Annexe A

Fondamentaux et compléments

Sommaire

A.1 Fondamentaux

A.1.1 Éléments de théorie des ensembles

Notations A.1.1. (◦) Soit X un ensemble, on note $\mathcal{P}(X)$ l'ensemble des parties de X, c'est à dire l'ensemble des sous-ensembles constitués d'éléments de X.

Soit A une partie de X. On note A^c le complémentaire de A dans X, c'est-à-dire l'ensemble des éléments de X qui ne sont pas dans A.

Soient A et B deux parties d'un ensemble X. On dit que A est inclus dans B, et l'on écrit $A \subset B$, si tout élément de A est aussi élément de B :

$$x \in A \Longrightarrow x \in B.$$

On a[1] $\emptyset \subset B$ pour toute partie B.

Soient A et B deux parties d'un ensemble X. On note $A \cap B$ l'*intersection* de A et B, c'est à dire l'ensemble des éléments qui appartiennent à la fois à A et à B :

$$A \cap B = \{x \in X\,,\ x \in A \text{ et } \ x \in B\}\ .$$

On note $A \cup B$ l'*union* de A et B, c'est à dire l'ensemble des éléments qui sont dans A ou dans B :

$$A \cup B = \{x \in X\,,\ x \in A \text{ ou } \ x \in B\}\ .$$

1. Cette assertion est à la fois évidente et troublante, du fait que l'ensemble vide est inclus dans toute partie de n'importe quel ensemble. Ce fait rend possible d'énoncer des propriétés comme : tout élément de l'ensemble vide est un *porte-clé*, ce qui signifie précisément que l'assertion : " $\forall x \in A$, x est un porte-clé"
est vraie pour $A = \emptyset$. De façon générale, toute propriété portant sur les éléments d'un ensemble est systématiquement vérifiée par l'ensemble vide.

Si $(A_i)_{i\in I}$ est une famille de parties disjointes de X, non vides, dont l'union est égale à X, on dit qu'elle réalise une *partition* de X.

On note $A \setminus B$ la *différence ensembliste* de A et B, c'est à dire l'ensemble des éléments qui sont dans A, mais pas dans B :

$$A \setminus B = \{x \in X\,,\ x \in A\,,\ X \notin B\} = A \cap B^c.$$

Soient X et Y deux ensembles. On appelle *produit cartésien* de X et Y, et l'on note $X \times Y$, l'ensemble des couples (x, y) avec $x \in X$ et $y \in Y$.

Soient X et Y deux ensembles, on note Y^X l'ensemble des applications de X vers Y. On peut représenter chaque application par une partie de $X \times Y$ qui, pour tout x, contient un unique couple du type (x, y) (l'élément y est l'image de x par l'application considérée). L'ensemble des couples $(x, f(x))$, qui est une partie de $X \times Y$, est appelé *graphe* de l'application f.

On peut identifier une partie A d'un ensemble X à sa *fonction indicatrice*[2] $\mathbb{1}_A$, qui à chaque élément x de X associe la valeur 1 ou 0, selon que x soit dans A ou pas. Chaque partie pouvant ainsi être représentée (ou "codée", pour utiliser un terme informatique) à une application de X dans $\{0, 1\}$, on note parfois 2^X l'ensemble des parties de X.

Définition A.1.2. (Autour de la notion d'application ($\circ$))
Soient X et Y deux ensembles, et f une application de X dans Y. Pour tout $y \in Y$, on appelle image réciproque de y, et l'on note[3] $f^{-1}(\{y\})$ (ou $f^{-1}(y)$) l'ensemble des antécédents de y :

$$f^{-1}(y) = \{x\,,\ y = f(x)\}\,.$$

On définit de la même manière l'image réciproque d'un ensemble $B \subset Y$ par

$$f^{-1}(B) = \{x\,,\ f(x) \in B\}\,.$$

On dit que f est *injective* si deux éléments de X ne peuvent avoir la même image, c'est-à-dire si l'image réciproque de tout $y \in Y$ contient au plus un élément.

On dit que f est *surjective* si tout élément de l'espace d'arrivée Y a au moins un antécédent, c'est-à-dire si l'image réciproque de tout $y \in Y$ contient au moins un élément.

On dit que f est *bijective* si elle est injective et surjective, c'est-à-dire si l'image réciproque de tout élément de l'ensemble d'arrivée contient exactement un élément.

Remarque A.1.3. Toutes les opérations ensemblistes peuvent être traduite en terme de fonctions indicatrices. Par exemple si, pour $C \subset X$, on définit $\mathbb{1}_C$ comme la fonction qui vaut 1 sur C, 0 à l'extérieur de C, on a

$$\mathbb{1}_{A\cup B} = \max(\mathbb{1}_A, \mathbb{1}_B)\,,\ \mathbb{1}_{A\cap B} = \min(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{1}_A \times \mathbb{1}_B\,,\ \mathbb{1}_{A^c} = 1 - \mathbb{1}_A.$$

A.1.2 Structures fondamentales : relations et structures algébriques

Relations

Définition A.1.4. (Relation, relation d'équivalence, classes d'équivalence ($\circ$))
Soit X un ensemble. Une relation est la donnée d'une partie R de $X \times X$, dénotée par le symbole $\mathcal{R}$ selon la convention

$$(x, x') \in R \Longleftrightarrow x\,\mathcal{R}\,x'.$$

On parle de *relation d'équivalence* si elle vérifie les les propriétés suivantes :

2. Le terme de fonction *caractéristique* est parfois utilisé, mais nous l'évitons ici car il prend un autre sens dans le contexte des probabilités. On prendra néammoins garde au fait que le terme de fonction indicatrice prend lui aussi un autre sens en *analyse convexe*, et donc en optimisation, désignant une fonction associée à un ensemble qui prend la valeur 0 dans l'ensemble, et $+\infty$ à l'extérieur.

3. On prendra garde à la confusion possible avec l'application inverse d'une bijection, notée également f^{-1}. Pour distinguer l'application considérée ici de cet inverse défini (quand c'est possible) de Y dans X, on utilise en général la notation ensembliste $f^{-1}(\{y\})$, qui rappelle que l'on considère ici une application qui à un ensemble (une partie de Y) associe un ensemble (une partie de X), qui peut être vide, ou non réduite à un singleton.

(i) *(réflexivité)* Pour tout $x \in X$, $x \mathcal{R} x$.
(ii) *(symétrie)* Pour tous x, $y \in X$, $x \mathcal{R} y \Longleftrightarrow y \mathcal{R} x$.
(iii) *(transitivité)* Pour tous x, y, $z \in X$,

$$x \mathcal{R} y \text{ et } y \mathcal{R} z \Longrightarrow x \mathcal{R} z.$$

Pour tout $x \in X$, on appelle classe d'équivalence l'ensemble

$$\overline{x} = \{y \in X\ ,\ y \mathcal{R} x\} \in \mathcal{P}(X).$$

L'ensemble $\overline{X}$ constitué de ces classes est appelé espace quotient, ce que l'on note $\overline{X} = X/\mathcal{R}$.

L'application qui à $x \in X$ associe sa classe $\overline{x}$ est par construction une surjection, appelée *surjection canonique.*

Remarque A.1.5. La notion de classes d'équivalence semble se limiter à formaliser différemment la notion de *partition* d'un ensemble. De fait, à toute partition d'un ensemble, i.e. $X = \bigcup X_i$ (union disjointe), on peut associer canoniquement la relation d'équivalence $x \mathcal{R} y$ si x et y appartiennent au même X_i. Cette notion est beaucoup plus féconde que cette version ensembliste dès que X est muni d'une structure, et que la relation d'équivalence respecte cette structure (dans un sens qui dépend de la structure en question). L'espace $\overline{X}$ des classes d'équivalence hérite alors de la structure de l'espace initial, c'est un espace de même type (groupe, espace vectoriel, espace métrique, ...), qui est "plus petit" puisqu'il existe une surjection de X vers $\overline{X}$ (la surjection canonique). L'exemple le plus simple est $\mathbb{Z}$ quotienté par la relation : $x \mathcal{R} y$ si et seulement si $x - y$ est pair. On a deux classes d'équivalences, notés $\overline{0}$ et $\overline{1}$. On peut définir sur l'espace quotient une addition $\overline{x} + \overline{y} = \overline{x+y}$ (on peut vérifier que ça ne dépend pas du représentant choisi : la somme de deux entiers de même parité est paire, impaire si les parités sont différentes), de telle sorte que l'espace quotient, noté $\mathbb{Z}/2\mathbb{Z}$, est aussi un groupe additif. Dans un tout autre contexte, celui des fonctions mesurables (voir le chapitre B dédié à ces questions), il sera extrêmement fécond d'introduire la relation d'équivalence $f \mathcal{R} g$ si l'ensemble des points en lesquels f et g diffèrent est négligeable. L'espace quotient contient des classes de fonctions, et il hérite des structures de l'espace initial (en particulier la structure d'espace vectoriel).

Définition A.1.6. (Relation d'ordre, majorant ($\circ$))
Soit X un ensemble. Une relation d'ordre sur X est la donnée d'une partie $\mathcal{O}$ de $X \times X$, dénotée par le symbole $\leq$ selon la convention

$$(x, x') \in \mathcal{O} \Longleftrightarrow x \leq x',$$

qui vérifie les propriétés suivantes :
(i) *(réflexivité)* pour tout $x \in X$, $x \leq x$,
(ii) *(antisymétrie)* si $x \leq y$ et $y \leq x$ alors $x = y$,
(iii) *(transitivité)* pour tous x, y, $z \in X$,

$$x \leq y \text{ et } y \leq z \Longrightarrow x \leq z.$$

On écrit $x < y$ si $x \leq y$ et $x \neq y$. On dit que l'ordre est *total* si, pour tout $x \neq y$, on a $x < y$ ou $y < x$. On dit qu'il est *partiel* dans le cas contraire. Lorsque l'ordre est partiel, deux éléments peuvent ne pas être comparables.

On dit que M est un *majorant* de $A \subset X$ si $x \leq M$ pour tout $x \in A$. Si $A \subset M$ admet un plus petit majorant, on l'appelle *borne supérieure* de A. On définit de la même manière un *minorant* d'un ensemble, et une *borne inférieure.*

Exercice A.1.1. a) Montrer que la relation d'inclusion sur l'ensemble des parties d'un ensemble X est une relation d'ordre. À partir de quel cardinal de X l'ordre n'est-il que partiel ?

b) Proposer une relation d'ordre sur l'ensemble des partitions d'un ensemble.

c) Décrire les éléments maximaux et minimaux des relations d'ordre évoquées ci-dessus.

Remarque A.1.7. Les relations d'équivalence et d'ordre peuvent être encodées par des *graphes*. Pour la relation d'équivalence, on peut considérer l'ensemble $E \subset X \times X$ des points en relation comme décrivant les arêtes d'un graphe. Cet ensemble est symétrique, de telle sorte que l'on peut identifier $(x, y) \in E$ et $(y, x) \in E$, E contient toutes les boucles (x, x) (réflexivité), et les composantes connexes du graphes sont des *cliques* (i.e. le sous-graphe correspondant est complet : il contient toutes les arêtes possibles entre les sommets).

Pour une relation d'ordre, l'ensemble $E \subset X \times X$ d'arêtes n'est pas symétrique (ou plutôt il ne l'est que dans le cas d'un graphe éclaté qui ne contient que des boucles, qui représente une relation d'ordre partiel d'un type extrême, où deux éléments distincts ne sont jamais comparables), on parle de graphe *orienté*. Il contient aussi toutes les boucles (réflexivité), vérifie la propriété de transitivité $(x, y) \in E$ et $(y, z) \in E$ implique $(x, z) \in E$, et ne contient *aucun cycle*, i.e. il n'est pas possible, partant d'un point, de se déplacer en suivant les flèches pour se retrouver au point de départ. On dit que le graphe est *acyclique*.

A.1.3 Cardinalité

Définition A.1.8. (Équipotence)
On dit que deux ensembles X et Y sont *équipotents* s'il existe une bijection de X dans Y. On écrira alors $X \simeq Y$ ou [4] $\mathrm{Card}(X) = \mathrm{Card}(Y)$.

Notation A.1.9. S'il existe une injection de X dans Y, on écrit $X \lesssim Y$, ou $\mathrm{Card}(X) \leq \mathrm{Card}(Y)$. Si de plus il n'existe pas de bijection entre les deux ensembles, on écrira $X < Y$ ou $\mathrm{Card}(X) < \mathrm{Card}(Y)$.

Théorème A.1.10. On note $\mathcal{P}(X)$ l'ensemble des parties de X. On a

$$\mathrm{Card}(X) < \mathrm{Card}(\mathcal{P}(X)).$$

Démonstration. Supposons qu'il existe une surjection φ de X dans $\mathcal{P}(X)$. On introduit

$$A = \{x \in X\ ,\ x \notin \varphi(x)\}\ .$$

Comme φ est surjective, il doit exister x tel que $\varphi(x) = A$. Si $x \in A$, alors $x \in \varphi(x)$ d'où $x \notin A$. Si $x \notin A = \varphi(x)$, alors $x \in A$. On a donc contradiction dans les deux cas, ce qui exclut l'existence d'une telle application φ. □

Remarque A.1.11. Malgré le caractère rudimentaire de sa démonstration, la proposition précédente est très générale et profonde. Dans le cas d'un ensemble fini, elle exprime simplement l'inégalité $n < n!$. Dans le cas d'ensembles infinis, elle permet de construire différents niveaux d'infini arbitrairement "grands" : partant d'un ensemble X infini, son ensemble de parties est d'une cardinalité strictement plus grand puisqu'il n'existe pas de bijection entre les deux. On peut itérer en considérant l'ensemble des parties de l'ensemble des parties, etc ..., pour construire formellement une suite d'ensembles infinis de "cardinaux" strictement croissants.

Définition A.1.12. (Ensemble dénombrable)
On dit que l'ensemble X est dénombrable si X est fini [5], ou si $X \simeq \mathbb{N}$, i.e. s'il est en bijection avec $\mathbb{N}$. Un ensemble infini dénombrable est donc énumérable : si l'on note φ la bijection de $\mathbb{N}$ vers X, et $x_n = \varphi(n)$, l'ensemble X est exactement la collection des x_n, et l'on note $(x_n)_{n\in\mathbb{N}}$, ou simplement (x_n), la suite associée.

Proposition A.1.13. Une union dénombrable d'ensembles infinis dénombrables est dénombrable.

4. On prendra garde à cette notation $\mathrm{Card}(X) = \mathrm{Card}(Y)$ qui exprime simplement l'existence d'une bijection entre deux ensembles. Dans le cas d'ensembles finis, cela correspond bien à l'identité des cardinaux, mais pour des ensembles infinis, il faut lire comme un tout cette identité, qui implique des "quantités" ($\mathrm{Card}(X)$ et $\mathrm{Card}(Y)$) qui n'ont pas été définies.

5. Certains auteurs considèrent que l'attribut dénombrable est restreint aux ensemble infinis. Nous faisons ici le choix de considérer qu'un ensemble fini est dénombrable, ce qui permet de simplifier l'énoncé d'un grand nombre de propriétés. Ce choix impose de préciser *infini dénombrable* pour un ensemble qui est en bijection avec $\mathbb{N}$.

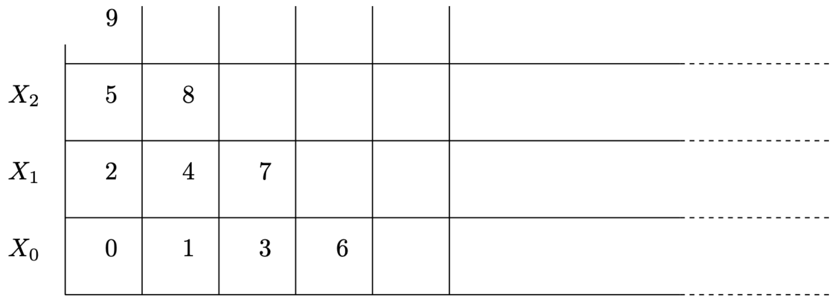

FIGURE A.1.1 – Énumération d'une réunion dénombrable d'ensembles dénombrables

Démonstration. Nous établissons la propriété dans le cas où l'union et les ensembles sont infinis. Soit $(X_n)_{n\in\mathbb{N}}$ une famille d'ensembles infinis dénombrables. On peut énumérer les éléments de chaque X_n : $X_n = \left\{x_n^k\,,\ k \in \mathbb{N}\right\}$. On peut énumérer les éléments de la réunion de la façon suivante :

$$x_0^0\,,\ x_0^1\,,\ x_1^0\,,\ x_0^2\,,\ x_1^1\,,\ x_2^0\,,\ x_0^3\,,\ \ldots$$

comme illustré par la figure A.1.1. Dans le cas où certains des ensembles sont finis, ou si la réunion est finie, ou si les ensembles partagent certains de leurs éléments, la construction précédente permet d'établir une bijection entre la réunion et une partie de $\mathbb{N}$, cette réunion est donc finie ou infinie dénombrable. □

Proposition A.1.14. Un produit fini d'ensembles dénombrables est dénombrable.

Démonstration. On considère N ensembles dénombrables X_1, ..., X_N, pour lesquels on se donne une énumération, et l'on note $P_k \subset X_1 \times \ldots X_N$ les éléments du produits qui ne font intervenir que les k premiers termes de chacun des X_i dans l'énumération choisie. Son cardinal est le nombre de mots de N lettres que l'on peut constituer à partir d'un alphabet de cardinal k, qui est k^N, il est donc fini. Le produit des X_i est inclus dans la réunion des P_k, il est donc dénombrable comme union dénombrable d'ensembles finis (proposition A.1.13). □

Exercice A.1.2. (••) On considère l'ensemble $X = \{0,1\}^{\mathbb{N}}$ des suites infinies de 0 ou 1.

1) Montrer que X n'est pas dénombrable.

2) Montrer que le sous-ensemble X_0 des suites constantes au delà d'un certain rang est dénombrable.

3) Montrer que l'ensemble X_{per} des suites périodiques au delà d'un certain rang est dénombrable.

4) On définit l'application φ_N qui à tout $x \in X$ associe la valeur moyenne des N premiers termes. Montrer que, pour tout $x \in X_{per}$ (et donc a fortiori tout $x \in X_0$), la quantité $\varphi_N(x)$ admet une limite quand N tend vers $+\infty$. Montrer que cette propriété n'est pas vraie pour tous les éléments de X.

5) L'ensemble des éléments de X pour lesquels $\varphi_N(x)$ converge lorsque N tend vers $+\infty$ est-il dénombrable ?

Structures algébriques élémentaires

Définition A.1.15. (Loi de composition interne)
Soit X un ensemble, une *loi de composition interne* est une application de $X \times X$ dans X. On note en général $x \star y$ (ou $x + y$, ou $x \bullet y$, ou simplement xy, selon le contexte) l'image de (x, y) par cette application.
La loi est dite *commutative* si $x \star y = y \star x$ pour tout couple $(x, y) \in X \times X$.
La loi est dite associative si $(x \star y) \star z = x \star (y \star z)$ pour tous x, y, $z \in X$.

Définition A.1.16. (Loi de composition externe)

Soient X et Y deux ensemble, une *loi de composition externe* est une application de $X \times Y$ (ou $Y \times X$) dans X

L'archétype de la loi de composition externe est la multiplication d'un vecteur par un réel (ou un élément d'un corps), qui est, avec l'addition entre deux vecteurs (loi de composition interne), à la base de la notion d'*espace vectoriel.*

Définition A.1.17. (Magma)
Un magma est un ensemble muni d'une loi de composition interne (sans aucune condition sur cette loi). Ce magma est dit *unifère* s'il possède un élément neutre, i.e. un élément $e \in X$ tel que $e \star x = x \star e = x$ pour tout $x \in X$.

Exercice A.1.3. Soit X un ensemble de cardinal N fini. Quel est le nombre de magmas sur X ? De magmas commutatifs sur X ?

Définition A.1.18. (Monoïde)
Un monoïde est un ensemble muni d'une loi de composition interne associative, qui admet un élément neutre.

Définition A.1.19. (Groupe)
Un *groupe* est un ensemble G muni d'une loi de composition interne, qui vérifie les propriétés suivantes
(i) La loi est *associative*, i.e. $(x \star y) \star z = x \star (y \star z)$ pour tous x, y, $z \in G$
(ii) Il existe un élément neutre :

$$\exists e \in G\,,\ x \star e = e \star x = x \quad \forall x \in G.$$

(iii) Tout élément admet un inverse :

$$\forall x \in G\,,\ \exists y \in G\,,\ x \star y = y \star x = e.$$

On notera x^{-1} l'inverse de x.
Le groupe est dit commutatif (ou *abélien*) si $x \star y = y \star x$ pour tous x, $y \in G$.

Exercice A.1.4. Montrer que, dans la définition ci-dessus, il suffit de demander que la loi soit associative, qu'il existe un élément neutre à gauche ($e \star x = x$) pour tout x, et que tout élément admette un inverse à gauche (pour tout x, il existe y tel que $y \star x = e$).
Montrer que l'inverse d'un élément est unique

Définition A.1.20. (Morphisme / isomorphisme de groupe)
Soient $(G, \star)$ et $(G, *)$ deux groupes . On dit que l'application f de G dans G' est un morphisme (ou homomorphisme) si elle respecte la structure de groupe, i.e.

$$f(x \star y) = f(x) * f(x)$$

On parle d'*isomorphisme* s'il s'agit d'une bijection.

Définition A.1.21. (Sous-groupe)
Soit $(G, \star)$ un groupe. et $H \subset G$ une partie qui contient e, et telle que

$$x \star y \in H \quad \forall x\,,\ y \in H \text{ et } \ x^{-1} \in H \quad \forall x \in H.$$

On appelle H un sous-groupe de G, c'est un groupe pour la même loi $\star$.

Proposition A.1.22. (Sous-groupe engendré)
Soit $(G, \star)$ un groupe et S une partie de G. On appelle sous-groupe engendré par S, et l'on note $\langle S \rangle$, le plus petit sous-groupe contenant S, i.e. l'intersection de tous les sous-groupe contenant S. Ce groupe $\langle S \rangle$ est constitué des produits d'éléments ou d'inverses d'éléments de S.

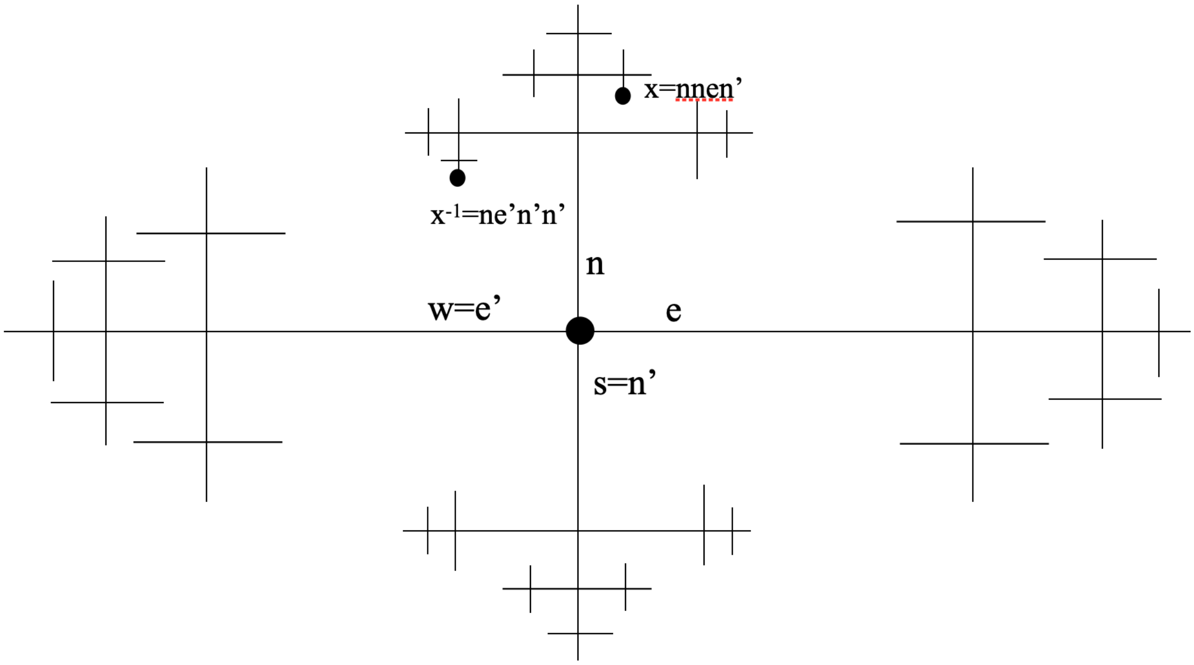

FIGURE A.1.2 – Groupre libre engendré par 2 éléments

Définition A.1.23. (Partie génératrice)
Soit $(G, \star)$ un groupe. On a qu'une partie S de G est *génératrice* si $G = \langle S \rangle$. On dit que G est de type fini s'il admet une partie génératrice finie.

La richesse de la structure de groupe repose sur les relations entre éléments, du type $x \star y = z$. De ce point de vue, le *groupe libre* engendré par un ensemble, tel que décrit ci-après joue un rôle singulier, un cas extrême dans cette grande famille, comme un groupe qui ne repose sur aucune relation autre que celle imposée par la structure de groupe.

Exemple A.1.1. (Groupe libre)
Soit S un ensemble, et S' un ensemble disjoint de S en bijection avec S. On va considérer $\cup S$ comme un *alphabet* permettant d'écrire des mots de longueur finie arbitraire, avec une loi de composition qui est la concaténation, en considérant que chaque lettre s de S a un unique s' dans S', sorte d'inverse, de telle sorte que l'on peut faire disparaître les chaînes ss' ou $s's$ qui apparaissent au sein d'un mot. Plus précisément, on considère l'ensemble X des mots finis constitué de lettres de S et S', auquel on adjoint le mot vide, et l'on considère la relation d'équivalence suivante : deux mots x et y sont équivalents si l'on peut passer de l'un à l'autre en ajoutant ou enlevant des chaines de type ss' ou $s's$. L'ensemble quotient, noté F_S, est un groupe (non commutatif) pour la loi de concaténation, d'élément neutre le mot vide.

Exercice A.1.5. Décrire le groupe libre engendré par un singleton.

Exemple A.1.2. (Groupe libre engendré par une paire)
Si S est fini, on peut représenter le groupe libre associé par un arbre. La figure A.1.2 propose une telle représentation dans le cas d'un cardinal 2. On introduit l'alphabet "cardinal" $S = \{e, n\}$ (est - nord), $S' = \{e', n'\} = \{w, s\}$ (ouest - sud). Le mot $nnen'$ est une feuille de l'arbre représenté par un point, ainsi que son inverse $x^{-1} = ne'n'n' = nwss$. Noter que si l'on représente l'inverse en retournant l'ordre de ses lettres, on trouve le symétrique de x par rapport au point central, qui correspond au mot vide. On distinguera bien F_S du groupe additif $\mathbb{Z}^2$. Chaque élément de ce dernier peut être représenté par une suite finie de directions

(en partant de l'origine), mais il existe plusieurs (une infinité) chemins conduisant à une même destination. Ainsi *ne* est différent de *en* dans le groupe libre, alors que ces chemins correspondent au même élément $(1, 1)$ de $\mathbb{Z}^2$. De ce point de vue, on peut identifier F_S à l'ensemble des chemins finis issus de 0. On notera d'ailleurs que $\mathbb{Z}^2$ est dénombrable, alors que F_S ne l'est pas (voir exercice A.1.2).

Définition A.1.24. (Anneau / corps)
Un *anneau* est un ensemble A muni de deux lois de composition internes, notées $+$ et $\star$, avec :

(i) $(A, +)$ est un groupe abélien,

(ii) La loi $\star$ est associative et distributive vis à vis de la loi $+$, i.e.

$$x \star (y + z) = x \star y + x \star z\,,\ (y + z) \star x = y \star x + z \star x \quad \forall x,\, y,\, z \in G$$

(iii) La loi $\star$ admet un élément neutre [6]

On parle de *corps* si tout élément admet un inverse pour la loi $\star$, c'est à dire que $(A, \star)$ est un groupe.

Définition A.1.25. (Espace vectoriel)
Soit $(K, +, \star)$ un corps. Un K-*espace vectoriel* E est un groupe abélien (loi $'+'$) munie d'une loi de composition externe à gauche, notée ici par simple juxtaposition :

$$(\lambda, x) \in K \times E \longmapsto \lambda x,$$

telle que l'on ait, pour tous x, y dans E, λ, μ dans K,

$$\lambda(x + y) = \lambda x + \lambda y\,,\ (\lambda + \mu)x = \lambda x + \mu x\,,\ \lambda(\mu x) = (\lambda \star \mu)x\,,\ 1_K x = x,$$

où $1_K \in K$ est l'élément neutre de la loi $\star$.

Étant donnés un corps K et un ensemble S, on peut construire un espace vectoriel, en quelque sorte minimisaliste [7], constitué des *combinaisons linéaires formelles* d'éléments e S, comme décrit ci-après.

Définition A.1.26. (Combinaisons linaires formelles, espace vectoriel libre.)
Soit K un corps et S un ensemble. On appelle combinaison linéaire formelle une expression du type

$$\sum_{x \in S} \lambda_x x\,,\ \lambda_x \in K \quad \forall x \in S,$$

où les λ_x sont tous nuls sauf un nombre fini (ce qui, signifie que les termes correspondants n'apparaissent pas dans la somme). La multiplication d'une telle expression par $\mu \in K$ est obtenue en multipliant tous les coefficients par μ. Pour sommer deux expressions, on garde les termes qui n'apparaissent que dans l'une ou l'autre, et si les deux partagent un éléments de S, on somme simplement les coefficients.
On obtient de cette façon un espace vectoriel appelé *espace vectoriel libre sur K engendré par S*. Chaque élément de K peut être identifié à une application de S dans K qui affecte une valeur non nulle à un nombre fini d'éléments de K.
La somme ci-dessus doit être vue comme une expression globale, il ne s'agit pas à proprement parler d'une loi de composition interne.

Exemple A.1.3. L'espace vectoriel des polynômes réels en X peut être vu comme le $\mathbb{R}$-espace vectoriel libre engendré par les $(X^n)_{n\in\mathbb{N}}$.

6. Cette condition n'est pas demandée par certains auteurs. Pour éviter toute ambigüité, on pourra appeler anneau *unifère* un anneau pour lequel on a bien un élément neutre pour la première loi. On pourra parler de pseudo-anneau quand cette propriété quand il n'y a pas d'élément neutre.

7. On pourra se convaincre que cette démarche donne un sens à ce que le bon sens rechigne à envisager : *l'addition de choux et de carottes*, selon le principe imparable '1 chou + 1 carotte = 1 chou + 1 carotte', en préservant la correspondance rassurante '1 chou + 1 chou = 2 choux'. Nous laissons en exercice l'addition de 2 carottes, et en question subsidiaire l'addition d'une moissonneuse-batteuse et de π flans aux pruneaux.

Lois de composition internes et modélisation

Les lois de compositions internes et les structures qu'elle induisent (groupes, anneaux, corps, ...) constituent le socle de l'algèbre, dont une bonne part des développement n'est pas directement reliée la représentation du monde réel (en dehors de la notion d'espace vectoriel, à la base de l'analyse fonctionnelle). Il peut néanmoins être fécond d'avoir une idée claire de ces notions, qui "résonnent" parfois avec le monde réel, même si le cœur de l'algèbre en est assez éloigné. Il est par ailleurs sain, dans l'approche de modélisation, d'avoir conscience des structures que l'on utilise, qui représentent en quelque sorte l'essence des objet que l'on manipule. Il nous paraît en particulier fécond, lorsque l'on manipule un ensemble multiplement structuré [8] comme l'ensemble des nombres réels (construit dans la section A.1.4 ci-après), de garder à l'esprit que l'on n'utilise qu'une partie des structures associées à cet ensemble. À titre d'illustration, lorsque l'on définit une mesure comme une application qui à une partie associe un nombre réel, l'espace d'arrivée de cette application est $\mathbb{R}_+$, qu'il est pertinent de considérer comme un *monoïde* muni de l'addition. À aucun moment (tout du moins tant qu'on ne s'intéresse pas aux mesures sur les espaces produit) on n'est amené à multiplier deux mesures, on ne fait que les sommer. On peut définir sur ce monoïde une soustraction (on retire d'une certaine quantité une quantité inférieure), mais il n'est ni nécessaire ni pertinent dans ce contexte de considérer que $m_1 - m_0$, avec $m_0 \leq m_1$ correspond à la somme de m_1 avec $-m_0$, "inverse" de m_0 pour la loi $'+'$. Concernant la structure de groupe, qui n'intervient a priori que de façon très élémentaire en analyse ($\mathbb{R}$ vu comme groupe additif abélien), précisons qu'elle peut néanmoins intervenir de façon non triviale dans des situations concrètes, lorsque l'on considère un ensemble de transformation qui laissent un ensemble invariant. L'exemple le plus élémentaire est le *groupe symétrique* $\mathbb{S}_N$, ensemble des bijections d'un ensemble à N éléments muni de la loi de composition. Un autre exemple fécond correspond aux groupe des transformations du plan qui laissent invariant une figure géométrique, ou plus généralement un ensemble de points du plan.

A.1.4 L'ensemble des réels : construction et structures afférentes

Il existe de multiples manières de construire l'ensemble $\mathbb{R}$ des réels munis de ses structures principales. La plupart des ouvrages privilégient une approche axiomatique et abstraite, nous décrivons ici une démarche plus ancrée sur la pratique quotidienne des nombres réels et leur utilisation effective, en nous en tenant ici à ce qui est strictement utile pour ce cours.

La construction proposée peut sembler périlleuse : on utilisera ci-dessous des propriétés métriques de cet ensemble, en particulier la complétude, pour définir certaines opérations comme la multiplication. Or la notion même de distance, qui est une application à image dans $\mathbb{R}$, nécessite que la droite des réels soit bien définie. On pourra cependant vérifier que la notion de métrique et de convergence d'une suite ne nécessite qu'une structure d'ordre sur $\mathbb{R}$ (qui est définie dès la proposition A.1.31), la notion de valeur absolue (définie d'emblée), et l'addition entre deux réels (définition A.1.34), qui ne nécessite pas de structure métrique sur $\mathbb{R}$ lui-même.

Au-delà de ces questions de cohérence de la construction, les paragraphes qui suivent contiennent des développement assez fastidieux visant à définir des opérations du type de celles pratiquées par les écoliers dès leur plus jeune âge, en particulier l'addition (ou la soustraction) entre nombres décimaux. Nous avons ici une petite difficulté supplémentaire liée au fait que ces opérations posées commencent *par la droite*, c'est à dire du côté où un nombre réel non décimal est infini, ce qui nécessite une adaptation de la procédure.

Nous supposons construit l'ensemble $\mathbb{Z}$ muni des deux lois $'+'$ et $'\times'$ qui en font un anneau (voir définition A.1.24), muni de la relation d'ordre total usuel. Et nous définissons l'ensemble $\mathbb{R}$ à l'aide de la représentation décimale décrite ci-dessous.

Définition A.1.27. (Ensemble des réels)
On définit l'ensemble $\mathbb{R}$ comme $\{+,-\} \times \mathbb{N} \times [\![0,9]\!]^{\mathbb{N}^\star}$, c'est-à-dire de l'ensemble des

$$\pm a_0, a_1 a_2 \dots a_k \dots$$

8. L'ensemble $\mathbb{R}$ peut être muni de la totalité des structures présentées précédemment, ainsi que d'autres qui sont présentées dans ce document (ordre total, groupe, anneau, corps, espace vectoriel, structure d'espace métrique complet, espace mesurable, espace mesuré, ...).

avec $a_0 \in \mathbb{N}$, et les a_k (appelées *décimales*) sont des entiers entre 0 et 9. On appelle *nombre réel* l'un de ces objets. On appelle nombre *décimal* un nombre dont l'écriture décimale est finie, c'est à dire que a_n est identiquement nul au-delà d'un certain rang, et l'on note $\mathbb{D}$ l'ensemble de ces nombres. On exclut *a priori* les nombres dont l'écriture finit par une infinité de 9 consécutifs, mais on prendra la liberté d'autoriser ponctuellement cette pathologie d'écriture [9], qui concerne les nombres décimaux. Tout nombre décimal peut en effet s'écrire

$$\pm\, a_0, a_1 \dots a_r 000 \dots \text{ avec } a_r \geq 1, \text{ ou } \pm\, a_0, a_1 \dots (a_r - 1) 999 \dots,$$

On appellera *propre* l'écriture d'un décimal sans l'infinité de 9.

Pour tout $a \in \mathbb{R}$, on note $-a$ le nombre obtenu en changeant le signe de a, qu'on appelle l'*opposé* de a, et $|a|$ (valeur absolue de a) le nombre obtenu en remplaçant le signe par '+'. On dira qu'un réel est strictement positif si son signe est '+' et que ses décimales ne sont pas toutes nulles, et strictement négatif si son opposé est strictement positif. 'Positif' signifie strictement positif ou nul, de même pour 'négatif'. Le nombre $0 = +0.0000\dots$ peut s'écrire aussi $0 = -0.0000\dots$, c'est le seul nombre à la fois positif et négatif.

Définition A.1.28. (Troncature entière, partie entière)
On appelle *troncature entière* de $a = \pm a_0, a_1 \dots$ l'entier relatif $\pm a_0 \in \mathbb{Z}$, et *partie entière* de a l'entier $+a_0$ si a est positif, et $-a_0 - 1$ si a est strictement négatif. La partie entière de -1.3 est ainsi -2, et sa troncature entière est -1.

Remarque A.1.29. La représentation décimale traditionnelle des réels décrite ci-dessus privilégie la notion de troncature, $\mathbb{R}$ est ainsi représenté en miroir, symétriquement par rapport à l'origine 0, qui se voit jouer de fait un rôle singulier. Nous verrons que ce choix est très adapté à la multiplication, mais moins à l'addition (définir l'addition entre nombres de signes différents demande un peu de soin). Signalons que l'on pourrait imaginer une autre convention, plus respectueuse de l'addition, invariante par translation, en représentant un nombre par un entier relatif (la partie entière), plus un nombre du type $0, a_0 a_1 \dots$, de telle sorte que par exemple $-1,23$ s'écrirait $(-2) + 0.77$. Selon cette écriture, il serait non trivial de construire l'opposé d'un nombre (alors que c'est immédiat avec la représentation-miroir que nous privilégions), ainsi que le produit entre deux nombres de signes différents. Nous utiliserons ponctuellement cette vision des choses lors de la construction de l'addition entre deux nombres de signes distincts.

Théorème A.1.30. L'ensemble $\mathbb{D}$ des nombres décimaux est infini dénombrable, l'ensemble $\mathbb{R}$ des nombres réels est infini non dénombrable.

Démonstration. L'ensemble $\mathbb{D}$ contenant $\mathbb{N}$, il est au moins infini dénombrable. Il s'écrit par ailleurs comme union des $\mathbb{D}_n$, qui sont les nombres décimaux dont les décimales sont nulles au-delà du rang n. Chacun de ces ensembles étant dénombrable (en multipliant par 10^n on retrouve les entiers), $\mathbb{D}$ est dénombrable comme réunion d'ensembles dénombrables (proposition A.1.13).

Montrons maintenant, en suivant une démarche proche de la démonstration du théorème A.1.10, que l'intervalle $[0, 1[$ n'est pas dénombrable. Supposons qu'il le soit, on peut alors énumérer ses éléments

$$\begin{array}{rcl} r_1 & = & 0, \mathbf{a_1^1}\, a_1^2\, a_1^3 \dots \\ r_2 & = & 0, a_2^1\, \mathbf{a_2^2}\, a_2^3 \dots \\ r_3 & = & 0, a_3^1\, a_3^2\, \mathbf{a_3^3} \dots \\ \vdots & = & 0, \vdots \end{array}$$

On construit alors un nombre par le procédé d'*extraction diagonale de Cantor* (décimales indiquées en gras ci-dessus), chaque décimale de ce nombre étant définie selon le principe suivant : si la n-ième décimale de r_n est différente de 1, on la fixe à 1, si elle est égale à 1, on la fixe à 2 (par exemple). On construit ainsi un nombre réel en écriture propre qui par construction ne peut pas figurer dans la liste ci-dessus, qui est pourtant une énumération exhaustive de $[0, 1[$. On en déduit par contradiction que $[0, 1[$ n'est pas dénombrable. $\square$

9. Il est prudent de garder cette possibilité, du fait que ces objets sont susceptibles d'apparaître spontanément, comme lorsque l'on définira des sommes du type $0.111 \cdots + 0.888 \dots$.

Proposition A.1.31. (Relation d'ordre total)
L'ensemble $\mathbb{R}$ admet une relation d'ordre total $\leq$.

Démonstration. Soient $a = a_0, a_1 \dots$ et $b = b_0, b_1 \dots$ deux éléments de $\mathbb{R}$ positifs et proprement écrits. S'ils sont différents, il existe un plus petit $k \in \mathbb{N}$ tel que $a_k \neq b_k$. Si $a_k < b_k$, on dit que $a < b$, et $b < a$ dans le cas contraire. On dit que tout négatif est inférieur à tout positif, et que, pour deux nombres a et b négatifs, on a $a < b$ si et seulement si $-b < -a$. Cette relation d'ordre permet de définir la notion d'intervalles de type $[a, b]$, $]a, b[$, $[a, b[\dots$ □

Proposition A.1.32. Toute partie non vide majorée de $\mathbb{R}$ admet une borne supérieure (voir définition A.1.6).

Démonstration. Soit A une partie de $\mathbb{R}$ majorée. On suppose dans un premier temps que A contient au moins un nombre strictement positif. Il existe

$$m_0 = \max_{a \in A} \{P_0(a)\} \geq 0,$$

où P_k associe à un réel sa k-ième décimale (ou sa troncature entière pour $k = 0$). On introduit $A_0 = \{a \in A\,,\ P_0(a) = m_0\}$, et l'on pose

$$m_1 = \max_{a \in A_0} \{P_1(a)\} \in [\![0, 9]\!].$$

On construit ainsi par récurrence le nombre $M = m_0, m_1 m_2 \dots$ qui est une borne supérieure de A par construction. Si A ne contient que des nombres négatifs, on procède de façon analogue en considérant l'ensemble $-A$ et en remplaçant le max par un min. □

Définition A.1.33. (Supremum - infimum - maximum - minimum)
Soit A une partie de $\mathbb{R}$. On note $\sup(A)$ sa borne supérieure (on dit qu'elle vaut $+\infty$ si A n'est pas majoré). Si elle appartient à A, on l'appelle le plus grand élément de A, ou maximum de A, qu'on écrit $\max(A)$.

Si A est majoré, $M = \sup A$ si et seulement si M majore A et s'il existe une suite x_n d'éléments de A qui tend vers M. On appellera x_n *une suite maximisante.*

On définit symétriquement les notions d'infimum, de plus petit élément, et de suite minimisante.

Définition A.1.34. (Addition sur les réels en écriture décimale)
On définit l'addition sur $\mathbb{R}$ de la façon suivante : soient $a = a_0, a_1 \dots$ et $b = b_0, b_1 \dots$ deux éléments de $\mathbb{R}$ positifs et proprement écrits. Pour alléger les notations, nous proposons de présenter la construction de la somme $c = c_0, c_1 \dots$.. comme un algorithme informatique, c'est à dire en gardant la notation c_k pour désigner une quantité dont la valeur est susceptible de changer au fil de la construction. On pose dans un premier temps $c_k = a_k + b_k$, et l'on définit $\alpha = (\alpha_k)$ comme une suite identiquement nulle au départ. Si $c_k \geq 10$, on remplace sa valeur par $c_k - 10$ et l'on pose $\alpha_{k-1} = 1$. Noter que dans ce cas la nouvelle valeur de c_k est inférieure ou égale à 8 (car $a_k + b_k$ est inférieur ou égal à 18). Si c se termine par une infinité de 9 consécutifs (à partir d'un rang k), alors d'après la remarque précédente la zone en question est vierge de toute retenue, on nettoie l'écriture en mettant à 0 tous les 9, et l'on rajoute 1 à c_{k-1}, qui est donc au maximal égal à 9, et sans menace de retenue puisque α_{k-1} est nécessairement égal à 0 d'après ce qui précède. On obtient donc un $c = (c_k)$ qui est soit décimal, soit non décimal et non stationnaire à 9. L'étape finale consiste à prendre en compte les retenues dans la somme finale. On considère pour cela les suites (nécessairement finies) de 9 consécutifs. Considérons une de ces suites de longueur maximale [10] $c_k c_{k+1} \dots c_\ell = 99 \dots 9$, encadrée donc par des décimales non égales à neuf. Toujours selon le même argument que la somme de deux naturels ≤ 9 est ≤ 18, on a nécessairement

$$\alpha_{k-1} = \alpha_k = \dots = \alpha_{\ell-1} = 0.$$

Si $\alpha_\ell = 0$, on laisse la suite de 9 inchangée, et si $\alpha_\ell = 1$, on remplace tous les 9 par des 0, et l'on rajoute 1 à c_{k-1} (qui est ≤ 8 par hypothèse). En dehors de ces paquets de 0 l'ajout de α_k à c_k peut se faire directement. On obtient ainsi un nombre réel en écriture décimale, que l'on définit comme la somme de a et b.

10. C'est-à-dire qu'elle n'est pas contenue dans une suite de 9 strictement plus longue.

Il s'agit maintenant de définir la somme entre un nombre positif et un nombre négatif. On commence par considérer un nombre de l'intervalle $]-1,0[$, $a=-0,a_1a_2\dots$ (écriture propre), auquel on ajoute $+1$. Si a est décimal, $a=-0,a_1a_2\dots a_p$, la somme est définie comme

$$1+a=0,c_1c_2\dots c_p\,,\ c_i=9-a_i \text{ pour } i=1,\dots,p-1\,,\ c_p=10-a_p.$$

Si le nombre est non décimal, on définit simplement la différence par $b_i=9-b_i$ pour tout i. On considère maintenant un réel $a>0$ et un entier $b<0$. On utilise la décomposition formelle (formelle car on n'a pas encore défini l'addition entre a et b)

$$a+b=a_0+0,a_1a_2\dots-b_0=a_0-b_0+0,a_1a_2\dots.$$

Si l'entier a_0-b_0 est positif ou nul, on se ramène à la somme de deux réels positifs déjà définie. Si cet entier est strictement négatif, on définit la somme comme

$$a+b=-c_0,c_1\dots \text{ avec } c_0=|a_0-b_0|-1 \text{ et } 0,c_1c_2\dots=1-0,a_1a_2\dots$$

(cette dernière somme a déjà été définie précédemment). Pour finir, on considère maintenant $a>0$ et $b<0$ non entier. On écrit

$$a+b=a_0+0,a_1a_2\dots-b_0-1+(1-0,b_1\dots)=a_0-b_0-1+0,a_1a_2\dots+(1-0,b_1b_2\dots).$$

Le terme entre parenthèses a été défini précédemment comme un réel de l'intervalle $]0,1[$. On sait effectuer sa somme avec $0,a_1a_2\dots$ (tous deux positifs). Si cette somme est dans l'intervalle $]0,1[$, on se retrouve dans la situation précédente. Sinon, on l'écrit $1+0,d_1d_2\dots$, et l'on est une nouvelle fois ramené à donner un sens à la somme d'un entier $a_0-b_0-1+1=a_0-b_0$ avec un nombre du type $0,d_1d_2\dots$, cas qui a déjà été traité.

Proposition A.1.35. Soit (x_n) une suite de réels majorée et croissante. Alors (x_n) converge vers une limite $\ell\in\mathbb{R}$, qui est la borne supérieure de l'ensemble des termes de la suite.

Démonstration. L'ensemble X des termes de la suite est majoré, il admet donc une borne supérieure $\ell\in\mathbb{R}$, telle que $x_n\le\ell$ pour tout n. Pour tout $\varepsilon>0$, il existe un terme de la suite supérieur à $\ell-\varepsilon$. la suite étant croissante, tous les termes de la suite sont dans l'intervalle $]\ell-\varepsilon,\ell]$, d'où la convergence de x_n vers ℓ. □

Pour toute suite (x_n) de réels, le supremum des x_k pour k plus grand que n est soit identiquement égal à $+\infty$ (si la suite n'est pas majorée), soit décroissant en n. Dans le second cas cette quantité converge donc vers $-\infty$ ou une limite réelle. De même l'infimum des x_k pour k plus grand que n est identiquement $-\infty$, ou réel croissant, et converge dans ce cas vers $+\infty$ ou une limite réelle. Ces conséquences directes de la proposition précédente permettent de définir les notions de lim sup et lim inf.

Définition A.1.36. (Limite inférieure, limite supérieure (●●))
Soit x_n une suite réele, on définit

$$\limsup_{n\to+\infty} x_n=\lim_{n\to+\infty}\left(\sup_{k\ge n} x_k\right)\in[-\infty,+\infty]\,,\ \liminf_{n\to+\infty} x_n=\lim_{n\to+\infty}\left(\inf_{k\ge n} x_k\right)\in[-\infty,+\infty]\,,$$

Exercice A.1.6. (Limites supérieure et inférieure (●))
a) Donner les lim sup et lim inf de la suite (x_n) dans les cas suivants

$$x_n=\frac{1}{n}\,,\ x_n=n\,,\ x_n=(-1)^n\,,\ x_n=\sin(n).$$

b) Que peut-on dire d'une suite dont la lim inf et la lim sup ont la même valeur finie ?

Exercice A.1.7. (●) Soient (x_n) et (y_n) deux suites réelles. Montrer que

$$\sup_n(x_n+y_n)\le\sup(x_n)+\sup(y_n).$$

Donner un exemple pour lequel il y a en fait égalité, et un exemple pour lequel l'inégalité est stricte.

Définition A.1.37. (Convergence d'une suite de réels (•))
On dit qu'une suite (a_n) de réels tend vers 0 si sa valeur absolue peut être rendue arbitrairement petite au-delà d'un certain rang, c'est à dire :

$$\forall \varepsilon > 0\,,\ \exists N \in \mathbb{N}\,,\ |a_n| < \varepsilon \quad \forall n \geq N.$$

On dit qu'une suite (a_n) de réels tend vers une limite a si $|a_n - a|$ tend vers 0.

Le fait d'avoir défini une relation d'ordre total et une addition sur $\mathbb{R}$ permet de donner un sens à la définition générale d'une métrique, selon la définition I.2.1, page 11. Cette définition peut être appliquée à $\mathbb{R}$ lui-même, pour définir la distance canonique basée sur la valeur absolue de la différence entre deux nombres.

Proposition A.1.38. (Métrique sur $\mathbb{R}$)
L'application $(x, y) \in \mathbb{R} \longmapsto d(x, y) = |x - y|$ définit une distance sur $\mathbb{R}$.

Démonstration. La séparation est immédiate, ainsi que la symétrie. Soient maintenant 3 réels x, y, et z. Si $x - y$ et $y - z$ sont de même signe, on a

$$|x - y| + |y - z| = |x - z|\,,$$

et s'il sont de signes opposés, on a

$$|x - z| = |x - y + (y - z)| \leq \max(|x - y|\,, |y - z|) \leq |x - y| + |y - z|\,.$$

dans les deux cas, l'inégalité triangulaire est vérifiée. □

Proposition A.1.39. L'espace métrique $(\mathbb{R}, d)$ est complet (voir définition I.5.3, page 20).

Démonstration. On considère une suite de Cauchy dans $\mathbb{R}$ (défini par A.1.27), notée[11] (a^n). La petite difficulté pour montrer que la suite converge est que, pour $n \in \mathbb{N}$, il est possible que a_k^n oscille entre deux valeurs successives, puisque deux nombres qui ne s'identifient que sur les $k - 1$ premières décimales peuvent être arbitrairement proches. Plus précisément, si l'on fixe $r \in \mathbb{N}$, alors $|a - b| < 10^{-r}$ impose l'alternative suivante :

1. les décimales de a et b s'identifient jusqu'au rang $r - 1$, et diffèrent d'une unité au r-ème rang ;
2. ces décimales diffèrent d'une unité dès un certain rang $\ell < r$, par exemple $b_\ell = a_\ell + 1$, et dans ce cas $a_i = 9$ et $b_i = 0$ pour $\ell + 1 \leq i \leq r$.

Considérons maintenant une suite de Cauchy dans $\mathbb{R}$ en écriture décimale, $(a^n) = (a_k^n)$ (l'indice k représente comme précédemment les décimales, et n l'indice du terme de la suite). Si toutes les décimales se stabilisent au delà d'un certain rang, c'est qu'il existe $\ell = (\ell_k)$ tel que

$$\forall k \in \mathbb{N}\,,\ \exists N_k\,,\ \forall n \geq N_k\,,\ a_k^n = \ell_k,$$

on a alors convergence de a^n vers ℓ. Si ce n'est pas le cas, notons k_0 le plus petit des indices k tels que a_k^n ne se stabilise jamais. Comme a^p et a^q deviennent arbitrairement proches, d'après l'alternative énoncée ci-dessus, la décimale a_k^n oscille nécessairement entre deux valeurs distantes de 1, disons ℓ_k et $\ell_k + 1$, les indices précédents se stabilisant. D'après la remarque faite en préambule de cette démonstration, pour tout $r \in \mathbb{N}$ grand, tous les termes de la suite pour $n \geq r$ prennent nécessairement l'une ou l'autre des formes

$$\ell_0, \ell_1 \dots \ell_{k-1}\ell_k 999999 \dots 99\underbrace{9}_{r} \dots \quad \text{ou} \quad \ell_0, \ell_1 \dots \ell_{k-1}(\ell_k + 1)0000000 \dots 00\underbrace{0}_{r} \dots,$$

qui implique la convergence vers le décimal $\ell_0, \ell_1 \dots \ell_{k-1}(\ell_k + 1)000\dots$ (la première forme correspond à l'écriture impropre de ce même décimal). □

11. On prendra garde à la notation a^n, où n ne représente pas une puissance mais un indice : $a^n \in \mathbb{R}$ désigne simplement le $n-$ième terme de la suite. Chaque terme a^n est lui-même défini par une infinité de chiffres, les a_k^n pour $k = 0, 1, \dots$

Définition A.1.40. (Produit de deux réels)
On définit le produit de deux réels par passage à la limite à partir du produit entre deux décimaux, qui lui-même découle du produit entre deux entiers. La démarche repose sur les trois étapes décrites suivantes :

1. (Produit par une puissance de 10)
   En premier lieu, nous définissons le produit entre un réel et une puissance de 10. Pour tout réel $a = a_0, a_1a_2\dots$, tout entier naturel n, on définit $10^n \times a$ comme le réel obtenu en décalant vers la droite la virgule de n pas. On définit $10^{-n} \times a$ comme le nombre obtenu en décalant la virgule de n pas vers la gauche.
2. (Produit entre deux décimaux)
   Soient $a = a_0, a_1 \dots a_n$ et $b = b_0, b_1 \dots b_m$ deux nombres décimaux. On définit le produit $a \times b$ comme
   $$a \times b = (\underbrace{10^n \times a}_{\in\mathbb{N}}) \times (\underbrace{10^m \times b}_{\in\mathbb{N}}) \times 10^{-m-n}.$$
3. (Produit entre deux réels)
   Soient $a = +a_0, a_1a_2\dots$ et $b = +b_0, b_1b_2\dots$ deux nombres réels. Pour tous $n \geq 0$, $m \geq n$, on note $a_{n,m}$ le nombre décimal obtenu en ne conservant que les décimales entre n et m. Le nombre a est ainsi limite de la suite $(a_{0,n})$ quand n tend vers $+\infty$, de même pour b. Montrons que la suite de décimaux $(a_{0,n} \times b_{0,n})$ est de Cauchy dans $\mathbb{R}$. On a, pour tous p, q avec $p < q$,
   $$\begin{aligned} a_{0,q} \times b_{0,q} - a_{0,p} \times b_{0,p} &= (a_{0,p} + a_{p+1,q}) \times (b_{0,p} + b_{p+1,q}) - a_{0,p} \times b_{0,p} \\ &= a_{p+1,q} \times b_{0,p} + a_{0,p} \times b_{p+1,q} + a_{p+1,q} \times b_{p+1,q}, \end{aligned}$$
   qui est majoré en valeur absolue par $|b| \times 10^{-p} + |a| \times 10^{-p} + 10^{-2p}$, qui tend vers 0 quand p et q tendent vers $+\infty$. La suite $(a_{0,n} \times b_{0,n})$, de Cauchy, converge donc vers une limite. Le produit $a \times b$ est défini comme cette limite.

Définition A.1.41. (Droite numérique achevée)
On appelle droite numérique achevée, et l'on note $\mathbb{R}$, l'ensemble $\overline{\mathbb{R}}$ auquel on rajoute deux points noté $+\infty$ et $-\infty$.

La relation d'ordre sur $\mathbb{R}$ est étendue à $\overline{\mathbb{R}}$ par $-\infty < +\infty$ et, pour tout réel a, $-\infty < a < +\infty$.

A.1.5 Inégalités fondamentales

Proposition A.1.42. (Inégalité arithmético-géométrique)
Soient x_1, …, x_n des réels positifs ou nuls, et $(\alpha_n) \in]0, +\infty[^n$ une famille de poids. On a

$$(x_1^{\alpha_1} \dots x_n^{\alpha_n})^{1/\alpha} \leq \frac{1}{\alpha}\sum_{i=1}^{n} \alpha_i x_i,$$

avec $\alpha = \sum \alpha_i$.

Démonstration. Par concavité de la fonction logarithme, on a

$$\frac{1}{\alpha}\sum \alpha_n \log x_n \leq \log\left(\frac{1}{\alpha}\sum \alpha_n x_n\right),$$

d'où l'inégalité en prenant l'exponentielle. □

Proposition A.1.43. (Inégalité de Young)
Soient a et b deux réels positifs où nuls, et p, q deux réels > 0 conjugués, i.e. tels que $\frac{1}{p} + \frac{1}{q} = 1$. On a alors

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Démonstration. C'est une conséquence de l'inégalité arithmético-géométrique (proposition A.1.42), avec $\alpha_1 = 1/p$, $\alpha_2 = 1/q$, $x_1 = a^p$, et $x_2 = b^q$. □

Proposition A.1.44. (Inégalité de Hölder)
Soient p et q deux réels positifs conjugués, i.e. tels que $1/p + 1/q = 1$, et $\theta = (\theta_i) \in [0, +\infty[^d$. Pour tous $x = (x_i)$, $y = (y_i) \in \mathbb{R}^d$, on a

$$\sum_{i=1}^{d} |\theta_i x_i y_i| \leq \left(\sum_{i=1}^{d} \theta_i \, |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{d} \theta_i \, |y_i|^q \right)^{1/q}.$$

Démonstration. Remarquons en premier lieu que cette inégalité est $1-$ homogène vis à vis de x et y : si elle est valable pour x et y, elle est aussi valable pour λx et μy, quels que soient les réels λ et μ. Il suffit donc de la démontrer dans le cas particulier où $\sum \theta_i \, |x_i|^p = \sum \theta_i \, |y_i|^p = 1$, c'est-à-dire de montrer que, pour x et y ainsi normalisés, on a

$$\sum_{i=1}^{d} |\theta_i x_i y_i| \leq 1.$$

Cette inégalité résulte directement de l'inégalité de Young (proposition A.1.43) :

$$\sum_{i=1}^{d} |\theta_i x_i y_i| \leq \sum_{i=1}^{d} \theta_i \left(\frac{|x_i|^p}{p} + \frac{|y_i|^q}{q} \right) = \frac{1}{p} \sum_{i=1}^{d} \theta_i \, |x_i|^p + \frac{1}{q} \sum_{i=1}^{d} \theta_i \, |y_i|^q = \frac{1}{p} + \frac{1}{q} = 1,$$

ce qui conclut la preuve. □

Proposition A.1.45. (Inégalité de Minkovski)
Soit $p \in [1 + \infty]$, et $\theta = (\theta_i) \in [0, +\infty[^d$. Pour tous $x = (x_i)$, $y = (y_i) \in \mathbb{R}^d$, on a

$$\left(\sum_{i=1}^{d} \theta_i \, |x_i + y_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^{d} \theta_i \, |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^{d} \theta_i \, |y_i|^p \right)^{1/p}.$$

Démonstration. L'inégalité est immédiate pour le cas $p = +\infty$. Pour $p \in [1, +\infty[$, on écrit

$$\sum_{i=1}^{d} \theta_i \, |x_i + y_i|^p = \sum_{i=1}^{d} \theta_i \, |x_i + y_i| \, |x_i + y_i|^{p-1} \leq \sum_{i=1}^{d} \theta_i (|x_i| + |y_i|) \, |x_i + y_i|^{p-1}$$

$$= \sum_{i=1}^{d} \theta_i \, |x_i| \, |x_i + y_i|^{p-1} + \sum_{i=1}^{d} \theta_i \, |y_i| \, |x_i + y_i|^{p-1}.$$

On applique alors l'inégalité de Hölder (proposition A.1.44) à chacun des deux termes de la somme avec les indices p et $p/(p-1)$, pour obtenir

$$\begin{aligned} \sum_{i=1}^{d} \theta_i \, |x_i + y_i|^p \quad & \leq \quad \left(\sum_{i=1}^{d} \theta_i \, |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{d} \theta_i \, |x_i + y_i|^p \right)^{(p-1)/p} \\ & + \quad \left(\sum_{i=1}^{d} \theta_i \, |y_i|^p \right)^{1/p} \left(\sum_{i=1}^{d} \theta_i \, |x_i + y_i|^p \right)^{(p-1)/p}. \end{aligned}$$

On divise les deux membres de cette inégalité par $\left(\sum \theta_i \, |x_i + y_i|^p \right)^{1-1/p}$ (si cette quantité est nulle, l'inégalité à démontrer est trivialement vérifiée), pour obtenir l'inégalité annoncée. □

A.2 Pour aller plus loin (●●●●)

A.2.1 Théorie des ensembles, cardinalité

Axiome A.2.1. (Axiome du choix (● ● ●))
Soit $(X_i)_{i\in I}$ une famille d'ensembles. Il existe une application qui à chaque ensemble X_i associe un élement x_i de cet ensemble.

Théorème A.2.2. (Cantor-Bernstein (●●●●))
Si $X \lesssim Y$ et $Y \lesssim X$, on a $X \simeq Y$. En d'autres termes, s'il existe une injection de X dans Y, et une injection de Y dans X, alors il existe une bijection entre X et Y.

A.2.2 Complété d'un espace métrique (●●●●)

Définition A.2.3. Soit (X, d) un espace métrique. On appelle complété de cet espace la donnée d'un espace métrique complet $(\bar{X}, \bar{d})$ muni d'une isométrie

$$T \ : \ (X, d) \longrightarrow (\bar{X}, \bar{d})$$

dont l'image est dense dans $\bar{X}$.

Théorème A.2.4. (Complété d'un espace métrique)
Tout espace métrique admet un complété, qui est unique à isométrie près.

Démonstration. Soit (X, d) un espace métrique. On munit l'espace des suites dans X de la relation d'équivalence

$$x = (x_n)_{n\in\mathbb{N}}\ ,\ x' = (x'_n)_{n\in\mathbb{N}}\ ,\ x \ \mathcal{R}\ x' \Longleftrightarrow d(x_n, x'_n) \longrightarrow 0$$

On note C l'ensemble des suites de Cauchy dans X, et $\overline{X} = C/\ \mathcal{R}$ l'espace quotient. Pour $\overline{x}$, $\overline{x}'$ dans $\overline{X}$, $(x_n) \in \overline{x}$, $(x'_n) \in \overline{x}'$, la quantité $d(x_n, x'_n)$ converge vers une limite qui ne dépend pas des représentant choisis. En effet, on a

$$d(x_p, x'_p) - d(x_q, x'_q) \le d(x_p, x_q) + d(x_q, x'_q) + d(x'_q, x'_p) - d(x_q, x'_q) = d(x_p, x_q) + d(x'_q, x'_p)$$

qui tend vers 0 quand p, q tendent vers $+\infty$. On montre de la même manière que son opposé $d(x_q, x'_q) - d(x_p, x'_p)$ est majoré par une quantité qui tend vers 0. La valeur absolue de $d(x_p, x'_p) - d(x_q, x'_q)$ tend donc vers 0 quand p et q tendent vers $+\infty$. La suite $d(x_n, x'_n)$ est donc de Cauchy dans $\mathbb{R}$, donc converge vers une limite $\ell \in \mathbb{R}$. On montre immédiatement que cette limite ne dépend pas du représentant choisi du fait de l'adjacence des suites d'une même classe. On note $\bar{d}(\overline{x}, \overline{x}')$ cette limite.

On montre tout aussi immédiatement que $\bar{d}(\,\cdot\,,\,\cdot\,)$ est une distance sur $\overline{X}$.

On note T l'application qui à une suite constante dans X (donc de Cauchy) associe sa classe dans $\overline{X}$. Cette application est par construction une isométrie de X vers $\overline{X}$. Montrons que son image est dense dans $\overline{X}$. Soit $\overline{x}$ une classe de $\overline{X}$, et (x_n) l'un de ses représentants. On note $\overline{x_n}$ la classe de la suite constante égale à x_n. On a

$$\bar{d}(\overline{x_n}, \overline{x}) = \lim_q \bar{d}(x_n, x_q),$$

qui tend vers 0 quand n tend vers l'infini du fait du caractère de Cauchy de (x_n).

Montrons maintenant que $\overline{X}$ muni de $\bar{d}$ est un espace métrique complet. On considère une suite $(\overline{x}^n)$ de Cauchy dans $\overline{X}$. Pour tout n, comme on l'a vu précédemment, la classe $\overline{x}^n$ peut être approchée par la classe d'une suite constante écrite $\overline{u_n}$, avec $u_n \in X$, à $1/n$ près. On considère maintenant la suite $u = (u_n)$. Cette suite est de Cauchy par construction. En effet, on a

$$d(u_p, u_q) = \bar{d}(\overline{u_p}, \overline{u_q}) \le \bar{d}(\overline{u_p}, \overline{x}^p) + \bar{d}(\overline{x}^p, \overline{x}^q) + \bar{d}(\overline{x}^q, \overline{u_q}) \le \frac{1}{p} + \bar{d}(\overline{x}^p, \overline{x}^q) + \frac{1}{q}.$$

On note $\overline{u}$ sa classe. On a

$$\bar{d}(\overline{x}^n, \overline{u}) \leq \bar{d}(\overline{x}^n, \overline{u_n}) + \bar{d}(\overline{u_n}, \overline{u}),$$

qui tend vers 0 par construction. □

A.2.3 Topologie générale (●●●●)

Définition A.2.5. (Topologie, ouverts, fermés)
Soit X un ensemble. On appelle topologie sur X la donnée d'une famille $\mathcal{T}$ de parties de X, qu'on appelle les *ouverts*, telle que

(i) l'ensemble vide et X appartiennent à $\mathcal{T}$,
(ii) toute union d'ouverts est un ouvert,
(iii) toute intersection finie d'ouverts est un ouvert.

On appelle fermé le complémentaire d'un ouvert.

On appelle le couple $(X, \mathcal{T})$ un *espace topologique*.

Définition A.2.6. (Finesse)
Soient $\mathcal{T}$ et $\mathcal{T}'$ deux topologie sur X. On dit que $\mathcal{T}'$ est *plus fine* que $\mathcal{T}$ si $\mathcal{T} \subset \mathcal{T}'$, i.e. si tout ouvert de $\mathcal{T}$ est ouvert de $\mathcal{T}'$.

Définition A.2.7. (Topologies discrète et grossière)
Tout ensemble X peut être muni de la topologie *discrète*, pour laquelle tout singleton, et donc toute partie, est un ouvert. Toute partie est donc à la fois ouverte et fermée pour la topologie discrète. C'est la plus fine des topologies dont on puisse équiper X. À l'opposé, pour la topologie *grossière*, seuls $\emptyset$ et X sont des ouverts. C'est la topologie la moins fine.

Définition A.2.8. (Voisinage)
Soit $(X, \mathcal{T})$ un espace topologique, et $x \in X$. On appelle voisinage de x toute partie de X qui contient un ouvert contenant x. On note $\mathcal{V}(x)$ l'ensemble des voisinage de X.

Définition A.2.9. (Espace topologique séparé)
Un espace topologique $(X, \mathcal{T})$ est dit *séparé* si pour tous x, x' dans X, distincts, il existe des voisinage U et U' de x et x', respectivement, avec $U \cap U' = \emptyset$.

La topologie discrète est séparée, la topologie grossière ne l'est pas (dès que l'ensemble comporte au moins 2 éléments).

Proposition A.2.10. Tout espace métrique muni de la topologie associée est séparé.

Définition A.2.11. Une application d'un espace topologique $(X, \mathcal{T})$ dans un espace topologique $(X', \mathcal{T}')$ est dite *continue* si l'image réciproque de tout ouvert est un ouvert.

Exercice A.2.1. Montrer qu'une application d'un espace topologique $(X, \mathcal{T})$ dans un espace topologique $(X', \mathcal{T}')$ est continue si et seulement si l'image réciproque de tout fermé est un fermé.

Proposition A.2.12. L'application identité de $(X, \mathcal{T}')$ dans $(X, \mathcal{T})$ est continue si et seulement si $\mathcal{T}'$ est plus fine que $\mathcal{T}$.

Exercice A.2.2. a) Décrire l'ensemble des applications continues de $(X, \mathcal{T})$ dans $(X', \mathcal{T}')$ lorsque $\mathcal{T}'$ est la topologie grossière sur X'.

b) Décrire l'ensemble des applications continues de $(X, \mathcal{T})$ dans $(X', \mathcal{T}')$ lorsque $\mathcal{T}$ est la topologie discrète sur X'.

Définition A.2.13. (Topologie induite)
Soit X un espace topologique et $A \subset X$. L'ensemble des intersections d'ouverts de X avec A munit A d'une topologie, appelée *topologie induite*.

Définition A.2.14. (Connexité)
Soit X un espace topologique. On dit que X est *connexe* si les seules parties de X à la fois ouvertes et fermées sont X et $\emptyset$. On dit que $A \subset X$ est connexe si A muni de la topologie induite est connexe.

Proposition A.2.15. L'espace X est connexe s'il n'admet aucune partition [12] en deux ouverts.

Proposition A.2.16. Une union de parties connexes d'intersection non vide est connexe.

Démonstration. Soit C l'union d'une famille $(C_i)_{i\in I}$ de parties connexes, et $x \in \cap C_i$. Considérons une partition de C pour la topologie induite :

$$C = (U_1 \cup U_2) \cap C,$$

où U_1 et U_2 sont des ouverts de X. Le point x est nécessairement dans l'un des deux ouverts, par exemple $x \in U_1$. Pour tout C_i, l'union disjointe des ouverts U_1 et U_2 recouvre C_i. Comme C_i est connexe, l'intersection d'un des deux ensembles avec C_i est nécessairement vide, comme ça ne peut pas être U_1 qui contient $x \in C_i$, c'est U_2. On a donc $U_2 \cap C_i = \emptyset$ pour tout i. Ainsi C n'admet aucune partition en deux ouverts (non vides), C est donc connexe. □

Définition A.2.17. (Composantes connexes)
Soit X un espace topologique. Pour tout $x \in X$ on note C_x la plus grande partie connexe contenant x, définie comme l'union des connexes contenant x (qui est bien connexe d'après la proposition précédente). Pour $x \neq y$, on a $C_x = C_y$ ou $C_x \cap C_y = \emptyset$, toujours d'après la proposition précédente. On peut donc introduire la relation d'équivalence suivante : $x \mathcal{R} y$ si $C_x = C_y$. Les classe d'équivalence de cette relation sont appelée *composantes connexes* de X.

Remarque A.2.18. Un espace connexe est un espace qui ne possède qu'une seule composante connexe, qui est lui-même tout entier. Pour la topologie discrète, tout ensemble X est *totalement discontinu*, c'est à dire que la composante connexe de chaque point est réduite à lui-même. Pour la topologie grossière, tout ensemble est connexe.

Suites

Définition A.2.19. (Suites convergentes)
Soit (x_n) une suite d'éléments de $(X, \mathcal{T})$. On dit que cette suite converge vers $x \in X$ si, pour tout ouvert U contenant x, il existe N tel que, pour tout $n \geq N$, $x_n \in U$.

Exercice A.2.3. Décrire l'ensemble des suites admettant une limite dans le cas où la topologie est discrète, et dans le cas où la topologie est grossière.

Proposition A.2.20. (Unicité de la limite dans un espace séparé)
Soit $(X, \mathcal{T}')$ un espace séparé. Alors tout suite convergente converge vers une limite unique.

Compacité

Définition A.2.21. (Espace topologique compact (•))
Soit $(X, \mathcal{T})$ un espace topologique , et K une partie de X (qui peut être X lui-même). On dit que K est

12. On rappelle que les membres d'une partition doivent être non vides.

compact s'il vérifie la propriété de *Borel–Lebesgue* : de tout recouvrement de K par des ouverts on peut extraire un recouvrement fini :

$$K \subset \bigcup_{i\in I} U_i\,,\ U_i \text{ ouvert } \ \forall i \in I \Longrightarrow \exists J \subset I\,,\ J \text{ fini, tel que } K \subset \bigcup_{i\in J} U_i.$$

Exercice A.2.4. Décrire les compacts de $(X, \mathcal{T})$ lorsque $\mathcal{T}$ est la topologie discrète (respectivement grossière)

L'exercice précédent illustre de façon caricaturale que moins il y a d'ouverts (c'est à dire plus la topologie est grossière), plus il y a de compacts. C'est cette dualité qui motive la construction de topologies les plus grossières possibles [13] de façon à avoir le plus de compacts possibles. On prendra néanmoins garde au fait qu'une suite dans un compact (au sens de la topologie générale), n'admet pas nécessairement de sous-suite convergente.

13. Il est néanmoins nécessaire de conserver la propriété de séparation pour pouvoir espérer construire des objets comme limites de suites.

Annexe B

Compléments sur la théorie de la mesure et de l'intégration

Sommaire

B.1 Motivations, vue d'ensemble

Cette première section précise au travers d'exemples la nature des objets abstraits construits dans les sections suivantes, et les difficultés associées à cette construction. La notion centrale est celle de *mesure*. Comme cadre conceptuel d'appréhension du réel, cette notion unique de mesure répond à deux enjeux, qu'il nous paraît important de distinguer malgré le fait qu'ils correspondent à la même notion mathématique.

En premier lieu, une mesure permet de structurer le fond d'un espace destiné à accueillir de la matière. Par espace nous entendons par exemple l'espace euclidien usuel, qui est en dimension 3 un modèle de l'espace physique dans lequel nous vivons, sur lequel il peut être pertinent de définir des *champs* (champ de densité, de concentration d'un polluant, de température, de densité de population, ...). Considérons par exemple un milieu occupant une certaine zone de l'espace euclidien, milieu dont on connait la densité. Si l'on suppose la densité constante sur une zone A, la masse portée par A est le produit entre cette valeur de densité et le volume de la zone. Il est donc essentiel de savoir estimer le volume des zones susceptibles d'accueillir de la matière, pour pouvoir estimer la masse correspondante. Définir une mesure consiste précisément à concevoir une procédure pour associer à une zone son volume. Même s'il n'est pas dans les usages d'affecter une unité physique aux grandeurs mathématiques, on pourra concevoir cette mesure comme s'exprimant en unité de volume (ou d'aire s'il s'agit de l'espace bi-dimensionnel, ou de longueur s'il s'agit d'un espace à une dimension). Il s'agit d'une donnée *statique* associée à l'espace considéré. Dans le cas de l'espace euclidien, ce volume est canoniquement défini dans le cas de formes simples : longueur d'un intervalle, aire d'un rectangle, volume d'un parallélogramme. La notion d'intégration d'une fonction constante sur de tels ensembles est basée sur le simple produit de la valeur à intégrer par le volume. Si, suivant l'intuition associée à la notion de volume, on décrète que le volume de la réunion de deux zones disjointes est la somme des volumes des zones élémentaires, on peut estimer le volume de toutes les zones qui peuvent se construire comme réunion disjointe finie de ces formes simples. Définir le volume de n'importe quel ensemble est plus délicat et même, d'une certaine manière, impossible, comme nous le verrons. La construction de la mesure de Lebesgue, qui est un point essentiel des sections qui suivent, permettra de définir un tel volume pour une classe très générale de zones de l'espace euclidien, et permettra de construire un cadre définissant la notion d'intégrales pour des fonctions très générales.

Les mesures ont également vocation à représenter des quantités absolues de matière (fluides, matériau solide, cellules, individus, ...), distributions d'une certaine substance susceptible d'évoluer en temps, d'être transportée, supprimée, développée. L'objet mathématique associé est le même, mais la nature de la réalité qu'il a vocation à représenter est différente. Il sera ici naturel de penser la mesure associée comme exprimée en kg, en moles ...., qui mesurent des quantités de matières associées à des principes de conservation.

Nous proposons dans les paragraphes qui suivent quelques exemples de situations réelles qui illustrent les deux types de mesures évoqués ci-dessus et les liens qu'elles entretiennent : mesures de type *volume*, qui formalisent la capacité de parties de l'espace sous-jacent à accueillir de la matière, et mesures de type *masse*, qui représentent des quantités de choses réelles. Nous nous restreignons dans ces exemples à des ensembles finis, de telle sorte que les objets mathématiques sont très simples à définir. Nous évoquerons dans la suite de cette introduction les difficultés posées par la construction de telles mesures pour des ensembles infinis.

Superficies, densités, et nombre d'habitants.
Considérons l'ensemble $X = [\![1, N]\!]$ des grandes villes françaises, numérotées de 1 à N. On note $\mu_i > 0$ la superficie de la ville i. À la collection des μ_i est naturellement associée une application μ de l'ensemble $\mathcal{P}(X)$ des parties de X dans $\mathbb{R}_+$:

$$\mu \ : \ A \in \mathcal{P}(X) \longmapsto \mu(A) = \sum_{i \in A} \mu_i \in \mathbb{R}_+. \tag{B.1.1}$$

Cette application est *additive* au sens où, si A et B sont disjoints, $\mu(A \cup B) = \mu(A) + \mu(B)$. Pour reprendre une terminologie physique, cette application définit une variable *extensive*.

Il s'agit d'une mesure au sens volumique évoqué ci-dessus, qui structure l'ensemble des villes en termes de capacité d'accueil. Notons maintenant ρ_i la densité d'habitants dans la ville i. Il s'agit là d'une variable *intensive*[1]. Le produit $m_i = \rho_i \mu_i$ est le nombre d'habitants dans cette ville i. On peut, comme précédemment pour les μ_i, associer à la collection des villes une application m de $\mathcal{P}(X)$ dans $\mathbb{R}$, additive par construction. Cette application est une nouvelle mesure sur X, de type "masse". Le nombre total d'habitants dans le sous-ensemble $A \subset X$ de villes peut s'écrire comme un produit de dualité[2] noté $\langle \rho \, , \mu \rangle_A$ entre les collections de superficies et de densités

$$m_A = \langle \rho \, , \mu \rangle_A = \sum_{i \in A} \rho_i \mu_i.$$

Il s'agit là de la version discrète d'une intégrale, construite par mise en dualité d'une mesure volume (μ, version discrète de la mesure de Lebesgue construite plus loin) et d'une variable intensive (densité ρ, qui joue le rôle d'une fonction à intégrer sur un domaine). La mesure masse m est la variable sommable, produit de la variable extensive μ et la variable intensive ρ.

1. La densité associée à la réunion de deux villes de même densité ρ est ρ, et pas 2ρ.
2. Un produit de dualité entre deux espaces vectoriels E et F de même dimension est simplement une application bilinéaire de $E \times F$ dans $\mathbb{R}$. On dit que cette application met les espaces en dualité. L'exemple le plus simple est le cas d'un espace euclidien, qui est en dualité avec lui même par le biais de son produit scalaire.

On peut aussi définir, dans le cas présent d'une collection finie de villes, des mesures qui correspondent à des probabilités. Prenons l'exemple d'un crime commis à Paris à l'heure H d'un jour J. Vingt-quatre heures après, l'assassin court toujours, et les enquêteurs cherchent à estimer dans quelle ville il pourrait être. L'état de leur opinion concernant la position du fugitif peut être encodé par une mesure $m = (m_i)$. Si l'on sait qu'il ne dispose pas de véhicule et que l'on considère que prendre le train était risqué pour lui, on considèrera que la probabilité associée à Paris est de 0.75. Si l'on sait qu'il a des contacts à Lyon, on évaluera à 0.15 la probabilité qu'il y soit, le complément étant distribué sur le reste du pays en fonction des informations que l'on peut avoir. On a ici l'exemple typique d'une mesure (ici de probabilité, c'est à dire normalisée à 1) qui évolue au cours du temps, en fonction des informations reçues.

L'intérêt d'introduire la notion de *mesure* pour les exemples ci-dessus, alors que les objets manipulés se ramènent à des tableaux de nombres réels, n'est pas immédiat. Nous verrons qu'il est néanmoins fécond de considérer par exemple la collection $\mu = (\mu_i)$ des superficies comme une application qui, à un ensemble de villes $I \subset [\![1, N]\!]$, associe la population totale des villes concernées, selon l'expression (B.1.1). Cette application attribue 0 à l'ensemble vide, et vérifie par construction la règle de sommation suivante : l'image de la réunion de deux ensembles disjoints est la somme des images (on dira que l'application est *additive*), ce qui peut s'écrire

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

Nous définirons une mesure comme une application qui à une partie associe un réel positif, et qui vérifie des conditions du type de celles qui précèdent.

Aérosols.
On considère maintenant une collection de N micro-gouttelettes sphériques flottant dans l'air. Si l'on note μ_i le volume de la gouttelette i, on peut définir une application de l'ensemble des parties de $X = [\![1, N]\!]$ dans $\mathbb{R}^+$ associant à une sous-collection de gouttelettes son volume total. Si l'on note ρ la densité du fluide considéré, on peut associer à la collection une nouvelle mesure, de type masse, simplement définie par ses valeurs en chaque entité, $m_i = \rho\mu_i$, la mesure associée, définie comme application de $\mathcal{P}(X)$ dans $\mathbb{R}_+$, s'en déduisant simplement par additivité. On a ainsi construit une nouvelle mesure exprimant une variable extensive, construite comme produit d'une première mesure volume avec une variable intensive. On peut dans ce contexte continuer l'empilement des mesures en considérant que chaque particule est animée d'une vitesse u_i. Cette collection de vitesses peut être vue comme une fonction sur X. Cette variable vectorielle intensive peut être adossée avec la mesure m (extensive) pour former une nouvelle variable extensive (la quantité de mouvement), construite selon $p_i = m_i u_i$. Il s'agit de la version discrète de ce que l'on appellera une *mesure vectorielle*. C'est une variable extensive (la quantité de mouvement d'un système est la somme des quantités de mouvement de ses constituants). Dans ce contexte, on dira que la vitesse est mesurable m-presque partout. Ici, l'ensemble étant fini, cela signifie simplement que cela n'a pas de sens de définir la vitesse d'un objet qui n'a pas de masse, puisque cette vitesse sans masse ne pourrait intervenir d'aucune manière dans un modèle mécanique cohérent.

On remarquera que la variable intensive vitesse peut être intégrée selon cette nouvelle mesure vectorielle, pour former une quantité scalaire qui représente l'énergie cinétique

$$E_A = \langle u \, , p \rangle_A = \frac{1}{2} \sum_{i \in A} m_i u_i^2.$$

Vers l'infini : le cas de l'intervalle $]0, 1[$
Les cadres présentés ci-dessus peuvent être étendu assez naturellement à des ensembles dénombrables, on remplace alors les sommes finies par des sommes infinies, des séries, de nombres positifs, en acceptant éventuellement que la série puisse prendre la valeur $+\infty$. On remarquera néanmoins que, s'il est possible d'affecter une masse à chaque point d'une collection dénombrable de façon à ce que la masse totale soit *finie*, la distribution est forcément inégalitaire, ou identiquement nulle. En effet, si chaque point de notre ensemble dénombrable a une masse m, on a l'alternative suivante : si $m > 0$ la masse totale est infinie, et si $m = 0$ la masse totale est nulle. Une version temporelle de cet énoncé, qui évoque le paradoxe d'Achille et de la tortue, pourrait être : disposant d'un temps fini, on peut faire une infinité de choses qui chacune prend un certain temps, mais c'est impossible en attribuant un temps identique à chacune des tâches. On retrouvera cet argument très simple au cœur de la construction d'un des ensembles pathologiques évoqués ci-après.

Les véritables difficultés commencent lorsque l'on s'intéresse à des ensembles qui ont ce que l'on appelle la *puissance du continu*, comme la droite réelle, ou l'espace physique $\mathbb{R}^3$. Considérons pour fixer les idées le cas de l'intervalle réel $X =]0,1[$. On cherche à définir sur cet ensemble une notion de volume (il s'agit plutôt en l'occurrence d'une notion de longueur, que nous verrons ici comme un volume monodimensionnel). Plus précisément, on cherche à construire une *mesure*, c'est-à-dire une application μ qui à une partie A de $]0,1[$ associe un nombre réel positif ou nul, et qui généralise à des ensembles quelconques la notion de longueur. On souhaite donc en particulier que $\mu(]a,b[) = b - a$. Le caractère extensif de la notion de longueur impose une propriété d'additivité. On demande donc que la mesure d'une union d'ensembles disjoints soit égale à la somme des mesures des ensembles. Comme nous le verrons plus loin, il est nécessaire pour aboutir à une notion "utilisable" que cette propriété s'étende à des collections dénombrables de parties, on parlera de *$\sigma-$additivité*. L'intervalle fermé $[a,b]$ étant l'intersection des intervalles$]a-1/n, b+1/n[$, sa longueur est la même que celle de l'intervalle ouvert. On en déduit que la mesure des singletons (comme les extrémités de l'intervalle) est nulle. On peut étendre immédiatement cette mesure à des réunions dénombrables d'intervalles, mais on se heurte ensuite à un mur. Pour des raisons assez profondes qui tiennent à la nature même de la droite réelle, et malgré l'apparente simplicité du problème, il est *impossible* de définir une telle application, qui affecterait aux intervalles leurs longueurs, qui serait σ-additive (manière distinguée de dire que cela correspond à une variable extensive), qui affecterait à une partie quelconque de l'intervalle [3] $]0,1[$ ce qu'il conviendrait alors d'appeler sa longueur. On peut contourner le problème par le haut en suivant un principe inhérent à la notion intuitive de volume : si un ensemble est inclus dans un autre, ce dernier a un plus gros volume. Si l'on se donne $A \subset]0,1[$, on peut considérer l'ensemble des collections dénombrables d'intervalles (on s'affranchit du caractère disjoint des collections) qui recouvrent A. Si l'on était capable de définir une mesure pour A, cette mesure serait inférieure où égale à la mesure de toute collection qui recouvre A, qui est elle-même inférieure à la somme des longueurs des intervalles. Il est ainsi naturel de considérer la quantité $\mu^\star(A)$ définie comme l'infimum de la somme des longueurs des intervalles, infimum sur l'ensemble des collections qui recouvrent A. On appellera cette quantité la *mesure extérieure* de Lebesgue de A. Cette démarche conduit néanmoins à un problème : il apparaît qu'il existe des parties de X qui vérifient des propriétés que nous qualifierons de *bizarres*. Il existe en effet des ensembles B, dont le complémentaire dans X est noté B^c, qui conduisent à une violation de la propriété d'additivité que l'on souhaite voir vérifier par la mesure. Plus précisément, il existe certaines parties B telles que, pour certaines parties A, l'identité

$$\mu^\star(A) = \mu^\star(A \cap B) + \mu^\star(A \cap B^c)$$

n'est *pas vérifiée*. Plus précisément $\mu^\star(A)$ est strictement inférieur à la somme des mesures des parties disjointes $A \cap B$ et $A \cap B^c$ qui le constituent. Le mathématicien se retrouve dans la position d'un arpenteur étudiant une région A, composée exclusivement de 2 propriétés A_1 et A_2 sans recouvrement, imbriquées l'une dans l'autre de façon extrêmement complexe, et telle que l'aire estimée de A selon la méthode évoquée ci-dessus est *strictement inférieure* à la somme des aires de A_1 et A_2.

Il n'existe pas de manière complètement satisfaisante de régler ce nouveau problème. La démarche conduisant à des "monstres", on choisit simplement de les exclure de l'approche, et de se concentrer sur les parties B pour lesquelles l'identité ci-dessus est vérifiée pour toute partie A (parties appelées *mesurables*, et dont la collection s'appelle une *tribu* comme on le verra) pour définir une mesure. Cette mesure, qui est la restriction de la mesure extérieure ci-dessus à la collection $\mathcal{A}$ des ensembles mesurables, vérifie alors de bonnes propriétés, au prix de *l'exclusion de certains ensembles pathologiques*, qu'il est d'ailleurs impossible de décrire explicitement [4]. Une fois cette construction réalisée, la définition de la notion d'intégrale s'ensuit naturellement. L'intégrale d'une fonction constante égale à ρ (que l'on peut voir ici comme une densité) sur une partie A est simplement le produit $\rho \times \mu(A)$, qui est alors la masse de la matière contenue dans A. On peut étendre facilement cette définition aux fonctions qui prennent un nombre fini de valeurs (fonctions dites *simples*, ou *étagées* dans le cadre de la théorie de la mesure) sur des parties mesurables, en sommant simplement les différentes contributions, comme pour calculer la masse d'un objet composite à partir des densités de ses constituants, et des volumes des différentes zones qu'ils occupent. On peut alors étendre cette notion d'intégrale à une classe très générale de fonctions (nous ne considérons pour l'instant que des fonctions positives),

3. On peut aussi formuler ce problème dans le plan $\mathbb{R}^2$ en considérant à la place des intervalles des rectangles, dont on sait calculer l'aire, ou dans l'espace physique $\mathbb{R}^3$ en considérant des pavés (i.e. parallélépipèdes), dont on sait calculer le volume.

4. La construction de ces contre-exemples nécessite l'*axiome du choix*, ce qui confère un caractère très abstrait à ces contre-exemples.

appelées *mesurables*, en définissant l'intégrale comme le supremum des intégrales des fonctions étagées qui sont partout inférieures ou égales à la fonction considérée.

Cadre général.
La démarche décrite précédemment s'inscrit dans un cadre général qui dépasse le cas particulier de la droite des réels, et qui constitue les bases de la théorie des probabilités. Les sections qui suivent présentent ce cadre abstrait, et en parallèle la construction progressive de l'intégrale de Lebesgue. Le point de départ est la notion de tribu déjà évoquée ci-dessous : une tribu sera définie comme une famille de parties d'un ensemble X qui vérifient un certain nombre de propriétés, essentiellement de stabilité (par complémentarité et par union dénombrable). On définira ensuite la notion de mesure sur une tribu, qui est une application à valeurs dans $\mathbb{R}^+$, et a vocation à affecter à une partie de X son volume. On demandera assez naturellement à ce que "rien" ne prenne pas de place ($\mu(\emptyset) = 0$), et l'on exigera par ailleurs, pour respecter le caractère extensif de la notion que l'on souhaite définir, une propriété d'additivité : la mesure d'une union disjointe (dénombrable) de parties est la somme des mesures de ces parties. On donnera un sens très général à la notion de mesure extérieure, déjà évoquée plus haut dans le cas de l'intervalle $]0,1[$, définie sur l'ensemble des parties d'un ensemble, en relaxant la propriété d'additivité (remplacée par une propriété de sous-additivité), et en imposant la monotonie (qui n'est plus garantie sinon, du fait que l'on a relaxé l'additivité). On qualifie alors de mesurable une partie qui vérifie la propriété d'additivité évoquée précédemment, et l'on peut montrer une propriété très générale : la familles des parties mesurables est une tribu, et la mesure extérieure restreinte à cette tribu est une *mesure*. C'est ce résultat qui permettra de définir la mesure de Lebesgue à partir de la mesure de Lebesgue extérieure introduite dans le paragraphe précédent.

Terminons cette longue introduction par quelques mots sur la théorie des probabilités, qui constitue une motivation important à l'étude détaillée des notions de tribu, classe monotone, mesure, ..., même si elle n'est pas centrale dans ce cours. Dans ce contexte l'ensemble X est vu comme un ensemble d'*éventualités* (on parle de l'univers des possibles, issues possibles d'une expérience). Définir une tribu consiste à définir un sous-ensemble de parties que l'on souhaite considérer comme des *événements*, c'est-à-dire comme des propriétés vérifiées par le résultat de l'expérience, exprimées au travers de l'appartenance à une des sous-parties de la tribu. Par exemple si l'on sait qu'une météorite est tombée en Europe, on concevra X comme l'ensemble des positions géographiques de cette zone (que l'on peut identifier à une carte au sens usuel du terme). On peut imaginer une tribu comme un ensemble d'assertions potentiellement pertinentes. Par exemple : 'La météorite est tombée en Alsace', 'La météorité n'est pas tombée en Alsace', 'La météorite est tombée à l'ouest de Berlin ou au nord d'Helsinki', 'La météorite est tombée en zone urbaine', 'La métérorite n'est pas tombée en Europe' ....
Noter que l'on peut choisir de structurer l'ensemble des assertions potentiellement pertinentes de façon plus ou moins détaillée (ce qui correspondra à la notion de tribu plus ou moins *fine*). Si l'on ne s'intéresse qu'au pays atteint, on se limitera à des assertions du type : 'La météorite est tombée dans un pays d'Europe du nord', ce qui correspond à une tribu finie (un élément de la tribu est un sous-ensemble de pays, éventuellement vide). A l'autre extrême, on peut envisager l'ensemble des assertions possibles correspondant à une localisation exacte du point d'impact. Comme nous l'avons évoqué plus haut, c'est cette volonté de structurer un espace ayant la puissance du continu qui soulève des difficultés profondes, qui sont abordées dans les sections qui suivent. Dans le contexte des probabilités, la définition d'une mesure de masse totale 1 sur la tribu choisie permettra d'affecter à chacun de ses membres A un nombre quantifiant la probabilité que l'assertion $x \in A$ soit vérifiée. L'événement X, de probabilité 1, est certain, et l'événement vide, de probabilité nulle, est impossible (il correspond à la dernière assertion ci-dessus). Les opérations sur les parties - événements sont associées à des opérations logiques : $A \cup B$ correspond à l'événement $x \in A$ *ou* $x \in B$, $A \cap B$ correspond à l'événement $x \in A$ *et* $x \in B$.

B.2 Tribus, espaces mesurables

B.2.1 Tribus

Définition B.2.1. (Tribu / σ-algèbre (•))
Soit X un ensemble. On appelle tribu (ou σ-algèbre) sur X un ensemble $\mathcal{A}$ de parties de X qui vérifie les propriétés suivantes :

(i) $\emptyset \in \mathcal{A}$,
(ii) $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$,
(iii) Si (A_n) est une collection dénombrable d'éléments de $\mathcal{A}$, alors

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}.$$

Si la propriété (*iii*) est restreinte aux collections finies, on dira que $\mathcal{A}$ est une *algèbre*. On appelle $(X, \mathcal{A})$ (ensemble muni de sa tribu) un *espace mesurable*.

Définition B.2.2. (Finesse)
On considère deux tribus $\mathcal{A}$ et $\mathcal{A}'$ sur un même ensemble X. On dit que la tribu $\mathcal{A}'$ est *plus fine* que la tribu $\mathcal{A}$ si $\mathcal{A} \subset \mathcal{A}'$.

Proposition B.2.3. (•) Toute tribu est stable par intersection dénombrable.

Démonstration. Soit (A_n) une collection dénombrable d'éléments d'une tribu $\mathcal{A}$. On a

$$\bigcap_{n \in \mathbb{N}} A_n = \left(\bigcup_{n \in \mathbb{N}} A_n^c \right)^c,$$

qui appartient à $\mathcal{A}$ par complémentarité et union dénombrable. □

Exemples B.2.1. Nous donnons ici quelques exemples de tribus associées à un ensemble X quelconque.

1. (Tribu discrète). Pour tout ensemble X, l'ensemble $\mathcal{P}(X)$ des parties de X est une tribu. Comme on le verra, dès que X est non dénombrable, par exemple sur $\mathbb{R}$, cette tribu est essentiellement *inutilisable*, car il est impossible de lui associer une mesure non triviale qui possède de bonnes propriétés.
2. (Tribu grossière). Pour tout ensemble X, $\{\emptyset, X\}$ est une tribu à 2 membres.

Exercice B.2.1. (Tribu trace)
Soit $\mathcal{A}$ une tribu sur un ensemble X, et F une partie de X. Montrer que

$$\mathcal{A}_F = \{A \cap F\,,\ A \in \mathcal{A}\}$$

est une tribu sur F (appelée tribu trace de $\mathcal{A}$ sur F).

Proposition B.2.4. (Une intersection de tribus est une tribu (•))
Soit X un ensemble. Toute intersection de tribus sur X est une tribu.

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

Comme pour toute propriété stable par intersection [5], on peut définir la notion de plus petite tribu contenant une collection de parties de X.

Définition B.2.5. (Tribu engendrée (•))
Soit $\mathcal{C} \subset \mathcal{P}(X)$ une collection de parties de X. On appelle tribu engendrée par $\mathcal{C}$, et l'on note $\sigma(\mathcal{C})$, la plus petite tribu contenant $\mathcal{C}$. Elle est définie comme l'intersection de toutes les tribus contenant $\mathcal{C}$.

Exercice B.2.2. Soit X un ensemble et A une partie de X. Montrer que la tribu engendrée par $\{A\}$ est de cardinal 2 ou 4.

5. On pourra penser par exemple au fait, pour une partie de l'espace $\mathbb{R}^d$, d'être un sous-espace vectoriel, un sous-espace affine, d'être convexe, d'être fermée, d'être conique, ... On parle en général d'*enveloppe* linéaire, affine, convexe, fermée, conique. Le terme d'enveloppe n'est pas utilisé dans le cas des tribus, mais le principe de construction est le même.

Définition B.2.6. (Tribu borélienne sur $\mathbb{R}(\bullet)$)
On appelle tribu borélienne sur $\mathbb{R}$ la tribu engendrée par les ouverts de $\mathbb{R}$. On la note $\mathcal{B}(\mathbb{R})$.

Proposition B.2.7. ($\bullet$) La tribu $\mathcal{B}(\mathbb{R})$ des boréliens de $\mathbb{R}$ est engendrée par les intervalles de la forme $]-\infty, a]$, avec $a \in \mathbb{R}$.

Démonstration. Notons en premier lieu que le fermé $]-\infty, a]$ est le complémentaire d'un ouvert, tous ces intervalles sont donc dans $\mathcal{B}(\{\mathbb{R}\})$, la tribu engendrée est donc contenue dans la tribu des boréliens. Pour montrer l'inclusion réciproque, tout ouvert de $\mathbb{R}$ étant réunion dénombrables d'intervalles ouverts (voir proposition I.3.12), il suffit de montrer que la tribu engendrée par les $]-\infty, a]$ contient les intervalles ouverts. Par complémentarité, cette tribu contient les intervalles du type $]a, +\infty[$. Par ailleurs, l'union des $]-\infty, b-1/n]$ est l'intervalle $]-\infty, b[$. La tribu contient donc (d'après la stabilité par intersection assurée par la proposition B.2.3), pour tous $a < b$, l'intervalle $]a, +\infty[\, \cap\,]-\infty, b[\, =\,]a, b[$ ce qui termine la démonstration. $\square$

Il sera utile lors de la construction de l'intégrale de considérer des fonctions réelles qui peuvent prendre des valeurs infinies ($+\infty$ ou $-\infty$), c'est-à-dire à valeurs dans la droite réelle achevée $\overline{\mathbb{R}}$ (voir définition I.3.13, page 18). Rappelons que les ouverts de cette droite réelle achevée sont les ensembles de type U, $U \cup]a, +\infty]$, $U \cup [-\infty, b[$, ou $U \cup]a, +\infty] \cup [-\infty, b[$, où U est un ouvert de $\mathbb{R}$ (voir proposition I.3.14).

Proposition B.2.8. ($\bullet$) La tribu $\mathcal{B}(\overline{\mathbb{R}})$ des boréliens de $\overline{\mathbb{R}}$ est engendrée par les intervalles de la forme $[-\infty, b]$.

Démonstration. Par union dénombrable, $\mathcal{B}(\overline{\mathbb{R}})$ contient les intervalles du type $[-\infty, b[$, donc les intervalles ouverts de $\mathbb{R}$ $]a, b[= [-\infty, b[\setminus[-\infty, a]$. Cette tribu contient également les $]b, +\infty] =]-\infty, b]^c$. On a montré que la tribu engendrée par les $[-\infty, b]$ contient les intervalles du type $[-\infty, a[$, $]a, c[$, et $]c, +\infty]$, elle contient donc tous les ouverts de $\overline{\mathbb{R}}$ par union dénombrable. $\square$

Tribus et applications

Nous terminons cette section par des premières propriétés impliquant des applications entre ensembles. dans ce qui suit f est une application de X dans X'. Si X' est muni d'une tribu $\mathcal{A}'$, on montre que l'image réciproque de $\mathcal{A}'$ est une tribu sur X. Si X est muni d'une tribu $\mathcal{A}$, on peut vérifier que l'image de $\mathcal{A}$ par f n'est *pas* en général une tribu sur X'. On définit ci-dessous une notion qui permet de pousser une tribu vers l'avant en utilisant la réciproque de f, il s'agit de la notion de *tribu image*, qui elle est bien une tribu sur l'espace d'arrivée. Cette notion s'étendra directement aux mesures, que l'on peut voir comme une distribution de masse sur un ensemble, si f est vu comme une application de transport, la mesure image correspond à la distribution des masses transportées.

Proposition B.2.9. (Image réciproque d'une tribu($\bullet$))
Soit f une application d'un ensemble X vers un ensemble X' muni d'une tribu $\mathcal{A}'$. L'image réciproque de $\mathcal{A}'$ par f, c'est-à-dire la famille $\mathcal{A}$ des parties A de X qui s'écrivent

$$A = f^{-1}(A') = \{x \in X\,,\ f(x) \in A'\}\,,$$

avec $A' \in \mathcal{A}'$, est une tribu sur X.

Démonstration. On a $\emptyset = f^{-1}(\emptyset)$. Par ailleurs, pour tout $A' \in \mathcal{A}'$,

$$f^{-1}(A')^c = \{x \in X\,,\ f(x) \notin A'\} = \{x \in X\,,\ f(x) \in (A')^c\} = f^{-1}((A')^c) \in \mathcal{A} \text{ car } (A')^c \in \mathcal{A}'$$

Enfin, pour toute collection (A'_n) de $\mathcal{A}'$, on a

$$\bigcup_{n\in\mathbb{N}} f^{-1}(A'_n) = f^{-1}\left(\bigcup_{n\in\mathbb{N}} A'_n\right) \text{ avec } \bigcup_{n\in\mathbb{N}} A'_n \in \mathcal{A}',$$

qui appartient bien à $\mathcal{A}$. $\square$

L'image directe d'une tribu par une application n'est en général pas une tribu, comme on peut s'en convaincre en considérant par exemple une application constante vers un ensemble de cardinal ≥ 2. On peut en revanche pousser en avant une tribu par une application pour obtenir une tribu, alors appelée *tribu image*, comme exprimé par la proposition suivante.

Proposition B.2.10. (Tribu image)
Soit f une application de X dans X', et $\mathcal{A}$ une tribu sur X. La collection de parties

$$\mathcal{A}' = f_\sharp \mathcal{A} = \left\{A' \subset X' \,,\; f^{-1}(A') \in \mathcal{A}\right\}$$

est une tribu sur X', appelée tribu-image de $\mathcal{A}$ par f.

Démonstration. On a $f^{-1}(\emptyset) = \emptyset \in \mathcal{A}$ et $f^{-1}(X') = X \in \mathcal{A}$. Par ailleurs, si $A' \in \mathcal{A}'$,

$$f^{-1}(A'^c) = \left(f^{-1}(A')\right)^c \in \mathcal{A},$$

d'où $A'^c \in \mathcal{A}'$. Enfin, pour toute famille (A'_n) de $\mathcal{A}'$,

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} (A'_n)\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(A'_n) \in \mathcal{A},$$

d'où l'on déduit que l'union est dans $\mathcal{A}'$. □

Exercice B.2.3. Soit $f \,: X \longrightarrow X'$ une application constante et $\mathcal{A}$ une tribu sur X. Identifier

$$f_\sharp \mathcal{A} = \left\{A' \subset X' \,,\; f^{-1}(A') \in \mathcal{A}\right\}.$$

Si maintenant $\mathcal{A}'$ est une tribu sur X', identifier $f^{-1}(\mathcal{A}')$.

B.2.2 Applications mesurables

Définition B.2.11. (Application mesurable (•))
Soit f une application d'un espace ensemble X vers un ensemble X'. On suppose X et X' munis de tribus $\mathcal{A}$ et $\mathcal{A}'$, respectivement. On dit que f est mesurable[6] de $(X, \mathcal{A})$ vers $(X', \mathcal{A}')$ si l'image réciproque de toute partie de $\mathcal{A}'$ est dans $\mathcal{A}$:

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

On notera que, par définition, une application f de X dans $(X', \mathcal{A}')$ est toujours mesurable, si l'on munit l'espace de départ de la tribu $f^{-1}(\mathcal{A}')$. Cette propriété est d'un intérêt limité du fait que la tribu sur l'ensemble de départ dépend de l'application.

Exercice B.2.4. Soit f une application de X dans X'.

Si l'on se donne une tribu $\mathcal{A}'$ sur X, montrer que $f^{-1}(\mathcal{A})$ est la plus petite tribu sur X telle que f soit mesurable, c'est-à-dire que, si f est $\mathcal{A} - \mathcal{A}'$ mesurable, alors $\mathcal{A}$ contient $f^{-1}(\mathcal{A})$.

Si l'on se donne maintenant une tribu $\mathcal{A}$ sur X, montrer que $f_\sharp \mathcal{A}$ est la plus grande tribu sur X' telle que f soit mesurable, c'est-à-dire que, si f est $\mathcal{A} - \mathcal{A}'$ mesurable, alors $\mathcal{A}'$ est contenue dans $f_\sharp(\mathcal{A})$.

Exercice B.2.5. Montrer qu'une application constante (qui envoie tous les éléments de l'espace de départ vers un même point de l'espace d'arrivée), est toujours mesurable.

Exercice B.2.6. Soit Id l'application identité de $(X, \mathcal{A})$ vers $(X, \mathcal{A}')$. A quelle condition cette application est-elle mesurable ?

6. On pourra écrire que f est $\mathcal{A}$ - $\mathcal{A}'$ mesurable, ou simplement mesurable s'il n'y a pas d'ambigüité sur les tribus qui structurent X et X'.

Proposition B.2.12. (Critère de mesurabilité d'une application (●●))
Soit f une application de X (muni d'une tribu $\mathcal{A}$) vers X' (muni d'une tribu $\mathcal{A}'$). On suppose que la tribu $\mathcal{A}'$ de l'espace d'arrivée est engendrée par $\mathcal{C}' \subset \mathcal{P}(X')$.

L'application f est mesurable si et seulement si $f^{-1}(C') \in \mathcal{A}$ pour tout $C' \in \mathcal{C}'$.

Démonstration. On introduit

$$\mathcal{B}' = \left\{B' \in \mathcal{A}' \,,\ f^{-1}(B') \in \mathcal{A}\right\} = f_\sharp \mathcal{A} \cap \mathcal{A}'.$$

Il s'agit d'une tribu comme intersection de tribus (la mesure image $f_\sharp \mathcal{A}$ est une tribu d'après la proposition B.2.10. Et cette tribu contient $\mathcal{C}'$ par hypothèse, elle contient donc la tribu engendrée par $\mathcal{C}'$, c'est à dire $\mathcal{A}'$. □

Exercice B.2.7. Soit $(X, \mathcal{A})$ et $(X', \mathcal{A}')$ deux espaces mesurables. Décrire l'ensemble des applications mesurables de X vers X', dans le cas où X est muni de la tribu grossière $\mathcal{A} = \{\emptyset, X\}$. Même question si X est muni de la tribu discrète $\mathcal{A} = \mathcal{P}(X)$. Que peut-on dire si X' est muni de la tribu grossière ? De la tribu discrète ?

B.2.3 Classes monotones

Définition B.2.13. (Classe monotone (●))
Soit X un ensemble. On appelle classe monotone (ou λ-système) sur X un ensemble $\mathcal{D}$ de parties de X qui vérifie les propriétés suivantes :

(i) $X \in \mathcal{D}$,
(ii) $A,\ B \in \mathcal{D}$, $A \subset B \Longrightarrow B \setminus A \in \mathcal{D}$,
(iii) si (A_n) est une suite croissante d'éléments de $\mathcal{D}$ (i.e. $A_n \subset A_{n+1}$ pour tout n), alors

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}.$$

Proposition B.2.14. Toute tribu sur X est une classe monotone.

Démonstration. Soit $\mathcal{A}$ une tribu sur X. On a par définition $X = \emptyset^c \in \mathcal{A}$. Pour tous A, B dans $\mathcal{D}$, avec $A \subset B$, on a $B \setminus A = B \cap A^c \in \mathcal{A}$. Enfin toute réunion dénombrable d'éléments de $\mathcal{A}$ est dans $\mathcal{A}$. □

Proposition B.2.15. Toute intersection de classes monotones est une classe monotone.

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

Définition B.2.16. (Classe monotone engendrée par un ensemble de parties (●))
La propriété étant stable par intersection, et l'ensemble $\mathcal{P}(X)$ de toutes les parties étant une classe monotone, on peut définir la notion de classe monotone engendrée par un ensemble $\mathcal{C}$ de parties, définie comme l'intersection de toutes les classes monotones qui contiennent $\mathcal{C}$.

Définition B.2.17. (π-système (●))
On appelle π–système sur un ensemble X un sous-ensemble $\mathcal{C}$ non vide de parties de X stable par intersection finie :

$$A \in \mathcal{C}\,,\ B \in \mathcal{C} \Longrightarrow A \cap B \in \mathcal{C}.$$

Remarque B.2.18. Certains auteurs ajoutent la condition qu'un π–système doit contenir X.

Exemples B.2.2. Les ensembles de parties suivants sont des π–systèmes :

1. L'ensemble $\mathcal{P}(X)$ des parties de X.

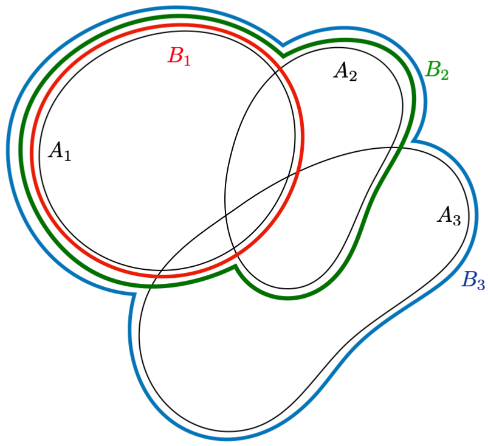

FIGURE B.2.1 – Construction de la suite croissante $B1$, B_2, ...

2. L'ensemble des ouverts d'un espace topologique.
3. L'ensemble des fermés d'un espace topologique.
4. La famille $\{]-\infty, c]\ ,\ c \in \mathbb{R}\}$ de parties de $\mathbb{R}$.
5. L'ensemble d'intervalles ouverts $\{]a, b[\ ,\ -\infty \leq a_i \leq b_i \leq +\infty\}$.
6. Les rectangles ouverts de $\mathbb{R}^2$, de type $]a_1, a_2[\times]b_1, b_2[$, avec $-\infty \leq a_i \leq b_i \leq +\infty$, $i = 1, 2$.
7. Les rectangles fermés.
8. Les rectangles semi-ouverts / semi-fermés de $\mathbb{R}^2$, de type $[a_1, a_2[\times[b_1, b_2[$, avec $-\infty \leq a_i \leq b_i \leq +\infty$, $i = 1, 2$.
9. Tout ensemble de singletons : pour $A \subset X$, $\{\{x\}\ ,\ x \in A\}$ auquel on rajoute la partie vide est un π-système.

Proposition B.2.19. ($\bullet$) Soit $\mathcal{D}$ une classe monotone stable par intersection finie (i.e. $\mathcal{D}$ est aussi un $\pi-$ système). Alors $\mathcal{D}$ est une tribu.

Démonstration. On a $\emptyset = X \setminus X \in \mathcal{D}$ et, pour tout $A \in \mathcal{D}$, $A^c = X \setminus A \in \mathcal{D}$. Considérons maintenant une famille (A_n) d'éléments de $\mathcal{D}$. Il s'agit de montrer que l'union des A_n est dans $\mathcal{D}$. On pose $B_0 = A_0$ et, considérant que B_n est construit, et qu'il appartient à $\mathcal{D}$, on définit B_{n+1} comme $B_n \cup A_{n+1} = A_0 \cup A_1 \cup \ldots \cup A_{n+1}$ (voir figure B.2.1). Montrons par récurrence que $B_{n+1} \in \mathcal{D}$. Supposons B_n dans $\mathcal{D}$. On a

$$B_{n+1} = B_n \cup A_{n+1} = \left(\underbrace{B_n^c \cap A_{n+1}^c}_{\in \mathcal{D}} \right)^c$$

qui est dans $\mathcal{D}$ comme complémentaire d'un élément de $\mathcal{D}$.

La suite (B_n) d'éléments de $\mathcal{D}$, est croissante par construction, d'où $\cup B_n \in \mathcal{D}$, et cette union s'identifie par construction à l'union des A_n. La famille $\mathcal{D}$ est donc bien une tribu. □

Proposition B.2.20. (Lemme de classe monotone)
($\bullet\bullet$) Soit $\mathcal{C}$ un $\pi-$ système sur l'ensemble X. La classe monotone $\mathcal{D}$ engendrée par $\mathcal{C}$ est égale à la tribu $\mathcal{A} = \sigma(\mathcal{C})$ engendrée par $\mathcal{C}$.

Démonstration. La tribu $\mathcal{A}$ contient $\mathcal{C}$, et c'est une classe monotone (proposition B.2.14), elle contient donc $\mathcal{D}$ qui est la plus petite classe monotone contenant $\mathcal{C}$. Pour montrer l'inclusion inverse, nous allons montrer que $\mathcal{D}$ est une tribu (qui alors contient nécessairement $\mathcal{A}$, qui est la plus petite). D'après la proposition B.2.19, il suffit de montrer que $\mathcal{D}$ est un π − système, i.e. qu'elle est stable par intersection finie. On considère dans un premier temps

$$\mathcal{D}' = \{A \in \mathcal{D}\,,\ A \cap C \in \mathcal{D} \quad \forall C \in \mathcal{C}\}\,.$$

Nous allons montrer que $\mathcal{D}'$ est une classe monotone qui contient $\mathcal{C}$. Comme elle est incluse dans $\mathcal{D}$ par définition, nous en déduirons qu'elle s'identifie à $\mathcal{D}$.

Cet ensemble de parties contient de façon évidente le π − système $\mathcal{C}$. Par ailleurs, pour tous A, $B \in \mathcal{D}'$, $A \subset B$, on a

$$(B \setminus A) \cap C = (\underbrace{B \cap C}_{\in \mathcal{D}}) \setminus (\underbrace{A \cap C}_{\in \mathcal{D}}) \in \mathcal{D}.$$

Montrons que $\mathcal{D}'$ est également stable par union croissante. Pour toute suite croissante (A_n) dans $\mathcal{D}'$, on a

$$\left(\bigcup A_n\right) \cap C = \bigcup \left(\underbrace{A_n \cap C}_{\in \mathcal{D}}\right) \in \mathcal{D}.$$

L'ensemble $\mathcal{D}'$ est donc une classe monotone, qui contient $\mathcal{C}$, et qui est contenue dans la classe monotone $\mathcal{D}$ engendrée par $\mathcal{C}$, elle s'identifie donc à $\mathcal{D}$. On a ainsi montré que, pour tout $A \in \mathcal{D}$, $C \in \mathcal{C}$, $A \cap C \in \mathcal{D}$. Il reste à montrer la stabilité par intersection finie, et pas seulement avec les éléments de $\mathcal{C}$. Pour cela on introduit

$$\mathcal{D}'' = \{A \in \mathcal{D}\,,\ A \cap A' \in \mathcal{D} \quad \forall A' \in \mathcal{D}\}\,.$$

Cet ensemble contient $\mathcal{C}$ comme on vient de le montrer, et c'est une classe monotone (la démonstration est la même que précédemment) contenue dans $\mathcal{D}$, elle s'identifie donc à $\mathcal{D}$. On a ainsi montré que $\mathcal{D}$ est un π − système. Comme c'est aussi une classe monotone, c'est une tribu d'après la proposition B.2.14, qui contient $\mathcal{C}$, elle contient donc $\mathcal{A} = \sigma(\mathcal{C})$, ce qui termine la preuve. □

Ce lemme est très utile pour montrer que deux mesures (voir section B.3 ci-après) définies sur une même tribu sont égales. Si cette tribu commune est engendrée par un π − système, il suffira de montrer que les deux mesures s'identifient sur ce π − système (voir proposition B.3.12).

B.3 Mesures

Définition B.3.1. (•) Soit X un ensemble, et $\mathcal{A}$ une tribu sur X. On appelle mesure une application de $\mathcal{A}$ dans $[0, +\infty]$ telle que

(i) $\mu(\emptyset) = 0$,

(ii) si (A_n) est une collection dénombrable d'éléments de $\mathcal{A}$ deux à deux disjoints, alors

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mu\left(A_n\right).$$

Le triplet $(X, \mathcal{A}, \mu)$ (ensemble X muni d'une tribu $\mathcal{A}$ et d'une mesure associée) est appelé *espace mesuré.*
On dit que la mesure est *finie* si $\mu(X) < +\infty$.
On dit que la mesure est $\sigma-$ *finie* si X est réunion dénombrable d'éléments de $\mathcal{A}$ de mesure finie.

Remarque B.3.2. On notera que le (i) de la définition n'est pas une équivalence, comme l'axiome de séparation pour une distance : la mesure d'un ensemble non vide peut être nulle. On verra en particulier que les singletons pour la mesure de Lebesgue sont de mesure nulle, ainsi que d'autres ensembles a priori beaucoup plus “gros”, par exemple l'ensemble triadique de Cantor, qui est non dénombrable (voir exercice B.6.6, page 175). Noter par ailleurs que la mesure identiquement nulle est bien une mesure.

Exemples B.3.1. (•) Nous donnons ici quelques exemples de mesures associées à un ensemble X quelconque.

1. (Mesure de comptage). Soit X un ensemble et $\mathcal{A}$ une tribu sur X. On définit $\mu(A)$ comme le cardinal de A.
2. (Masse ponctuelle) Soit X un ensemble, $\mathcal{A}$ une tribu sur X, et $x \in X$. L'application δ_x qui à $A \in \mathcal{A}$ associe 1 si $x \in A$, 0 sinon, est une mesure appelée masse ponctuelle en x.
3. Toute combinaison positive de masses ponctuelles

$$\mu = \sum_{i=1}^{n} \alpha_i \delta_{x_i} \, , \;\; \alpha_i \geq 0 \quad \forall i = 1, \ldots, n$$

est également une mesure.
4. (Mesure grossière) Soit X un ensemble de cardinal ≥ 2, et $\mathcal{A}$ une tribu sur X. On pose $\mu(\emptyset) = 0$ et $\mu(A) = +\infty$ dès que $A \neq \emptyset$. Il s'agit d'une manière quelque peu grossière (d'où le nom), binaire, d'appréhender le monde, en distinguant deux types d'ensembles (disons des zones de l'espace physique pour fixer les idées) : l'ensemble vide, de volume nul, et tout ensemble non vide, auquel on attribuerait par convention une quantité infinie. Cette mesure extérieure ne distingue en quelque sorte que le "rien" et le "quelque chose". On peut associer à cette mesure extérieure une addition [7] primitive basée sur :

rien + rien = rien, rien + quelque chose = quelque chose, et
quelque chose + quelque chose = quelque chose.

Proposition B.3.3. (Monotonie (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. La mesure μ est *monotone*, c'est-à-dire que si $A \in \mathcal{A}$ est inclus dans $B \in \mathcal{A}$, alors la mesure de A est plus petite que celle de B :

$$\forall A \, , \; B \in \mathcal{A} \, , \; A \subset B \Longrightarrow \mu(A) \leq \mu(B).$$

Si la mesure de A est finie, on a de plus $\mu(B \setminus A) = \mu(B) - \mu(A)$.

Démonstration. On écrit simplement B comme réunion disjointe $B = A \cup (B \setminus A)$, qui implique, d'après la définition,

$$\mu(B) = \mu(A) + \mu(B \setminus A),$$

d'où les propriétés annoncées. □

Proposition B.3.4. Soit μ une mesure $\sigma -$ finie. Il existe une partition (B_n) de X constituée d'ensembles mesurables et de mesures finies.

Démonstration. Par définition X s'écrit comme réunion d'ensembles A_n de mesures finies (non nécessairement disjoints). On prend $B_0 = A_0$, et l'on définit B_n comme $A_n \setminus (A_{n-1} \cup \ldots \cup A_0)$. On a

$$B_n = A_n \setminus \bigcup_{k=0}^{n-1} A_k \subset A_n,$$

d'où

$$\mu(B_n) \leq \mu(A_n) < +\infty,$$

et, par construction, (B_n) est une partition [8] de X. □

7. Il s'agit bien d'une loi de composition interne sur l'ensemble à deux éléments {'rien', 'quelque chose'}, que l'on peut obtenir en quotientant X par la relation d'équivalence basée sur la notion d'être vide ou pas. Pour les lecteurs sensibles au plaisir de désigner des choses simples par des termes abscons, rajoutons que cette loi est associative, commutative, possède un élément neutre ('rien'), et munit donc notre petit univers à deux éléments d'une structure de *magma associatif unifère abélien*, ou plus sobrement *monoïde abélien*.

8. A strictement parler, une partition est censée ne contenir que des parties non vides, on se ramène à cette situation en ôtant de la liste les indices n tels que $B_n = \emptyset$.

Proposition B.3.5. (Sous-additivité (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et (A_n) une suite dans $\mathcal{A}$. On a

$$\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) \leq \sum_{n=0}^{+\infty} \mu\left(A_n\right).$$

Démonstration. La démonstration est basée, comme dans la démonstration de la proposition B.3.4 ci-dessus, sur la construction d'une suite (B_n) d'éléments disjoints de $\mathcal{A}$, telle que l'union des n premiers termes s'identifie à l'union des n premiers A_k. On pose $B_0 = A_0$ et l'on construit par récurrence

$$B_n = A_n \setminus \left(\bigcup_{k=0}^{n-1} A_k\right) = A_n \cap A_1^c \cap \ldots A_{n-1}^c,$$

qui appartient bien à la tribu $\mathcal{A}$, et qui est tel que $\mu(B_n) \leq \mu(A_n)$. On a donc

$$\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) = \mu\left(\bigcup_{n=0}^{+\infty} B_n\right) = \sum_{n=0}^{+\infty} \mu\left(B_n\right) \leq \sum_{n=0}^{+\infty} \mu\left(A_n\right),$$

qui est l'inégalité annoncée. □

Proposition B.3.6. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et (A_n) une suite de parties de $\mathcal{A}$, croissante pour l'inclusion. On a alors

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mu\left(A_n\right) \in [0, +\infty].$$

Démonstration. On construit la même suite B_n en posant $B_0 = A_0$, et $B_n = A_n \setminus A_{n-1}$. Les B_n sont disjoints par construction, et A_n est l'union des B_j pour $j \leq n$. On a donc

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{j=0}^{+\infty} \mu(B_j) = \lim_{n\to+\infty} \sum_{j=0}^{n} \mu(B_j)$$

$$= \lim_{n\to+\infty} \mu\left(\bigcup_{n=0}^{n} B_j\right) = \lim_{n\to+\infty} \mu\left(A_n\right).$$

qui termine la preuve. □

Définition B.3.7. (Partie négligeable, mesure complète (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. On dit que l'ensemble $N \subset X$ est *négligeable* s'il est inclus dans une partie $A \in \mathcal{A}$ de mesure nulle. On dit qu'une propriété est vérifiée *μ-presque partout*, ou qu'elle est vérifiée pour μ-presque tout x, si elle est vérifiée en dehors d'un ensemble négligeable. On dit que la mesure μ est *complète* si tous les ensembles négligeables sont dans $\mathcal{A}$.

Remarque B.3.8. Une propriété vérifiée presque partout est en particulier vérifiée sur un membre de la tribu dont le complémentaire est de mesure nulle. En effet, elle est vérifiée en dehors d'un ensemble négligeable N inclus dans $A \in \mathcal{A}$ de mesure nulle. Elle est donc a fortiori vérifiée sur $A^c \in \mathcal{A}$.

Définition B.3.9. (Ensembles de mesure pleine (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. On dit que $B \in \mathcal{A}$ est de mesure pleine si, pour tout $A \in \mathcal{A}$, $\mu(A \cap B) = \mu(A)$.

Définition B.3.10. (Absolue continuité d'une mesure par rapport à une autre (••))
Soit $(X, \mathcal{A})$ un espace mesurable, λ et μ deux mesures sur $\mathcal{A}$. On dit que μ est *absolument continue* par rapport à λ, et l'on écrit $\mu \ll \lambda$, si pour tout $A \in \mathcal{A}$ tel que $\lambda(A) = 0$, on a aussi $\mu(A) = 0$.

Remarque B.3.11. Cette propriété qui caractérise d'une certaine manière le positionnement relatif de deux mesures joue un rôle essentiel dans les applications. La situation typique est la suivante : on a une mesure définie sur un espace mesurable, notée λ (il s'agira en général de la mesure de Lebesgue sur $\mathbb{R}$ ou $\mathbb{R}^d$ qui va être définie dans la section suivante) qui, pour reprendre l'esprit des remarques introductives à cette partie, *tapisse* en quelque sorte l'espace sous-jacent, en permettant d'affecter à une zone son volume (que l'on peut concevoir comme une capacité à accueilir de la matière). Cette mesure sera en général *statique*, au sens où elle est définie une fois pour toutes. On fera vivre sur ce même espace d'autres mesures, que nous appellerons μ, qui représentent typiquement une distribution de matière, susceptible d'évoluer au cours du temps. Dire que l'on a absolue continuité de μ par rapport à λ signifie que l'on n'a pas de concentration : une zone de volume nul ne peut contenir qu'une masse elle-même nulle. Dans cette situation, le théorème de Radon-Nykodim (qui dépasse le cadre de ce cours sous sa forme présente) assurera que l'on peut représenter la mesure μ par une densité adossée à la mesure λ, ce qui permet de représenter une distribution de matière comme une *fonction*.

La proposition suivante donne un critère très utile en pratique d'égalité entre deux mesures, elle jouera un rôle essentiel dans la construction des mesures-produits.

Proposition B.3.12. (••) Soit $(X, \mathcal{A})$ un espace mesurable, et $\mathcal{C}$ un π − système sur X (définition B.2.17), qui engendre $\mathcal{A}$. Soient μ et ν deux mesures finies sur $\mathcal{A}$, telles que $\mu(X) = \nu(X)$ et $\mu(C) = \nu(C)$ pour tout C dans $\mathcal{C}$. Alors $\mu = \nu$.

Démonstration. On considère l'ensemble $\mathcal{D}$ de parties défini par

$$\mathcal{D} = \{A \in \mathcal{A}\,,\ \mu(A) = \nu(A)\}\ .$$

On souhaite montrer que $\mathcal{D} = \mathcal{A}$. Il suffit pour cela de montrer que $\mathcal{D}$ est une classe monotone qui contient $\mathcal{C}$. En effet, cela impliquera que $\mathcal{D}$ contient la classe monotone engendrée par $\mathcal{C}$ qui, d'après la proposition B.2.20, est la tribu engendrée par $\mathcal{C}$.
On a par hypothèse $X \in \mathcal{D}$. Pour tous A, B dans $\mathcal{D}$, avec $A \subset B$, on a

$$\mu\left(B \setminus A\right) = \mu\left(B\right) - \mu\left(A\right) = \nu\left(B\right) - \nu\left(A\right) = \nu\left(B \setminus A\right),$$

d'où $B \setminus A \in \mathcal{D}$. Pour toute suite croissante dans $\mathcal{D}$, on a, d'après la proposition B.3.6 ;

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mu\left(A_n\right) = \lim_{n\to+\infty} \nu\left(A_n\right) = \nu\left(\bigcup_{n\in\mathbb{N}} A_n\right).$$

Nous avons donc montré que $\mathcal{D}$ est une classe monotone. Comme évoqué ci-dessus, contenant $\mathcal{C}$, elle contient la classe monotone engendrée par $\mathcal{C}$, qui est la tribu engendrée par $\mathcal{C}$, c'est-à-dire $\mathcal{A}$. On a donc $\mu(A) = \nu(A)$ pour tout $A \in \mathcal{A}$, les mesures sont donc les mêmes. □

Corollaire B.3.13. On se place dans les hypothèses de la proposition précédente, en ne supposant plus les mesures finies. On suppose en revanche qu'il existe une suite croissante (C_n) d'éléments de $\mathcal{C}$, dont l'union est égale à X, et telle que $\mu(C_n) < +\infty$ et $\nu(C_n) < +\infty$ pour tout n. Alors $\mu = \nu$.

Démonstration. Pour tout n on définit

$$\mu_n(A) = \mu(A \cap C_n)\,,\ \nu_n(A) = \nu(A \cap C_n),$$

D'après la proposition B.3.12, on a $\mu_n = \nu_n$ pour tout n, d'où, pour tout $A \in \mathcal{A}$,

$$\mu(A) = \lim_n \mu_n(A) = \lim_n \nu_n(A) = \nu(A),$$

qui exprime l'identité des mesures μ et ν. □

Noter que les hypothèses de la proposition précédente imposent que μ et ν soient σ-finie, mais l'on demande en outre que la famille de parties de mesures finies qui recouvre X puisse être composée d'éléments du $\pi-$ système $\mathcal{C}$.

B.4 Mesures extérieures

Cette section présente la notion de mesure extérieure, qui correspond à une mesure à la laquelle on aurait ôté la condition de σ–additivité, remplacée par une notion affaiblie de σ–*sous*-additivité. Cette notion constituera une étape essentielle dans la construction de la mesure de Lebesgue. Comme on le verra, il est assez facile de construire explicitement une telle mesure extérieure définie sur l'ensemble des parties de $\mathbb{R}$ ou $\mathbb{R}^d$, et c'est en dégrossissant cette mesure extérieure que nous aboutirons à la mesure de Lebesgue.

B.4.1 Définitions, premières propriétés

Définition B.4.1. (•) Soit X un ensemble. On appelle *mesure extérieure* sur X une application $\mu^\star$ de $\mathcal{P}(X)$ dans $[0, +\infty]$ telle que

(i) $\mu^\star(\emptyset) = 0$,
(ii) si $A \subset B$ alors $\mu^\star(A) \leq \mu^\star(B)$,
(iii) si (A_n) est une collection au plus dénombrable de parties de X, alors

$$\mu^\star\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=0}^{+\infty} \mu^\star(A_n).$$

Exemples B.4.1. Nous donnons ici quelques exemples de mesures extérieures associées à un ensemble X quelconque. Précisons en premier lieu que toute mesure définie sur la tribu constituée de toutes les parties d'un ensemble est une mesure extérieure, du fait que la monotonie est une propriété des mesures (voir proposition B.3.3), ainsi que la sous-additivité (voir proposition B.3.5). La mesure de comptage, ou toute masse ponctuelle (voir exemples B.3.1), si on les définit sur la tribu discrète $\mathcal{P}(X)$, sont donc des mesures extérieures.
Autres exemples :

1. Soit X un ensemble. On pose $\mu^\star(\emptyset) = 0$ et $\mu^\star(A) = 1$ pour tout $A \in \mathcal{P}(X)$, $A \neq \emptyset$ (voir exercice B.4.1).
2. Soit X un ensemble. Pour tout $A \in \mathcal{P}(X)$, on pose $\mu^\star(A) = 0$ si A est dénombrable, et $\mu^\star(A) = 1$ dès que A est non dénombrable.

Une mesure extérieure n'est pas en général une mesure sur $\mathcal{P}(X)$. Nous verrons néanmoins que sa restriction à une sous-partie de $\mathcal{P}(X)$ est bien une mesure. Cette sous-partie est constituée des ensembles appelés $\mu^\star-$ mesurables. Il s'agit d'ensemble qui, avec leur complémentaire, constituent une partition de l'espace complet vis-à-vis de laquelle la mesure extérieure est *additive*, au sens précisé ci-dessous.

Définition B.4.2. (Ensembles mesurables pour une mesure extérieure (•))
Soit X un ensemble, et $\mu^\star$ une mesure extérieure sur X. On dit que A est $\mu^\star$−mesurable si, pour tout $B \subset X$, on a

$$\mu^\star(B) = \mu^\star(B \cap A) + \mu^\star(B \cap A^c). \tag{B.4.1}$$

Du fait de la sous-additivité de $\mu^\star$, l'inégalité

$$\mu^\star(B) \geq \mu^\star(B \cap A) + \mu^\star(B \cap A^c)$$

pour tout B suffit pour assurer la $\mu^\star-$ mesurablité de A.

Exercice B.4.1. Soit X un ensemble non vide et $\mu^\star$ l'application de $\mathcal{P}(X)$ dans $\mathbb{R}_+$ définie par $\mu^\star(\emptyset) = 0$ et $\mu^\star(A) = 1$ pour tout $A \in \mathcal{P}(X)$, différent de $\emptyset$. Montrer que $\mu^\star$ est une mesure extérieure. À quelles conditions $\mu^\star$ est-elle une mesure ? Quels sont les parties de X mesurables pour $\mu^\star$?

Nous verrons que, si l'on s'en tient à la définition, il se peut qu'une mesure extérieure admette très peu d'ensembles mesurables (voir exercice B.4.1 ci-après). La proposition suivante établit que, dans tous les cas, les ensembles *très petits* (de mesure extérieure nulle), ou *très gros* (i.e. de complémentaire très petit), sont mesurables au sens de la définition précédente.

Proposition B.4.3. (•) Soit X un ensemble et $\mu^\star$ une mesure extérieure sur X (définition B.4.2). Tout $B \subset X$ tel que $\mu^\star(B) = 0$ ou $\mu^\star(B^c) = 0$ est $\mu^\star-$ mesurable.

Démonstration. Soit B une telle partie. Comme indiqué dans la définition, il suffit de vérifier que, pour tout $A \subset X$, $\mu^\star(A) \geq \mu^\star(A \cap B) + \mu^\star(A \cap B^c)$. Or, si $\mu^\star(B) = 0$, alors $\mu^\star(A \cap B) \leq \mu^\star(B) = 0$ par monotonie, et $\mu^\star(A \cap B^c) \leq \mu^\star(A)$ par monotonie également. Le cas $\mu^\star(B^c) = 0$ se traite de la même manière. □

B.4.2 D'une mesure extérieure à une mesure

Le théorème suivant constitue un outil très général pour construire des couples tribus - mesures à partir de la donnée d'une mesure extérieure. Nous l'utiliserons pour construire la mesure de Lebesgue à partir de sa version extérieure (définie par la proposition B.4.6 ci-après).

Théorème B.4.4. (••) Soit X un ensemble et $\mu^\star$ une mesure extérieure sur X (définition B.4.2). On note $\mathcal{A}_{\mu^\star}$ la collection des parties $\mu^\star-$ mesurables (selon la définition B.4.2).

Alors $\mathcal{A}_{\mu^\star}$ est une tribu, et la restriction de $\mu^\star$ à $\mathcal{A}_{\mu^\star}$ est une mesure.

Démonstration. (•••) Montrons dans un premier temps que $\mathcal{A}_{\mu^\star}$ est une algèbre, c'est-à-dire qu'elle vérifie les conditions de la définition B.2.13, avec condition (*iii*) limitée aux collections finies. La proposition B.4.3 assure que $\emptyset$ et X sont dans $\mathcal{A}_{\mu^\star}$. Par ailleurs l'identité (B.4.1) est inchangée si l'on échange les rôles de B et B^c, $\mathcal{A}_{\mu^\star}$ est donc stable par complémentarité.

Montrons maintenant la stabilité par union. Soient B_1, B_2 deux parties de $\mathcal{A}_{\mu^\star}$ (la démarche ci-dessous est illustrée par la figure B.4.1). Comme B_1 est $\mu^\star$–mesurable, on a, pour tout $A \subset X$,

$$\mu^\star(A \cap (B_1 \cup B_2)) = \mu^\star(A \cap (B_1 \cup B_2) \cap B_1) + \mu^\star(A \cap (B_1 \cup B_2) \cap B_1^c)$$

$$= \mu^\star(A \cap B_1) + \mu^\star(A \cap B_1^c \cap B_2).$$

On a donc, en utilisant $(B_1 \cup B_2)^c = B_1^c \cap B_2^c$ et la $\mu^\star-$ mesurabilité de B_2,

$$\begin{aligned} & \mu^\star(A \cap (B_1 \cup B_2)) + \mu^\star(A \cap (B_1 \cup B_2)^c) \\ = \quad & \mu^\star(A \cap B_1) + \mu^\star((A \cap B_1^c) \cap B_2) + \mu^\star((A \cap B_1^c) \cap B_2^c) \\ = \quad & \mu^\star(A \cap B_1) + \mu^\star(A \cap B_1^c) \end{aligned}$$

qui est égal à $\mu^\star(A)$ du fait que $B_1 \in \mathcal{A}_{\mu^\star}$. L'ensemble $\mathcal{A}_{\mu^\star}$ est donc stable par unions finies, il s'agit bien d'une algèbre.

Pour montrer que c'est une tribu, il reste à montrer qu'elle est stable par union dénombrable. Considérons une suite (B_k) d'éléments de $\mathcal{A}_{\mu^\star}$, supposés deux à deux *disjoints*. Nous allons montrer par récurrence que, pour tout n, la partition finie (recouvrement de X par $n + 1$ parties disjointes)

$$B_1\,,\ B_2\,,\ \ldots\,,\ B_n\,,\ \left(\bigcup_{k=1}^{n} B_k\right)^c$$

respecte $\mu^\star$, au sens où, pour tout A

$$\mu^\star(A) = \sum_{k=1}^{n} \mu^\star(A \cap B_k) + \mu^\star\left(A \cap \left(\bigcup_{k=1}^{n} B_k\right)^c\right).$$

Nous allons en fait démontrer par récurrence l'identité équivalente

$$\mu^\star(A) = \sum_{k=1}^{n} \mu^\star(A \cap B_k) + \mu^\star\left(A \cap \left(\bigcap_{k=1}^{n} B_k^c\right)\right). \tag{B.4.2}$$

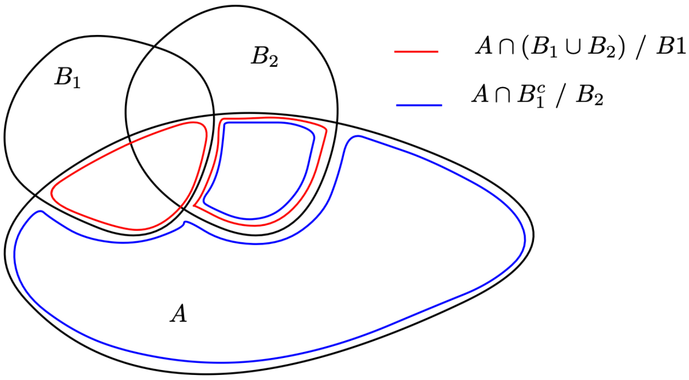

FIGURE B.4.1 – Mesurabilité de $B_1 \cup B_2$. La notation C / D de la légende indique que l'on écrit C comme l'union de $C \cap D$ et de $C \cap D^c$, où D est un ensemble mesurable.

L'identité pour $n = 1$ exprime simplement la $\mu^\star-$ mesurablité de B_1. Supposons maintenant qu'elle est vraie jusqu'au rang n. L'appartenance de B_{n+1} à $\mathcal{A}_{\mu^\star}$ implique

$$\begin{aligned} & \mu^\star\left(A \cap \left(\bigcap_{k=1}^{n} B_k^c\right)\right) \\ = \quad & \mu^\star\left(A \cap \left(\bigcap_{k=1}^{n} B_k^c\right) \cap B_{n+1}\right) + \mu^\star\left(A \cap \left(\bigcap_{k=1}^{n} B_k^c\right) \cap B_{n+1}^c\right) \\ = \quad & \mu^\star(A \cap B_{n+1}) + \mu^\star\left(A \cap \left(\bigcap_{k=1}^{n+1} B_k^c\right)\right) \end{aligned}$$

qui établit (B.4.2). On remarque maintenant que, par monotonie, le second terme de (B.4.2) diminue (au sens large) si l'on remplace l'intersection finie par l'intersection de tous les B_k^c. On a donc, en faisant tendre n vers $+\infty$,

$$\mu^\star(A) \geq \sum_{k=1}^{+\infty} \mu^\star(A \cap B_i) + \mu^\star\left(A \cap \left(\bigcap_{k=1}^{+\infty} B_k^c\right)\right), \tag{B.4.3}$$

d'où, par σ-sous-additivité de $\mu^\star$,

$$\mu^\star(A) \geq \mu^\star\left(A \cap \left(\bigcup_{k=1}^{+\infty} B_k\right)\right) + \mu^\star\left(A \cap \left(\bigcup_{k=1}^{+\infty} B_k\right)^c\right),$$

qui est supérieur ou égal à $\mu^\star(A)$ par sous-additivité. On a donc identité entre $\mu^\star(A)$ et l'expression ci-dessus, pour tout A, ce qui prouve que l'union des B_k est dans $\mathcal{A}_{\mu^\star}$. On en déduit immédiatement la stabilité par union dénombrable générale (sans le caractère disjoint deux à deux), en notant que l'union des B_k dans $\mathcal{A}_{\mu^\star}$ peut s'écrire comme union d'ensembles disjoints

$$B_1\,,\ B_2 \cap B_1^c\,,\ B_3 \cap B_2^c \cap B_1^c\,,\ \ldots\,,\ B_n \cap B_{n-1}^c \cap B_{n-2}^c \cap \ldots \cap B_1^c\,,\ \ldots$$

Nous avons ainsi démontré que $\mathcal{A}_{\mu^\star}$ est une tribu.

Il reste à vérifier que la restriction de $\mu^\star$ à $\mathcal{A}_{\mu^\star}$ est bien une mesure. Il suffit pour cela de vérifier l'additivité dénombrable. Considérons une suite (B_k) d'éléments de $\mathcal{A}_{\mu^\star}$ disjoints deux à deux. On écrit simplement l'inégalité (B.4.3) en prenant pour A l'union des B_k. Il vient

$$\mu^\star(A) \geq \sum_{k=1}^{+\infty} \mu^\star(B_k) + 0$$

qui est supérieur à $\mu^\star(A)$ par sous-additivité, d'où l'identité entre les deux expressions. □

Remarque B.4.5. Noter que ce théorème peut aboutir dans certains cas à un résultat très "pauvre" pour certaines mesures extérieures. Comme l'illustre l'exercice B.4.1, si la mesure extérieure ne présente pas de bonnes propriétés d'additivité, les seuls ensembles mesurables sont X et $\emptyset$.

B.4.3 Mesure de Lebesgue

La première étape dans la construction de la mesure de Lebesgue est la définition d'une mesure *extérieure* de Lebesgue, selon le principe décrit dans l'introduction : on sait la valeur que l'on veut à la longueur d'un intervalle, la longueur totale d'une réunion d'intervalles (avec possibles recouvrements) est supérieure à la sommes des longueurs. On considère donc que la mesure d'un ensemble, telle que l'on souhaite la définir, est inférieure à la somme des longueurs des intervalles, pour tout recouvrement de l'ensemble. Ces considérations conduisent à la définition de ce que l'on appelle la mesure extérieure de Lebesgue, donnée par la proposition qui suit (la proposition établit que l'objet défini est bien une mesure extérieure au sens de la définition B.4.2).

Proposition B.4.6. (Mesure de Lebesgue extérieure sur $\mathbb{R}$ (••))
Pour tout $A \subset \mathbb{R}$, on note C_A l'ensemble des suites d'intervalles ouverts dont l'union recouvre A :

$$C_A = \left\{ (]a_i, b_i[)_{i\in\mathbb{N}} \ , \ A \subset \bigcup_{\mathbb{N}}]a_i, b_i[\right\}.$$

On autorise les intervalles à être vides (i.e. a_i peut être égal à b_i), ce qui revient à autoriser les collections finies. On définit alors $\lambda^\star \ : \ \mathcal{P}(X) \longrightarrow [0, +\infty]$ par

$$\lambda^\star(A) = \inf_{C_A} \left(\sum_i (b_i - a_i) \right). \tag{B.4.4}$$

Cette application est une mesure extérieure, appelée *mesure extérieure de Lebesgue*, et elle attribue à tout intervalle sa longueur.

Démonstration. (• • •) Montrons que $\lambda^\star$ vérifie les trois conditions de la définition B.4.2.
(*i*) En premier lieu, l'ensemble vide est recouvert par une réunion d'intervalles vides. On a donc $\lambda^\star(\emptyset) = 0$.
(*ii*) Ensuite, si $A \subset B$, alors toute suite d'intervalles qui recouvre B recouvre aussi A, l'infimum de (B.4.4) qui définit $\lambda^\star(A)$, porte donc sur un ensemble plus grand que celui associé à B, d'où $\lambda^\star(A) \leq \lambda^\star(B)$.
(*iii*) Il s'agit maintenant de démontrer la σ–sous-additivité. On considère une suite (A_n) de parties de X. Il s'agit de montrer que la mesure extérieure de l'union est inférieure à la somme des mesures. Remarquons tout d'abord que si la somme des mesures est infinie, alors l'inégalité est immédiatement vraie. On peut donc supposer que toutes les mesures sont finies. Pour tout $\varepsilon > 0$, il existe une collection $(]a_i^n, b_i^n[)$ qui réalise (B.4.4) à $\varepsilon/2^n$ près, i.e.

$$\sum (b_i^n - a_i^n) \leq \lambda^\star(A_n) + \frac{\varepsilon}{2^n}.$$

L'union de toutes ces collections d'intervalles est elle-même une collection dénombrable d'intervalles ouverts, dont la longueur totale majore par définition la mesure de l'union des A_n. On a donc

$$\lambda^\star\left(\bigcup A_n\right) \leq \sum_{n=0}^{+\infty} \left(\lambda^\star(A_n) + \frac{\varepsilon}{2^n}\right) = \sum_{n=0}^{+\infty} \lambda^\star(A_n) + 2\varepsilon,$$

pour tout ε, d'où la σ–sous-additivité.

Il reste à montrer que $\lambda^\star$ affecte aux intervalles (ouverts, fermés, ou mixtes) leur longueur. Considérons l'intervalle fermé $[a, b]$. Pour tout ε on peut recouvrir cet intervalle par $]a - \varepsilon, b + \varepsilon[$ et des intervalles vides, la quantité $\lambda^\star(A)$ est donc majorée par une quantité arbitrairement proche de $b - a$, on a donc $\lambda^\star(]a, b[) \leq b - a$. Montrons l'inégalité inverse. On considère pour cela un recouvrement de $[a, b]$ par des intervalles ouverts.

Comme $[a,b]$ est compact, on peut en extraire un recouvrement fini, que l'on note $(]a_i,b_i[)_{1\leq i\leq n}$ (on suppose que l'on ne garde dans ce recouvrement que des intervalles "utiles", i.e. qui rencontrent $]a,b[$). Il existe nécessairement i_1 tel que $a_{i_1} < a$ (sinon a ne serait pas couvert), et on a $a < b_{i_1}$ (les intervalles inutiles ont été exclus). Si $b_{i_1} > b$, l'intervalle recouvre $]a,b[$, sinon il existe nécessairement i_2 tel que $a_{i_2} < b_{i_1}$ (sinon $b_{i_1} \in]a,b[$ ne serait pas couvert). On construit ainsi une suite $]a_{i_k}, b_{i_k}[$, avec $a_{i_k} < b_{i_{k+1}}$. Comme la collection finie recouvre $]a,b[$, on finit par arriver à $b_{i_k} > b$, on arrête alors la construction et l'on note n le rang atteint (voir figure B.4.2). On a alors

$$\begin{aligned} b-a &\leq b_{i_n} - a_{i_1} = b_{i_n} - a_{i_n} + \underbrace{a_{i_n}}_{\leq b_{i_{n-1}}} - \cdots - a_{i_2} + \underbrace{a_{i_2}}_{\leq b_{i_1}} - a_{i_1} \\ &\leq b_{i_n} - a_{i_n} + \cdots + b_{i_2} - a_{i_2} + b_{i_1} - a_{i_1}, \end{aligned}$$

qui est inférieur ou égal à la longueur totale de la collection d'intervalles initiale (car on en a enlevé certains). On a donc $b - a \leq \lambda^\star(A)$. Pour finir si l'on considère l'intervalle $]a,b[$, on peut encadrer (pour l'inclusion) cet intervalle par des intervalles fermés $[a+\varepsilon, b-\varepsilon]$ et $[a-\varepsilon, b+\varepsilon]$, dont la mesure tend vers $b-a$, on a donc également $\lambda^\star(]a,b[) = b-a$. Le raisonnement est analogue pour les intervalles de type $[a,b[$ et $]a,b]$. $\square$

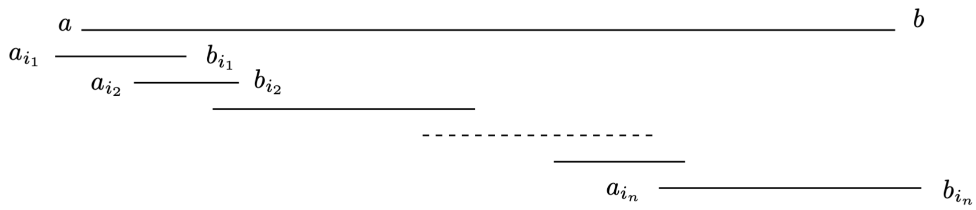

FIGURE B.4.2 – Recouvrement d'un intervalle

Définition B.4.7. (Mesure de Lebesgue extérieure sur $\mathbb{R}^d$ (• • •))
On définit de la même manière une mesure extérieure sur $\mathbb{R}^d$ en remplaçant les collections d'intervalles par des collections de pavés

$$I_1 \times I_2 \times \ldots I_d = \left\{(x_1,\ldots,x_d) \in \mathbb{R}^d\,,\ x_k \in I_k \text{ pour } \ k = 1,\ldots,d\right\}$$

où les I_k sont des intervalles ouverts de $\mathbb{R}$. Le volume du pavé est le produit des longueurs des intervalles qui le définissent. On vérifie de façon analogue que l'application ainsi construite de $\mathcal{P}(\mathbb{R}^d)$ dans $[0,+\infty]$ est bien une mesure extérieure.

Proposition B.4.8. (••) Les parties boréliennes de $\mathbb{R}$ sont mesurables pour la mesure extérieure de Lebesgue $\lambda^\star$ définie ci-dessus.

Démonstration. Montrons dans un premier temps que les intervalles de type $]-\infty,b]$ sont mesurables (au sens de la définition B.4.2). Soit B un tel intervalle. Il suffit de montrer que, pour tout $A \subset \mathbb{R}$,

$$\lambda^\star(A) \geq \lambda^\star(A\cap B) + \lambda^\star(A \cap B^c). \tag{B.4.5}$$

Si $\lambda^\star(A) = +\infty$, l'inégalité est automatiquement vérifiée. On se place maintenant dans le cas $\lambda^\star(A) < +\infty$. Soit $\varepsilon > 0$. Par définition de $\lambda^\star$, il existe une collection $(]a_n,b_n[)$ d'intervalles (bornés) dont l'union contient A telle que

$$\lambda^\star(A) \geq \sum (b_n - a_n) - \varepsilon.$$

Dans les lignes qui suivent, nous allons construire, à partir de ce recouvrement, des recouvrements de $A\cap B$ et $A \cap B^c$. Plus précisément, nous allons pour chaque intervalle $]a_n,b_n[$, construire deux intervalles ouverts

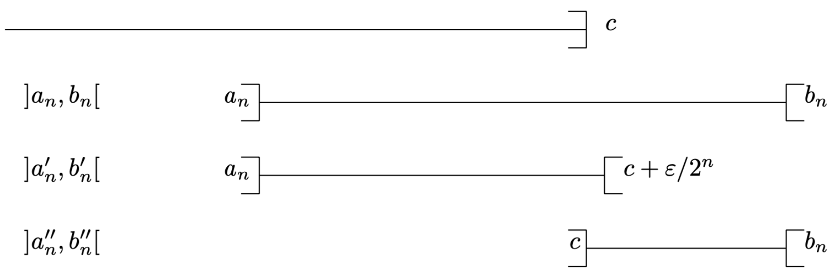

FIGURE B.4.3 – Décomposition des $]a_n, b_n[$

contenant respectivement $]a_n, b_n[\cap B$ et $]a_n, b_n[\cap B^c$. La seul difficulté est que l'on ne peut pas prendre directement ces intersections comme intervalles ouverts, car certains sont du type $]a_n, b]$. L'idée est alors de recouvrir par un intervalle ouvert du type $]a_n, b + \varepsilon/2^n[$, de façon à ce que que, même en sommant ensuite toutes les longueurs, on ne rajoute d'une quantité contrôlé par ε arbitrairement petit.

Plus précisément, pour tout n, l'intersection de $]a_n, b_n[$ avec $B =]-\infty, b]$ est soit vide (si $a_n \geq b$), soit un intervalle du type $]a_n, b]$ (si $a_n < b < b_n$), soit $]a_n, b_n[$ (si $b_n \leq b$). Dans le premier cas (intersection vide), on pose $a'_n = b'_n = 0$, de telle sorte de $]a'_n, b'_n[= \emptyset$. Dans le cas où il s'agit d'un intervalle du type $]a_n, b]$, ou si c'est l'intervalle $]a_n, b_n[$ lui-même, on pose $a'_n = a_n$ et $b'_n = b + \varepsilon/2^n$. De la même manière, l'intersection de $]a_n, b_n[$ avec $B^c =]b, +\infty[$ est soit vide (on pose alors $a''_n = b''_n = 0$), soit l'intervalle $]b, b_n[$, soit $]a_n, b_n[$. Dans ces deux derniers cas, on pose alors simplement $a''_n = \max(a_n, b)$, $b''_n = b_n$. Par construction (voir figure B.4.3), on a

$$b'_n - a'_n + b''_n - a''_n \leq b_n - a_n + \varepsilon/2^n.$$

L'union $]a'_n, b'_n[$ recouvre $A \cap B$, et l'union $]a''_n, b''_n[$ recouvre $A \cap B^c$. On a donc

$$\lambda^\star(A \cap B) \leq \sum (b'_n - a'_n) \text{ et } \lambda^\star(A \cap B^c) \leq \sum (b''_n - a''_n),$$

et ainsi

$$\lambda^\star(A \cap B) + \lambda^\star(A \cap B^c) \leq \sum (b'_n - a'_n) + \sum (b''_n - a''_n) \leq \sum (b_n - a_n) + 2\varepsilon \leq \lambda^\star(A) + 3\varepsilon,$$

pour tout $\varepsilon > 0$, ce qui prouve l'inégalité (B.4.5). L'intervalle $]-\infty, b]$ est donc mesurable pour tout $b \in \mathbb{R}$.

D'après le théorème B.4.4, la famille des parties mesurables pour $\lambda^\star$ est une tribu, qui contient les $]-\infty, b]$ d'après ce que l'on vient de voir. Cette famille contient donc la tribu engendrée par ces intervalles, qui est la tribu borélienne d'après la proposition B.2.7. □

Définition B.4.9. (Mesure de Lebesgue sur $\mathbb{R}$ et $\mathbb{R}^d$ (•))
On définit la mesure de Lebesgue λ comme la mesure construite, selon le théorème B.4.4, à partir de la mesure extérieure de Lebesgue $\lambda^\star$ définie par la proposition B.4.6. Cette mesure est définie sur la tribu des parties mesurables pour la mesure extérieure $\lambda^\star$. On appelle *tribu de Lebesgue* cette tribu. Elle contient la tribu des boréliens d'après la proposition B.4.8.

La mesure de Lebesgue sur $\mathbb{R}^d$ est définie de la même manière à partir de la mesure de Lebesgue extérieure sur $\mathbb{R}^d$ (proposition B.4.7).

Exercice B.4.2. (Caractère σ-fini de la mesure de Lebesgue (•))
Montrer que la mesure de Lebesgue sur $\mathbb{R}$ n'est pas finie, mais qu'elle est σ–finie.

Proposition B.4.10. (Invariance par translation de λ)
La mesure de Lebesgue est invariante par translation sur $\mathbb{R}$ (et sur $\mathbb{R}^d$) : pour tout A mesurable, tout $c \in \mathbb{R}$, $A + c$ est mesurable, et

$$\lambda(A + c) = \lambda(A).$$

Démonstration. Les collections d'intervalles $(]a_i, b_i[)_{i\in\mathbb{N}}$ qui recouvrent $A+c$ sont obtenues à partir de celles recouvrant A en translatant tous les intervalles de c. Comme les translations ne changent pas les longueurs des intervalles, la mesure extérieure $\lambda^\star(A+c)$ est égale à $\lambda^\star(A)$. La mesure de Lebesgue λ étant égale à cette mesure extérieure sur un sous-ensemble de $\mathcal{P}(X)$, on en déduit l'invariance par translation pour λ. $\square$

Remarque B.4.11. La propriété précédente implique une certaine forme d'*unicité* de la mesure de Lebesgue : si une mesure définie sur les boréliens affecte aux intervalles leur longueurs, alors il s'agit de la mesure de Lebesgue.

B.5 Compléments

La question de savoir si toutes les parties de X sont mesurables pour $\lambda^\star$ n'a pas encore été abordée. Si c'était le cas, la tribu associée $\mathcal{A}$ (voir théorème B.4.4) serait $\mathcal{P}(X)$ tout entier, et $\lambda^\star$ serait une mesure sans qu'il soit nécessaire d'élaguer la tribu discrète de ses membres non mesurables. Nous allons voir que ça n'est pas le cas : il existe bien des parties de $\mathbb{R}$ qui *ne sont pas mesurables* pour $\lambda^\star$, et qui sont donc exclues de la tribu des parties sur laquelle λ est définie. La proposition suivante établit directement l'existence d'un ensemble non mesurable pour λ. Le principe de cette démonstration reprend une idée simple évoquée dans l'introduction : nous allons en substance décomposer l'intervalle $]0,1[$ en une infinité dénombrable de parties qui, *si elles sont mesurables*, ont nécessairement pour mesure une même valeur. On est alors confronté à une alternative sans issue : si cette valeur est nulle, alors $]0,1[$ est de mesure nulle, et si la valeur est strictement positive, la mesure de $]0,1[$ est infinie.

Proposition B.5.1. (Ensemble de Vitali (• • •))
Il existe une partie de $\mathbb{R}$ qui n'est pas λ–mesurable .

Démonstration. On se propose de construire une partie de $I =]0,1[$ qui n'est pas mesurable. On introduit sur I la relation d'équivalence suivante :

$$x \,\mathcal{R}\, y \iff y - x \in \mathbb{Q}.$$

On choisit[9] un représentant de chaque classe, et l'on note C l'ensemble des représentants ainsi choisis. On considère maintenant une énumération (q_n) des rationnels de l'intervalle $]-1,1[$, et l'on s'intéresse à la collection des ensembles $q_n + C$, dont nous allons montrer qu'elle vérifie trois propriétés :

(i) Les $q_n + C$ sont disjoints. En effet, si $q_n + x = q_m + y$, avec x et y dans C, on a $y - x = q_n - q_m \in \mathbb{Q}$ donc y et x appartiennent à la même classe. Mais comme C est constitué de représentants uniques de chaque classe, cela implique $x = y$, d'où nécessairement $q_n = q_m$.

(ii) L'union des $q_n + C$ est incluse dans l'intervalle $]-1,2[$. C'est une conséquence directe du fait que $C \subset]0,1[$ et $q_n \in]-1,1[$ pour tout n.

(iii) L'intervalle $]0,1[$ est inclus dans l'union des $q_n + C$. Tout $y \in]0,1[$ appartient à sa classe $\overline{y}$, qui admet un (unique) représentant x dans C, donc $y = x + q$, avec q rationnel de l'intervalle $]-1,1[$, il s'agit donc de l'un des q_n de notre énumération.

Si C est mesurable, alors les $q_n + C$ sont mesurables, avec $\lambda(q_n + C) = \lambda(C)$ (d'après l'invariance par translation établie dans la proposition B.4.10). D'après (i) et l'additivité de la mesure λ, on a alors

$$\lambda\left(\bigcup_{n\in\mathbb{N}} (q_n + C)\right) = \sum_{n\in\mathbb{N}} \lambda\left(q_n + C\right) = \sum_{n\in\mathbb{N}} \lambda\left(C\right).$$

9. Il s'agit d'un point très délicat de la construction. Lorsque l'on dispose d'une infinité d'ensemble, il existe parfois une manière de *choisir* un élément de chacun des ensembles. Par exemple s'il s'agit d'intervalles fermés bornés, on peut prendre la borne inférieure. Ici il s'agit d'ensembles qui n'admettent pas de plus petit (ni de plus grand) élément, la procédure n'est donc pas applicable. De fait, on peut se convaincre qu'il n'existe pas de procédure systématique pour effectuer ce choix, et la possibilité d'extraire de la collection d'ensemble une collection de représentants repose sur ce qu'on appelle l'*axiome du choix* (axiome A.2.1, page 146), qui est une assertion que l'on ne peut pas démontrer à partir des axiomes de base de la théorie des ensembles, que l'on doit donc rajouter aux fondements de la théorie pour disposer de cette propriété.

Si $\lambda(C) = 0$, alors la somme ci-dessus est nulle, ce qui est absurde car, d'après (iii), $\cup(q_n + C)$ contient l'intervalle $]0, 1[$, qui est de mesure 1. Si $\lambda(C) > 0$, alors la somme est infinie, ce qui est absurde aussi car, d'après (ii), l'union des $q_n + C$ est contenue dans $] - 1, 2[$, qui est de mesure finie. L'ensemble C n'est donc pas mesurable. □

La démarche menée ci-dessus permet en fait de démontrer un résultat plus général qui, si l'on se base sur l'axiome du choix, assure l'impossibilité de construire une mesure sur la tribu discrète de $\mathbb{R}$, qui affecte aux intervalles leurs longueurs, et qui soit invariante par translation.

Proposition B.5.2. Il n'existe aucune mesure sur $\mathbb{R}$ définie sur la tribu discrète qui soit invariante par translation et qui affecte aux intervalles leur longueur.

Théorème B.5.3. (Régularité de la mesure de Lebesgue (•••))
La mesure de Lebesgue sur $\mathbb{R}^d$ est dite *régulière* au sens où, pour tout A mesurable,

$$\begin{aligned}\lambda(A) &= \inf\left(\lambda(U)\,,\ U \text{ ouvert } \ ,\ A \subset U\right)\\ &= \sup\left(\lambda(K)\,,\ K \text{ compact } \ ,\ K \subset A\right).\end{aligned}$$

Démonstration. Notons en premier lieu que, par monotonie, on a

$$\begin{aligned}\lambda(A) &\leq \inf\left(\lambda(U)\,,\ U \text{ ouvert}\,,\ A \subset U\right)\\ \lambda(A) &\geq \sup\left(\lambda(K)\,,\ K \text{ compact}\,,\ K \subset A\right).\end{aligned}$$

Le fait que la première inégalité soit une égalité découle directement de la définition : pour tout $\varepsilon > 0$, il existe une collection de pavés dont la réunion contient A, et telle que la somme des volumes est inférieure à $\lambda(A) + \varepsilon$, et cette réunion est un ouvert, l'infimum est donc égal à $\lambda(A)$.

Pour l'approximation intérieure, supposons dans un premier temps que A est borné, et considérons un fermé borné C qui contient A. Soit $\varepsilon > 0$. D'après ce qui précède, il existe U ouvert contenant $C \setminus A$ tel que

$$\lambda(U) \leq \lambda(C \setminus A) + \varepsilon.$$

Soit $K = C \setminus U = C \cap U^c$. Il s'agit d'un fermé borné par construction, donc d'un compact, et il est inclus dans A. On a $C \subset K \cup U$, donc $\lambda(C) \leq \lambda(K) + \lambda(U)$. On a donc finalement (du fait que $\lambda(C) = \lambda(C \setminus A) + \lambda(A)$),

$$\begin{aligned}\lambda(K) \geq \lambda(C) - \lambda(U) &= \lambda(C \setminus A) + \lambda(A) - \underbrace{\lambda(U)}_{\leq \lambda(C\setminus A)+\varepsilon}\\ &\geq \lambda(C \setminus A) + \lambda(A) - \lambda(C \setminus A) - \varepsilon\\ &= \lambda(A) - \varepsilon.\end{aligned}$$

On a donc bien égalité entre $\lambda(A)$ et le supremum.

Si A n'est pas borné, on l'écrit comme union croissante $\cup_j A_j$, où les A_j sont les intersections de A avec les boules fermée de rayon j. La mesure de A est la limite des $\lambda(A_j)$ d'après la proposition B.3.6, page 163. Si $\lambda(A) < +\infty$, il existe donc j tel que $\lambda(A_j) \geq \lambda(A) - \varepsilon$, et la construction précédente appliquée à A_j assure l'existence d'un compact K tel que $\lambda(K) \geq \lambda(A) - 2\varepsilon$. Si $\lambda(A) = +\infty$, alors $\lambda(A_j)$ tend vers $+\infty$, et l'on peut construire une suite de compacts $K_j \subset A$ tels que $\lambda(K_j) \geq \lambda(A_j) - 2\varepsilon$, qui tend vers $+\infty$. □

B.6 Exercices

Exercice B.6.1. (Tribu image / mesure image)
a) Soit $(X, \mathcal{A})$ un espace mesuré, X' un ensemble, et f une application de X dans X'. On définit

$$\mathcal{A}' = \left\{A' \subset X' \,,\; f^{-1}(A') \in \mathcal{A}\right\}.$$

Montrer que $\mathcal{A}'$ est une tribu, que l'on appelle tribu – image de $\mathcal{A}$ par f.

b) Montrer que $\mathcal{A}'$ est la *plus grande* tribu que l'on puisse mettre sur X', qui rende f mesurable.

c) On considère maintenant $(X, \mathcal{A}, \mu)$ un espace mesuré, $(X', \mathcal{A}')$ un espace mesurable [10], et f une application mesurable de X vers X'. On définit l'application

$$\nu \;:\; A' \in \mathcal{A}' \longmapsto \mu(f^{-1}(A)).$$

Montrer que l'on définit ainsi une mesure sur $(X', \mathcal{A}')$, appelée mesure – image de μ par f, ou *poussé en avant* de μ par f, ce que l'on note $\nu = f_\sharp\mu$. Montrer que la masse est conservée par cette opération de transport, c'est-à-dire que $\nu(X') = \mu(X)$.
Si l'on suppose que la mesure μ est la loi de probabilité d'une variable aléatoire X, quelle est l'interprétation de ν ?

c bis) Montrer que l'opération dans l'autre sens n'est pas pertinente. Plus précisément, on considère une application f mesurable d'une espace mesurable $(X, \mathcal{A})$ dans un espace mesuré $(X', \mathcal{A}', \nu)$, et l'on définit sur $\mathcal{A}$ l'application η sur $\mathcal{A}$ définie par $\eta(A) = \nu(f(A))$. Montrer par un ou plusieurs contre-exemples que η n'est pas une mesure en général.

d) On se place maintenant sur $X = X' = \mathbb{R}$ muni de la tribu borélienne, et de la mesure de Lebesgue λ. Décrire la mesure image $\nu = f_\sharp\lambda$ dans les cas suivants :

(i) $f(x) = x + c$, où $c \in \mathbb{R}$.

(ii) $f(x) = \alpha x$, avec $\alpha \neq 0$.

(iii) $f(x) = E(x)$ (partie entière de x).

(iv) $f(x) = 0$.

(v) $f(x) = \exp(x)$ (on précisera la mesure des intervalles $]a, b[$ dans l'espace d'arrivée).

$(\star)$ *On considère maintenant deux mesures μ et ν sur $(\mathbb{R}, \mathcal{B})$, de même masse totale finie, et l'on note $\Lambda_{\mu,\nu}$ l'ensemble des fonctions f mesurables de $\mathbb{R}$ dans $\mathbb{R}$ qui transportent μ vers ν :*

$$\Lambda_{\mu,\nu} = \{f \,,\; f_\sharp\mu = \nu\}.$$

e) $(\star)$ Montrer que, si f dans $\Lambda_{\mu,\nu}$, alors toute fonction g mesurable égale à f μ – presque partout est aussi dans $\Lambda_{\mu,\nu}$.

g) $(\star)$ Montrer au travers d'exemples que $\Lambda_{\mu,\nu}$ peut-être vide, ou réduit à un singleton (en identifiant les fonctions égales μ – presque partout), ou contenir un nombre fini d'élément, ou contenir une infinité non dénombrable d'éléments.

Exercice B.6.2. (Peigne de Dirac)
On se place sur la tribu des boréliens de $\mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, et l'on définit la mesure de comptage μ_0 de la façon suivante : pour toute partie A de $\mathbb{R}$, on définit $\mu_0(A) = \mathrm{Card}(A \cap \mathbb{Z})$ comme le nombre d'entiers relatifs que A contient. On écrit, en utilisant la notation introduite dans le deuxième des exemples B.3.1 :

$$\mu_0 = \sum_{k \in \mathbb{Z}} \delta_k.$$

a) Montrer que l'on définit ainsi une mesure sur $\mathbb{R}$ qui est σ–finie, sans être finie.
Quels sont les ensembles de mesure pleine pour μ_0 ? (voir définition B.3.9).

10. La tribu $\mathcal{A}'$ peut être la tribu image de $\mathcal{A}$ par f, auquel cas f est automatiquement mesurable, mais pas forcément.

Donner un exemple de fonctions égales μ_0– presque partout, qui sont pourtant très différentes au sens usuel du terme.

Construire une mesure du même type, en modifiant les coefficients de la somme ci-dessus, qui soit finie.

b) On définit maintenant, pour tout $n \in \mathbb{N}$,

$$\mu_n = \frac{1}{2^n} \sum_{k \in \mathbb{Z}} \delta_{k/2^n}.$$

Montrer qu'il s'agit encore d'une mesure infinie, et que l'on n'a ni $\mu_n \ll \lambda$ (mesure de Lebesgue), ni $\lambda \ll \mu_n$.

Montrer que, pour tout intervalle $]a, b[\subset \mathbb{R}$, on a

$$\lim_{n \to +\infty} \mu_n(]a, b[) \longrightarrow \lambda(]a, b[) = b - a,$$

où λ est la mesure de Lebesgue sur $\mathbb{R}$.

c) Montrer que, malgré la propriété de la question précédente, on n'a pas convergence de $\mu_n(A)$ vers $\mu(A)$ pour tout $A \in \mathcal{B}$.

Exercice B.6.3. Soit μ une mesure finie sur $(X, \mathcal{A})$.

a) Montrer que l'on a, pour tous A, $B \in \mathcal{A}$.

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

b) Montrer que l'on a, pour tous A, B, $C \in \mathcal{A}$.

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

c) Proposer une formule analogue pour une réunion d'un nombre quelconque (mais fini) d'éléments de $\mathcal{A}$.

Exercice B.6.4. On se place sur $\mathbb{R}$ muni de sa tribu borélienne, et l'on définit $\mu(A)$ comme le cardinal de l'ensemble des nombres rationnels contenus dans A.

a) Montrer qu'il s'agit d'une mesure sur $\mathcal{B}(\mathbb{R})$, qui est σ – finie, et qui donne une mesure infinie ou nulle à tout ouvert de $\mathbb{R}$.

b) Montrer que μ n'est pas comparable à la mesure de Lebesgue λ (on n'a ni $\mu \ll \lambda$, ni $\lambda \ll \mu$).

Exercice B.6.5. (Un réel sur deux)

On se propose ici de montrer qu'il n'existe aucun sous-ensemble de $\mathbb{R}$ qui contiendrait d'une certaine manière "un réel sur deux" quelle que soit l'échelle à laquelle on le regarde, et qui serait ainsi une sorte d'équivalent continu de l'ensemble des entiers pairs dans $\mathbb{N}$.

On note λ la mesure de Lebesgue sur $\mathbb{R}$.

a) Pour $n \geq 1$, on définit A_n comme l'ensemble des réels dont la n-ème décimale (en écriture propre) est entre 0, 1, 2, 3, ou 4. Montrer que pour tout intervalle I dont la longueur est un multiple entier de 10^{-n+1}, on a $\lambda(I \cap A_n) = \lambda(I)/2$.

Montrer que, pour tout intervalle $I =]a, b[$, on a

$$\lim_{n \to +\infty} \lambda(I \cap A_n) = \frac{1}{2} \lambda(I).$$

b) Montrer qu'il n'existe aucune partie A de $\mathbb{R}$ mesurable telle que l'on ait

$$\lambda(A \cap]a, b[) = \frac{b - a}{2} \quad \forall a\,,\ b\,,\ a < b.$$

Exercice B.6.6. (Ensemble de Cantor)
On pose $K_0 = [0,1]$, et l'on définit $K_1 = K_0 \backslash]1/3, 2/3[$, qui est donc la réunion de 2 intervalles fermés. On construit K_2 en retirant de la même manière le tiers central aux deux intervalles qui composent K_1. On construit ainsi K_3, $\ldots K_n$, qui est la réunion de 2^n intervalles de même longueur $1/3^n$. On définit l'ensemble de Cantor K comme l'intersection de K_n.

a) Montrer que K est un compact de $\mathbb{R}$, d'intérieur vide.

b) Montrer K a la puissance du continu, c'est-à-dire qu'il est infini non dénombrable, plus précisément équipotent à $\mathbb{R}$.

c) Montrer que K est Lebesgue – mesurable, de mesure nulle.

B.7 Fonctions mesurables, intégrale de Lebesgue

Nous décrivons dans cette section une procédure permettant de définir la notion d'intégrale pour une classe très générale de fonctions.

B.7.1 Fonctions mesurables

Rappelons qu'une application d'un espace mesurable $(X, \mathcal{A})$ dans $(X', \mathcal{A}')$ est dite mesurable si l'image réciproque par f de tout élément de $\mathcal{A}'$ est dans $\mathcal{A}$.

Dans le cas où l'espace d'arrivée est $\mathbb{R}$, on le considèrera par défaut muni de la tribu des boréliens, engendrée par les intervalles de type $]-\infty, c]$ (voir proposition B.2.7, page 157).

Dans le cas où l'espace d'arrivée est $\overline{\mathbb{R}} = [-\infty, \infty]$, on le considèrera aussi, sans qu'il soit besoin de le préciser, muni de sa tribu borélienne, engendrée par les $[-\infty, c]$ (voir proposition B.2.8, page 157).

On parlera donc simplement de fonction mesurable de $(X, \mathcal{A}, \mu)$ à valeurs dans $\mathbb{R}$ ou dans $\overline{\mathbb{R}}$, en gardant en tête que ces espaces sont munis de leurs tribus boréliennes. Si l'espace de départ est lui-même $\mathbb{R}$ (ou $\mathbb{R}^d$), on parle de fonction mesurable de $\mathbb{R}$ dans $\mathbb{R}$, où l'on considère l'espace de départ muni de la tribu de Lebesgue [11] (voir définition B.4.9, page 170).

Le caractère mesurable de telles fonctions se caractérise de façon élémentaire, comme l'exprime la proposition suivante.

Proposition B.7.1. (•) : Soit $(X, \mathcal{A})$ un espace mesurable, et f une fonction de X dans $\mathbb{R}$. La fonction f est mesurable si et seulement si, pour tout c réel

$$f^{-1}(]-\infty, c]) \in \mathcal{A}.$$

Pour une fonction à valeurs dans $\overline{\mathbb{R}}$, la condition est la même, pour les intervalles du type $[-\infty, c]$.

Démonstration. C'est une conséquence directe de la proposition B.2.12, page 159, qui donne un critère simple de mesurabilité d'une application : si la tribu sur l'espace d'arrivée est engendrée par une certaine famille, il suffit de vérifier que l'image réciproque de chaque élément de cette famille est dans la tribu sur l'espace de départ. □

Exercice B.7.1. Soit f une fonction monotone de $\mathbb{R}$ dans $\mathbb{R}$. Montrer que f est mesurable.

Proposition B.7.2. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et (f_n) une suite de fonctions mesurables de X dans $[-\infty, +\infty]$. On a alors

a) Les fonctions $\sup f_n$ et $\inf f_n$ sont mesurables.

b) Les fonctions $\limsup f_n$ et $\liminf f_n$ (voir définition A.1.36) sont mesurables.

c) Si (f_n) converge simplement vers f, alors f est mesurable.

Démonstration. a) Soit (f_n) une suite de fonctions mesurables. On définit

$$f_{\sup} = \sup(f_n) \text{ et } \ f_{\inf} = \inf(f_n).$$

On a, pour tout c dans $\mathbb{R}$,

$$f_{\sup}(x) = \sup(f_n(x)) \leq c \Longleftrightarrow f_n(x) \leq c \quad \forall n,$$

d'où

$$f_{\sup}^{-1}([-\infty, c]) = \bigcap_n f_n^{-1}([-\infty, c]),$$

11. Il peut sembler surprenant, lorsque l'on considère des fonctions de $\mathbb{R}$ dans $\mathbb{R}$, de munir les espaces d'arrivée et de départ de tribus différentes. L'intérêt de ce choix est de rendre le plus possible de fonctions mesurables, du fait que le critère de mesurabilité est d'autant plus laxiste que la tribu d'arrivée est grossière, et la tribu de départ fine. Noter que, en pratique, on montrera en général que $f^{-1}(]-\infty, b])$ est un borélien, donc *a fortiori* membre de la tribu de Lebesgue, de telle sorte que pour les situations usuelles, munir l'espace de départ de la tribu des boréliens ne changerait pas grand' chose.

qui est mesurable comme intersection de mesurables.

On a par ailleurs
$$f_{\inf}(x) = \inf(f_n(x)) \le c \Longleftrightarrow \forall N\,,\ \exists n\,,\ f_n(x) \le c + 1/N,$$
d'où
$$f_{\inf}^{-1}([-\infty, c]) = \bigcap_N \bigcup_n f_n^{-1}([-\infty, c+1/N]),$$
qui est mesurable comme intersections dénombrable de mesurables (eux même mesurables comme union dénombrable de mesurables).

b) Soit maintenant $f_{\limsup}$ définie par
$$f_{\limsup}(x) = \limsup f_n(x) = \lim_{n\to+\infty} \sup_{k\ge n} f_k(x).$$
On introduit $g_n = \sup_{k\ge n} f_k$ d'après ce qui précède, les g_n sont mesurables. La suite $g_n(x)$ étant décroissante pour tout x, elle converge dans $[-\infty, +\infty[$, et l'on a, pour tout x
$$\limsup f_n = \lim g_n = \inf g_n,$$
qui est mesurable toujours d'après ce qui précède.

On procède de la même manière pour la $\liminf$ en introduisant $h_n = \inf_{k\ge n}\ f_k$ qui est croissante.

c) Du fait que $\lim f_n = \limsup f_n$, la mesurabilité de la limite simple est conséquence immédiate de ce qui précède. □

Proposition B.7.3. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. Pour tous f, g mesurables, pour tout $\alpha \in \mathbb{R}$, αf et $f+g$ sont mesurables. L'ensemble des fonctions mesurables est donc un espace vectoriel.

Démonstration. Si $\alpha = 0$, $(\alpha f)^{-1}([-\infty, c]$ est soit vide, soit X tout entier. Pour $\alpha > 0$
$$(\alpha f)^{-1}([-\infty, c]) = \{\ x\,,\ f(x) \le c/\alpha\}\,,$$
qui est dans $\mathcal{A}$. Si $\alpha < 0$, on a
$$(\alpha f)^{-1}([-\infty, c] = \{\ x\,,\ \alpha f(x) \le c\} = \{f(x) \ge c/\alpha\} =$$
$$\{x\,,\ f(x) < c/\alpha\}^c = \left(\bigcup_n \{x\,,\ f(x) \le c/\alpha - 1/2^n\}\right)^c$$
qui est bien dans $\mathcal{A}$ par mesurabilité de f.

Considérons maintenant f et g mesurables. Montrons que $f(x) + g(x) < c$ si et seulement s'il existe un nombre rationnel q tel que
$$f(x) + q < c \text{ et } g(x) < q.$$
La condition suffisante est immédiate. Pour la condition nécessaire, on choisit un rationnel q tel que $g(x) < q < c - f(x)$ (il existe par densité des rationnels dans $\mathbb{R}$). Si l'on note (q_n) une énumération des rationnels, on a
$$\{x\,,\ f(x) + g(x) \le c\} = \bigcap_{n=0}^{+\infty} \{x\,,\ f(x) + g(x) < c + 1/2^n\}\ .$$
Chacun des ensembles ci-dessus est du type
$$\{x\,,\ f(x) + g(x) < c'\} = \bigcup_m (\{x\,,\ f(x) + q_m < c'\} \cap \{x\,,\ g(x) < q_m\})\,,$$
qui est mesurable comme union dénombrable de parties mesurables. L'ensemble
$$\{x\,,\ f(x) + g(x) \le c\}$$
est donc mesurable comme intersection dénombrable d'ensembles mesurables. □

Proposition B.7.4. (•) Soit un espace topologique, et $\mathcal{B}$ sa tribu des boréliens (engendrée par les ouverts, ou de façon équivalente par les fermés). Toute fonction f continue de $(X, \mathcal{B})$ dans $\overline{\mathbb{R}}$ est mesurable.

Démonstration. Pour tout $b \in \mathbb{R}$, l'intervalle $[-\infty, b]$ étant un fermé, son image réciproque par f est un fermé, il est donc dans $\mathcal{B}$ (et a fortiori dans la tribu de Lebesgue). □

Fonctions simples, fonctions étagées

Définition B.7.5. (Fonction simple, fonction étagée (•))

Soit X un ensemble. On appelle fonction *simple* une application de X dans $\mathbb{R}$ qui prend un nombre fini de valeurs $\alpha_1, \ldots, \alpha_n$.

Si $(X, \mathcal{A})$ est un espace mesurable, et que l'application simple f est mesurable, ce qui est équivalent à dire que $f^{-1}(\{\alpha_i\}) \in \mathcal{A}$ pour tout i, on parle de fonction *étagée*.

On notera

$$f(x) = \sum_{i=1}^{N} \alpha_i \mathbb{1}_{A_i}, \tag{B.7.1}$$

où les A_i sont mesurables, disjoints, et les α_i sont des réels (non nécessairement distincts).

On vérifie immédiatement que αf est étagée pour tout α réel, toute fonction f étagée. Pour la somme, on enrichit l'expression (B.7.1) de f en rajoutant le terme $\beta_0 \mathbb{1}_{A_0}$, avec $\beta_0 = 0$, et où A_0 est le complémentaire de l'union des A_i, de telle sorte que les $A_0, \ldots, A_N$ forment une partition de X. On fait de même pour g. La somme est une fonction étagée [12]

$$f + g = \sum_{i,j} (\alpha_i + \beta_j) \mathbb{1}_{A_i \cap B_j},$$

L'ensemble des fonctions étagée est donc un espace vectoriel, que l'on notera $\mathcal{E}(X)$ ou simplement $\mathcal{E}$. On notera $\mathcal{E}^+$ le sous-ensemble (qui est un cône convexe) des fonctions étagées à valeurs positives.

Nous terminons cette section par une propriété d'approximation des fonctions mesurables positives par des fonctions étagées.

Proposition B.7.6. (Approximation d'une fonction mesurable par des fonctions étagées (••))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et f une fonction mesurable de X dans $[0, +\infty]$. Il existe une suite (f_n) de fonctions de $\mathcal{E}^+$, croissante, avec $f_n \leq f$ pour tout n, qui converge simplement vers f, c'est à dire que

$$f(x) = \lim_n f_n(x) \quad \forall x \in X.$$

Démonstration. La démonstration repose sur la construction explicite d'une fonction étagée, qui reproduit de façon abstraite ce que ferait un logiciel de traitement d'image pour échantillonner les niveaux de gris, de façon à limiter l'espace mémoire nécessaire pour stocker l'image. L'idée est simplement de pratiquer cet échantillonnage avec une précision arbitrairement grande (dans le cas d'une image, il s'agirait de faire tendre vers l'infini le nombre de *bits* utilisés pour encoder les niveaux de gris). La petite différence avec ce cadre informatique est qu'ici on ne peut pas supposer que les valeurs de la fonction sont bornées, on doit donc construire une approximation de plus en plus fine, mais qui s'étale aussi sur une plage de valeurs de plus en plus grande. Pour tout entier $n \geq 1$, tout $k = 1, \ldots, n2^n$, on définit dans cet esprit (voir figure B.7.1 pour $n = 2$)

$$A_{n,k} = \{x\,,\ (k-1)/2^n \leq f(x) < k/2^n\}\ .$$

Pour tout n, les $A_{n,k}$ sont disjoints, et sont mesurables par mesurabilité de f. On définit maintenant la fonction f_n en affectant la valeur $(k-1)/2^n$ pour tout $x \in A_{n,k}$, et la valeur n pour les x qui ne sont dans aucun des $A_{n,k}$ (là où la valeur de f dépasse n). Les fonctions f_n sont étagées, la suite est croissante, et on a convergence simple de f_n vers f. □

12. Noter que, dans l'écriture qui suit, il est possible que certaines des sommes $\alpha_i + \beta_j$ soient égales (même si les α_i et les β_j sont distincts entre eux), ce que nous n'avons pas interdit dans l'écriture (B.7.1). En revanche les $A_i \cap B_j$ sont bien disjoints deux à deux. s

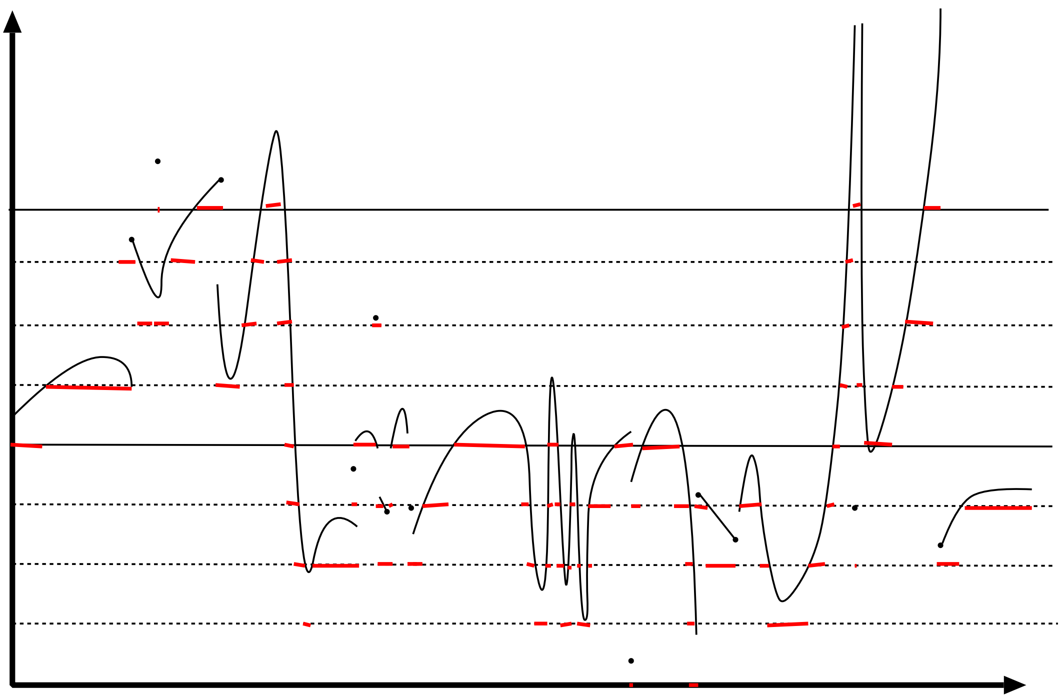

FIGURE B.7.1 – Approximation inférieure d'une fonction f (en noir) par une suite croissante de fonctions étagées.

Remarque B.7.7. La construction proposée dans la preuve précédente est *monotone* : si f et g sont deux fonctions mesurables avec $f \leq g$, f_n et g_n les suites associées, on a $f_n \leq g_n$ pour tout n.

B.7.2 Intégrale de fonctions étagées

Définition B.7.8. (Intégrale d'une fonction étagée positive (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, c'est-à-dire un ensemble muni d'une tribu $\mathcal{A}$ (définition B.2.13). Soit f une fonction étagée positive :

$$f(x) = \sum_{i=1}^{N} \alpha_i \mathbb{1}_{A_i},$$

où les A_i sont mesurables, disjoints, et les α_i sont des réels strictement positifs [13]. On définit [14] l'intégrale de f sur X comme la quantité

$$\int_X f(x) d\mu(x) = \sum_{i=1}^{N} \alpha_i \mu(A_i). \tag{B.7.2}$$

Pour tout $A \in \mathcal{A}$, on définit de la même manière

$$\int_A f(x) d\mu(x) = \sum_{i=1}^{N} \alpha_i \mu(A_i \cap A).$$

13. Cette stricte positivité n'était pas imposée dans l'écriture B.7.1 pour anticiper la situation où la somme de deux fonction étagées signées puisse conduire à des $\alpha_i + \beta_j$ nuls, mais nous considérons ici que nous ne gardons que les valeurs > 0, pour éviter d'avoir à préciser qu'une valeur nulle sur une partie de mesure infinie (qui prend une forme indéterminée $0 \times +\infty$) apporte une contribution nulle à l'intégrale.

Remarque B.7.9. On peut illustrer cette approche dans le contexte des images telles que celles qui sont stockées sur ordinateur. On peut voir une telle image (disons en noir et blanc pour simplifier) comme un tableau à $N \times N$ nombres dans l'intervalle $[0,1]$, qui correspondent aux niveaux de gris. Ces niveaux de gris sont en général stockés en format 8 bits, ce qui signifie que chaque niveau peut prendre l'une des 256 valeurs de la subdivision uniforme de $[0,1]$. Si l'on cherche à calculer la somme des niveaux de gris sur l'ensemble de l'image, l'approche usuelle (qui correspond à la philosophie de l'intégrale de Riemann) consiste à sommer les valeurs des pixels successifs :

$$S = \sum_{i=1}^{N} \sum_{j=1}^{N} u_{ij}.$$

L'approche suivie ici pour définir l'intégrale correspondrait à la démarche suivante, structurée par l'espace d'arrivée (les niveaux de gris), et pas l'espace de départ (les pixels) : pour chaque valeur g_k de niveau de gris, on considère l'ensemble A_k des pixels qui réalisent cette valeur. La somme est alors estimée selon la formule

$$S = \sum_{k=0}^{255} g_k \times \mathrm{Card}(A_k).$$

Cette approche repose implicitement sur l'*histogramme* de l'image, qui est la représentation de la distribution des niveaux de gris : en abscisse les 256 niveaux de gris, et en ordonnée les cardinaux des ensembles A_k correspondants.

Proposition B.7.10. (•) Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, f et g deux fonctions de $\mathcal{E}^+$, et $\alpha \geq 0$. On a

$$\int_X (\alpha f)\, d\mu = \alpha \int_X f\, d\mu\,, \quad \int (f+g)\, d\mu = \int_X f\, d\mu + \int_X g\, d\mu,$$

et

$$f(x) \leq g(x) \quad \forall x \in X \quad \Longrightarrow \quad \int_X f\, d\mu \leq \int_X f\, d\mu.$$

Démonstration. La première identité est conséquence directe de la définition. Pour la somme, on enrichit l'expression (B.7.1) de f en rajoutant le terme $\beta_0 \mathbb{1}_{A_0}$, avec $\beta_0 = 0$, et où A_0 est le complémentaire de l'union des A_i, de telle sorte que les A_0, …, A_N forment une partition de X. On fait de même pour g. On a [15]

$$f + g = \sum_{i,j} (\alpha_i + \beta_j) \mathbb{1}_{A_i \cap B_j},$$

d'où

$$\int (f+g) = \sum_{i,j} (\alpha_i + \beta_j) \mu(A_i \cap B_j) = \sum_i \alpha_i \sum_j \mu(A_i \cap B_j) + \sum_j \beta_j \sum_i \mu(A_i \cap B_j)$$

14. Si l'on n'impose pas $A_i = f^{-1}(\{\alpha_i\})$, l'écriture (B.7.1) de f n'est pas unique. On peut néanmoins vérifier que la quantité définie par (B.7.2) ne dépend pas de l'écriture choisie. En effet, si l'on considère une autre écriture

$$\sum_{i=1}^{N'} \alpha'_i \mu(A'_i),$$

on a, par additivité de la mesure, et du fait que $\alpha_i = \alpha'_j$ sur $A_i \cap A'_j$,

$$\sum_{i=1}^{N} \alpha_i \mu(A_i) = \sum_{i=1}^{N} \sum_{j=1}^{N'} \alpha_i \mu(A_i \cap A'_j) = \sum_{j=1}^{N'} \sum_{i=1}^{N} \alpha'_i \mu(A_i \cap A'_j) = \sum_{j=1}^{N'} \alpha'_i \mu(A'_i).$$

15. Noter que dans l'écriture qui suit, il est possible que certaines des sommes $\alpha_i + \beta_j$ soient égales (même si les α_i et les β_j sont distincts entre eux), ce que nous n'avons pas interdit dans l'écriture (B.7.1). En revanche les $A_i \cap B_j$ sont bien disjoints deux à deux. s

$$= \sum_i \alpha_i \mu(A_i) + \sum_j \beta_j \mu(B_j) = \int f + \int g$$

par additivité de la mesure. □

Proposition B.7.11. (•) Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, f une fonction de $\mathcal{E}^+$, et (f_n) une suite de fonctions de $\mathcal{E}^+$ (fonctions étagées positives). On suppose que (f_n) est croissante, c'est-à-dire que $(f_n(x))$ est une suite croissante pour tout x de X, et que f_n converge simplement vers f, c'est à dire que

$$\lim_{n\to+\infty} f_n(x) = f(x) \quad \forall x \in X.$$

L'intégrale de f est alors la limite des intégrales des f_n :

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

Démonstration. On a, d'après la proposition B.7.10, $\int f_n \leq \int f$ pour tout $n \in \mathbb{N}$. La suite des intégrales, croissante, converge donc vers une valeur $\lim \int f_n \leq \int f$. Montrons que cette inégalité est en fait une égalité. On sait que f peut s'écrire

$$f = \sum_{i=1}^N a_i \mathbb{1}_{A_i},$$

où les A_i sont des éléments disjoints de $\mathcal{A}$, et les a_i des réels strictement positifs. Soit $\varepsilon > 0$. Pour $i = 1, \ldots, N$, on introduit

$$A_i^n = \{x \in A_i\,,\ f_n(x) \geq (1-\varepsilon)a_i\} \in \mathcal{A}.$$

Pour tout i, la suite des (A_i^n) est croissante d'après la croissance de f_n, et l'union des A_i^n est égale à A_i par convergence simple de f_n vers f. On a donc, d'après la proposition B.3.6, page 163,

$$\lim_n \mu\left(A_i^n\right) = \mu(A_i).$$

On considère maintenant la fonction g_n définie par

$$g_n = \sum_{i=1}^N (1-\varepsilon) a_i \mathbb{1}_{A_i^n}.$$

C'est une fonction étagée, qui vérifie $g_n \leq f_n \leq f$, et la suite (g_n) est croissante. La suite réelle $(\int g_n)$ converge donc, et l'on a

$$\lim_n \int f_n \geq \lim_n \int g_n = \lim_n \left(\sum_{i=1}^N (1-\varepsilon) a_i \mu(A_i^n)\right) = (1-\varepsilon) \sum_{i=1}^N a_i \mu(A_i) = (1-\varepsilon) \int f.$$

Cette inégalité étant vérifiée pour tout $\varepsilon > 0$, on a bien $\lim_n \int f_n \geq \int f$, ce qui termine la preuve. □

B.7.3 Intégrale de fonctions mesurables

Cette définition de l'intégrale pour les fonctions étagées peut être étendue à une fonction f positive plus générale en considérant le supremum de l'ensemble des valeurs prises par les intégrales des fonctions étagées qui sont inférieures à f en tout point, comme le précise la définition suivante.

Définition B.7.12. (Intégrale d'une fonction mesurable positive (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et f une fonction mesurable de X dans $[0, +\infty]$. On définit l'intégrale de f sur X comme la quantité

$$\int_X f(x)\, d\mu = \sup_{g\in\mathcal{E}^+, g\leq f} \left(\int_X g(x) d\mu\right) \in [0, +\infty].$$

On définit de la même manière l'intégrale de f sur toute partie A mesurable.

Proposition B.7.13. (Monotonie de l'intégrale)
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, f et g deux fonctions mesurables de X dans $[0, +\infty]$. On a

$$f \leq g \quad \Longrightarrow \quad \int_X f\, d\mu \leq \int_X f\, d\mu.$$

Démonstration. Si $f \leq g$, alors toute fonction h de $\mathcal{E}^+$ admissible dans le sup définissant l'intégrale de f est admissible pour celui définissant g, d'où l'inégalité sur les intégrales. □

Exercice B.7.2. Montrer, en utilisant la définition précédente, que l'intégrale de la fonction indicatrice de l'ensemble des rationnels dans $\mathbb{R}$ est d'intégrale nulle.

L'intégrale définie ci-dessus ne "voit" pas les ensembles négligeables :

Proposition B.7.14. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, f et g des fonctions mesurables de X dans $[0, +\infty]$. On suppose que $f(x) = g(x)$ presque partout. Alors

$$\int_X f(x)\, d\mu(x) = \int_X g(x)\, d\mu(x).$$

Démonstration. On introduit $A \in \mathcal{A}$ sur lequel f et g s'identifient, tel que $\mu(A^c) = 0$. Toute fonction de $h \in \mathcal{E}^+$, inférieure à f, s'écrit

$$h = \sum_{i=1}^{N} a_i \mathbb{1}_{A_i} = \sum_{i=1}^{N} a_i \mathbb{1}_{A_i \cap A} + \sum_{i=1}^{N} a_i \mathbb{1}_{A_i \cap A^c}.$$

On a

$$\int h = \sum_{i=1}^{N} a_i\, \mu(A_i \cap A) + \sum_{i=1}^{N} a_i \mu\left(A_i \cap A^c\right).$$

Le second terme est nul car $\mu(A_i \cap A^c) \leq \mu(A^c) = 0$ pour tout i. Le premier terme est l'intégrale d'une fonction étagée qui est inférieure à f, donc à g (là où la fonction ne s'annule pas, f et g s'identifient). Cette quantité est donc inférieure à $\int g \in [0, +\infty]$, et ce pour tout h de $\mathcal{E}^+$ inférieure à f. On a donc $\int f \leq \int g$. Les rôles de f et g étant interchangeables, on montre de la même manière $\int g \leq \int f$, d'où l'identité des valeurs des deux intégrales. □

La proposition suivante, qui étend la proposition B.7.11 à une fonction mesurable quelconque (non nécessairement étagée), peut être vue comme une version préliminaire du théorème de convergence monotone, fondamental, qui sera énoncé plus loin.

Proposition B.7.15. (••) Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, f une fonction mesurable de X dans $[0, +\infty]$, et (f_n) une suite de fonctions de $\mathcal{E}^+$ (fonctions étagées positives). On suppose que (f_n) est croissante, c'est-à-dire que $(f_n(x))$ est une suite croissante pour tout x de X, et que f_n converge simplement vers f, c'est-à-dire que

$$\lim_{n \to +\infty} f_n(x) = f(x) \quad \forall x.$$

L'intégrale de f est alors la limite des intégrales des f_n :

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

Démonstration. On a de façon évidente

$$\int_X f_1 d\mu \leq \int_X f_2 d\mu \leq \cdots \leq \int_X f d\mu,$$

d'où l'on déduit que la limite de $\int f_n$ existe, et vérifie $\lim \int f_n \, d\mu \leq \int f \, d\mu \in [0, +\infty]$. Établissons maintenant l'inégalité inverse. L'intégrale de f étant (définition B.7.12) le supremum des intégrales $\int g$, pour g décrivant l'ensemble des fonctions de $\mathcal{E}^+$ inférieures à f, il suffit de montrer que pour toute fonction g de ce type, on a $\int g \leq \lim \int f_n$. Soit g une telle fonction de $\mathcal{E}^+$, inférieure à f. On considère la fonction $g_n = \min(g, f_n)$. La suite (g_n) est croissante car (f_n) l'est, et g_n converge simplement vers g. On a donc, d'après la proposition B.7.11,

$$\int g \, d\mu = \lim_n \int g_n.$$

Or on a $g_n \leq f_n$ pour tout n, d'où l'on déduit que la limite ci-dessus est majorée par $\lim \int f_n$, d'où finalement

$$\int g \, d\mu \leq \lim_n \int f_n,$$

qui conclut la preuve. □

Proposition B.7.16. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, f et g deux fonctions mesurables de X dans $[0, +\infty]$, et $\alpha \geq 0$. On a

$$\int_X (\alpha f) \, d\mu = \alpha \int_X f \, d\mu \,, \quad \int_X (f + g) \, d\mu = \int_X f \, d\mu + \int_X g \, d\mu.$$

Démonstration. La première identité est conséquence directe de la définition. Pour l'identité sur la somme, on utilise la proposition B.7.6, qui assure l'existence de suite de fonctions étagées (f_n) et (g_n) croissantes qui convergent simplement vers f et g respectivement . La suite $f_n + g_n$ est également dans $\mathcal{E}^+$, elle est croissante, et converge simplement vers $f + g$. On a

$$\int (f_n + g_n) = \int f_n + \int g_n$$

d'après la proposition B.7.10. On fait maintenant tendre n vers $+\infty$, pour obtenir grâce à la proposition B.7.15 que l'intégrale de la somme est la somme des intégrales. □

Intégrabilité des fonctions

Définition B.7.17. (Partie positive / négative d'une fonction (•))
Soit f une fonction d'un ensemble X dans $\overline{\mathbb{R}}$. On appelle *partie positive* de f, et l'on note f^+, la fonction qui à x associe $f^+(x) = \max(f(x), 0) = (f(x) + |f(x)|)/2$. La partie négative de f, notée f^-, est la partie positive de l'opposé de f, de telle sorte que l'on a

$$f = f^+ - f^-.$$

Définition B.7.18. (Intégrabilité (•))
Soit f une fonction mesurable de $(X, \mathcal{A}, \mu)$ dans $\overline{\mathbb{R}}$. On dit que f est intégrable si $\int f^+$ et $\int f^-$ sont finies. On définit alors l'intégrale de f comme

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

Si une seule des deux quantités $\int f^+$ et $\int f^-$ est finie, on dit que l'intégrale existe, et prend la valeur $-\infty$ si $\int f^+$ est finie, et $+\infty$ dans le cas contraire.
Si l'espace de départ est $\mathbb{R}^d$, on dira simplement que f est intégrable au sens de Lebesgue.

Exercice B.7.3. On note g la fonction $\sin(x)/x$ sur $\mathbb{R}_+$.

a) Que peut-on dire de l'intégrale généralisée (au sens vu en classes préparatoires) de g entre 0 et $+\infty$?

b) On se place maintenant dans le cadre de l'intégrale de Lebesgue. La fonction g est-elle intégrable sur $]0, +\infty[$? L'intégrale de g existe-t-elle ?

Proposition B.7.19. Soit f une fonction mesurable de $(X, \mathcal{A}, \mu)$ dans $\overline{\mathbb{R}}$. Alors f est intégrable si et seulement si $|f|$ l'est, et l'on a

$$\left|\int f\, d\mu\right| \leq \int |f|\, d\mu.$$

Démonstration. Si f est intégrable, alors les intégrales de f^+ et f^- sont finies, l'intégrale de $f^+ + f^- = |f|$ est donc finie. Inversement, l'intégrabilité de $|f| = f^+ + f^-$ assure l'intégrabilité de f^+ et f^-. On a

$$\left|\int f\, d\mu\right| = \left|\int f^+\, d\mu - \int f^-\, d\mu\right| \leq \int f^+\, d\mu + \int f^-\, d\mu$$

qui est égal à $\int |f|\, d\mu$. □

Proposition B.7.20. (•) Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, f et g deux fonctions mesurables de X dans $\overline{\mathbb{R}}$. On suppose que f et g sont égales presque partout, alors les fonctions sont indiscernables du point de vue de l'intégration, c'est-à-dire que f est intégrable si et seulement si g l'est , et alors $\int_A f = \int_A g$ pour tout $A \in \mathcal{A}$.

Démonstration. Si f et g s'identifient presque partout, il en est de même de leurs parties positives et négatives. La propriété est donc conséquence directe de la proposition B.7.14. □

Proposition B.7.21. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré et f une fonction intégrable à valeurs dans $[0, +\infty]$. Pour tout $t \in]0, +\infty[$ on introduit $A_t = \{x, f(x) \geq t\}$. On a

$$\mu(A_t) \leq \frac{1}{t}\int_{A_t} f(x) d\mu \leq \frac{1}{t}\int_X f(x) d\mu.$$

Démonstration. On a $0 \leq t\mathbb{1}_{A_t} \leq f\mathbb{1}_{A_t} \leq f$, d'où

$$t\mu(A_t) \leq \int_{A_t} f(x) \leq \int_X f(x),$$

d'où l'on tire les inégalités annoncées en divisant par t. □

Proposition B.7.22. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré et f une fonction intégrable à valeurs dans $[-\infty, +\infty]$. Alors f est finie μ–presque partout, i.e.

$$\mu\left(\{x\,,\ |f(x)| = +\infty\}\right) = 0.$$

Démonstration. C'est une conséquence de la proposition B.7.21. On a effet pour tout $n \in \mathbb{N}$

$$\mu\left(\{x\,,\ |f(x)| = +\infty\}\right) \leq \mu\left(\{x\,,\ |f(x)| \geq n\}\right) \leq \frac{1}{n}\int_X |f|\, d\mu.$$

La quantité positive $\mu\left(\{x\,,\ |f(x)| = +\infty\}\right)$ est donc majorée par des réels arbitrairement petits, elle est donc nulle. □

B.7.4 Théorèmes fondamentaux

Nous pouvons maintenant démontrer le théorème de convergence monotone, qui constitue l'aboutissement des propositions B.7.11 et B.7.15.

Théorème B.7.23. (Convergence monotone (••))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, (f_n) une suite de fonctions mesurables et positives de X dans $[0, +\infty]$. On suppose que

1. la suite $(f_n(x))$ est croissante pour presque tout x (voir définition B.3.7),
2. f_n converge simplement vers f.

Alors f est mesurable, et l'intégrale de f est la limite des intégrales des f_n :

$$\int f\,d\mu = \lim_{n\to+\infty} \int f_n\,d\mu \in [0,+\infty].$$

Démonstration. La mesurabilité de f est une conséquence de la proposition B.7.2, page 177. On suppose dans un premier temps que les propriétés de monotonie et de convergence ponctuelle sont vérifiées pour tout x dans X. La monotonicité de l'intégrale (proposition B.7.16) assure que

$$\int f_1 d\mu \leq \int f_2 d\mu \leq \cdots \leq \int f d\mu,$$

On a donc convergence de la suite $(\int f_n)$ vers un réel inférieur ou égal à $\int f$. Montrons maintenant l'inégalité inverse. Pour tout n, la fonction f_n peut être approchée inférieurement par une suite $(g_{n,j})_j$ dans $\mathcal{E}^+$ (voir proposition B.7.6, page 179). On considère que la suite approchante est construite selon le procédé proposé dans la preuve, de telle sorte que l'on a toujours $g_{n,j} \leq g_{m,j}$, pour tous $m \leq n$, tout j (voir remarque B.7.7, page 180). On définit maintenant la fonction $h_n = g_{n,n} \in \mathcal{E}^+$, qui est croissante par construction, et telle que $h_n(x) \leq f(x)$.

Montrons enfin la convergence de $h_n(x)$ vers $f(x)$. Pour tout $\varepsilon > 0$, il existe n tel que $f_n(x) \geq f(x) - \varepsilon$. Il existe $j \geq n$ tel que $g_{n,j}(x) \geq f_n(x) - \varepsilon$. On a donc

$$h_j(x) \geq g_{n,j}(x) \geq f(x) - 2\varepsilon,$$

d'où la convergence de $h_j(x)$ vers $f(x)$. D'après la proposition B.7.15, on a donc

$$\int f = \lim_n \int h_n \leq \lim_n \int f_n,$$

ce qui conclut la première partie de la preuve.

On suppose maintenant que les propriétés de croissance et de convergence simple ne sont vérifiées que presque partout : il existe un ensemble $A \in \mathcal{A}$, dont le complémentaire est de mesure nulle, et sur lequel les propriétés sont vérifiées. La suite $f_n \mathbb{1}_A$ (qui met à 0 toutes les valeurs sur A^c) vérifie les hypothèses vis-à-vis de la fonction cible $f\mathbb{1}_A$. On a donc, d'après ce qui précède, convergence de la suite des intégrales vers l'intégrale de f. Or, comme $f_n\mathbb{1}_A$ s'identifie à f_n presque partout, de même pour $f\mathbb{1}_A$ et f, les intégrales sont les mêmes (d'après la proposition B.7.14), ce qui conclut la preuve. □

Lemme B.7.24. (Fatou)
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré et (f_n) une suite de fonctions mesurables de X dans $[0,+\infty]$. On a

$$\int \liminf_n f_n\,d\mu \leq \liminf_n \int f_n\,d\mu.$$

Démonstration. Pour tout n on définit g_n par $g_n(x) = \inf_{k\geq n} f_k(x)$. D'après la proposition B.7.2, chacune de ces fonctions est mesurable. La suite des g_n, croissante, et converge simplement vers $\liminf_n f_n$ par définition de la $\liminf$.

D'après le théorème de convergence monotone B.7.23, on a donc

$$\int \liminf_n f_n = \lim_n \int g_n \leq \liminf \int f_n$$

car $g_n \leq f_n$ pour tout n. Noter qu'il s'agit bien d'une $\liminf$ dans le membre de droite, car, la suite f_n n'ayant pas de propriété de monotonie, la suite $\int f_n$ peut ne pas converger. □

Théorème B.7.25. (Convergence dominée)
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, g une fonction intégrable de X dans $[0,+\infty]$, et (f_n) une suite de fonctions mesurables de X dans $[-\infty,+\infty]$. On suppose

$$f(x) = \lim_n f_n(x) \quad \text{pour presque tout } x,$$

et que, pour tout n,

$$|f_n(x)| \leq g(x)$$

pour presque tout x dans X. Alors les fonctions f et f_n pour tout n sont intégrables sur X, et l'on a

$$\lim \int |f - f_n| \, d\mu = 0,$$

d'où en particulier $\lim \int f_n = \int f$.

Démonstration. Il existe un ensemble A dans $\mathcal{A}$, dont le complémentaire est de mesure nulle [16], tel que toutes les propriétés soient vérifiées. Pour tout x dans A, on a $|f_n(x)| \leq g(x)$ et, par passage à la limite, $|f(x)| \leq g(x)$. On a donc $\int |f_n| \leq \int g < +\infty$ et $\int |f| \leq \int g < +\infty$, qui exprime l'intégrabilité de f et des f_n. On a par ailleurs $|f_n - f| \leq |f_n| + |f| \leq 2g$, qui est donc intégrable pour tout n. On applique le lemme de Fatou B.7.24 à la suite de fonctions positives $(2g - |f_n - f|)$:

$$\int \liminf (2g - |f_n - f|) \leq \liminf \int (2g - |f_n - f|),$$

d'où l'on déduit, par linéarité de l'intégrale (et prenant garde de transformer les lim inf en lim sup quand on fait sortir le signe $-$),

$$\limsup \int |f_n - f| \leq \int \limsup |f_n - f| \, .$$

Or, comme f_n converge vers f sur A, la fonction $\limsup |f_n - f|$ est identiquement nulle presque partout, d'où la nullité de son intégrale, ce qui exprime la convergence de $\int |f_n - f|$ vers 0. □

B.7.5 Intégrales multiples

Définition B.7.26. (Rectangles)
Soient $(X_1, \mathcal{A}_1)$ et $(X_2, \mathcal{A}_2)$ deux espaces mesurables. On appelle *rectangle* de $X_1 \times X_2$ un ensemble de la forme $A_1 \times A_2$, avec $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$, et l'on note $\mathcal{R}$ l'ensemble de ces rectangles.

Proposition B.7.27. Soient $(X_1, \mathcal{A}_1)$ et $(X_2, \mathcal{A}_2)$ deux espaces mesurables. L'ensemble $\mathcal{R}$ des rectangles est un π – système (définition B.2.17), c'est à dire qu'il est stable par intersection finie.

Démonstration. Pour tous rectangles $A_1 \times A_2$ et $A_1' \times A_2'$ de $\mathcal{A}_1 \times \mathcal{A}_2$, on a

$$(A_1 \times A_2) \cap (A_1' \times A_2') = \left(\underbrace{A_1 \cap A_1'}_{\in \mathcal{A}_1} \right) \times \left(\underbrace{A_2 \cap A_2'}_{\in \mathcal{A}_2} \right),$$

qui appartient $\mathcal{A}_1 \times \mathcal{A}_2$. □

Définition B.7.28. (Tribu-produit)
Soient $(X_1, \mathcal{A}_1)$ et $(X_2, \mathcal{A}_2)$ deux espaces mesurables. On appelle *tribu-produit* de $\mathcal{A}_1$ et $\mathcal{A}_2$ la tribu de $X_1 \times X_2$ engendrée par les rectangles. On la note $\mathcal{A}_1 \otimes \mathcal{A}_2$.

Définition B.7.29. (Sections)
Soient X_1 et X_2 deux ensembles et $E \in X_1 \times X_2$. Pour $x_1 \in X_1$, on définit la *section* associée à X_1 par

$$E_{x_1} = \{x_2 \in X_2 \, , \; (x_1, x_2) \in E\}$$

On définit de la même manière, pour $x_2 \in X_2$, la section $E^{x_2} = \{x_1 \in X_1 \, , \; (x_1, x_2) \in E\}$.

16. Chacune des propriétés énoncées est vraie sur un ensemble dont le complémentaire est de mesure nulle. On exclut ici la réunion de tous ces ensembles sur lesquels les propriétés sont vérifiées, comme il s'agit d'une réunion dénombrable, cet ensemble reste de mesure nulle.

Proposition B.7.30. Soient $(X_1, \mathcal{A}_1)$ et $(X_2, \mathcal{A}_2)$ deux espaces mesurables. Soit $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Toute section E_{x_1} est dans $\mathcal{A}_2$, et toute section E^{x_2} est dans $\mathcal{A}_1$.

Démonstration. Soit $x_1 \in X_1$. On définit $\mathcal{F}$ comme l'ensemble des parties E de $X_1 \times X_2$ telles que E_{x_1} est élément de $\mathcal{A}_2$. Pour tout rectangle $E = A_1 \times A_2$, avec $A_i \in \mathcal{A}_i$, on a soit $E_{x_1} = \emptyset$ (si $x_1 \notin A_1$), soit $E_{x_1} = A_2$ (si $x_1 \in A_1$), d'où l'on déduit que $\mathcal{F}$ contient tous les rectangles $A_1 \times A_2$. On a par ailleurs, pour toute partie E de l'espace produit,

$$(E^c)_{x_1} = (E_{x_1})^c$$

et, pour toute collection (E_n),

$$\left(\bigcup E_n\right)_{x_1} = \bigcup (E_n)_{x_1},$$

d'où l'on déduit que $\mathcal{F}$ est stable par complémentarité et par union dénombrable. Il s'agit donc d'une tribu, qui contient donc la tribu engendrée par les rectangles, qui est $A_1 \otimes A_2$. Pour tout $E \subset \mathcal{A}_1 \otimes \mathcal{A}_2$, on a donc $E_{x_1} \in \mathcal{A}_2$. On démontre de la même manière que toute section E^{x_2} d'un ensemble $E \subset \mathcal{A}_1 \otimes \mathcal{A}_2$ est dans $\mathcal{A}_1$. □

Définition B.7.31. (Section d'une application)
Soit f une fonction définie sur un espace produit $X_1 \times X_2$. On note f_{x_1} la fonction (appelée *section*) définie sur X_2 par

$$f_{x_1}(x_2) = f(x_1, x_2).$$

On définit de la même manière $x_1 \longmapsto f^{x_2}(x_1) = f(x_1, x_2)$.

Proposition B.7.32. Soit f une application $\mathcal{A}_1 \otimes \mathcal{A}_2$ – mesurable à valeurs dans $[-\infty, +\infty]$, alors pour tout $x_1 \in X_1$, la section f_{x_1} est $\mathcal{A}_2$ – mesurable, et pour tout $x_2 \in X_2$, la section f_{x_2} est $\mathcal{A}_1$ – mesurable.

Démonstration. Pour tout $x_1 \in X_1$, tout $A_2 \in \mathcal{A}_2$, tout borélien D de $\overline{\mathbb{R}}$, on a

$$(f_{x_1})^{-1}(D) = (f^{-1}(D))_{x_1},$$

Or $f^{-1}(D) \in \mathcal{A}_1 \otimes \mathcal{A}_2$ d'après l'hypothèse de mesurabilité de f, et donc $(f^{-1}(D))_{x_1} \in \mathcal{A}_2$ d'après la proposition B.7.30. On montre de la même manière que, pour tout $x_2 \in X_2$, la section f_{x_2} est $\mathcal{A}_1$ – mesurable. □

Proposition B.7.33. Soient $(X_1, \mathcal{A}_1, \mu_1)$ et $(X_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés, tels que μ_1 et μ_2 sont σ – finies. Pour tout $E \subset \mathcal{A}_1 \otimes \mathcal{A}_2$, les applications

$$x_1 \longmapsto \mu_2(E_{x_1}) \text{ et } x_2 \longmapsto \mu_1(E^{x_2})$$

sont respectivement $\mathcal{A}_1$ – mesurable et $\mathcal{A}_2$ – mesurable.

Démonstration. On suppose dans un premier temps que μ_2 est finie. D'après la proposition B.7.30, pour tout $x_1 \in X_1$, tout $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$, la section E_{x_1} est dans $\mathcal{A}_2$, la quantité $\mu_2(E_{x_1})$ est donc bien définie. On introduit l'ensemble $\mathcal{D}$ des éléments E de $\mathcal{A}_1 \otimes \mathcal{A}_2$ tels que la fonction $x_1 \longmapsto \mu_2(E_{x_1})$ est $\mathcal{A}_1$ – mesurable. Nous allons montrer que $\mathcal{D}$ est une classe monotone qui contient le π – système des rectangles, dont nous déduirons que $\mathcal{D}$ est la tribu produit toute entière. Pour tout rectangle $E = A_1 \times A_2$, cette fonction s'écrit

$$\mu_2(E_{x_1}) = \mu_2(A_2)\, \mathbb{1}_{A_1}(x_1),$$

elle est donc μ_1 – mesurable. En particulier, $X_1 \times X_2 \in \mathcal{D}$. Si maintenant E et F sont dans $\mathcal{D}$, avec $E \subset F$, on a

$$\mu_2((F \setminus E)_{x_1}) = \mu_2(F_{x_1}) - \mu_2(E_{x_1}),$$

d'où la mesurabilité de $x_1 \longmapsto \mu_2((F \setminus E)_{x_1})$. Si maintenant (E_n) est une suite croissante d'éléments de $\mathcal{D}$, on a

$$\mu_2\left(\left(\bigcup E_n\right)_{x_1}\right) = \lim \mu_2\left((E_n)_{x_1}\right) = \sup \mu_2\left((E_n)_{x_1}\right),$$

qui est mesurable d'après la proposition B.7.2, page 177. L'ensemble $\mathcal{D}$ est donc une classe monotone, qui contient le π - système $\mathcal{R}$ des rectangles. Il contient donc la tribu engendrée par $\mathcal{R}$, qui est $\mathcal{A}_1 \otimes \mathcal{A}_2$ (définition B.7.28). Or $\mathcal{D}$ a été défini comme l'ensemble des parties E telles que $x_1 \longmapsto \mu_2(E_{x_1})$ est $\mathcal{A}_1$ – mesurable. Cette propriété est donc vraie pour tout $E \in \mathcal{A}$. On montre symétriquement que $x_2 \longmapsto \mu_1(E^{x_2})$ est $\mathcal{A}_2$ – mesurable pour tout $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$.

Si maintenant μ_2 est σ – finie, on introduit une partition (D_n) de X_2, constituée de parties de mesure finie (voir proposition B.3.4). Chacune des mesures μ_2^n définie par $\mu_2^n(A) = \mu_2(A \cap D_n)$ est donc finie. D'après ce qui précède, la fonction $x_1 \longmapsto \mu_2^n(E_{x_1})$ est μ_1 – mesurable, d'où

$$x_1 \longmapsto \mu_2(E_{x_1}) = \sum_{n=0}^{+\infty} \mu_2^n(E_{x_1})$$

est mesurable. On démontre de la même manière la propriété symétrique. □

Théorème B.7.34. (Mesure – produit)
Soient $(X_1, \mathcal{A}_1, \mu_1)$ et $(X_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés, avec μ_1 et μ_2 des mesures que l'on suppose σ – finies. Il existe une unique mesure sur $(X_1 \times X_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, appelée *mesure produit* de μ_1 et μ_2, notée $\mu_1 \otimes \mu_2$, telle que

$$(\mu_1 \otimes \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),$$

pour tous $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$. Cette mesure vérifie en outre, pour tout $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$,

$$(\mu_1 \otimes \mu_2)(E) = \int_{X_1} \mu_2(E_{x_1}) d\mu_1(x_1) = \int_{X_2} \mu_1(E^{x_2}) d\mu_2(x_2).$$

Démonstration. D'après la proposition B.7.33, les fonctions $x_1 \longmapsto \mu_2(E_{x_1})$ et $x_2 \longmapsto \mu_1(E^{x_2})$ sont respectivement $\mathcal{A}_1$ – mesurable et $\mathcal{A}_2$ – mesurable. On peut ainsi définir deux fonctions de $\mathcal{A}_1 \otimes \mathcal{A}_2$ dans $\mathbb{R}_+$ comme suit

$$(\mu_1 \otimes \mu_2)_1(E) = \int_{X_1} \mu_2(E_{x_1}) d\mu_1(x_1)\,,\ (\mu_1 \otimes \mu_2)_2(E) = \int_{X_2} \mu_1(E^{x_2}) d\mu_2(x_2).$$

On vérifie immédiatement que ce sont bien des mesures sur la tribu-produit $\mathcal{A}_1 \otimes \mathcal{A}_2$. Ces mesures prennent les mêmes valeurs sur les rectangles : pour tous $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$,

$$(\mu_1 \otimes \mu_2)_1(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) = (\mu_1 \otimes \mu_2)_2(A_1 \times A_2).$$

Elles s'identifient donc sur l'ensemble $\mathcal{R}$ des rectangles, qui constituent un π – système d'après la proposition B.7.27. La mesure μ_1 étant σ – finie, X_1 s'écrit comme union croissante dénombrable d'ensembles A_n^1 de mesure finie, de même X_2 est réunion croissante des A_n^2, avec $\mu_2(A_n^2) < +\infty$ pour tout n. L'union des $C_n = A_n^1 \times A_n^2$, recouvre donc $X_1 \times X_2$, et l'on peut utiliser le corollaire B.3.13, page 164, qui assure que ces mesures s'identifient sur la tribu engendrée par $\mathcal{R}$, qui est par définition la tribu-produit $\mathcal{A}_1 \otimes \mathcal{A}_2$. □

Exercice B.7.4. a) Soient f_1 et f_2 deux applications mesurables de $(X_1, \mathcal{A}_1, \mu_1)$ et $(X_2, \mathcal{A}_2, \mu_2)$ vers $(X_1', \mathcal{A}_1', \mu_1')$ et $(X_2', \mathcal{A}_2', \mu_2')$, respectivement. Montrer que l'application

$$F\ :\ (x_1, x_2) \longmapsto (f_1(x_1), f_2(x_2))$$

est mesurable pour les tribus produits sur les espaces d'arrivée et de départ.
b) On considère maintenant f_1 et f_2 deux applications mesurables de $(X, \mathcal{A}, \mu)$ vers $(X_1', \mathcal{A}_1', \mu_1')$ et $(X_2', \mathcal{A}_2', \mu_2')$, respectivement. Montrer que l'application

$$G\ :\ x \in X \longmapsto (f_1(x), f_2(x))$$

est mesurable pour les tribus produits sur les espaces d'arrivée et de départ.

Théorème B.7.35. (Fubini – Tonelli)
Soient $(X_1, \mathcal{A}_1)$ et $(X_2, \mathcal{A}_2)$ deux espaces mesurés, avec μ_1 et μ_2 des mesures σ-finies. Soit f une fonction $\mathcal{A}_1 \otimes \mathcal{A}_2$–mesurable de (X_1, X_2) dans $[0, +\infty]$. Alors, pour μ_1–presque tout x_1, la section f_{x_1} est $\mathcal{A}_2$ mesurable sur X_2 et pour μ_2–presque tout x_2, la section f^{x_2} est $\mathcal{A}_1$ – mesurable sur X_1, et l'on a

$$\begin{aligned}\int_{X_1\times X_2} f(x_1,x_2)\, d(\mu_1\otimes\mu_2) &= \int_{X_1}\left(\int_{X_2} f_{x_1}(x_2)\, d\mu_2(x_2)\right) d\mu_1(x_1)\\ &= \int_{X_2}\left(\int_{X_1} f^{x_2}(x_1)\, d\mu_1(x_1)\right) d\mu_2(x_2).\end{aligned}$$

Démonstration. On considère dans un premier temps le cas où f est la fonction indicatrice d'une partie $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Les sections f_{x_1} et f^{x_2} sont alors les fonctions indicatrices de E_{x_1} et E^{x_2}, respectivement :

$$f_{x_1}(x_2) = f(x_1,x_2) = \mathbb{1}_E(x_1,x_2) = \mathbb{1}_{E_{x_1}}(x_2)\,,\ \ f^{x_2}(x_1) = \mathbb{1}_{E^{x_2}}(x_1).$$

elles sont donc respectivement $\mathcal{A}_2$ – mesurable et $\mathcal{A}_1$ – mesurable d'après la proposition B.7.33, et l'on a

$$\int_{X_2} f_{x_1}(x_2) d\mu_2(x_2) = \mu_2(E_{x_1}) \text{ et } \int_{X_1} f^{x_2} d\mu_1(x_2) = \mu_1(E^{x_2}).$$

On a d'après le théorème B.7.34, qui définit la mesure-produit,

$$\begin{aligned}\int_{X_1}\left(\int_{X_2} f_{x_1} d\mu_2(x_2)\right) d\mu_1(x_1) &= \int_{X_1} \mu_2(E_{x_1}) d\mu_1(x_1)\\ &= (\mu_1\otimes\mu_2)(E)\\ &= \int_{X_2} \mu_1(E^{x_1}) d\mu_2(x_2)\\ &= \int_{X_2}\left(\int_{X_1} f^{x_2}(x_1) d\mu_1(x_1)\right) d\mu_2(x_2).\end{aligned}$$

La propriété est donc vérifiée pour les fonctions indicatrices d'éléments de $\mathcal{A}_1 \otimes \mathcal{A}_2$. Elle donc vérifiée, par linéarité de l'intégrale, pour les fonctions étagées. Or toute fonction mesurable sur $\mathcal{A}_1 \otimes \mathcal{A}_2$ est limite croissante d'une suite de fonctions étagées (proposition B.7.6, page 179). Pour toute fonction étagée g sur $\mathcal{A}_1 \otimes \mathcal{A}_2$, la section g_{x_1} est également étagée :

$$g(x_1,x_2) = \sum \alpha_i \mathbb{1}_{C_i}(x_1,x_2)\,,\ \ g_{x_1}(x_2) = \sum \alpha_i \mathbb{1}_{(C_i)_{x_1}}(x_2).$$

Pour toute suite croissante de telles fonctions, les sections sont également croissantes, et la convergence simple implique la convergence de toute section vers la section correspondante de la limite. La proposition B.7.2, page 177, assure la mesurabilité des sections. Le théorème B.7.23 de convergence monotone assure la convergence des intégrales, ce qui conclut la preuve. □

Théorème B.7.36. (Fubini – Lebesgue)
Soient μ_1 et μ_2 deux mesures σ–finies sur les espaces mesurables $(X_1, \mathcal{A}_1)$ et $(X_2, \mathcal{A}_2)$, respectivement. Soit f une fonction $\mathcal{A}_1 \otimes \mathcal{A}_2$–mesurable de (X_1, X_2) dans $[-\infty, +\infty]$. On suppose que f est intégrable pour la mesure produit $\mu_1 \otimes \mu_2$. Alors

(i) Pour μ_1–presque tout x_1, la section f_{x_1} est μ_2– intégrable sur X_2 et pour μ_2–presque tout x_2, la section f^{x_2} est μ_1 – intégrable sur X_1 ;

(ii) Les fonctions

$$x_1 \in X_1 \longmapsto I^1_f(x_1) = \left|\begin{array}{ll}\displaystyle\int_{X_2} f_{x_1}(x_2)\, d\mu_2 & \text{si } \ f_{x_1} \text{ est } \ \mu_2 - \text{intégrable}\\ 0 & \text{sinon}\end{array}\right.$$

et

$$x_2 \in X_2 \longmapsto I_f^2(x_2) = \left| \begin{array}{ll} \displaystyle\int_{X_1} f^{x_2}(x_1)\, d\mu_1 & \text{si } \ f^{x_2} \text{ est } \ \mu_1 - \text{intégrable} \\ 0 & \text{sinon} \end{array} \right.$$

sont respectivement μ_1 – intégrable et μ_2 – intégrable.

(iii) On a

$$\int_{X_1 \times X_2} f(x_1, x_2)\, d(\mu_1 \otimes \mu_2) = \int_{X_1} I_f^1(x_1) d\mu_1 = \int_{X_2} I_f^2(x_2) d\mu_2.$$

Démonstration. Soient f^+ et f^- les parties positive et négative de f. D'après la proposition B.7.32, les sections $(f^+)_{x_1}$, $(f^-)_{x_1}$ sont $\mathcal{A}_2$ mesurables. D'après le théorème B.7.35, les fonctions

$$x_1 \longmapsto \int_{X_2} (f^+)_{x_1}\, d\mu_2 \quad \text{et} \qquad x_1 \longmapsto \int_{X_2} (f^-)_{x_1}\, d\mu_2$$

sont $\mathcal{A}_1$– mesurables et μ_1–intégrables. D'après la proposition B.7.22, ces fonctions sont donc finies μ_1 presque partout. La section f_{x_1} est donc intégrable pour presque tout x_1. Soit N l'ensemble des x_1 tels que l'une ou l'autre des fonctions ci-dessus est infinie. L'ensemble N est dans $\mathcal{A}_1$ car

$$N = \left(\bigcap_n \left\{ x_1 \,,\ \int_{X_2} (f^+)_{x_1}\, d\mu_2 > n \right\} \right) \bigcup \left(\bigcap_n \left\{ x_1 \,,\ \int_{X_2} (f^-)_{x_1}\, d\mu_2 > n \right\} \right).$$

la fonction I_f^1 vaut 0 sur N, et prend la valeur

$$\int_{X_2} (f^+)_{x_1}\, d\mu_2 - \int_{X_2} (f^-)_{x_1}\, d\mu_2$$

sur son complémentaire. La fonction I_f^1 est donc μ_1–intégrable. On a donc, d'après le théorème B.7.35 et la proposition B.7.20, page 185,

$$\begin{aligned} \int_{X_1 \times X_2} f\, d(\mu_1 \otimes \mu_2) \quad &= \quad \int_{X_1 \times X_2} f^+\, d(\mu_1 \otimes \mu_2) - \int_{X_1 \times X_2} f^-\, d(\mu_1 \otimes \mu_2) \\ &= \quad \int_{X_1} \int_{X_2} (f^+)_{x_1}\, d\mu_2 - \int_{X_1} \int_{X_2} (f^-)_{x_1}\, d\mu_2 = \int_{X_1} I_f^1 d\mu_1. \end{aligned}$$

La même démarche appliquée aux sections $(f^+)^{x_2}$ et $(f^-)^{x_2}$ permet de conclure. □

B.8 Exercices

Exercice B.8.1. Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$, dérivable en tout point. Montrer que la dérivée de f est mesurable.

Exercice B.8.2. Soit f une fonction mesurable de $(X, \mathcal{A}, \mu)$ dans $\mathbb{R}$, avec μ mesure *finie*. Pour tout $n \in \mathbb{N}$, on introduit

$$A_n = \{x \in X \,,\ |f(x)| \geq n\} \ ,\ B_n = \{x \in X \,,\ |f(x)| \in [n, n+1[\} \,.$$

Montrer que les trois assertions suivantes sont équivalentes

(i) La fonction f est intégrable

(ii) La série $\sum n\mu(B_n)$ est convergente.

(iii) La série $\sum \mu(A_n)$ est convergente.

(On pourra montrer $(i) \Longleftrightarrow (ii)$ *et* $(ii) \Longleftrightarrow (iii)$*.)*

Exercice B.8.3. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et (f_n) une suite de fonctions mesurables à valeurs dans $\mathbb{R}+$, telle que

$$\int_X f_n(x)\, d\mu \leq M \quad \forall n.$$

On définit $f = \liminf f_n$. Montrer que

$$\int_X f(x)\, d\mu \leq M.$$

Exercice B.8.4. Soit f une fonction mesurable de $(X, \mathcal{A}, \mu)$ à valeurs dans $\mathbb{R}$, intégrable. Pour tout $n \in \mathbb{N}$, on introduit

$$A_n = \{x \in X\,,\ |f(x)| \geq n\}\,.$$

a) Montrer que

$$\lim_{n \to +\infty} \int_{A_n} |f(x)|\, d\mu = 0$$

b) Montrer que, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que

$$\forall A \in \mathcal{A} \text{ tel que } \ \mu(A) \leq \delta\,, \quad \text{on a } \int_A |f(x)|\, d\mu < \varepsilon.$$

(On pourra utiliser la décomposition d'une partie A en $A \cap A_n$ et $A \cap A_n^c$.)

Exercice B.8.5. (Intégrale dépendant d'un paramètre)
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré (par exemple $(\mathbb{R}, \mathcal{B}, \lambda)$), et f une application de $X \times I$ dans $\mathbb{R}$, où I est un intervalle de $\mathbb{R}$.

a) On suppose dans un premier temps que f vérifie les propriétés suivantes :
  (i) Pour tout $t \in I$, la fonction $x \longmapsto f(x,t)$ est mesurable.
  (ii) Pour tout $x \in X$, la fonction $t \longmapsto f(x,t)$ est continue.
  (iii) Il existe une fonction g sur X, intégrable, telle que pour tout $(x,t) \in X \times I$,

$$|f(x,t)| \leq g(x).$$

Montrer que l'application

$$F\ : t \in I \longmapsto \int_X f(x,t)\, d\mu(x)$$

est bien définie et continue sur I.

b) On renforce les hypothèses sur f de la façon suivante :
  (iv) Pour tout $x \in X$, la fonction $t \longmapsto f(x,t)$ est continûment différentiable.
  (v) Il existe une fonction h sur X, intégrable, telle que pour tout $(x,t) \in X \times I$,

$$\left|\frac{\partial f}{\partial t}(x,t)\right| \leq h(x).$$

Montrer que l'application F définie ci-dessus est bien définie et dérivable sur $\mathbb{R}$, de dérivée

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

Exercice B.8.6. Soient $(X_1, \mathcal{A}_1)$ et $(X_2, \mathcal{A}_2)$ deux espaces mesurables, et γ une mesure sur $\mathcal{A}_1 \otimes \mathcal{A}_2$. On note π_1 la projection sur X_1, définie par $\pi_1(x_1, x_2) = x_1$. On appelle première marginale de γ la mesure transportée $\mu_1 = (\pi_1)_\sharp \gamma$ (voir exercice B.6.1). On définit de même la deuxième marginale μ_2 sur $\mathcal{A}_2$.

a) Pour $A_i \in \mathcal{A}_i$, donner l'expression de $\mu_i(A_i)$, et montrer que μ_1 et μ_2 ont même masse totale. A-t-on en général $\gamma = \mu_1 \otimes \mu_2$?

Si γ est la loi d'une variable aléatoire $(Y_1, Y_2) \in X_1 \times X_2$, interpréter μ_1 et μ_2.

b) Dans le cas où $X_1 = X_2 = [\![1, N]\!]$, munis de leurs tribus discrètes, montrer que toute mesure γ sur $\mathcal{A}_1 \otimes \mathcal{A}_2$ peut se représenter par une matrice carrée $A \in \mathcal{M}_N(\mathbb{R})$, et préciser comment construire μ_1 et μ_2 à partir de A.

Introduction au transport optimal

c) Pour μ_1 et μ_2 mesures de même masse finie $M > 0$ sur $\mathcal{A}_1$ et $\mathcal{A}_2$, respectivement, on définit

$$\Pi_{\mu_1,\mu_2} = \{\gamma \in \mathcal{M}_M(X_1 \times X_2)\,,\ (\pi_i)_\sharp\gamma = \mu_i\,,\ i = 1, 2\}\,,$$

où $\mathcal{M}_M(X_1 \times X_2)$ est l'espace des mesures sur $X_1 \times X_2$ de masse M. Montrer que $\Pi_{\mu,\nu}$ est non vide.

d) On se place dans le cas $X_1 = X_2 = \mathbb{R}^d$, muni de la tribu borélienne, et l'on considère μ_1 et μ_2 des mesures sur $\mathbb{R}^d$ définies par

$$\mu_1 = \sum_{i=1}^{n} \alpha_i \delta_{x_i}, \mu_2 = \sum_{j=1}^{m} \beta_j \delta_{y_j}\,,\ \sum \alpha_i = \sum \beta_j = M\,,\ \alpha_i \geq 0\,,\ \beta_j \geq 0.$$

Décrire l'ensemble Π_{μ_1,μ_2}. Dans quel cas cet ensemble est-il réduit à un singleton ?

e) On se place dans le cadre de la question précédente, et l'on se donne une collection $(c_{ij}) \in \mathbb{R}_+^{n\times m}$ de *coûts*. Pour tout couple (i, j), le nombre c_{ij} correspond à ce que coûte le transport d'une quantité unitaire de matière de x_i vers y_j. On pourra prendre $c_{ij} = \|y_j - x_i\|$ pour fixer les idées.

Montrer que le problème

$$\min_{\gamma \in \Pi_{\mu_1,\mu_2}} \sum_{i,j} \gamma_{ij}\, c_{ij}$$

admet une solution. Cette solution est-elle unique en général ?

f) ($\star$) Imaginer une situation de la vie réelle (dans un contexte de logistique de transport), où les x_i correspondraient à des lieux de production, et les y_j des lieux de vente ou de consommation. Discuter du choix des (c_{ij}) en terme de pertinence.

g) ($\star$) On considère une population de n employés sur une période donnée, et l'on note μ_i le temps de travail de l'employé i durant cette période. On considère qu'il y a m tâches à accomplir, associée chacune, pour fixer les idées, à une machine j, et l'on suppose que le temps de disponibilité de la machine durant la période considérée est égal à ν_j. On note u_{ij} la *productivité* de l'employé i vis-à-vis de la tâche j, de telle sorte que $u_{ij}\mu_i$ est la valeur ajoutée résultant du travail de i à la tâche j pendant le temps μ_i. On suppose que le temps de travail total est égal au temps total de disponibilité des machines (on ne se préoccupera pas des questions de répartition effective des tâches durant la période de temps considérée, en supposant par exemple que cette période est très grande). Montrer que chercher à maximiser la valeur ajoutée totale conduit à un problème d'*affectation* du type de celui étudié dans les questions précédentes.

Exercice B.8.7. Soit f une fonction d'un espace mesuré $(X, \mathcal{A}, \mu)$ dans $\mathbb{R}_+$, mesurable.

a) On munit $X \times \mathbb{R}_+$ de la tribu produit. Montrer que la fonction F à valeurs dans $\mathbb{R}$ qui à (x, t) associe $f(x) - t$ est mesurable.

b) Montrer que la fonction $\mathbb{1}_{\{(x,t), f(x) \geq t\}}$ est mesurable sur $X \times \mathbb{R}_+$.

c) Montrer que

$$\int_X f(x)\, d\mu = \int_0^{+\infty} \mu(\{(x\,,\ f(x) \geq t\})\, dt.$$

d) Interpréter graphiquement l'identité de la fonction précédente dans le cas où X est un intervalle réel et f une fonction régulière. Expliquer en quoi cette formule suggère une méthode numérique d'estimation effective de l'intégrale d'une fonction régulière de $\mathbb{R}$ dans $\mathbb{R}_+$, que l'on pourrait appeler méthode des "rectangles horizontaux".

Exercice B.8.8. Soit f une fonction intégrable de $(X, \mathcal{A}, \mu)$ dans $\mathbb{R}$. Montrer que presque tous les ensembles de niveaux sont de mesure nulle, c'est-à-dire que, pour presque tout t réel, on a

$$\mu\left(\{x\,,\ f(x) = t\}\right) = 0.$$

(On pourra s'inspirer de la démarche proposée à l'exercice B.8.7).

Exercice B.8.9. On considère la fonction définie sur $[-1,1]\times[-1,1]\setminus\{(0,0)\}$ par

$$f(x,y)=\frac{x^2-y^2}{(x^2+y^2)^2}.$$

a) Montrer que

$$\int_{-1}^{1}\left(\int_{-1}^{1} f(x,y)\,dx\right)dy \neq \int_{-1}^{1}\left(\int_{-1}^{1} f(x,y)\,dy\right)dx$$

b) Expliquer en quoi cette propriété n'est pas en contradiction avec le théorème de Fubini-Lebesgue.
(On pourra montrer que f n'est pas intégrable au voisinage de 0 en calculant l'intégrale de $|f|$ sur la couronne $\left\{(x,y)\,,\ \varepsilon<\sqrt{x^2+y^2}<1\right\}$.)

Index