T. Blossier & M. Carrizosa& J. Melleray

Université Lyon I
Semestre d'automne 2017-2018

Table des matières

Chapitre 1

Espaces métriques

1.1 Distances

La notion de *distance* a été introduite pour formaliser les propriétés d'une façon de mesurer l'écart entre des éléments d'un même ensemble ; ces propriétés sont modelées sur celles de la longueur d'un vecteur dans $\mathbb{R}^2$ ou $\mathbb{R}^3$.

Définition 1.1. Soit X un ensemble. Une *distance* sur X est une fonction $d\colon X \times X \to [0, +\infty[$ satisfaisant les propriétés suivantes :

1. $\forall x, y \in X \quad d(x, y) = 0 \Leftrightarrow x = y$.
2. $\forall x, y \in X \quad d(x, y) = d(y, x)$.
3. $\forall x, y, z \in X \quad d(x, z) \leq d(x, y) + d(y, z)$.

On dit alors que (X, d) est un *espace métrique*.

La première des trois propriétés ci-dessus est appelée *axiome de séparation* : elle dit en particulier que deux points distincts sont nécessairement à distance strictement positive. La deuxième propriété est l'*axiome de symétrie*. Enfin, la troisième est appelée *inégalité triangulaire*. C'est peut-être la moins intuitive ; dans $\mathbb{R}^2$, muni de sa notion usuelle de distance, elle correspond au fait que la longueur d'un côté d'un triangle est toujours inférieure à la somme des longueurs des deux autres côtés.

Propriété 1.2. Dans un espace métrique (X, d), pour tout $x, y, z \in X$, on a $|d(x, z) - d(y, z)| \leq d(x, y)$. (Preuve en exercice.)

L'exemple le plus important d'espace métrique est $\mathbb{R}$ muni de la distance $d(x, y) = |x - y|$; plus généralement, presque toutes les distances que nous rencontrerons proviennent d'une *norme*, dont nous rappelons la définition maintenant.

Définition 1.3. Soit X un espace vectoriel sur $\mathbb{R}$ (resp. sur $\mathbb{C}$). Une *norme* sur X est une fonction $\|\cdot\|\colon X \to [0, +\infty[$ satisfaisant les conditions suivantes :

1. $\forall x \in X \quad \|x\| = 0 \Leftrightarrow x = 0$.
2. $\forall x, y \in X \quad \|x + y\| \leq \|x\| + \|y\|$.
3. $\forall \lambda \in \mathbb{R}$ (resp. $\in \mathbb{C}$) $\forall x \in X \quad \|\lambda x\| = |\lambda| \cdot \|x\|$.

On dit alors que $(X, \|\cdot\|)$ est un *espace normé*.

Exemple. La norme associée au produit scalaire d'un espace euclidien, appelée norme euclidienne.

Exemple. Pour $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, on pose

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \text{ et } \|x\|_\infty = \max(|x_1|, \ldots, |x_n|).$$

Les fonctions $\|\cdot\|_1$ et $\|\cdot\|_\infty$ sont des normes sur $\mathbb{R}^n$. (Preuve en exercice.)

Propriété 1.4. Si $(X, \|\cdot\|)$ est un espace normé, alors la fonction $d\colon X \times X \to [0, +\infty[$ définie par $d(x, y) = \|x - y\|$ est une distance sur X. (Preuve en exercice.)

On dira que d définie comme ci-dessus est la distance induite par $\|\cdot\|$. Ce sont les distances avec lesquelles nous travaillerons principalement ; l'ensemble X ne sera pas forcément un espace vectoriel tout entier, c'est pourquoi la remarque suivante sera importante : si (X, d) un espace métrique, et $A \subseteq X$, alors la restriction de d à A munit A d'une structure d'espace métrique.

Exercice 1.5. Soit X un ensemble. On définit une fonction $d\colon X \times X \to [0, +\infty[$ en posant

$$\forall x, y \in X \quad d(x, y) = \begin{cases} 0 & \text{si } x = y \\ 1 & \text{sinon} \end{cases} .$$

Montrer que d est une distance sur X, appelée *distance discrète*.

Définition 1.6. Soit (X, d) un espace métrique, $x \in X$ et $r \geq 0$. On définit :

— La *boule ouverte* de centre x et de rayon r par

$$B(x, r) = \{y \in X\colon d(x, y) < r\}.$$

— La *boule fermée* de centre x et de rayon r par

$$\overline{B}(x, r) = \{y \in X\colon d(x, y) \leq r\}.$$

Montrer que, s Pour $\mathbb{R}$ muni de la distance usuelle, $B(x, r) =]x - r, x + r[$ et $\overline{B}(x, r) = [x - r, x + r]$.

Exercice 1.7. Représenter les boules ouvertes/fermées dans $\mathbb{R}^2$ pour les distances associées aux normes $\|\cdot\|_1$, $\|\cdot\|_2$ et $\|\cdot\|_\infty$ de centre $(1, 0)$ et de rayon 1.

Exercice 1.8. Déterminer les boules ouvertes et les boules fermées d'un ensemble X muni de la distance discrète.

Proposition 1.9. *Soit (X, d_X) et (Y, d_Y) deux espaces métriques. Alors d_1 et d_∞ définies pour (x_1, y_1) et (x_2, y_2) dans $X \times Y$ par*

$$d_1((x_1, y_1), (x_2, y_2)) = d_X(x_1, x_2) + d_Y(y_1, y_2) \qquad \textit{et}$$
$$d_\infty((x_1, y_1), (x_2, y_2)) = \max(d_X(x_1, x_2), d_Y(y_1, y_2))$$

sont des distances sur $X \times Y$. On appellera d_∞ la distance produit *de d_X et d_Y.*

Preuve:

Les applications d_1 et d_∞ sont à valeurs dans $[0, +\infty[$ et satisfont de manière évidente les deux premières propriétés (axiome de séparation et axiome de symétrie).
Vérifions l'inégalité triangulaire : soit (x_1, y_1), (x_2, y_2) et (x_3, y_3) dans $X \times Y$. Alors,

$$\begin{aligned} d_1((x_1, y_1), (x_3, y_3)) = d_X(x_1, x_3) + d_Y(y_1, y_3) &\leq d_X(x_1, x_2) + d_X(x_2, x_3) + d_Y(y_1, y_2) + d_Y(y_1, y_3) \\ &= d_1((x_1, y_1), (x_2, y_2)) + d_1((x_2, y_2), (x_3, y_3)). \end{aligned}$$

On a

$$d_X(x_1, x_3) \leq d_X(x_1, x_2) + d_X(x_2, x_3) \leq \max(d_X(x_1, x_2), d_Y(y_1, y_2)) + \max(d_X(x_2, x_3), d_Y(y_2, y_3)) \ ,$$

d'où

$$d_X(x_1, x_3) \leq d_\infty((x_1, y_1), (x_2, y_2)) + d_\infty((x_2, y_2), (x_3, y_3)).$$

De même,

$$d_Y(y_1, y_3) \leq d_\infty((x_1, y_1), (x_2, y_2)) + d_\infty((x_2, y_2), (x_3, y_3)) \ .$$

Ainsi,

$$d_\infty((x_1, y_1), (x_3, y_3)) = \max(d_X(x_1, x_3), d_Y(y_1, y_3)) \leq d_\infty((x_1, y_1), (x_2, y_2)) + d_\infty((x_2, y_2), (x_3, y_3)) \ .$$

1.2 Suites dans un espace métrique

L'intérêt principal de la notion de distance, pour nous, est de pouvoir formaliser la notion de suites convergentes : une suite $(x_n)_{n\geq 0}$ converge vers x si et seulement si la distance $d(x_n, x)$ tend vers 0 quand n tend vers l'infini. Avec des quantificateurs, on obtient la définition suivante.

Définition 1.10. Soit (X, d) un espace métrique, $(x_n)_{n\geq 0}$ une suite d'éléments de X et $x \in X$. On dit que la suite $(x_n)_{n\geq 0}$ converge vers x si :

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n \geq N\ d(x_n, x) \leq \varepsilon\ ,$$

c'est-à-dire si :

$$d(x_n, x) \underset{n\to+\infty}{\longrightarrow} 0\ .$$

Propriété 1.11. Soit (X, d) un espace métrique et (x_n) une suite d'éléments de X, qui converge à la fois vers x et x'. Alors $x = x'$, autrement dit, la limite d'une suite convergente est unique. (Preuve en exercice.)

Proposition 1.12. *Soit (X, d_X) et (Y, d_Y) deux espaces métriques. On munit $X \times Y$ de la distance produit d_∞ définie à la Proposition 1.9. Alors une suite $(x_n, y_n)_{n\geq 0}$ d'éléments de $X \times Y$ converge vers (x, y) si, et seulement si, $(x_n)_{n\geq 0}$ converge vers x et $(y_n)_{n\geq 0}$ converge vers y.* (Preuve en exercice.)

Ce résultat reste vrai si on remplace d_∞ par d_1. (Preuve en exercice.)

Ainsi, dans $\mathbb{R}^n$ muni de la distance induite par $\|\cdot\|_\infty$, on retrouve le fait qu'une suite est convergente si et seulement si chaque suite de coordonnées converge.

A priori, la notion de convergence dépend de la distance considérée sur X ; mais il arrive que des distances différentes aient les mêmes suites convergentes.

Définition 1.13. Soit X un espace vectoriel, et $\|\cdot\|_1, \|\cdot\|_2$ deux normes sur X. On dit que $\|\cdot\|_1$ et $\|\cdot\|_2$ sont *équivalentes* s'il existe des constantes m, M strictement positives et telles que :

$$\forall x \in X \quad m\|x\|_1 \leq \|x\|_2 \leq M\|x\|_1\ .$$

Exercice 1.14. Montrer que la norme du sup et la norme euclidienne sur $\mathbb{R}^n$ sont équivalentes.

On reverra plus tard le théorème, normalement déjà connu et qu'il est en tout cas sans doute utile d'avoir en tête, selon lequel toutes les normes sur $\mathbb{R}^n$ sont équivalentes.

Cette définition a un analogue pour les distances :

Définition 1.15. Soit X un ensemble. Deux distances d_1, d_2 sur X sont dites *équivalentes* s'il existe des constantes m, M strictement positives et telles que :

$$\forall x, y \in X \quad md_1(x, y) \leq d_2(x, y) \leq Md_1(x, y)\ .$$

Bien sûr, si d_1, d_2 sont les distances associées à des normes équivalentes, alors elles sont elles-mêmes équivalentes. Cette définition est importante pour nous à cause de la propriété suivante.

Proposition 1.16. *Soit X un ensemble et d_1, d_2 deux distances sur X. Si d_1 et d_2 sont équivalentes, alors une suite $(x_n)_{n\geq 0}$ converge dans (X, d_1) si et seulement si elle converge dans (X, d_2).*

Autrement dit : deux distances équivalentes ont les mêmes suites convergentes.

Preuve:

Soit deux constantes m, M strictement positives telles que :

$$\forall x, y \in X \quad md_1(x, y) \leq d_2(x, y) \leq Md_1(x, y)\ .$$

Soit (x_n) une suite et x un élément de X. On a alors pour tout $n \in \mathbb{N}$,

$$md_1(x_n, x) \leq d_2(x_n, x) \leq Md_1(x_n, y)\ ,$$

ce qui entraîne que :

$$d_1(x_n, x) \underset{n\to+\infty}{\longrightarrow} 0 \text{ si et seulement si } d_2(x_n, x) \underset{n\to+\infty}{\longrightarrow} 0\ .$$

Quand on peut considérer des suites, on peut aussi considérer des suites extraites...

Définition 1.17. Soit X un ensemble, et $(x_n)_{n\geq 0}$ une suite d'éléments de X. Une *suite extraite* de $(x_n)_{n\geq 0}$ est une *sous-suite* de la forme $(x_{n_k})_{k\geq 0}$ où $(n_k)_{k\geq 0}$ est une suite strictement croissante d'entiers, ou, de manière équivalente, de la forme $(x_{\varphi(k)})_{k\geq 0}$, où $\varphi\colon \mathbb{N} \to \mathbb{N}$ est une fonction *strictement croissante* (il suffit de poser pour $k \geq 0$, $\varphi(k) = n_k$).

Intuitivement : une suite extraite est obtenue en ne gardant que certains termes de la suite $(x_n)_{n\geq 0}$ et en oubliant les autres ; par exemple, la suite $(x_{2k})_{k\geq 0}$ est une suite extraite de la suite $(x_n)_{n\geq 0}$.

Propriété 1.18. Soit $\varphi\colon \mathbb{N} \to \mathbb{N}$ une fonction strictement croissante. Alors pour tout $k \in \mathbb{N}$ on a $\varphi(k) \geq k$ (et en particulier $\varphi(k)$ tend vers $+\infty$!). (Preuve en exercice.)

Proposition 1.19. *Soit (X, d) un espace métrique. Si $(x_n)_{n\geq 0}$ est une suite convergente alors toute suite extraite $(x_{\varphi(k)})_{k\geq 0}$ est convergente.* (Preuve en exercice.)

Exercice 1.20. Soit (X, d) un espace métrique et (x_n) une suite d'éléments de X.

Supposons d'abord que les suites (x_{2k}) et (x_{2k+1}) convergent vers la même limite. Montrer que (x_n) est convergente. Ce résultat reste-t-il vrai si l'on ne suppose pas que les limites de (x_{2k}) et (x_{2k+1}) sont égales ?

Montrer que si (x_{2k}), (x_{2k+1}) et (x_{3k}) sont toutes les trois convergentes alors (x_n) est convergente.

Proposition 1.21. *Toute suite de réels admet une sous-suite monotone.*

Preuve:

Soit (x_n) une suite de réels. Pour cette preuve, on dira qu'un entier n est un pic, si pour tout $m > n$, $x_n > x_m$. S'il y a une infinité de pics $n_0 < n_1 < n_2 < \ldots < n_k < \ldots$, alors la suite extraite (x_{n_k}) est strictement décroissante.

Sinon, il existe un entier N tel qu'aucun entier $n \geq N$ n'est un pic. On construit alors par récurrence une suite strictement croissante d'entiers (n_k) tel que la sous-suite (x_{n_k}) soit croissante : pour l'initialisation, on pose $n_0 = N$. Supposons que l'on a choisit $N = n_0 < n_1 < \ldots < n_k$, tel que $x_{n_0} \leq \ldots \leq x_{n_k}$ (condition vide si $k = 0$). Comme $n_k \geq N$, n_k n'est pas un pic et par conséquent il existe $n_{k+1} > n_k$ tel que $x_{n_k} \leq x_{n_{k+1}}$.

1.3 Ouverts et fermés

Si on est intéressé plus par la notion de "proximité" induite par une distance que par les valeurs exactes de la distance, on est amené à la notion d'ouvert : $A \subseteq X$ est ouvert ssi, dès que $a \in A$, tout point suffisamment proche de a appartient à A. Autrement dit, A contient une petite boule non vide autour de chacun des ses points.

Définition 1.22. Soit (X, d) un espace métrique. Une partie A de X est un *ouvert* si

$$\forall a \in A\ \exists r > 0\ B(a, r) \subseteq A.$$

Une boule ouverte est bien un ouvert ! L'ensemble X tout entier et l'ensemble vide $\emptyset$ sont des ouverts. Dans $\mathbb{R}$, muni de la distance usuelle, tout intervalle ouvert est un ouvert (intervalles de la forme $]a, b[$, $]-\infty, a[$, $]b, +\infty[$ et $]-\infty, +\infty[$, avec $a, b \in \mathbb{R}$).

Exercice 1.23. Vérifier qu'une partie A d'un espace métrique est un ouvert si et seulement si

$$\forall a \in A\ \exists n \in \mathbb{N}^*\ \overline{B}(a, 1/n) \subseteq A.$$

Exercice 1.24. Vérifier qu'une partie A d'un espace métrique est un ouvert si et seulement si

$$\forall a \in A\ \exists r > 0\ \overline{B}(a, r) \subseteq A.$$

La notion d'ouvert dépend de la distance sur X ; par contre pour deux distances équivalentes les ouverts sont les mêmes.

Exercice 1.25. 1. Donner un exemple de deux distances sur $\mathbb{R}$ qui ne définissent pas les mêmes ouverts.

2. Soit X un ensemble et d_1, d_2 deux distances équivalentes sur X. Montrer qu'une partie A de X est un ouvert de (X, d_1) si et seulement si elle est un ouvert de (X, d_2).

Théorème 1.26. *Soit A une partie d'un espace métrique (X, d). Alors, A est un ouvert si et seulement pour toute suite (x_n) d'éléments de X qui converge vers un élément de A, alors x_n appartient à A pour tout n suffisamment grand.*

Preuve:

Supposons A ouvert et considérons une suite (x_n) d'éléments de X qui converge vers $a \in A$. Comme A est ouvert, il existe $r > 0$ tel que $B(a, r) \subseteq A$. La convergence de la suite vers a, entraîne qu'il existe un entier N tel que pour tout $n \geq N$, $d(x_n, a) < r$, c'est-à-dire tel que pour tout $n \geq N$, $x_n \in B(a, r) \subseteq A$.
Réciproquement, si A n'est pas ouvert, il existe $a \in A$ tel que pour tout $n \in \mathbb{N}$, $B(a, \frac{1}{n+1}) \not\subset A$. On choisit ainsi pour tout n, un élément $x_n \in B(a, \frac{1}{n+1}) \setminus A$. Alors aucun élément de la suite x_n n'appartient à A mais elle converge vers a qui appartient à A.

Définition 1.27. Une partie A d'un espace métrique est un *fermé* si son complémentaire $X \setminus A$ est ouvert.

Une boule fermée est bien un fermé ! L'ensemble X tout entier et l'ensemble vide sont des fermés. Dans $\mathbb{R}$, muni de la distance usuelle, tout intervalle fermé est un fermé.

Les fermés sont caractérisés par la propriété suivante : la limite de toute suite convergente d'éléments d'un fermé est dans ce fermé. Notons que toute propriété des ouverts se traduit, par passage au complémentaire, en une propriété des fermés.

Proposition 1.28. *Soit A une partie d'un espace métrique (X, d). Alors, A est un fermé si et seulement pour toute suite (x_n) d'éléments de A qui converge vers un élément x de X, alors x appartient à A.*

Preuve:

Supposons A fermé. Soit (x_n) une suite d'éléments de A qui converge vers un élément x. Comme aucun x_n n'est dans l'ouvert $X \setminus A$, (encore moins pour tout n suffisamment grand), d'après le théorème 1.26, x ne peut appartenir à l'ouvert $X \setminus A$ et donc appartient à A.
Réciproquement, si A n'est pas fermé, c'est-à-dire si $X \setminus A$ n'est pas ouvert, alors le théorème 1.26 garantit l'existence d'une suite (x_n) d'éléments n'appartenant pas à $X \setminus A$, c'est-à-dire appartenant à A, qui converge vers un élément de $X \setminus A$.

Proposition 1.29. *Dans un espace métrique (X, d) donné,*

1. *si $(O_i)_{i \in I}$ est une famille d'ouverts, alors $\bigcup\limits_{i \in I} O_i$ est également un ouvert (une* **union quelconque** *d'ouverts est un ouvert) ;*
2. *si $O_1, \ldots, O_n$ sont des ouverts, alors $O_1 \cap \ldots \cap O_n$ est également un ouvert (une* **intersection finie** *d'ouverts est un ouvert) ;*
3. *si $(F_i)_{i \in I}$ est une famille de fermés, alors $\bigcap\limits_{i \in I} F_i$ est également un fermé (une* **intersection quelconque** *de fermés est un fermé) ;*
4. *si $F_1, \ldots, F_n$ sont des fermés, alors $F_1 \cup \ldots \cup F_n$ est également un fermé (une* **union finie** *de fermés est un fermé).*

Preuve:

(1) Soit $a \in \bigcup\limits_{i \in I} O_i$, alors il existe $i_0 \in I$ tel que $a \in O_{i_0}$. Comme O_{i_0} est ouvert, il existe $r > 0$, tel que $B(a, r) \subseteq O_{i_0} \subseteq \bigcup\limits_{i \in I} O_i$, ce qui permet de conclure.
(2) Soit $a \in O_1 \cap \ldots \cap O_n$, alors pour chaque $i \in \{1, \ldots, n\}$, on a $a \in O_i$ qui est un ouvert. Ainsi, pour chaque $i \in \{1, \ldots, n\}$, il existe $r_i > 0$ tel que $B(a, r_i) \subseteq O_i$. En posant $r = \min(r_1, \ldots, r_n)$, on en déduit que la boule ouverte $B(a, r)$ est incluse dans chacun des O_i et donc que $B(a, r) \subseteq O_1 \cap \ldots \cap O_n$, ce qui permet de conclure.
(3) On a $X \setminus \bigcap\limits_{i \in I} F_i = \bigcup\limits_{i \in I} (X \setminus F_i)$, qui est ouvert par (1). Ainsi, $\bigcap\limits_{i \in I} F_i$ est fermé.
(4) On a $X \setminus (F_1 \cup \ldots \cup F_n) = (X \setminus F_1) \cap \ldots \cap (X \setminus F_n)$, qui est ouvert par (2), et donc $F_1 \cup \ldots \cup F_n$ est fermé.

Exercice 1.30.

1. Donner un exemple d'une intersection d'ouverts qui n'est pas un ouvert.
2. Donner un exemple d'une union de fermés qui n'est pas un fermé.

Définition 1.31. Étant donné un espace métrique (X, d), et une partie $A \subseteq X$, l'*intérieur* de A, dénoté $\mathring{A}$, est la réunion de tous les ouverts contenus dans A, c'est-à-dire

$$\mathring{A} = \bigcup_{\substack{O \text{ ouvert} \\ O \subseteq A}} O.$$

Par définition, $\mathring{A}$ est un ouvert, est contenu dans A, et il contient tous les autres ouverts contenus dans A : c'est le plus grand ouvert contenu dans A. Attention, $\mathring{A}$ peut tout à fait être vide même si A ne l'est pas !

Propriété 1.32. Soit A une partie d'un espace métrique (X, d). Alors pour tout $x \in X$, on a

$$x \in \mathring{A} \iff \exists r > 0 \text{ tel que } B(x, r) \subseteq A.$$

(Preuve en exercice.)

Définition 1.33. Étant donné un espace métrique (X, d), et une partie $A \subseteq X$, l'*adhérence* de A, notée $\overline{A}$, est l'intersection de tous les fermés contenant A, c'est-à-dire

$$\overline{A} = \bigcap_{\substack{F \text{ fermé} \\ F \supseteq A}} F.$$

Cette fois, $\overline{A}$ est un fermé, contient A, et est contenu dans tous les autres fermés qui contiennent A : c'est le plus petit de tous les fermés contenant A.

Propriété 1.34. Soit (X, d) un espace métrique, et $A \subseteq X$. On a $\overline{X \setminus A} = X \setminus \mathring{A}$ et $\overbrace{X \setminus A}^{\circ} = X \setminus \overline{A}$. (Preuve en exercice.)

Propriété 1.35. Soit (X, d) un espace métrique, $A \subseteq X$ et $x \in X$. Il existe une suite d'éléments de A qui converge vers x si, et seulement si, $B(x, r) \cap A \neq \emptyset$ pour tout $r > 0$. (Preuve en exercice.)

Ceci nous permet d'établir la caractérisation suivante de l'adhérence, qui est fondamentale.

Proposition 1.36. *Soit (X, d) un espace métrique, A une partie de X et $x \in X$. Alors $x \in \overline{A}$ si, et seulement si, il existe une suite d'éléments de A qui converge vers x.*

Preuve:

Supposons qu'il existe une suite d'éléments de A qui converge vers $x \in X$. Alors, pour tout fermé F contenant A, il existe aussi une suite (la même !) d'éléments de F qui converge vers x ; par conséquent, x appartient à F. Donc x appartient à tous les fermés qui contiennent A : $x \in \overline{A}$.
Réciproquement, supposons qu'il n'existe pas de suite d'éléments de A qui converge vers $x \in X$. Alors, d'après la propriété 1.35, il doit exister $r > 0$ tel que $B(x, r) \cap A = \emptyset$. Autrement dit, $X \setminus B(x, r)$ est un fermé de X, contenant A, mais auquel x n'appartient pas : $x \notin \overline{A}$.

Exercice 1.37. Déterminer l'intérieur et l'adhérence de $\mathbb{Q}, [0, 1] \cap \mathbb{Q},]0, 1[\cap \mathbb{Q}$ (dans $\mathbb{R}$ muni de sa distance usuelle).

Proposition 1.38. *Soit $(E, \|\cdot\|)$ un espace vectoriel normé (on considère E muni de la distance induite par sa norme). Soit $a \in E$ et $r > 0$.*

1. *L'intérieur de la boule fermée $\overline{B}(a, r)$ est la boule ouverte $B(a, r)$.*
2. *L'adhérence de la boule ouverte $B(a, r)$ est la boule fermée $\overline{B}(a, r)$.*

(Preuve en exercice.)

Exercice 1.39. Les deux propriétés de la proposition ci-dessus sont elles vraies pour tout espace métrique ?

Définition 1.40. Soit (X, d) un espace métrique, et $A \subseteq X$. On dit que A est *dense* dans X si $\overline{A} = X$.

Par exemple, il est bien connu que $\mathbb{Q}$ et $\mathbb{R} \setminus \mathbb{Q}$ sont tous les deux denses dans $\mathbb{R}$.

Chapitre 2

Fonctions continues entre espaces métriques

2.1 Fonctions continues et uniformément continues

Maintenant qu'on sait ce qu'est une distance, on peut définir la continuité pour des fonctions entre espaces métriques, plutôt que de $\mathbb{R}$ dans $\mathbb{R}$; c'est essentiellement la même chose, en remplaçant $|x-y|$ (qui n'a a priori pas de sens dans un espace métrique) par $d(x,y)$.

Définition 2.1. Soit (X,d) et (Y,D) deux espaces métriques, $f\colon X\to Y$ et $x\in X$. On dit que f est *continue en* x si :

$$\forall\varepsilon>0\ \exists\delta>0\ \forall x'\in X\ d(x,x')<\delta\Rightarrow D(f(x),f(x'))<\varepsilon\ .$$

On dit que f est *continue sur* X si elle est continue en x pour tout $x\in X$, autrement dit :

$$\forall x\in X\ \forall\varepsilon>0\ \exists\delta>0\ \forall x'\in X\ d(x,x')<\delta\Rightarrow D(f(x),f(x'))<\varepsilon\ .$$

Ou encore (l'ordre dans lequel on écrit les deux $\forall$ ne change pas le sens de l'énoncé) :

$$\forall\varepsilon>0\ \forall x\in X\ \exists\delta>0\ \forall x'\in X\ d(x,x')<\delta\Rightarrow D(f(x),f(x'))<\varepsilon\ .$$

Il faut bien comprendre que, ci-dessus, δ dépend de ε *et du point* x *où l'on se place*. Une définition plus forte imposerait que le même δ fonctionne pour tous les $x\in X$ simultanément ; dans ce cas, on dit que f est *uniformément continue*.

Définition 2.2. Soit (X,d) et (Y,D) deux espaces métriques, et $f\colon X\to Y$. On dit que f est *uniformément continue sur* X si

$$\forall\varepsilon>0\ \exists\delta>0\ \forall x\in X\ \forall x'\in X\ d(x,x')<\delta\Rightarrow D(f(x),f(x'))<\varepsilon\ .$$

Par rapport à la définition de la continuité, on a remplacé "$\forall x\in X\ \exists\delta>0$" par "$\exists\delta>0\ \forall x\in X$" : δ dépend toujours de ε, mais ne dépend plus de x. Toute fonction uniformément continue est continue, mais la réciproque est fausse.

Exercice 2.3. Montrer que la fonction $x\mapsto x^2$ n'est pas uniformément continue sur $\mathbb{R}$.

Exercice 2.4. Pour chacun des énoncés suivants, déterminer toutes les fonctions $f\colon\mathbb{R}\to\mathbb{R}$ qui le satisfont :

1. $\exists\delta>0\ \forall\varepsilon>0\ \forall x\in\mathbb{R}\ \forall x'\in\mathbb{R}\ |x-x'|<\delta\Rightarrow|f(x)-f(x')|<\varepsilon$.
2. $\forall\varepsilon>0\ \forall x\in\mathbb{R}\ \forall x'\in\mathbb{R}\ \exists\delta>0\ |x-x'|<\delta\Rightarrow|f(x)-f(x')|<\varepsilon$.

Définition 2.5. Soit (X,d) et (Y,D) deux espaces métriques. On dit que $f\colon X\to Y$ est *lipschitzienne* s'il existe $K>0$ tel que

$$\forall x,x'\in X\ D(f(x),f(x'))\le Kd(x,x')\ .$$

Propriété 2.6. Toute fonction lipschitzienne est uniformément continue. (Preuve en exercice.)

Exercice 2.7. Soit $f\colon \mathbb{R} \to \mathbb{R}$ une fonction dérivable, et telle qu'il existe M satisfaisant $|f'(x)| \le M$ pour tout $x \in \mathbb{R}$. Montrer que f est lipschitzienne.

Théorème 2.8. *Soit (X,d) et (Y,D) deux espaces métriques, $f\colon X \to Y$ une fonction et $x \in X$. Les propriétés suivantes sont équivalentes :*

— *f est continue en x.*

— *Pour toute suite (x_n) d'éléments de X qui converge vers x, la suite $(f(x_n))$ converge vers $f(x)$.*

Preuve:

Supposons tout d'abord que f est continue en x, et fixons une suite (x_n) qui converge vers x ainsi que $\varepsilon > 0$. D'une part il existe $\delta > 0$ tel que $D(f(x), f(x')) \le \varepsilon$ dès que $d(x,x') < \delta$; d'autre part il existe $N \in \mathbb{N}$ tel que $d(x_n, x) \le \delta$ pour tout $n \ge N$. Alors, pour tout $n \ge N$, on a $d(f(x_n), f(x)) \le \varepsilon$, ce qui prouve que $(f(x_n))$ converge vers $f(x)$.

Réciproquement, supposons que f ne soit pas continue en x :

$$\exists \varepsilon > 0 \ \forall \delta > 0 \ \exists y \in X \ d(x,y) < \delta \text{ et } d(f(x), f(y)) \ge \varepsilon \ .$$

Fixons $\varepsilon > 0$ comme ci-dessus, et appliquons la propriété pour $\delta = \frac{1}{n}$: ceci nous donne une suite (y_n) telle que $d(x, y_n) < \frac{1}{n}$ pour tout $n \in \mathbb{N}^*$ (en particulier, (y_n) converge vers x) mais $d(f(y_n), f(x)) \ge \varepsilon$ (par conséquent, $f(y_n)$ ne converge pas vers $f(x)$).

Théorème 2.9. *Soit (X,d) et (Y,D) deux espaces métriques et $f\colon X \to Y$ une fonction. Les propriétés suivantes sont équivalentes :*

1. *f est continue.*
2. *Pour tout ouvert O de Y, $f^{-1}(O)$ est un ouvert de X.*
3. *Pour tout fermé F de Y, $f^{-1}(F)$ est un fermé de X.*

On rappelle que $f^{-1}(A) = \{x \in X \colon f(x) \in A\}$ désigne l'*image inverse* de A par f.

Preuve:

Supposons que f est continue, et soit O un ouvert de Y. Fixons $x \in f^{-1}(O)$, et considérons une suite (x_n) qui tend vers x. Alors $f(x_n)$ tend vers $f(x)$ puisque f est continue, donc $f(x_n)$ appartient à O pour n suffisamment grand puisque $f(x) \in O$ et O est ouvert. Par conséquent, $x_n \in f^{-1}(O)$ pour n suffisamment grand, ce qui nous montre que $f^{-1}(O)$ est ouvert, et on a montré que (1) $\Rightarrow$ (2).

Si (2) est vrai et F est fermé dans Y, alors $Y \setminus F$ est ouvert et par hypothèse on obtient que $f^{-1}(Y \setminus F) = X \setminus f^{-1}(F)$ est ouvert dans X, autrement dit $f^{-1}(F)$ est fermé dans X. Ceci établit l'implication (2) $\Rightarrow$ (3), et en fait le même argument de passage au complémentaire donne l'implication réciproque (3) $\Rightarrow$ (2).

Il nous reste à prouver que (2) $\Rightarrow$ (1) ; supposons donc de nouveau que (2) soit vérifié, et considérons $x \in X$ et $\varepsilon > 0$. Puisque $B(f(x), \varepsilon)$ est un ouvert contenant $f(x)$, son image inverse est par hypothèse un ouvert contenant x, par conséquent il existe $\delta > 0$ tel que $B(x,\delta) \subseteq f^{-1}(B(f(x),\varepsilon))$, c'est-à-dire :

$$\forall x' \in X \ d(x,x') < \delta \to D(f(x), f(x')) < \varepsilon \ .$$

On a bien montré que f est continue.

On voit dans cette preuve qu'il vaut mieux être à l'aise avec les propriétés de l'image inverse par une fonction... Ce sera aussi très important dans la partie du cours consacrée à la théorie de la mesure !

Exercice 2.10. Soit X, Y deux ensembles, $f\colon X \to Y$ une fonction. Montrer que, pour tout $A, B \subseteq Y$ on a $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$ et $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.

Exercice 2.11. Soit l'application de $\mathbb{R}$ dans $\mathbb{R}$, $f\colon x \mapsto x^2$.

1. Déterminer les ensembles suivants : $f([-3,-1])$, $f([-2,1])$, $f([-3,-1] \cup [-2,1])$ et $f([-3,-1] \cap [-2,1])$. Les comparer.
2. Mêmes questions avec les ensembles $f^{-1}(]-\infty, 2])$, $f^{-1}([1,+\infty[)$, $f^{-1}(]-\infty,2] \cup [1,+\infty[)$ et $f^{-1}(]-\infty,2] \cap [1,+\infty[)$.

Proposition 2.12. *Soit (X, d_X), (Y, d_Y) et (Z, d_Z) trois espaces métriques, ainsi que $f\colon Y \to Z$ et $g\colon X \to Y$ deux fonctions continues. Alors $f \circ g\colon X \to Z$ est continue.*

Preuve:

Fixons $\varepsilon > 0$. Comme f est continue, il existe $\delta_1 > 0$ tel que pour tout $y, y' \in Y$ on ait $d_Y(y, y') < \delta_1 \Rightarrow d_Z(f(y), f(y')) < \varepsilon$. Puis, comme g est continue, il existe δ_2 tel que pour tout $x, x' \in X$ on ait $d_X(x, x') < \delta_2 \Rightarrow d_Y(g(x), g(x')) < \delta_1$. On a alors, pour tout $x, x' \in X$:

$$d_X(x, x') < \delta_2 \Rightarrow d_Y(g(x), g(x')) < \delta_1 \Rightarrow d_Z(f(g(x)), f(g(x')) < \varepsilon \ .$$

On vient de prouver que $f \circ g$ est continue.

Exercice 2.13. On munit $\mathbb{R}^2$ de la distance induite par $\| \cdot \|_\infty$, et $\mathbb{R}$ de sa distance usuelle. Montrer que les fonctions $(x, y) \mapsto x + y$ et $(x, y) \mapsto xy$ sont continues.

Exercice 2.14. Soit (X, d) un espace métrique et $f, g \colon X \to \mathbb{R}$ deux fonctions continues. Montrer que la somme $f + g$ et le produit fg sont également des fonctions continues.

Exercice 2.15. Soit (X, d_X), (Y, d_Y) et (Z, d_Z) trois espaces métriques, ainsi que $f \colon Y \to Z$ et $g \colon X \to Y$ deux fonctions uniformément continues. Montrer que $f \circ g \colon X \to Z$ est uniformément continue.

2.2 Suites de fonctions

Tout comme la continuité, les notions de convergence simple/uniforme de suites de fonctions qu'on connaît pour des fonctions de $\mathbb{R}$ dans $\mathbb{R}$ s'étendent sans difficultés aux fonctions entre espaces métriques.

Définition 2.16. Soit (X, d), (Y, D) deux espaces métriques, et (f_n) une suite de fonctions de X dans Y. On dit que (f_n) converge *simplement* vers une fonction $f \colon X \to Y$ si pour tout $x \in X$ la suite $(f_n(x))$ converge vers $f(x)$; autrement dit :

$$\forall x \in X \ \forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n \geq N \ D(f_n(x), f(x)) < \varepsilon \ .$$

Ci-dessus, N dépend à la fois de ε et de x ; comme dans la définition de la continuité, on pourrait demander que N ne dépende que de ε, et on obtient ainsi la définition de la convergence *uniforme*.

Définition 2.17. Soit (X, d), (Y, D) deux espaces métriques, et (f_n) une suite de fonctions de X dans Y. On dit que (f_n) converge *uniformément* vers une fonction $f \colon X \to Y$ si

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall x \in X \ \forall n \geq N \ D(f_n(x), f(x)) < \varepsilon \ .$$

Bien entendu, la convergence uniforme entraîne la convergence simple. La réciproque est fausse, comme le montre l'exercice suivant.

Exercice 2.18. Pour tout $n \in \mathbb{N}^*$, on définit $f_n \colon [0, 1] \to [0, 1]$ en posant

$$f_n(x) = \begin{cases} 0 & \text{si } x \geq \frac{1}{n} \\ 1 - nx & \text{si } 0 \leq x \leq \frac{1}{n} \end{cases} .$$

Pour $n \in \mathbb{N}^*$, représenter le graphe de la fonction f_n, puis montrer que (f_n) converge simplement vers une fonction f que l'on déterminera. La convergence est-ele uniforme ? f est-elle continue ?

On voit donc que la convergence simple ne préserve pas la continuité (ce qui sera une bonne raison, plus tard, pour travailler avec des fonctions *mesurables* plutôt que des fonctions continues).

Proposition 2.19. *Soit (X, d) et (Y, D) deux espaces métriques, et (f_n) une suite de fonctions continues de X dans Y. Si (f_n) converge uniformément vers $f \colon X \to Y$ alors f est continue.*

Preuve:

Fixons $\varepsilon > 0$ et $x \in X$. Il existe $N \in \mathbb{N}$ tel que, pour tout $n \geq N$, on ait $D(f_n(x), f(x)) \leq \varepsilon$ pour tout $x \in X$.
Fixons un tel N ; comme f_N est continue, il existe $\delta > 0$ tel que pour tout $x' \in X$,

$$d(x, x') < \delta \Rightarrow D(f_N(x), f_N(x')) < \varepsilon.$$

Alors on a, pour tout $x' \in X$ tel que $d(x, x') < \delta$:

$$\begin{aligned} D(f(x), f(x')) &\leq D(f(x), f_N(x)) + D(f_N(x), f_N(x')) + D(f_N(x'), f(x')) \\ &\leq \varepsilon + \varepsilon + \varepsilon \\ &= 3\varepsilon\ . \end{aligned}$$

Comme ε était quelconque, ceci suffit à démontrer que f est continue en x.

Propriété 2.20. Une limite uniforme de fonctions uniformément continues est uniformément continue. (Preuve en exercice.)

Chapitre 3

Compacité

Définition 3.1. Soit (X, d) un espace métrique. On dit que (X, d) est *compact* s'il a la propriété suivante : pour toute suite (x_n) d'éléments de X, il existe une sous-suite (x_{n_k}) qui converge dans X.

Un exemple fondamental d'espace compact est donné par un intervalle fermé borné (un *segment*) de $\mathbb{R}$ ou, plus généralement n'importe quelle partie fermée bornée de $\mathbb{R}$. On verra plus loin, Théorème 3.9 et Corollaire 3.13, qu'en fait les compacts de $\mathbb{R}^n$ sont exactement ses parties fermées bornées.

Théorème 3.2. *Toute partie fermée bornée de $\mathbb{R}$, en particulier tout segment, est compacte.*

Preuve:

Considérons une suite (x_n) d'éléments d'une partie F fermée bornée de $\mathbb{R}$. Par la proposition 1.21, il existe une suite extraite (x_{n_k}) qui est monotone. Comme cette sous-suite est bornée et monotone, elle converge. Sa limite est dans F car F est fermé.

Proposition 3.3. *Soit (X, d), (Y, D) deux espaces métriques compacts. Alors $X \times Y$, muni de la distance produit, est encore un espace métrique compact.*

Preuve:

Soit (x_n, y_n) une suite d'éléments de $X \times Y$. Par compacité de (X, d), on peut extraire une sous-suite $(x_{\varphi_1(k)})$ qui converge vers $x \in X$. Et par compacité de (Y, D), on peut extraire de $(y_{\varphi_1(k)})$ [i] une nouvelle sous-suite $y_{\varphi_1(\varphi_2(l))}$ [ii] qui converge vers $y \in Y$.
Notons $\psi = \varphi_1 \circ \varphi_2$. Alors la suite $(x_{\psi(l)}, y_{\psi(l)})$ est telle que $x_{\psi(l)}$ converge vers x et $y_{\psi(l)}$ converge vers y, autrement dit, cette suite est une suite extraite de (x_n, y_n) qui converge vers (x, y).

Corollaire 3.4. Pour tout $n \in \mathbb{N}^*$, et tout $a_1, \dots, a_n, b_1, \dots, b_n$, l'ensemble $\prod [a_i, b_i]$ est un compact de $\mathbb{R}^n$ muni de la distance induite par $\| \cdot \|_\infty$.

Preuve:

Chacun des espaces $[a_i, b_i]$ est compact, et par récurrence on montre facilement à partir de la proposition précédente qu'un produit fini d'espaces métriques compacts est compact.

Théorème 3.5. *Soit (X, d) un espace métrique, et A un sous-ensemble compact de X (c'est-à-dire que (A, d) est un espace métrique compact). Alors A est fermé dans X.*

Preuve:

Soit (x_n) une suite d'éléments de A qui converge vers $x \in X$; on doit prouver que $x \in A$. Comme A est compact, il existe une sous-suite (x_{n_k}) qui converge vers $a \in A$; et (x_{n_k}) converge toujours vers x. Par unicité de la limite, $x = a \in A$.

Réciproquement, on a le résultat suivant.

i. Pourquoi n'extrait-on pas de (y_n) ?
ii. Pourquoi $\varphi_1(\varphi_2(l))$ et pas $\varphi_2(\varphi_1(k))$?

Proposition 3.6. *Soit (X,d) un espace métrique compact et A une partie fermée de X. Alors (A,d) est un espace métrique compact.*

Preuve:

> Soit (a_n) une suite d'éléments de A. Comme X est compact et (a_n) est aussi une suite d'éléments de X, il existe une sous-suite (a_{n_k}) qui converge vers $x \in X$; comme A est fermé, $x \in A$, ce qui montre que (a_n) a une sous-suite convergente dans A : (A,d) est compact.

La compacité est importante pour nous en particulier parce que les fonctions continues sur les espaces compacts ont des propriétés très fortes.

Théorème 3.7. *Soit (X,d) un espace métrique compact, et $f\colon X \to \mathbb{R}$ une fonction continue. Alors f est bornée et atteint ses bornes.*

Preuve:

> Montrons que f atteint sa borne supérieure $M = \sup(\{f(x) \colon x \in X\})$ (l'argument pour la borne inférieure est symétrique, ou découle du résultat pour la borne supérieure appliqué à $-f$). Par définition d'une borne supérieure, il existe une suite (x_n) d'éléments de X telle que $f(x_n)$ converge vers M. Comme (X,d) est compact, (x_n) admet une sous-suite convergente (x_{n_k}) ; appelons x sa limite. Alors $f(x_{n_k})$ converge à la fois vers M (c'est une sous-suite d'une suite qui converge vers M) et vers $f(x)$ (par continuité de f). Par conséquent, $f(x) = M$.

La proposition suivante est une conséquence facile de ce théorème.

Proposition 3.8. *Tout espace métrique compact (X,d) est* borné*, c'est-à-dire qu'il existe M tel que pour tout $x, x' \in X$ on ait $d(x,x') \leq M$.*

Preuve:

> La fonction $d\colon X \times X \to \mathbb{R}$ est continue lorsqu'on munit $X \times X$ de la distance produit D ; en effet, elle est lipschitzienne :
>
> $$\begin{aligned} |d(x_1,x_2) - d(x_1',x_2')| &= |d(x_1,x_2) - d(x_1,x_2') + d(x_1,x_2') - d(x_1',x_2')| \\ &\leq |d(x_1,x_2) - d(x_1,x_2')| + |d(x_1,x_2') - d(x_1',x_2')| \\ &\leq d(x_2,x_2') + d(x_1,x_1') \\ &\leq 2D((x_1,x_2),(x_1',x_2')). \end{aligned}$$
>
> Comme $(X \times X, D)$ est compact, la proposition précédente nous permet de conclure que d est bornée, ce qui revient à dire que (X,d) est un espace métrique borné.

Théorème 3.9. *Dans $\mathbb{R}^n$ muni de la distance d_∞ induite par $\|\cdot\|_\infty$, les compacts sont les fermés bornés.*

Preuve:

> Si $A \subseteq \mathbb{R}^n$ est tel que (A,d) soit compact, alors on sait que (A,d) doit être fermé dans $\mathbb{R}^n$, et borné d'après la proposition précédente.
>
> Réciproquement, si A est fermé borné dans $\mathbb{R}^n$, alors il existe M tel que A soit contenu dans $[-M,M]^n$; on a vu que cet ensemble est compact, et A y est fermé, donc (A,d) est compact.

Théorème 3.10. *Soit (X,d) un espace métrique compact, (Y,D) un espace métrique, et $f\colon X \to Y$ une fonction continue. Alors $f(X)$ est un sous-ensemble compact de Y.*

Notons que, une fois qu'on sait que les sous-ensembles compacts de $\mathbb{R}$ sont les fermés bornés, ce résultat généralise le théorème 3.7.

Preuve:

> Soit (y_n) une suite d'éléments de $f(X)$. Pour tout n on peut choisir x_n tel que $f(x_n) = y_n$, et ensuite on peut extraire une sous-suite convergente (x_{n_k}) de la suite (x_n). Par continuité de f, la suite $y_{n_k} = f(x_{n_k})$ converge vers $y = f(x)$.

Théorème 3.11 (Théorème de Heine)**.** *Soit (X,d) un espace métrique compact, (Y,D) un espace métrique, et $f\colon X \to Y$ une fonction continue. Alors f est uniformément continue.*

Preuve:

On va montrer la contraposée : si f n'est pas uniformément continue, alors f n'est pas continue. Supposons donc que f ne soit pas uniformément continue, c'est-à-dire qu'il est faux que

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x, x' \in X\ d(x, x') < \delta \Rightarrow D(f(x), f(x')) < \varepsilon\ .$$

Autrement dit, notre hypothèse est que

$$\exists \varepsilon > 0\ \forall \delta > 0\ \exists x, x' \in X\ d(x, x') < \delta \text{ et } D(f(x), f(x')) \geq \varepsilon\ .$$

Fixons $\varepsilon > 0$ comme ci-dessus. Pour tout $n \in \mathbb{N}^*$, on sait qu'il existe x_n, x'_n tels que $d(x_n, x'_n) < \frac{1}{n}$ et $D(f(x_n), f(x'_n)) \geq \varepsilon$.
Comme l'espace produit $X \times X$ est compact, on peut extraire une sous-suite convergente (x_{n_k}, x'_{n_k}) de la suite (x_n, x'_n) ; la suite x_{n_k} converge vers $x \in X$, et la suite x'_{n_k} converge vers $x' \in X$.
Puisque $n_k \to +\infty$ et $d(x_{n_k}, x'_{n_k}) \leq \frac{1}{n_k}$, on doit avoir $x = x'$. Mais on a aussi $D(f(x_{n_k}), f(x'_{n_k})) \geq \varepsilon$ pour tout k : il est impossible que les deux suites $f(x_{n_k})$ et $f(x'_{n_k})$ convergent vers $f(x)$, par conséquent f n'est pas continue en x.

Théorème 3.12. *Toutes les normes sur $\mathbb{R}^n$ sont équivalentes.*

Preuve:

Soit N une norme sur $\mathbb{R}^n$. On va montrer que N est équivalente à $\|\cdot\|_\infty$; alors toutes les normes sont équivalentes à $\|\cdot\|_\infty$, donc elles sont toutes équivalentes.
Notons $e_1, \ldots, e_n$ la base canonique de $\mathbb{R}^n$, et $M = \max(\{N(e_i) \colon 1 \leq i \leq n\})$. Alors on a, pour $x = (x_1, \ldots, x_n)$:

$$\begin{array}{rcl} N(x) & = & N(x_1 e_1 + \ldots + x_n e_n) \\ & \leq & |x_1| N(e_1) + \ldots + |x_n| N(e_n) \\ & \leq & M(|x_1| + \ldots + |x_n|) \\ & \leq & M.n.\|x\|_\infty\ . \end{array}$$

Cela nous donne une des deux inégalités nécessaires pour montrer que N est équivalente à $\|\cdot\|_\infty$; cette inégalité implique aussi que $N \colon (\mathbb{R}^n, d_\infty) \to \mathbb{R}$ est lipschitzienne et donc continue :

$$\forall x, x' \in \mathbb{R}^n \quad |N(x) - N(x')| \leq N(x - x') \leq Mn\|x - x'\|_\infty\ .$$

On a vu que les fermés bornés de $(\mathbb{R}^n, d_\infty)$ sont compacts ; par conséquent, l'ensemble

$$S = \{x \in [-1, 1]^n \colon \exists k,\ x_k = \pm 1\}$$

(la sphère unité pour $\|\cdot\|_\infty$) est compact, et comme N y est continue, elle y atteint son minimum m ; notons que, comme $0 \notin S$, $m > 0$. Soit maintenant $x \in \mathbb{R}^n$ un vecteur non nul ; le vecteur $\frac{x}{\|x\|_\infty}$ appartient à S, et on a donc $N(\frac{x}{\|x\|_\infty}) \geq m$.
Autrement dit, pour tout $x \in \mathbb{R}^n$ on a $m\|x\|_\infty \leq N(x)$ (cette inégalité est bien sûr aussi vraie pour $x = 0$), et on a fini de prouver que N et $\|\cdot\|_\infty$ sont équivalentes.

Corollaire 3.13. Pour n'importe quelle distance induite par une norme sur $\mathbb{R}^n$, les compacts de $\mathbb{R}^n$ sont les fermés bornés.

Exercice 3.14. Soit (X, d) un espace métrique, et (x_n) une suite d'éléments de X qui converge vers $x \in X$. Montrer que l'ensemble $\{x_n \colon n \in \mathbb{N}\} \cup \{x\}$ est compact.

Définition 3.15. Soit (X, d), (Y, D) deux espaces métriques. Une fonction $f \colon X \to Y$ est un *homéomorphisme* si f est une bijection telle que les fonctions f et f^{-1} soient continues.

Proposition 3.16. *Soit (X, d), (Y, D) deux espaces métriques compacts, et $f \colon X \to Y$ une bijection continue. Alors f^{-1} est nécessairement continue, autrement dit, f est un homéomorphisme.*

Preuve:

On doit montrer que la fonction $g = f^{-1}$ est continue. Soit F un fermé de X ; il nous suffit de montrer que $g^{-1}(F)$ est fermé dans Y. Pour cela, notons que

$$g^{-1}(F) = \{y \in Y : g(y) \in F\} = \{y \in Y : f^{-1}(y) \in F\} = f(F) \ .$$

D'après le théorème 3.10, $f(F)$ est compact ; comme les compacts sont fermés, on en déduit bien que $g^{-1}(F) = f(F)$ est fermé, ce qu'il fallait démontrer.

Chapitre 4

Une introduction à la mesure de Lebesgue et aux grands théorèmes de théorie de la mesure.

4.1 Mesures ; le presque partout

Définition 4.1 (première tentative de définition d'une mesure)**.** Une mesure sur un ensemble X est une application $\mu\colon \mathcal{P}(X) \to [0, +\infty]$ avec les propriétés suivantes :

1. $\mu(\emptyset) = 0$.
2. Si $(A_i)_{i\in\mathbb{N}}$ est une suite de parties de X deux à deux disjointes, c'est-à-dire si $A_i \cap A_j = \emptyset$ pour tout $i \neq j$, alors $\mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i=0}^{+\infty} \mu(A_i)$.

Remarque 4.2. Ci-dessus, $\mathcal{P}(X)$ désigne l'ensemble des parties de X. La définition donnée plus haut est très exigeante : en général, on ne demande pas que *toutes* les parties de X aient une mesure. Ceci amène à introduire une notion un peu plus précise (en particulier, dire quelles propriétés on attend de l'ensemble des parties qui admettent une mesure, qu'on appelle les parties *mesurables*). Dans cette partie du cours, on va prétendre que toutes les parties sont mesurables ; pour que les théorèmes présentés dans cette partie soient vraiment corrects, on va ajouter en gris les hypothèses nécessaires de mesurabilité - que vous pouvez ignorer en première lecture, et sur lesquelles vous pourrez revenir après avoir lu la suite du cours.

Définition 4.3. Comme on autorise que des parties aient une mesure infinie (intuitivement, la mesure généralise les notions de longueur/d'aire/de volume, et on « voit »bien que $\mathbb{R}$, par exemple, est de longueur infinie), il faut donner quelques précisions :

— Pour tout $t \in [0, +\infty]$, $+\infty + t = t + \infty = +\infty$.
— Pour tout t *fini*, $+\infty - t = +\infty$.
— $(+\infty) - (+\infty)$ n'est pas défini.

Dans l'énoncé des propriétés de l'intégrale, on va aussi devoir écrire des multiplications, avec les conventions suivantes :

— $0 \cdot (+\infty) = (+\infty) \cdot 0 = 0$.
— Si $t > 0$, $t \cdot (+\infty) = (+\infty) \cdot t = +\infty$.

Proposition 4.4. *Soit μ une mesure sur un ensemble X.*

— *Si $A_0, \dots, A_n$ sont des parties deux à deux disjointes et mesurables, alors*

$$\mu(\bigcup_{i=0}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i) \ .$$

— *Si $A \subseteq B$ et A et B sont mesurables alors $\mu(A) \leq \mu(B)$.*

Preuve:

Pour voir que la première propriété est vraie, il suffit de définir $A_{n+1} = A_{n+2} = \ldots = \emptyset$ et d'appliquer la définition d'une mesure :

$$\mu(\bigcup_{i=0}^{n} A_i) = \mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i=0}^{+\infty} \mu(A_i) = \sum_{i=0}^{n} \mu(A_i) + \sum_{i=n+1}^{+\infty} \mu(A_i) = \sum_{i=0}^{n} \mu(A_i) + 0 = \sum_{i=0}^{n} \mu(A_i) \ .$$

Pour la deuxième propriété on utilise le fait que $B = A \cup (B \setminus A)$ et que $B \setminus A$ est mesurable pour écrire :

$$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A) \ .$$

Attention : si $A \subseteq B$ et A et B sont mesurables, on a envie d'écrire $\mu(A) = \mu(B) - \mu(B \setminus A)$. Ceci peut ne pas avoir de sens, si $\mu(B \setminus A)$ est infini ! Avant d'écrire une telle égalité, il faut vérifier que $\mu(B \setminus A) < +\infty$.

Pour l'instant, on va se concentrer sur une mesure, sur un ensemble très particulier : la mesure de Lebesgue sur $\mathbb{R}$.

Définition 4.5. Il existe une unique mesure λ sur $\mathbb{R}$ (définie sur la tribu borélienne de $\mathbb{R}$), appelée *mesure de Lebesgue* et telle que pour tout segment $[a, b]$ on ait $\lambda([a, b]) = b - a$.

Remarque 4.6. Par convention, à chaque fois qu'on écrira $[a, b]$, on supposera que a, b sont des réels et $a \leq b$.

Intuitivement, $b - a$ correspond à la longueur du segment $[a, b]$; la mesure de Lebesgue permet d'associer une notion de longueur à des parties plus compliquées de $\mathbb{R}$ (au moins, à toutes les parties boréliennes).

Notons quelques conséquences immédiates de cette définition.

Proposition 4.7.
— *Pour tout* $x \in \mathbb{R}$ *on a* $\lambda(\{x\}) = 0$.
— *Si* (x_n) *est une suite de nombres réels, alors* $\lambda(\{x_n\}_{n\in\mathbb{N}}) = 0$.
— *Pour tous réels* a, b *avec* $a \leq b$ *on a* $\lambda([a, b]) = \lambda(]a, b[) = \lambda([a, b[) = \lambda(]a, b]) = b - a$.

Une idée fondamentale en théorie de la mesure est qu'on peut, la plupart du temps, *négliger* les parties de mesure nulle : elle n'ont pas d'influence sur la valeur d'une intégrale, par exemple.

Définition 4.8. Une partie borélienne $A \subseteq \mathbb{R}$ est dite *négligeable* si $\lambda(A) = 0$. On dit aussi que *presque tout* x appartient à $\mathbb{R} \setminus A$, ou encore que $x \in \mathbb{R} \setminus A$ est *vrai presque partout.*

En particulier, si $f, g \colon \mathbb{R} \to \mathbb{R}$ sont des fonctions boréliennes, on dit que f et g sont égales *presque partout* si $\lambda(\{x \colon f(x) \neq g(x)\}) = 0$ (c'est-à-dire, si l'ensemble de tous les x tels que $f(x) \neq g(x)$ est négligeable, ou encore si $f(x) = g(x)$ pour presque tout x).

Exemple. La fonction f définie sur $\mathbb{R}$ par $f(x) = \begin{cases} 0 & \text{si } x \notin \mathbb{N} \\ x & \text{sinon} \end{cases}$ est borélienne et nulle presque partout.

4.2 L'intégrale de Lebesgue

L'idée de base de la construction de l'intégrale de Lebesgue est la suivante : si $f \colon \mathbb{R} \to \mathbb{R}$ est borélienne et ne prend qu'un nombre fini de valeurs $\alpha_1, \ldots, \alpha_n$, alors on voudrait avoir $\displaystyle\int_{\mathbb{R}} f d\lambda = \sum_{i=1}^{n} \alpha_i \lambda(\{x \colon f(x) = \alpha_i\})$. Par exemple, si f est une fonction en escalier (qui est continue par morceaux et donc borélienne), cette définition nous redonne la valeur habituelle de l'intégrale de f.

Dans le cadre de la théorie de la mesure, on commence par utiliser un procédé de passage à la limite (qu'on ne détaillera pas) pour étendre la définition ci-dessus à l'intégrale de toutes les fonctions positives et mesurables ; une fois qu'on sait intégrer toutes les fonctions à valeurs positives, on peut définir facilement l'intégrale de n'importe quelle fonction mesurable. En effet, étant donnée une fonction f, on peut définir sa partie positive f^+ et sa partie négative f^- en posant

$$f^+(x) = \begin{cases} f(x) & \text{si } f(x) \geq 0 \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad f^-(x) = \begin{cases} -f(x) & \text{si } f(x) \leq 0 \\ 0 & \text{sinon} \end{cases} .$$

Alors, f^+ et f^- sont deux fonctions mesurables et à valeurs positives, et par définition $f = f^+ - f^-$. Si on sait intégrer les fonctions mesurables et à valeurs positives, il nous reste donc simplement à poser $\int_{\mathbb{R}} f d\lambda = \int_{\mathbb{R}} f^+ d\lambda - \int_{\mathbb{R}} f^- d\lambda$ (et ce processus s'applique pour toute mesure, pas seulement pour la mesure de Lebesgue).

On va admettre qu'on peut construire une notion d'intégrale ayant les propriétés suivantes :

1. Si f est la fonction caractéristique d'une partie A de $\mathbb{R}$ borélienne, alors $\int_{\mathbb{R}} f d\lambda = \lambda(A)$.
2. Pour toutes fonctions boréliennes $f, g \colon \mathbb{R} \to [0, +\infty]$ et toute constante $c \geq 0$, on a :

$$\int_{\mathbb{R}} c f d\lambda = c \int_{\mathbb{R}} f d\lambda \text{ et } \int_{\mathbb{R}} (f+g) d\lambda = \int_{\mathbb{R}} f d\lambda + \int_{\mathbb{R}} g d\lambda \ .$$

(En particulier, l'intégrale de la fonction nulle vaut 0.)

3. Pour toutes fonctions boréliennes $f, g \colon \mathbb{R} \to [0, +\infty]$,

$$\text{si } f = g \text{ presque partout alors } \int_{\mathbb{R}} f d\lambda = \int_{\mathbb{R}} g d\lambda.$$

4. Pour toutes fonctions boréliennes $f, g \colon \mathbb{R} \to [0, +\infty]$,

$$\text{si } f \leq g \text{ (presque partout) alors } 0 \leq \int_{\mathbb{R}} f d\lambda \leq \int_{\mathbb{R}} g d\lambda.$$

Remarque 4.9. Il est pratique d'autoriser le fait que les fonctions prennent la valeur $+\infty$. Si $\lambda(\{x \colon f(x) = +\infty\}) > 0$, et $f \colon \mathbb{R} \to [0, +\infty]$ est borélienne, alors on a toujours $\int_{\mathbb{R}} f d\lambda = +\infty$. En particulier, si $f \colon \mathbb{R} \to [0, +\infty]$ est d'intégrale finie alors $f(x) < +\infty$ presque partout, et on peut se ramener à considérer l'intégrale d'une fonction à valeurs finies en intégrant f sur l'ensemble $\{x \colon f(x) < +\infty\}$, ce que permet la définition suivante.

Définition 4.10. Étant donnée une partie A de $\mathbb{R}$ borélienne, et $f \colon \mathbb{R} \to [0, +\infty]$ borélienne, on définit

$$\int_A f d\lambda = \int_{\mathbb{R}} f \mathbf{1}_A d(\lambda) \ .$$

(En particulier, pour toute constante $c \geq 0$, $\int_A c d\lambda = c\lambda(A)$.)

Remarque 4.11. $\mathbf{1}_A$ désigne la *fonction caractéristique* de A, c'est-à-dire que $\mathbf{1}_A(x)$ vaut 1 si $x \in A$ et 0 sinon. C'est une fonction borélienne si, et seulement si, A est borélien. Ainsi, $f\mathbf{1}_A$ est la fonction qui est égale à f sur A, et à 0 ailleurs.

Proposition 4.12.
5. *Si A est borélien et $\lambda(A) = 0$ alors $\int_A f d\lambda = 0$ pour toute fonction $f \colon \mathbb{R} \to [0, +\infty]$ borélienne.*
6. *Si A_1, A_2 sont des parties de $\mathbb{R}$ boréliennes et disjointes, alors pour toute fonction $f \colon \mathbb{R} \to [0, +\infty]$ borélienne, on a*

$$\int_{A_1 \cup A_2} f d\lambda = \int_{A_1} f d\lambda + \int_{A_2} f d\lambda \ .$$

Exercice 4.13. Déduire des propriétés énoncées plus haut que si A_1, A_2 sont des parties de $\mathbb{R}$ boréliennes et telles que $\lambda(A_1 \cap A_2) = 0$, alors pour toute fonction $f \colon \mathbb{R} \to [0, +\infty]$ borélienne, on a

$$\int_{A_1 \cup A_2} f d\lambda = \int_{A_1} f d\lambda + \int_{A_2} f d\lambda \ .$$

Remarque 4.14. Pour intégrer une fonction à valeurs positives sur une partie A, il suffit qu'elle soit au moins définie sur A : si $f \colon X \subseteq \mathbb{R} \to \mathbb{R}$ est une fonction borélienne et A une partie borélienne de X, on définit $\int_A f d\lambda = \int_A g d\lambda$ où g est un (n'importe quel) prolongement borélien de f à $\mathbb{R}$ (par exemple g définie par $g(x) = f(x)$ pour $x \in X$ et $g(x) = 0$ pour $x \in \mathbb{R} \setminus X$).

Attention : on peut définir l'intégrale $\int_{\mathbb{R}} f d\lambda$ pour *toute* fonction borélienne f à valeurs positives ; *mais cette intégrale peut valoir* $+\infty$. Une idée intuitive est que, dans le cas d'une fonction à valeurs positives, l'intégrale de f coïncide avec « l'aire sous la courbe représentative de f » ; et cette aire peut éventuellement être infinie : par exemple, si f est la fonction constante égale à 1 sur $\mathbb{R}$, alors le domaine limité par le graphe de f et l'axe des abscisses est un « rectangle »dont un côté est de longueur 1 est l'autre de longueur infinie, et on voit bien que l'aire d'un tel domaine est infinie. Mais si f change de signe, alors la notion d'« aire sous la courbe »devient plus compliquée à interpréter.

Définition 4.15. Soit $f\colon X \subseteq \mathbb{R} \to \mathbb{R}$ une fonction borélienne. On dit que f est *intégrable* sur une partie A borélienne de X si $\int_A |f| d\lambda < +\infty$.

Dans ce cas,f^+ et f^- sont également intégrables sur A (par la propriété (4)) et on définit $\int_A f d\lambda = \int_A f^+ d\lambda - \int_A f^- d\lambda$.

Attention, dans le cadre de la théorie de Lebesgue, on définit $\int_A f d\lambda$ uniquement quand f est *intégrable* sur A (c.à.d. $\int_A |f| d\lambda < +\infty$) et on ne considère pas les intégrales généralisées semi-convergentes qui apparaissent dans la théorie de l'intégrale de Riemann (voir la première feuille de TD).

On retrouve les propriétés (1), (2), (3), (4), (5), (6) ci-dessus qui sont cette fois vraies pour des fonctions boréliennes intégrables.

Proposition 4.16. *Pour toutes fonctions boréliennes $f, g\colon X \subseteq \mathbb{R} \to \mathbb{R}$ intégrables sur une partie borélienne $A \subseteq X$ on a*

1. $\int_A 1 d\lambda = \lambda(A)$;
2. *pour toutes constantes réelles α et β, la fonction $\alpha f + \beta g$ est également intégrable sur A et*
$$\int_A (\alpha f + \beta g) d\lambda = \alpha \int_A f d\lambda + \beta \int_A g d\lambda \text{ (linéarité de l'intégrale);}$$
3. *si $f = g$ presque partout, alors* $\int_A f d\lambda = \int_A g d\lambda$;
4. *si $f \geq 0$ presque partout, alors* $\int_A f d\lambda \geq 0$ (positivité de l'intégrale) ;
si $f \leq g$ presque partout, alors $\int_A f d\lambda \leq \int_A g d\lambda$ (monotonie de l'intégrale) ;
5. *si $\lambda(A) = 0$ alors* $\int_A f d\lambda = 0$;
6. *si A_1, A_2 sont des parties boréliennes de A telles que $\lambda(A_1 \cap A_2) = 0$, alors*
$$\int_{A_1 \cup A_2} f d\lambda = \int_{A_1} f d\lambda + \int_{A_2} f d\lambda \ .$$

On a de plus

7. L'inégalité triangulaire
$$\left| \int_A (f+g) d\lambda \right| \leq \int_A |f| d\lambda + \int_A |g| d\lambda \ .$$

4.3 Théorèmes d'échanges limite-intégrale

La théorie de Lebesgue nous permet d'intégrer plus de fonctions que la théorie de Riemann. Cela simplifie l'énoncé des théorèmes d'échange limite/intégrale, qui sont plus puissants dans ce contexte.

Théorème 4.17 (Théorème de convergence monotone)**.** *Soit A une partie de $\mathbb{R}$ borélienne et $f_i\colon A \to [0, +\infty[$ une suite de fonctions boréliennes telle pour presque tout $x \in A$ la suite $(f_i(x))$ soit une suite croissante. Alors il existe une fonction borélienne $f\colon A \to [0, +\infty]$ telle que $f_i(x)$ converge vers $f(x)$ presque partout, et on a*
$$\lim_{i \to +\infty} \int_A f_i d\lambda = \int_A f d\lambda \ .$$

Remarque 4.18. Comme, pour presque tout x, la suite $(f_i(x))$ est croissante, il est immédiat que cette suite admet une limite dans $[0, +\infty]$. Ainsi, l'existence de la fonction f est une conséquence immédiate de l'hypothèse de monotonie. Ce qui n'est pas trivial, c'est de pouvoir échanger limite et intégrale !

Attention : l'hypothèse du théorème *n'est pas* que les fonctions f_i sont des fonctions croissantes ; c'est que, pour presque tout x *fixé*, la suite $(f_i(x))$ est une suite croissante.

Remarque 4.19. Comme on ne travaille dans cette partie qu'avec la mesure de Lebesgue, on n'a écrit ce théorème que pour la mesure de Lebesgue ; mais il serait vrai aussi pour toute autre mesure (en remplaçant « borélien »par « mesurable ») ; il en va de même de la définition de l'intégrale qu'on a vue plus haut, et du théorème de convergence dominée qu'on va maintenant présenter.

L'avantage de ce théorème est qu'il s'applique même si la suite des intégrales tend vers $+\infty$; mais l'hypothèse demandant que les fonctions soient à valeurs positives, et que la suite $(f_i(x))$ soit croissante pour presque tout x, n'est souvent pas vérifiée dans les exemples. Néanmoins, à partir de ce théorème on peut déduire un autre théorème fondamental.

Théorème 4.20 (Théorème de convergence dominée)**.** *Soit A une partie de $\mathbb{R}$ borélienne, et $f_i\colon A \to \mathbb{R}$ une suite de fonctions boréliennes et satisfaisant les hypothèses suivantes :*

1. *Pour presque tout x, la suite $f_i(x)$ est convergente vers une limite qu'on appelle $f(x)$ (alors f est borélienne).*
2. *Il existe une fonction $g\colon A \to \mathbb{R}$ borélienne à valeurs positives, telle que pour presque tout x on ait $|f_i(x)| \le g(x)$ et $\int_A g d\lambda < +\infty$.*

Alors f est intégrable, et on a

$$\int_A f d\lambda = \lim_{i\to+\infty} \int_A f_i d\lambda \ .$$

Remarque 4.21. Dans l'énoncé du théorème ci-dessus, il est fondamental que *la fonction g ne dépend pas de i.*

En appliquant les théorèmes de convergence monotone et de convergence dominée sur les sommes partielles d'une suite de fonctions, on obtient le corollaire suivant.

Corollaire 4.22 (Échanges série/intégrale)**.** Soit A une partie de $\mathbb{R}$ borélienne.

1. Si $(f_n)_{n\ge 0}$ est une suite de fonctions positives boréliennes définies sur A, alors

$$\int_A \sum_{n=0}^{+\infty} f_n \ d\lambda = \sum_{n=0}^{+\infty} \int_A f_n d\lambda \qquad \text{(convergence monotone).}$$

2. Si $(f_n)_{n\ge 0}$ est une suite de fonctions boréliennes définies sur A telle que

$$\sum_{n=0}^{+\infty} \int_A |f_n| d\lambda < +\infty$$

alors la série $\sum_{n=0}^{+\infty} f_n(x)$ converge absolument pour presque tout $x \in A$ vers une fonction f intégrable, et on a

$$\int_A f d\lambda = \sum_{n=0}^{+\infty} \int_A f_n d\lambda \qquad \text{(convergence dominée).}$$

Remarque 4.23. La théorie de l'intégrale de Cauchy-Riemann est basée sur la stratégie suivante : d'abord, on définit l'intégrale des fonctions en escalier (qui coïncide avec leur intégrale pour la mesure de Lebesgue). Ensuite, on utilise le fait que toute fonction continue par morceaux f définie sur un segment I est une limite uniforme de fonctions en escalier f_i, et on définit $\displaystyle\int_I f(x)dx = \lim_{i\to+\infty} \int_I f_i(x)dx$.

Comme toute suite de fonctions en escalier convergeant uniformément vers une fonction continue par morceaux sur un segment est bornée par une constante M, et que les fonctions constantes sont intégrables sur les segments, on peut appliquer le théorème de convergence dominée pour conclure que $\displaystyle\int_I f d\lambda = \lim \int_I f_i d\lambda$. Ainsi, on déduit que, pour une fonction continue par morceaux, son intégrale au sens de Riemann et son intégrale au sens de Lebesgue coïncident.

En utilisant le théorème de convergence monotone, on en déduit que l'intégrale généralisée au sens de Riemann d'une application continue positive sur un intervalle ouvert est égale à son intégrale au sens de Lebesgue. Enfin, une fonction continue qui a une intégrale généralisée absolument convergente au sens de Riemann, est également intégrable au sens de Lebesgue et à nouveau par le théorème de convergence dominée, ces intégrales coïncident. Plus précisément :

Théorème 4.24 (Comparaison avec l'intégrale de Riemann)**.** *Les intégrales au sens de Riemann et au sens de Lebesgue coïncident (en particulier) dans les situations suivantes :*

1. *Si $-\infty < a < b < +\infty$ et $f : [a, b] \to \mathbb{R}$ est une fonction continue par morceaux, alors f est intégrable au sens de Riemann et au sens de Lebesgue, et*

$$\int_a^b f(x)dx = \int_{[a,b]} f d\lambda.$$

2. *Si $-\infty \leq a < b \leq +\infty$ et $f :]a, b[\to [0, +\infty[$ est une fonction continue par morceaux, alors*

$$\int_a^b f(x)dx = \int_{]a,b[} f d\lambda.$$

(Rappelons que ces intégrales peuvent valoir $+\infty$.)

3. *Si $-\infty \leq a < b \leq +\infty$ et $f :]a, b[\to \mathbb{R}$ est une fonction continue par morceaux dont l'intégrale généralisée au sens de Riemann est absolument convergente, c'est-à-dire telle que $\displaystyle\int_a^b |f(x)|dx < +\infty$, alors f est intégrable au sens de Lebesgue et*

$$\int_a^b f(x)dx = \int_{]a,b[} f d\lambda.$$

4.4 Continuité et dérivabilité des intégrales à paramètres

Dans cette section, on fixe une partie X borélienne de $\mathbb{R}$, et I un intervalle. Étant donnée une fonction $f: X \times I \to \mathbb{R}$ telle que pour tout $t \in I$ $x \mapsto f(x, t)$ soit intégrable sur X, on peut définir la fonction $F: t \mapsto \int_X f(x, t)d\lambda(x)$. On s'intéresse à des hypothèses sur f permettant de conclure que F est continue, ou dérivable.

Théorème 4.25 (Théorème de continuité des intégrales à paramètres)**.** *Soit $f: X \times I \to \mathbb{R}$ une fonction telle que*

1. *Pour presque tout $x \in X$, la fonction $t \mapsto f(x, t)$ est continue sur I.*
2. *Il existe une fonction $g: X \to [0, +\infty[$ intégrable et telle que pour tout $t \in I$ on ait $|f(x, t)| \leq g(x)$ pour presque tout $x \in X$.*
3. *Pour tout t la fonction $x \mapsto f(x, t)$ est borélienne.*

Alors la fonction $F: t \mapsto \displaystyle\int_X f(x, t)d\lambda(x)$ est bien définie et continue sur I.

Remarque 4.26. Dans la deuxième condition ci-dessus, *la fonction g ne dépend pas de t*. Dans les applications, il n'est pas toujours facile de trouver une fonction g satisfaisant les conditions du théorème ! La première condition est, elle, normalement facile à vérifier.

Preuve:

Déjà, notons que, pour tout t fixé, la fonction $x \mapsto f(x, t)$ est borélienne et telle que $|f(x, t)| \leq g(x)$ pour presque tout $x \in X$; comme g est intégrable, on en déduit que $\int_X |f(x, t)|d\lambda(x) \leq \int_X g(x)d\lambda(x) < +\infty$, donc la fonction $x \mapsto f(x, t)$ est intégrable, et F est donc bien définie.

Fixons maintenant $t \in I$. Pour montrer que F est continue en t, on doit vérifier que, pour toute suite (t_n) d'éléments de I qui converge vers t, la suite $F(t_n)$ converge vers $F(t)$. Fixons donc une telle suite (t_n). Par définition, on a $F(t_n) = \displaystyle\int_X f(x, t_n)d\lambda(x)$.

La première hypothèse ci-dessus nous permet d'affirmer que la suite de fonctions $f_n : x \mapsto f(x, t_n)$ est une suite de fonctions boréliennes et telle que $f_n(x)$ converge vers $f(x, t)$ pour presque tout $x \in X$; et on a aussi $|f_n(x)| \leq g(x)$ pour presque tout x. Comme g est intégrable, on peut appliquer le théorème de convergence dominée pour conclure que

$$\lim_{n\to+\infty} \int_X f_n(x)d\lambda(x) = \int_X f(x, t)d\lambda(x) \ .$$

On vient de montrer que $F(t_n) = \int_X f_n(x)d\lambda(x)$ converge vers $F(t)$, ce qui prouve que F est bien continue en t.

Remarque 4.27. En regardant la preuve, on voit que si, dans les hypothèses du théorème, on avait remplacé « pour presque tout $x \in X$, la fonction $t \mapsto f(x,t)$ est continue sur I »par « pour presque tout $x \in X$, la fonction $t \mapsto f(x,t)$ est continue en un certain $t_0 \in I$ »(en laissant les autres hypothèses inchangées), alors on aurait pu conclure que F est continue en t_0.

Par ailleurs, notons que si X et I sont des segments, et f est une fonction continue, alors f est bornée, c'est-à-dire qu'il existe M tel que $|f(x,t)| \leq M$ pour tout $(x,t) \in X \times I$. Comme les fonctions constantes sont intégrables sur les segments, on voit qu'on peut poser $g(x) = M$ et appliquer le théorème de continuité : si X et I sont des segments, et f est une fonction continue des deux variables (x,t), alors F est continue. Ce corollaire du théorème de continuité peut se démontrer simplement sans utiliser le théorème de convergence dominée.

Théorème 4.28 (Théorème de dérivabilité des intégrales à paramètres)**.** *Soit $f \colon X \times I \to \mathbb{R}$ une fonction telle que*

1. *Pour tout $t \in I$, la fonction $x \mapsto f(x,t)$ est intégrable sur X.*
2. *Pour presque tout $x \in X$, la fonction $t \mapsto f(x,t)$ est dérivable sur I (on note sa dérivée $\frac{\partial f}{\partial t}(x,t)$).*
3. *Il existe une fonction $g \colon X \to [0,+\infty[$ intégrable et telle que pour tout $t \in I$ et tout $x \in X$ on ait $\left|\frac{\partial f}{\partial t}(x,t)\right| \leq g(x)$.*
4. *Pour tout $t \in I$ la fonction $x \mapsto \frac{\partial f}{\partial t}(x,t)$ est borélienne.*

Alors la fonction $F \colon t \mapsto \int_X f(x,t)d\lambda(x)$ est bien définie sur I, dérivable, et on a pour tout $t \in I$ que

$$F'(t) = \int_X \frac{\partial f}{\partial t}(x,t)d\lambda(x) \ .$$

Remarque 4.29. En fait, le théorème reste vrai si l'on remplace la première hypothèse par le fait qu'il existe un $t \in I$ tel que $x \mapsto f(x,t)$ est intégrable sur I (bien sûr, on doit toujours supposer que cette fonction est borélienne pour tout $t \in I$).

Remarque 4.30. Si on ajoute l'hypothèse selon laquelle la fonction $t \mapsto \frac{\partial f}{\partial t}(x,t)$ est continue sur I pour presque tout $x \in X$, alors le théorème de continuité des intégrales à paramètres nous permet de conclure que F' est continue sur I, c'est-à-dire que F est de classe $\mathcal{C}^1$ sur I.

Preuve:

Fixons $t \in I$. Pour vérifier que F est dérivable en t, on doit étudier le taux d'accroissement de F en t ; fixons donc une suite (t_n) d'éléments de I différents de t, et considérons le taux d'accroissement

$$\frac{F(t_n) - F(t)}{t_n - t} = \int_X \frac{f(x,t_n) - f(x,t)}{t_n - t} d\lambda(x) \ .$$

Comme dans la preuve du théorème de continuité, considérons la suite de fonctions $f_n \colon x \mapsto \frac{f(x,t_n) - f(x,t)}{t_n - t}$. Le fait que $t \mapsto f(x,t)$ soit dérivable en t pour presque tout $x \in X$ nous dit que $f_n(x)$ converge vers $\frac{\partial f}{\partial t}(x,t)$ pour presque tout $x \in X$. De plus, l'inégalité des accroissements finis nous dit que, pour presque tout $x \in X$, on a

$$\left|\frac{f(x,t_n) - f(x,t)}{t_n - t}\right| \leq \sup_{s \in [t_n,t]} \frac{\partial f}{\partial t}(x,s) \leq g(x) \ .$$

On peut donc appliquer le théorème de convergence dominée à la suite f_n pour conclure que

$$\lim_{n \to +\infty} \int_X f_n(x)d\lambda(x) = \int_X \lim_{n \to +\infty} f_n(x)d\lambda(x) \ .$$

Autrement dit, on a montré que $\frac{F(t_n) - F(t)}{t_n - t}$ converge vers $\int_X \frac{\partial f}{\partial t}(x,t)d\lambda(x)$ quand n tend vers $+\infty$, c'est-à-dire que F est dérivable en t et $F'(t) = \int_X \lim_{n \to +\infty} f_n(x)d\lambda(x)$.

Chapitre 5

Intégrale de fonctions de plusieurs variables réelles

5.1 La mesure de Lebesgue sur $\mathbb{R}^n$

De même que la mesure de Lebesgue sur $\mathbb{R}$ généralise la notion de longueur, on peut généraliser la notion d'aire (dans $\mathbb{R}^2$), de volume (dans $\mathbb{R}^3$)...

Théorème 5.1. *Soit $n \geq 1$ un entier. Il existe une unique mesure sur $\mathbb{R}^n$ -ou plutôt, sur la tribu borélienne de $\mathbb{R}^n$), notée λ_n, telle que*

$$\lambda_n([a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]) = \prod_{i=1}^{n} (b_i - a_i) \ .$$

On appelle cette mesure la mesure de Lebesgue sur $\mathbb{R}^n$.

Quand $n = 2$, la formule ci-dessus nous dit que la mesure d'un rectangle (dont les côtés sont parallèles aux axes de coordonnées) est égale au produit des longueurs de ses côtés, c'est-à-dire à son aire ; quand $n = 3$ on retrouve la formule pour le volume d'un pavé, etc.

Exactement de la même façon que pour la définition de l'intégrale de Lebesgue sur $\mathbb{R}$ (ou, plus généralement, de l'intégrale par rapport à une mesure quelconque), on peut définir tout d'abord l'intégrale pour les fonctions boréliennes et à valeurs dans $[0, +\infty]$, et ensuite pour les fonctions $f \colon \mathbb{R}^n \to \mathbb{R}$ boréliennes et telles que $\int_{\mathbb{R}^n} |f| d\lambda_n < +\infty$. Cette notion d'intégrale a les mêmes propriétés que celles qu'on a énoncées au chapitre précédent : positivité, linéarité, inégalité triangulaire... Elle est définie en posant tout d'abord $\displaystyle\int_{\mathbb{R}^n} \mathbf{1}_A d\lambda_n = \lambda_n(A)$, si A est une partie borélienne de $\mathbb{R}^n$ et $\mathbf{1}_A$ est sa fonction caractéristique ; puis en étendant par linéarité aux combinaisons linéaires de fonctions caractéristiques, puis par un procédé de passage à la limite pour définir l'intégrale de toutes les fonctions boréliennes et à valeurs positives. Comme dans le cas de la mesure de Lebesgue sur $\mathbb{R}$, on dit que f est *intégrable* si f est borélienne et $\displaystyle\int_{\mathbb{R}^n} |f| d\lambda_n$ est finie.

On voit donc que, exactement comme dans le cas de la mesure de Lebesgue sur $\mathbb{R}$, les fonctions boréliennes et à valeurs positives jouent un rôle particulier ; on peut toujours calculer leur intégrale, et celle-ci peut valoir $+\infty$.

5.2 Théorèmes de Fubini-Tonelli et Fubini

La méthode de base pour calculer une intégrale d'une fonction de n variables est de se ramener à n calculs successifs d'intégrales de fonctions de 1 variable.

Théorème 5.2 (Théorème de Fubini-Tonelli)**.** *Soit m, p deux entiers et $f \colon \mathbb{R}^m \times \mathbb{R}^p \to [0, +\infty]$ une fonction* *borélienne. Alors on a*

$$\begin{aligned}\int_{\mathbb{R}^{m+p}} f d\lambda_{m+p} &= \int_{\mathbb{R}^m}\left(\int_{\mathbb{R}^p} f(x,y)d\lambda_p(y)\right)d\lambda_m(x)\\ &= \int_{\mathbb{R}^p}\left(\int_{\mathbb{R}^m} f(x,y)d\lambda_m(x)\right)d\lambda_p(y)\end{aligned}$$

En pratique, dans les exercices, m et p vaudront le plus souvent 1 ou 2.

Remarque 5.3. L'énoncé ci-dessus sous-entend que les fonctions apparaissant dans les intégrales (par exemple, $x \mapsto \int_{\mathbb{R}^p} f(x,y)d\lambda_p(y)$) sont boréliennes...

Remarque 5.4. Comme dans le cas de $\mathbb{R}$, intégrer une fonction de n variables sur une partie borélienne A de $\mathbb{R}^n$ revient à intégrer la fonction $f \cdot \mathbf{1}_A$ sur $\mathbb{R}^n$, donc l'énoncé ci-dessus peut être utilisé pour calculer les intégrales de fonctions boréliennes définies sur des sous-parties boréliennes de $\mathbb{R}^n$.

Exemple. Calculons l'aire du disque unité $D = \{(x,y) \in \mathbb{R}^2 \colon x^2+y^2 \leq 1\}$.

Par définition, l'aire d'une partie D est l'intégrale de la fonction caractéristique de D sur $\mathbb{R}^2$; on a donc

$$\begin{aligned}\text{aire}(D) &= \int_{\mathbb{R}^2} \mathbf{1}_D(x,y)d\lambda_2(x,y)\\ &= \int_{x=-1}^{1}\left(\int_{y=-\sqrt{1-x^2}}^{\sqrt{1-x^2}} d\lambda(y)\right)d\lambda(x)\\ &= \int_{-1}^{1} 2\sqrt{1-x^2}dx\\ &= 2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2(t)dt\\ &= \pi\end{aligned}$$

On voit que, dans le calcul ci-dessus, on n'a pas pris en compte les points en dehors de D, puisque $\mathbf{1}_D$ vaut 0 en ces points : en pratique, on détermine dans quel domaine varie x (ici, $[-1,1]$) puis, *à x fixé*, dans quel domaine varie y, et on calcule les intégrales correspondantes.

Bien sûr, on aurait pu aussi faire le calcul « dans l'autre sens », en intégrant d'abord par rapport à x puis par rapport à y ; ici ça ne changerait rien (la fonction intégrée et le domaine sont symétriques en x,y) mais dans d'autres cas un des calculs peut être beaucoup plus facile que l'autre.

Dans le cas où f n'est pas à valeurs positives, on a d'abord besoin de s'assurer que $|f|$ est intégrable, ce qui peut être fait en appliquant le théorème de Fubini-Tonelli.

Théorème 5.5 (Théorème de Fubini)**.** *Soit $f \colon \mathbb{R}^{m+p} \to \mathbb{R}$ une fonction borélienne telle que $|f|$ est intégrable. Alors on a :*

$$\begin{aligned}\int_{\mathbb{R}^{m+p}} f d\lambda_{m+p} &= \int_{\mathbb{R}^m}\left(\int_{\mathbb{R}^p} f(x,y)d\lambda_p(y)\right)d\lambda_m(x)\\ &= \int_{\mathbb{R}^p}\left(\int_{\mathbb{R}^m} f(x,y)d\lambda_m(x)\right)d\lambda_p(y)\end{aligned}$$

Remarque 5.6. Comme dans le cas du théorème de Fubini-Tonelli, il est sous-entendu que les fonctions apparaissant dans l'intégrable sont boréliennes (en fait, elles ne sont définies que presque partout, mais ça n'affecte pas la valeur de l'intégrale ; on reverra cela plus en détail quand on énoncera le théorème de Fubini pour des mesures plus générales)

Remarque 5.7. Dans le cas particulier d'une fonction continue sur un compact de $\mathbb{R}^{m+p}$, on sait qu'elle est toujours intégrable ; il n'y a donc pas besoin dans ce cas d'utiliser au préalable le théorème de Fubini-Tonelli, par contre il faut bien vérifier que la fonction f est continue, et surtout que le domaine d'intégration est compact...

Pour essayer de se convaincre que le théorème de Fubini est « raisonnable », donnons-en une preuve dans le cas où on intègre une fonction *continue* f sur un rectangle $[a,b]\times[c,d]$. Notons d'abord que sous ces hypothèses $|f|$ est bien intégrable puisqu'elle est bornée et que le domaine d'intégration est de mesure finie. On veut montrer que :

$$\int_a^b \left(\int_c^d f(x,y)dy\right) dx = \int_c^d \left(\int_a^b f(x,y)dx\right) dy\ .$$

Pour cela, on introduit deux fonctions $F,G\colon [c,d]\to\mathbb{R}$ en posant

$$F(t) = \int_a^b \left(\int_c^t f(x,y)dy\right) dx \quad \text{et} \quad G(t) = \int_c^t \left(\int_a^b f(x,y)dx\right) dy\ .$$

Commençons par étudier F ; elle est de la forme $F(t)=\int_a^b h(x,t)dx$, avec $h(x,t)=\int_c^t f(x,y)dy$. A x fixé, la fonction $t\mapsto h(x,t)=\int_c^t f(x,y)dy$ est dérivable, de dérivée $f(x,t)$; de plus cette dérivée est bornée. On peut donc appliquer le théorème de dérivation des intégrales à paramètres pour conclure que F est dérivable sur $[c,d]$, de dérivée $F'(t)=\int_a^b f(x,t)dx$.

Passons à G : le théorème de continuité des intégrales à paramètres nous permet de voir que $g\colon y\mapsto \int_a^b f(x,y)dx$ est une fonction continue sur $[c,d]$. Puisque $G(t)=\int_c^t g(y)dy$, on voit que G est dérivable et que $G'(t)=g(t)=\int_a^b f(x,t)dx$.

On voit donc que F et G sont dérivables, et que $F'=G'$ sur $[c,d]$; de plus on a $F(c)=G(c)=0$. On en conclut que F et G sont égales sur $[c,d]$, en particulier $F(d)=G(d)$, ce qu'on voulait démontrer.

5.3 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. Toutes les fonctions ne sont pas acceptables pour ce théorème : les fonctions que l'on peut utiliser pour faire un changement de variable sont les *difféomorphismes de classe* $\mathcal{C}^1$.

Définition 5.8. Étant donné U un ouvert de $\mathbb{R}^n$, et $\varphi\colon U\to\mathbb{R}^n$ une fonction de classe $\mathcal{C}^1$, on note $J\varphi(x)$ la matrice jacobienne de φ en un point $x\in U$; rappelons qu'il s'agit de la matrice dont le coefficient sur la i-ième ligne et la j-ième colonne vaut $\dfrac{\partial\varphi_i}{\partial x_j}(x)$.

Définition 5.9. Soit U,V deux ouverts de $\mathbb{R}^n$. Une fonction $\varphi\colon U\to V$ est un difféomorphisme de classe $\mathcal{C}^1$ si :

1. φ est une bijection de U sur V.
2. φ est de classe $\mathcal{C}^1$ sur U, c'est-à-dire que chaque dérivée $\dfrac{\partial\varphi_i}{\partial x_j}$ existe et est continue sur U.
3. La bijection réciproque φ^{-1} est également de classe $\mathcal{C}^1$.

Remarque 5.10. Comme $\varphi^{-1}\circ\varphi(x)=x$ pour tout $x\in U$ par définition, on obtient en appliquant la règle de la chaîne que $J\varphi^{-1}(\varphi(x))J\varphi(x)=I_n$ (la matrice identité) pour tout $x\in U$; en particulier $J\varphi(x)$ doit être inversible pour tout $x\in U$ si φ est un difféomorphisme (ce qui revient à dire que son déterminant est non nul).

Un exemple fondamental de difféomorphisme de classe $\mathcal{C}^1$ de $\mathbb{R}^n$ dans $\mathbb{R}^n$ est donné par une application linéaire et inversible ; rappelons que si φ est linéaire, sa matrice jacobienne est simplement la matrice de φ dans la base canonique de $\mathbb{R}^n$.

Souvent, pour vérifier qu'une application est un difféomorphisme de classe $\mathcal{C}^1$, on utilise la caractérisation suivante.

Théorème 5.11 (Théorème d'inversion globale)**.** *Soit U un ouvert de $\mathbb{R}^n$, et $\varphi\colon U \to \mathbb{R}^n$ une application de classe $\mathcal{C}^1$. Si φ est injective sur U et $\det(J\varphi(x)) \neq 0$ pour tout $x \in U$, alors $\varphi(U)$ est un ouvert de $\mathbb{R}^n$ et φ est un difféomorphisme de classe $\mathcal{C}^1$ de U sur $\varphi(U)$.*

Remarque 5.12. Toute la difficulté du théorème précédent est dans la démonstration que $\varphi(U)$ est ouvert. Si on suppose que $\varphi(U)$ est ouvert dans l'énoncé ci-dessus, alors le fait que φ soit un difféomorphisme vient simplement du fait qu'on sait calculer « explicitement »l'inverse d'une matrice inversible (via le lien entre inverse et comatrice).

Nous pouvons maintenant énoncer le théorème de changement de variables.

Théorème 5.13 (Théorème de changement de variables)**.** *Soit 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*
$$\int_V f(x) d\lambda_n(x) = \int_U f \circ \varphi(y) |\det(J\varphi(y))| d\lambda_n(y) \ .$$

La définition de l'intégrale fait que (3) est une conséquence immédiate de (2) ; et (1) est un cas particulier de (2) appliqué à la fonction caractéristique de $\varphi(B)$. La définition fait aussi qu'il n'est pas difficile de déduire (2) une fois qu'on a démontré (1).

Ce théorème est difficile à démontrer, et on ne va pas essayer de donner une idée de la preuve dans le cas où φ n'est pas une application linéaire. Discutons un peu la preuve de (1) dans le cas où φ est une application linéaire. Étudions quelques cas particuliers :

1. La matrice M de φ dans la base canonique est la matrice diagonale $\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & t & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$. Alors on a $\varphi([0,1]^n) = [0,1] \times [0,1] \times \ldots [0,t] \times [0,1] \times \ \ldots [0,1]$ si $t > 0$, et $[0,1] \times [0,1] \times \ldots [t,0] \times [0,1] \times \ \ldots [0,1]$ si $t < 0$. Dans les deux cas, on a par définition de la mesure de Lebesgue que
$$\lambda_n(\varphi([0,1]^n) = |t| = |\det(M)| \lambda_n([0,1]^n) \ .$$
2. La matrice M est la matrice par blocs $\begin{pmatrix} N & \\ & I_{n-2} \end{pmatrix}$, où $N = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Alors on a $\varphi([0,1]^n) = D \times [0,1]^{n-2}$, où D est le parallélogramme de sommets $(0,0)$, $(0,1)$, (1.1) et $(1,2)$ (faites un dessin !). Par conséquent $\lambda_n(\varphi([0,1]^n) = \text{aire}(D) = 1$, et $\det(M) = \det(N) = 1$. Donc dans ce cas on a bien comme espéré que $\lambda_n(\varphi([0,1]^n) = |\det(M)| \lambda_n(\varphi([0,1]^n)$.
3. La matrice M est une *matrice de permutation*, c'est-à-dire que φ est une application linéaire permutant les vecteurs de base. Alors la déterminant de M vaut ± 1 (ce dont on peut se convaincre par récurrence, en développant le déterminant de M par rapport à la première ligne par exemple, et en utilisant que M a exactement un 1 sur chaque ligne et chaque colonne, et des 0 ailleurs ; ou bien en revenant à la formule définissant le déterminant), et $\varphi([0,1]^n) = [0,1]^n$. On a donc encore $\lambda_n(\varphi([0,1]^n) = |\det(M)| \lambda_n(\varphi([0,1]^n)$

Pour l'instant, on a le résultat désiré pour des matrices de trois types, et $B = [0,1]^n$; par pavage, on peut se convaincre qu'on a alors le résultat désiré pour une matrice d'un de ces trois types et B n'importe quelle partie borélienne de $\mathbb{R}$. Ensuite, un théorème d'algèbre linéaire nous dit que l'on peut écrire toute matrice inversible

M sous la forme $M = M_1 M_2 \dots M_p$, où chaque matrice M_i est d'un des trois types ci-dessus. On peut donc écrire, si φ est une bijection linéaire de matrice M et B une partie borélienne de $\mathbb{R}^n$:

$$\begin{aligned}
\lambda_n(\varphi(B)) &= \lambda_n(M_1 \dots M_p(B)) \\
&= |\det(M_p)|\lambda_n(M_1 \dots M_{p-1}(B)) \\
&= |\det(M_p)| \dots |\det(M_1)|\lambda_n(B) \\
&= |\det(M)|\lambda_n(B).
\end{aligned}$$

Exemple (changement de variables en 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 $\phi(U) = \mathbb{R}^2 \setminus [0, +\infty[= V$.

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.

Par exemple, calculons

$$I = \int_D (x+y)^2 dxdy, \text{ où } D = \{(x, y) \colon x^2 + y^2 < 1\}.$$

En utilisant le théorème de changement de variables avec les coordonnées polaires (et le théorème de Fubini), on obtient $\phi^{-1}(D) =]0, 1[\times]0, 2\pi[$ et

$$\begin{aligned}
I &= \int_{\phi^{-1}(D)} (r\cos\theta + r\sin\theta)^2 r dr d\theta \\
&= \int_0^1 \left(\int_0^{2\pi} r^3(\cos^2\theta + \sin^2\theta + 2\cos\theta\sin\theta) d\theta \right) dr \\
&= \int_0^1 r^3 \left(\int_0^{2\pi} (1 + \sin 2\theta) d\theta \right) dr \\
&= \int_0^1 2\pi r^3 dr \\
&= \frac{\pi}{2}.
\end{aligned}$$

Chapitre 6

Tribus et mesures

On va maintenant prendre en compte les questions de *mesurabilité* qui ont été éludées dans les chapitres précédents. En pratique, on se heurte à des problèmes si l'on souhaite vraiment que *toutes* les parties d'un ensemble X soient mesurables. Par exemple, considérons le cas où X est le cercle unité $\{e^{i\alpha} \colon \alpha \in \mathbb{R}\}$ et T l'application $e^{i\alpha} \mapsto e^{i(\alpha+1)} = e^i e^{i\alpha}$. C'est une rotation : si on espère pouvoir définir une mesure μ sur X qui étend la notion de longueur d'un arc de cercle, elle devrait être préservée par T, c'est-à-dire que pour tout A on aurait $\mu(A) = \mu(T(A))$.

Par ailleurs, il n'est pas dur de vérifier que, pour tout x, on a $T^n(x) \neq T^m(x)$ dès que $n \neq m$; ainsi chaque ensemble $\{T^n(x) \colon n \in \mathbb{Z}\}$ (l'*orbite* de x) est infini. D'autre part, pour tout x, y, soit x et y ont la même orbite, soit leurs orbites sont disjointes ; imaginons un ensemble A obtenu en choisissant exactement un point dans chaque orbite. Alors, pour tout $n \neq 0$, $A \cap T^n(A) = \emptyset$: si $x \in A$ et $x \in T^n(A)$, alors x et $T^{-n}(x)$ sont deux éléments distincts de la même orbite, qui appartiennent à A tous les deux, ce qui est interdit. Et pour tout x on doit avoir un élément de l'orbite de x qui appartient à A, ce qui revient à dire que x appartient à un $T^n(A)$ pour un $n \in \mathbb{Z}$.

Tout ce qu'on vient de dire signifie que

$$X = \bigsqcup_{n \in \mathbb{Z}} T^n(A) .$$

Mais on voudrait avoir $\mu(T^n(A)) = \mu(A)$ pour tout n ; alors, si $\mu(A) > 0$, la définition d'une mesure (et le fait qu'on peut écrire $\mathbb{Z} = \{n_k \colon k \in \mathbb{N}\}$, ce sur quoi on reviendra) nous donne $\mu(X) = \sum_{n\in\mathbb{Z}} \mu(A) = +\infty$. Ce n'est pas raisonnable pour une mesure supposée étendre la notion de longueur : on devrait avoir $\mu(X) = 2\pi$. Mais si $\mu(A) = 0$ la même équation nous donne cette fois-ci $\mu(X) = \sum_{n\in\mathbb{Z}} 0 = 0$.

On doit donc abandonner l'espoir d'avoir une mesure sur le cercle qui étende la notion de longueur, qui soit invariante par les rotations, et pour laquelle tout ensemble soit mesurable.

Heureusement, on n'a pas vraiment besoin de mesurer absolument toutes les parties : celles que l'on est amené à rencontrer ne sont pas quelconques, mais ont une définition plus ou moins explicite. Mais, si l'on ne peut pas mesurer toutes les parties, alors quelles propriétés doit satisfaire l'ensemble des parties mesurables ?

6.1 Tribus

Définition 6.1. Soit X un ensemble, et $\mathcal{A}$ une famille de parties de X. On dit que $\mathcal{A}$ est une *tribu* si $\mathcal{A}$ satisfait les conditions suivantes :

1. $\emptyset \in \mathcal{A}$ et $X \in \mathcal{A}$.
2. Si $A \in \mathcal{A}$ alors $X \setminus A \in \mathcal{A}$.
3. Si $(A_i)_{i\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$ alors $\bigcup_i A_i \in \mathcal{A}$.

On dit alors que $(X, \mathcal{A})$ est un *espace mesurable*

Par exemple, la famille $\mathcal{P}(X)$ formée par toutes les parties de X est une tribu. La famille $\{\emptyset, X\}$ n'ayant que $\emptyset$ et X comme éléments en est une aussi. Mais il y a bien d'autres exemples ; décrivons-en un.

Exemple. Soit X un ensemble, et B une partie de X. Alors
$$\mathcal{A} = \{A \subseteq X : B \subseteq A \text{ ou } B \subseteq X \setminus A\}$$
est une tribu.

Notons que, si $\mathcal{A}$ est une tribu, alors $\mathcal{A}$ est stable par union finie : si $A_0, \ldots, A_n \in \mathcal{A}$, alors la suite obtenue en posant $A_i = \emptyset$ pour $i \geq n+1$ est une suite d'éléments de $\mathcal{A}$, et
$$\bigcup_{i=0}^{n} A_i = \bigcup_{i=0}^{+\infty} A_i \in \mathcal{A}.$$

On voit aussi que, si (A_i) est une suite de parties de $\mathcal{A}$, alors $\bigcap_{i\in\mathbb{N}} A_i \in \mathcal{A}$: en effet,
$$\bigcap_{i=0}^{+\infty} A_i = X \setminus \bigcup_{i=0}^{+\infty} (X \setminus A_i) \ .$$

Enfin, on voit que si A, B appartiennent à une tribu $\mathcal{A}$, alors $A \setminus B = A \cap (X \setminus B)$ y appartient aussi.

On peut produire beaucoup de tribus grâce à la propriété suivante.

Proposition 6.2. *Soit X un ensemble, et $(\mathcal{A}_i)_{i\in I}$ une famille de tribus sur X. Alors $\mathcal{A} = \bigcap \mathcal{A}_i$ est une tribu.*

La preuve est facile pour peu que vous compreniez ce que peut bien être l'intersection d'une famille de tribus... Elle est donc laissée en exercice (recommandé !).

En particulier, si $\mathcal{E}$ est un ensemble de parties de X, il existe toujours une plus petite tribu contenant $\mathcal{E}$: l'intersection de toutes les tribus contenant $\mathcal{E}$. On l'appelle *tribu engendrée* par $\mathcal{E}$ et on la note $\sigma(\mathcal{E})$. Si $\mathcal{A}$ est une tribu, et $\mathcal{E} \subseteq \mathcal{A}$ est tel que $\mathcal{A} = \sigma(\mathcal{E})$, on dit que $\mathcal{A}$ est *engendrée par* $\mathcal{E}$.

Définition 6.3. Soit (X, d) un espace métrique. La *tribu borélienne* sur X est la tribu engendrée par les ouverts, c'est-à-dire la plus petite tribu contenant tous les ouverts, et est notée $\mathcal{B}(X)$.

En pratique, on va surtout manipuler les tribus boréliennes sur les espaces $\mathbb{R}^n$.

On voudrait maintenant expliquer le théorème suivant.

Théorème 6.4. *Soit O un ouvert de $\mathbb{R}$. Alors il existe une suite (I_n) d'intervalles ouverts telle que*
$$O = \bigcup_{n\in\mathbb{N}} I_n \ .$$

Avant de passer à sa démonstration, notons une conséquence.

Exercice 6.5. La tribu borélienne sur $\mathbb{R}$ est la tribu engendrée par les intervalles ouverts.

Définition 6.6. Un ensemble A est *dénombrable* s'il existe une bijection $f : \mathbb{N} \to A$, c'est-à-dire si l'on peut écrire $A = \{a_n : n \in \mathbb{N}\}$ avec des a_i deux à deux distincts.

Théorème 6.7.
1. *Si A est infini et il existe une injection de A dans $\mathbb{N}$, alors A est dénombrable.*
2. *Si A est infini et il existe une surjection de $\mathbb{N}$ sur A, alors A est dénombrable.*
3. *Si A, B sont dénombrables alors $A \times B$ est dénombrable.*
4. *Si $(A_i)_{i\in\mathbb{N}}$ est une suite d'ensembles dénombrables alors $\bigcup_{i\in\mathbb{N}} A_i$ est dénombrable.*

Corollaire 6.8. $\mathbb{Q}$ est dénombrable. (Preuve en exercice.)

Passons maintenant à la preuve du théorème 6.4.

Preuve:

Soit O un ouvert de $\mathbb{R}$. Commençons par fixer une énumération de tous les intervalles ouverts dont les deux extrémités sont rationnelles : $\{]a, b[: a, b \in \mathbb{Q}\} = \{I_n\}_n \in \mathbb{N}$ (c'est possible exactement parce que $\mathbb{Q}$, et donc $\mathbb{Q} \times \mathbb{Q}$, est dénombrable). Soit $A = \{n \in \mathbb{N} : I_n \subseteq O\}$. Pour conclure, il nous suffit de montrer que $O = \bigcup_{n\in A} I_n$; l'inclusion de droite à gauche est immédiate (pour $n \in A$, $I_n \subseteq O$, et donc leur réunion est aussi contenue dans O). Pour voir l'inclusion réciproque, fixons $x \in O$. Comme O est ouvert, il existe $r > 0$ tel que $]x - r, x + r[\subseteq O$. Par densité de $\mathbb{Q}$ dans $\mathbb{R}$, il existe q_1, q_2 tels que $x - r < q_1 < x < q_2 < x + r$; alors l'intervalle $]q_1, q_2[$ est un intervalle d'extrémités rationnelles, qui contient x, ce qui montre comme espéré que $x \in \bigcup_{n\in A} I_n$.

Il n'est peut-être pas inutile de rappeler ici le théorème suivant (qui est essentiel pour que la théorie de la mesure puisse avoir un intérêt...).

Théorème 6.9. $\mathbb{R}$ *n'est pas dénombrable.*

6.2 Mesures

Évidemment, un espace mesurable est là pour être mesuré ! Et maintenant, on peut donner la “vraie” définition d'une mesure.

Définition 6.10. Soit $(X, \mathcal{A})$ un espace mesurable. Une *mesure* sur X est une application $\mu\colon \mathcal{A} \to [0, +\infty]$ vérifiant les propriétés suivantes :

1. $\mu(\emptyset) = 0$.
2. Si $(A_i)_{i\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$ et $A_i \cap A_j = \emptyset$ pour tout $i \neq j$ alors
$$\mu(\bigcup_{i=0}^{+\infty} A_i) = \sum_{i=0}^{+\infty} \mu(A_i) \ .$$

Dans le cas particulier où $\mu(X) = 1$, on dit que μ est une *mesure de probabilité*.

La seconde propriété ci-dessus est appelée *σ-additivité* de la mesure.

Définition 6.11. Comme on autorise que des parties aient une mesure infinie (intuitivement, la mesure généralise les notions de longueur/d'aire/de volume, et on « voit »bien que $\mathbb{R}$, par exemple, est de longueur infinie), il faut donner quelques précisions :

— Pour tout $t \in [0, +\infty]$, $+\infty + t = t + \infty = +\infty$.
— Pour tout t *fini*, $+\infty - t = +\infty$.
— $(+\infty) - (+\infty)$ n'est pas défini.

Dans l'énoncé des propriétés de l'intégrale, on va aussi devoir écrire des multiplications, avec les conventions suivantes :

— $0 \cdot (+\infty) = (+\infty) \cdot 0 = 0$.
— Si $t > 0$, $t \cdot (+\infty) = (+\infty) \cdot t = +\infty$.

Exemple. On peut définir une mesure sur $(X, \mathcal{P}(X))$ en posant $\mu(A) = 0$ pour tout A, ou encore $\mu(\emptyset) = 0$ et $\mu(A) = +\infty$ dès que A est non vide.

Ces mesures ne sont pas très intéressantes ! Un peu plus utile : étant donné $x \in X$ fixé, on peut considérer la *mesure de Dirac* δ_x définie sur $\mathcal{P}(X)$ par :
$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases} \ .$$

Sur $\mathcal{P}(\mathbb{N})$, on peut aussi définir la *mesure de comptage* μ, en définissant :
$$\mu(A) = \begin{cases} \mathrm{card}(A) & \text{si A est fini} \\ +\infty & \text{sinon} \end{cases} \ .$$

Proposition 6.12. *Soit μ une mesure sur un ensemble X.*

1. *Si $A_0, \dots, A_n$ sont des parties deux à deux disjointes et mesurables, alors*
$$\mu(\bigcup_{i=0}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i) \ .$$
2. *Si $A \subseteq B$ et A et B sont mesurables alors $\mu(A) \leq \mu(B)$.*

Preuve:

Pour voir que la première propriété est vraie, il suffit de définir $A_{n+1} = A_{n+2} = \ldots = \emptyset$ et d'appliquer la définition d'une mesure :
$$\mu(\bigcup_{i=0}^{n} A_i) = \mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i=0}^{+\infty} \mu(A_i) = \sum_{i=0}^{n} \mu(A_i) + \sum_{i=n+1}^{+\infty} \mu(A_i) = \sum_{i=0}^{n} \mu(A_i) + 0 = \sum_{i=0}^{n} \mu(A_i) \ .$$

Pour la deuxième propriété on utilise le fait que $B = A \cup (B \setminus A)$ et que $B \setminus A$ est mesurable pour écrire :
$$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A) \ .$$

Attention : si $A \subseteq B$ et A et B sont mesurables, on a envie d'écrire $\mu(A) = \mu(B) - \mu(B \setminus A)$. Ceci peut ne pas avoir de sens, si $\mu(B \setminus A)$ est infini ! Avant d'écrire une telle égalité, il faut vérifier que $\mu(B \setminus A) < +\infty$.

Regroupons quelques propriétés utiles des mesures.

Proposition 6.13. *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré.*

1. *Si $A, B \in \mathcal{A}$ et $A \subseteq B$ alors $\mu(B) \geq \mu(A)$ (on dit que μ est* croissante*).*
2. *Si $A, B \in \mathcal{A}$, on a toujours $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$; si $\mu(A \cap B) < +\infty$ alors on peut aussi écrire $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.*
3. *Si $(A_i)_{i \in \mathbb{N}}$ sont telles que $A_i \in \mathcal{A}$ pour tout $i \in \mathbb{N}$, alors on a*
$$\mu\left(\bigcup_{i=0}^{+\infty} A_i\right) \leq \sum_{i=0}^{+\infty} \mu(A_i) \ .$$
4. *Si $(A_i)_{i \in \mathbb{N}}$ sont telles que $A_i \in \mathcal{A}$ et $A_i \subseteq A_{i+1}$ pour tout $i \in \mathbb{N}$; alors*
$$\mu\left(\bigcup_{i=0}^{+\infty} A_i\right) = \lim_{i \to +\infty} \mu(A_i) \ .$$
5. *Si $(A_i)_{i \in \mathbb{N}}$ sont telles que $A_i \in \mathcal{A}$ et $A_{i+1} \subseteq A_i$, et $\mu(A_0)$ est finie, alors*
$$\mu\left(\bigcap_{i=0}^{+\infty} A_i\right) = \lim_{i \to +\infty} \mu(A_i) \ .$$

Attention, pour la dernière propriété, il faut supposer que A_0 (ou un des A_i) soit de mesure finie pour que le résultat soit vrai en général : par exemple, dans $\mathbb{R}$ muni de la mesure de Lebesgue, considérons $A_i = [i, +\infty[$. Alors $\mu(A_i) = +\infty$ pour tout i, mais $\bigcap A_i = \emptyset$ et donc $\mu(\bigcap A_i) = 0$.

Preuve:

On a déjà démontré (1).

Pour démontrer (2), on commence par écrire que $A \cup B$ est la réunion disjointe de $A \setminus B$, $A \cap B$, et $B \setminus A$, pour obtenir

$$\begin{aligned} \mu(A \cup B) &= \mu(A \setminus B) + \mu(A \cap B) + \mu(B \setminus A) \\ &= \mu(A) + \mu(B \setminus A) \ . \end{aligned}$$

La deuxième égalité vient du fait que A est la réunion disjointe de $A \setminus B$ et $A \cap B$. En ajoutant $\mu(A \cap B)$, on obtient

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B \setminus A) + \mu(A \cap B) = \mu(A) + \mu(B) \ .$$

Pour démontrer (3), on peut par exemple commencer par introduire, pour tout $i \in \mathbb{N}$,

$$B_i = A_i \setminus \bigcup_{j<i} A_j \ .$$

Alors chaque B_i appartient à $\mathcal{A}$, les B_i sont deux à deux disjointes et leur réunion est égale à $\bigcup_i A_i$ (pourquoi ?) ; enfin, $B_i \subseteq A_i$ donc $\mu(B_i) \leq \mu(A_i)$ pour tout i. En utilisant la σ-additivité, on obtient :

$$\mu(\bigcup_{i=0}^{+\infty} A_i) = \mu(\bigcup_{i=0}^{+\infty} B_i) = \sum_{i=0}^{+\infty} \mu(B_i) \leq \sum_{i=0}^{+\infty} \mu(A_i) \ .$$

Pour (4), on procède de la même façon, en introduisant les mêmes B_i que ci-dessus et en utilisant que, comme la suite (A_i) est croissante, on a pour tout k

$$A_k = \bigcup_{i=0}^{k} A_i = \bigcup_{i=0}^{k} B_i \ .$$

Mais alors on a aussi, pour tout k :

$$\mu(A_k) = \sum_{i=0}^{k} \mu(B_i) \ .$$

Et alors, quand k tend vers $+\infty$, on voit que $\mu(A_k)$ tend vers

$$\sum_{k=0}^{+\infty} \mu(B_i) = \mu\left(\bigcup_{i=0}^{+\infty} B_i\right) = \mu\left(\bigcup_{i=0}^{+\infty} A_i\right) \ .$$

Enfin, pour montrer (5), on peut par exemple poser $C_i = A_0 \setminus A_i$, pour tout i ; alors $C_i \in \mathcal{A}$, la suite C_i est croissante, $\mu(C_i) = \mu(A_0) - \mu(A_i)$ pour tout i, et

$$\bigcup_{i=0}^{+\infty} C_i = A_0 \setminus \bigcap_{i=0}^{+\infty} A_i \ .$$

De plus, le résultat précédent nous dit que

$$\lim_{i\to+\infty} \mu(C_i) = \mu(\bigcup_{i=0}^{+\infty} C_i) \ .$$

Autrement dit,

$$\lim_{i\to+\infty} (\mu(A_0) - \mu(A_i)) = \mu(A_0) - \mu(\bigcap_{i=0}^{+\infty} A_i) \ .$$

Étant donnée une mesure, on a à disposition une notion de « grosseur » : plus la mesure est grande, plus la partie est grosse. Si la mesure est nulle, alors la partie est même *négligeable* : du point de vue de la mesure, elle n'est pas différente de l'ensemble vide. Bien sûr, différentes mesures donnent lieu à différentes notions de grosseur : la mesure de Dirac δ_x pense qu'une partie est négligeable si elle ne contient pas x, et aussi grosse que l'espace entier si elle contient x ; la mesure de comptage prend en compte le nombre d'éléments d'un ensemble ; et la mesure de Lebesgue généralise la notion de longueur, d'une façon qu'il est parfois difficile de se représenter.

Définition 6.14. La mesure de Lebesgue sur $\mathbb{R}^n$ est l'unique mesure définie sur la tribu borélienne de $\mathbb{R}^n$ et telle que

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

On admettra à la fois qu'une telle mesure existe, et qu'elle est unique !

Proposition 6.15. *1. Pour tout $x \in \mathbb{R}$ on a $\lambda(\{x\}) = 0$.*

2. Si (x_n) est une suite de nombres réels, alors $\lambda(\{x_n : n \in \mathbb{N}\}) = 0$.

3. Pour tous réels a, b avec $a \leq b$ on a $\lambda([a,b]) = \lambda(]a,b[) = \lambda([a,b[) = \lambda(]a,b]) = b - a$.

Une idée fondamentale en théorie de la mesure est qu'on peut, la plupart du temps, *négliger* les parties de mesure nulle : elle n'ont pas d'influence sur la valeur d'une intégrale, par exemple.

Définition 6.16. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. Une partie $A \subseteq X$ est dite *négligeable* si A est contenue dans ensemble de mesure nulle, c'est-à-dire s'il existe $B \in \mathcal{A}$ tel que $A \subseteq B$ et $\mu(B) = 0$.

Inversement, on dit que *presque tout* x appartient à A, ou encore que $x \in A$ est *vrai presque partout*, si $X \setminus A$ est négligeable, c'est-à-dire s'il existe $B \in \mathcal{A}$ tel que $B \subseteq A$ et $\mu(X \setminus B) = 0$.

Exercice 6.17. Montrer que $\mathbb{Q}$ est négligeable dans $\mathbb{R}$.

Chapitre 7

Fonctions mesurables et intégration

On va maintenant décrire brièvement quelles fonctions définies sur un espace mesuré on peut intégrer. L'idée est que, si $(X, \mathcal{A}, \mu)$ est un espace mesuré et f est la fonction caractéristique d'une partie $A \in \mathcal{A}$, alors on voudrait poser $\int_X f d\mu = \mu(A)$. Une fois qu'on a fait ça, on sait comment intégrer toutes les fonctions qui s'écrivent sous la forme $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$: l'intégrale d'une telle fonction devra être $\sum \alpha_i \mu(A_i)$. Et un procédé de passage à la limite nous permettra d'intégrer encore plus de fonctions - les fonctions pour lesquelles ce processus de passage à la limite fonctionne bien sont les *fonctions mesurables.*

7.1 Fonctions mesurables

Définition 7.1. Soit $(X, \mathcal{A})$ un espace mesurable. On dit que $f \colon X \to \mathbb{R}$ est *mesurable* si, pour tout ouvert O de $\mathbb{R}$, $f^{-1}(O) \in \mathcal{A}$.

Plus généralement, si $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ sont des espaces mesurables, on dit que $f \colon X \to Y$ est mesurable si $f^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{B}$.

Exercice 7.2. Montrer que, si $(X, \mathcal{A})$ est un espace mesurable, $f \colon X \to \mathbb{R}$ est mesurable si et seulement si $f \colon (X, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable.

Montrer qu'on obtient encore la même définition en demandant simplement que $f^{-1}(I) \in \mathcal{A}$ pour tout intervalle ouvert I.

Exercice 7.3. Soit $(X, \mathcal{A})$ un espace mesurable, et f la fonction caractéristique d'une partie $A \in \mathcal{A}$, vue comme une fonction de X dans $\mathbb{R}$. Montrer que f est mesurable.

Un cas particulier est spécialement important : celui où f est une fonction de $\mathbb{R}^n$, muni de sa tribu borélienne, dans $\mathbb{R}$, muni de sa tribu borélienne.

Définition 7.4. Soit $n, m \in \mathbb{N}^*$ et $f \colon \mathbb{R}^n \to \mathbb{R}^m$. On dit que f est *borélienne* si f est mesurable de $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ dans $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$.

Rappelons qu'une fonction $f \colon \mathbb{R} \to \mathbb{R}$ est continue si, et seulement si, $f^{-1}(O)$ est ouvert pour tout ouvert O de $\mathbb{R}$. Par conséquent, *toute fonction continue est borélienne* ; la réciproque n'est pas vraie : la fonction caractéristique de $[0, 1]$ (ou de n'importe quelle autre borélien) est borélienne, mais pas du tout continue.

Une bonne raison de s'autoriser à considérer ces fonctions plus générales est qu'on est souvent confronté à des limites simples de suites de fonctions, qui en général ne sont pas continues...

Proposition 7.5. *Soit $(X, \mathcal{A})$ un espace mesurable, et $f_n \colon X \to \mathbb{R}$ une suite de fonctions mesurables. On suppose que $(f_n(x))$ converge pour tout $x \in X$, vers un réel que l'on note $f(x)$. Alors f est mesurable.*

Démonstration. Soit $I =]a, b[$ un intervalle ouvert. Si $f(x) \in I$, alors $f_n(x)$ appartient à I pour tout n assez grand, par définition de la limite. Réciproquement, si $f_n(x) \in I$ pour tout n suffisamment grand, alors $f(x) \in [a, b]$. Ainsi, on voit que $f(x) \in I$ si, et seulement si, il existe $\varepsilon > 0$ tel que $f_n(x) \in]a + \varepsilon, b - \varepsilon[$ pour tout n suffisamment grand ; menant à l'égalité suivante :

$$f^{-1}(]a, b[) = \bigcup_{N \in \mathbb{N}^*} \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \left\{ x \colon f_m(x) \in \left] a + \frac{1}{N}, b - \frac{1}{N} \right[\right\} .$$

Chacun des ensembles $\{x\colon f_m(x) \in \left]a+\frac{1}{N}, b-\frac{1}{N}\right[\}$ appartient à $\mathcal{A}$ puisque f_m est mesurable ; une intersection d'une suite d'éléments de $\mathcal{A}$ est encore un élément de $\mathcal{A}$, et de même pour une réunion, ce qui nous donne la conclusion souhaitée. $\square$

Proposition 7.6. *Soit $(X,\mathcal{A})$, $(Y,\mathcal{B})$ et $(Z,\mathcal{C})$ trois espaces mesurables. Supposons que $f\colon X \to Y$ et $g\colon Y \to Z$ sont mesurables. Alors $g \circ f$ est mesurable.*

Preuve:

Soit $C \in \mathcal{C}$. Alors

$$(g \circ f)^{-1}(C) = \{x \in X\colon g(f(x)) \in C\} = \{x \in X\colon f(x) \in g^{-1}(C)\} = f^{-1}(g^{-1}(C)) \ .$$

Puisque g est mesurable, $g^{-1}(C) \in \mathcal{B}$; et alors le fait que f est mesurable nous donne que $f^{-1}(g^{-1}(C)) \in \mathcal{A}$. On vient de montrer que $(g\circ f)^{-1}(C) \in \mathcal{A}$ pour tout $C \in \mathcal{C}$, autrement dit que $g\circ f$ est mesurable.

Exercice 7.7. Montrer que l'addition $(x,y) \mapsto x+y$ et la multiplication $(x,y) \mapsto xy$ de $\mathbb{R}^2$ dans $\mathbb{R}$ sont boréliennes.

Exercice 7.8. Soit $(X,\mathcal{A})$ un espace mesurable, et f,g deux fonctions mesurables de X dans $\mathbb{R}$. Montrer que $f+g$ et fg sont mesurables.

Rappelons que, si $f,g\colon \mathbb{R} \to \mathbb{R}$ sont des fonctions boréliennes, on dit que f et g sont égales *presque partout* si $\lambda(\{x\colon f(x) \neq g(x)\}) = 0$ (c'est-à-dire, si l'ensemble de tous les x tels que $f(x) \neq g(x)$ est négligeable, ou encore si $f(x) = g(x)$ pour presque tout x)

Exemple. La fonction f définie sur $\mathbb{R}$ par $f(x) = \begin{cases} 0 & \text{si } x \notin \mathbb{N} \\ x & \text{sinon} \end{cases}$ est borélienne et nulle presque partout.

Par la suite, il sera utile pour nous d'autoriser certaines fonctions à prendre la valeur $+\infty$.

Définition 7.9. Soit $(X,\mathcal{A})$ un espace mesurable. Une *fonction mesurable à valeurs positives* est une fonction $f\colon X \to [0,+\infty]$ telle que :
— $f^{-1}(\{+\infty\})$ soit mesurable, et
— Pour tout borélien B de $[0,+\infty[$, $f^{-1}(B)$ soit mesurable.

7.2 Intégrale des fonctions mesurables

Dans cette partie, on fixe un espace mesuré $(X,\mathcal{A},\mu)$. On commence par le cas particulier des fonctions mesurables à valeurs positives, qui sont exactement les fonctions de la forme

$$f = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i} \ ,$$

où chaque A_i appartient à $\mathcal{A}$ et $\alpha_i \geq 0$ (attention, éventuellement, un des α_i peut être égal à $+\infty$!). Pour une telle fonction, on pose

$$\int_X f d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i) \ .$$

Il faut faire un peu attention, et vérifier que la définition précédente ne dépend pas de la façon dont on écrit f sous la forme $\sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$...

Ensuite, un procédé de passage à la limite que l'on ne décrira pas ici nous permet de définir l'intégrale d'une fonction mesurable et à valeurs dans $[0,+\infty]$; cette intégrale peut valoir $+\infty$! Puis, en écrivant une fonction mesurable quelconque f sous la forme $f = f^+ - f^-$, on arrive à définir l'intégrale des fonctions mesurables f, à la condition que $\int_X |f| d\mu < +\infty$. On dit que ces fonctions sont *intégrables*.

Étant donnée une partie mesurable $A \subseteq X$, et une fonction mesurable $f\colon X \to \mathbb{R}$,

$$\int_A f d\mu = \int_X f \mathbf{1}_A d\mu.$$

Résumons ici les propriétés de l'intégrale ainsi obtenue, que l'on admettra :

Proposition 7.10. *Pour toutes fonctions mesurables $f, g\colon X \to \mathbb{R}$ intégrables sur une partie mesurable $A \subseteq X$ on a*

1. $\int_A 1 d\mu = \mu(A)$;
2. *pour toutes constantes réelles α et β, la fonction $\alpha f + \beta g$ est également mesurable et intégrable sur A et*
$$\int_A (\alpha f + \beta g) d\mu = \alpha \int_A f d\mu + \beta \int_A g d\mu \text{ (linéarité de l'intégrale);}$$
3. *si $f = g$ presque partout, alors $\int_A f d\mu = \int_A g d\mu$;*
4. *si $f \geq 0$ presque partout, alors $\int_A f d\mu \geq 0$* (positivité de l'intégrale) *;*
   si $f \leq g$ presque partout, alors $\int_A f d\mu \leq \int_A g d\mu$ (monotonie de l'intégrale) *;*
5. *si $\mu(A) = 0$ alors $\int_A f d\mu = 0$;*
6. *si A_1, A_2 sont des parties mesurables de A telles que $\mu(A_1 \cap A_2) = 0$, alors*
$$\int_{A_1 \cup A_2} f d\mu = \int_{A_1} f d\mu + \int_{A_2} f d\mu \ .$$

On a de plus

7. L'inégalité triangulaire
$$\left| \int_A (f+g) d\mu \right| \leq \int_A |f| d\mu + \int_A |g| d\mu \ .$$

Remarque 7.11. Rappelons que l'intégrale est d'abord définie pour les fonctions mesurables à valeurs positives ($+\infty$ compris) et possède les propriétés ci-dessus pour ces fonctions. Répétons également que, quand on écrit que f est une fonction mesurable de X à valeurs dans l'intervalle fermé $[0, +\infty]$, on entend que $A = \{x \in X\colon f(x) = +\infty\}$ est mesurable, et que pour tout intervalle ouvert I de $\mathbb{R}$, $f^{-1}(I)$ est mesurable. Dans ce cas, l'intégrale de f peut valoir $+\infty$. Si f est intégrable (c.à.d. si $\int_X f d\mu < +\infty$) alors, par monotonie, la partie A où f vaut $+\infty$ est négligeable (c.à.d. $\mu(A) = 0$).

Proposition 7.12 (Inégalité de Tchebychev)**.** *Soit $f\colon X \to [0, +\infty]$ une fonction mesurable. Alors, pour tout $\alpha > 0$, on a*
$$\mu(\{x\colon f(x) \geq \alpha\}) \leq \frac{1}{\alpha} \int_X f d\mu \ .$$

Preuve:

Notons $B = \{x \in X\colon f(x) \geq \alpha\}$. Alors $\alpha \mathbf{1}_B \leq f$ et par monotonie de l'intégrale, on a
$$\alpha \mu(B) \leq \int_X f d\mu \ .$$

Les propriétés de l'intégrale rappelées dans la proposition 7.10 se vérifient directement à partir de la définition de l'intégrale, à l'exception de la linéarité ou plus précisément de la propriété $\int_A (f+g) d\mu = \int_A f d\mu + \int_A g d\mu$ qui est un corollaire du théorème de convergence monotone. Pour montrer ce dernier, nous allons utiliser la propriété supplémentaire suivante que nous admettrons. Cette propriété se vérifie tout d'abord pour les fonctions simples positives de la forme $\sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$, puis pour toutes les fonctions positives intégrables par le procédé de passage à la limite permettant de définir l'intégrale.

Proposition 7.13 (Mesures à densité)**.** *Soit $f\colon X \to [0, +\infty]$ une fonction mesurable. On définit une application $\nu : \mathcal{A} \to [0, +\infty]$ par*
$$\nu(A) = \int_A f d\mu \ .$$
Alors, ν est une mesure sur X, appelée mesure de densité *f par rapport à μ.*

Exemple. 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*.

Théorème 7.14 (Théorème de convergence monotone)**.** *Soit $f_n\colon X \to [0,+\infty]$ une suite de fonctions mesurables telle pour presque tout $x \in X$ la suite $(f_n(x))$ soit une suite croissante. Alors il existe une fonction mesurable $f\colon X \to [0,+\infty]$ telle que $f_n(x)$ converge vers $f(x)$ presque partout, et on a*

$$\lim_{n\to+\infty} \int_X f_n d\mu = \int_X f d\mu \ .$$

Preuve:

On peut supposer que pour tout $x \in X$, la suite $(f_n(x))$ est une suite croissante (la partie où ce n'est pas le cas étant négligeable). Dans ce cas, pour chaque $x \in X$, la suite $(f_n(x))$ a une limite (finie ou infinie) que l'on note $f(x)$. Cette limite f est alors mesurable (voir la preuve de la proposition 7.5 et la remarque 7.11).

De plus, pour tout $n \in \mathbb{N}$, on a $f_n \leq f_{n+1} \leq f$ et donc, par monotonie de l'intégrale, la suite $\int_X f_n d\mu$ a une limite vérifiant

$$\lim_{n\to+\infty} \int_X f_n d\mu \leq \int_X f d\mu \ .$$

Pour l'inégalité réciproque, considérons $\alpha \in]0,1[$. Pour $n \in \mathbb{N}$, posons $E_n = \{x \in X\colon f_n(x) \geq \alpha f(x)\}$. Alors, par linéarité, chaque E_n est mesurable. De plus, par hypothèse, on a $E_n \subseteq E_{n+1}$ pour tout $n \in \mathbb{N}$, et $\cup_{n\in\mathbb{N}} E_n = X$. Comme $A \mapsto \int_A \alpha f d\mu$ définie une mesure, on en déduit par la proposition 6.13 4 que

$$\int_X \alpha f d\mu = \lim_{n\to+\infty} \int_{E_n} \alpha f d\mu \leq \lim_{n\to+\infty} \int_{E_n} f_n d\mu \leq \lim_{n\to+\infty} \int_X f_n d\mu \ .$$

En faisant tendre α vers 1, on conclut que

$$\int_X \alpha f d\mu \leq \lim_{n\to+\infty} \int_X f_n d\mu \ .$$

Théorème 7.15 (Théorème de convergence dominée)**.** *Soit $f_n\colon X \to \mathbb{R}$ une suite de fonctions mesurables satisfaisant les hypothèses suivantes :*

1. *Pour presque tout $x \in X$, la suite $f_n(x)$ est convergente vers une limite qu'on appelle $f(x)$.*
2. *Il existe une fonction $g\colon X \to [0,+\infty]$ mesurable et intégrable, telle que pour presque tout x on ait*

$$\forall n \in \mathbb{N}\ ,\ |f_n(x)| \leq g(x).$$

Alors f est intégrable (ainsi que les f_n), et on a

$$\int_X f d\mu = \lim_{n\to+\infty} \int_X f_n d\mu \ .$$

Rappelons que dans le théorème ci-dessus, il est fondamental que *la fonction g ne dépende pas de n* et que g soit intégrable, c'est-à-dire $\int_X g d\mu < \infty$.

Dans l'énoncé, l'application f est définie presque partout et on la définit par 0 (ou par une fonction mesurable quelconque) sur le complémentaire qui est négligeable. Par la proposition 7.5, la fonction f est mesurable.

Preuve:

Pour tout $n \in \mathbb{N}$, on a $|f_n| \leq g$ presque partout et de même $|f| \leq g$ presque partout par passage à la limite. Comme g est intégrable, par monotonie les f_n et f le sont également.

Pour simplifier, on suppose par la suite la convergence simple et les inégalités vraies partout. On pose pour chaque $n \in \mathbb{N}$,

$$h_n = \inf_{k\geq n} (2g - |f - f_k|).$$

On peut alors vérifier que les fonctions h_n sont mesurables, positives et majorées par $2g$. De plus, pour tout x, la suite $h_n(x)$ est croissante et converge vers $2g$. Par le théorème de convergence monotone, on obtient que $\lim\limits_{n\to+\infty} \int_X h_n d\mu = \int_X 2g d\mu$.

Par définition, on a pour tout n :

$$\int_X h_n d\mu \leq \int_X (2g - |f - f_n|) d\mu \leq \int_X 2g d\mu \ .$$

Ainsi,

$$\left|\int_X f d\mu - \lim_{n\to+\infty}\int_X f_n d\mu\right| = \lim_{n\to+\infty}\left|\int_X (f-f_n)d\mu\right| \leq \lim_{n\to+\infty}\int_X |f-f_n|d\mu = 0.$$

Ces théorèmes ont pour corollaires les théorèmes analogues d'échanges séries/intégrales et de continuité ou dérivabilité des intégrales à paramètres vus dans la première partie pour la mesure de Lebesgue.

Discutons quelques exemples.

1. Si $X=\mathbb{N}$, $\mathcal{A}=\mathcal{P}(\mathbb{N})$ et μ est la mesure de comptage sur $\mathbb{N}$, que vaut l'intégrale dans ce cas ? Pour le comprendre, commençons par le cas d'une fonction positive $f\colon \mathbb{N}\to[0,+\infty[$, et même, par un cas simple, où il existe N tel que $f(n)=0$ pour tout $n\geq N$. Dans ce cas, f ne prend qu'un nombre fini de valeurs $x_1,\ldots,x_k$ et on a par définition de l'intégrale

$$\int_{\mathbb{N}} f d\mu = \sum_{i=1}^{k} x_i \left|\{p\in\mathbb{N}\colon f(p)=x_i\}\right|$$

Dans la somme ci-dessus, x_i apparaît exactement autant de fois qu'il existe d'entiers p pour lesquels $f(p)=x_i$; autrement dit, la somme ci-dessus est simplement la somme de toutes les valeurs de f, c'est-à-dire

$$\int_{\mathbb{N}} f d\mu = \sum_{p=0}^{N} f(p)\ .$$

On vient de voir que, au sens de la mesure de comptage sur $\mathbb{N}$, on a $\int_N f d\mu = \sum_{p=0}^{+\infty} f(p)$, du moment que f est à valeurs positives et nulle pour n suffisamment grand. Maintenant, si jamais $f\colon\mathbb{N}\to[0,+\infty[$ est une fonction quelconque, alors on peut considérer la suite (f_N) définie par

$$f_N(i)=\begin{cases} f(i) & \text{si } i\leq N\\ 0 & \text{sinon}\end{cases}$$

Pour tout i fixé, la suite $(f_N(i))$ est croissante vers $f(i)$; on peut donc appliquer le théorème de convergence monotone pour déduire que

$$\int_{\mathbb{N}} f d\mu = \lim_{N\to+\infty}\int_{\mathbb{N}} f_n d\mu = \lim_{N\to+\infty}\sum_{i=0}^{N} f(i) = \sum_{i=0}^{+\infty} f(i)\ .$$

Ainsi, on a $\displaystyle\int_{\mathbb{N}} f d\mu = \sum_{i=0}^{+\infty} f(i)$ pour toute fonction positive f ; on en déduit qu'une fonction $f\colon\mathbb{N}\to\mathbb{R}$ est intégrable si, et seulement si, la série $\displaystyle\sum_{i=0}^{+\infty}|f(i)|$ converge (ou encore la série $\displaystyle\sum_{i=0}^{+\infty} f(i)$ est absolument convergente) , et que dans ce cas-là, on a encore $\displaystyle\int_{\mathbb{N}} f d\mu = \sum_{i=0}^{+\infty} f(i)$.

Par conséquent, la sommation de séries absolument convergentes peut être vue comme un cas particulier de calcul d'intégrale, relativement à la mesure de comptage sur $\mathbb{N}$.

2. Supposons maintenant que $(X,\mathcal{A},\mu)$ est un espace mesuré quelconque, $f\colon X\to[0,+\infty]$ est mesurable et ν est la mesure de densité f par rapport à μ. Alors, on sait que, pour toute partie mesurable A, on a

$$\int_X \mathbf{1}_A d\nu = \nu(A) = \int_A f d\mu = \int_X f\mathbf{1}_A d\mu\ .$$

En suivant le même cheminement que ci-dessus (et en admettant le fait que toute fonction mesurable, à valeurs positives, est une limite presque partout de fonctions mesurables, positives et ne prenant qu'un nombre fini de valeurs) on peut alors montrer qu'on a, pour toute fonction mesurable $g\colon X\to[0,+\infty]$:

$$\int_X g d\nu = \int_X fg d\mu\ .$$

Cette formule permet de déduire que les fonctions ν-intégrables sont les fonctions $g\colon X\to\mathbb{R}$ telles que $\displaystyle\int_X f|g|d\mu<+\infty$, et que pour ces fonctions on a aussi $\displaystyle\int_X g d\nu = \int_X fg d\mu$.

7.3 Mesure produit et théorèmes de Fubini

Définition 7.16. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. On dit que $(X, \mathcal{A}, \mu)$ est *σ-fini* s'il existe une suite de parties $(A_n)_{n\in\mathbb{N}}$ telles que $\mu(A_n) < +\infty$ pour tout n, et $X = \bigcup_n A_n$.

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

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 7.17. Soit $(X, \mathcal{A}, \mu_1)$ 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 $\mathcal{B}$.

Alors il existe une unique mesure ν sur $\mathcal{A} \otimes \mathcal{B}$ vérifiant $\nu(A \times B) = \mu_1(A)\mu_2(B)$ pour tout $A \in \mathcal{A}$ et tout $B \in \mathcal{B}$. Cette mesure est notée $\mu_1 \otimes \mu_2$, et est σ-finie.

On n'essaiera pas de rentrer dans le détail de la construction de cette mesure ; notons que $\mathcal{B}(\mathbb{R}^{n+m}) = \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m)$ et que , si λ_n désigne la mesure de Lebesgue sur $\mathbb{R}^n$, alors on a toujours $\lambda_{n+m} = \lambda_n \otimes \lambda_m$.

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 aille de même de l'intégrale d'une fonction mesurable relativement à $\mu_1 \otimes \mu_2$. Et c'est effectivement le contenu des théorèmes de Fubini.

Théorème 7.18 (Fubini–Tonelli)**.** *Soit $(X_1, \mathcal{T}_1, \mu_1)$ et $(X_2, \mathcal{T}_2, \mu_2)$ deux espaces mesurés σ-finis. Soit $f : X_1 \times X_2 \to [0, +\infty]$ une fonction $\mathcal{T}_1 \otimes \mathcal{T}_2$-mesurable. Alors :*

1. *$y \mapsto f(x, y)$ est une fonction mesurable (sur $(X_2, \mathcal{T}_2)$ dans $[0, +\infty]$) pour tout $x \in X_1$, et $x \mapsto \int_{X_2} f(x, y) d\mu_2(y)$ est une fonction mesurable (sur $(X_1, \mathcal{T}_1)$).*
2. *$x \mapsto f(x, y)$ est une fonction mesurable (sur $(X_1, \mathcal{T}_1)$ dans $[0, +\infty]$) pour tout $y \in X_2$, et $y \mapsto \int_{X_1} f(x, y) d\mu_1(x)$ est une fonction mesurable (sur $(X_2, \mathcal{T}_2)$).*
3. *On a*
$$\int_{X_1\times X_2} f(x, y) d\mu_1 \otimes \mu_2(x, y) = \int_{X_1} \left(\int_{X_2} f(x, y) d\mu_2(y) \right) d\mu_1(x) = \int_{X_2} \left(\int_{X_1} f(x, y) d\mu_1(x) \right) d\mu_2(y) \ .$$

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 7.19 (Fubini)**.** *Soit $(X_1, \mathcal{T}_1, \mu_1)$ et $(X_2, \mathcal{T}_2, \mu_2)$ deux espaces mesurés σ-finis. Soit $f : X_1 \times X_2 \to \mathbb{R}$ une fonction* intégrable*. Alors :*

1. *$y \mapsto f(x, y)$ est une fonction intégrable (sur X_2) pour presque tout $x \in X_1$, et $x \mapsto \int_{X_2} f(x, y) d\mu_2(y)$ est une fonction intégrable (sur X_1).*
2. *$x \mapsto f(x, y)$ est une fonction intégrable (sur X_1) pour preseque tout $y \in X_2$, et $y \mapsto \int_{X_1} f(x, y) d\mu_1(x)$ est une fonction intégrable (sur X_2)*
3. *On a*
$$\int_{X_1\times X_2} f(x, y) d\mu_1 \otimes \mu_2(x, y) = \int_{X_1} \left(\int_{X_2} f(x, y) d\mu_2(y) \right) d\mu_1(x) = \int_{X_2} \left(\int_{X_1} f(x, y) d\mu_1(x) \right) d\mu_2(y) \ .$$

Chapitre 8

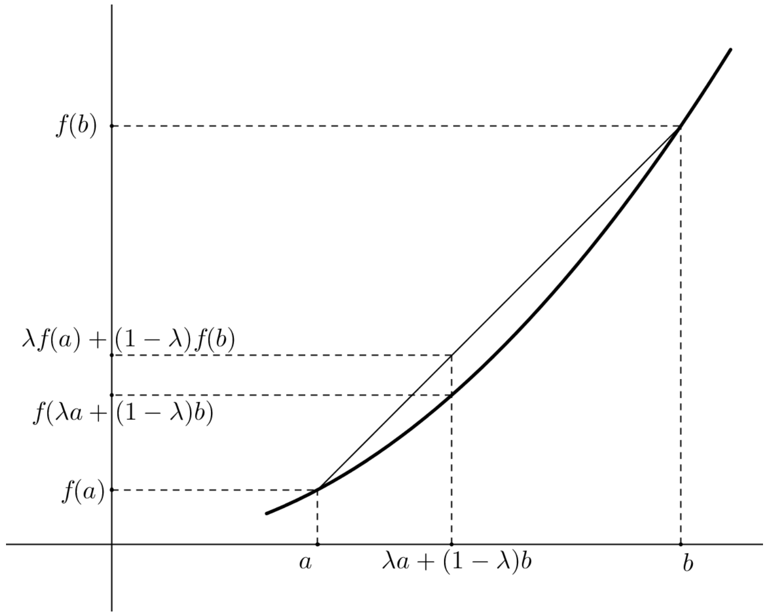
Convexité

8.1 Convexité : définition et premières propriétés

On commence par la définition générale d'une fonction convexe. Nous verrons dans la partie suivante des caractérisations plus simples dans le cas où f est dérivable ou deux fois dérivable.

Définition 8.1. Une fonction $f : I \to \mathbb{R}$ où I est intervalle de $\mathbb{R}$ est dite *convexe* si elle vérifie l'*inégalité de convexité* suivante :

$$\forall a,\, b \in I,\, \forall \lambda \in [0,1],\; f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda) f(b).$$



En observant la représentation graphique ci-dessus, on obtient la traduction géométrique suivante de la convexité d'une fonction :

Propriété 8.2. Soit $f : I \to \mathbb{R}$ et $\mathcal{C}$ sa courbe représentative. Alors f est convexe si et seulement si pour tout couple (A, B) de points distincts de $\mathcal{C}$, l'arc $\overset{\frown}{AB}$ sur $\mathcal{C}$ est au-dessous de la corde $[AB]$.

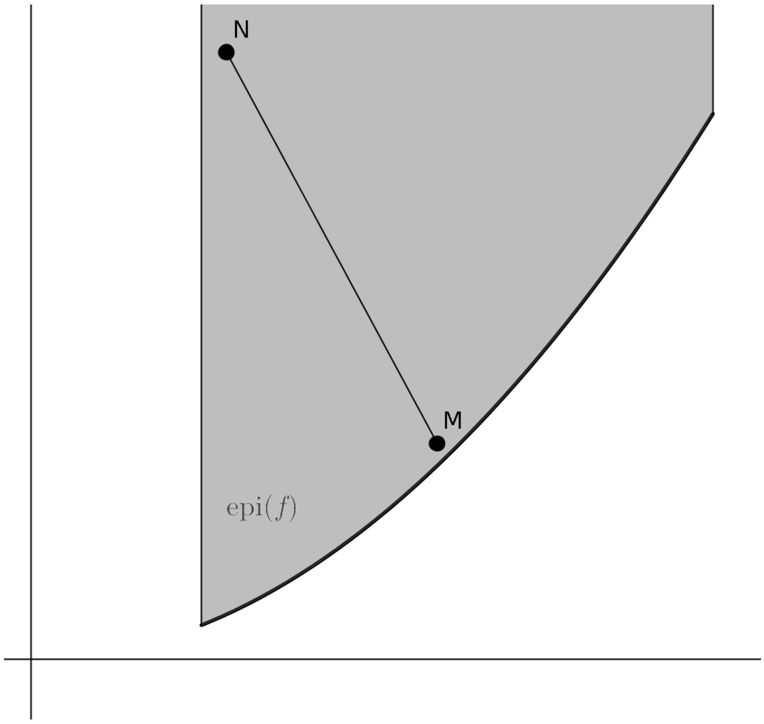
Définition 8.3. Une fonction $g : I \to \mathbb{R}$ où I est intervalle de $\mathbb{R}$ est dite *concave* si $-g$ est convexe, ou encore si elle vérifie l'*inégalité de concavité* suivante :

$$\forall a,\, b \in I,\, \forall \lambda \in [0,1],\; g(\lambda a + (1-\lambda)b) \geq \lambda g(a) + (1-\lambda) g(b);$$

ou de manière équivalente si pour tout couple (A, B) de points distincts de sa courbe représentative, l'arc $\overset{\frown}{AB}$ est au-dessus de la corde $[AB]$.

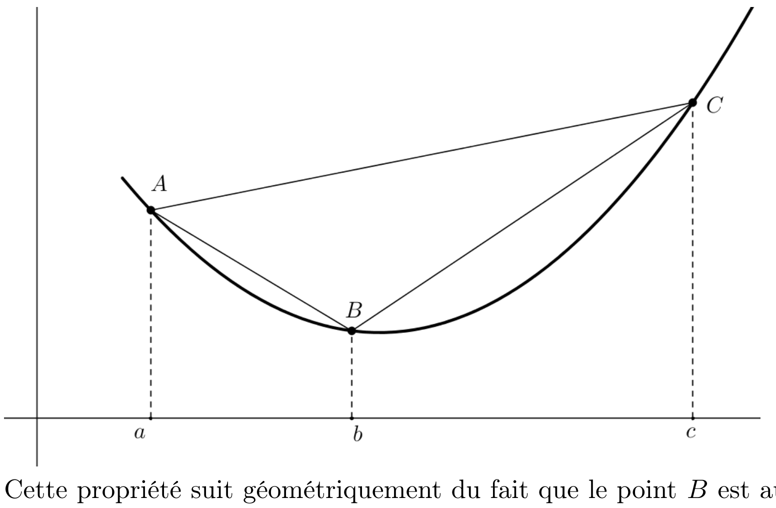
On peut également traduire la convexité d'une fonction par la convexité de son *épigraphe* en tant que partie de $\mathbb{R}^2$:

Propriété 8.4. Soit $f : I \to \mathbb{R}$. On appelle *épigraphe* de f la partie de $\mathbb{R}^2$ notée $\mathrm{epi}(f)$ et définie par $\mathrm{epi}(f) = \{(x, y) \in \mathbb{R}^2 \colon y \geq f(x)\}$. Alors f est convexe si et seulement si $\mathrm{epi}(f)$ est une partie convexe de $\mathbb{R}^2$, c'est-à-dire pour tout couple (M, N) de points de $\mathrm{epi}(f)$, on a $[M, N] \subseteq \mathrm{epi}(f)$.



Propriété 8.5. Une fonction convexe $f : I \to \mathbb{R}$ vérifie l'*inégalité des pentes* :

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



Cette propriété suit géométriquement du fait que le point B est au-dessous de la corde $[AC]$, et est en fait équivalente à la convexité de f.

Exercice 8.6. Démontrer algébriquement l'inégalité des pentes, en utilisant la définition 8.1.

Pour une fonction $f : I \to \mathbb{R}$ et $a \in I$, on considère la fonction $\Delta_a f$ définie par $\Delta_a f(x) = \dfrac{f(x) - f(a)}{x - a}$ pour tout $x \in I \setminus \{a\}$. Les inégalités précédentes (inégalité des pentes) sur une fonction f se traduisent en terme de croissance des fonctions $\Delta_a f$:

Propriété 8.7. 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 $I \setminus \{a\}$.

Exercice 8.8. Vérifier que l'inégalité des pentes est équivalente au fait que pour tout $a \in I$, la fonction $\Delta_a f$ est croissante sur $I \setminus \{a\}$.

Nous allons maintenant démontrer un lemme un peu plus difficile qui permettra de montrer que toute fonction convexe sur un intervalle ouvert est continue et admet des dérivées à gauche et à droite en tout point de cet intervalle ouvert.

Lemme 8.9. *Soit I un intervalle de $\mathbb{R}$ et $a \in \mathring{I}$. Si g est une fonction à valeurs réelles définie et croissante sur $I \setminus \{a\}$ alors g admet en a une limite à gauche et une limite à droite finies telles que*

$$\lim_{\substack{x \to a \\ x < a}} g(x) \leq \lim_{\substack{x \to a \\ x > a}} g(x).$$

Démonstration. Comme $a \in \mathring{I}$, il existe $\eta > 0$ tel que $[a-\eta, a+\eta] \subset I$ et cela a donc un sens d'étudier l'existence d'éventuelles limites à gauche et à droite de g. On considère

$$X = \{g(x) \colon x \in I \text{ et } x < a\}.$$

Par croissance de g, cette partie X est majorée par $g(a+\eta)$. Notons l_1 sa borne supérieure. Encore par croissance de g, il suit que l_1 est la limite à gauche de g en a : en effet, pour $\varepsilon > 0$, il existe $x_0 \in X$ tel que $l_1 - \varepsilon < g(x_0) \leq l_1$ et alors pour tout x tel que $x_0 < x < a$, on a $l_1 - \varepsilon < g(x_0) \leq g(x) \leq l_1$ et en particulier $|l_1 - g(x)| < \varepsilon$.

De manière analogue, l'ensemble $X' = \{g(x') \colon x' \in I \text{ et } x' > a\}$ est minoré et sa borne inférieure l_2 correspond à la limite à droite de g en a. Enfin, comme pour tout $x \in X$ et tout $x' \in X'$, on a $x < a < x'$ et donc $g(x) \leq g(x')$, il suit que $l_1 \leq l_2$.

□

Proposition 8.10. *Soit I un intervalle ouvert et f une fonction à valeurs réelles, définie et convexe sur I. Alors f est continue sur I et admet des dérivées à gauche et à droite en tout point a de I, vérifiant de plus $f'_g(a) \leq f'_d(a)$.*

De plus, pour tout $a, b \in I$ tels que $a < b$, on a $f'_d(a) \leq \frac{f(b)-f(a)}{b-a} \leq f'_g(b)$.

Démonstration. Soit $a \in I$. Comme I est ouvert, on a $a \in \mathring{I}$, et comme f est convexe, la fonction $\Delta_a f$ est croissante. On peut appliquer le lemme précédent sur cette fonction, ce qui permet de conclure car les limites à gauche et droite de $\Delta_a f$ en a correspondent exactement aux dérivées à gauche et à droite de f en a. Bien sûr, la dérivabilité à droite et à gauche en a impliquent en particulier que $f(x) - f(a)$ tend vers 0 quand x tend vers a, et donc que f est continue en a.

Pour démontrer les inégalités apparaissant à la dernière ligne de l'énoncé ci-dessus, choisissons $x \in]a, b[$; la croissance de la fonction $\Delta_x f$ nous donne $\dfrac{f(a) - f(x)}{a - x} \leq \dfrac{f(b) - f(x)}{b - x}$. En faisant tendre x vers a et en utilisant la continuité de f, on obtient l'inégalité de gauche ; en faisant tendre x vers b on obtient celle de droite.

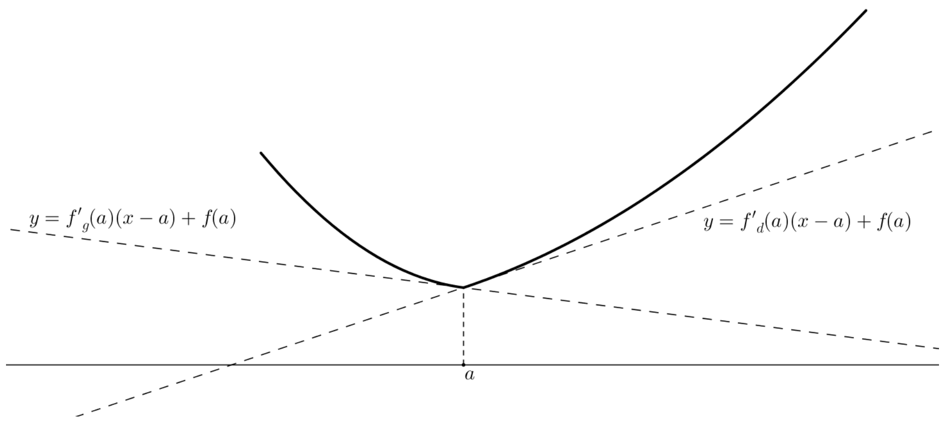
□

Notons qu'il est immédiat que $f'_g(a) \leq f'_d(a)$ pour tout $a \in \mathbb{R}$ (la limite à gauche d'une fonction croissante est toujours inférieure à sa limite à droite !) ; en particulier, si f est convexe alors f'_g et f'_d sont toutes deux des fonctions croissantes.

Théorème 8.11. *Soit I un intervalle ouvert de $\mathbb{R}$, et $f \colon I \to \mathbb{R}$ une fonction convexe. Alors pour tout $a \in I$ on a pour tout $x \in I$ que $f(x) \geq f'_d(a)(x - a) + f(a)$ 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 $x \in I$.*

Démonstration. Soit $a \in I$. D'après les inégalités à la fin de la proposition 8.10 on a 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, on sait que $f'_g(a) \leq f'_d(a)$; 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 $x \in \mathbb{R}$. □



Proposition 8.12. *Toute fonction convexe sur $\mathbb{R}$ et majorée est constante.*

Démonstration. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction convexe majorée. D'après le résultat précédent, on a $f(x) \geq f'_d(a)(x-a) + f(a)$ pour tout $a \in \mathbb{R}$; comme f est majorée ceci impose que $f'_d(a) = 0$ pour tout $a \in \mathbb{R}$ (faites un dessin et considérez les limites en $+\infty$ ou $-\infty$ selon le signe de $f'_d(a)$). De même on doit avoir $f'_g(a) = 0$ pour tout $a \in \mathbb{R}$; par conséquent $f'_d(a) = f'_g(a) = 0$ pour tout $a \in \mathbb{R}$. Autrement dit, f est dérivable et $f' = 0$: f est constante. □

Exercice 8.13. Donner un exemple de fonction convexe sur $[0, 1]$ qui n'est ni continue en 0, ni continue en 1.

8.2 Fonctions convexes dérivables

Proposition 8.14. *Soit I un intervalle ouvert et $f : I \to \mathbb{R}$ dérivable sur I. Alors les propriétés suivantes sont équivalentes :*

1. *La fonction f est convexe sur I.*
2. *Pour tout $x, y \in I$ $f(x) \geq (x-y)f'(y) + f(y)$.*
3. *La fonction f' est croissante sur I*

Notons l'interprétation géométrique du second point ci-dessus : le graphe de f est au-dessus de ses tangentes.

Démonstration. Supposons f convexe et dérivable. Alors en tout $x \in I$ on a $f'(x) = f'_d(x) = f'_g(x)$, donc la deuxième condition est vraie d'après le théorème 8.11.

Supposons maintenant la deuxième condition satisfaite, et considérons $x < y \in I$. Alors on a à la fois $f(y) \geq f(x) + (y-x)f'(x)$ et $f(x) \geq f(y) + (x-y)f'(y)$, ce qui nous donne $f'(x)(y-x) \leq f(y) - f(x) \leq f'(y)(y-x)$. On en déduit que $f'(x) \leq f'(y)$ et donc f' est croissante.

Enfin, supposons f' croissante, fixons $x < y \in I$ et considérons la fonction $g : [0, 1] \to \mathbb{R}$ définie par

$$g(t) = tf(x) + (1-t)f(y) - f(tx + (1-t)y) \ .$$

Notre but est de démontrer que g est à valeurs positives. Notons qu'on a $g(0) = g(1) = 0$, et que g est dérivable, de dérivée donnée par la formule

$$g'(t) = f(x) - f(y) + (y-x)f'(tx + (1-t)y) \ .$$

L'hypothèse selon laquelle f' est croissante entraîne donc que g' est décroissante ; et le théorème de Rolle nous garantit l'existence de $c \in]0, 1[$ tel que $g'(c) = 0$. Ainsi, g' est positive sur $[0, c]$ et négative sur $[c, 1]$, ce qui signifie que g est croissante sur $[0, c]$ et décroissante sur $[c, 1]$. Comme $g(0) = g(1) = 0$, on vient d'établir que g ne prend que des valeurs positives. □

Exercice 8.15. Soit I un intervalle ouvert de $\mathbb{R}$, f une fonction convexe sur I et dérivable. Montrer que $a \in I$ est un minimum de f sur I si et seulement si $f'(a) = 0$.

Proposition 8.16. *Soit I un intervalle ouvert et $f\colon I \to \mathbb{R}$ une fonction deux fois dérivable. Alors f est convexe si et seulement si f'' est positive sur I.*

Démonstration. Il suffit de remarquer que f'' est positive si et seulement si f' est croissante, et d'appliquer le résultat précédent. □

Remarquons ici qu'une fonction dérivée vérifie toujours la propriété des valeurs intermédiaires (c'est le contenu du théorème de Darboux) ; et une fonction croissante vérifiant la propriété des valeurs intermédiaires est nécessairement continue. Ainsi, si une fonction convexe définie sur un intervalle ouvert est dérivable, sa dérivée est nécessairement continue.

8.3 Inégalités de convexité

La convexité (ou la concavité) est souvent utilisée pour établir des inégalités. Citons un exemple important.

Proposition 8.17. *Soit $f\colon [0,+\infty[\to [0,+\infty[$ une fonction concave. Alors 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. □

Voyons maintenant l'inégalité de convexité la plus importante de notre cours.

Théorème 8.18 (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. Posons $m = \int_X g d\mu$. Notons que $m \in I$; 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 8.11 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, on en déduit que la partie négative de $\varphi \circ g$ est d'intégrale finie ; et en intégrant cette inégalité, on obtient aussi que

$$\int_X \varphi \circ g d\mu \geq \int_X \varphi(m) d\mu + \alpha \int_X (g-m) d\mu = \varphi(m) + \alpha(\int_X g d\mu - m) = \varphi(m) \ .$$

□

Le corollaire suivant est un cas (très) particulier de l'inégalité de Jensen, qui peut se montrer élémentairement, sans théorie de la mesure.

Corollaire 8.19. 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 $x_1, \ldots, x_n \in I$ et on considère l'espace mesuré d'ensemble sous-jacent $X = \{x_1, \ldots, x_n\}$, 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 $\displaystyle\int_X \varphi \circ g d\mu = \sum_{i=1}^n \alpha_i \varphi(x_i)$, et $\displaystyle\int_X g d\mu = \sum_{i=1}^n \alpha_i x_i$. 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 8.20. 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.

Exercice 8.21. Donner une démonstration du corollaire ci-dessus qui n'utilise que la définition d'une fonction convexe (et le principe de récurrence).

Le corollaire 8.19 permet d'établir plusieurs inégalités classiques, comme par exemples celles de l'exercice suivant.

Exercice 8.22.

1. Montrer que la fonction $x \mapsto -\ln(x)$ est convexe sur $]0, +\infty[$. En déduire l'*inégalité arithmético-géométrique* : pour tout $x_1, \ldots, x_n$ positifs on a

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

2. Montrer que pour $p \geq 1$ la fonction $x \mapsto x^p$ est convexe sur $[0, +\infty[$ et en déduire que pour tout $x_1, \ldots, x_n$ positifs on a

$$\frac{1}{n}\sum_{k=1}^n x_k \leq \left(\frac{1}{n}\sum_{k=1}^n x_k^p\right)^{\frac{1}{p}} \ .$$

Remarque 8.23. Dans l'exercice ci-dessus, pour montrer que $x \mapsto x^p$ est convexe sur $[0, +\infty[$, on peut montrer qu'elle est convexe sur $]0, +\infty[$ et continue sur $[0, +\infty[$, conditions qui impliquent la convexité sur $[0, +\infty[$.

Proposition 8.24. *Supposons que $0 < p < q < +\infty$ et que $(X, \mathcal{A}, \mu)$ est un espace de probabilité. Alors, pour toute fonction positive mesurable f, on a*

$$\left(\int_X f^p d\mu\right)^{\frac{1}{p}} \leq \left(\int_X f^q d\mu\right)^{\frac{1}{q}}$$

Démonstration. Posons $r = \frac{q}{p}$ et $g = f^p$. Comme $x \mapsto x^r$ est convexe sur $[0, +\infty[$, l'inégalité de Jensen nous dit que quand g est intégrable on a

$$\left(\int_X g d\mu\right)^r \leq \int_X g^r d\mu \ .$$

Autrement dit, $\left(\int_X f^p d\mu\right)^{\frac{q}{p}} \leq \int_X f^q d\mu$, et cette inégalité est équivalente à celle qu'on souhaitait établir.

Reste à traiter le cas où g n'est pas intégrable ; dans ce cas, on peut par exemple poser $f_n = \min(f, n)$, et noter que les suites (f_n^p) et (f_n^q) sont croissantes vers f^p et f^q respectivement ; on peut donc appliquer le théorème de convergence monotone pour conclure que

$$\int_X f^p d\mu = \lim_{n\to+\infty} \int_X f_n^p d\mu \quad \text{et} \int_X f^q d\mu = \lim_{n\to+\infty} \int_X f_n^q d\mu \ .$$

Comme chaque f_n^p est intégrable, la première étape de notre raisonnement nous donne $\left(\int_X f_n^p d\mu\right)^{\frac{1}{p}} \leq \left(\int_X f_n^q d\mu\right)^{\frac{1}{q}}$, et l'inégalité souhaitée s'en déduit par passage à la limite. □

Notation. Si $(X, \mathcal{A}, \mu)$ est un espace mesuré, f est une fonction mesurable sur X et $0 < p < +\infty$, on notera

$$\|f\|^p = \left(\int_X |f|^p d\mu\right)^{\frac{1}{p}} \ .$$

Ainsi, la Proposition 8.24 consiste à dire que, sur un espace de probabilité, on a pour toutes fonctions f, g mesurables positives l'inégalité $\|f\|_p \leq \|f\|_q$ dès que $0 < p < q < +\infty$.

Définition 8.25. Soit $p, q \in [1, +\infty]$. On dit que p et q sont des *exposants conjugués* si on a $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

On dit qu'alors q est l'exposant conjugué de p.

Notons que, si q est l'exposant conjugué de p, alors p est l'exposant conjugué de q ; 1 est l'exposant conjugué de $+\infty$, tandis que l'exposant conjugué de 2 est 2.

Cette notion est importante en grande partie à cause de l'inégalité suivante.

Théorème 8.26 (Inégalité de Hölder)**.** *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, $p, q \in]1, +\infty[$ deux exposants conjugués et f, g deux fonctions mesurables à valeurs réelles telles que $|f|^p$ et $|g|^q$ sont intégrables. Alors le produit fg est intégrable, et on a*

$$\int_X |fg| d\mu \leq \left(\int_X |f|^p d\mu\right)^{\frac{1}{p}} \left(\int_X |g|^q d\mu\right)^{\frac{1}{q}} \ .$$

De manière plus condensée : $\|fg\|_1 \leq \|f\|_p \|g\|_q$ dès que $\|f\|_p$ et $\|g\|_q$ sont finies.

Démonstration. Comme seules les valeurs absolues de f, g interviennent dans l'énoncé, on peut supposer f, g à valeurs positives. Notons ensuite que, si f ou g est nulle presque partout alors il en va de même du produit fg et l'inégalité désirée est vraie. On peut donc supposer que g n'est pas nulle presque partout, auquel cas $\|g\|_q > 0$. Considérons l'espace mesurable $(X, \mathcal{A})$ muni de la mesure à densité ν définie par $\nu(A) = \displaystyle\int_A \frac{g^q}{\int_X g^q d\mu} d\mu$. C'est un espace de probabilité.

Considérons la fonction $h = fg^{1-q}$; alors en utilisant la proposition 8.24 et le fait que, comme p et q sont conjugués, on a $p(1-q)+q=0$, on obtient :

$$\begin{aligned}
\int_X fg d\mu &= \left(\int_X g^q d\mu\right)\left(\int_X fg^{1-q}\frac{g^q}{\int_X g^q d\mu}d\mu\right) \\
&= \left(\int_X g^q d\mu\right)\int_X h d\nu \\
&\leq \left(\int_X g^q d\mu\right)\left(\int_X h^p d\nu\right)^{\frac{1}{p}} \\
&\leq \left(\int_X g^q d\mu\right)\left(\int_X f^p g^{p(1-q)}\frac{g^q}{\int_X g^q d\mu}d\mu\right)^{\frac{1}{p}} \\
&= \left(\int_X g^q d\mu\right)\left(\int_X \frac{f^p}{\int_X g^q d\mu}d\mu\right)^{\frac{1}{p}} \\
&= (\|g\|_q)^{q(1-\frac{1}{p})}\|f\|_p \\
&= \|g\|_q\|f\|_p \ .
\end{aligned}$$

□

L'inégalité de Hôlder est une des inégalités fondamentales concernant les espaces L^p, qui constituent la matière du prochain chapitre ; dans le cas particulier où $p=q=2$, on retrouve l'inégalité de Cauchy-Schwarz, valable dès que f^2 et g^2 sont intégrables :

$$\int_X |fg| d\mu \leq \left(\int_X f^2 d\mu\right)^{\frac{1}{2}}\left(\int_X g^2 d\mu\right)^{\frac{1}{2}} \ .$$

Chapitre 9

Introduction aux espaces L^p

Dans ce chapitre, $(X, \mathcal{A}, \mu)$ désignera un espace mesuré ; on va travailler en identifiant les fonctions si elles coïncident μ-presque partout. Autrement dit, on écrira $f = g$ quand $\mu(\{x\colon f(x) \neq g(x)\}) = 0$; 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 $f = 0$.

9.1 L'espace L^∞

Définition 9.1. Soit $f\colon X \to [0, +\infty]$ une fonction mesurable. On dit que $M \in [0, +\infty]$ est un *majorant essentiel* de f si $\mu(\{x\colon f(x) > M\}) = 0$, autrement dit, si $f \leq M$ presque partout.

On définit $\|f\|_\infty$ comme le plus petit majorant essentiel de $|f|$.

Définition 9.2. L'espace $L^\infty(X, \mathcal{A}, \mu)$, qu'on notera simplement $L^\infty(X)$ quand il n'y a pas de risque de confusion, est l'espace vectoriel formé par toutes les fonctions f telles que $\|f\|_\infty < +\infty$ [i].

Proposition 9.3. $(L^\infty(X), \|\cdot\|_\infty)$ *est un espace normé.*

Démonstration. Commençons par vérifier l'axiome de séparation : $\|f\|_\infty = 0$ est équivalent à dire que $\mu(\{x\colon |f(x)| > 0\}) = 0$, autrement dit que $f = 0$ (presque partout).

Ensuite, notons que si $\lambda \in \mathbb{R}$, alors M est un majorant essentiel de $f \in L^\infty(X)$ si et seulement si $|\lambda|M$ est un majorant essentiel de λf. Il suit que $\|\lambda f\|_\infty = |\lambda| \|f\|_\infty$.

Enfin, vérifions l'inégalité triangulaire : soit $f, g \in L^\infty(X)$. Alors on a

$$\mu(\{x\colon |f(x)| > \|f\|_\infty \text{ ou } |g(x)| > \|g\|_\infty\}) = 0$$

puisque cet ensemble est la réunion de deux ensembles de mesure nulle. Par conséquent, μ-presque partout on a $|f(x)| \leq \|f\|_\infty$ et $|g(x)| \leq \|g\|_\infty$, donc aussi $|f(x)+g(x)| \leq |f(x)|+|g(x)| \leq \|f\|_\infty + \|g\|_\infty$. On vient de montrer que $\|f\|_\infty + \|g\|_\infty$ est un majorant essentiel de $f + g$, ce qui revient à dire que $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$. □

Proposition 9.4. *L'inégalité de Hölder reste vraie pour les exposants conjugués* $1, +\infty$ *: si* f, g *sont mesurables,* $\|f\|_1 < +\infty$ *et* $\|g\|_\infty < +\infty$*, alors* fg *est intégrable et* $\|fg\|_1 \leq \|f\|_1 \|g\|_\infty$.

Démonstration. Il suffit de noter que, μ-presque partout, on a $|g(x)| \leq \|g\|_\infty$, et donc $|f(x)g(x)| \leq |f(x)| \|g\|_\infty$. En intégrant cette inégalité, on obtient bien

$$\|fg\|_1 = \int_X |f(x)g(x)| d\mu \leq \int_X |f(x)| \|g\|_\infty d\mu = \|f\|_1 \|g\|_\infty \ .$$

□

i. Répétons pour la dernière fois que deux fonctions sont identifiées si elles coïncident presque partout ; notons que si $f = g$ presque partout alors $\|f\|_\infty = \|g\|_\infty$.

9.2 Les espaces L^p

Définition 9.5. Soit $p \in [1, +\infty[$. L'espace $L^p(X, \mathcal{A}, \mu)$, noté $L^p(X)$ quand il n'y a pas de risque de confusion, est l'espace vectoriel formé par les fonctions f telles que $\|f\|_p < +\infty$.

On voudrait montrer que $\|\cdot\|_p$ est une norme sur $L^p(X)$; l'axiome de séparation n'est pas difficile à montrer : on a bien $\|0\|_p = 0$; et réciproquement, si $\|f\|_p = 0$ alors $\int_X |f(x)|^p d\mu = 0$, ce qui n'est possible (comme $|f(x)|^p \geq 0$ pour tout x) que si $|f(x)|^p = 0$ μ-presque partout, c'est-à-dire si $f = 0$ (presque partout).

L'axiome d'homogénéité se vérifie également très facilement : pour $f \in L^p(X)$ et $\lambda \in \mathbb{R}$,

$$|\lambda f|_p = \left(\int_X |\lambda f(x)|^p d\mu\right)^{\frac{1}{p}} = \left(|\lambda|^p \int_X |f(x)|^p d\mu\right)^{\frac{1}{p}} = |\lambda| \|f\|_p.$$

L'inégalité triangulaire est plus difficile à établir.

Théorème 9.6 (Inégalité de Minkowski). *Soit $p \in [1, +\infty]$ et $f, g \in L^p(X)$. Alors $f + g \in L^p(X)$ 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 $f, g \in L^p(X)$.

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

$$\left\||f + g|^{p-1}\right\|_q = \left(\int_X |f + g|^{(p-1)q} d\mu\right)^{\frac{1}{q}} = \left(\int_X |f + g|^p\right)^{1-\frac{1}{p}} = \|f + g\|_p^{p-1} .$$

Alors on a

$$\begin{aligned}
\|f + g\|_p^p &= \int_X |f + g|^p d\mu \\
&\leq \int_X (|f| + |g|)|f + g|^{p-1} d\mu \\
&= \int_X |f||f + g|^{p-1} d\mu + \int_X |g||f + g|^{p-1} d\mu \\
&\leq \|f\|_p \left\||f + g|^{p-1}\right\|_q + \|g\|_p \left\||f + g|^{p-1}\right\|_q \\
&= (\|f\|_p + \|g\|_p) \left\||f + g|^{p-1}\right\|_q \\
&= (\|f\|_p + \|g\|_p)\|f + g\|_p^{p-1}
\end{aligned}$$

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 $\|f + g\|_p \leq \|f\|_p + \|g\|_p$. □

Il y aurait beaucoup de choses à dire sur les espaces L^p ; en particulier, étudier la convergence de suites dans les espaces L^p et ses relations avec d'autres modes de convergence (en proba, convergence simple) est un sujet riche et important. Dans la fin de ce chapitre, on va supposer que $(X, \mathcal{A}, \mu)$ est un espace de probabilité (pour se simplifier la vie), justifier le choix de la notation $\|f\|_\infty$ et voir que la convergence L^p entraîne la convergence en probabilité.

Proposition 9.7. *Soit $(X, \mathcal{A}, \mu)$ un espace de probabilité et $f : X \to [0, +\infty]$ une fonction mesurable. Alors on a*

$$\|f\|_\infty = \lim_{p \to +\infty} \|f\|_p \ .$$

Démonstration. Commençons par remarquer que l'on a toujours

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}} \leq (\|f\|_\infty^p \mu(X))^{\frac{1}{p}} = \|f\|_\infty \ .$$

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 $A_t = \{x : |f(x)| > t\}$ 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 $\|f\|_\infty - \varepsilon \leq \|f\|_p \leq \|f\|_\infty$. ☐

Définition 9.8. Soit $(X, \mathcal{A}, \mu)$ un espace de probabilité. Une *variable aléatoire* est une fonction mesurable $f : X \to [0, +\infty]$.

Remarque 9.9. Ici nos notations sont différentes de celles qui sont communément utilisées en probabilités : souvent la variable aléatoire est notée X, l'espace de départ Ω, et la mesure de probabilité $\mathbb{P}$. Cela ne jouera pas vraiment de rôle ici donc on garde les notations utilisées précédemment dans le cours.

Définition 9.10. Soit $(X, \mathcal{A}, \mu)$ un espace de probabilité, et (f_n) une suite de variables aléatoires. On dit que f_n *converge en probabilité* vers une variable aléatoire f si, pour tout $\varepsilon > 0$, on a

$$\lim_{n \to +\infty} \mu(\{x : |f_n(x) - f(x)| \geq \varepsilon\}) = 0 \ .$$

Théorème 9.11. *Soit $p \in [1, +\infty]$, $(X, \mathcal{A}, \mu)$ un espace de probabilité, et (f_n) une suite de variables aléatoires telles que $f_n \in L^p(X)$ pour tout n et la suite (f_n) converge vers f dans $L^p(X, \mu)$. Alors (f_n) converge vers f en probabilité.*

Démonstration. Si $p = +\infty$ le résultat est immédiat puisque quel que soit $\varepsilon > 0$, pour n suffisamment grand on a $\|f_n - f\|_\infty < \varepsilon$ et donc $\mu(\{x : |f_n(x) - f(x)| \geq \varepsilon\}) = 0$.

Supposons donc $p < +\infty$, fixons $\varepsilon > 0$ et considérons l'ensemble $A_n = \{x : |f_n(x) - f(x)\| \geq \varepsilon\}$. Sur A_n, on a $|f_n - f|^p \geq \varepsilon^p$, et donc

$$\varepsilon^p \mu(A_n) \leq \int_X |f_n - f|^p d\mu \ .$$

Autrement dit, $\mu(A_n) \leq \dfrac{\|f_n - f\|_p^p}{\varepsilon^p}$, et le terme de droite tend vers 0 quand n tend vers $+\infty$ puisque $\|f_n - f\|_p$ tend vers 0. ☐

On pourrait aussi se demander quel est le rapport entre convergence L^p et convergence presque sûre. Le résultat suivant est une conséquence facile du théorème de convergence dominée, et la preuve est laissée en exercice.

Proposition 9.12. *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, $p \in [1, +\infty[$, et (f_n) une suite de fonctions dans $L^p(X)$ telles que $f_n(x)$ converge simplement vers une limite $f(x)$ (presque partout), et supposons qu'il existe une fonction $g \in L^p(X)$ telle que pour tout n on ait $|f_n(x)| \leq g(x)$ presque partout. Alors $f \in L^p(X)$ et $\|f_n - f\|_p \to 0$ quand p tend vers $+\infty$.*

Réciproquement, on a le résultat suivant, qu'on se contente de mentionner ici sans démonstration.

Théorème 9.13. *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, $p \in [1, +\infty[$, et (f_n) une suite d'éléments de $L^p(X)$ qui converge vers f dans $(L^p(X), \|\cdot\|_p)$. Alors il existe une suite extraite (f_{n_k}) telle que $(f_{n_k}(x))$ tend vers $f(x)$ presque partout.*

Exercice 9.14. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et (f_n) une suite d'éléments de $L^\infty(X)$ qui converge vers f dans $(L^\infty(X), \|\cdot\|_\infty)$. Montrer que $(f_n(x)$ converge vers $f(x)$ presque partout. Montrer que la réciproque est fausse.

Chapitre 10

Espaces de Hilbert ; bases hilbertiennes

Dans tout ce chapitre, E désignera un $\mathbb{R}$-espace vectoriel.

10.1 Vocabulaire et premiers résultats sur les produits scalaires

Définition 10.1. Un *produit scalaire* sur E est une application $\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R}$ qui est :
- — *symétrique*, c'est-à-dire que pour tout $x, y \in E$ on a $\langle x, y \rangle = \langle y, x \rangle$;
- — *bilinéaire*, c'est-à-dire que pour tout $x, y, z \in E$ et tout $\lambda \in \mathbb{R}$ on a $\langle \lambda x + y, z \rangle = \lambda \langle x, z \rangle + \langle y, z \rangle$ (et donc aussi $\langle x, \lambda y + z \rangle = \lambda \langle x, y \rangle + \langle y, z \rangle$ par symétrie) ;
- — *définie positive*, c'est-à-dire que pour tout $x \in E$ on a $\langle x, x \rangle \geq 0$ et $\langle x, x \rangle = 0 \Leftrightarrow x = 0_E$.

Exercice 10.2. Montrer que, si $(X, \mathcal{A}, \mu)$ est un espace mesuré, alors $\langle \cdot, \cdot \rangle$ défini par

$$\langle f, g \rangle = \int_X f g d\mu$$

est un produit scalaire sur $L^2(X)$.

Dans la suite, la notation $\langle \cdot, \cdot \rangle$ désignera toujours un produit scalaire sur E.

Notons qu'ici on travaillera avec des espaces vectoriels réels ; on peut aussi définir un produit scalaire pour des espaces vectoriels complexes mais dans ce cas-là on doit demander qu'il soit *sesquilinéaire* plutôt que bilinéaire.

Pour nous habituer aux calculs avec les produits scalaires, notons deux conséquences faciles. La première est qu'on connaît les valeurs de $\langle x, y \rangle$ pour tout x, y dès qu'on connaît la valeur de $\langle x, x \rangle$ pour tout $x \in E$.

Proposition 10.3 (Identité de polarisation)**.** *Pour tout $x, y \in E$ on a*

$$4\langle x, y \rangle = \langle x + y, x + y \rangle - \langle x - y, x - y \rangle \ .$$

Démonstration. Par bilinéarité et symétrie, on a

$$\begin{aligned} \langle x + y, x + y \rangle &= \langle x, x + y \rangle + \langle y, x + y \rangle \\ &= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\ &= \langle x, x \rangle + \langle y, y \rangle + 2\langle x, y \rangle \ . \end{aligned}$$

Un calcul similaire (ou le résultat précédent appliqué à $y' = -y$) donne

$$\langle x - y, x - y \rangle = \langle x, x \rangle + \langle y, y \rangle - 2\langle x, y \rangle$$

d'où le résultat. □

Proposition 10.4 (Identité du parallélogramme)**.** *Pour tout $x, y \in E$ on a*

$$\langle x + y, x + y \rangle + \langle x - y, x - y \rangle = 2\langle x, x \rangle + 2\langle y, y \rangle \ .$$

La preuve est laissée en exercice.

Théorème 10.5 (Inégalité de Cauchy-Schwarz)**.** *Pour tout $x, y \in E$ on a*

$$\langle x, y\rangle^2 \leq \langle x, x\rangle\langle y, y\rangle \ .$$

Avant de donner la preuve de l'inégalité de Cauchy-Schwarz, notons tout de suite une conséquence fondamentale.

Corollaire 10.6. L'application $x \mapsto \sqrt{\langle x, x\rangle}$ est une norme sur E.

On notera dans la suite

$$\|x\| = \sqrt{\langle x, x\rangle} \ .$$

Preuve du corollaire. L'axiome de séparation se déduit immédiatement du fait que le produit scalaire est défini positif. Celui d'homogénéité suit de sa bilinéarité. Il reste à vérifier l'inégalité triangulaire : soit $x, y \in E$. On a

$$\begin{aligned}
\langle x+y, x+y\rangle &= \langle x, x\rangle + \langle y, y\rangle + 2\langle x, y\rangle \\
&\leq \langle x, x\rangle + \langle y, y\rangle + 2\sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle} && \text{(Cauchy-Schwarz)} \\
&= \left(\sqrt{\langle x, x\rangle} + \sqrt{\langle y, y\rangle}\right)^2
\end{aligned}$$

Comme toutes les quantités en jeu sont positives, on peut passer à la racine carrée de chaque côté de l'inégalité et obtenir comme espéré

$$\sqrt{\langle x+y, x+y\rangle} \leq \sqrt{\langle x, x\rangle} + \sqrt{\langle y, y\rangle} \ .$$

□

Preuve de l'inégalité de Cauchy-Schwarz. Fixons $x, y \in E$, et considérons l'application définie sur $\mathbb{R}$ par

$$t \mapsto \langle x+ty, x+ty\rangle = \langle x, x\rangle + 2t\langle x, y\rangle + t^2\langle y, y\rangle.$$

Il s'agit d'une application polynomiale, de degré 2, à valeurs positives. Le polynôme en question ne peut avoir qu'une racine réelle au plus (sans quoi il changerait de signe) : son discriminant doit donc être négatif ou nul. Autrement dit, on doit avoir

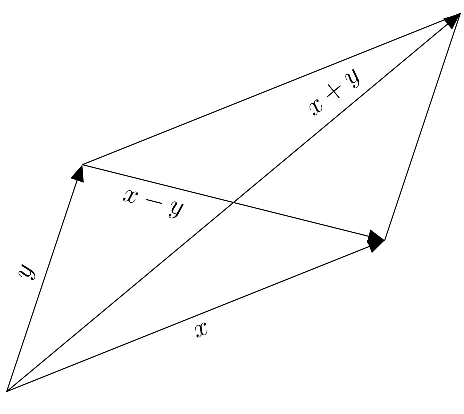
$$4\langle x, y\rangle^2 - 4\langle x, x\rangle\langle y, y\rangle \leq 0 \ .$$

C'est exactement l'inégalité qu'on cherchait à démontrer. □

Jusqu'à nouvel ordre, E désigne un espace vectoriel réel muni du produit scalaire $\langle\cdot,\cdot\rangle$ et de la norme $\|\cdot\|$ associée au produit scalaire.

En utilisant la norme, les identités de polarisation et du parallélogramme s'écrivent ainsi :

$$\begin{aligned}
&\forall x, y \in E \quad 4\langle x, y\rangle = \|x+y\|^2 - \|x-y\|^2 && \text{(polarisation)} \\
&\forall x, y \in E \quad 2\|x\|^2 + 2\|y\|^2 = \|x+y\|^2 + \|x-y\|^2 && \text{(parallélogramme)}
\end{aligned}$$



Exercice 10.7. Montrer que le produit scalaire définit une application continue de $E \times E$ vers $\mathbb{R}$.

Définition 10.8. Soit A un sous-ensemble de E. On définit

$$A^{\perp} = \{x \in E \colon \forall a \in A \ \langle a, x\rangle = 0\} \ .$$

On appelle $A^{\perp}$ l'*orthogonal* de A.

Exercice 10.9. Montrer que, pour $A \subseteq E$:

1. $A^{\perp}$ est un sous-espace vectoriel de E.
2. $A^{\perp}$ est fermé dans E.
3. $A^{\perp} = (\overline{A})^{\perp}$.

Définition 10.10. Une famille $(a_i)_{i\in I}$ de vecteurs non nuls de E est dite *orthogonale* si pour tout $i \neq j$ on a $\langle a_i, a_j\rangle = 0$.

La famille est dite *orthonormale* si elle est orthogonale et de plus $\|a_i\| = 1$ pour tout $i \in I$.

Proposition 10.11. *Toute famille orthogonale est une famille libre.*

Démonstration. Soit $(a_i)_{i\in I}$ une famille orthogonale et $a_{i_1}, \dots, a_{i_n}$, $\lambda_1, \dots, \lambda_n$ tels que $\sum_{k=1}^{n} \lambda_k a_{i_k} = 0$.

Observons que pour tout $j \in \{1, \dots, n\}$ on a

$$\begin{aligned}
\left\langle \sum_{k=1}^{n} \lambda_k a_{i_k}, a_{i_j} \right\rangle &= \sum_{k=1}^{n} \lambda_k \langle a_{ik}, a_{i_j}\rangle \\
&= \sum_{k\neq j} \lambda_k \langle a_{ik}, a_{i_j}\rangle + \lambda_j \|a_{i_j}\|^2 \\
&= \lambda_j \|a_{i_j}\|^2
\end{aligned}$$

On en déduit que $\lambda_j \|a_{i_j}\|^2 = 0$ pour tout j puis, puisque tous les a_{i_j} sont non nuls, que $\lambda_j = 0$ pour tout j : la famille est bien libre. □

Exercice 10.12. Montrer que, si $(a_i)_{i\in I}$ est une famille orthogonale de vecteurs de E on a pour tout $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ et tout $i_1, \dots, i_n \in I$

$$\| \sum_{k=1}^{n} \lambda_k a_{i_k} \|^2 = \sum_{k=1}^{n} \lambda_k^2 \|a_{i_k}\|^2 \ .$$

Exercice 10.13. On se place dans l'espace $L^2([0,1], \mathcal{B}([0,1]), \lambda)$ muni du produit scalaire $\langle f, g\rangle = \int_0^1 f(x)g(x)dx$. On considère la famille $(u_k)_{k\geq 1}$ définie par $u_k(x) = \sin(k\pi x)$ pour tout $x \in [0,1]$ et tout $k \geq 1$. Montrer que c'est une famille orthogonale.

10.2 Espaces métriques complets

Définition 10.14. Soit (X, d) un espace métrique, et (x_n) une suite d'éléments de X. On dit que (x_n) est une *suite de Cauchy* si la propriété suivante est vérifiée :

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

Autrement dit : à partir d'un certain rang, les termes de la suite deviennent arbitrairement proches les uns des autres.

Exercice 10.15. Soit (X, d) un espace métrique. Montrer que toute suite convergente est de Cauchy, et que la réciproque est fausse en général (on pourra par exemple considérer la suite définie par $x_n = 2^{-n}$ dans l'espace $X =]0, +\infty[$ muni de sa distance usuelle).

Exercice 10.16. Soit (X, d) un espace métrique, et (x_n) une suite de Cauchy d'éléments de X. Montrer que (x_n) est bornée.

Définition 10.17. Soit (X, d) un espace métrique. On dit que (X, d) est *complet* si toute suite de Cauchy d'éléments de X est convergente.

Proposition 10.18. *Soit (X, d) un espace métrique complet, et $F \subseteq X$ un sous-ensemble fermé. Alors (F, d) est complet.*

Démonstration. Soit (x_n) une suite de Cauchy d'éléments de F. C'est en particulier une suite de Cauchy d'éléments de X, elle converge donc vers un certain $x \in X$ puisque (X, d) est complet. Comme F est fermé dans X, on doit avoir $x \in F$. On vient de montrer que toute suite de Cauchy d'éléments de F converge dans F : (F, d) est complet. □

Cette proposition admet une forme de réciproque.

Proposition 10.19. *Soit (X, d) un espace métrique et $F \subseteq X$ un sous-ensemble tel que (F, d) est complet. Alors F est fermé dans X.*

Démonstration. Soit (x_n) une suite d'éléments de F qui converge vers $x \in X$. Alors (x_n) est de Cauchy puisque toute suite convergente est de Cauchy ; comme F est complet (x_n) converge dans F, ce qui montre que $x \in F$. □

Proposition 10.20. *Soit (X, d) un espace métrique, et (x_n) une suite de Cauchy d'éléments de X telle que (x_n) admette une sous-suite (x_{n_k}) convergente. Alors (x_n) converge.*

Démonstration. Soit (x_n) une suite de Cauchy d'éléments de X, et $x \in X$ tel qu'il existe une sous-suite (x_{n_k}) qui converge vers x. Fixons $\varepsilon > 0$. Alors on sait qu'il existe N tel que $d(x_n, x_m) \leq \varepsilon$ pour tout $n, m \geq N$. De plus, il existe aussi K tel que $d(x_{n_k}, x) \leq \varepsilon$ pour tout $k \geq K$. Comme n_k tend vers $+\infty$, on peut trouver k_0 tel qu'on ait à la fois $k_0 \geq K$ et $n_{k_0} \geq N$. Par conséquent, pour tout $n \geq N$ on a à la fois $d(x_n, x_{n_{k_0}}) \leq \varepsilon$ et $d(x_{n_{k_0}}, x) \leq \varepsilon$, donc aussi $d(x_n, x) \leq 2\varepsilon$. Ceci prouve que (x_n) converge vers x. □

Proposition 10.21. *Tout espace métrique compact est complet.*

Démonstration. Soit (X, d) un espace métrique compact, et (x_n) une suite de Cauchy d'éléments de X. Par définition de la compacité, on peut extraire une sous-suite de (x_n) qui converge vers $x \in X$. La proposition précédente nous permet donc de conclure que (x_n) converge vers x. □

Ceci nous fournit nos premiers exemples d'espaces complets. Mais on en connaît beaucoup d'autres !

Théorème 10.22. *Soit $n \in \mathbb{N}^*$ et $\|\cdot\|$ une norme sur $\mathbb{R}^n$. Alors $(\mathbb{R}^n, \|\cdot\|)$ est complet.*

Démonstration. Soit (x_n) une suite de Cauchy d'élements de $(\mathbb{R}^n, \|\cdot\|)$. Alors (x_n) doit être bornée, et le théorème de Bolzano–Weierstrass nous permet donc d'en extraire une sous-suite convergente. La proposition 10.20 nous permet de conclure que (x_n) converge. □

Exercice 10.23. Soit $(E, \|\cdot\|)$ un $\mathbb{R}$-espace vectoriel, et $F \subseteq E$ un sous-espace de dimension finie. Montrer que F est fermé dans E.

Théorème 10.24 (Théorème de Riesz–Fisher)**.** *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et $p \in [1, +\infty]$. Alors l'espace $L^p(X)$ est un espace complet.*

On admettra ce résultat dont la preuve est difficile !

Définition 10.25. On dit qu'un espace vectoriel E muni d'un produit scalaire $\langle \cdot, \cdot \rangle$ est un *espace de Hilbert* si E, muni de la norme induite par son produit scalaire, est un espace métrique complet.

Un exemple fondamental d'espace de Hilbert : l'espace $L^2(X)$ pour tout espace mesuré $(X, \mathcal{A}, \mu)$.

10.3 Projection sur un convexe fermé

Maintenant on suppose que E est un espace de Hilbert.

Définition 10.26. Soit A une partie non vide de E et $x \in E$. On dit que $a_0 \in A$ est une *projection* de x sur A si on a

$$\|x - a_0\| = \inf\{\|x - a\| \colon a \in A\} \ .$$

Remarque 10.27. Il n'existe pas toujours une projection ; et même quand elle existe, elle n'est pas nécessairement unique. Les deux exercices suivants donnent des exemples de ce phénomène.

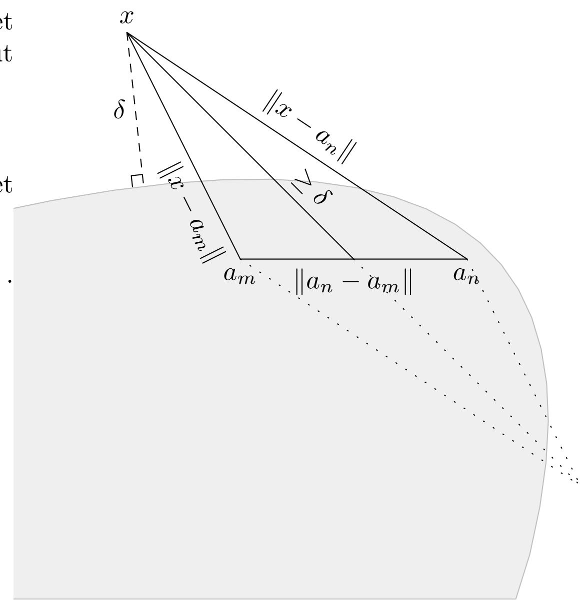
Exercice 10.28. Soit E un $\mathbb{R}$-espace vectoriel muni d'un produit scalaire possédant une famille orthonormale $(e_n)_{n>0}$. On considère $F = \left\{(1+\dfrac{1}{n})e_n \colon n > 0\right\}$.

1. Montrer que 0_E n'a pas de projection sur F.
2. Montrer que pour tout $n \neq m$, $\|(1+\dfrac{1}{n})e_n - (1+\dfrac{1}{m})e_m\| \geq 1$. En déduire que F est fermé.
3. Expliciter un tel exemple pour l'espace d'Hilbert $L^2([0,1])$.

Exercice 10.29. On considère un triangle équilatéral dans $\mathbb{R}^2$. Montrer que son centre de gravité a trois projections.

Théorème 10.30. *Soit E un espace de Hilbert, $A \subseteq E$ un sous-ensemble convexe et fermé. Alors tout point $x \in E$ admet une unique projection $p_A(x)$ sur A.*

Démonstration. Fixons $x \in E$, et notons $\delta = \inf_{a \in A} \|x - a\|$. On peut trouver une suite (a_n) d'éléments de A tels que $\|x - a_n\|$ converge vers δ. On va prouver que (a_n) est une suite de Cauchy ; alors on pourra conclure (comme E est complet) que (a_n) converge vers a, qui appartient à A puisque A est fermé. Par continuité de la norme on aura alors $\|x - a\| = \lim_{n \to +\infty} \|x - a_n\| = \delta$. Ceci montrera l'existence d'une projection (mais pas encore son unicité).



Pour justifier que (a_n) est de Cauchy, on va utiliser la convexité de A et utiliser le fait que $\frac{a+b}{2} \in A$ pour tout $a, b \in A$. En particulier, pour tout $a, b \in A$ on doit avoir

$$\|x - \frac{a+b}{2}\| \geq \delta \ .$$

Fixons $n, m \in \mathbb{N}$ et appliquons l'identité du parallélogramme à $x - a_n$ et $x - a_m$:

$$\|(x-a_n)+(x-a_m)\|^2 + \|(x-a_n)-(x-a_m)\|^2 = 2\|x-a_n\|^2 + 2\|x-a_m\|^2 \ .$$

On en déduit

$$\begin{aligned}
\|a_n - a_m\|^2 &= 2\|x-a_n\|^2 + 2\|x-a_m\|^2 - \|2x - a_n - a_m)\|^2 \\
&= 2\|x-a_n\|^2 + 2\|x-a_m\|^2 - 4\|x - \frac{a_n+a_m}{2}\|^2 \\
&\leq 2\|x-a_n\|^2 + 2\|x-a_m\|^2 - 4\delta^2
\end{aligned}$$

Fixons $\varepsilon > 0$. On sait qu'il existe N tel que pour tout $n, m \geq N$ on a $\|x - a_n\|^2 \leq \delta^2 + \varepsilon$ et $\|x - a_m\|^2 \leq \delta^2 + \varepsilon$, et donc aussi $\|a_n - a_m\|^2 \leq 4\varepsilon$. On vient de montrer comme promis que (a_n) est de Cauchy.

Pour montrer que le projeté est unique, considérons $a, b \in A$ tels que $\|x - a\| = \|x - b\| = \delta$. Le même calcul que ci-dessus nous donne

$$\|a - b\|^2 \leq 2\|x - a\|^2 + 2\|x - b\|^2 - 4\delta^2 \leq 0 \ .$$

Ceci n'est possible que si $a = b$. $\square$

Théorème 10.31. *Soit E un espace de Hilbert, A un convexe fermé de E et $x \in E$. Le projeté $p_A(x)$ de x sur A est l'unique $a \in A$ tel que*

$$\forall b \in A \quad \langle x-a, b-a \rangle \leq 0 \ .$$

On en déduit que l'application $x \mapsto p_A(x)$ est 1-lipschitzienne.

Démonstration. Notons $a = p_A(x)$. Pour tout $b \in A$, on a $\|x-a\|^2 \leq \|x-b\|^2$ par définition du projeté. En utilisant la relation

$$\begin{aligned} \|x-b\|^2 &= \|(x-a)+(a-b)\|^2 \\ &= \|x-a\|^2 + \|a-b\|^2 + 2\langle x-a, a-b \rangle \end{aligned}$$

On en déduit que pour tout $b \in A$

$$\begin{aligned} 2\langle x-a, b-a \rangle &= \|x-a\|^2 - \|x-b\|^2 + \|a-b\|^2 \\ &\leq \|a-b\|^2 \ . \end{aligned}$$

Fixons maintenant $b \in A$, différent de a ; pour tout $t \in [0,1]$ on peut appliquer l'inégalité précédente à $b_t = tb + (1-t)a$, qui appartient à A, et obtenir en utilisant le fait que $b_t - a = t(b-a)$

$$2\langle x-a, t(b-a) \rangle \leq t^2 \|b-a\| \ .$$

En divisant des deux côtés par t et en faisant tendre t vers 0, on voit que ceci n'est possible que si $\langle x-a, b-a \rangle \leq 0$, ce qu'on voulait démontrer.

Supposons maintenant qu'on a $a' \in A$ tel que $\langle x-a', b-a' \rangle \leq 0$ pour tout $b \in A$. Avec un calcul similaire au précédent, on a

$$\|x-a'\|^2 = \|x-a\|^2 + 2\langle x-a', a-a' \rangle - \|a-a'\|^2 \ .$$

Puisque $\langle x-a', a-a' \rangle \leq 0$ par hypothèse sur a', on a donc que $\|x-a'\|^2 \leq \|x-a\|^2$, ce qui n'est possible que si $a = a'$ par définition d'un projeté.

Enfin, il nous reste à démontrer que $x \mapsto p_A(x)$ est 1-lipschitzienne. Soit $x, x' \in X$. D'après la propriété qu'on vient d'établir, on a à la fois

$$\langle x - p_A(x), p_A(x') - p_A(x) \rangle \leq 0 \quad \text{ et } \quad \langle x' - p_A(x'), p_A(x) - p_A(x') \rangle \leq 0 \ .$$

En notant $r = x - x' + p_A(x') - p_A(x)$ on obtient en sommant les deux inégalités précédentes que

$$\langle p_A(x) - p_A(x'), x' - x + p_A(x) - p_A(x') \rangle \leq 0 \ ,$$

autrement dit $\langle p_A(x) - p_A(x'), r \rangle \geq 0$. Finalement,

$$\begin{aligned} \|x-x'\|^2 &= \|p_A(x) - p_A(x') + r\|^2 \\ &= \|p_A(x) - p_A(x')\|^2 + \|r\|^2 + 2\langle p_A(x) - p_A(x'), r \rangle \\ &\geq \|p_A(x) - p_A(x')\|^2 \ . \end{aligned}$$

On a bien montré que p_A est 1-lipschitzienne. □

Remarque 10.32. Reprenons les notations du théorème précédent, et supposons de plus que $x \notin A$. L'ensemble $H = \{z \in E \colon \langle z-a, x-a \rangle = 0\}$ est un sous-espace affine de E (égal à $a + (\mathrm{Vect}(x-a))^\perp$), qui admet un supplémentaire de dimension 1 (la droite de vecteur directeur $x-a$) ; on dit que H est un *hyperplan*. L'hyperplan H "coupe" E en deux demi-espaces : l'ensemble H^-des z tels que $\langle z-a, x-a \rangle \leq 0$ et l'ensemble H^+ des z tels que $\langle z-a, x-a \rangle > 0$. Le théorème ci-dessus nous dit que A est entièrement contenu dans H^-, tandis que x appartient à H^+. On dit alors que H *sépare* x et A. L'existence d'un hyperplan séparant un point d'un convexe fermé est une conséquence importante du théorème de projection. Par exemple, elle nous permet de voir que tout convexe fermé dans un espace de Hilbert est une intersection de demi-espaces fermés.

Un cas particulier très important est celui de la projection sur un sous-espace vectoriel fermé de E (ce qui est le cas, rappelons-le, de tout sous-espace vectoriel de dimension finie).

Théorème 10.33. *Soit E un espace de Hilbert et $F \subseteq E$ un sous-espace vectoriel fermé. Alors, pour tout $x \in E$, le projeté $p_F(x)$ est l'unique élément de F tel que $x - p_F(x)$ soit orthogonal à F. On le nomme* projeté orthogonal *de x sur F.*

La projection orthogonale sur F est une application linéaire continue.

Démonstration. Soit $x \in E$. Par définition, $p_F(x)$ appartient à F, et on a vu que $p_F(x)$ est l'unique élément de F tel que $\langle x - p_F(x), f - p_F(x)\rangle \leq 0$ pour tout $f \in F$. Mais comme $f + p_F(x) \in F$, on en déduit que $\langle x - p_F(x), f\rangle = 0$ pour tout $f \in F$. Réciproquement, si $a \in F$ est tel que $\langle x - a, f\rangle = 0$ pour tout $f \in F$, alors on a en particulier $\langle x - a, a\rangle = 0$ et donc $\langle x - a, f - a\rangle = 0$ pour tout $f \in F$, et d'après le théorème précédent cette propriété impose que $a = p_F(x)$.

Reste à montrer que p_F est linéaire (on sait déjà qu'elle est continue puisqu'elle est 1-lipschitzienne). Considérons $x, y \in E$ et $\lambda \in \mathbb{R}$. Alors pour tout $f \in F$ on a

$$\begin{aligned}\langle x + \lambda y - p_F(x) - \lambda p_F(y), f\rangle &= \langle x - p_F(x), f\rangle + \langle \lambda y - \lambda p_F(y), f\rangle \\ &= \langle x - p_F(x), f\rangle + \lambda\langle y - p_F(y), f\rangle \\ &= 0\end{aligned}$$

Par conséquent, $x+\lambda y-(p_F(x)+\lambda p_F(y))$ est orthogonal à tous les éléments de F, ce qui montre que $p_F(x+\lambda y) = p_F(x) + \lambda p_F(y)$. $\square$

Corollaire 10.34. Soit E un espace de Hilbert, et F un sous-espace vectoriel fermé. Alors on a $E = F \oplus F^\perp$. De plus, $F^\perp$ est l'unique supplémentaire de F orthogonal à F : si $E = F \oplus G$ et G est orthogonal à F alors $G = F^\perp$.

Démonstration. On sait déjà que $F^\perp$ est un sous-espace vectoriel (fermé) de E ; de plus, si $x \in F \cap F^\perp$ alors $\|x\|^2 = \langle x, x\rangle = 0$. Enfin, pour tout $x \in E$ on a $x = (x - p_F(x)) + p_F(x)$, ce qui prouve que $x \in F^\perp + F$.

Soit G un supplémentaire de F orthogonal à F, c'est-à-dire $E = F \oplus G$ et $G \subset F^\perp$. Prenons $x \in F^\perp$. Alors, $x = f + g$ avec $f \in F$ et $g \in G$, d'où $f = x - g \in F \cap F^\perp = \{0_E\}$ et donc $x \in G$. $\square$

Exercice 10.35. Montrer qu'un sous-espace vectoriel F de E est dense dans E si, et seulement si, $F^\perp = \{0_E\}$.

Exercice 10.36. Soit $(a_i)_{i\in I}$ une famille de vecteurs de E. Montrer que le sous-espace vectoriel engendré par les $(a_i)_{i\in I}$ est dense dans E si, et seulement si,

$$\forall x \in E \quad (\forall i \in I\ \langle x, a_i\rangle = 0) \Leftrightarrow x = 0_E\ .$$

Exercice 10.37. Soit F un sous-espace vectoriel de E. Montrer que $(F^\perp)^\perp = \overline{F}$.

Définition 10.38. Soit $(a_i)_{i\in I}$ une famille de vecteurs de E. On dit que $(a_i)_{i\in I}$ est une *base hilbertienne* de vecteurs de E si, et seulement si, la famille $(a_i)_{i\in I}$ est orthonormale et engendre un sous-espace vectoriel dense.

On rappelle que la famille engendre un sous-espace vectoriel dense si, et seulement si le seul vecteur qui est orthogonal à chacun des a_i est le vecteur nul. Attention, si E est de dimension infinie, une base hilbertienne de E n'est pas nécessairement une base algébrique ! Par contre, si E est de dimension finie, c'est-à-dire si E est un espace euclidien, alors toute base hilbertienne est une base de E, c'est-à-dire engendre E. Autrement dit, en dimension finie, aucun sous-espace vectoriel propre n'est dense dans E (ceci suit du fait qu'en dimension finie, tout sous-espace vectoriel est fermé).

10.4 Procédé de Gram-Schmidt, inégalité de Bessel, identités de Parseval

Notons 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 orthonormale de F (preuve laissée en exercice) :

Proposition 10.39. *Soit E un espace de Hilbert et F un sous-espace vectoriel de dimension finie avec $(e_1, \ldots, e_n)$ une base orthonormale de F. Alors pour tout $x \in E$, on a*

$$p_F(x) = \sum_{i=1}^{n} \langle x, e_i\rangle e_i.$$

Rappelons que le procédé de Gram-Schmidt permet de calculer une base orthonormale d'un espace euclidien à partir d'une base donnée :

Proposition 10.40 (Procédé de Gram-Schmidt)**.** *Soit E un espace euclidien et $(e_1, \ldots, e_n)$ une base de E. Pour chaque $0 < i < n$, notons F_i le sous-espace vectoriel $\mathrm{Vec}(e_1, \ldots, e_i)$ engendré par $e_1, \ldots, e_i$. Alors, la famille $(e'_1, \ldots, e'_n)$ définie de la manière suivante est une base orthonormale de E :*

$$\begin{aligned} e'_1 &= \frac{e_1}{\|e_1\|} \\ e'_i &= \frac{e_i - p_{F_{i-1}}(e_i)}{\|e_i - p_{F_{i-1}}(e_i)\|} = \frac{e_i - \sum_{k=1}^{i-1} \langle e_i, e'_k \rangle e'_k}{\|e_i - \sum_{k=1}^{i-1} \langle e_i, e'_k \rangle e'_k\|} \quad \textit{pour } 1 < i \leq n \end{aligned}$$

Exercice 10.41. Vérifier que les vecteurs $e_1 = (1, 1, 1)$, $e_2 = (1, 1, -1)$ et $e_3 = (0, 1, 2)$ forment une base de $\mathbb{R}^3$. Utiliser le procédé de Gram-Schmidt sur cette base pour obtenir une base orthonormale.

Maintenant, étant donnée une base orthonormale d'un espace euclidien, on a les identités suivantes (à vérifier en exercice) :

Proposition 10.42. *Soit E un espace euclidien et $(e_1, \ldots, e_n)$ une base orthonormale de E. Alors pour tous x, y de E, on a :*

$$\begin{aligned} x &= \sum_{i=1}^{n} \langle x, e_i \rangle e_i \\ \|x\|^2 &= \sum_{i=1}^{n} |\langle x, e_i \rangle|^2 \\ \langle x, y \rangle &= \sum_{i=1}^{n} \langle x, e_i \rangle \langle y, e_i \rangle \end{aligned}$$

Dans cette partie, nous allons voir que ces identités se généralisent aux espaces de Hilbert ayant une base hilbertienne dénombrable.

Définition 10.43. Un espace métrique est dit *séparable* s'il contient une partie dénombrable dense.

Exercice 10.44. Un théorème d'approximation de Weierstrass montre que toute fonction réelle définie et continue sur $[-1, 1]$ est limite uniforme de fonctions polynomiales à coefficients rationnels. Notons que l'ensemble des fonctions polynomiales à coefficients rationnels est dénombrable. En déduire que l'espace de Hilbert $(C([-1, 1], \mathbb{R}), \| \cdot \|_2)$ est séparable.

Proposition 10.45. *Un espace de Hilbert séparable de dimension infinie possède une base hilbertienne dénombrable $(e_i)_{i \in \mathbb{N}}$.*

Démonstration. La preuve est laissée en exercice. Indication : on considère une énumération $(a_n)_{n \in \mathbb{N}}$ d'une partie dénombrable A dense, puis on en extrait une base du sous-espace vectoriel engendré par A, et enfin, en utilisant le procédé de Gram-Schmidt, on construit une base orthonormale de ce sous-espace vectoriel. □

Exercice 10.46 (Polynômes de Legendre)**.** On se place dans l'espace de Hilbert $(C([-1, 1], \mathbb{R}), \| \cdot \|_2)$.

1. En utilisant l'exercice précédent et le procédé de Gram-Schmidt, montrer qu'il existe une base hilbertienne $(p_n)_{n \in \mathbb{N}}$, telle que pour chaque $n \in \mathbb{N}$, la fonction p_n est une fonction polynomiale de degré n.
2. Pour $n \in \mathbb{N}$, on considère la fonction polynomiale l_n définie par $l_n(t) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(t^2 - 1)^n$ pour $t \in [-1, 1]$ (polynômes de Legendre).
   (a) Montrer que la famille $(l_n)_{n \in \mathbb{N}}$ est une famille orthogonale.
   (b) Déterminer le degré de l_n pour chaque n.

(c) En déduire que pour chaque n on a $p_n = \dfrac{l_n}{\|l_n\|}$.

Théorème 10.47 (Inégalité de Bessel)**.** *Soit E un espace de Hilbert et $(e_i)_{i\in\mathbb{N}}$ une famille orthonormale. Alors pour tout $x \in E$,*

$$\sum_{i=0}^{+\infty} |\langle x, e_i\rangle|^2 \leq \|x\|^2.$$

Démonstration. Soit $n \geq 0$. La projection de x sur l'espace vectoriel engendré par $e_0, \dots, e_n$ est égale à $\sum_{i=0}^n \langle x, e_i\rangle e_i$ et est orthogonale à $x - \sum_{i=0}^n \langle x, e_i\rangle e_i$.

Cette orthogonalité entraîne que :

$$\|x\|^2 = \left\|\sum_{i=0}^n \langle x, e_i\rangle e_i\right\|^2 + \left\|x - \sum_{i=0}^n \langle x, e_i\rangle e_i\right\|^2 = \sum_{i=0}^n |\langle x, e_i\rangle|^2 + \left\|x - \sum_{i=0}^n \langle x, e_i\rangle e_i\right\|^2 \geq \sum_{i=0}^n |\langle x, e_i\rangle|^2.$$

On conclut en passant à la limite. □

Théorème 10.48 (Identités de Parseval)**.** *Soit E un espace de Hilbert possédant une base hilbertienne dénombrable $(e_i)_{i\in\mathbb{N}}$. Alors, pour tous x, y de E, on a :*

$$\begin{aligned} x &= \sum_{i=0}^{+\infty} \langle x, e_i\rangle e_i \\ \|x\|^2 &= \sum_{i=0}^{+\infty} |\langle x, e_i\rangle|^2 \\ \langle x, y\rangle &= \sum_{i=0}^{+\infty} \langle x, e_i\rangle\langle y, e_i\rangle \end{aligned}$$

Démonstration. Notons F le sous-espace vectoriel engendré par la famille $(e_i)_{i\in\mathbb{N}}$ et pour chaque n, notons F_n le sous-espace vectoriel engendré par $e_0, \dots, e_n$. Comme F est dense dans E, il existe une suite (x_n) d'éléments de F convergeant vers x. Quitte à ajouter au début de la suite suffisamment de termes nuls et à répéter des termes, on peut supposer que pour chaque n, le terme x_n appartient à F_n. Comme $\sum_{i=0}^n \langle x, e_i\rangle e_i$ est la projection de x sur F_n, on obtient ainsi l'inégalité

$$\|x - \sum_{i=0}^n \langle x, e_i\rangle e_i\| \leq \|x - x_n\|,$$

ce qui entraîne que $\displaystyle\sum_{i=0}^n \langle x, e_i\rangle e_i$ converge vers x.

En passant à la limite dans l'égalité

$$\|x\|^2 = \left\|\sum_{i=0}^n \langle x, e_i\rangle e_i\right\|^2 + \left\|x - \sum_{i=0}^n \langle x, e_i\rangle e_i\right\|^2 = \sum_{i=0}^n |\langle x, e_i\rangle|^2 + \left\|x - \sum_{i=0}^n \langle x, e_i\rangle e_i\right\|^2$$

on obtient la seconde identité du théorème.

Notons que par bilinéarité du produit scalaire on a,

$$\left\langle \sum_{i=0}^n \langle x, e_i\rangle e_i, \sum_{i=0}^n \langle y, e_i\rangle e_i \right\rangle = \sum_{i=0}^n \langle x, e_i\rangle\langle y, e_i\rangle.$$

Par continuité du produit scalaire, le terme de gauche converge vers $\langle x, y\rangle$ ce qui donne la dernière identité. □

Remarque 10.49. Le théorème ci-dessus est valide quel que soit l'ordre de l'énumération de la base de Hilbert et donc indépendamment de l'ordre avec lequel l'on somme les séries.

Les identités de Parseval permettent de montrer réciproquement que si un espace de Hilbert a une base hilbertienne dénombrable alors il est séparable :

Exercice 10.50. Soit E un espace de Hilbert possédant une base hilbertienne dénombrable $(e_i)_{i\in\mathbb{N}}$. On considère la partie A des combinaisons linéaires finies de $(e_i)_{i\in\mathbb{N}}$ à coefficients rationnels. Notons que A est dénombrable. Montrer que A est dense dans E

Terminons cette partie en montrant l'unicité du développement en série dans le théorème 10.48. On étudiera un exemple important dans la partie suivante, les séries de Fourier.

Proposition 10.51 (unicité du développement)**.** *Soit $(e_i)_{i\in\mathbb{N}}$ une base hilbertienne d'un espace de Hilbert E séparable. Soit $x \in E$ et une famille $(\lambda_i)_{i\in\mathbb{N}}$ de scalaires dans $\mathbb{R}$ tels que la série $\sum_{i=0}^{+\infty} \lambda_i e_i$ converge vers x. Alors pour tout $i \in \mathbb{N}$, on a $\lambda_i = \langle x, e_i\rangle$. On appellera ces scalaires, les coefficients de x dans la base $(e_i)_{i\in\mathbb{N}}$.*

Démonstration. Pour chaque n, posons $x_n = \sum_{i=0}^{n} \lambda_i e_i$. Par hypothèse, on a (x_n) qui converge vers x. Notons que pour $i \leq n$, on a $\langle x_n, e_i\rangle = \lambda_i$ et par continuité du produit scalaire, en faisant tendre n vers l'infini, on obtient $\langle x, e_i\rangle = \lambda_i$.

□

10.5 Une application : les séries de Fourier

On munit l'espace $L^2([-\pi,\pi])$ du produit scalaire défini par

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{[-\pi,\pi]} fg d\lambda \ .$$

Pour nous simplifier un peu les notations ci-dessous, notons $s_n(t) = \sin(nt)$ pour $n \geq 1$, et $c_n(t) = \cos(nt)$ pour $n \geq 0$ (en particulier c_0 est la fonction constante égale à 1). On a déjà vu que les fonctions (s_n) forment une famille orthogonale ; de plus le changement de variable $u = \frac{\pi}{2} - x$ (ou un dessin des graphes...) permet de voir que pour $n \geq 1$ on a $\|s_n\| = \|c_n\|$ (la norme étant celle associée à notre produit scalaire : la constante multiplicative est là pour que $\|1\| = 1$...), et comme $s_n^2 + c_n^2 = 1$, on déduit que pour $n \geq 1$ on a $\displaystyle\int_{-\pi}^{\pi} s_n^2 d\mu = \pi$, et donc $\|s_n\| = \frac{\sqrt{2}}{2}$, et de même pour c_n.

Le même type de calculs que ceux effectués pour la famille (s_n) permet de voir que la famille (c_m) est orthogonale, et aussi que pour tout n, m on a $\langle s_n, c_m\rangle = 0$. A titre d'exemple, montrons cette dernière égalité : pour $n \geq 1$ et $m \geq 0$ on a

$$\sin(nt)\cos(mt) = \frac{\sin((n+m)t) + \sin((n-m)t)}{2} \ .$$

De plus, $\displaystyle\int_{-\pi}^{\pi} \sin(kt) = 0$ pour tout k, ce qui prouve que $\displaystyle\int_{-\pi}^{\pi} \sin(nt)\cos(mt)dt = 0$ pour tout n, m.

Il ressort des calculs précédents que la famille de fonctions $((s_n)_{n\geq 1}, (c_m)_{m\geq 0}$ est une famille orthogonale ; cette famille n'est pas orthonormale mais le serait si pour $n \geq 1$ on remplaçait (s_n) par $\sqrt{2}s_n$ et c_n par $\sqrt{2}c_n$. L'espace vectoriel engendré par cette famille est appelé l'ensemble des *polynômes trigonométriques*. Les formules de linéarisation nous permettent de vérifier qu'un produit de polynômes trigonométriques est encore un polynôme trigonométrique.

Théorème 10.52. *L'espace des polynômes trigonométriques est dense dans $L^2([-\pi,\pi])$.*

Démonstration. On doit montrer que, si $f \in L^2([-\pi,\pi]$ est telle que $\langle f, c_n\rangle = 0$ pour tout $n \geq 0$ et $\langle f, s_n\rangle = 0$ pour tout $n \geq 1$ alors f est la fonction nulle.

Commençons par supposer que f est une fonction continue et non nulle ; on souhaite montrer que f n'est pas orthogonale à tous les polynômes trigonométriques. Pour simplifier les notations, on va se contenter de traiter le cas où il existe $h > 0$ tel que $f(x) > 0$ pour tout $x \in]-h, h[$ (la preuve ci-dessous s'adapte facilement au cas général). Considérons alors le polynôme trigonométrique

$$P_n \colon x \mapsto (1 + \cos(x) - \cos(h))^n \ .$$

Pour $x \in]-h, h[$ on a que $P_n(x)$ croît (à x fixé) vers $+\infty$ quand $n \to +\infty$, tandis que pour $x \in [-\pi, -h[\cup]h, \pi[$ on a $P_n(x) \to 0$ et $|P_n(x)| \leq 1$. Mais alors :

— Sur $]-h,h[$, on peut appliquer le théorème de convergence monotone pour conclure que $\displaystyle\int_{]-h,h[} fP_n d\lambda$ tend vers $+\infty$.

— Sur $]-\pi,-h[\cup]h,\pi[$, on peut utiliser le fait que $|f(x)P_n(x)| \leq |f(x)| \leq \|f\|_\infty$ pour conclure, à l'aide du théorème de convergence dominée, que $\displaystyle\int_{]-\pi,-h[\cup]h,\pi[} fP_n d\lambda$ tend vers 0.

Tout ceci nous dit que $\displaystyle\int_{]-\pi,\pi[} fP_n d\lambda$ tend vers $+\infty$; en particulier on ne peut avoir $\langle f, P_n\rangle = 0$ pour tout n, par conséquent f n'est pas orthogonale à l'espace des polynômes trigonométriques.

Si maintenant on ne suppose plus f continue, mais seulement L^2, et que f est orthogonale aux polynômes trigonométriques, remarquons d'abord que $x \mapsto F(x) = \int_{-\pi}^x f d\lambda$ est bien définie (d'après Hölder, f est intégrable puisque la fonction constante 1 appartient à $L^2([\pi,\pi])$ et continue : pour tout $x,y \in [-\pi,\pi]$ on a

$$|F(x)-F(y)| = \left|\int_y^x f d\lambda\right| \leq \sqrt{\int_y^x f^2 d\lambda}\sqrt{\int_y^x 1 d\lambda} \leq \|f\|_2\sqrt{y-x} \ .$$

Soit maintenant P un polynôme trigonométrique, et Q une primitive de P (qui est de nouveau un polynôme trigonométrique)

$$\begin{aligned}
\langle F, P\rangle &= \frac{1}{2\pi}\int_{-\pi}^{\pi} F(x)P(x)dt \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\int_{-\pi}^{x} f(t)dt\right)P(x)dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\int_{t}^{\pi} P(x)dx\right)f(t)dt \text{ (par Fubini)} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}(Q(\pi)-Q(t))f(t)dt \\
&= \langle Q(\pi)-Q, f\rangle \\
&= 0 \ .
\end{aligned}$$

Ci-dessus, l'application de Fubini est justifiée par le fait qu'on intègre une fonction continue sur un compact (l'ensemble $\{(x,t) \in [-\pi,\pi]^2 \colon t \leq x\}$) ; et la dernière égalité vient du fait que f est supposée orthogonale à tous les polynômes trigonométriques. Comme F est continue, la première partie de notre raisonnement nous permet de conclure que $F = 0$. On vient donc de conclure que, pour toute fonction $f \in L^2([-\pi,\pi])$ qui est orthogonale aux polynômes trigonométriques, on a pour tout $x,y \in [-\pi,\pi]$ que $\displaystyle\int_{[x,y]} f d\lambda = 0$. Ceci entraîne que $f = 0$ presque partout [i], ce qu'on souhaitait démontrer.

□

Une autre façon d'énoncer le théorème précédent est de dire que la famille $(c_0, (\sqrt{2}s_n)_{n\geq 1}, (\sqrt{2}c_n)_{n\geq 1})$ est une base hilbertienne de $L^2([-\pi,\pi])$. Pour $f \in L^2([-\pi,\pi])$, introduisons ses *coefficients de Fourier* :

$$a_0(f) = \langle f, c_0\rangle = \frac{1}{2\pi}\int_{[-\pi,\pi]} f d\lambda$$

$$\forall n \geq 1 \ a_n(f) = 2\langle f, c_n\rangle = \frac{1}{\pi}\int_{[-\pi,\pi]} f(t)\cos(nt)d\lambda(t)$$

$$\forall n \geq 1 \ b_n(f) = 2\langle f, s_n\rangle = \frac{1}{\pi}\int_{[-\pi,\pi]} f(t)\sin(nt)d\lambda(t)$$

i. Ici, on triche un peu : la preuve de ce fait n'est pas immédiate. On peut démontrer que deux mesures μ,ν sur $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ telles que pour tout a,b on ait $\mu([a,b]) = \nu([a,b]) < +\infty$ doivent être égales, ce qui permet de conclure ici, et aussi de montrer l'unicité de la mesure de Lebesgue sur $\mathbb{R}$.

Puisque la famille $(c_0, (\sqrt{2}s_n)_{n\geq 1}, (\sqrt{2}c_n)_{n\geq 1})$ est une base hilbertienne, on sait que pour tout $f \in [-\pi, \pi]$ on peut écrire :

$$f = \langle f, c_0\rangle + \sum_{n\geq 1} \langle f, \sqrt{2}c_n\rangle\sqrt{2}c_n + \langle f, \sqrt{2}s_n\rangle\sqrt{2}s_n$$

$$\|f\|^2 = \langle f, c_0\rangle^2 + \sum_{n\geq 1} (\langle f, \sqrt{2}c_n\rangle)^2 + (\langle f, \sqrt{2}s_n\rangle)^2 \ .$$

En termes de coefficients de Fourier, on a donc, pour toute fonction $f \in L^2([-\pi, \pi])$ et presque tout $t \in [-\pi, \pi]$:

1. $f(t) = a_0(f) + \sum_{n\geq 1} a_n(f)\cos(nt) + b_n(f)\sin(nt)$;
2. $\frac{1}{2\pi}\int_{-\pi}^{\pi} f^2 d\lambda = a_0(f)^2 + \frac{1}{2}\sum_{n\geq 1} (a_n(f))^2 + (b_n(f))^2.$

La série apparaissant dans la première égalité ci-dessus est appelée *série de Fourier* de f. Remarquons que, même si f est continue, on n'a justifié la première égalité ci-dessus que presque partout (c'est une égalité au sens des fonctions L^2) : il peut a priori y avoir des points où on n'a pas égalité entre la fonction et sa série de Fourier ; effectivement, on a besoin d'hypothèses supplémentaires (par exemple, si f est C^1 par morceaux) pour conclure que la série de Fourier de f converge en tout point vers f.

La deuxième égalité ci-dessus est souvent appelée égalité de Parseval (et c'est bien une conséquence de ce que nous avons appelé égalité de Parseval plus tôt dans ce cours).

Exercice 10.53. Calculer les coefficients de Fourier de la fonction f définie sur $[-\pi, \pi]$ par $f(x) = x$. A l'aide de l'égalité de Parseval, en déduire l'égalité :

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \ .$$

10.6 Une autre application : le théorème du minimax

On va conclure ces notes en discutant un théorème qu'on peut montrer à l'aide du théorème de projection sur un convexe fermé (ou, dans la preuve qu'on va présenter, son corollaire sur la séparation d'un fermé convexe et d'un point) : le théorème du minimax de von Neumann.

Il s'agit d'un théorème de *théorie des jeux*, qu'on va ici essayer de présenter dans un contexte assez simple. Imaginons deux joueurs X et Y jouant à un jeu J ; X a un nombre fini de choix possibles $A = \{a_1, \ldots, a_n\}$ et Y un nombre fini de choix possibles $B = \{b_1, \ldots, b_m\}$. On a également une fonction de gain $g\colon A \times B \to \mathbb{R}$; si X joue a_i et Y joue b_j, alors X gagne $g(a_i, b_j)$ et Y perd $g(a_i, b_j)$ (on dit que le jeu est à somme nulle : ce que X gagne est perdu par Y). Attention : g peut prendre des valeurs négatives (et on considère que gagner -10 euros revient à perdre 10 euros...).

Par exemple, X et Y pourraient jouer à pierre-papier-ciseaux, ou à un jeu un peu plus subtil appelé *pair ou impair* : X et Y jouent chacun un entier entre 1 et 2 ; si la somme des deux est impaire X gagne un montant égal à la somme des deux chiffres joués, et si la somme est paire c'est Y qui gagne. Comment doivent jouer X et Y s'ils sont rationnels ?

Comme ni X ni Y n'a d'information sur ce que l'autre va faire, on va considérer qu'une stratégie pour un des deux joueurs consiste à jouer au hasard, suivant une loi de probabilité p qu'il s'est fixé à l'avance : X va jouer a_i avec la probabilité p_i, alors que Y va jouer b_j avec la probabilité q_j. Lorsque X suit p, et Y suit q, l'espérance de gain pour X (et donc, l'espérance de perte pour Y) est

$$g(p, q) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_i g(a_i, b_j) q_j \ .$$

Si X se fixe une stratégie p, le pire qui peut lui arriver est de gagner $\inf_q g(p, q)$ (si q choisit une stratégie aussi bien adaptée que possible pour contrer p) ; notons que cet inf est en fait un min, puisque $q \mapsto g(p, q)$ est une fonction continue sur le compact

$$\{(q_1, \ldots, q_m)\colon \forall i \ q_j \geq 0 \text{ et } \sum q_j = 1\} \ .$$

Ceci signifie que, au minimum, X peut trouver une stratégie lui garantissant un gain [ii] supérieur ou égal à

$$\underline{V}(J) = \max_p \min_q g(p,q) \ .$$

En suivant le même raisonnement en échangeant les rôles de X et Y, on voit que Y peut se garantir une perte inférieure ou égale à

$$\overline{V}(J) = \min_q \max_p g(p,q) \ .$$

La définition de $\underline{V}(J)$ et $\overline{V}(J)$ devrait rendre claire l'inégalité $\underline{V}(J) \le \overline{V}(J)$.

Théorème 10.54 (Théorème du minimax de von Neumann)**.** *Pour tout jeu J à somme nulle, on a $\underline{V}(J) = \overline{V}(J)$. Cette valeur commune est appelée la valeur du jeu J et notée $V(J)$.*

Avant de donner une preuve de ce résultat, observons une conséquence : on peut trouver des stratégies p_0, q_0 telles que $g(p_0,q_0) = V(J)$. Le fait que $g(p_0,q_0) = \underline{V}(J)$ nous dit que, pour toute stratégie q de Y, on a $g(p_0,q) \ge g(p_0,q_0) = \min g(p_0,q)$: autrement dit, si X joue p_0, toute autre stratégie que q_0 est moins bonne (au sens large) pour Y que q_0. Comme jouer q_0 garantit à Y sa perte minimale, Y doit jouer cette stratégie (ou une autre stratégie telle que $g(p_0,q) = V(J)$). De même, X doit jouer p_0 (ou une autre stratégie lui garantissant le même gain) s'il est rationnel. On voit ainsi apparaître un équilibre : si les deux joueurs sont rationnels, ils vont jouer des stratégies p,q telles que $g(p,q) = V(J)$.

Preuve du théorème du minimax. Une petite observation préliminaire :

$$\overline{V}(J) = \min_q \max_i g(a_i,q) \ .$$

En effet, X peut choisir la stratégie de toujours jouer a_i ; ceci montre que pour tout q, on a $\max_p g(p,q) \ge \max_i g(a_i,q)$; Réciproquement, si on choisit i_0 tel que $g(a_i,q) \le g(a_{i_0},q)$ pour tout i, alors pour toute stratégie p on a

$$g(p,q) = \sum_{i=1}^n p_i g(a_i,q) \le \sum_{i=1}^n p_i g(a_{i_0},q) = g(a_{i_0},q) \ .$$

Ceci nous donne l'inégalité $\max_p g(p,q) \le \max_i g(a_i,q)$, et donc l'égalité entre ces deux quantités, qui nous sera utile dans la suite de la preuve.

On va raisonner par l'absurde et supposer $\underline{V}(J) < \overline{V}(J)$; on peut alors trouver c entre ces deux valeurs et, quitte à remplacer la fonction g par $g-c$, se ramener au cas où $\underline{V}(J) < 0 < \overline{V}(J)$, dans lequel on se place dans la suite.

Introduisons

$$E = \{(e_1,\dots,e_n) \in \mathbb{R}^n \colon \exists q \ \forall i \ e_i \ge g(a_i,q)\} \ .$$

Notons d'abord que E est convexe : si $(e_1,\dots,e_n)$ et $(f_1,\dots,f_n)$ appartiennent à E, et $t \in [0,1]$, on a des stratégies q_e, q_f pour Y telles que $e_i \ge g(a_i,q_e)$ et $f_i \ge g(a_i,q_f)$ pour tout i. Mais alors $q' = tq_e + (1-t)q_f$ est une stratégie pour Y, et on a bien pour tout i

$$te_i + (1-t)f_i \ge tg(a_i,q_e) + (1-t)g(a_i,q_f) = g(a_i, tq_e + (1-t)q_f) = g(a_i,q') \ .$$

Ensuite, montrons que E est fermé : soit (x_k) une suite d'éléments de E qui converge vers $x \in \mathbb{R}^n$; on veut prouver que $x \in E$. Pour tout k on peut trouver une stratégie q_k pour Y telle que $g(a_i,q_k) \le x_k(i)$. Comme on l'a déjà dit précédemment, l'ensemble des stratégies pour Y est $\{(q_1,\dots,q_m) \colon \forall j \ q_j \ge 0 \text{ et } \sum q_j = 1\}$, qui est fermé et borné dans $\mathbb{R}^m$, et donc compact. Par conséquent, quitte à extraire on peut supposer que (q_k) converge vers une stratégie q. Mais alors, par continuité de g, on a pour tout i que $g(a_i,q_k)$ tend vers $g(a_i,q)$, et par définition de la convergence dans $\mathbb{R}^k$ la suite $x_k(i)$ converge vers x_i. Comme pour tout k on a $g(a_i,q_k) \le x_k(i)$ on conclut par passage à la limite que $g(a_i,q) \le x_i$, ce qui montre comme espéré que $x \in E$.

Ensuite, notons que $0 \notin E$: en effet, par l'observation faite au début de la preuve, pour toute stratégie q il existe i tel que $g(a_i,q) \ge \overline{V}(J) > 0$.

D'après la remarque 10.32, on peut trouver $u = (u_1,\dots,u_n) \in \mathbb{R}^n$ tel que $\langle u,e \rangle > 0$ pour tout $e \in E$. En particulier, pour toute stratégie q et tout uplet $(r_1,\dots,r_n)$ de réels positifs, on a $\sum_{i=1}^n u_i(g(a_i,q)+r_i) > 0$. Ceci n'est possible que si $u_i \ge 0$ pour tout i (sans quoi, faire tendre r_i vers $+\infty$ sans toucher aux autres coordonnées

ii. notons ici encore que ce "gain" peut-être négatif !

et regarder la limite...). De plus, on ne peut avoir $u_i = 0$ pour tout i ; par conséquent, on peut considérer la stratégie p où X joue a_i avec probabilité

$$p_i = \frac{u_i}{\sum_{i=1}^n u_i} \ .$$

Alors, le fait que $\langle u, e\rangle > 0$ pour tout $e \in E$ nous permet de conclure que pour toute stratégie q on a

$$\sum_{i=1}^n p_i g(a_i, q) = \frac{1}{\sum_{i=1}^n u_i} \sum_{i=1}^n u_i g(a_i, q) > 0 \ .$$

Autrement dit, la stratégie p que nous venons de produire est telle que $g(p, q) > 0$ pour toute stratégie q, ce qui démontre que $\underline{V}(J) > 0$, contredisant notre hypothèse de départ. $\square$

Index