda_i)_{i\in I}$ dans $\ell^2(I)$, on voit que $\sum \lambda_i e_i$ converge par complétude comme ci-dessus et on obtient ainsi la surjectivité. □

3 Correction de l'exercice sur les polynômes de Hermite

Soit $H = L^2(\mathbb{R}, \mu)$ l'espace de Hilbert réel des fonctions de carrés intégrables pour la mesure gaussienne standard $\mu(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, muni de la norme usuelle :

$$||f||_2 = \sqrt{\int_{\mathbb{R}} |f(x)|^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx}.$$

Soit

$$H_n(x) = (-1)^n \frac{e^{x^2/2}}{\sqrt{n!}} \left(\frac{d}{dx}\right)^n (e^{-x^2/2})$$

(et donc $H_0(x) = 1$)

1. Montrons par récurrence que pour $n \geq 1$, H_n est un polynôme de la forme :

$$\sqrt{n!}H_n(x) = x^n + \sum_{k=0}^{n-1} a_k x^k.$$

   En effet $H_1(x) = (-1)e^{x^2/2}(-xe^{-x^2/2}) = x$ et si on suppose l'hypothèse au rang n Or $\left(\frac{d}{dx}\right)(e^{-x^2/2}x^k) = -x^{k+1}e^{-x^2/2} + kx^{k-1}e^{-x^2/2}$ donc l'hyp de rec donne qui a la forme souhaitée.
2. Montrons que $(H_n)_{n\geq 0}$ est une famille orthonormale de H.
   On calcule pour $m \geq n$:

$$\langle H_n, H_m\rangle = (-1)^m \frac{1}{\sqrt{2\pi}\sqrt{m!}} \int H_n(x) \left(\frac{d}{dx}\right)^m (e^{-x^2/2}) dx$$

En intégrant par partie

$$\begin{aligned}\int H_n(x)\left(\frac{d}{dx}\right)^m (e^{-x^2/2})dx \\ &= [H_n(x)\left(\frac{d}{dx}\right)^{m-1} (e^{-x^2/2})]_{-\infty}^{\infty} \\ &\quad - \int H_n'(x)\left(\frac{d}{dx}\right)^{m-1} (e^{-x^2/2})dx\end{aligned}$$

le crochet est 0 vu que $P(x)e^{-x^2/2}$ pour P polynome tend vers 0 en $\pm\infty$.
Par induction si $m > n$

$$\langle H_n, H_m\rangle = (-1)^{m-n}\frac{1}{\sqrt{2\pi}\sqrt{m!}}\int H_n^{(n+1)}(x)\left(\frac{d}{dx}\right)^{m-n+1}(e^{-x^2/2})dx = 0$$

et si m= n vu $H_n^{(n)}(x) = \sqrt{n!}$ en appliquant le 1.

$$\begin{aligned}\langle H_n, H_n\rangle \\ &= \frac{(-1)^{m-n}}{\sqrt{2\pi}\sqrt{m!}}\int H_n^{(n)}(x)\left(\frac{d}{dx}\right)^{m-n}(e^{-x^2/2})dx \\ &= \frac{1}{\sqrt{2\pi}}\int e^{-x^2/2}dx = 1\end{aligned}$$

comme voulue.

4 Théorème d'injectivité de la transformée de Fourier

Définition E.1. La **fonction caractéristique** (f.c. ou transformée de Fourier) du v.a. $(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ est définie par

$$\Phi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) = \mathbf{E}[e^{i\langle t,X\rangle}],$$

pour tout $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$ et en notant le produit scalaire $\langle t, X\rangle := \sum_{i=1}^{n} t_i X_i$.

La fonction φ_X caractérise la loi de X par le théorème d'injectivité de la transformée de Fourier/ théorème d'inversion de la transformée de Fourier ci-dessous. On utilisera aussi plus tard au chapitre 2 la fonction caractéristique pour caractériser une notion de convergence, au chapitre 3 pour l'introduction des vecteurs gaussiens qui seront la base du chapitre 5 sur le mouvement brownien. C'est une notion FONDAMENTALE...

Lemme E.3. Soit $X \sim \mathcal{N}(m, \sigma^2)$ de loi normale alors $\Phi_X(t) = exp(-\frac{t^2\sigma^2}{2} + imt)$.

Démonstration. On a vu une preuve à l'exercice 8 du TD 3 de MASS 31 utilisant que la partie imaginaire est nulle par parité et le calcul de la partie réelle en établissant une équation différentielle par intégration dépendant d'un paramètre.

On donne ici une autre preuve par prolongement analytique. Par transfert, on doit montrer $\int \frac{1}{\sigma\sqrt{2\pi}} e^{ixt-\frac{(x-m)^2}{2\sigma^2}} = exp(-\frac{t^2\sigma^2}{2} + imt)$ en faisant le changement de variables $u = (x-m)/\sigma$ on se ramène au cas $\sigma = 1$, $m = 0$.

En prenant $m = z$ dans le calcul de la densité, on a pour $z \in \mathbb{R}$

$$\int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2+z^2-2xz}{2}} = \int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-z)^2}{2}} = 1.$$

Pour $z \in \mathbb{C}$, en appliquant le résultat précédent

$$\begin{aligned}\sum_{n=0}^{\infty} \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \frac{|zx|^n}{n!} e^{-\frac{x^2}{2}} &= \lim_{N\to\infty} \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{N} \frac{|zx|^n}{n!} e^{-\frac{x^2}{2}} \\ &\leq \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}+|zx|} \\ &\leq exp(\frac{|z|^2}{2}) < \infty\end{aligned}$$

La première bornitude permet d'appliquer le TCD pour les séries (ou Fubini pour la mesure discrète) et intervertir somme et série :

$$\sum_{n=0}^{\infty} z^n \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \frac{x^n}{n!} e^{-\frac{x^2}{2}} = \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}+zx}$$

la fonction de droite est donc la somme d'une série entière $exp(\frac{z^2}{2})$ pour $z \in \mathbb{R}$, donc par identification des coefficients, elle vaut cette valeur pour tout $z \in \mathbb{C}$, en particulier pour $z = it$ et on trouve le résultat. □

On démontrera le théorème suivant dans la prochaine section puisque la preuve utilise des propriétés générales de l'indépendance importante à noter pour elles-mêmes :

Théorème E.4 (Théorème d'injectivité de la transformation de Fourier)**.** Deux v.a. $(X_1, \ldots, X_n), (Y_1, \ldots, Y_n)$ tels que

$$\Phi_{(X_1,\ldots,X_n)}(t) = \Phi_{(Y_1,\ldots,Y_n)}(t) \forall t \in \mathbb{R}^n$$

sont égales en loi, c'est à dire :

$$P_{(X_1,\dots,X_n)} = P_{(Y_1,\dots,Y_n)}.$$

De plus, si $\Phi_X \in L^1(\mathbb{R}^n, Leb)$ alors $P_{(X_1,\dots,X_n)}$ a une densité par rapport à la mesure de Lebesgue donnée par (la transformée de Fourier inverse) qui est une fonction continue :

$$f_{(X_1,\dots,X_n)}(x) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \Phi_{(X_1,\dots,X_n)}(t)exp(-i\langle x, t\rangle)dt.$$

Sommes de variables aléatoires indépendantes (Rappels)

Vous avez probablement vu en TD de théorie de la mesure la définition de la convolution que l'on rappelle ici et relie aux sommes de variables aléatoires indépendantes.

Définition E.2 (Convolution)**.** Soit μ une mesure de Proba sur $S \subset \mathbb{R}^d$ et $f : \mathbb{R} \to \mathbb{R}$ une fonction mesurable telle que pour tout $x \in S$, $y \mapsto f(x-y)$ est dans $L^1(\mathbb{R}^d, \mu)$, la convolution de f et μ est la fonction $f * \mu$ définie par :

$$(f * \mu)(x) = \int_{\mathbb{R}^d} f(x-y)d\mu(y).$$

Si μ est absolument continue par rapport à la mesure de Lebesgue de densité g, on note aussi $f * g$.

Proposition E.5. Soient $X, Y : \Omega \to \mathbb{R}^d$ des v.a. indépendantes :

3. $\forall t \in \mathbb{R}^d$, $\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t)$
4. Si X_i, Y_i sont dans $L^2(\Omega)$, $Cov(X_i + Y_i, X_j + Y_j) = Cov(X_i, X_j) + Cov(Y_i, Y_j)$.
5. Si $P_X(dx) = f(x)dx$, $P_Y(dx) = g(y)dy$ alors P_{X+Y} est absolument continue par rapport à Lebesgue (sur $\mathbb{R}^d$) de densité $f * g$ définie Lebesgue p.p. :

$$P_{X+Y}(dz) = (f * g)(z)dz.$$

6. Si seulement X est de loi absolument continue mais de densité continue bornée f, alors quel que soit Y, P_{X+Y} est absolument continue par rapport à Lebesgue (sur $\mathbb{R}^d$) de densité $f * P_Y$ (définie partout). De plus, pour tout h continue bornée :

$$E((h * f)(Y)) = E(h(X+Y)).$$

Démonstration. 1. On a $\Phi_{X+Y}(t) = \mathbf{E}[e^{it(X+Y)}] = \mathbf{E}[e^{itX}e^{itY}] = \mathbf{E}[e^{itX}]\mathbf{E}[e^{itY}] = \Phi_X(t)\Phi_Y(t)$ l'avant dernière égalité par indépendance car $f(x) = e^{itx}$ est bornée donc intégrable (par rapport à une probabilité).

2. En général par bilinéarité $Cov(X_i + Y_i, X_j + Y_j) = Cov(X_i, X_j) + Cov(Y_i, Y_j) + Cov(Y_i, X_j) + Cov(Y_i, X_j)$, mais ici par indépendance les deux derniers termes sont nuls.

3.Il faut d'abord vérifier que $f * g$ est bien définie. Par Fubini–Tonelli vu le caractère positif :

$$\int_{\mathbb{R}^n} dx \int_{\mathbb{R}^n} dy f(x-y)g(y) = \int_{\mathbb{R}^n} dy (\int_{\mathbb{R}^n} dx f(x-y))g(y) = \int_{\mathbb{R}^n} dy g(y) = 1$$

donc $\int_{\mathbb{R}^n} dy f(x-y)g(y)$ existe et est fini p.p.

En prenant h mesurable positive et en appliquant le transfert, on obtient par changement de variables $z = x + y$ dans l'intégrale sur y obtenue par Fubini :

$$\begin{aligned} E(h(X+Y)) &= \int_{\mathbb{R}^{2d}} h(x+y)f(x)dxP_Y(dy) \\ &= \int_{\mathbb{R}^{2d}} h(z)f(z-y)dzP_Y(dy) = \int_{\mathbb{R}^d} h(z)(f * P_Y)(z)dz \end{aligned}$$

ce qui donne le calcul de densité (égalité de la loi avec seulement le cas $h = 1_B$). Dans le cas de 4. on raisonne pareil sauf que f continue bornée donne $x \mapsto f(x-y)$ intégrable par rapport à la proba P_Y directement. L'application de Fubini vient de $\int_{\mathbb{R}^{2d}} |h(z)f(z-y)|dzP_Y(dy) \leq ||h||_\infty$. L'égalité intermédiaire donne aussi $E(h(X+Y)) = \int_{\mathbb{R}^d}(h * f)(y)P_Y(dy) = E((h * f)(Y))$ par transfert. □

Preuve [Facultative] du Thm d'injectivité de la transformée de Fourier

On va utiliser les lois gaussiennes pour se ramener au cas avec densité tout en exploitant leurs propriétés de stabilité par cette transformée.

Lemme E.6. Soit g_σ la densité sur $\mathbb{R}^n$ d'un n–uplet de variable gaussienne i.i.d. $\mathcal{N}(0, \sigma^2)$. Pour tout $h : \mathbb{R}^n \to \mathbb{R}$ continue bornée, $(h * g_\sigma)(x) \to_{\sigma \to 0} h(x)$. On a même convergence uniforme sur tout compact.

En terme de convergence en loi, cela signifiera au chapitre 2 que si $(X_1(\sigma), \ldots, X_n(\sigma))$ sont les variables de densités g_σ, alors $x + (X_1(\sigma), \ldots, X_n(\sigma)) \to_{\sigma \to 0} x$ en loi en utilisant la proposition E.5.(4) au cas $Y = x$.

Démonstration. Par transfert et changement de variables

$$(h * g_\sigma)(x) - h(x) = \int_{\mathbb{R}^d} (h(x - \sigma z) - h(x)) g_1(z) dz.$$

En prenant, en prenant le supremum sur un compact K :

$$\sup_{x \in K} |(h * g_\sigma)(x) - h(x)| \le \int_{\mathbb{R}^d} \sup_{x \in K} |(h(x - \sigma z) - h(x))| g_1(z) dz$$

la limite vient de la convergence dominée par une constante $2||h||_\infty$ puisque une constante est intégrable par rapport à une probabilité comme $g_1(z)dz$, et la limite ponctuelle en z vient de la continuité de h qui est donc uniformément continue sur $K + B(0, |z|)$ et donc pour $|\sigma| < 1$, $x - \sigma z, x$ sont dans ce compact de distance $\sigma|z|$ tendant vers 0. Si h est uniformément continue sur $\mathbb{R}^d$ on a même convergence uniforme sur $\mathbb{R}^d$. □

On a aussi besoin de la conséquence suivante du lemme de classe monotone :

Proposition E.7. Soient $X, Y : \Omega \to \mathbb{R}^n$ des variables aléatoires. Les propriétés suivantes sont équivalentes

3. X, Y sont égales en loi : $P_X = P_Y$.
4. Pour tout $h : \mathbb{R}^n \to \mathbb{R}$, continue bornée, $\int h(X)dP = \int h(Y)dP$
5. Pour tout ouvert O de $\mathbb{R}^n$, $P_X(O) = P_Y(O)$.
6. pour tout $(x_1, \ldots, x_n) \in \mathbb{R}^n$:

$$P_X(]-\infty, x_1] \times \ldots \times]-\infty, x_n]) = P_Y(]-\infty, x_1] \times \ldots \times]-\infty, x_n]).$$

La fonction $F_X(x_1, \ldots, x_n) = P_X(]-\infty, x_1] \times \ldots \times]-\infty, x_n])$ appelée **fonction de répartition** caractérise donc la loi.

Démonstration. Les produits d'intervalles $]-\infty, x_1] \times \ldots \times]-\infty, x_n]$ et les ouverts sont des familles stables par intersection finie et engendrent la tribu des boréliens de $\mathbb{R}^n$ (car par intersection et complémentaire on obtient les boules carrées de la norme infini et que tout ouvert de $\mathbb{R}^n$ est union dénombrable de telles boules, de centre un point de $\mathbb{Q}^n$ par densité de $\mathbb{Q}^n$.) On applique donc le lemme de classe monotone pour obtenir les 2 dernières équivalences. 1 implique 2 vient du th de transfert plus bas comme l'équivalence de 2 avec : Pour tout $h : \mathbb{R}^d \to \mathbb{R}$, $\int_{\mathbb{R}^d} h(x)dP_X(x) = \int_{\mathbb{R}^d} h(x)dP_Y(x)$.

Pour montrer 3 à partir de 2 et conclure, il suffit de remarquer que $h_n(x) = \max(1, nd(., O^c))$ sont des fonctions continues bornées par 1 (car la distance à un fermé $x \mapsto d(x, O^c) = \inf\{d(x, y), y \in O^c\}$ est continue, cf. MASS 31). Si $x \in O^c$, $h_n(x) = 0$ et sinon, h_n est une suite croissante qui tend vers $h_n(x) \to 1_O(x)$ (car si $x \in O$, $nd(., O^c) \to \infty$

donc ≥ 1 pour n assez grand donc $h_n(x) = 1$ pour n assez grand). Donc par convergence monotone, $\int_{\mathbb{R}^d} h_n(x) dP_X(x) \to P_X(O)$ d'où l'égalité du 3. par celle du 2. $\square$

Preuve du Thm E.4. Pour montrer l'injectivité, par le lemme E.7, il suffit de montrer que l'égalité des transformée de Fourier implique égalité de $\mathbf{E}(h(X))$ pour tout h continue bornée.

Or par le lemme précédent, $(h * g_\sigma)(x) \to h(x)$ tout en étant borné par $||h||_\infty$ donc par TCD :

$$\mathbf{E}(h(X)) = \lim_{\sigma \to 0} \mathbf{E}((h * g_\sigma)(X)) = \lim_{\sigma \to 0} \mathbf{E}(h(X + Y_\sigma))$$

la dernière égalité avec Y_σ de densité g_σ et indépendant de X par la proposition E.5 (4) puisque la densité g_σ est continue bornée. Or la transformée de Fourier de $X + Y_\sigma$ est $\Phi_{X+Y_\sigma}(t) = \Phi_X(t)\Phi_{Y_\sigma}(t)$ par la proposition E.5 (2) et donc

$$\Phi_{X+Y_\sigma}(t) = \Phi_X(t) exp(-\frac{||t||_2^2 \sigma^2}{2})$$

par le calcul du lemme E.3. Comme ceci est intégrable, on s'attend à avoir la formule d'inversion de Fourier de la deuxième partie qui va donner $\mathbf{E}(h(X + Y_\sigma))$ en fonction de $\Phi_{X+Y_\sigma}(t)$, nous allons donc la montrer à la main dans ce cas pour conclure la preuve.

Or en interprétant la densité comme une variante de la transformée de Fourier dans le cas gaussien :

$$\begin{aligned}
(g_\sigma * P_X)(x) &= \int_{\mathbb{R}^d} \frac{1}{\sigma^d (2\pi)^{d/2}} exp(-\frac{||x-y||_2^2}{2\sigma^2}) P_X(dy) \\
&= \int_{\mathbb{R}^{2d}} P_X(dy) dv \frac{1}{\sigma^d (2\pi)^d} exp(-\frac{||v||^2}{2} + i\langle \frac{y-x}{\sigma}, v\rangle))
\end{aligned}$$

soit par le changement de variables $u = v/\sigma$ de jacobien σ^{-d} on obtient

$$\begin{aligned}
\mathbf{E}(h(X + Y_\sigma)) &= \int_{\mathbb{R}^d} dx h(x) (g_\sigma * P_X)(x) \\
&= \int_{\mathbb{R}^{3d}} dx P_X(dy) dv \\
&\qquad h(x) \frac{1}{(2\pi)^d} exp(-\frac{\sigma^2 ||v||^2}{2} + i\langle y - x, v\rangle))
\end{aligned}$$

soit en appliquant Fubini sur les intégrales en y, v

$$\begin{aligned}
\mathbf{E}(h(X + Y_\sigma)) &= \int_{\mathbb{R}^{2d}} dx dv \frac{h(x)}{(2\pi)^d} exp(-\frac{\sigma^2 ||v||^2}{2} - i\langle x, v\rangle)) \Phi_X(v) \\
&= \int_{\mathbb{R}^{2d}} dx dv \frac{h(x)}{(2\pi)^d} exp(-i\langle x, v\rangle)) \Phi_{X+Y_\sigma}(v)
\end{aligned}$$

qui est la formule souhaitée qui ne dépend bien que de la transformée de Fourier Φ_X et conclut l'injectivité.

Maintenant si Φ_X est intégrable $|h(x)\Phi_{X+Y_\sigma}(v)| \leq h(x)|\Phi_X(v)|$ est une domination (si h est à support compacte) et puisque $\Phi_{X+Y_\sigma}(v) \to_{\sigma\to 0} \Phi_X(v)$ par les formules précédentes, on obtient par le TCD la formule souhaitée pour la densité à la limite. La continuité de la densité vient du Théorème de continuité des intégrales à paramètres. On remarque qu'en utilisant $\mathbf{E}(h(X))) = \int_{\mathbb{R}^d} dxh(x)f_X(x)$ pour tout h positive continue à support compact, on déduit f_X positive (sinon par continuité elle est négative sur un ouvert dans lequel on peut prendre le support de h pour contredire positivité de l'intégrale) et par convergence monotone et faisant tendre $h \to 1$, on déduit f_X intégrable et densité de proba. D'où on peut utiliser $\mathbf{E}(h(X))) = \int_{\mathbb{R}^d} dxh(x)f_X(x)$ (maintenant valable pour h continue bornée car f_X peut servir de domination) pour identifier $P_X(dx) = f_X(x)dx$ en utilisant le lemme E.7. □

5 Théorème de Radon–Nikodym et Théorème de Dunford–Pettis (Niveau M1–M2)

Ce complément pourrait pour l'essentiel être ajouté comme application du théorème de Riesz ou du théorème de dualité des espaces L^p. Nous expliquons un théorème de théorie de la mesure qui permet de dire quand une mesure provient d'une densité dans $L^1(\Omega,\mu)$. On en déduit une application à un théorème de compacité qui est utile pour la preuve du cas uniformément continue du théorème de convergence des martingale dans L^1, le théorème de Dunford–Pettis E.9.

Définition E.3. Si μ, ν sont des mesures de probabilités sur $(\Omega, \mathcal{T})$, on dit que μ est absolument continue par rapport à ν et on note $\mu \ll \nu$ si pour tout $A \in \mathcal{T}$, $\nu(A) = 0$ implique que $\mu(A) = 0$

Définition E.4. Si μ, ν sont des mesures de probabilités sur $(\Omega, \mathcal{T})$, on dit que μ admet une densité $h \in L^1(\Omega, \nu)$ par rapport à ν et on note $h = \frac{d\mu}{d\nu}$, si $h \geq 0$ p.s. et pour tout $A \in \mathcal{T}$:

$$\int_\Omega 1_A h d\nu = \mu(A).$$

Les définitions s'étendent aux mesures σ–finies, mais on considère seulement ici le cas de probabilités.

Théorème E.8 (de Radon–Nikodym)**.** Pour toutes mesures de probabilités μ, ν sur $(\Omega, \mathcal{T})$, il y a équivalence entre $\mu \ll \nu$ et l'existence d'une densité $h = \frac{d\mu}{d\nu} \in L^1(\Omega, \nu)$ de μ par rapport à ν, et la densité est alors unique ν–p.s.

Démonstration. Si on a deux densités h, k, $\int_\Omega 1_A(h-k)d\nu = 0$ pour tout A $\mathcal{T}$ mesurable, donc par la construction de l'intégrale aussi $\int_\Omega fhd\nu = \int_\Omega fkd\nu$ d'abord pour f mesurable positive (par TCM) puis pour f mesurable bornée donc par dualité $h - k = 0$ dans $L^1(\Omega, \nu)$ donc ν–p.s.

De plus, si on a existence d'une densité et si $\nu(A) = 0$, par TCM, $\int_\Omega 1_A h = \lim_{n\to\infty} \int_\Omega 1_A(h \wedge n) = 0$ car $|\int_\Omega 1_A(h\wedge n)d\nu| \leq ||(h\wedge n)||_2||1_A||_2 \leq n\nu(A)^{1/2} = 0$ par Cauchy–Schwartz. Donc $\mu(A) = 0$ c'est à dire on a montré $\mu \ll \nu$.

La partie difficile est l'existence d'une densité si $\mu \ll \nu$. On va utiliser le théorème de représentation de Riesz (ou sa variante pour la dualité de L^1, le théorème D.8). Soit $\mu_\alpha = \mu + \alpha\nu$ avec $\alpha > 0$. L'idée est simple on s'attend à avoir une densité $\frac{d\mu_\alpha}{d\nu} = \alpha + h$ strictement positive et donc $\frac{d\nu}{d\mu_\alpha} = \frac{1}{\alpha+h}$ bornée par $1/\alpha$ donc dans L^2 ensuite $\alpha(1 - \frac{\alpha}{\alpha+h}) = \alpha\frac{h}{\alpha+h} \to_{\alpha\to\infty} h$ et on devrait pouvoir retrouver h ainsi.

Appliquons cette idée, si $f \in L^1(\Omega, d\mu_\alpha)$, on a

$$\int |f|d\nu = \frac{1}{\alpha}\int |f|d\alpha\nu \leq \frac{1}{\alpha}\int |f|d\mu_\alpha$$

Donc $f \in L^1(\Omega, d\nu)$ et $f \mapsto \int fd\nu$ définit une forme linéaire continue sur $L^1(\Omega, d\mu_\alpha)$, donc par le théorème D.8, il existe $h_\alpha \in L^\infty(\Omega, d\mu_\alpha)$ telle que pour tout $f \in L^1(\Omega, d\mu_\alpha)$ on a

$$\int fd\nu = \int fh_\alpha d\mu_\alpha.$$

Et de plus, on a $||h_\alpha||_{L^\infty(\mu_\alpha)} \leq 1/\alpha$. Si $f = 1_{\{h_\alpha<0\}}$, on obtient $\int \max(0, h_\alpha)d\mu_\alpha \geq 0$ donc vaut 0, donc

$$\nu(\{h_\alpha < 0\}) \leq \frac{1}{\alpha}\mu_\alpha(\{h_\alpha < 0\}) = 0$$

donc $h_\alpha \geq 0$, ν p.s.

On montre maintenant la monotonie attendue pour h_α (si on veut qu'elle soit égale à un $\frac{1}{\alpha+h}$) Si $\beta > \alpha$, on a pour f positive bornée en utilisant $\mu_\alpha(g) \leq \mu_\beta(g)$ pour g positive ν–p.s,

$$\int fh_\beta d\mu_\beta = \int fd\nu = \int fh_\alpha d\mu_\alpha \leq \int fh_\alpha d\mu_\beta$$

car fh_α posivite ν–p.s. par le résultat précédent, donc comme c'est valable pour tout $f \geq 0$, on a $h_\beta \leq h_\alpha \mu_\beta$–p.s. donc ν–p.s.

Finalement, on a l'identité
$$\int f d\mu = \int f d\mu_\alpha - \int f\alpha d\nu = \int f(1-\alpha h_\alpha)d\mu_\alpha = \int f\alpha(1-\alpha h_\alpha)d\nu + \int f(1-\alpha h_\alpha)d\mu.$$

Par $||h_\alpha||_{L^\infty(\mu_\alpha)} \leq 1/\alpha$. on a $1-\alpha h_\alpha \geq 0$ μ_α–p.s. donc ν–p.s. En raisonnant comme avant on obtient $(1-\alpha h_\alpha) \geq (1-\beta h_\beta)$ ν–p.s. Donc, par l'égalité précédente, après simplification de f (et toujours pour f positive en utilisant la croissance de $\alpha \to \alpha h_\alpha$ ν–p.s. par ce qu'on vient de voir donc μ–p.s. par l'hypothèse $\mu \ll \nu$) , on obtient
$$\int f\alpha(1-\alpha h_\alpha)d\nu = \int f\alpha h_\alpha d\mu \leq \int f\beta h_\beta d\mu = \int f\beta(1-\beta h_\beta)d\nu$$
soit $\alpha(1-\alpha h_\alpha) \leq \beta(1-\beta h_\beta)$, ν–p.s. donc converge vers un h en croissant et par convergence monotone et l'égalité avant on obtient
$$\int fhd\nu = \lim_{\alpha\to\infty}\int f\alpha(1-\alpha h_\alpha)d\nu = \lim_{\alpha\to\infty}\int fd\mu - \int f(1-\alpha h_\alpha)d\mu \leq \int fd\mu.$$
Donc pour $f=1$ on trouve $h \in L^1(\Omega, d\nu)$. Or par la monotonie de la limite définissant h, on a
$$(1-\alpha h_\alpha) = \frac{\alpha(1-\alpha h_\alpha)}{\alpha} \leq \frac{h}{\alpha} \to_{\alpha\to\infty} 0$$
ν–p.s. puisque h est fini ν–p.s. donc en utilisant encore l'hypothèse, aussi μ–p.s. Comme on a vu la monotonie en α par convergence monotone, on déduit $\int f(1-\alpha h_\alpha)d\mu \to 0$ et donc finalement l'égalité attendue qui conclut la preuve :
$$\int fhd\nu = \lim_{\alpha\to\infty}\int fd\mu - \int f(1-\alpha h_\alpha)d\mu = \int fd\mu.$$
□

On peut maintenant rappeler et prouver le théorème E.9 :

Théorème E.9 (Dunford–Pettis)**.** Soit une suite (X_n) dans $L^1(\Omega, \mathcal{T}, P)$ avec $\mathcal{T}$ une tribu dénombrablement engendrée (donc $\mathcal{T} = \mathcal{T}(\mathcal{E})$ avec $\mathcal{E}$ dénombrable, en particulier $\mathcal{T} = \mathcal{B}(\mathbb{R}^n)$). On a l'équivalence entre

3. (X_n) est uniformément intégrable
4. (X_n) admet une sous-suite (X_{n_k}) ayant pour limite faible $X \in L^1$, c'est-à-dire :
$$\forall f \in L^\infty(\Omega), \quad \mathbf{E}((X_{n_k} - X)f) \to 0.$$
5. (X_n) est bornée dans L^1 et pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que si $A \in \mathcal{T}$ vérifie $P(A) \leq \eta$ alors pour tout n, $\mathbf{E}(1_A|X_n|) \leq \epsilon$.

C'est surtout l'équivalence entre 1. et 2. qui est difficile et porte le nom de théorème de Dunford–Pettis. L'hypothèse “dénombrablement engendrée" n'est pas nécessaire (cf. Delacherie–Meyer **Probabilités et Potentiel** Vol 1 p 27) mais nous la faisons pour simplifier.

Démonstration. On commence par l'équivalence entre 1 et 3. Supposons 3. et fixons $\epsilon > 0$, η t.q. $P(A) \leq \eta$ implique $\mathbf{E}(1_A|X_n|) \leq \epsilon$. Par l'inégalité de Markov $P(|X_n| \geq c) \leq \frac{\sup_{n\in\mathbb{N}} \mathbf{E}(|X_n|)}{c} \leq \eta$ dès que $c \geq \frac{\sup_{n\in\mathbb{N}} \mathbf{E}(|X_n|)}{\eta}$, en appliquant alors à $A = \{|X_n| \geq c\}$, on déduit $\sup_n \mathbf{E}(1_{\{}|X_n| \geq c\}|X_n|) \leq \epsilon$. Et donc $\lim_{c\to\infty} \mathbf{E}(1_{\{}|X_n| \geq c\}|X_n|) = 0$ qui est l'uniforme intégrabilité recherchée.

Réciproquement, pour $\epsilon < 0$ fixé, on prend $c > 0$ tel que $\sup_n \mathbf{E}(1_{\{|X_n|\geq c\}}|X_n|) \leq \epsilon/2$, (en particulier

$$\mathbf{E}(|X_n|) = \mathbf{E}(1_{\{|X_n|\geq c\}}|X_n|) + \mathbf{E}(1_{\{|X_n|<c\}}|X_n|) \leq c + \epsilon/2$$

donc X_n et bornée dans L^1, de sorte que

$$\begin{aligned}\mathbf{E}(1_A|X_n|) &= \mathbf{E}(1_A 1_{\{|X_n|\geq c\}}|X_n|) + \mathbf{E}(1_A 1_{\{|X_n|<c\}}|X_n|)\\ &\leq \mathbf{E}(1_{\{|X_n|\geq c\}}|X_n|) + \mathbf{E}(1_A 1_{\{|X_n|<c\}} c) \leq \epsilon/2 + P(A)c\end{aligned}$$

qui est borné par ϵ dès que $P(A) \leq \eta = \epsilon/2c$ qui convient.

On suppose maintenant 3 et on montre 2. Si $\mathcal{T} = \sigma(A_n, n \in \mathbb{N})$, $\mathcal{A}$ l'algèbre engendré par les A_n c'est à dire les unions finis d'intersections finis de A_n, A_n^c (qui n'est en général pas une σ algèbres) qui est stable par, complémentaire union finie et intersection finie. Il est facile de voir que $\mathcal{A}$ est dénombrable.

En séparant les parties positives,négatives, on peut supposer $X_n \geq 0$ et par extraction diagonale, on trouve n_k telle que $\mathbf{E}[X_{n_k} 1_A] \to \mu(A)$ converge pour tout $A \in \mathcal{A}$.

Il est facile de voir que $\mu(\Omega) < \infty$ vu que (X_n) est bornée dans L^1 (par 3.) μ est additive sur les unions disjointes finies (par additivité de $1 \mapsto \mathbf{E}[X_{n_k} 1_A]$ qui est une mesure et passage à la limite). De plus, par 3., soit ϵ positive, on a un η tel que $P(A) \leq \eta$ implique $\mathbf{E}[X_{n_k} 1_A] \leq \epsilon$ donc $\mu(A) \leq \epsilon$.

En particulier si $P(A) = 0$, on a $\mu(A) = 0$.

Un résultat classique de théorie de la mesure dit que μ s'étend de façon unique sur $\sigma(\mathcal{A})$ en une mesure μ^* (cf. par exemple Barbe–Ledoux Th 1.49). Il est facile de voir que l'on a encore si $P(A) = 0$, on a $\mu^*(A) = 0$. Donc, $\mu^* \ll P$ et par le théorème de Radom–Nikodym, il existe $X \in L^1$ telle que $\mathbf{E}(X 1_A) = \mu(A) = \lim_{n\to_i nfty} \mathbf{E}[X_{n_k} 1_A]$. Il en est donc de même pour toute fonction étagée f_m (resp. g_m) d'une suite décroissante (resp. croissante) convergeant vers f mesurable positive bornée

D'où on a les deux inégalités donnant l'égalité

$$\limsup_{n\to\infty} \mathbf{E}[X_{n_k} f] \leq \lim_{n\to\infty} \mathbf{E}[X_{n_k} f_m] = \mathbf{E}(X f_m) \to \mathbf{E}(Xf)$$

$$\liminf_{n\to\infty} \mathbf{E}[X_{n_k} f] \geq \lim_{n\to\infty} \mathbf{E}[X_{n_k} g_m] = \mathbf{E}(X g_m) \to \mathbf{E}(Xf).$$

On a donc obtenu 2.

On laisse en exercice l'implication de 3. vers 1. que l'on n'a pas utilisé dans le cours.

$\square$