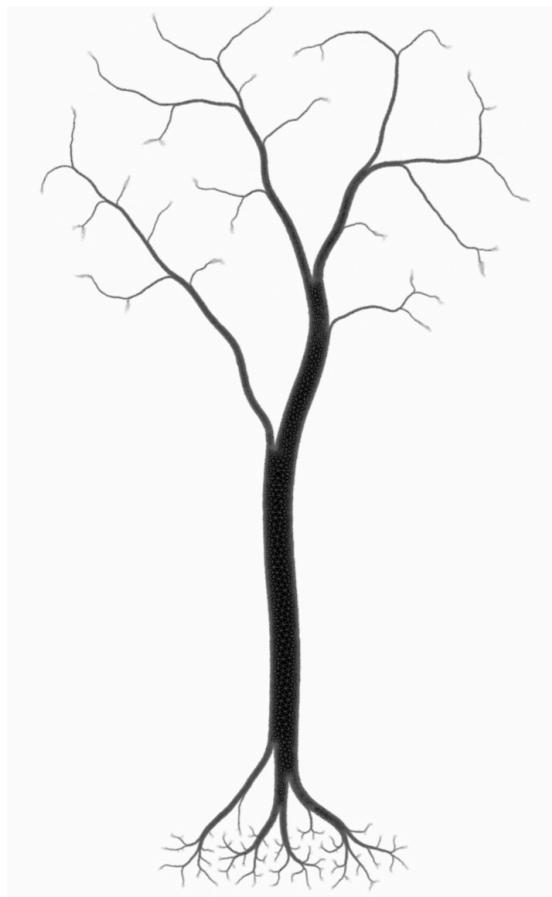


UE 11
TOPOLOGIE, ESPACES FONCTIONNELS,
CALCUL DIFFÉRENTIEL



ÉCOLE DES MINES PSL

B. MAURY

AVANT-PROPOS

Ce document accompagne un cours donné aux élèves de première année de l’École des Mines de Paris. Il s’agit d’un cours de mathématiques générales, orienté vers l’analyse. Il commence par des notions de topologie dans les espaces métriques, qui sont à la base de tous les développements en analyse, et qui sont essentiellement des rappels pour les élèves provenant de filières les plus chargées en mathématiques.

Le chapitre II est principalement dédié à la construction des espaces fonctionnels qui permettent d’aborder de façon rigoureuse et efficace l’étude des équations aux dérivées partielles. Il s’agit de construire des espaces de fonctions qui vérifient de bonnes propriétés, en particulier la *complétude*, de façon à ce qu’ils puissent permettre de conduire à des résultats d’existence de solutions à des problèmes d’optimisation et des problèmes de type équations aux dérivées partielles (équation de transport, de la chaleur, des ondes, équation de Navier-Stokes pour la modélisation des fluides visqueux, équation de Schrödinger en mécanique quantique, ...). Les espaces de fonctions continues (ou continûment différentiables) sont construits dans un premier temps, comme espaces vectoriels normés complets. Malgré leur caractère “naturel” (ils sont construits sur la notion de convergence uniforme, en continuité avec le programme de classes préparatoire), ils sont inadaptés à la plupart des problèmes venant de la physique. Le sup de la valeur absolue d’un champ n’est en effet pas, en général, une quantité physique pertinente, même s’il peut avoir du sens en sûreté des procédés, par exemple. Il est essentiel de construire des normes sur les espaces fonctionnels qui correspondent à des quantités pertinentes physiquement, comme une quantité d’énergie thermique qui se conserve pour l’équation de la transport, ou l’énergie cinétique d’un fluide qui a tendance à décroître sans forçage extérieur. Ces quantités sont basées sur des intégrales de fonctions sur un domaine. Nous montrons dans la suite du chapitre II que l’intégrale de Riemann, qui garde tout son intérêt dans certains cas, ne permet pas de définir des espaces fonctionnels munis de bonnes propriétés (en particulier la complétude). Cette impasse a conduit il y plus d’un siècle à l’élaboration d’une nouvelle forme d’intégration, dite de *Lebesgue*, qui permet la construction d’espaces fonctionnels adaptés à la représentation de champs physiques, et à la résolution numérique étayée des équations aux dérivées partielles. cette construction s’appuie sur la théorie de la mesure, qui repose elle-même sur la notion de *tribus*. Nous décrivons les grandes étapes de cette construction dans ce même chapitre II, en reportant en annexe la démarche détaillée (annexe B), notamment pour ceux qui souhaiteraient poursuivre dans l’étude des mathématiques fondamentales. La théorie de la mesure constitue également le socle de la théorie des probabilités. Nous encourageons donc les lecteurs à intégrer les éléments essentiels. Cette démarche nous conduit à la construction des espaces fonctionnels de type L^p , qui termine ce chapitre.

Le cas $p = 2$ est d’un intérêt particulier, puisqu’un grand nombre de quantités physiques sont basées sur l’intégrale du carré d’un champ ou du gradient d’un champ (énergie cinétique d’un fluide ou d’un solide déformable, énergie de déformation d’un solide déformable, intégrale du module au carré de la fonction d’onde comme densité de probabilité de présence d’une particule quantique, ...). Ces espaces fonctionnels, appelés *espaces de Hilbert*, peuvent être vus comme des extensions à la dimension infinie des espaces euclidiens. Le chapitre III présente leurs principales propriétés d’un point de vue abstrait. Nous verrons en particulier que la projection sur un convexe fermé est bien définie, ce qui a une importance considérable en optimisation, et que la plupart de ces espaces admettent ce que l’on appelle une base hilbertienne, qui permet de formaliser de façon très féconde certains problèmes (en traitement du signal, EDP de type chaleur ou ondes, équation de Schrödinger...). Le chapitre IV, en bordure du programme de ce cours, propose une introduction à l’étude générale des espaces vectoriels normés complets (appelés *espaces de Banach*), qui correspond au cas des espaces de fonctions continues, et aux espaces évoqués ci-dessus de type L^p pour $p \neq 2$.

Le chapitre V revient sur les fonctions régulières : calcul différentiel, dérivées partielles dans \mathbb{R}^d et notion de différentielle, théorème des fonctions implicites, théorème d’inversion locale, dérivées d’ordre supérieur (matrice hessienne, opérateur laplacien).

Ce cours s’adresse à des élèves-ingénieurs, et l’auteur de ses lignes a une activité de recherche tournée vers les mathématiques appliquées. Il nous a pourtant paru pertinent de construire ce cours non pas comme une suite de techniques, voire de “recettes”, à appliquer directement, mais de façon à faire ressortir la construction de concepts abstraits et généraux. Outre le plaisir toujours renouvelé de jouer avec des concepts abstraits et parfois “inutiles” (au sens noble du terme), il nous paraît en effet essentiel de bien maîtriser la nature et l’origine des concepts, du fait que la démarche de *modélisation mathématique*, qui repose notamment sur une

démarche de formalisation mathématique de problèmes réels, exige une bonne maîtrise des concepts abstraits pré-existants, et la capacité à en construire de nouveaux si cela s'avère pertinent.

Table des matières

I Topologie des espaces métriques	9
I.1 Motivations, vue d'ensemble	9
I.2 Distance, espace métrique	11
I.3 Topologie des espaces métriques	16
I.4 Suites	21
I.5 Complétude	22
I.6 Compacité	23
I.7 Applications entre espaces métriques	27
I.8 Compléments	29
I.8.1 Théorème de point fixe de Banach	29
I.8.2 Compléments sur les e.v.n. de dimension finie	30
I.9 Exercices	34
II Espaces fonctionnels, théorie de la mesure et de l'intégration	45
II.1 Convergences simple & uniforme	45
II.2 Espaces de fonctions continues	47
II.3 De Riemann à Lebesgue	48
II.3.1 L'intégrale de Riemann et ses limites	48
II.3.2 Théorie de la mesure et intégrale de Lebesgue : un aperçu	51
II.3.3 Mesure et intégration	52
II.4 Les espaces L^p	60
II.4.1 L'espace $L^\infty(X)$	61
II.4.2 Les espaces $L^p(X)$, pour $p \in [1, +\infty[$	62
II.4.3 Les espaces $L^p(\mathbb{N}) = \ell^p$ et $L^p(\mathbb{R}^d)$	64
II.5 Compléments	66
II.6 Exercices	67
III Espaces de Hilbert	75
III.1 Définitions, principales propriétés	75
III.2 Convergence faible	80
III.3 Sommes hilbertiennes, bases hilbertiennes	81
III.4 Théorie de Lax-Milgram	83
III.5 Opérateurs	85
III.6 Exercices	87

IV Éléments d'analyse fonctionnelle	91
IV.1 Définitions	91
IV.2 Compléments sur la dualité	93
V Calcul Différentiel	97
V.1 Dérivées partielles, notion de différentielle	97
V.1.1 Définitions, premières propriétés	97
V.1.2 Compléments	104
V.1.3 Théorème fondamental de l'analyse	109
V.1.4 Formules de changement de variable pour les transformations régulières	111
V.2 Exercices	111
V.3 Théorèmes des fonctions implicites et d'inversion locale	119
V.4 Problème adjoint	124
V.5 Exercices	128
V.6 Dérivées d'ordre supérieur	134
V.6.1 Dérivées partielles d'ordre supérieur pour les fonctions scalaires	134
V.6.2 Différentielles d'ordre supérieur pour les fonctions de \mathbb{R}^n dans \mathbb{R}^m	137
V.7 Exercices	138
VI Développements	141
VI.1 Entiers p -adiques, espaces ultramétriques	141
VI.2 Dendrogrammes	149
VI.3 Introduction au transport optimal	151
VI.3.1 Problème d'affectation et problème de Monge Kantorovich discret	151
VI.3.2 Transport optimal, cas général	153
VI.4 Distance de Gromov-Wasserstein	154
VI.5 Propagation d'opinion et flot de gradient	155
VI.6 Modèles macroscopiques de trafic routier	161
VI.7 Autour de la notion de complétude, théorème de Banach-Steinhaus	164
VI.7.1 Lemme de Baire	164
VI.7.2 Théorème de Banach Steinhaus	165
VI.7.3 Théorème de l'application ouverte, théorème du graphe fermé	166
A Fondamentaux et compléments	169
A.1 Fondamentaux	169
A.1.1 Éléments de théorie des ensembles	169
A.1.2 Structures fondamentales : relations et structures algébriques	170
A.1.3 Cardinalité	172
A.1.4 L'ensemble des réels : construction et structures afférentes	178
A.1.5 Inégalités fondamentales	183
A.2 Pour aller plus loin (••••)	185
A.2.1 Théorie des ensembles, cardinalité	185
A.2.2 Complété d'un espace métrique (••••)	185
A.2.3 Topologie générale (••••)	186

B Compléments sur la théorie de la mesure et de l'intégration	189
B.1 Motivations, vue d'ensemble	189
B.2 Tribus, espaces mesurables	193
B.2.1 Tribus	193
B.2.2 Applications mesurables	197
B.2.3 Classes monotones	198
B.3 Mesures	200
B.4 Mesures extérieures	204
B.4.1 Définitions, premières propriétés	204
B.4.2 D'une mesure extérieure à une mesure	205
B.4.3 Mesure de Lebesgue	207
B.5 Compléments	210
B.6 Exercices	213
B.7 Fonctions mesurables, intégrale de Lebesgue	219
B.7.1 Fonctions mesurables	219
B.7.2 Intégrale de fonctions étagées	222
B.7.3 Intégrale de fonctions mesurables	225
B.7.4 Théorèmes fondamentaux	228
B.7.5 Intégrales multiples	230
B.8 Exercices	234

Chapitre I

Topologie des espaces métriques

Sommaire

I.1	Motivations, vue d'ensemble	9
I.2	Distance, espace métrique	11
I.3	Topologie des espaces métriques	16
I.4	Suites	21
I.5	Complétude	22
I.6	Compacité	23
I.7	Applications entre espaces métriques	27
I.8	Compléments	29
I.8.1	Théorème de point fixe de Banach	29
I.8.2	Compléments sur les e.v.n. de dimension finie	30
I.9	Exercices	34

I.1 Motivations, vue d'ensemble

Les sections qui suivent portent pour l'essentiel sur des questions de topologie dite métrique, c'est-à-dire sur la manière dont une *distance* structure un ensemble. Cette notion est très intuitive, du fait que la distance euclidienne de l'espace physique \mathbb{R}^3 , ou tout du moins l'ordre de grandeur de la distance entre deux objets, est directement accessible aux sens. Il est néanmoins intéressant, y compris pour appréhender le monde réel, d'aborder cette notion d'un point de vue général, abstrait. En effet, la modélisation d'un très grand nombre de phénomènes réels repose sur la définition d'une métrique adaptée. Ce rôle central joué par la métrique s'est encore accru ces dernières années avec l'explosion de la science des données. Nous décrivons ci-dessous quelques exemples de situations dans lesquelles une part essentielle de la démarche réside dans la définition d'une distance adaptée, en nous limitant aux 3 grandes classes de contextes qui nous paraissent essentielles.

1) En premier lieu une distance a vocation à structurer un espace donné, et sa définition dépend du regard que l'on souhaite porter sur cet espace, et des facteurs que l'on souhaite prendre en compte. Ainsi, pour deux points d'une zone géographique représentée sur un plan, la distance euclidienne canonique correspond à la distance “à vol d'oiseau” du langage commun. Mais si l'on s'intéresse à une notion de proximité respectueuse de la difficulté effective de se rendre d'un point à un autre, il peut être pertinent de privilégier par exemple le temps qu'il faut pour se rendre du point x au point y , selon une modalité donnée, en prenant en compte les contraintes associées à la topographie du terrain. Un exemple archétypal de cette approche est la célèbre *distance de Manhattan*, appelée aussi *distance du chauffeur de taxi*, adaptée à une zone géographique dans laquelle les axes de circulation sont structurés en réseau orthogonal. À l'échelle d'un pays comme la France, il peut être fécond de définir la distance entre deux points comme le temps minimal qu'il faut pour aller d'un point à un autre en utilisant les transports en communs (complétés par de la marche à pied au départ et à l'arrivée). L'espace métrique associé (ce terme est défini plus loin, il s'agit simplement de la carte munie de la

distance que l'on vient de définir) est difficile à représenter graphiquement, mais il est une représentation plus fidèle et féconde du territoire si l'on s'intéresse à la manière dont le réseau de transport structure l'espace. On pourra considérer par exemple le “disque” (dont la forme peut être très éloignée de l'image que l'on se fait d'un disque) centré en la position d'une entreprise et de rayon une demie-heure, qui couvrira la zone dans laquelle habiteront les salariés qui souhaitent garder leur temps de déplacement journalier en dessous de cette durée. On pourra aussi considérer la distance associée au temps de parcours sur route, et s'intéresser par exemple à la part du territoire constituée des points situés à moins d'une demie-heure d'un hôpital ou d'une maternité. Ce type d'approche conduit assez naturellement à des problèmes délicats d'optimisation, comme par exemple : comment positionner de façon optimale des centres de soin dans un territoire de façon à minimiser la distance (= temps) maximale pour se rendre dans l'un de ces centres ? Ces problèmes complexes et la forme de leurs solutions dépendent étroitement de la distance choisie.

2) Une distance a également vocation à quantifier une différence, un écart, entre deux entités de même nature. Prenons l'exemple d'une collection d'images. Une image (disons carrée, et en noir et blanc) d'une résolution donnée peut être vue comme une matrice $N \times N$ de valeurs de l'intervalle $[0, 1]$ (niveaux de gris). La proximité entre deux images (par exemple parmi des images représentant des formes géométriques, ou des objets de la vie courante) peut être estimée assez efficacement par un enfant de 3 ans, mais il est extrêmement délicat de définir une métrique adaptée à cette situation, et la réponse peut dépendre du type d'images que l'on considère. On pourra se convaincre rapidement qu'identifier des images à des vecteurs de $[0, 1]^{N^2}$, et utiliser la distance euclidienne standard sur \mathbb{R}^{N^2} , est très loin d'être pertinent. Dans un contexte différent, on peut penser à une population d'individus à laquelle on associe un vecteur de caractéristiques (poids, âge, taille, taux de diverses substances dans le sang, ...). Structurer la population en catégories pertinentes passe par la définition d'une distance : peut-on associer à tout couple de personnes (représentées ici par leurs vecteurs de paramètres) un nombre qui quantifie leur éloignement ? Là encore une réponse pertinente à cette question peut être assez éloignée de la distance euclidienne. Un écart de 2 mois entre individus peut ainsi être considéré comme plus significatif s'il s'agit de nouveaux-nés que s'il s'agit d'adultes en pleine force de l'âge. Il peut être aussi pertinent de prendre en compte des couplages forts entre composantes, ce que la distance euclidienne ne permet pas. À titre d'exemple, une différence de poids de 5 kg est plus significative pour des nouveaux-nés que pour des quinquagénaires. Ces questions sont également cruciales dans le contexte de l'apprentissage statistique supervisé : l'apprentissage passe par une étape de minimisation d'une fonctionnelle de coût (appelée fonction *loss*), basée sur la définition d'une distance entre ce qui est produit par la fonction prédictive et la sortie provenant des données labellisées de la base d'apprentissage, et l'efficacité de l'approche repose en grande partie sur le choix de la distance.

3) L'expression des lois de la physique (systèmes de particules, mécanique des fluides, du solide, électromagnétisme, mécanique quantique), ou la modélisation de phénomènes réels (en biologie, dynamique des populations, économie, ...) conduit à des équations différentielles ordinaires (EDO) ou à des équations aux dérivées partielles (EDP), qui font intervenir dans le premier cas (EDO) des fonctions d'un intervalle en temps dans \mathbb{R}^d , dans le second cas (EDP) des fonctions à la fois du temps et d'une variable d'espace dans \mathbb{R}^d . L'analyse théorique de ces équations nécessite une structuration de l'espace dans lequel vivent leurs solutions. Il s'agit d'espaces de dimension infinie¹ sur lesquels il s'agit de définir une distance. On choisit en général une distance qui respecte la structure linéaire, construite à partir de ce que l'on appellera une *norme*. L'étude de tels espaces, qui constitue ce que l'on appelle l'*Analyse Fonctionnelle*, repose sur les notions qui sont introduites dans les sections qui suivent (espace vectoriel normé, ouverts et fermés, suites, complétude, compacité).

Dans un contexte plus directement lié aux sciences de l'ingénieur, la résolution numérique des problèmes de type EDO ou EDP mentionnés ci-dessus passe par des approximations : on effectue ce que l'on appelle une *discrétisation* du temps, en subdivisant l'intervalle continu en petits tronçons, qui vont permettre de transformer le problème de départ, qui est en dimension infinie, en un problème en dimension finie. Une approche analogue est faite en espace dans le cas des EDP². Étudier ces méthodes d'approximation (ce

1. On ne peut par exemple pas décrire l'ensemble des fonctions continues d'un intervalle en temps $[0, T]$ dans \mathbb{R}^d comme combinaison linéaire d'un nombre fini de fonctions particulières. C'est encore pire, si l'on ose dire, dans le cas des EDP, où l'on est amené à considérer, par exemple pour représenter un champ scalaire dynamique comme une température, des applications de $\mathbb{R} \times \mathbb{R}^d$ dans \mathbb{R} .

2. Il s'agit là d'un domaine extrêmement vaste, qui dépasse largement le cadre de ce cours. Citons simplement les trois grandes méthodes de discrétisation en espace : différences finies, éléments finis (très utilisés par exemple en mécanique du

qu'on appelle l'*Analyse Numérique*) consiste à montrer que les solutions approchées *convergent* vers la solution exacte de l'équation. Cette dernière vivant dans un espace de dimension infinie, sur lequel toutes les normes ne sont pas équivalentes, le choix de la norme conditionne encore de façon essentielle la nature du résultat de convergence.

I.2 Distance, espace métrique

Notions principales : distance, espace métrique, boules ouvertes et fermées, sphères, norme, espace vectoriel normé.

Définition I.2.1. (Distance, espace métrique (•))

Soit X un ensemble non vide. On appelle distance (ou métrique) sur X une application de $X \times X$ dans \mathbb{R}_+ vérifiant les propriétés suivantes :

- (i) (*Séparation*) Pour tous x, y dans X , on a $d(x, y) = 0 \iff x = y$.
- (ii) (*Symétrie*) $d(x, y) = d(y, x)$ pour tous x, y dans X .
- (iii) (*Inégalité triangulaire*) Pour tous x, y, z dans X , on a

$$d(x, z) \leq d(x, y) + d(y, z).$$

Le couple (X, d) est appelé espace métrique.

Exercice I.2.1. (•) On considère l'ensemble fini $\llbracket 1, N \rrbracket$, pour N entier ≥ 1 . À une distance sur X on peut associer une matrice carrée $D = (d_{ij}) \in \mathcal{M}_N(\mathbb{R})$, avec $d_{ij} = d(i, j)$. Décrire l'ensemble \mathcal{D} des matrices $D \in \mathcal{M}_N(\mathbb{R})$ qui correspondent à une métrique sur X . Quelles sont les propriétés de cet ensemble ?

CORRECTION.

Une matrice est dans \mathcal{D} si et seulement si :

1. ses éléments diagonaux sont nuls, tous les autres étant strictement positifs ;
2. elle est symétrique : $d_{ij} = d_{ji}$ pour tous $i \neq j$;
3. ses coefficients vérifient les inégalités triangulaires : pour tout triangle (i, j, k) dans X^3 , on a

$$d_{ij} \leq d_{ik} + d_{kj}.$$

L'ensemble \mathcal{D} est un cône de sommet 0 (on a $D \in \mathcal{D} \implies \lambda D \in \mathcal{D}$ pour tout $\lambda > 0$) épointé (il ne contient pas son sommet). Ce cône est par ailleurs convexe : pour tous D_0, D_1 dans \mathcal{D} , tout $t \in [0, 1]$, on a

$$(1 - \lambda)D_0 + \lambda D_1 \in \mathcal{D}.$$

Noter que l'on peut associer à ce cône un ordre partiel sur l'ensemble des matrices carrées, défini par

$$A < B \iff B - A \in \mathcal{D}.$$

Vocabulaire afférent aux espaces métriques

Définition I.2.2. (Diamètre (•))

Soit (X, d) un espace métrique, et $A \subset X$ une partie non vide de X . On appelle diamètre de A le nombre (potentiellement égal à $+\infty$)

$$\text{diam}(A) = \sup_{x, y \in A} d(x, y) \in [0, +\infty].$$

solide), et volumes finis (très adaptés à la discrétisation de lois de conservation dites hyperboliques, utilisés par exemple pour modéliser les écoulements compressibles). Selon des procédures diverses, chacune de ces méthodes est basée sur un espace de dimension finie dont les éléments (fonctions particulières, par exemple affines par morceaux pour certaines méthodes d'éléments finis), ont vocation à *approcher* les solutions des systèmes de départ.

Définition I.2.3. (Distance à une partie (•))

Soit (X, d) un espace métrique, $A \subset X$ non vide, et $x \in X$. On définit la distance de x à A par

$$d(x, A) = \inf_{y \in A} d(x, y).$$

On dit que cette distance est *atteinte* s'il existe $z \in A$ tel que $d(x, A) = d(x, z)$.

Exercice I.2.2. On se place sur $X = \mathbb{R}$ muni de la distance usuelle. Donner un exemple de partie $A \subset X$ pour laquelle la distance à A est atteinte pour n'importe quel point $x \in \mathbb{R}$, et un exemple de partie pour laquelle cette distance n'est atteinte que pour les points de A lui-même.

CORRECTION.

Si A est un singleton, ou une collection finie de points, ou un intervalle de type $[a, b]$, la distance est toujours atteinte. En revanche si l'on prend $A =]0, 1[$, la distance $d(x, A)$ n'est atteinte que pour $x \in A$.

Exercice I.2.3. On se place sur un espace métrique X . Montrer que la fonction

$$x \in A \longmapsto d(x, A)$$

ne caractérise pas A en général, en donnant un exemple de $A \subset X$ pour lequel il existe d'autres ensembles conduisant à la même fonction distance. Montrer en particulier que cette fonction peut être identiquement nulle sans que A ne s'identifie à X .

CORRECTION.

La connaissance de la fonction distance à A dit beaucoup de chose sur A , mais ne suffit pas à le déterminer, comme le suggère l'exercice précédent : les intervalles $[0, 1]$, $[0, 1[$, $]0, 1[$, $]0, 1]$ ont la même "signature" en termes de fonction distance, ils sont pourtant différents. Comme on le verra, la distance permet d'identifier l'adhérence de A .

Pour $A = \mathbb{D} \subset X = \mathbb{R}$, la distance à A est identiquement nulle (on peut approcher avec une précision arbitraire un réel par un décimal). De façon plus générale, comme on le verra, la distance est nulle dès que A est dense dans X .

Définition I.2.4. (Boules ouvertes, boules fermées, sphères (•))

Soit (X, d) un espace métrique, on appelle boule ouverte de centre x et de rayon $r \geq 0$ l'ensemble

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

La boule fermée, notée³ $B_f(x, r)$, est obtenue en remplaçant l'inégalité stricte par l'inégalité large $d(x, y) \leq r$. La sphère de centre x et de rayon r est l'ensemble des points situés à distance r de x :

$$S(x, r) = \{y \in X, d(x, y) = r\}.$$

Définition I.2.5. (Ensemble discret (•))

Soit (X, d) un espace métrique, et A une partie de X . On dit que l'ensemble A est *discret* si pour tout élément x de A il existe une boule de centre x qui ne contient pas d'autre élément de A :

$$\forall x \in A, \exists \varepsilon > 0, B(x, \varepsilon) \cap A = \{x\}.$$

Remarque I.2.6. On prendra garde au fait que, dans la définition ci-dessus, le ε dépend x , de telle sorte que l'ensemble $\{1/n, n \in \mathbb{N}\} \subset \mathbb{R}$ est discret, bien que deux points distincts de cet ensemble puissent être arbitrairement proches.

3. On trouvera parfois la notation $\overline{B}(x, r)$ pour désigner la boule fermée de centre x et de rayon r . Dans \mathbb{R}^d , la boule fermée $B_f(x, r)$ s'identifie en effet à l'adhérence (au sens précisé plus loin) de la boule ouverte $B(x, r)$. Il est cependant préférable d'éviter cette notation, du fait que cette correspondance peut être invalidée. Considérer par exemple l'espace métrique \mathbb{N} muni de la distance canonique. La boule fermée $B_f(0, 1)$ de centre 0 et de rayon 1 s'identifie à $\{-1, 0, 1\}$, la boule ouverte $B(0, 1)$ est réduite à $\{0\}$, dont l'adhérence est $\{0\}$.

Espaces métriques particuliers : les espaces vectoriels normés

Lorsque l'ensemble considéré a une structure d'espace vectoriel, il est en général fécond de choisir des distances particulières qui respectent la structure linéaire sous-jacente. On souhaite en particulier que la distance entre deux points x et y ne dépende que du vecteur $y - x$. Ces distances particulières sont construites à partir de *normes*, selon $d(u, v) = \|v - u\|$, où $\|\cdot\|$ est une norme selon la définition qui suit.

Définition I.2.7. (Norme, espace vectoriel normé (\bullet))

Soit E un espace vectoriel. On appelle norme une application de E dans \mathbb{R}_+ , notée $u \mapsto \|u\|$ qui vérifie les propriétés suivantes :

- (i) (*Séparation*) Pour tout x dans E , on a $\|x\| = 0 \iff x = 0$.
- (ii) (*Homogénéité*) Pour tout $x \in E$, tout $\lambda \in \mathbb{R}$, $\|\lambda x\| = |\lambda| \|x\|$.
- (iii) (*Inégalité triangulaire*) Pour tous x, y dans E , on a

$$\|x + y\| \leq \|x\| + \|y\|$$

Le couple $(E, \|\cdot\|)$ est appelé *espace vectoriel normé*. On vérifie immédiatement que c'est un espace métrique pour

$$(u, v) \mapsto d(u, v) = \|v - u\|.$$

Exercice I.2.4. Soit E un e.v.n. Montrer que, pour tous x, y dans E , on a

$$\|\|x\| - \|y\|\| \leq \|x - y\|.$$

CORRECTION.

On a

$$\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\| \implies \|y\| - \|x\| \leq \|x - y\|.$$

On démontre la même inégalité sur $\|x\| - \|y\|$ en intervertissant les rôles de x et y , d'où l'inégalité sur $\|y\| - \|x\|$.

Proposition I.2.8. (Normes $\|\cdot\|_p$ sur \mathbb{R}^d (\bullet))

L'ensemble des réels muni de la valeur absolue est un e.v.n. Pour tout d entier ≥ 1 , tout $p \in [1, +\infty]$, les expressions

$$\|x\|_p = \left(\sum_{k=1}^d |x_k|^p \right)^{1/p} \quad \text{si } p \in [1, +\infty[, \quad \|x\|_\infty = \max_{1 \leq i \leq d} |x_i|,$$

définissent des normes sur \mathbb{R}^d , qu'on appelle norme p , ou norme ℓ^p . Le cas $p = 2$ correspond à la norme dite *euclidienne*, associée au produit scalaire canonique

$$\langle x | y \rangle = \sum_{i=1}^d x_i y_i,$$

de telle sorte que $\|x\|_2 = \langle x | x \rangle^{1/2}$.

Démonstration. On sait déjà que $|x - y|$ définit une distance sur \mathbb{R} (proposition A.1.38), on a donc la séparation et l'inégalité triangulaire. Et on a par ailleurs $|\lambda x| = |\lambda| |x|$ pour tous x et λ réels par définition de la valeur absolue et du produit entre réels.

En dimension supérieure, les propriétés de séparation et d'homogénéité sont immédiates. L'inégalité triangulaire pour $p = +\infty$ découle de celle de la valeur absolue :

$$\|x + y\|_\infty = \max_{1 \leq i \leq d} |x_i + y_i| \leq \max_{1 \leq i \leq d} (|x_i| + |y_i|) \leq \max_{1 \leq i \leq d} |x_i| + \max_{1 \leq i \leq d} |y_i|.$$

Pour $p \in [1, +\infty[$ est une conséquence directe de l'inégalité dite de *Minkovski* (voir proposition A.1.45, page 184). \square

Remarque I.2.9. (•••) Comme nous le verrons plus loin (section I.8.2), toutes ces normes sont *équivalentes* (théorème I.8.4), ce qui implique en particulier que la convergence d'une suite pour l'une de ces normes implique la convergence pour toutes les autres. Cela ne signifie pas pour autant que l'on peut utiliser indifféremment l'une ou l'autre quel que soit le contexte. Les espaces correspondants, s'ils ont comme nous le verrons la même *topologie*, ont des propriétés très différentes, et leur légitimité pour formaliser tel ou tel problème dépend de la nature du problème considéré. La norme utilisée par défaut pour décrire l'espace qui nous entoure est la norme ℓ^2 , l'espace résultant de ce choix est dit euclidien, c'est à dire que l'on peut y définir un *produit scalaire* dont la norme est issue. Le chapitre III est consacré à l'étude théorique de tels espaces, appelés *espaces de Hilbert* si l'on n'impose pas à la dimension d'être finie, qui vérifient un grand nombre de propriétés qu'on ne trouve pas dans les autres espaces. Cette norme euclidienne canonique présente aussi l'avantage de ne pas dépendre des axes choisis, elle est *isotrope* (un vecteur ne change pas de norme si on lui fait subir une rotation), comme l'espace physique de la mécanique classique. Cette norme a aussi un fort intérêt dans la modélisation de phénomènes physiques, le "2" du ℓ^2 faisant écho au carré qui intervient dans l'expression de l'énergie cinétique, ce qui donnera une forte légitimité aux normes de type quadratique dans le contexte de la dynamique des fluides et des solides inertiels. On retrouve également un carré dans l'expression de l'énergie potentielle élastique d'un ressort linéaire, ce qui légitimera aussi les normes de ce type pour estimer l'énergie de déformation d'un solide élastique. On retrouve également ce "2", pour des raisons plus subtiles, en mécanique quantique⁴. Hors du contexte physique, ce cadre euclidien (ou hibertien) est aussi très naturel en statistique, pour définir notamment la notion de variance. De façon général, lorsque l'on cherche à estimer un écart entre deux collections de valeurs, il est souvent pertinent d'estimer cet écart (ou parle de *loss* dans le contexte de l'apprentissage statistique) au sens des moindres carrés, ce qui mobilise la norme ℓ^2 .

Les cas extrêmes $p = 1$ et $+∞$ conduisent à des espaces qui ont des propriétés que l'on peut qualifier de pathologiques (voir l'exercice I.2.5 ci-après, et aussi le chapitre IV sur les espaces de dimension infinie construits sur des normes de ce type). Ils ont pourtant une forte légitimité dans certains contextes. Le cas $p = 1$ correspond en quelque sorte à un principe de conservation. On pense par exemple à une situation où un vecteur de \mathbb{R}^d encode les quantités d'une substance dans les n compartiments respectifs d'une zone de l'espace, la conservation de la masse totale s'exprime alors naturellement comme l'appartenance du vecteur à la sphère unité de ℓ^1 . Cette "manière de mesurer les choses" est de fait très adaptée aux variables *extensives*, et elle jouera de fait un rôle essentiel en théorie de la mesure, dont l'un des fruits est la construction d'un espace (L^1) qui est le pendant en dimension infinie de ℓ^1 . Noter que, dans l'exemple des compartiments ci-dessus, les coefficients du vecteur correspondent à des quantités sommables, dont les valeurs elles-mêmes n'ont pas de pertinence intrinsèque (si l'on prend, pour une même situation physique, des compartiments plus petits, on aura des valeurs plus petites). La norme ℓ^∞ valorise les valeurs elles-mêmes plus que les quantités, c'est la norme naturelle pour mesurer des variables *intensives*, comme une température⁵, dont on s'intéresse à la valeur maximale. C'est ainsi une norme très "naturelle" en ingénierie ou en sûreté, puisqu'elle se focalise sur les valeurs maximales, qui correspondent souvent à des prescriptions ou normes de fonctionnement de systèmes industriels.

Les cas extrêmes $p = 1$ et $p = +∞$, qui sont pourtant légitimes dans certains cas (voir remarque précédente), sont d'une certaine manière dégénérés, comme l'illustre l'exercice suivant.

Exercice I.2.5. On se place sur \mathbb{R}^2 muni de la distance ℓ^1 . On définit la longueur d'une ligne brisée de \mathbb{R}^2 (ligne continue constituée d'un nombre fini de segments consécutifs) comme la somme des longueurs des segments qui la constituent. Montrer qu'il existe une infinité de lignes brisées qui joignent l'origine 0 et le point $A = (1, 1)$, et dont la longueur est égale à la distance entre ces deux points.

Décrire une situation de la vie réelle qui manifeste cette dégénérescence.

Montrer que le cas $p = ∞$ est aussi dégénéré en termes d'unicité du plus court chemin.

4. La densité de probabilité de présence d'une particule quantique est égale au *carré* du module de la fonction d'onde.

5. Reconnaissions que la situation est parfois ambiguë. Dans l'esprit de la théorie de la mesure qui sera abordée ensuite, on peut voir une température comme une chaleur divisée par une capacité thermique. La température est la variable intensive associée à la variable extensive de chaleur, elle jouera le rôle d'une fonction que l'on intègre contre la mesure / capacité calorifique, le résultat de l'intégration étant l'énergie thermique (ou chaleur), qui est elle une variable extensive.

CORRECTION.

Le segment qui joint les deux points est de longueur 2, c'est la distance entre les points. Mais on peut aussi passer par le point (1,0), le chemin a aussi une longueur de 2. On peut de façon générale de déplacer par petits pas successivement horizontaux (vers la droite) et verticaux (vers le haut), avec des points anguleux arbitraires, pour trouver toujours une longueur 2. Il y a donc une infinité (manifestement non dénombrable) de tels chemins. On peut les représenter précisément par un couple de suites de termes positif dont la somme vaut 1 , le première suite correspond aux longueurs successives des segments horizontaux, la seconde aux longueurs des segments verticaux. Si l'on s'en tient aux chemins qui ne contiennent qu'un nombre fini de segments, comme indiqué dans l'énoncé, on imposera que ces suites soient nulles au delà d'un certain rang. Noter que (contrairement à l'exemple de Manhattan indiqué ci-après) les segments ne sont pas nécessairement horizontaux ou verticaux, on peut par exemple passer par le point (1, 1/2).

Cette situation se rencontre par exemple lorsque l'on se promène dans un quartier d'une ville américaine comme Manhattan (on appelle d'ailleurs parfois la distance ℓ^1 la distance de Manhattan). Si l'on cherche à rejoindre un point situé sur l'avenue dans laquelle on est, la ligne droite est l'unique plus court chemin, mais si la cible n'est ni dans la même avenue, ni dans la même rue, alors il existe plusieurs plus courts chemins. Pour la norme ℓ^∞ il y a un unique plus court chemin brisé de 0 à A. Si l'on considère ici 0 et B = (0, 1), on retrouve une infinité de chemin, qui sont obtenus en rabatant des segments qui font un angle inférieur ou égal à $\pi/4$ avec l'axe des ordonnées.

Les espaces décrits ci-dessus jouent un rôle essentiel en analyse, en particulier l'espace euclidien (\mathbb{R}^d, ℓ^2) , mais toutes les distances ne sont pas issues d'une norme. À titre d'illustration extrême, la distance triviale définie ci-après peut être définie sur n'importe quel ensemble, y compris bien sûr l'espace vectoriel \mathbb{R}^d .

Définition I.2.10. (Distance triviale / discrète)

Soit X un ensemble. On définit $d : X \times X \rightarrow \mathbb{R}_+$ par $d(x, x) = 0$ pour tout x , et $d(x, y) = 1$ si $x \neq y$. On vérifie immédiatement qu'il s'agit d'une distance (appelée distance triviale, ou distance discrète).

L'exercice suivant définit une distance qu'il est tentant d'appeler une distance lexicographique entre mots (ou plus généralement chaîne de caractère)

Exercice I.2.6. (Distance lexicographique)

On considère A un ensemble ("A" pour Alphabet, que l'on peut voir comme la collection de lettres dont on dispose), et l'on considère un ensemble X de mots constitués de lettres de A . Un élément de X peut donc s'écrire comme une suite (x_i) finie ou infinie de lettres. Soient x et y deux mots. On note n_{xy} le plus petit indice pour lequel x_n et y_n diffèrent (avec la convention $n = +\infty$ si les mots sont les mêmes), et l'on définit $d(x, y) = 2^{-n}$. Montrer que l'on définit ainsi une distance sur X , qui une inégalité triangulaire renforcée, appelée *ultramétrique* :

$$d(x, y) \leq \max(d(x, z), d(z, y)) \quad \forall x, y, z \in X.$$

Montrer l'on peut munir \mathbb{R}^d , et même $\mathbb{R}^{\mathbb{N}}$ (ensemble des suite de réels) d'une distance de ce type.

Dans l'hypothèse (non faite jusqu'à présent) où A et X sont finis, proposer une autre distance sur X plus conforme à la distance entre les mots dans un dictionnaire.

CORRECTION.

La séparation et la symétrie sont immédiates. Considérons maintenant x, y et z trois mots de X. Les lettres de x et y sont les mêmes pour tout indice inférieur ou égal à min(n_{xz}, n_{yz}) - 1, on a donc $n_{xy} > \min(n_{xz}, n_{yz}) - 1$, soit $n_{xy} \geq \min(n_{xz}, n_{yz})$, d'où l'inégalité triangulaire renforcée

$$d(x, y) \leq \max(d(x, z), d(z, y)) \quad \forall x, y, z \in X,$$

qui implique l'inégalité triangulaire usuelle.

N.B. : un espace muni d'une telle distance vérifie des propriétés très fortes et contre-intuitives, comme détaillé dans la section VI.1, voir par exemple la proposition VI.1.6 qui établit que tout point d'une boule est centre de cette boule.

On peut appliquer directement cette approche à \mathbb{R}^d ou $\mathbb{R}^{\mathbb{N}}$ en considérant un vecteur de \mathbb{R}^d ou une suite infinie comme un mot constitués de lettres qui sont des réels.

Si X est fini, on peut associer à n'importe quelle ordre sur A (ordre alphabétique, qui est l'un des $|A|!$ ordres possibles sur A) un ordre lexicographique. Si X est fini, on peut numérotter les éléments de X selon cet ordre, et définir la distance entre deux mots comme la valeur absolue de la différence de leur indices dans cette numérotation. On obtient ainsi une distance qui correspond exactement à la distance entre deux mots dans le dictionnaire, vu comme une "liste" de tous les mots rangé par ordre alphabétique.

Noter que, pour la distance lexicographique définie précédemment, un "chapitre" de dictionnaire (qui rassemble l'ensemble des mots commençant par une même lettre) est une boule fermée de rayon $1/2$, dont chaque mot est le centre.

Exercice I.2.7. Proposer une métrique sur $X =] -1, 1[$ qui en fasse un espace de diamètre infini.

CORRECTION.

On peut considérer par exemple la distance

$$(x, y) \in] -1, 1[^2 \longrightarrow d(x, y) = |g(y) - g(x)|, \quad \text{avec } g(u) = \frac{u}{1 - u^2}.$$

Le caractère strictement croissant de cette fonction assure la séparation, la symétrie est immédiate, et l'inégalité triangulaire résulte de celle de la valeur absolue :

$$|g(u) - g(v)| = |g(u) - g(w) + g(w) - g(v)| \leq |g(u) - g(w)| + |g(w) - g(v)|.$$

Exercice I.2.8. (Des boules de toutes les formes)

- a) Donner l'allure de la sphère unité (sphère de centre l'origine et de rayon 1) de \mathbb{R}^2 pour les différentes normes p (on distinguera les cas $p = 1$, $p \in]1, 2[$, $p = 2$, $p \in]2, +\infty[$, et $p = +\infty$).
- b) On considère maintenant sur \mathbb{R}^3 la distance lexicographique qui fait l'objet de l'exercice I.2.6 : Pour x et y dans \mathbb{R}^3 , on note n_{xy} le plus petit indice (on numérote à partir de 0) pour lequel x_n et y_n diffèrent (avec la convention $n = +\infty$ si $x = y$), et l'on définit $d(x, y) = 2^{-n_{xy}}$. Quel est le diamètre de \mathbb{R}^3 ? Décrire les boules pour cette distance.
- c) Soit X un ensemble muni de la distance triviale (voir définition I.2.10). Quel est le diamètre de X ? Quelles sont les boules ouvertes et fermées pour cette distance ?

CORRECTION.

- a) La sphère unité pour la norme infinie est le carré $[-1, 1] \times [-1, 1]$. La sphère unité pour la norme 1 est aussi un carré centré en l'origine, dont l'un des côtés est le segment $[(1, 0), (0, 1)]$ (les autres sont obtenus par rotations d'angles $\pi/2$, π , et $3\pi/2$). Pour $p = 2$, on a le cercle de centre $(0, 0)$ et de rayon 1. Entre 2 et ∞ , la sphère à la forme d'un carré aux coins arrondis, d'autant plus proche du carré (en restant intérieure au carré limite) que p tend vers $+\infty$. On a le même type de convergence, quand p passe de 2 à 1, vers le carré du cas $p = 1$ (par l'extérieur).
- b) L'espace \mathbb{R}^3 est une boule fermée de rayon 1, dont tout point est centre. Le rayon de cette boule est aussi son diamètre (cette propriété surprenante est commune à tous les espaces ultramétriques), le diamètre de \mathbb{R}^3 est donc 1. Deux points quelconques dont les premières composantes diffèrent sont "diamétralement opposés". Pour tout $x = (x_0, x_1, x_2)$, $B_f(x, 1)$ est donc \mathbb{R}^3 , $B_f(x, 1/2)$ (ou $B_f(x, r)$ avec $r \geq 1/2$) est le plan

$$\{(x_0, \lambda_1, \lambda_2), \lambda_1, \lambda_2 \in \mathbb{R}\},$$

et $B_f(x, 1/4)$ (ou $B_f(x, r)$ avec $1/4 \leq r < 1/2$) est la droite $\{(x_0, x_1, x_2), \lambda_2 \in \mathbb{R}\}$. Enfin $B_f(x, r)$ avec $r < 1/4$ est réduite au singleton $\{x\}$.

- c) Si X n'est pas un singleton son diamètre est 1 (sinon le diamètre est nul), et les seules boules (qui sont à la fois ouvertes et fermées) sont les singletons et l'espace X tout entier. Noter que tout point est centre de l'espace, vu comme boule fermée de rayon 1.

I.3 Topologie des espaces métriques

Notions principales : ouvert, fermé, adhérence, intérieur, frontière, voisinage, partie dense.

Définition I.3.1. (Ouverts et fermés d'un espace métrique (•))

Soit (X, d) un espace métrique. On dit que $U \subset X$ est ouvert s'il est vide⁶ ou si tout x dans U est centre d'une boule ouverte non vide incluse dans U , i.e.

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

On dit qu'une partie de X est fermée si son complémentaire est ouvert.

Exercice I.3.1. Préciser dans les cas suivants si $A \subset X$ vu comme partie de l'espace métrique X (muni dans les exemples de la distance euclidienne canonique) est un ouvert, un fermé, ou ni l'un ni l'autre.

$$A_1 = \{0\} \subset \mathbb{R}, A_2 = [0, 1] \subset \mathbb{R}, A_3 = [0, 1[\subset \mathbb{R}, A_4 = [0, +\infty[\subset \mathbb{R}, A_5 = \mathbb{Q} \subset \mathbb{R},$$

$$A_6 = \bigcup_{n=1}^{+\infty}]n, n + 1/n[\subset \mathbb{R}, A_7 = \bigcup_{n=1}^{+\infty} [n, n + 1/n] \subset \mathbb{R},$$

$$A_8 =]0, 1[\times]0, 1[\subset \mathbb{R}^2, A_9 =]0, 1[\times]0, 1[\times \{0\} \subset \mathbb{R}^3, A_{10} = \mathbb{N} \times \mathbb{R} \subset \mathbb{R}^2.$$

CORRECTION.

On a A_1 fermé, A_2 fermé, A_3 ni fermé ni ouvert, A_4 fermé, A_5 ni fermé ni ouvert, A_6 est ouvert, A_7 est fermé, A_8 ouvert, A_9 ni fermé ni ouvert, A_{10} fermé.

Proposition I.3.2. (•) Soit (X, d) un espace métrique.

L'ensemble vide et X sont à la fois fermés et ouverts.

Toute union d'ouverts est un ouvert, et toute intersection finie d'ouverts est un ouvert.

Toute intersection de fermés est fermée, et toute union finie de fermés est fermée.

Démonstration. L'ensemble vide est un ouvert par définition, et X est lui-même ouvert car toutes les boules sont dans X .

Soit $(U_i)_{i \in I}$ une famille d'ouverts, et U leur union. Tout $x \in U$ est dans l'un des U_i , il existe donc une boule $B(x, r) \subset U_i \subset \bigcup U_j$.

Soient maintenant U_1, U_2, \dots, U_N des ouverts de X . Pour x dans l'intersection, pour tout i il existe $r_i > 0$ tel que $B(x, r_i) \subset U_i$. Si l'on prend $r = \min(r_i)$ (qui est bien strictement positif car la famille est finie), alors $B(x, r)$ est dans l'intersection des U_i .

Les propriétés sur les fermés se déduisent des propriétés sur les ouverts par complémentarité. Par exemple, pour toute collection (F_i) d'ouverts, si l'on note $U_i = F_i^c$ le complémentaire de F_i (qui est un ouvert par définition), on a

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

Une union finie de fermés se ramène de la même manière à une intersection finie d'ouverts. □

Exercice I.3.2. On se place sur \mathbb{R} muni de la distance usuelle. Montrer qu'une intersection infinie d'ouverts peut ne pas être ouverte, et qu'une union infinie de fermés peut ne pas être fermée.

CORRECTION.

L'intersection des intervalles $] -1/n, 1/n[$ est le singleton $\{0\}$, qui n'est pas ouvert car il ne contient aucune boule ouverte de centre 0.

L'union des $[1/n, 1]$ est $]0, 1[$, qui n'est pas fermé.

6. Cette précision n'est pas à strictement parler nécessaire, car l'ensemble vide vérifie automatiquement toute condition du type : "Pour tout x dans \emptyset, \dots ".

Remarque I.3.3. (Vers la topologie générale)

Les propriétés de la proposition précédente permettent de définir ce que l'on appelle une topologie (voir section A.2.3, page 186), dans un cadre général (sans recourir à la notion de métrique). Elles sont utilisées comme définition d'une famille d'ouverts, qui définit une topologie générale sur un ensemble. Dans le présent contexte des espaces métriques, la proposition précédente permet donc de s'assurer que la définition I.3.1 correspond bien à une topologie au sens général.

Remarque I.3.4. (Très importante, et un peu délicate)

Il est important de garder à l'esprit que, pour une métrique donnée, les caractéristiques topologiques d'un ensemble ne sont pas intrinsèques, mais dépendent de l'espace topologique dans lequel il est inclus. Ainsi on dira que l'intervalle $I = [0, 1]$ n'est pas ouvert, en considérant implicitement qu'il est considéré comme sous-ensemble de \mathbb{R} muni de la topologie usuelle. Toute boule centrée en 1 contient des réels strictement supérieurs à 1, ce qui exclut que I soit ouvert. Si l'on considère en revanche que l'univers se réduit à I lui-même, considéré comme un espace à part entière (que l'on pourrait donc noter X dans l'esprit de ce qui précède), alors $I = X = [0, 1]$ est bien un ouvert, puisque les réels strictement supérieurs à 1 considérés ci-dessus ne sont "pas dans le paysage". Noter qu'il est aussi fermé, conformément à la première assertion de la proposition I.3.2.

Du fait qu'une union d'ouverts est un ouvert, et qu'une intersection de fermés est fermée, on peut définir les notions de plus grand ouvert contenu dans un ensemble (intérieur), et de plus petit fermé qui contienne un ensemble (adhérence).

Définition I.3.5. (Intérieur, adhérence, frontière (•))

Soit (X, d) un espace métrique, et A une partie de X .

On appelle *adhérence* de A , et l'on note \bar{A} , le plus petit fermé contenant A , i.e. l'intersection de tous les fermés qui contiennent A .

On appelle *intérieur* de A , et l'on note \mathring{A} , le plus grand ouvert contenu dans A , i.e. l'union de tous les ouverts que A contient.

On appelle *frontière* de A l'ensemble $\partial A = \bar{A} \setminus \mathring{A} = \bar{A} \cap (\mathring{A})^c$.

On appelle *voisinage* d'un point x toute partie de X qui contient un ouvert contenant x .

Exercice I.3.3. a) Montrer qu'un ouvert est un ensemble qui s'identifie à son intérieur, et qu'un fermé est un ensemble qui s'identifie à son adhérence.

b) Comment caractériser un ensemble qui s'identifie à sa frontière ?

CORRECTION.

a) L'intérieur étant un ouvert par définition, un ensemble qui s'identifie à son intérieur est ouvert. Par ailleurs un ouvert se contient, et c'est évidemment le plus grand des ouverts qu'il contient. Le raisonnement est analogue pour les fermés.

b) Un ensemble qui s'identifie à sa frontière est un fermé (il l'est comme adhérence de quelque chose), d'intérieur vide.

Exercice I.3.4. On se place dans \mathbb{R}^2 muni de la distance euclidienne. Préciser \bar{A} , \mathring{A} et ∂A dans les cas suivants :

$$a) A = \{x_1, \dots, x_N\} \quad b) A = B(0, 1) \quad c) A = \{(x, y) \in \mathbb{R}^2, y \geq 0\}$$

$$d) A = \{(1/n, 0), n = 1, 2, \dots\} \quad e) A = \{(t, \sin(1/t)), t \in]0, +\infty[\}$$

CORRECTION.

a) $\bar{A} = A$, $\mathring{A} = \emptyset$, $\partial A = A$.

b) $\bar{A} = \overline{B}(0, 1)$, $\mathring{A} = A$, $\partial A = S(0, 1)$.

c) $\bar{A} = A$, $\mathring{A} = \{(x, y) \in \mathbb{R}^2, y > 0\}$, $\partial A = \{(x, 0), x \in \mathbb{R}\}$.

- d) $\bar{A} = A \cup \{0\}$, $\mathring{A} = \emptyset$, $\partial A = \bar{A}$.
e) $\bar{A} = A \cup \{0\} \times [-1, 1]$, $\mathring{A} = \emptyset$, $\partial A = \bar{A}$.

Remarque I.3.6. L'appellation la plus conforme à l'intuition commune est celle de frontière. Si l'on considère par exemple la zone occupée par un pays, la frontière topologique correspond bien à la frontière administrative, constituée de lignes marquant la séparation avec un pays voisin, à laquelle se rajoutent les bords naturels du pays (côtes). Dans ce cadre, conformément à l'intuition, la frontière est *petite* par rapport à l'objet lui-même, on serait tenté de dire qu'elle ne “pèse rien”⁷. On prendra cependant garde au fait que, si l'on sort du cadre de domaines réguliers comme des pays sur une carte, certains ensembles ont une frontière beaucoup plus “grosse” qu'eux mêmes. On se reporterà par exemple à l'exercice I.3.5, qui établit que l'ensemble des rationnels, qui est négligeable, admet pour frontière l'ensemble des réels⁸.

Le terme *ouvert* évoque le fait qu'un objet ne contienne aucun point de sa frontière, il est en quelque sorte directement exposé au monde extérieur, par opposition à un fermé qui contient sa frontière, et se trouve en quelque sorte *délimité*.

Remarque I.3.7. Il peut sembler étonnant que ces notions puissent être pertinentes dans une démarche de modélisation du réel, puisque ce qui distingue par exemple un ouvert de son adhérence consiste en des points qui sont infiniment proches (en deçà de toute longueur mesurable en pratique) de l'objet lui-même. Elles sont pourtant d'une importance considérable. On pourra penser par exemple à la modélisation d'un matériau déformable qui occupe une certaine zone de l'espace physique. On choisira de représenter cette zone par un ouvert⁹, souvent noté Ω (on parle de *domaine*). Ainsi, chaque point x de Ω est entouré d'une petite zone située à l'intérieur de l'objet modélisé, ce qui permet d'écrire l'équilibre des forces en x à partir des propriétés constitutives du matériau. On obtient ainsi une équation aux dérivées partielles qui a du sens en tout point de (l'ouvert) Ω . La frontière correspond à l'interface avec le monde extérieur (l'air libre, ou un autre matériau). Sur cette frontière l'équation constitutive du matériau n'est pas valide, mais on prescrira l'équilibre de l'interface qui impliquera les lois constitutives de part et d'autre, et on parlera de *condition aux limites* si l'extérieur est un milieu simple (comme du vide), ou de *conditions d'interface* s'il s'agit d'une zone de contact entre deux matériaux.

Remarque I.3.8. Dans l'esprit de la remarque précédente, on ne s'autorise à parler de la dérivée en un point x d'une fonction définie sur un sous-ensemble A de \mathbb{R} que lorsque x appartient à l'intérieur de cet ensemble. Pour que la dérivée puisse être définie en tout point de A , il est nécessaire que A soit ouvert. De la même manière, comme précisé dans le chapitre V, on ne pourra définir la différentielle d'une fonction définie sur $A \subset \mathbb{R}^d$ qu'en un point x *intérieur* à A . Il est en effet essentiel dans la définition de la dérivée ou de la différentielle que l'on puisse effectuer des petites variations, *dans toutes les directions*, autour du point considéré en restant dans l'ensemble sur lequel la fonction est définie.

Exercice I.3.5. Montrer que \mathbb{Q} , comme partie de l'espace métrique \mathbb{R} (muni de sa distance usuelle), est d'intérieur vide, d'adhérence et de frontière \mathbb{R} . En quel(s) sens peut-on dire que la frontière de cet ensemble est “beaucoup plus grosse” que l'ensemble lui-même ?

CORRECTION.

Toute boule $B(q, \varepsilon)$ contient un non rationnel (par exemple $q + \varepsilon\pi/4$), \mathbb{Q} est donc d'intérieur vide. Comme \mathbb{Q} est dense dans \mathbb{R} , son adhérence est \mathbb{R} tout entier, et sa frontière aussi. Pour un tel ensemble, la frontière est infinie non dénombrable, alors que l'ensemble lui-même est dénombrable. Nous verrons dans le chapitre suivant un autre point de vue, celui de la mesure de Lebesgue, qui conforte ce vocabulaire : on verra que \mathbb{Q} est négligeable, alors que sa frontière est de mesure pleine.

7. Mathématiquement (voir chapitre sur la théorie de la mesure et de l'intégration), on pourra établir qu'un domaine régulier est de mesure strictement positive, alors que sa frontière est de mesure nulle. Pour revenir au contexte géographique, ces considérations expriment de façon abstraite le fait que, si l'on prend au hasard un point en Europe, il y a peu de chance qu'il se trouve sur la frontière entre deux pays : on dira qu'il s'agit *presque sûrement* d'un point intérieur à l'un des pays.

8. Un pays construit de la sorte risquerait de consacrer l'essentiel de son budget à entretenir ses douaniers et ses gardes-côtes.

9. Ce choix est d'une certaine manière arbitraire, ou simplement justifié par ce qui suit. La question de savoir s'il est licite ou pas de considérer qu'un objet physique contienne sa frontière ou pas dépend du contexte de la modélisation.

Définition I.3.9. (Densité (•))

Soit (X, d) un espace métrique, et A une partie de X . On dit que A est *dense* dans X si $\overline{A} = X$, autrement dit si, pour tout $x \in X$, tout $\varepsilon > 0$, la boule $B(x, \varepsilon)$ contient au moins un élément de A .

Définition I.3.10. (Séparabilité)

On dit que (X, d) est *séparable* s'il contient un ensemble dénombrable dense.

Nous terminons cette section par quelques propriétés de \mathbb{R} , et de sa version *achevée* $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\} = [-\infty, +\infty]$.

Proposition I.3.11. L'ensemble $\mathbb{D} \subset \mathbb{R}$ des nombres décimaux et l'ensemble \mathbb{Q} des nombres rationnels sont denses dans \mathbb{R} .

Proposition I.3.12. Les ouverts de \mathbb{R} sont les réunions dénombrables d'intervalles ouverts.

Démonstration. Soit $U \subset \mathbb{R}$ un ouvert. Pour $x \in U$, on introduit

$$b_x = \sup \{y \geq x, [x, y] \subset U\}$$

Comme U est ouvert, il contient un intervalle du type $[x - \varepsilon, x + \varepsilon]$, avec $\varepsilon > 0$, on a donc $b_x \geq x + \varepsilon > x$. Si b_x est fini, b_x ne peut pas être dans U , sinon $[x, b + \varepsilon]$ serait dans U pour ε assez petit, et b_x serait battu. L'intervalle $[x, b_x]$ est donc dans U , et b_x ne l'est pas (qu'il soit infini ou fini). De la même manière on construit un intervalle $]a_x, x]$ dans U , avec $a_x \notin U$. L'intervalle $I_x =]a_x, b_x[$ est donc dans U , qui s'écrit ainsi comme réunion de ces intervalles¹⁰.

Par densité de \mathbb{Q} dans \mathbb{R} , chacun de ces intervalles contient un rationnel. On peut donc construire une injection de cette famille d'intervalle dans \mathbb{Q} , la famille est donc dénombrable. \square

La droite réelle achevée (••)

Dans certains contextes, en particulier en théorie de l'intégration, il est pertinent de compléter \mathbb{R} par des valeurs “aux bouts” :

Définition I.3.13. (Droite réelle achevée)

On appelle *droite réelle achevée*, et l'on note $\overline{\mathbb{R}}$, l'ensemble $\mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$, muni de la relation d'ordre canonique sur \mathbb{R} complétée par

$$-\infty < a < +\infty$$

pour tout $a \in \mathbb{R}$.

Proposition I.3.14. On peut munir $\overline{\mathbb{R}}$ d'une métrique

$$d(x, y) = |\arctan(y) - \arctan(x)|,$$

avec la convention $\arctan(\pm\infty) = \pm\pi/2$. Cette métrique induit une topologie sur $\overline{\mathbb{R}}$ telle que tout ouvert est soit un ouvert de \mathbb{R} soit du type $U \cup]a, +\infty]$, $U \cup [-\infty, b[$, ou $U \cup]a, +\infty] \cup [-\infty, b[$.

Démonstration. Le fait que $d(\cdot, \cdot)$ soit une distance se vérifie sans difficulté. On reprend maintenant l'argument utilisé dans la démonstration de la proposition I.3.12. Cette approche permet de décomposer tout ouvert U en une réunion dénombrables d'intervalles ouverts¹¹, la seule différence ici étant que l'un de ces intervalles peut être du type $]a, +\infty]$ ou $[-\infty, b[$. Cette décomposition assure la propriété annoncée.

Soit maintenant U un ouvert de $\overline{\mathbb{R}}$. Si U ne contient ni $+\infty$ ni $-\infty$, alors c'est aussi un ouvert de \mathbb{R} . S'il contient par exemple $+\infty$, alors il contient une boule $B(+\infty, \eta)$, avec $\eta > 0$. \square

10. Ces intervalles sont les classes d'équivalence de la relation \mathcal{R} d'équivalence (voir définition A.1.4, page 170, définie par $x \mathcal{R} y$ si $[x, y] \subset U$ (ou $[y, x] \subset U$ si $y < x$).

11. On appelle ces ouverts les *composantes connexes* de U .

Exercice I.3.6. Quel est le diamètre de $\overline{\mathbb{R}}$ muni de la métrique définie ci-dessus ? Quelle est la boule ouverte de centre 0 et de rayon $\pi/2$? La boule ouverte de centre 1 et de rayon $\pi/2$?

CORRECTION.

Le diamètre de $\overline{\mathbb{R}}$ pour cette métrique est π , et on a

$$B(0, \pi/2) =]-\infty, +\infty[, \quad B(1, \pi/2) =]-1, +\infty[,$$

Exercice I.3.7. Donner un autre exemple de métrique sur $\overline{\mathbb{R}}$ conduisant à la même topologie.

CORRECTION.

On peut par exemple remplacer la fonction arctan par $x \mapsto \frac{1}{1+|x|}$.

I.4 Suites

Définition I.4.1. (Suite (\bullet))

Une suite dans un ensemble X est une collection de points de X (non nécessairement distincts) indexée par les entiers. Plus formellement, on peut voir une suite comme une application de \mathbb{N} dans X . On note dans ce contexte x_n l'image de n par cette application.

Définition I.4.2. (Suite convergente (\bullet))

Soit (X, d) un espace métrique. On dit que la suite $(x_n)_{n \in \mathbb{N}}$ converge vers ℓ si la distance de x_n à ℓ peut être rendue arbitrairement petite pour n assez grand :

$$\forall \varepsilon, \exists N, \forall n \geq N, d(x_n, \ell) < \varepsilon.$$

Proposition I.4.3. (Unicité de la limite (\bullet))

Une suite convergente ne peut converger que vers une seule limite.

Démonstration. Soient ℓ et ℓ' deux limites de la suite (x_n) . Pour tout ε , il existe N et N' tels que, pour tout $n \geq \max(N, N')$, $d(x_n, \ell) < \varepsilon$ et $d(x_n, \ell') < \varepsilon$. On a donc, en prenant $n = \max(N, N')$,

$$0 \leq d(\ell, \ell') \leq d(\ell, x_n) + d(x_n, \ell') < 2\varepsilon.$$

On a donc $d(\ell, \ell') = 0$, d'où $\ell = \ell'$. □

Définition I.4.4. (Valeur d'adhérence (\bullet))

Soit (X, d) un espace métrique et $(x_n)_{n \in \mathbb{N}}$ une suite de points de X . On dit que x est valeur d'adhérence pour la suite s'il existe une suite extraite qui converge vers x , i.e. s'il existe une application φ strictement croissante de \mathbb{N} dans \mathbb{N} telle que la suite $(x_{\varphi(n)})$ converge vers x .

Proposition I.4.5. (•) Soit $(x_n)_{n \in \mathbb{N}}$ une suite de points de l'espace métrique X . Le point x est valeur d'adhérence pour la suite si et seulement si, pour tout $\varepsilon > 0$, l'ensemble $\{n \in \mathbb{N}, x_n \in B(x, \varepsilon)\}$ est infini.

Démonstration. Si x est valeur d'adhérence, on peut extraire une sous-suite $(x_{\varphi(n)})$ qui converge vers x . L'ensemble

$$\{n \in \mathbb{N}, x_{\varphi(n)} \in B(x, \varepsilon)\}$$

est infini par définition de la convergence d'une suite.

Réciproquement, on prend $\varepsilon = 1$. Comme l'ensemble $\{n \in \mathbb{N}, x_n \in B(x, 1)\}$ est infini, il est non vide, et on peut considérer k_1 son plus petit élément. On continue ensuite avec $\varepsilon = 1/2$, auquel on associe k_2 le plus petit élément de $\{n \in \mathbb{N}, x_n \in B(x, 1/2)\}$ strictement supérieur à k_1 . On construit ainsi une suite extraite (x_{k_n}) telle que $d(x_{k_n}, x) < 1/2^n$, donc qui converge vers x . □

Proposition I.4.6. (Caractérisations séquentielles (••))

Soit (X, d) un espace métrique, et A une partie de X . On a les équivalences :

1. Un point $x \in X$ est dans l'adhérence de A si et seulement si x est limite d'une suite de points (non nécessairement distincts, il peut s'agir de la suite constante égale à x) de A .
2. Un point $x \in A$ est dans l'intérieur de A si et seulement si toute suite convergeant vers x est dans A au delà d'un certain rang.
3. A est fermé si et seulement si, pour toute suite de points de A qui converge dans X , la limite est dans A .

Démonstration. 1. Soit x dans l'adhérence de A . Si x est dans A , il est limite de la suite constante égale à x . Si $x \notin A$, pour tout $n \in \mathbb{N}$, la boule $B(x, 1/n)$ rencontre nécessairement A en un point x_n . Si tel n'était pas le cas, alors $\bar{A} \setminus B(x, 1/n) = \bar{A} \cap B(x, 1/n)^c$ serait un fermé qui contient A , strictement plus petit que \bar{A} , ce qui est absurde. La suite (x_n) de points de A ainsi construite converge vers x par construction. Soit maintenant $x \notin \bar{A}$, il est dans son complémentaire, qui est un ouvert d'intersection vide avec A , il existe donc une boule ouverte $B(x, \varepsilon)$ qui ne contient aucun élément de A . Le point x ne peut donc être limite d'une suite de points de A .

2. Soit (x_n) une suite qui converge vers $x \in \bar{A}$. Il existe $r > 0$ tel que $B(x, r) \subset A$. Par convergence de la suite, il existe N tel que, pour tout $n \geq N$, on a $d(x, x_n) < r$. On a donc $x_n \in B(x, r) \subset A$ pour tout $n \geq N$. Réciproquement, si $x \in A$ n'est pas dans l'intérieur de A , pour tout $n > 0$, $B(x, 1/n)$ n'est pas incluse dans A , il existe donc $x_n \in B(x, 1/n) \cap A^c$. Cette suite converge vers x , mais ne rencontre pas A .

3. Soit $A \subset X$ fermé. Tout point de A^c , qui est ouvert, est dans une boule de rayon ε qui ne contient aucun élément de A , il ne peut donc être limite d'une suite de points de A . Réciproquement, si A n'est pas fermé, A^c n'est pas ouvert, il contient donc un élément x tel que pour tout ε , $B(x, \varepsilon)$ contient un élément qui n'est pas dans A^c , i.e. qui est dans A . On peut donc ainsi construire une suite de points de A qui converge vers $x \notin A$.

□

I.5 Complétude

La notion de complétude abordée dans cette section joue un rôle essentiel en analyse, pour montrer en particulier des résultats d'existence à toutes sortes de problèmes. Dans des espaces dits *complets*, au sens précisé ci-dessous, on dispose d'un critère de convergence d'une suite qui *n'utilise pas la limite elle-même* (contrairement à la définition I.4.2), mais seulement les termes de la suite.

Définition I.5.1. (Suite de Cauchy (•))

Soit (X, d) un espace métrique. On dit que la suite (x_n) de points de X est de Cauchy si la quantité $d(x_p, x_q)$ peut être rendue arbitrairement petite pour p et q assez grands :

$$\forall \varepsilon > 0, \exists N, \forall p, q \geq N, d(x_p, x_q) < \varepsilon.$$

Proposition I.5.2. (•) Toute suite convergente est de Cauchy.

Démonstration. Si une suite (x_n) converge vers une limite x , $d(x_n, x)$ peut être rendu arbitrairement petit pour n assez grand. Il en est donc de même pour

$$d(x_p, x_q) \leq d(x_p, x) + d(x, x_q),$$

d'où le caractère de Cauchy de la suite. □

Exercice I.5.1. Soit X un espace métrique, et (x_n) une suite de Cauchy qui ne prend qu'un nombre fini de valeurs. Montrer que la suite est nécessairement constante au delà d'un certain rang.

CORRECTION.

Soient a_1, a_2, \dots, a_p les valeurs prises par la suite de Cauchy. La quantité $d(a_i, a_j)$ admet un minimum $\varepsilon > 0$ sur l'ensemble des $1 \leq i < j \leq p$. On écrit le critère de Cauchy pour cet ε particulier : il existe N tel que, pour tous p, q plus grand que N , $d(x_p, x_q) < \varepsilon$. Les termes x_p pour $p \geq N$ s'identifient donc forcément à x_N .

Exercice I.5.2. Soit X un espace métrique, et (x_n) une suite de Cauchy. On note X_N l'ensemble des termes de la suite au delà du rang N :

$$X_N = \{x_n, n \geq N\}.$$

Montrer que la suite (x_n) est de Cauchy si et seulement si le diamètre de X_N tend vers 0 quand N tend vers $+\infty$.

CORRECTION.

Si la suite est de Cauchy, pour tout $\varepsilon > 0$, il existe N tel que, pour tous p, q plus grands que N , on a $d(x_p, x_q) < \varepsilon$, d'où $\text{diam}((X_N)) \geq N$, ce qui exprime exactement¹² la convergence de $\text{diam}((X_N))$ vers 0. La réciproque se démontre de la même manière.

Définition I.5.3. (Espace métrique complet (•))

On dit que (X, d) est complet si toute suite de Cauchy dans X converge vers un élément de X .

Proposition I.5.4. (•) Soit (X, d) un espace métrique complet. Une partie A de X est complète si et seulement si elle est fermée.

Démonstration. Si $A \subset X$ n'est pas fermée, il existe d'après la proposition I.4.6, page 22, une suite (x_n) de points de A qui converge vers $x \notin A$. Cette suite est de Cauchy (proposition I.5.2 ci-dessus), non convergente dans A , qui ne saurait donc être complet.

Si maintenant $A \subset X$ est fermée, toute suite de Cauchy dans A et de Cauchy dans X , elle converge donc vers $x \in X$. Cette limite est dans A , toujours d'après la proposition I.4.6. \square

Proposition I.5.5. (••) Soit $d \geq 1$ un entier. Pour tout $p \in [1, +\infty]$, l'espace \mathbb{R}^d muni de la norme $\|\cdot\|_p$ (voir proposition I.2.8 page 13), est complet.

Démonstration. La voie à suivre pour montrer la complétude de \mathbb{R} dépend de la manière dont on a construit l'ensemble des réels. On se reportera à la proposition A.1.39 pour une démonstration dans le cadre d'une construction basée sur l'écriture décimale (section A.1.4). Si l'on considère maintenant une suite de Cauchy $(x^n) = (x_k^n)_{1 \leq k \leq d}$ dans \mathbb{R}^d , la suite associée à l'une quelconque des composantes est aussi de Cauchy dans \mathbb{R} , donc converge vers une valeur x_k^∞ . On en déduit la convergence de (x^n) vers $(x^\infty) = (x_k^\infty)_{1 \leq k \leq d}$. \square

Exercice I.5.3. Montrer que l'ensemble \mathbb{D} des nombres décimaux, muni de la distance canonique $d(x, y) = |y - x|$, n'est pas complet.

CORRECTION.

La suite $x_n = 0.1111\dots 1100$ (décimales égales à 1 jusqu'à la n -ième) est de Cauchy, mais ne converge pas vers un décimal.

I.6 Compacité

La notion de compacité est essentielle en analyse, elle est en particulier à l'origine de l'essentiel des résultats d'existence de solution à des équations issues de la physique. Nous avons privilégié la définition la plus générale, qui pourrait s'appliquer à des espaces non métriques, car elle est basée sur la notion

12. Pour les élèves les formés en maths, penser à préciser que le fait que l'inégalité soit large en pose pas de problème, on peut même si on veut pinailler écrire les distances sont $< \varepsilon/2$, donc le diamètre $\geq \varepsilon/2 < \varepsilon$.

de recouvrement par des ouverts, notion très féconde également au cœur de la définition de la mesure extérieure de Lebesgue qui sera introduite dans la partie sur la théorie de la mesure. Dans le cas d'un espace métrique, la compacité peut se caractériser à l'aide de suites : un ensemble est compact si, de toute suite, on peut en extraire une sous-suite qui converge dans l'ensemble. Le lecteur désireux d'aller au plus simple pourra considérer que cette caractérisation est la définition première, en gardant à l'esprit qu'il en existe une formulation équivalente (et plus générale, puisqu'elle s'applique à des espaces topologiques non métriques) basée sur le recouvrement par des ouverts. L'équivalence entre les deux formulations fait l'objet du théorème I.6.2 ci-après, dit de Bolzano-Weierstrass, dont on pourra admettre le résultat.

Définition I.6.1. (Compact (•))

Soit (X, d) un espace métrique, et K une partie de X (qui peut être X lui-même). On dit que K est *compact* s'il vérifie la propriété de *Borel-Lebesgue* : de tout recouvrement de K par des ouverts on peut extraire un recouvrement fini :

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

On dit qu'une partie est *relativement compacte* si son adhérence est compacte.

Exercice I.6.1. (Compacts de \mathbb{R} (•))

- a) Montrer que ni \mathbb{R} (muni de sa métrique usuelle $d(x, y) = |y - x|$), ni $]0, 1[\subset \mathbb{R}$, ne sont compacts.
- b) Montrer que tout ensemble fini d'un espace métrique est compact.
- c) Montrer que l'ensemble des termes d'une suite strictement décroissante vers 0 n'est pas compact. Montrer que si l'on rajoute à cet ensemble la limite 0, alors l'ensemble est compact.

CORRECTION.

- a) \mathbb{R} n'est pas compact car de $\mathbb{R} = \bigcup_{\mathbb{Z}}]n-1, n+1[$ on ne peut extraire aucun recouvrement fini. Pour $]0, 1[$, on peut considérer

$$]0, 1[\subset \bigcup_{n \geq 2} \left[\frac{1}{n+1}, \frac{1}{n-1} \right].$$

d'où on ne peut extraire aucun recouvrement fini.

- b) Si un ensemble est fini, de cardinal N chaque point est contenu dans l'un des ouverts du recouvrement, on peut donc extraire un recouvrement par N ouverts (voire moins).
- c) Soit x_n décroissante vers 0. On peut recouvrir l'ensemble de ses termes par l'union des $]x_n/2, 2x_n[$, dont on ne peut extraire aucun recouvrement fini. Si l'on rajoute 0, alors pour tout recouvrement par des ouverts il existe un recouvrement qui contient 0. Les termes de la suite qui ne sont pas dans cet ouvert sont en nombre fini, on peut donc extraire un recouvrement fini de ce sous ensemble (voir (b)).

Exercice I.6.2. Montrer que la droite réelle achevée $\overline{\mathbb{R}} = [-\infty, +\infty]$ muni de la métrique définie par la proposition I.3.14, page 20, est *compacte*.

CORRECTION.

L'application

$$T : x \in \overline{\mathbb{R}} \mapsto T(x) = \arctan(x),$$

(avec la convention $\arctan(\pm\infty) = \pm\pi/2$), est une isométrie entre $[-\pi/2, \pi/2]$ muni de la distance usuelle et \mathbb{R} . Pout toute suite x_n dans $\overline{\mathbb{R}}$, on peut extraire une sous suite telle que $y_{\varphi(n)} = T(x_{\varphi(n)})$ converge vers $y \in [-\pi/2, \pi/2]$, d'où la convergence de $T^{-1}(x_{\varphi(n)})$ vers $T^{-1}(y)$.

Dans le cas des espaces métriques, on peut caractériser la compacité de façon séquentielle.

Théorème I.6.2. (Bolzano – Weierstrass (••))

Soit (X, d) un espace métrique, et $K \subset X$. L'ensemble K est compact (définition I.6.1) si et seulement si de toute suite de points de K on peut extraire une sous-suite qui converge vers un élément de K .

Démonstration. (•••) On suppose K compact au sens de la définition I.6.1. On considère une suite (x_n) d'éléments de K . Si cette suite n'admet aucune valeur d'adhérence, c'est-à-dire que l'on ne peut en extraire aucune sous-suite convergente, alors (d'après la proposition I.4.5) pour tout $y \in K$ il existe $r_y > 0$ tel que $B(y, r_y)$ ne contienne qu'un nombre fini de termes de la suite, plus précisément un nombre fini d'indices n tels que x_n est dans cette boule. La réunion de ces boules ouvertes recouvre K par construction, on peut donc en extraire un recouvrement fini :

$$K \subset \bigcup_{i \in J} B(y_i, r_i), \quad J \text{ fini.}$$

Le nombre total d'indices concernés est donc fini, car inférieur à la somme (finie) des cardinaux des indices affectés à chaque boule, ce qui est absurde.

Réciproquement, on suppose maintenant K séquentiellement compact, et l'on considère un recouvrement de K par des ouverts

$$K \subset \bigcup_{i \in I} U_i.$$

La démonstration se fait en trois étapes.

1) On montre dans un premier temps l'existence d'un $\rho > 0$ tel que, pour tout $x \in K$, il existe $i \in I$ tel que $B(x, \rho) \subset U_i$. Si tel n'est pas le cas, pour tout n , il existe $x_n \in K$ tel que $B(x_n, 1/n)$ n'est dans aucun des U_i . On peut extraire une sous-suite $(x_{\varphi(n)})$ qui converge vers $x \in K$ par compacité séquentielle de K . La limite x est dans un ouvert U_{i_0} . Il existe ε tel que $B(x, \varepsilon) \subset U_{i_0}$. Par convergence de $x_{\varphi(n)}$ vers x , il existe N tel que, pour tout $n \geq N$, $d(x_{\varphi(n)}, x) < \varepsilon/2$. On choisit maintenant n tel que $1/\varphi(n) < \varepsilon/2$. La boule $B(x_{\varphi(n)}, \varepsilon/2)$ est alors incluse dans U_{i_0} , ce qui contredit l'hypothèse initiale.

2) Montrons maintenant que K peut être recouvert par une collection finie de boules de rayon ρ . On raisonne une nouvelle fois par l'absurde. Si la propriété n'est pas vraie, on prend $x_1 \in X$ arbitraire. Comme $B(x_1, \rho)$ ne recouvre pas K , il existe $x_2 \in K \setminus B(x_1, \rho)$. Comme $B(x_1, \rho) \cup B(x_2, \rho)$ ne recouvre toujours pas K , il existe $x_3 \in K \setminus (B(x_1, \rho) \cup B(x_2, \rho))$. On construit ainsi par récurrence une suite (x_n) telle que tous les termes sont distants deux à deux d'au moins $\rho > 0$, on ne peut donc pas en extraire une sous-suite convergente, ce qui contredit la compacité séquentielle.

3) D'après le 2, il existe une collection finie de points x_1, \dots, x_N telles que

$$K \subset \bigcup_{n=1}^N B(x_n, \rho).$$

Comme chacune de ces boules $B(x_n, \rho)$ est dans l'un des ouverts U_{i_n} , on a bien un sous-recouvrement fini de K par les ouverts U_{i_n} , $n = 1, \dots, N$. \square

Proposition I.6.3. (•) Tout compact K d'un espace métrique X est fermé.

Démonstration. On utilise la caractérisation séquentielle du caractère fermé (proposition I.4.6) : si K est fermé, pour toute suite d'éléments de K qui converge dans X , la limite est dans K . On considère donc une telle suite. Si K est compact, on peut en extraire une sous-suite qui converge dans K , et la limite de la suite de départ s'identifie à la limite cette suite extraite. La limite est donc dans K , d'où le caractère fermé de K . \square

Exercice I.6.3. Montrer qu'une partie finie d'un espace métrique est toujours compacte.

CORRECTION.

Une suite d'éléments d'un ensemble fini visite nécessairement une infinité de fois au moins l'un de ces éléments. La sous-suite correspondant à ces indices est stationnaire en cet élément, elle converge donc dans K .

Exercice I.6.4. Montrer que l'intersection de deux compacts est compacte.

CORRECTION.

On considère une suite de $K_1 \times K_2$. On extrait une première suite qui converge dans K_1 . Cette sous-suite est une suite de K_2 , on peut donc en extraire une suite qui converge dans K_2 , donc dans $K_1 \cap K_2$.

Exercice I.6.5. Montrer que tout fermé inclus dans un compact d'un espace métrique est compact.

CORRECTION.

On considère une suite de points du fermé. Elle est dans le compact, on peut donc extraire une sous-suite qui converge dans le compact. La limite est dans le fermé, puisqu'il est fermé.

Théorème I.6.4. (Heine – Borel ou Borel – Lebesgue (••))

Les compacts de \mathbb{R}^d (pour toute norme $\|\cdot\|_p$, avec $1 \leq p \leq +\infty$) sont les fermés bornés.

Démonstration. Soit K un compact de \mathbb{R}^d . Si K n'est pas borné, on peut construire une suite d'éléments de K dont la norme tend vers $+\infty$. Il est de façon évidente impossible d'extraire d'une telle suite une sous-suite de convergence, K est donc nécessairement borné. La proposition I.6.3 ci-dessus assure par ailleurs que K est fermé.

Réciproquement, il s'agit de montrer que tout fermé borné de \mathbb{R}^d est compact. On considère d'abord le cas $d = 1$. Soit K un fermé borné de \mathbb{R} , et (x_n) une suite d'éléments de K . On suppose pour simplifier les notations que K est inclus dans l'intervalle $[0, 1]$. Au moins l'un des deux intervalles $[0, 1/2]$ et $]1/2, 1]$ contient une infinité de termes. On considère un tel sous-intervalle, et l'on choisit un terme de la suite, x_{n_1} , qui en fait partie. On subdivise en deux ce sous-intervalle, pour obtenir deux sous-intervalles de longueur $1/4$ dont l'un au moins contient une infinité de termes. On en prend un point x_{n_2} , avec $n_2 > n_1$. On construit ainsi par récurrence une suite extraite (x_{n_k}) , qui vérifie pour $p < q$,

$$|x_{n_q} - x_{n_p}| \leq \frac{1}{2^p},$$

elle est donc de Cauchy, donc converge dans \mathbb{R} (voir proposition A.1.39). Comme K est fermé, la limite est dans K (voir proposition I.4.6). On a donc pu extraire une sous suite qui converge dans K , ce qui assure sa compacité.

Dans le cas $d > 1$, on peut mettre en œuvre une approche analogue, en décomposant à chaque étape un cube de côté $1/2^k$ en 2^d cubes de côté $1/2^{k+1}$. On peut aussi appliquer ce qui précède à la première coordonnée, en extrayant une sous-suite convergente, puis passer à la seconde coordonnée en extrayant une sous-suite à cette première sous-suite, et continuer jusqu'à la d -ème coordonnée. \square

Remarque I.6.5. Le dictionnaire de l'Académie Française décrit comme compact un “objet dont les constituants sont serrés les uns contre les autres, pour former un substrat condensé”¹³. La définition mathématique dépasse largement cette acceptation commune, comme le suggère l'exercice I.6.1, en particulier du fait qu'un ensemble fini de points est compact. Pour s'en faire une idée intuitive, il est plus aisés d'identifier les propriétés qui font qu'un ensemble n'est *pas* compact. En premier lieu, comme l'indique la proposition I.6.3, la non compacité peut venir d'un défaut de fermeture : on peut extraire une sous-suite convergente, mais la limite n'est pas dans l'ensemble. Cela peut être corrigé en rajoutant les limites possibles de suites de l'ensemble, en considérant simplement l'adhérence de l'ensemble de départ (un tel ensemble dont l'adhérence est compacte est appelé *relativement compact*). Il y a des causes plus essentielles de non compacité. En premier lieu le caractère non borné de l'ensemble. L'ensemble \mathbb{N} des entiers naturels dans \mathbb{R} est bien fermé, mais de façon évidente non compact, car non borné (on peut recouvrir \mathbb{N} par des boules de rayon suffisamment petit pour que chaque boule ne contienne qu'un entier, il est alors évidemment impossible d'extraire un recouvrement fini). La dernière cause de non-compacité est plus profonde et moins facile à appréhender, elle porte sur le cœur de l'objet lui-même, ou plutôt de la nature de l'espace sous-jacent auquel il appartient. Dans un espace vectoriel normé de dimension infinie, on peut vérifier par exemple que la boule unité fermée, qui est bien fermée et bornée, n'est *pas compacte*. Considérons à titre d'exemple l'espace des polynômes¹⁴ muni de la norme définie comme le maximum des valeurs absolues des coefficients. La boule unité fermée de cet espace

13. Le terme est souvent utilisé à propos de *foules compactes*.

vectoriel normé de dimension est un fermé borné. Or la suite (X^n) est telle que la distance entre deux termes est égal à 1, on ne peut donc en extraire aucune sous suite convergente.

Définition I.6.6. (Relative compacité)

On dit que $K \subset X$ est relativement compact si \overline{K} est compact.

Définition I.6.7. (Précompacte)

On dit qu'une partie K d'un espace métrique (X, d) est *précompacte* si pour tout $\varepsilon > 0$ on peut recouvrir K par une nombre fini de boules de rayon ε .

Proposition I.6.8. Soit K une partie précompacte d'un espace complet X . Alors K est relativement compacte, i.e. d'adhérence compacte.

Démonstration. Soit (x_n) une suite dans K . On recouvre K par un nombre fini de boules de rayon 1. Au moins l'une de ces boules contient une infinité de termes de la suite, on extrait la sous-suite $x_{\varphi_1(n)}$ correspondante. On recouvre maintenant K par un nombre fini de boules de rayon $1/2$, on extrait $x_{\varphi_1 \circ \varphi_2(n)}$, etc ... On utilise ensuite le procédé d'extraction diagonale :

$$\psi(k) = \varphi_1 \circ \cdots \circ \varphi_k(k).$$

La suite $u_{\psi(k)}$ est telle que ses termes sont dans une boule de rayon $1/N$ pour $n \geq N$, elle est donc de Cauchy, donc converge vers une limite dans X , limite qui est dans \overline{K} , par définition de l'adhérence. \square

I.7 Applications entre espaces métriques

Définition I.7.1. (Application continue entre espaces métriques (•))

Une application f de (X, d) dans (X', d') est dite continue en x si $d'(f(x), f(y))$ peut être rendu arbitrairement petit pour tout y suffisamment proche de x , c'est-à-dire :

$$\forall \varepsilon > 0, \exists \eta > 0, \forall y \in X, d(x, y) < \eta \implies d'(f(x), f(y)) < \varepsilon.$$

On peut exprimer ce qui précède de façon séquentielle : pour toute suite (x_n) de X qui converge vers x , la suite $(f(x_n))$ des images converge vers $f(x)$.

On dit que F est continue sur X si elle est continue en tout point de X .

Cette définition est équivalente à une autre, plus abstraite, qui présente l'avantage de pouvoir s'appliquer à des espaces topologiques généraux (sans métrique). L'équivalence, dans le cas métrique, entre les deux, fait l'objet de la proposition suivante.

Proposition I.7.2. (Continuité d'une application, caractérisation générale (••))

Une application f de (X, d) dans (X', d') est continue si et seulement si l'image réciproque par f de tout ouvert de X' est un ouvert de X . De la même manière, une application f de (X, d) dans (X', d') est continue si et seulement si l'image réciproque par f de tout fermé de X' est un fermé de X .

Démonstration. Soit f une application de (X, d) dans (X', d') , continue au sens de la définition I.7.1 ci-dessus. On considère un ouvert U' de X' . Si $f^{-1}(U')$ est vide, il est ouvert. S'il n'est pas vide, pour tout x dans cette image réciproque, $f(x) = x' \in U'$ par définition de l'image réciproque. Comme U' est ouvert, il contient une boule $B(x', \varepsilon)$. Par continuité de f en x , il existe η tel que, pour tout y à distance de x inférieure à η , la distance de $f(y)$ à x' est inférieure à ε , ce qui signifie exactement $f(B(x, \eta)) \subset B(x', \varepsilon) \subset U'$. On a donc $B(x, \eta) \subset f^{-1}(U')$, d'où $f^{-1}(U')$ ouvert.

Montrons la réciproque. Soit f une application telle que l'image réciproque de tout ouvert de l'espace d'arrivée

14. On peut assimiler cet espace à l'espace F des suites finies (a_n) (qui s'annulent au-delà d'un certain rang), muni de la norme ℓ^∞ qui correspond au maximum des valeur absolue des termes.

est un ouvert de l'espace de départ. Soit $x \in X$, et $\varepsilon > 0$. L'image réciproque de $B(f(x), \varepsilon)$ est un ouvert, donc son image réciproque est un ouvert contenant x . Il existe donc $\eta > 0$ tel que $B(x, \eta) \subset f^{-1}(B(f(x), \varepsilon))$, c'est-à-dire $f(B(x, \eta)) \subset B(f(x), \eta)$.

Pour la caractérisation par les images réciproques de fermés, on utilise le fait que tout fermé F' de X' s'écrit $F' = U'^c$, où U' est ouvert. On a donc

$$f^{-1}(F') = f^{-1}(U'^c) = (f^{-1}(U'))^c,$$

qui est fermé si et seulement si $f^{-1}(U')$ est ouvert. \square

Exercice I.7.1. Montrer que l'image (directe) d'un ouvert par une application continue peut ne pas être ouverte, de même que l'image d'un fermé peut ne pas être fermée.

CORRECTION.

L'application $f : x \mapsto f(x) = \sin(x)$ est continue, mais l'image de l'ouvert \mathbb{R} (ou de l'ouvert $]-\pi, \pi[$ par f , égale à $[-1, +1]$, n'est pas ouverte.

L'application $f : x \mapsto f(x) = e^x$ est continue, mais l'image du fermé \mathbb{R} par f , égale à $]0, +\infty[$, n'est pas fermée.

On notera que cette deuxième application est injective, on pourrait penser que ça "marche dans les deux sens", mais comme elle n'est pas surjective, sa réciproque doit être vue comme application (le logarithme) de $]0, +\infty[$ dans \mathbb{R} , qui est aussi continue. L'image réciproque du fermé \mathbb{R} est $]0, +\infty[$, qui est bien fermé comme espace métrique à part entière (voir remarque I.3.4, page 18).

Proposition I.7.3. (Image d'un compact par une application continue (•))

Soit f une application de (X, d) dans (X', d') . Si f est continue, alors l'image d'un compact de X est compacte dans X' .

Démonstration. Soit K un compact de X . Une suite de $f(K)$ s'écrit $(f(x_n))$, avec $x_n \in K$ pour tout n . Comme K est compact, cette suite admet une sous-suite (x_{n_k}) qui converge vers $x \in K$, et la continuité de f assure que $f(x_{n_k})$ converge vers $f(x) \in f(K)$, d'où la compacité de $f(K)$. \square

Proposition I.7.4. (•) Soit f une fonction définie d'un compact K de (X, d) à valeurs dans \mathbb{R} , continue sur K . Alors f est bornée, et atteint ses bornes sur K .

Démonstration. L'image du compact K par f étant un compact de \mathbb{R} d'après la proposition I.7.3, il est borné (proposition I.6.4), la fonction f est donc majorée et minorée sur K . Notons M sa borne supérieure. Par définition il existe (x_n) dans K telle que

$$f(x_n) \rightarrow M = \sup_K f.$$

La suite maximisante (x_n) n'est pas nécessairement convergente, mais comme K est compact, on peut en extraire une sous-suite $(x_{\varphi(n)})$ qui converge vers $x \in K$. On a, par continuité de l'application, $f(x) = \lim f(x_{\varphi(n)}) = M$, la borne supérieure est donc atteinte. On montre de la même manière que la borne inférieure est atteinte. \square

Corollaire I.7.5. (•) Soit f une fonction continue d'un fermé borné $K \subset \mathbb{R}^d$, à valeurs dans \mathbb{R} . Alors f est bornée, et atteint ses bornes.

Exercice I.7.2. (•)

- a) Soit f une fonction continue de \mathbb{R}^d dans \mathbb{R} . Montrer qu'elle est bornée sur tout ensemble borné de \mathbb{R}^d .
- b) Montrer qu'une fonction définie d'un borné B de \mathbb{R}^d , continue, peut ne pas être bornée sur B .

CORRECTION.

- a) *Tout borné est d'adhérence compacte, d'où f est bornée sur l'adhérence, donc sur le borné.*
- b) *La fonction $x \mapsto 1/x$ est continue sur le borné $]0, 1[$ de \mathbb{R} , mais non bornée.*

Définition I.7.6. (Uniforme continuité (•))

Soit f une application de (X, d) dans (X', d') . On dit que f est uniformément continue sur X si

$$\forall \varepsilon > 0, \exists \eta > 0, \forall x, y \in X, d(x, y) < \eta \implies d'(f(x), f(y)) < \varepsilon.$$

Théorème I.7.7. (Heine (••))

Soient (X, d) et (X', d') deux espaces métriques, et f une application continue d'un compact $K \subset X$ dans X' . Alors f est uniformément continue.

Démonstration. On raisonne par contradiction. Si f n'est pas uniformément continue, il existe $\varepsilon > 0$ tel que, pour tout $\eta > 0$, il existe x et y tels que $d(x, y) < \eta$, et $d'(f(x), f(y)) \geq \varepsilon$. On prend $\eta = 1/n$, et l'on construit ainsi une suite (x_n, y_n) dans $X \times X$ telle que

$$d(x_n, y_n) < \eta, \quad d'(f(x_n), f(y_n)) \geq \varepsilon.$$

Par compacité de K on peut extraire de x_n une suite qui converge vers $x \in K$. On note toujours (x_n) la suite extraite (et par (y_n) la suite associée). La suite (y_n) étant adjacente à (x_n) , elle converge également vers x . On a

$$d(f(x_n), f(y_n)) \leq d(f(x_n), f(x)) + d(f(x), f(y_n)),$$

qui tend vers 0 par continuité de f en x , ce qui est en contradiction avec l'hypothèse.

□

Définition I.7.8. (Application lipschitzienne (•))

Une application de (X, d) dans (X', d') est dite lipschitzienne s'il existe $k \in \mathbb{R}_+$ tel que

$$d'(f(x), f(y)) \leq k d(x, y).$$

Exercice I.7.3. (Distance à une partie 1-lipschitzienne)

Soit (X, d) un espace métrique, $K \subset X$ non vide. Montrer que l'application qui à $x \in X$ associe sa distance à K (voir définition I.2.3) est une application 1-lipschitzienne.

CORRECTION.

Pour tout v dans K , on a

$$d(z, v) \leq d(z, z') + d(z', v) \leq d(z, z') + d(z', K),$$

d'où, en passant à l'infimum en v ,

$$d(z, K) - d(z', K) \leq d(z, z').$$

On démontre de la même manière que $d(z', K) - d(z, K) \leq d(z, z')$, d'où le résultat.

I.8 Compléments

I.8.1 Théorème de point fixe de Banach

Définition I.8.1. (Application contractante (•))

On dit qu'une application T de (X, d) dans lui-même est *contractante* s'il existe $\kappa \in [0, 1[$ tel que

$$d(T(x), T(y)) \leq \kappa d(x, y) \quad \forall x, y \in X.$$

Théorème I.8.2. (Théorème de point fixe de Banach (••))

Soit (X, d) un espace métrique complet et T une application de (X, d) dans lui-même, contractante. Elle admet alors un *point fixe* unique, c'est-à-dire qu'il existe un unique x dans X tel que $T(x) = x$.

Démonstration. L'unicité est immédiate : si x et y sont points fixes, on a

$$d(x, y) = d(T(x), T(y)) \leq \kappa d(x, y).$$

Comme $0 < \kappa < 1$, ça n'est possible que si $x = y$.

Pour l'existence, on considère un élément x_0 arbitraire de X , et l'on construit la suite des itérés par T :

$$x_n = T(x_{n-1}) = T^n(x_0) = \underbrace{T \circ T \circ \cdots \circ T}_{n \text{ fois}}(x_0).$$

On a

$$d(x_{n+1}, x_n) = d(T(x_n), T(x_{n-1})) \leq \kappa d(x_n, x_{n-1}) \leq \cdots \leq \kappa^n d(x_1, x_0).$$

On a donc, pour tous $p < q$,

$$\begin{aligned} d(x_p, x_q) &\leq d(x_p, x_{p+1}) + d(x_{p+1}, x_{p+2}) + \cdots + d(x_{q-1}, x_q) \\ &\leq (\kappa^p + \cdots + \kappa^{q-1}) d(x_1, x_0) \end{aligned}$$

qui tend vers 0 car la série $\sum \kappa^n$ est convergente. La suite (x_n) est donc de Cauchy, et converge vers un certain $x \in X$ car X est complet. Comme $d(x_{n+1}, x_n) = d(T(x_n), x_n)$ tend vers 0, on obtient en faisant tendre n vers $+\infty$, $d(T(x), x) = 0$, d'où $T(x) = x$. \square

I.8.2 Compléments sur les e.v.n. de dimension finie

Cette section est consacrée à une étude plus poussée des espaces vectoriels normés de dimension finie, qui se ramène à l'étude des espaces \mathbb{R}^n , pour n entier supérieur ou égal à 1. On rappelle la définition des normes usuelles sur \mathbb{R}^d (introduites dans la proposition I.2.8, page 13) :

$$\|x\|_p = \left(\sum_{k=1}^n |x_k|^p \right)^{1/p} \quad \text{pour } p \in [1, +\infty[,$$

ainsi que

$$\|x\|_\infty = \max_{1 \leq k \leq n} |x_k|.$$

Nous établissons ci-dessous des résultats généraux pour les normes sur \mathbb{R}^n . Il peut s'agir des normes ℓ_p rappelées ci-dessus, ou de variations directes de ces normes, obtenues par exemple en rajoutant des poids

$$\|x\|_{p,\mu} = \left(\sum_{k=1}^n \mu_k |x_k|^p \right)^{1/p},$$

avec $\mu = (\mu_k) \in]0, +\infty[^n$. La collection de poids μ joue le rôle *mesure* (voir II.3.2) sur l'ensemble fini $\llbracket 1, N \rrbracket$, mesure contre laquelle on intègre une certaine puissance des composantes du vecteurs. Mais on peut construire des normes d'autres types, par exemple¹⁵

$$\|x\|_{H^1}^2 = \sum_{k=1}^{n-1} \mu_k |x_{k+1} - x_k|^2 + \left| \sum_{k=1}^n x_k \right|^2.$$

On pourra se convaincre qu'il s'agit bien d'une norme sur \mathbb{R}^n en vérifiant que cette expression rentre dans le cadre de l'exercice I.8.1)

Exercice I.8.1. Soit F une application linéaire injective de \mathbb{R}^n dans \mathbb{R}^m , et $\|\cdot\|$ une norme sur \mathbb{R}^m . Alors

$$x \longmapsto \|F(x)\|$$

est une norme sur \mathbb{R}^n .

15. Il s'agit d'une version discrète d'une norme dite de *Sobolev*, qui pour des fonctions fait intervenir l'intégrale du carré de la dérivée.

CORRECTION.

La séparation est immédiate du fait de l'injectivité de F , l'homogénéité et l'inégalité triangulaire sont conséquences de la linéarité de F .

Lemme I.8.3. Toute application linéaire de $(\mathbb{R}^n, \|\cdot\|_\infty)$ dans $(\mathbb{R}^m, \|\cdot\|)$ (où $\|\cdot\|$ est une norme quelconque sur \mathbb{R}^m) est continue.

Démonstration. On écrit pour cela la décomposition d'un élément x de \mathbb{R}^n dans la base canonique, et l'on estime la norme de $F(x)$:

$$\left\| F \left(\sum_{i=1}^d x_i e_i \right) \right\| = \left\| \sum_{i=1}^d x_i F(e_i) \right\| \leq \sum_{i=1}^d |x_i| \|F(e_i)\| \leq M \max |x_i| = M \|x\|_\infty,$$

où $M = \sum \|F(e_i)\|$. On a donc continuité de F car $F(x+h) = F(x) + F(h)$, et $\|F(h)\|$ peut être contrôlé par $\|h\|_\infty$ d'après ce qui précède. \square

Théorème I.8.4. (Équivalence des normes en dimension finie (•))

Toutes les normes sur \mathbb{R}^n sont équivalentes, c'est à dire que, pour toutes¹⁶ normes $\|\cdot\|_\alpha$ et $\|\cdot\|_\beta$ sur \mathbb{R}^n , il existe deux constantes $M > m > 0$ telles que

$$m \|x\|_\alpha \leq \|x\|_\beta \leq M \|x\|_\alpha \quad \forall x \in \mathbb{R}^n$$

Démonstration. Nous allons montrer que toutes les normes sont équivalentes à la norme $\|\cdot\|_\infty$, ce qui établira l'équivalence de toutes les normes entre elles. On considère l'application identité de \mathbb{R}^n dans \mathbb{R}^n , en munissant l'espace de départ de la norme $\|\cdot\|_\infty$, et l'espace d'arrivée d'une norme quelconque $\|\cdot\|$. D'après ce qui précède il existe une constante M telle que $\|x\| \leq M \|x\|_\infty$.

On considère maintenant la fonction qui à x dans $(\mathbb{R}^n, \|\cdot\|_\infty)$ associe $\|x\|$. Cette application est continue car (voir exercice I.2.4)

$$\| \|x+h\| - \|x\| \| \leq \|h\| \leq M \|h\|_\infty.$$

Comme la sphère unité S de \mathbb{R}^n est compacte (proposition I.6.4), cette fonction continue atteint ses bornes sur S (proposition I.7.4, page 28), en particulier son infimum $m \geq 0$. Il existe donc un x_0 , de norme ∞ égale à 1, tel que $\|x_0\| = m$. Comme x_0 est non nul, on a $m > 0$, et ainsi, pour tout x non nul,

$$\left\| \frac{x}{\|x\|_\infty} \right\| \geq m > 0 \implies m \|x\|_\infty \leq \|x\|.$$

On a donc montré

$$m \|x\|_\infty \leq \|x\| \leq M \|x\|_\infty,$$

c'est-à-dire que les normes $\|\cdot\|_\infty$ et $\|\cdot\|$ sont équivalentes. Toutes les normes sont donc équivalentes à une même norme $\|\cdot\|_\infty$, elles sont donc équivalentes entre elles. \square

Proposition I.8.5. (•) On munit \mathbb{R}^n et \mathbb{R}^m de normes $\|\cdot\|_\alpha$ et $\|\cdot\|_\beta$, respectivement. Alors toute application F linéaire de \mathbb{R}^n dans \mathbb{R}^m est continue, et il existe une constante $C \geq 0$ telle que, pour tout $x \in \mathbb{R}^n$,

$$\|Fx\|_\beta \leq C \|x\|_\alpha.$$

Démonstration. Cette propriété a fait l'objet du lemme I.8.3, dans le cas où la norme sur l'espace de départ est la norme $\|\cdot\|_\infty$. Elle est donc vraie pour toute autre norme sur l'espace de départ d'après l'équivalence des normes qui vient d'être établie. \square

Le choix d'une norme¹⁷ sur \mathbb{R}^n et \mathbb{R}^m induit canoniquement une norme sur l'espace vectoriel des applications linéaires de \mathbb{R}^n vers \mathbb{R}^m .

16. Malgré la notation, on peut envisager des normes qui diffèrent des normes p définies précédemment.

17. Il peut s'agir de normes différentes, même si nous privilégierons ici le cas de normes de même type.

Proposition I.8.6. (Norme d'opérateur (•))

On note $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (ou simplement $\mathcal{L}(\mathbb{R}^n)$ si $m = n$) l'espace des applications linéaires de \mathbb{R}^n dans \mathbb{R}^m . On note¹⁸ Fx l'image par F d'un élément x de \mathbb{R}^n . Pour tout $p \in [1, +\infty]$, l'application

$$F \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \longmapsto \|F\|_p = \sup_{x \neq 0} \frac{\|Fx\|_p}{\|x\|_p} = \sup_{\|x\|_p=1} \|Fx\|_p$$

définit une norme sur $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, et l'on a, pour tout $x \in \mathbb{R}^n$,

$$\|Fx\|_p \leq \|F\|_p \|x\|_p.$$

On dit que cette norme d'opérateur est *subordonnée* à la norme p . Les normes ainsi définies sont compatibles avec le produit de composition, au sens suivant : pour tous $F \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^m)$, $G \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$

$$\|F \circ G\|_p \leq \|F\|_p \|G\|_p.$$

Dans le cas où l'on considère la norme euclidienne (cas $p = 2$), on omettra parfois l'indice, pour noter simplement $\|F\|$.

Démonstration. La propriété de séparation est immédiate, $\|F\|_p$ ne pouvant être nulle que si F est identiquement nulle. L'homogénéité résulte elle aussi directement de l'homogénéité de la norme p pour les vecteurs. Pour l'inégalité, on remarque que $\|F\|_p$ est le sup sur la sphère unité de $\|Fx\|_p$. Or on a, pour tous F_1, F_2 dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, tout x de norme unitaire,

$$\sup_{\|x\|_p=1} (\|F_1x + F_2x\|_p) \leq \sup_{\|x\|_p=1} (\|F_1x\|_p + \|F_2x\|_p) \leq \sup_{\|x\|_p=1} \|F_1x\|_p + \sup_{\|x\|_p=1} \|F_2x\|_p.$$

D'après la définition-même, on a $\|Fx\|_p / \|x\|_p \leq \|F\|_p$ pour tout x non nul, dont on déduit immédiatement $\|Fx\|_p \leq \|F\|_p \|x\|_p$.

Pour la composée d'applications, on écrit

$$\|F \circ G\|_p = \sup_{x \neq 0} \frac{\|F(G(x))\|_p}{\|x\|_p} \leq \sup_{x \neq 0} \frac{\|F\|_p \|G(x)\|_p}{\|x\|_p} = \|F\|_p \|G\|_p,$$

ce qui termine la preuve. \square

Remarque I.8.7. Nous avons défini les normes d'opérateurs en munissant les espaces de départ et d'arrivée d'une même norme, mais on peut bien sûr étendre cette approche au cas où l'on choisit des normes différentes, en notant alors $\|F\|_{p,q}$ la norme subordonnée (p pour l'espace de départ, q pour l'espace d'arrivée).

Exercice I.8.2. Montrer que toutes les normes d'opérateurs que l'on peut définir selon les principes décrits ci-dessus à partir de normes sur \mathbb{R}^n et \mathbb{R}^m sont équivalentes entre elles.

CORRECTION.

On peut le montrer “à la main” en établissant les inégalités d'équivalences pour les opérateurs à partir de celles pour les vecteurs, ou simplement remarquer que l'espace des applications linéaires de \mathbb{R}^n dans \mathbb{R}^m est de dimension finies, donc toutes les normes sont équivalentes entre elles, en particuliers les normes d'opérateur telles que définies précédemment.

Quelques mots sur les espaces de dimension infinie.

Le cas de la dimension infinie, qui correspond par exemple aux espaces de fonctions définies sur un domaine de \mathbb{R}^d , est beaucoup plus délicat. On se reportera au chapitre IV, page 91, pour un aperçu de ce domaine à part entière des mathématiques. Disons simplement ici que toutes les propriétés présentées ci-dessus, qui utilisent de façon essentielle le caractère compact des fermés bornés, ne sont plus vraies. De fait, la boule unité fermée d'un espace vectoriel normé de dimension infinie n'est jamais compacte. On peut avoir sur un même espace des normes qui ne sont pas équivalentes entre elles. L'espace peut être complet pour une norme, pas pour une autre. Une application linéaire peut ne pas être continue¹⁹.

18. Cette notation Fx plutôt que $F(x)$ (qui peut aussi être utilisée) rappelle que l'on peut représenter F par une matrice, et donc l'image d'un élément de \mathbb{R}^n par un produit matrice-vecteur.

19. Précisons tout de même que la construction d'une telle application linéaire non continue entre espaces complets nécessite

l'axiome du choix : on ne rencontrera pas beaucoup de tels objets dans la nature.

I.9 Exercices

Exercice I.9.1. (Suite décroissante d'ensembles (•))

- a) Donner un exemple de suite $(U_i)_{i \in \mathbb{N}}$ décroissante d'ouverts de \mathbb{R} , c'est-à-dire telle que $U_{i+1} \subset U_i$ pour tout i , qui soit telle que l'intersection des U_i est vide.
- b) Donner un exemple de suite $(F_i)_{i \in \mathbb{N}}$ décroissante de fermés de \mathbb{R} , c'est-à-dire telle que $F_{i+1} \subset F_i$ pour tout i , qui soit telle que l'intersection des F_i est vide.

CORRECTION.

a) Il suffit de considérer la suite des intervalles $]0, 1/n[$.

b) Comme on le verra, une suite décroissante de compacts a une intersection non vide (exercice I.9.8), il est donc sans espoir de chercher un exemple avec des bornés. On pourra considérer par exemple la suite des intervalles semi-infinis $[n, +\infty[$.

Exercice I.9.2. (Distance de Hamming)

On considère l'ensemble $H_N = \{0, 1\}^N$ des N -uplets de 0 ou de 1 (ensemble des mots de N bits).

- 1) On définit $d(x, y)$ comme le nombre de positions où les bits de x et y diffèrent, i.e.

$$x = (x_0 \dots, x_{N-1}), \quad y = (y_0 \dots, y_{N-1}), \quad d(x, y) = \sum_{n=0}^{N-1} |x_n - y_n|.$$

Montrer que l'on définit ainsi une distance (appelée distance de *Hamming*), qui fait de H_N un espace discret. Quel est le diamètre de H_N ?

- 2) Soit $x \in X$ et $r \in \mathbb{R}_+$. Donner le cardinal de la sphère de centre x et de rayon $r \geq 0$, en fonction de r .

CORRECTION.

a) La séparation et la symétrie sont immédiates. Si l'on considère maintenant x, y , et z dans H_N . Si aucun des bits qui diffèrent entre x et y ne correspond à l'un de ceux qui diffèrent entre y et z , on a $d(x, z) = d(x, y) + d(y, z)$. Si ces deux ensemble ont un ou des éléments communs, on a $d(x, z) < d(x, y) + d(y, z)$, d'où l'inégalité triangulaire.

Chaque point est seul dans sa boule ouverte de rayon $1/2$, il s'agit donc d'un espace discret.

Pour tout $x \in H_N$, le point le plus éloigné est celui obtenu en changeant tous les bits, il est donc à distance N . La diamètre est donc N , et l'on peut dire que tout point est "sur le bord de H_N , au sens où tout point admet un élément diamétralalement opposé.

2) On considère le cas $x = (0, 0, \dots, 0)$, la situation étant exactement la même si l'on part d'un autre éléments. Pour $r < 1$, la boule est réduite au centre. Pour $1 \leq r < 2$, on a tous les points qui diffèrent d'un bit, il y en a donc N . Pour $2 \leq r < 3$ on a $C_N^2 = N(N-1)/2$. De façon générale, pour $k \leq r < k+1$, le cardinal est C_N^k , et $C_N^N = 1$ pour $r \geq N$ (la sphère est constituée de l'unique point diamétralalement opposé).

Exercice I.9.3. (Distances ultramétriques (•••))

On considère l'ensemble $H_N = \{0, 1\}^N$ des N -uplets de 0 ou de 1 (ensemble des mots de N bits).

Pour $x = (x_1 \dots, x_N)$ et $y = (y_1 \dots, y_N) \neq x$, on note k le plus petit indice tel que les bits de x et y diffèrent, i.e.

$$k = \min \{k, x_k \neq y_k\},$$

et l'on définit $\delta(x, y) = 2^{-k}$. On pose $\delta(x, x) = 0$.

- 1) Montrer que $\delta(\cdot, \cdot)$ est une distance sur H_N , et que cette distance est *ultramétrique*, c'est à dire qu'elle vérifie l'inégalité triangulaire renforcée

$$d(x, z) \leq \max(d(x, y), d(y, z)).$$

- 2) Montrer que tout point d'une boule est centre de cette boule (on se gardera d'essayer de faire un dessin...).
- 3) Quel est le diamètre de H_N pour cette distance ?

- 4) a) Décrire la sphère de centre $0 = (0, 0, \dots, 0)$ et de rayon $r \in [0, 1]$, selon la valeur de r , et plus généralement la sphère de centre $x = (x_1, \dots, x_N)$ et de rayon $r \in [0, 1]$.
 b) Soient $x \in X$ et $r \in \mathbb{R}_+$. Donner le cardinal de la boule fermée de rayon x et de rayon r , en fonction de r .
 5) Étendre l'approche précédente à l'ensemble $H_\infty = \{0, 1\}^{\mathbb{N}}$ des suites infinies de 0 ou 1.
 6) Donner des exemples de contextes dans lesquels une distance ultramétrique apparaît naturellement.

CORRECTION.

1) La distance n'est nulle que si les points sont confondus, par définition. Les rôles joués par x et y dans la définition sont interchangeables, ce qui implique la symétrie. Pour l'inégalité triangulaire, considérons x, y , et z , et k_{xy} le plus petit indice pour lequel les bits de x et y diffèrent. On définit de même k_{xz} et k_{yz} . Les éléments x et y s'identifient sur les $k_{xy} - 1$ premiers bits, y et z sur les $k_{yz} - 1$, d'où x et z s'identifient par transitivité sur les $\min(k_{xy}, k_{yz}) - 1$ premiers bits, d'où l'on déduit, $k_{xz} \geq \min(k_{xy}, k_{yz})$, et l'inégalité triangulaire en appliquant la fonction décroissante $k \mapsto 2^{-k}$.

XXXXX. 2) 3) 4) 5) 6)

Exercice I.9.4. (••) Soit (X, d) un espace métrique, et $A \subset X$. Montrer que l'intérieur de A est égal au complémentaire de l'adhérence du complémentaire de A .

CORRECTION.

Le complémentaire de l'adhérence du complémentaire de A est un ouvert comme complémentaire d'un fermé. Montrons qu'il est contenu dans A : pour tout $x \in (\overline{A^c})^c$, on a $x \notin \overline{A^c}$ d'où $x \notin A^c$, et donc $x \in A$. Montrons pour finir que tout ouvert contenu dans A est contenu dans $x \in (\overline{A^c})^c$. Soit U un ouvert contenu dans A . Il existe une boule $B(x, \varepsilon) \subset U \subset A$. Le point x n'est donc pas dans l'adhérence du complémentaire de A (aucune suite du complémentaire de A ne peut tendre vers x), il appartient donc à $(\overline{A^c})^c$. Cet ouvert contient donc tous les ouverts contenus dans A , c'est donc le plus grand, à savoir l'intérieur de A .

Exercice I.9.5. (•) Soit (X, d) un espace métrique, et A une partie de X . La distance d'un point x à l'ensemble est notée $d(x, A)$ (voir définition I.2.3 de la distance à un ensemble). Montrer

$$\begin{aligned}\bar{A} &= \{x \in X, d(x, A) = 0\}, \quad \mathring{A} = \{x \in X, d(x, A^c) > 0\}, \\ \partial A &= \{x \in X, d(x, A) = d(x, A^c) = 0\}.\end{aligned}$$

CORRECTION.

Tout point de \bar{A} est limite d'une suite (x_n) de points de A (proposition I.4.6), on a donc

$$0 \leq d(x, A) \leq d(x, x_n) \rightarrow 0,$$

d'où $d(x, A) = 0$. Réciproquement, si $d(x, A) = 0$, il existe une suite minimisante (x_n) d'éléments de A telle que $d(x, x_n)$ tend vers 0, d'où $d(x, A) = 0$.

Si $x \in \mathring{A}$, il existe une boule $B(x, \varepsilon) \subset A$, la distance de x à tout point de A est donc supérieure à ε , d'où $d(x, A) \geq \varepsilon > 0$. Inversement, si $d(x, A^c) = \alpha > 0$, alors tous les points y tels que $d(x, y) < \alpha$ sont dans A , i.e. $B(x, \alpha) \subset A$.

Par définition, la frontière est l'ensemble des points de l'adhérence de A (donc tels que $d(x, A) = 0$ d'après ce qui précède), qui ne sont pas dans \mathring{A} (et donc tels que $d(x, A^c) = 0$).

Exercice I.9.6. (••) On se place sur $X = \mathbb{R}^2$ muni de la distance euclidienne.

- a) Donner un exemple de partie $A \subset \mathbb{R}^2$ telle que la distance de x est atteinte pour certains points, et pas pour d'autres.
 b) Donner un exemple de partie A , et d'un point $x \in \mathbb{R}^2$, tels que la distance de x à A est atteinte en plusieurs points.
 c) (Cellules de Voronoï)

On considère la situation où A est une collection finie de points : $A = \{x_i\}_{1 \leq i \leq N}$. On appelle A_i l'ensemble des points qui sont strictement plus près de x_i que des autres x_j , autrement dit les points x tels que la

distance de x à A est atteinte en x_i , et en x_i seulement. Décrire les A_i , appelées cellules de Voronoï dans les cas suivants

(i) Les x_i sont tous situés sur le premier axe de coordonnées.

(ii) Les x_i sont équidistribués sur le cercle unité.

Dans le cas général de points distribués de façon quelconque, faire un dessin de ces cellules de Voronoï.

d) (*) Pourquoi parle-t-on de téléphone *cellulaire* pour désigner un téléphone portable ?

CORRECTION.

a) Il suffit de considérer par exemple le segment $[0, 1] \times \{0\}$. La distance est atteinte pour tout point de $\{(x, y), x < 1\}$, non atteinte pour les autres points.

b) Il suffit de prendre un ensemble non convexe, par exemple la réunion de deux singletons disjoints. En chaque point de la médiatrice, la distance est atteint en deux points. On peut aussi avoir une distance atteinte en une infinité de points. Considérer par exemple le cercle unité. La distance est atteinte en un seul point pour tous les x non nuls, mais en tous les points de l'ensemble pour $x = (0, 0)$.

c)

d) On considère l'ensemble des stations-relais comme une famille de points du plan. Lorsque l'on passe un appel, le téléphone passe par la borne la plus proche. La zone couverte par une station données (en les supposant toutes de même puissance) est donc la cellule de Voronoï associée à la position de la station. Le terme de cellule provient du fait que la zone ressemble à une cellule organique (cellules de peau d'oignon par exemple).

Exercice I.9.7. (•)

a) Donner un exemple de suite réelle telle que $|x_{n+1} - x_n|$ tend vers 0, mais qui ne converge pas dans \mathbb{R} .

b) Proposer une procédure pour construire une telle suite, qui soit telle que l'ensemble de ses termes soit de plus dense dans \mathbb{R} .

CORRECTION.

a) On peut considérer par exemple la suite $x_n = \log(n)$.

b) On considère une suite obtenue en partant de $u_0 = 1$, puis $u_1 = u_0 + 1/2$, $u_2 = u_1 + 1/3$, $u_3 = u_2 + 1/4$, qui dépasse deux, on change alors de direction $u_4 = u_3 - 1/5$, etc..., jusqu'à passer en dessous de -3 , puis on repart vers la droite jusqu'à dépasser 4 , etc Cette suite balaie des intervalles de plus en plus grands, avec un pas qui tend vers 0, d'où la densité.

Exercice I.9.8. (Suite décroissante de compacts (•))

Soit $(K_n)_{n \in \mathbb{N}}$ décroissante de compacts non vides d'un espace métrique X . Montrer que l'intersection des K_n est non vide.

CORRECTION.

Pour tout n on choisit x_n dans K_n . La suite (x_n) est dans K_0 compact, elle admet donc une sous-suite $(x_{\varphi(n)})$ convergente vers $x \in K_0$. Pour tout n , la suite extraite est dans K_n au-delà d'un certain rang. Comme K_n est compact, il est fermé, la limite x est donc dans K_n . L'élément x appartient donc à tous les K_n .

Exercice I.9.9. (••)

Soit F un fermé non vide de \mathbb{R}^d . Montrer que la distance de tout x à F (définition I.2.3, page 12)) est atteinte.

CORRECTION.

a) L'application $y \mapsto d(x, y)$ est continue sur le compact K , elle atteint donc ses bornes, en particulier sa borne inférieure.

b) Si F est un fermé non vide, on choisit arbitrairement $y_0 \in F$. La distance de x à F est par définition inférieure ou égale à $d(x, y_0)$. Elle s'écrit donc comme infimum de $d(x, \cdot)$ sur

$$F \cap \{y, d(x, y) \leq d(x, y_0)\}.$$

Le second ensemble est fermé comme image réciproque du fermé $]-\infty, d(x, y_0)]$ par l'application $d(x, \cdot)$, c'est donc un fermé, et il est borné par définition. L'intersection ci-dessus est donc un fermé comme intersection de fermés, et borné, c'est un compact. La fonction distance atteint donc ses bornes, d'où l'on déduit que l'infimum est atteint.

Exercice I.9.10. (Distance de Hausdorff (•••))

Soit (X, d) un espace métrique. On note \mathcal{K} l'ensemble des parties compactes non vides de X . Pour tous K_1, K_2 dans \mathcal{K} , on définit la quantité

$$d_H(K_1, K_2) = \max \left(\sup_{x_1 \in K_1} d(x_1, K_2), \sup_{x_2 \in K_2} d(x_2, K_1) \right).$$

- a) Montrer que les sup dans l'expression ci-dessus sont en fait des max, et que $d_H(\cdot, \cdot)$ définit une distance sur \mathcal{K} .
- b) Explorer la possibilité de définir une telle quantité afférente à deux ensembles si l'on ne se restreint pas à des compacts.
- c) Donner l'expression de la distance d'un compact K à sa propre frontière ∂K pour les formes géométriques suivantes de \mathbb{R}^2 (muni de la distance euclidienne standard) : cercle, segment, disque, carré, rectangle, ellipse, triangle.
- d) Expliquer comment cette notion peut être utilisée pour *métriser* l'ensemble des fonctions continues de l'intervalle $[0, 1]$ dans \mathbb{R} , et vérifier que la distance ainsi construite diffère de celle issue la norme de la convergence uniforme définie par $d_\infty(f, g) = \max_{[0,1]} |f(x) - g(x)|$.
- e)(•••) Montrer que si X est complet, alors l'espace métrique (\mathcal{K}, d_H) l'est aussi.
- f) Montrer que tout compact K non vide peut être approché avec une précision arbitraire en distance de Hausdorff par un ensemble fini.

CORRECTION.

a) La distance d'un point à une partie non vide est une application continue (voir exercice I.7.3, page 29). L'application $x_1 \mapsto d(x_1, K_2)$ atteint donc ses bornes, en particulier le sup, sur le compact K_2 (proposition I.7.4, page 28). Le second sup est de la même manière un max.

Si le premier sup est nul, cela signifie que $K_1 \subset K_2$, si le second est nul on a l'inclusion inverse. L'annulation de la distance impose donc l'identité des ensembles. La symétrie est évidente par construction. Pour l'inégalité triangulaire, on considère 3 compacts K_1, K_2 , et K_3 . Pour tout $x_1 \in K_1$, tout $x_2 \in K_2$, on a

$$d(x_1, K_3) \leq d(x_1, x_2) + d(x_2, K_3) \leq d(x_1, x_2) + \sup_{K_2} d(x'_2, K_3),$$

d'où, en prenant l'inf en x_2 ,

$$d(x_1, K_3) \leq d(x_1, K_2) + \sup_{K_2} d(x'_2, K_3) \implies \sup_{K_1} d(x_1, K_3) \leq \sup_{K_1} d(x_1, K_2) + \sup_{K_2} d(x'_2, K_3)$$

On a donc (on majore de la même manière le second terme du max)

$$\begin{aligned} d_H(K_1, K_3) &\leq \max \left(\sup_{K_1} d(x_1, K_2) + \sup_{K_2} d(x'_2, K_3), \sup_{K_3} d(x_3, K_2) + \sup_{K_2} d(x'_2, K_1) \right) \\ &\leq \max \left(\sup_{K_1} d(x_1, K_2), \sup_{K_2} d(x_2, K_1) \right) + \max \left(\sup_{K_3} d(x_3, K_2), \sup_{K_2} d(x_2, K_3) \right) \\ &= d_H(K_1, K_2) + d_H(K_2, K_3). \end{aligned}$$

b) On peut définir une telle quantité pour 2 parties bornées non vides d'un espace métrique (si l'on est pas bornée, certaines des quantités peuvent être infinies), les propriétés de symétrie et d'inégalité triangulaire sont préservées en général, mais la séparation n'est plus assurée si l'on n'a pas la compacité.

c) Pour un cercle, c'est 0, un segment 0 aussi (c'est la demi-longueur si on considère un segment de \mathbb{R} , un disque son rayon, un rectangle son demi petit côté, une ellipse son demi petit axe, et un triangle le rayon du cercle inscrit).

d) Le graphe d'une application continue de $[0, 1]$ dans \mathbb{R} est un compact de \mathbb{R}^2 . En effet, si l'on se donne une suite $(x_n, f(x_n))$, on peut extraire un sous-suite $x_{n'}$ qui converge vers x , et alors $(x_{n'}, f(x_{n'}))$ converge vers $(x, f(x))$ par continuité de f . On peut ainsi définir la distance entre deux fonctions comme la distance de Hausdorff entre leurs graphes. Cette distance est toujours inférieure ou égale à celle issue de la norme uniforme, elles sont en général différentes, à tel point que l'ensemble X des graphes de fonctions continues sur $[0, 1]$ n'est pas complet pour cette distance. Considérer par exemple la suite

$$x \mapsto (1 - nx)_+.$$

Cette suite est de Cauchy pour la distance des graphes, mais converge vers un compact qui n'est pas le graphe d'une fonction continue.

e) La démonstration est un peu laborieuse. On considère une suite (K_n) de Cauchy dans \mathcal{K} . On définit K comme l'ensemble des $x \in X$ tels qu'il existe une suite (x_n) , avec $x_n \in K_n$, qui converge vers x .

Montrons pour commencer un petit lemme, à savoir que l'on peut se façon équivalente considérer les x tels qu'il existe une suite (n_k) d'entiers strictement croissante telle que x_{n_k} converge vers x , avec $x_{n_k} \in K_{n_k}$ (autrement dit : on n'est pas forcée d'avoir tous les K_n). La condition nécessaire est évidente. Maintenant si x_{n_k} converge vers x , on peut compléter les indices manquant en considérant, pour $n_k \leq j < n_{k+1}$, x_j un élément de K_j qui réalise la distance de x_{n_k} à K_j . On a

$$d(x_{n_k}, x_j) \leq d_H(K_j, K_{n_k}) \quad \forall j, n_k \leq j < n_{k+1},$$

d'où convergence de $d(x_{n_k}, x_j)$ vers 0 quand k tend vers $+\infty$, uniformément en j , du fait du caractère de Cauchy de la suite K_n .

Montrons maintenant que K est non vide. Il existe n_1 tel que, pour tous $p, q \geq n_1$, $d_H(K_p, K_q) < 1/2$. On choisit x_{n_1} dans K_{n_1} . Il existe n_2 tel que, pour tous $p, q \geq n_2$, $d_H(K_p, K_q) < 1/4$. On choisit x_{n_2} dans K_{n_2} tel que $d(x_{n_2}, x_{n_1}) < 1/2$. C'est possible car $d_H(K_{n_2}, K_{n_1}) < 1/2$. On construit ainsi itérativement une suite (x_{n_k}) , avec $x_{n_k} \in K_{n_k}$. Cette suite est de Cauchy par construction, elle converge donc vers $x \in X$. La limite x est dans K d'après le lemme qui précède.

Nous allons montrer que K est fermé, puis que de toute suite de K on peut extraire une sous-suite qui est de Cauchy, cela établira la compacité. On considère donc une suite (x^k) dans K . Par définition de K , pour tout k , x^k est limite d'une suite (x_n^k) , avec $x_n^k \in K_n$ pour tout n . On choisit n_1 tel que $d(x_{n_1}^1, x^1) < 1$, $n_2 > n_1$ tel que $d(x_{n_2}^2, x^2) < 1/2$, ... $n_{k+1} > n_k$ tel que $d(x_{n_{k+1}}^{k+1}, x^k) < 1/k$, ... On construit ainsi une suite $(x_{n_k}^k)$ qui est adjacente à x^k , avec $x_{n_k}^k \in K_{n_k}$. Sa limite est donc x , qui est donc dans K (toujours d'après le lemme précédent).

L'avant-dernière étape de la preuve consiste à montrer que la quantité

$$\sup_K d_H(K, K_n)$$

(même si l'il n'est pas encore établi que K est compact, l'expression a bien un sens pour les ensembles non vides) tend vers 0 quand n tend vers $+\infty$. Montrons tout d'abord que $\sup_K d(x, K_n)$ tend vers 0. Pour tout $\varepsilon > 0$, il existe N tel que, pour tous $p, q \geq N$, $d_H(K_p, K_q) < \varepsilon$. Tout $x \in K$ est limite d'une suite x_n , avec $x_n \in K_n$. Considérons un $m \geq N$ tel que $d(x, x_m) < \varepsilon$. Pour tout $n \geq N$, on a

$$d(x, K_n) \leq d(x, x_m) + d_H(K_m, K_n) < 2\varepsilon.$$

On a donc bien convergence de $\sup_K d(x, K_n)$ vers 0. Montrons maintenant que la quantité symétrique $\sup_{K_n} d(x, K)$ tend vers 0. Pour tout ε , il existe N tel que $d_H(K_p, K_q) < \varepsilon$ pour tous $p, q \geq N$. Pour tout $n \geq N$, tout $y \in K_n$, on cherche à construire une suite partant de y , qui reste proche de y , et qui converge vers un élément de K . On prend $y_0 = y$. Il existe $N_1 > N_0 = N$ tel que $d_H(K_{N_0}, K_{N_1}) < \varepsilon/2$ pour tous $p, q \geq N$. On prend y_1 dans K_{N_1} tel que $d(y_1, h_0) < \varepsilon$. On construit ainsi des suites N_k et y_k , avec $y \in K_{N_k}$, et $d(y_{k+1}, y_k) < \varepsilon/2^k$. La suite y_k est de Cauchy, elle converge donc vers un élément $x \in K$, à distance de y inférieure ou égale à 2ε par construction. On a donc bien montré la convergence de $\sup_{K_n} d(x, K)$ vers 0, d'où la convergence de K_n vers K en distance de Hausdorff (même si, comme évoqué précédemment, nous

n'avons pas encore montré de K était compact).

Il reste à montrer que, de toute suite (x^k) dans K , on peut extraire une sous-suite de Cauchy. On construit en premier lieu, grâce à ce qui précède, une suite strictement croissante d'indices (n_j) telle que $\sup_K d(x, K_{n_j}) < 1/j$. On construit maintenant une suite $(x_{n_1}^k)_k$ dans K_{n_1} , telle que $x_{n_1}^k$ réalise la distance de x^k à K_{n_1} , on a donc $d(x^k, x_{n_1}^k)$ d'après la définition de n_1 . Comme K_{n_1} est compact, on peut en extraire une suite qui converge dans K_{n_1} , il existe donc une application croissante $k \mapsto j_1(k)$ telle que $d(x_{n_1}^{j_1(p)}, x_{n_1}^{j_1(q)}) < \varepsilon$ pour tous p, q . On a

$$d(x^{j_1(p)}, x^{j_1(q)}) \leq d(x^{j_1(p)}, x_{n_1}^{j_1(p)}) + d(x_{n_1}^{j_1(p)}, x_{n_1}^{j_1(q)}) + d(x_{n_1}^{j_1(q)}, x^{j_1(q)}) < 3 \times 1.$$

On extrait de la même manière une suite $x^{j_1 \circ j_2(k)}$ telle que

$$d(x^{j_1 \circ j_2(p)}, x^{j_1 \circ j_2(q)}) < 3 \times \frac{1}{2},$$

puis itérativement des suites $x^{j_1 \circ j_2 \circ \dots \circ j_n(k)}$, avec

$$d(x^{j_1 \circ j_2 \circ \dots \circ j_n(p)}, x^{j_1 \circ j_2 \circ \dots \circ j_n(q)}) < 2 \times \frac{1}{n}.$$

Par extraction diagonale, on introduit $\psi(k) = j_1 \circ j_2 \circ \dots \circ j_k(k)$. La suite $x^{\psi(k)}$ ainsi extraite vérifie, pour tous $p < q$,

$$d(x^{\psi(p)}, x^{\psi(q)}) < 3 \times \frac{1}{p},$$

elle est donc de Cauchy.

La convergence de K_n vers le compact K pour la distance de Hausdorff a déjà été établie précédemment.

f) Soit K un compact non vide. Pour tout $\varepsilon > 0$, on considère le recouvrement de K par les boules $B(x, \varepsilon)$, pour x parcourant K . On peut en extraire un recouvrement fini. Notons K_ε l'ensemble des centres des boules constituant ce recouvrement fini. Par construction K_ε est à distance de Hausdorff de K inférieure à ε .

Exercice I.9.11. (••) Soit (X, d) un espace métrique, et A une partie non vide de X . La distance d'un point x de X à A est définie (voir définition I.2.3) comme l'infimum des distances de x à a , pour a parcourant A .

Montrer que l'application $d(\cdot, A)$ de X dans \mathbb{R}_+ qui à x associe $d(x, A)$ est $1 - lipschitzienne$, c'est à dire que

$$|d(x, A) - d(y, A)| \leq d(x, y).$$

CORRECTION.

On s'intéresse à la continuité de l'application

$$x \in X \longmapsto d(x, A) = \inf_{a \in A} d(x, a).$$

Pour tout $a \in A$, on a

$$d(x, A) \leq d(x, a) \leq d(x, y) + d(y, a).$$

En appliquant cette inégalité à une suite minimisante pour $d(y, A)$, c'est à dire une suite (a_n) dans A telle que $d(y, A) = \lim d(y, a_n)$, on obtient

$$d(x, A) \leq d(x, y) + d(y, A).$$

On montre en échangeant les rôles de x et y que

$$d(y, A) \leq d(x, y) + d(x, A).$$

On a donc montré

$$\max(d(x, A) - d(y, A), d(y, A) - d(x, A)) = |d(x, A) - d(y, A)| \leq d(x, y),$$

qui exprime le caractère $1 - lipchitzien$ de $d(\cdot, A)$.

Exercice I.9.12. On se place sur \mathbb{R}^2 . Analyser le problème consistant à minimiser $\|x\|_1$ sur un demi plan (on pourra faire un dessin).

Généraliser

CORRECTION.

On se place dans le cas où le demi-plan ne contient pas l'origine (sinon le minimiseur est 0), le mieux est de tracer la plus grosse "boule" centrée en 0, c'est à dire ici un carré donc les côtés sont orientés à $\pi/4$ et $3\pi/4$, dont l'intersection avec le demi-plan admissible soit d'intérieur vide. On voit que le minimum est en général réalisé par un coin du carré (donc un point sur l'un des axes), sauf si la frontière du demi plan est alignée avec l'un des côtés, auquel cas le minimum n'est pas unique, car réalisé par tous les points de l'un des côtés. Cela illustre l'intérêt de l'utilisation de rajouter cette norme ℓ^1 à certains problèmes de minimisation sous contraintes (approche de type lasso), car ce terme a tendance à diminuer le nombre de coefficients non nuls pour le minimiseurs (approche parcimonieuse, voir aussi exercice I.9.13). En dimension 3 le minimiseur est en général l'un des sommets d'un cube, mais cela peut-être aussi dans certains cas particuliers une arête, ou une face.

Exercice I.9.13. (••) a) Pourquoi a-t-on exclu le cas $p \in [0, 1[$ dans la définition des normes p ? Qu'advient-il de la quantité

$$C_P(x) = \sum_{k=1}^d |x_k|^p$$

quand, pour un $x \in \mathbb{R}^d$ donné, on fait tendre p vers 0? On notera $C_0(x)$ cette limite

b) Quel peut être l'intérêt de considérer de telles expressions pour p petit?

CORRECTION.

a) On peut bien sûr définir la quantité $\sum |x_k|^p$ pour $p < 1$, on garde les propriétés de séparation et de symétrie, mais on perd l'inégalité triangulaire. On pourra par exemple considérer le cas $p = 1/2$, et $x = (1, 0)$, $y = (0, 1)$, on a

$$\|x\|_{1/2} = 1, \|y\|_{1/2} = 1, \|x + y\|_{1/2} = 4 > \|x\|_{1/2} + \|y\|_{1/2}.$$

Lorsque l'on fait tendre p vers 0 dans $\sum |x_k|^p$, chaque terme nul reste nul, et les termes non nuls tendent vers 1, on converge donc vers un entier positif qui est le nombre de termes non nuls parmi les composantes de x .

b) Bien que cela sorte du cadre de la norme, il s'agit d'une quantité (on parle de norme 0), qui peut être très intéressante à utiliser dans un contexte d'optimisation. Si l'on cherche à minimiser une fonctionnelle par rapport à un vecteur x de \mathbb{R} , rajouter à la fonctionnelle un terme proportionnel à $\|x\|_0$ aura tendance à minimiser cette quantité, et donc à limiter le nombre de termes non nul dans le minimum trouvé, ce qui peut être très intéressant si le problème d'optimisation consistant par exemple à approcher une fonction par une combinaison de fonctions particulières, dont les coefficient sont les x_i . On parle d'approximation parcimonieuse quand on cherche de la sorte à approcher quelque chose par une combinaison de constituants élémentaires de façon en quelque sorte économique (l'approximant sera en particulier plus léger à stocker sur un ordinateur).

Exercice I.9.14. (••) Préciser les valeurs des constantes d'équivalences optimales entre la norme ∞ et les différentes normes p , pour $p \in [1, +\infty[$.

Qu'advient-il de ces constantes lorsque la dimension n de l'espace tend vers $+\infty$?

CORRECTION.

a) On a, pour tout $x \in \mathbb{R}^d$,

$$\|x\|_\infty \leq \|x\|_p \leq n^{1/p} \|x\|_\infty.$$

On réalise l'égalité à droite pour $x = (1, 1, \dots, 1)$, et l'égalité à gauche pour $x = (1, 0, \dots, 0)$.

b) La constante de l'inégalité de droite tend vers $+\infty$ quand n tend vers $+\infty$, ce qui suggère que les normes ne sont pas équivalentes en dimension infinies.

Exercice I.9.15. (••) La direction d'une école d'ingénieurs prestigieuse décide facétieusement de changer sa procédure de calcul des moyennes, en la remplaçant par (on désigne par x_1, \dots, x_n les notes d'un élève)

$$m_p = \frac{1}{n^{1/p}} \|x\|_p,$$

pour un certain $p \in]1, +\infty]$.

- a) Justifiez le fait qu'il s'agit bien d'une moyenne, et expliquer pourquoi l'on peut s'attendre à ce que les élèves se réjouissent à première vue de cette nouvelle.
- b) S'agissant d'un concours, seul le classement est véritablement important. Dans cette optique, expliquer pourquoi certains compétiteurs puissent se sentir lésés, d'autres au contraire avantagés, en précisant les profils de ces deux types d'élèves.

CORRECTION.

a) On a

$$m_p = \frac{1}{n^{1/p}} \|x\|_p = \frac{1}{n^{1/p}} \left(\sum x_k^p \right)^{1/p} \leq \frac{1}{n^{1/p}} (n \max(x_k)^p)^{1/p} = \max(x_k),$$

et de la même manière $m_p \geq \min(x_k)$. Il s'agit bien d'un nombre compris entre la valeur min et la valeur max.

On a par ailleurs, d'après l'inégalité de Hölder (proposition A.1.44, page 184),

$$\sum_{k=1}^n |x_k y_k| \leq \left(\sum_{k=1}^d \theta_i |x_k|^p \right)^{1/p} \left(\sum_{k=1}^d \theta_i |y_k|^q \right)^{1/q},$$

avec $1/p + 1/q = 1$. En appliquant cette inégalité à (x_k) et le vecteur constant égal à 1, on obtient

$$m_1 = \frac{1}{n} \sum_{k=1}^n |x_k y_k| \leq \frac{1}{n} \left(\sum_{k=1}^n \theta_i |x_k|^p \right)^{1/p} n^{1/q},$$

d'où

$$n^{1/q-1} \left(\sum_{k=1}^n \theta_i |x_k|^p \right)^{1/p} \geq m_1.$$

Comme $1/q - 1 = -1/p$, on en déduit $m_p \geq m_1$. Chaque élève voit donc sa moyenne augmenter par rapport au calcul classique.

b) Le fait d'utiliser une somme des notes à la puissance $p > 1$ renforce l'importance des notes élevées, d'autant plus que p est grand. La moyenne 1 habituellement utilisée favorise les élèves qui ne sacrifient aucune épreuve, alors qu'un p grand favorisera les profils plus spécialisés, qui brillent à certaines épreuves quitte à en sacrifier d'autres. Dans le cas extrême $p = +\infty$, on ne garde que la meilleure note. Dans le cas $n = 10$, un élève ayant 20 à une épreuve, 0 aux autres, aura une moyenne m_1 de 2, et une moyenne m_∞ de 20, alors qu'un élève ayant 10 à toutes les épreuves aura une moyenne m_1 de 10 (donc largement au-dessus du premier), et une moyenne m_∞ de 10 (très en-dessous du premier).

Exercice I.9.16. (••) Soit f une fonction continue de \mathbb{R}^n dans \mathbb{R} .

a) Montrer que si l'on a

$$\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$$

alors l'image réciproque par f de tout compact est un compact.

b) Montrer réciproquement que si l'image réciproque par f de tout compact est un compact, alors $|f(x)|$ tend vers $+\infty$ quand $\|x\|$ tend vers $+\infty$.

c)(•••) Montrer que, si $n \geq 2$, alors la conclusion de la question précédente peut être précisée : $f(x)$ converge vers $+\infty$ quand $\|x\|$ tend vers $+\infty$, ou $f(x)$ converge vers $-\infty$ quand $\|x\|$ tend vers $+\infty$. Qu'en est-il du cas $n = 1$?

d) Soit f une fonction continue qui vérifie la propriété du a) (on dit que f est coercive, ou parfois dans certains contextes que f est propre). Montrer que f est minorée sur \mathbb{R}^n , et qu'elle atteint son minimum.

CORRECTION.

- a) Soit K compact. En premier lieu, f étant continue, l'image réciproque de tout fermé est fermée (proposition I.7.2, page 27), l'image réciproque du fermé K est donc fermée. Si elle n'était pas bornée, on pourrait construire une suite (x_n) de points de $f^{-1}(K)$ non bornée, dont l'image serait non bornée par hypothèse, ce qui est absurde. L'image réciproque de K est donc compacte comme fermé borné de \mathbb{R}^n .
- b) Soit (x_n) une suite de \mathbb{R}^n telle que $\|x\|$ tend vers $+\infty$. Si la suite des images est bornée, alors on peut l'inclure dans un intervalle fermé (donc compact comme fermé borné), dont l'image réciproque est compacte donc bornée, alors qu'elle est censée contenir tous les x_n , ce qui est absurde.
- c) On sait que $|f(x)|$ tend vers $+\infty$ quand $\|x\|$ tend vers $+\infty$. Il existe donc en particulier un $R > 0$ tel que, pour tout x de norme plus grande que R , $|f(x)| \geq 1$. Si f prend des valeurs à la fois positives et négatives à l'extérieur de B_R , on peut trouver deux points x et x' de norme plus grande que R tels que, par exemple, $f(x) > 0$ et $f(x') < 0$. On relie alors les deux points par un chemin continu qui évite la boule (si le segment ne convient pas, on fait le tour). La restriction de f à cette ligne est continue, donc prend la valeur 0 en un certain point (d'après le théorème des valeurs intermédiaires), ce qui est absurde puisque l'on doit avoir $\|f(x)\| \geq 1$ en tout point de la ligne.

La dimension 1 est d'une certaine manière pathologique, car on ne peut pas faire le tour de 0 pour passer d'un nombre négatif à un nombre positif. De fait, la conclusion est invalidée : l'application identité vérifie bien l'hypothèse que l'image réciproque de tout compact est compacte, et pourtant la fonction tend bien vers $+\infty$ ou $-\infty$ selon la direction que l'on prend.

- d) Soit f une fonction coercive au sens du (a). Soit $x \in \mathbb{R}^n$ (par exemple 0). Par hypothèse, $f(x) > f(0)$ pour $\|x\| \geq C$. L'infimum $m \in [-\infty, +\infty[$ de f sur \mathbb{R}^n est donc l'infimum de f sur $\overline{B}(0, C)$. Comme f est continue sur le compact $\overline{B}(0, C)$, elle est bornée et atteint ses bornes, on a donc en particulier $m > -\infty$, et cet infimum est atteint.

Exercice I.9.17. (••) Donner un exemple d'application de \mathbb{R}_+ dans \mathbb{R}_+ qui vérifie $|T(x) - T(y)| < |x - y|$ pour tous $x \neq y$, mais qui n'est pas contractante.

CORRECTION.

L'application $x \mapsto f(x) = 1 - e^{-x}$ sur \mathbb{R}_+ est de dérivée positive, strictement inférieure à 1 sur $]0, +\infty[$. Pour tous $x \neq y$ dans \mathbb{R}_+ , il existe c entre x et y tel que, d'après le théorème des accroissements finis, on ait

$$|f(x) - f(y)| = f'(c) |x - y| < |x - y|.$$

On a aussi

$$\lim_{h \rightarrow 0} \frac{f(h) - f(0)}{h} = f'(0) = 1,$$

il ne peut donc exister aucun $\kappa \in [0, 1[$ tel que f soit κ -contractante.

Exercice I.9.18. (••) Soit (X, d) un espace métrique complet et T une application de (X, d) dans lui-même. On suppose qu'il existe $p \in \mathbb{N}$ tel que

$$T^p = \underbrace{T \circ T \circ \cdots \circ T}_{p \text{ fois}}$$

soit contractante. Montrer que T admet un unique point fixe.

CORRECTION.

L'application T^p étant contractante, elle admet un point fixe x unique : $T^p x = x$. On a donc, en composant cette identité par p , $T^p(T(x)) = T(x)$, d'où l'on tire que $T(x)$ est aussi point fixe de T^p . On a donc $T(x) = x$ par unicité du point fixe de T^p . Tout point fixe de T étant aussi point fixe de T^p , on en déduit l'unicité.

Exercice I.9.19. (•)

- a) Quelle est la norme (subordonnée à la norme euclidienne) d'une application de \mathbb{R}^n dans lui-même représentée dans la base orthonormée canonique de \mathbb{R}^n par une matrice diagonale $A = \text{diag}(a_1, \dots, a_n)$?
- b) Quelle est la norme d'une application représentée dans la base orthonormée canonique de \mathbb{R}^n par une matrice A symétrique ?

CORRECTION.

a) Pour tout $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, on a

$$\|Ax\|^2 = \sum a_i^2 x_i^2 \leq \max(|a_i|^2) \sum x_i^2.$$

Si l'on considère maintenant le j qui réalise le max des $|a_i|$, et que l'on prend $x_j = 1$, et $x_i = 0$ pour $i \neq j$, le vecteur ainsi défini réalise l'égalité. On a donc $\|A\| = \max(|a_i|)$.

b) Toute matrice symétrique est diagonalisable dans une base orthonormée (v_i). Tout vecteur x de \mathbb{R}^n peut s'écrire

$$x = \sum x_i v_i \implies Ax = \sum x_i \lambda_i v_i,$$

et donc

$$\|Ax\|^2 = \sum |x_i|^2 |\lambda_i|^2 \leq \max(|\lambda_i|^2) \sum |x_i|^2 = \max(|\lambda_i|^2) \|x\|^2,$$

et, comme précédemment, le vecteur $x = v_j$ (où j réalise le max des $|\lambda_i|$) réalise l'égalité. La norme d'opérateur de A est donc le max des valeurs absolues des valeurs propres de A .

Exercice I.9.20. (••)

On considère l'espace vectoriel des suites bornées

$$\ell^\infty = \{u = (u_n) \in \mathbb{R}^{\mathbb{N}}, \sup |u_n| < +\infty\},$$

muni de la norme de la convergence uniforme $\|u\| = \sup |u_n|$.

Montrer que cet espace est complet, et que sa boule unité fermée n'est pas compacte.

CORRECTION.

On considère une suite $(x^k)_k = ((x_n^k)_n)_k$. Pour tout n , la suite $(x_n^k)_k$ est de Cauchy dans \mathbb{R} complet, elle converge donc vers un x_n . Comme (x_k) est bornée dans ℓ^∞ , la suite réelle x_n est bornée, on a donc $x = (x_n) \in \ell^\infty$. La suite (x_k) étant de Cauchy, on a

$$\forall \varepsilon > 0 \quad \exists N, \forall p, q \geq N, \|x^p - x^q\| < \varepsilon,$$

c'est à dire

$$\|x_n^p - x_n^q\| < \varepsilon \quad n.$$

On fait tendre q vers l'infini, on a donc

$$\|x_n^p - x_n\| < \varepsilon \quad n$$

d'où $\|x^p - x\| \leq \varepsilon$. On a donc convergence de (x^k) vers x pour la norme $\|\cdot\|_\infty$.

On note $e_n = (0, 0, \dots, 0, 1, 0, \dots)$. La suite des e_n est telle que deux termes de la suite sont à distance toujours égale à 2, on ne peut donc en extraire aucun sous-suite qui serait de Cauchy, donc aucune sous-suite convergente.

Chapitre II

Espaces fonctionnels, théorie de la mesure et de l'intégration

Sommaire

II.1	Convergences simple & uniforme	45
II.2	Espaces de fonctions continues	47
II.3	De Riemann à Lebesgue	48
II.3.1	L'intégrale de Riemann et ses limites	48
II.3.2	Théorie de la mesure et intégrale de Lebesgue : un aperçu	51
II.3.3	Mesure et intégration	52
II.4	Les espaces L^p	60
II.4.1	L'espace $L^\infty(X)$	61
II.4.2	Les espaces $L^p(X)$, pour $p \in [1, +\infty[$	62
II.4.3	Les espaces $L^p(\mathbb{N}) = \ell^p$ et $L^p(\mathbb{R}^d)$	64
II.5	Compléments	66
II.6	Exercices	67

Nous introduisons dans ce chapitre les espaces fonctionnels (espaces vectoriels normés constitués de fonctions) utilisés en analyse et analyse numérique¹ des équations aux dérivées partielles, ainsi qu'en optimisation. Les espaces de fonctions continues, ou plusieurs fois continûment différentiables, sont décrits en premier lieu. Ces espaces ne fournissent pas un cadre satisfaisant pour la plupart des Équations aux Dérivées Partielles (EDP) de la physique. Il est ainsi nécessaire de construire des normes basées sur des intégrales en espace, qui correspondent pour la plupart de ces EDP à des quantités pertinentes physiquement. Nous précisons en quoi l'intégrale de Riemann ne permet pas de définir des espaces fonctionnels dotés de bonnes propriétés. Nous présentons ensuite une approche alternative, basée sur la théorie de la mesure et de l'intégrale de Lebesgue, qui permet de construire des espaces de fonctions ayant de bonnes propriétés, en particulier de complétude. La construction détaillée de cette intégrale de Lebesgue est reportée en annexe B, nous nous contentons ici d'en esquisser les principaux éléments, qui permettent la construction des espaces fonctionnels de type L^p développés à la fin de ce chapitre.

II.1 Convergences simple & uniforme

Nous rappelons dans cette section les notions très générales de convergence *uniforme* et de convergence *simple* pour des suite d'applications à valeurs dans un espace métrique. On notera que les définitions ne

1. L'analyse numérique consiste à étudier la convergence des méthodes d'approximation numérique des équations différentielles et équations aux dérivées partielles, qui permet d'assurer que les approximations numériques produites par calcul effectif sur ordinateur convergent bien vers la solution exacte de l'équation considérée. Ces solutions exactes étant des fonctions d'une ou plusieurs variables, elle nécessitent de définir des normes adaptées sur ces espaces de fonctions.

nécessitent aucune structure sur l'ensemble de départ. La propriété de convergence uniforme, très forte, jouera un rôle essentiel dans les espaces de fonctions continues introduits ci-après (l'espace de départ sera alors muni d'une métrique, et les applications considérée seront continues). Les espaces construits sur une norme intégrale, qui font l'objet des sections suivantes, feront jouer un rôle central à la notion plus faible de convergence simple (l'espace de départ sera alors *mesuré*, l'espace d'arrivée mesurable, et les applications considérées seront mesurables, dans un sens précisé ci-après).

Définition II.1.1. (Convergence simple, convergence uniforme (•))

Soit f une application d'un ensemble X dans un espace métrique (Y, d) , et (f_n) une suite d'applications de X dans Y . On dit que f_n converge *simplement* vers f si $f_n(x)$ converge vers $f(x)$ quand n tend vers $+\infty$, pour tout x , ce qui peut s'écrire

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

On dit que la convergence est *uniforme* si le N ne dépend pas du x , i.e.

$$\forall \varepsilon > 0, \exists N, \forall n \geq N, d(f_n(x), f(x)) < \varepsilon \quad \forall x \in X, .$$

Exercice II.1.1. Pour chacune des suites de fonctions de \mathbb{R} dans \mathbb{R} ci-dessous, indiquer sa limite simple si elle existe, et préciser si la convergence est uniforme (sur \mathbb{R} , sinon sur certaines parties de \mathbb{R}).

$$f_n^1(x) = \frac{1}{1 + (x - n)^2}, \quad f_n^2(x) = e^{-nx^2}, \quad f_n^3(x) = \frac{e^{-nx^2}}{1 + (x - n)^2}, \quad f_n^4(x) = \sin\left(\frac{2\pi x}{n}\right),$$

CORRECTION.

La suite (f_n^1) est la translatée de la fonction $1/(1+x^2)$, elle converge simplement vers 0. Mais pas uniformément : pour tout n il existe x (égale à n) en lequel f_n^1 vaut 1. Elle converge par contre uniformément vers 0 sur tout compact de \mathbb{R} et même sur toute intervalle du type $]-\infty, a]$.

La suite réelle $(f_n^2(x))$ converge vers 0 pour tout $x \neq 0$, et prend la valeur 1 pour $x = 0$. La convergence n'est pas uniforme car, f_n^2 étant continue en 0, il existe pour tout n un $x \neq 0$ tel que $f_n^2(x) > 1/2$. La convergence est uniforme sur tout ensemble à distance strictement positive de 0, c'est à dire sur tout ensemble dont l'adhérence ne contient pas 0, comme par exemple $]-\infty, -a] \cup [a, +\infty[$ pour $a > 0$. La suite (f_n^3) converge simplement vers 0, et cette convergence est uniforme. En effet, $0 \leq f_n^3(x) \leq 1/(1 + (1 - n)^2)$ sur $]-\infty, 1]$, et inférieur à $\exp(-n)$ sur $[1, +\infty[$.

La fonction f_n^4 est sinusoïdale périodique de période n , elle ne converge uniformément vers aucune fonction sur \mathbb{R} . En revanche, pour tout x , $\sin(2\pi x/n)$ converge vers 0. On a donc convergence simple vers la fonction nulle. Cette convergence vers la fonction nulle est uniforme sur tout ensemble borné.

Topologie de la convergence simple (• • •)

Même si nous n'utiliserons pas cette structure dans la suite du chapitre, précisons que la notion de convergence simple permet de définir une *topologie*² sur l'ensemble des applications d'un ensemble X dans un espace métrique (Y, d) , noté X^Y .

Proposition II.1.2. (Topologie de la convergence simple (• • •))

Soit X un ensemble et (Y, d) un espace métrique³. On dit que $F \subset X^Y$ est fermé si, pour toute suite (f_n) dans F qui converge simplement vers $f \in Y^X$, la limite f est dans F . On définit les ouverts comme complémentaires des fermés. On définit ainsi une *topologie* (au sens de la définition générale A.2.5, page 186) sur Y^X .

Démonstration. On considère une collection de fermés $(F_i)_{i \in I}$ de X^Y , et une suite (f^n) d'applications dans l'intersection de F_i , qui converge simplement vers $f \in Y^X$. Cette limite appartient à chacun des F_i , donc à l'intersection. On considère maintenant une collection *finie* de fermés $(F_i)_{i \in I}$, et une suite (f^n) d'applications dans l'union des F_i , qui converge simplement vers $f \in Y^X$. Comme I est fini, il existe au moins un i tel que F_i contient une infinité de termes de la suite. On extrait la sous-suite correspondante, qui converge simplement vers f , qui est donc nécessairement dans F_i , donc dans l'union. \square

2. Topologie au sens de la topologie générale (voir section A.2.3), qui n'est en général pas *métrisable*, c'est à dire qu'il n'existe pas de métrique qui conduirait à la même collection d'ouverts.

3. On pourrait immédiatement généraliser ce qui suit au cas d'un espace d'arrivée topologique, en se basant sur la notion de convergence d'une suite par les ouverts, voir définition A.2.19 page 187.

II.2 Espaces de fonctions continues

Proposition II.2.1. (Espace $C_b(X)$ (•))

Soit X un espace métrique. L'ensemble $C_b(X)$ des applications continues et bornées de X dans \mathbb{R} , muni de la norme

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

est un espace vectoriel normé *complet*, c'est à dire un *espace de Banach*⁴.

Démonstration. On vérifie immédiatement $\|\cdot\|$ est une norme. Considérons maintenant une suite de Cauchy (f_n) dans $C_b(X)$. Pour tout $x \in X$, $(f_n(x))$ est une suite de Cauchy dans \mathbb{R} , elle converge donc vers un réel que l'on note $f(x)$. La suite (f_n) étant de Cauchy, pour $\varepsilon > 0$ fixé, il existe N tel que pour tous p, q plus grands que N , on a

$$|f_p(x) - f_q(x)| < \varepsilon \quad \forall x \in X.$$

On fait tendre p vers $+\infty$, on obtient $|f(x)| \leq \varepsilon + M$ pour tout x , où M est un majorant de $|f_q|$, la fonction f est donc bornée. On peut donc étendre la norme du sup aux fonctions de type $g + f$, avec $g \in C_b(X)$, même si on n'a pas encore montré que f est continue.

De la même manière, pour tout $\varepsilon > 0$, il existe N tel que, pour tout $p \geq N$, on a

$$\|f_p - f\|_\infty \leq \varepsilon.$$

On a donc convergence de f_n vers f en norme $\|\cdot\|$.

Montrons maintenant que f est continue. Soit $x \in X$. Pour tout $\varepsilon > 0$, il existe N tel que, pour tout $n \geq N$,

$$\|f_n - f\|_\infty \leq \varepsilon.$$

La fonction f_N étant continue en x , il existe $\eta > 0$ tel que pour tout y à distance de x inférieure à η , $|f_N(x) - f_N(y)| < \varepsilon$. On a donc, pour un tel y

$$|f(x) - f(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| < 3\varepsilon,$$

d'où $f \in C_b(X)$. □

Noter que, lorsque X est compact, il n'est pas nécessaire d'imposer le caractère borné des fonctions. On écrira ainsi $C([0, 1])$ l'espace des applications continues de $[0, 1]$ dans \mathbb{R} (qui sont bornées sans qu'il soit nécessaire de le préciser).

Proposition II.2.2. (Espace $C_b^1([a, b])$ (••))

L'espace $C_b^1([a, b])$ des fonctions continûment différentiables sur $[a, b] \subset \mathbb{R}$, bornées et de dérivées bornées, est un espace vectoriel normé complet pour la norme

$$\|f\| = \sup_{x \in [a, b]} |f(x)| + \sup_{x \in [a, b]} |f'(x)|.$$

Démonstration. Soit f_n une suite de Cauchy dans $C_b([a, b])$. Comme dans la proposition précédente, f_n et f'_n convergent uniformément vers f et g , respectivement, fonctions continues et bornées. Montrons que f est continûment dérivable, et de dérivée g . Pour toute fonction φ continûment différentiable sur $[a, b]$, à support compact, on a

$$\int_a^b f'_n \varphi = - \int_a^b f_n \varphi' \quad \forall n \implies \int_a^b g \varphi = - \int_a^b f \varphi'.$$

Soit $x \in [a, b]$ et $h > 0$ tel que $[x - 2h, x + 2h] \subset [a, b]$. Pour $\varepsilon \in]0, h[$, on introduit une fonction régulière positive d'intégrale 1 supportée sur $[-\varepsilon, \varepsilon]$, et l'on définit φ_ε comme

$$\varphi_\varepsilon(y) = \int_a^b (\rho_\varepsilon(y) - \rho_\varepsilon(y - (x + h))) dy.$$

4. On se reportera au chapitre IV, page 91, pour une présentation générale de ce type d'espaces.

On a, pour tout $\varepsilon > 0$,

$$\int_a^b g\varphi_\varepsilon = - \int_a^b f\varphi'_\varepsilon,$$

d'où, en faisant tendre ε vers 0

$$\int_x^{x+h} g(y) dy = -f(x) + f(x+h),$$

d'où

$$\frac{f(x+h) - f(x)}{h} \rightarrow g(x).$$

On démontre de la même manière que la limite du taux de variation pour $h < 0$ est $g(x)$. La fonction f est donc dérivable en x , et de dérivée $g(x)$, pour tout $x \in]a, b[$, d'où la continuité différentiabilité du fait que g est continue. La convergence de f_n vers f pour la norme $\|\cdot\|$ se vérifie comme dans la preuve de la proposition précédente. \square

Noter que l'on peut définir de la même manière l'espace $C_b^1(\Omega)$ des fonctions bornées et de gradient (voir chapitre V) borné sur un ouvert Ω de \mathbb{R}^d , qui est complet pour la norme

$$\|f\|_{C^1} = \sup_{x \in \Omega} |f(x)| + \sup_{x \in \Omega} |\nabla f(x)|.$$

Exercice II.2.1. (Distance à un convexe fermé de $C([a, b])$ non atteinte)

On se place sur $E = C([0, 1])$, et l'on considère

$$F = \left\{ f \in E, \ f(0) = 1, \ \int_0^1 f(s) ds = 0 \right\}.$$

- a) Montrer que F est un convexe fermé de E .
- b) Montrer que la distance de $g = 1$ à F n'est pas atteinte.

CORRECTION.

a) *L'ensemble F est un sous-espace affine, donc convexe. Il est par ailleurs défini comme intersection des noyaux de deux formes linéaires continues, il est donc fermé.*

b) *Pout tout f dans F , il existe $x \in [0, 1]$ tel que $f(x) < 0$ (sinon l'intégrale serait strictement positive), on a donc $\|g - f\|_\infty > 1$. On considère maintenant, pour $n \in \mathbb{N}$, la fonction f_n continue affine par morceaux qui prend la valeur 1 en 0, de pente n sur $[0, a_n]$, et constante sur $[a_n, 1]$, où a_n est choisi de façon à ce que l'intégrale soit nulle, de telle sorte. La distance de f_n à g tend vers 1, on a donc $d(g, F) = 1$, et cette distance n'est pas atteinte d'après ce qui précède.*

Exercice II.2.2. Montrer que la boule unité fermée de $C([0, 1])$ n'est pas compacte.

CORRECTION.

On considère une fonction F de \mathbb{R} dans \mathbb{R} , continue, nulle sur $[0, 1/2] \cup [1, +\infty[$, et strictement positive sur $]1/2, 1[$. Pour $n \in \mathbb{N}$ on définit f_n comme la restriction à $[0, 1]$ de $x \mapsto F(2^n x)$. Pour tous $p \neq q$, la quantité $\|f_q - f_p\|_\infty$ est égale au max de F , qui est non nul (les supports sont disjoints), il est donc impossible d'extraire de (f_n) une sous-suite convergente.

II.3 De Riemann à Lebesgue

II.3.1 L'intégrale de Riemann et ses limites

Définition II.3.1. (Sommes de Riemann, Intégrale de Riemann)

On se place sur l'intervalle $[a, b] \subset \mathbb{R}$. On note Λ l'ensemble des *subdivisions marquées* de l'intervalle I , c'est à dire l'ensemble des couples (σ, t) , avec

$$\sigma = (x_0, \dots, x_n), \ a = x_0 < x_1 < \dots < x_n = b, \ t = (t_1, \dots, t_n), \ t_j \in [x_{j-1}, x_j] \quad \forall j = 1, \dots, n,$$

où $n \geq 1$ est un entier. Pour toute fonction f de $[a, b]$ dans \mathbb{R} , pour tout $(\sigma, t) \in \Lambda$, on note $S(f, \sigma, t)$ la somme de Riemann

$$S(f, \sigma, t) = \sum_{j=1}^n (x_j - x_{j-1}) f(t_j). \quad (\text{II.3.1})$$

On appelle *pas* d'une subdivision le max des $x_j - x_{j-1}$, que l'on note h_σ . On note Λ_h l'ensemble des subdivisions marquées de pas inférieur à $h > 0$. On dit que f est Riemann intégrable si la limite de $S_{(\sigma, t)}$ lorsque h_σ tend vers 0 existe, c'est-à-dire

$$\exists I \in \mathbb{R}, \forall \varepsilon > 0, \exists \eta, \forall (\sigma, t) \in \Lambda_\eta, |S_{(\sigma, t)} - I| < \varepsilon.$$

On appelle alors I l'intégrale (de Riemann) de f , et l'on note

$$I = \int_a^b f(x) dx.$$

Proposition II.3.2. Une fonction Riemann intégrable sur $[a, b]$ est nécessairement bornée sur cet intervalle.

Démonstration. Soit f une fonction non majorée sur $[a, b]$. Nous allons montrer que l'on peut trouver une suite de subdivisions marquées de pas tendant vers 0 telle que $S_{(\sigma_\eta, t_\eta)}$ tends vers $+\infty$, ce qui exclut la Riemann intégrabilité. Il existe une suite (x_n) de $[a, b]$ telle que $f(x_n)$ tend vers $+\infty$. L'intervalle $[a, b]$ étant compact, on peut en extraire une suite qui converge vers un élément $x \in [a, b]$. On note toujours (x_n) cette suite extraite. Pour tout $\eta > 0$, on complète l'intervalle $[x - \eta/2, x + \eta/2] \cap [a, b]$ pour former une subdivision marquée (σ_η, t_η) de pas $\leq \eta$, dont l'intervalle centré en x est le k -ème. On choisit arbitrairement des marques t_j pour $j \neq k$ (par exemple la borne inférieure du sous-intervalle considéré). Comme f n'est pas bornée sur $[x - \eta/2, x + \eta/2] \cap [a, b]$, il existe y dans cet intervalle tel que

$$S_{(\sigma_\eta, t_\eta)} = \eta f(y) + \sum_{j \neq k}^n (x_j - x_{j-1}) f(t_j) \geq \frac{1}{\eta}$$

On construit ainsi une suite (σ_η, t_η) de subdivisions marquées dont le pas tend vers 0, telle que les sommes de Riemann associées tendent vers $+\infty$, cette fonction n'est donc pas R-intégrable. On démontre de la même manière qu'une fonction non minorée n'est pas Riemann intégrable. \square

Remarque II.3.3. (Sommes de Riemann / de Darboux)

Il existe une construction alternative de l'intégrale de Riemann, basée sur les sommes dites de *Darboux* (équations (II.6.1)(II.6.2), page 67). L'équivalence des deux constructions fait l'objet de l'exercice II.6.1, page 67.

Proposition II.3.4. (Structure d'e.v.n. des fonctions Riemann-intégrables)

On note E l'ensemble des fonctions Riemann-intégrables sur $[a, b] \subset \mathbb{R}$. On munit E de la relation d'équivalence

$$f \mathcal{R} g \iff \int_a^b |f - g| = 0,$$

et l'on note $\bar{E} = E/\mathcal{R}$ l'espace quotient. La quantité $\int_a^b |f|$ ne dépend pas du représentant choisi pour \bar{f} , on la note $\|\bar{f}\|$, elle confère à \bar{E} une structure d'espace vectoriel normé.

Exercice II.3.1. Montrer que la fonction caractéristique de \mathbb{Q} n'est pas Riemann-intégrable sur $[0, 1]$.

CORRECTION.

L'ensemble \mathbb{Q} des rationnels ainsi que son complémentaire sont dense dans $[0, 1]$. Pour toute subdivision de pas $h > 0$, on peut donc choisir des marques telles que la somme de Riemann associée vaille 1, ou alternativement 0, ce qui exclut la convergence de ces sommes de Riemann lorsque le pas tend vers 0.

Proposition II.3.5. L'espace E des (classes de) fonctions Riemann-intégrables sur $[a, b] \subset \mathbb{R}$ n'est pas complet.

Démonstration. On se place sur l'intervalle $[0, 1]$, et l'on considère la suite de fonctions définies par

$$f_n(x) = \frac{1}{\sqrt{x}} \text{ sur } [1/n, 1], \quad f_n(x) = 0 \text{ sur } [0, 1/n[.$$

On a, pour tous $p < q$

$$\|f_q - f_p\| = \int_{1/q}^{1/p} \frac{1}{\sqrt{t}} dt = \frac{2}{\sqrt{p}} - \frac{2}{\sqrt{q}},$$

qui tend vers 0 quand p et q tendent vers $+\infty$. Il s'agit donc d'une suite de Cauchy. Cette suite converge uniformément vers $1/\sqrt{x}$ sur tout intervalle du type $[\eta, 1]$, avec $\eta > 0$. Si elle converge dans \overline{E} , elle converge donc nécessairement vers la classe de cette fonction, qui n'est pas dans \overline{E} d'après la proposition II.3.2. \square

Le contre-exemple de la démonstration ci-dessus repose sur le caractère borné des fonctions Riemann-intégrables. On pourrait espérer contourner ce problème en intégrant la notion d'intégrale généralisée absolument convergente. Mais cette notion n'est pas suffisante : on peut construire de tels contre-exemples avec des fonctions qui explosent en des points intérieurs à l'intervalle, et même en une infinité de points intérieurs à l'intervalle. Par ailleurs le caractère borné des fonctions R-intégrables n'est pas le seul obstacle à la complétude, comme l'exercice II.3.2 ci-dessous le montre.

Exercice II.3.2. (••)

On considère une numérotation (r_n) des rationnels de $[0, 1]$, et l'on considère la suite de fonctions (f_n) définie par

$$x \in [0, 1] \longmapsto f_n(x) = \sum_{k=1}^n \mathbf{1}_{[r_k, r_k + \varepsilon/2^n]}(x),$$

avec $\varepsilon > 0$.

a) Montrer que (f_n) (plus précisément la suite des classes de fonctions associées aux f_n) est une suite de Cauchy dans l'e.v.n. E des fonctions R-intégrables sur $[0, 1]$ (défini par la proposition II.3.4 ci-dessus).

b) Montrer que, pour tout n , $0 \leq \int f_n \leq \varepsilon$.

On suppose que f_n converge vers une fonction f de E .

c) Montrer que, pour tout $h > 0$, il existe une subdivision marquée (σ, t) de pas plus petit que h telle que

$$S(f, \sigma, t) \geq 1.$$

d) Conclure.

CORRECTION.

a) Pour tous $p < q$, on a

$$f_q - f_p = \sum_{k=p+1}^q \mathbf{1}_{[r_k, r_k + \varepsilon/2^n]},$$

donc la norme (intégrale de la valeur absolue) est inférieure à $\varepsilon/2^p$. Il s'agit donc bien d'une suite de Cauchy.

b) On a de la même manière

$$0 \leq \int f_n \leq \sum_{k=1}^n \int_0^1 \mathbf{1}_{[r_k, r_k + \varepsilon/2^n]}(x) dx \leq \varepsilon.$$

c) On prend h de la forme $1/N$, et l'on considère la subdivision uniforme de ce pas ($x_j = jh$ pour $j = 0, \dots, N$). Par densité des r_k , pour n assez grand, tout intervalle $[x_j - 1, x_j]$ contient au moins l'un des r_k . On prend ce r_k pour la marque t_j . Pour cette subdivision marquée particulière, on a $f(t_j) = 1$ pour tout j , d'où

$$S(f_n, \sigma, t) \geq 1.$$

Comme $f \geq f_n$, on a $S(f, \sigma, t) \geq 1$, pour une suite de subdivision dont le pas tend vers 0 d'où $S(f, \sigma, t) \geq 1$.
d) Comme f_n converge vers f pour la norme de l'intégrale, on a d'après la question précédente $\int f \geq 1$. Mais d'après la question b), cette intégrale est inférieure à ε , d'où une contradiction dès que $\varepsilon > 1$. On a donc non convergence dans E de la suite de Cauchy (f_n) , d'où la non complétude de l'espace.

Comme nous l'avons vu ci-dessus (proposition II.3.5), les espaces de fonctions (ou de classes de fonctions) canoniquement associés à l'intégrale de Riemann ne sont pas complets. L'élaboration de cadres fonctionnels adaptés à l'étude des équations aux dérivées partielles (existence / unicité de solution, étude de convergence des méthodes numériques) passe par une autre construction de la notion d'intégrale, appelée intégrale de Lebesgue, qui fait l'objet des sections suivantes. Comme nous le verrons, cette construction présente un certain nombre d'avantages par rapport à celle de Riemann, parmi lesquels :

- 1) La nouvelle notion généralise strictement la précédente au sens où toutes les fonctions Riemann-intégrables seront Lebesgue-intégrables, mais il y a "beaucoup plus" de fonctions intégrables dans ce sens nouveau. En particulier les fonctions intégrables ne seront pas nécessairement bornées, elle peuvent aussi être aussi très pathologique en termes de continuité : une fonction peut n'être continue en aucun point, et pourtant Lebesgue-intégrable⁵.
- 2) Comme nous l'avons indiqué comme motivation principale de cette nouvelle construction, cette intégrale permettra de définir une norme sur des espaces de fonctions intégrables (ou dont une certaine puissance est intégrable), normes qui feront de ces espaces des espaces *complets*.
- 3) Cette nouvelle théorie de l'intégration apporte des théorèmes puissants très utiles pour les démonstrations, en particuliers deux théorèmes permettant, sous certaines conditions, de déduire de la convergence simple d'une suite de fonctions intégrables l'intégrabilité de la fonction limite et la convergence des intégrales vers l'intégrale de la limite⁶.
- 4) La notion de *tribu* évoquée ci-dessus, au fondement de la construction de la notion de mesure et de l'intégrale de Lebesgue, est aussi à la base de la théorie des probabilités, permettant de formaliser de façon rigoureuse la notion d'espace probabilisé et d'*événement*.
- 5) La construction de Riemann peut s'étendre aux espaces \mathbb{R}^d , mais cette construction est un peu laborieuse, en particulier si l'on souhaite considérer des fonctions définies sur des domaines peu réguliers. La construction de Lebesgue est beaucoup plus souple et générale, et, même si sa construction sur \mathbb{R} passe par les longueurs des segments⁷, elle peut se définir sur des espaces sans structure affine ni métrique.

Précisons néanmoins que la construction Riemannienne reste intéressante en elle-même. En particulier le calcul effectif de valeurs approchées d'intégrales suit en général une démarche de type sommes de Riemann. Par ailleurs la convergence des sommes de Riemann (pour des subdivisions uniformes) vers l'intégrale permet d'estimer certaines limites de sommes de façon très performante.

Précisons que cette nouvelle approche sera d'une certaine manière plus restrictive selon un certain aspect : à aucun moment un principe de compensation⁸ ne sera utilisé. De fait, la construction de Lebesgue ne repose que sur une structure assez rudimentaire sur l'espace de départ des fonctions que l'on cherche à intégrer, plus précisément la notion de *tribu*, ensemble de parties dont on sait définir la *mesure*, indépendamment de toute structure d'ordre. On aura donc équivalence entre l'intégrabilité d'une fonction et l'intégrabilité de sa valeur absolue. On pourra se reporter à l'exercice II.6.2, page 67, pour se convaincre de la fragilité des constructions d'intégrales (il s'agit de sommes en l'occurrence) basées sur un principe de compensation.

II.3.2 Théorie de la mesure et intégrale de Lebesgue : un aperçu

La section qui suit donne un aperçu synthétique de la théorie de la mesure, ainsi que de l'intégration de Lebesgue, sur lesquelles se fonde la construction des espaces fonctionnels proposée dans les sections

5. À titre d'exemple la fonction de l'exercice II.3.1, qui n'est pas Riemann intégrable, sera Lebesgue intégrable.

6. Il s'agit des théorèmes de convergence monotone (Th. B.7.23, page 228), et de convergence dominée (Th. B.7.25, page 229). Noter qu'il existe des versions de ces théorèmes dans le cadre de l'intégrale de Riemann (voir par exemple <http://alain.troesch.free.fr/2018/Fichiers/sujet12.pdf>), mais il est nécessaire de supposer que la fonction limite est elle-même intégrable, ce qui n'est pas le cas dans le cadre Lebesgue.

7. En particulier dans la construction de la mesure extérieure de Lebesgue, qui fait l'objet de la proposition B.4.6, page 207.

8. Qui permet par exemple de définir l'intégrale sur \mathbb{R}_+ de la fonction $x \mapsto \sin(x)/x$.

suivantes. Cette construction⁹ est basée sur la notion de mesure. Le point de départ est le suivant¹⁰ : si l'on se donne une partie A de \mathbb{R}^d dont on connaît le volume, il est naturel de définir l'intégrale d'une fonction constante sur A comme le produit de cette constante par le volume¹¹. Pour toute fonction qui s'écrit comme somme finie de fonctions constantes sur des parties disjointes deux à deux dont on connaît le volume (on parlera plus loin de fonction *étagée*), on a aussi une manière canonique de définir l'intégrale comme somme des différentes contributions. On peut alors définir l'intégrale d'une fonction positive quelconque comme le supremum des intégrales des fonctions étagées positives inférieure ou égale à f . L'extension à des fonctions de signe quelconque se fait alors aisément en considérant les parties positives et négatives. Le point délicat ici est le sens que l'on donne à la notion de "volume" pour des parties quelconques. Comme précisé ci-après, on introduira la notion de *mesure*, qui généralise la notion de volume. Sur \mathbb{R}^d , on demandera à cette mesure de vérifier des propriétés conformes aux attentes, c'est à dire que le vide a un volume nul, est les volumes (ou mesures) de parties disjointes s'additionnent¹² pour donner le volume de la réunion. Pour que la définition de l'intégrale conduise à des propriétés exploitables, on aura besoin que cette propriété de sommation s'étendent aux unions infinies dénombrables. On souhaite enfin que cette mesure donne les valeurs attendues pour des ensembles simples, c'est à dire que la mesure d'un segment est sa longueur (pour $d = 1$), la mesure d'un rectangle est le produit des côtés pour $d = 2$, etc ... Cette construction est plus délicate qu'il n'y paraît, du fait qu'il est *impossible* de construire une mesure sur l'ensemble des parties de \mathbb{R}^d , qui vérifie les propriétés précédentes. On est contraint de se restreindre à des sous-familles de l'ensemble des parties. Pour que l'intégrale résultant de cette démarche ait de bonnes propriétés, ces sous-familles doivent vérifier un certain nombre de conditions (en particulier de stabilité vis-à-vis de l'union et du passage au complémentaire), ce qui conduit à la notion de *tribu*. Cette notion (en particulier au travers de la tribu des boréliens, et de la tribu de Lebesgue sur \mathbb{R}) joue un rôle central dans le théorie de l'intégration de Lebesgue. Elle constitue par ailleurs le fondement de la théorie moderne des probabilités. Nous en précisons donc ci-dessous les principales propriétés, dans un cadre très général, en renvoyant au chapitre B en annexe (page 189) pour un exposé plus détaillé et formalisé de ces questions.

II.3.3 Mesure et intégration

Tribus

Soit X un ensemble. Une tribu \mathcal{A} sur X est un ensemble de parties ($\mathcal{A} \subset \mathcal{P}(X)$) qui vérifie les trois propriétés suivantes (voir définition B.2.13) :

- (i) \mathcal{A} contient l'ensemble vide,
- (ii) si A est dans \mathcal{A} , son complémentaire l'est aussi,
- (iii) si (A_n) est une collection dénombrable d'éléments de \mathcal{A} , alors leur union est dans \mathcal{A} .

On appelle (X, \mathcal{A}) un *espace mesurable*.

Noter qu'une tribu est également stable par intersection dénombrable (proposition B.2.3), du fait que

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

L'ensemble des parties $\mathcal{P}(X)$ est une tribu, appelée tribu discrète, c'est la tribu la plus fine sur X (elle contient toutes les autres). La tribu la moins fine est $\{\emptyset, X\}$, appelée tribu *grossière*.

L'intersection de tribus étant une tribu, on peut définir la tribu *engendrée* par un ensemble de parties \mathcal{C} , que l'on notera $\sigma(\mathcal{C})$, comme la plus petite tribu contenant \mathcal{C} , c'est-à-dire l'intersection des tribus contenant \mathcal{C} .

Exercice II.3.3. Donner un exemple de tribu sur \mathbb{Z} (ainsi que sur \mathbb{R}), autre que la tribu grossière, qui ne contient qu'un nombre fini de parties.

9. La démarche présentée ci-après a fait l'objet des travaux de thèse de Henri Lebesgue, au tout début du vingtième siècle.
10. Ces considérations informelles sur la démarche générale sont détaillées dans la section B.1, page 189.

11. On peut penser à la valeur de la fonction comme une *densité*, en kg m^{-d} . Le résultat de cette "intégration" est simplement la masse de la matière contenue dans cette partie.

12. La mesure est la version mathématique d'une variable *extensive*.

CORRECTION.

On peut par exemple considérer Z_0 l'ensemble des nombres pairs, et \mathcal{A} la tribu engendrée par Z_0 qui contient les 4 parties

$$\emptyset, Z_0, Z_1, \mathbb{Z},$$

où Z_1 est l'ensemble des nombres impairs. Sur \mathbb{R} on pourra considérer par exemple la tribu engendrée par \mathbb{N} , qui contient

$$\emptyset, \mathbb{N}, \mathbb{R} \setminus \mathbb{N}, \mathbb{R}.$$

Exercice II.3.4. Les familles ci-dessous engendent-elles la tribu discrète sur \mathbb{N} ?

- (i) La famille des singletons.
- (ii) La famille des parties du type $\{n, n+1\}$, pour n décrivant \mathbb{N} .
- (iii) La famille des parties du type $\{2n, 2n+1\}$, pour n décrivant \mathbb{N} .

CORRECTION.

(i) oui : toute partie de \mathbb{N} est réunion dénombrable de singleton (ses propres éléments).

(ii) oui : on a $\{n, n+1\} \cap \{n+1, n+2\} = \{n+1\}$, la tribu engendrée contient donc tous les singletons pour les entiers ≥ 1 , et $\{0\} = \{0, 1\} \cap \{1\}^c$ y est aussi,

(iii) non : deux entiers consécutifs pair-impair ne sont pas distingués par la famille d'origine, ils ne sont pas distingués par la tribu engendrées. La tribu engendrée est simplement la famille des réunions quelconques de paires $\{2n, 2n+1\}$.

Exercice II.3.5. Soit X un ensemble et A, B et C des parties de \mathcal{A} , non vides, et disjointes deux à deux. Quels sont les cardinaux possibles pour la tribu engendrée par $\{A, B, C\}$?

CORRECTION.

Si l'union des parties est X (elles réalisent alors une partition), le cardinal est celui de l'ensemble des parties de l'ensemble à 3 points, donc $2^3 = 8$. Si l'union ne recouvre pas X , alors la tribu est celle engendrée par la partition $\{A, B, C, (A \cup B \cup C)^c\}$, qui est donc de cardinal $2^4 = 16$.

Sur un espace topologique X , on définit la tribu des boréliens $\mathcal{B}(X)$ comme la tribu engendrée par les ouverts de X .

Sur \mathbb{R} cette tribu est engendrée par les intervalles $] -\infty, b]$, pour b décrivant \mathbb{R} (prop. B.2.7, page 195).

Sur $\overline{\mathbb{R}}$ cette tribu est engendrée par les intervalles $[-\infty, b]$, pour b décrivant \mathbb{R} . (prop. B.2.8, page 195).

Noter que, si la tribu des boréliens sur \mathbb{R} ou $\overline{\mathbb{R}}$ est très facile à définir, il n'est pas aisé de se représenter ce qu'elle contient. Nous verrons plus loin qu'elle est strictement incluse dans l'ensemble des parties, mais contient néanmoins des ensembles beaucoup plus exotiques que les ouverts qui l'engendent.

Une application f entre deux espaces mesurables (X, \mathcal{A}) et (X, \mathcal{A}') est dite *mesurable* si l'image réciproque de tout élément de \mathcal{A}' est dans \mathcal{A} (définition B.2.11, page 197). On peut montrer (proposition B.2.12, page 197) que, si la tribu \mathcal{A}' en engendrée par \mathcal{C}' , alors l'application est mesurable si et seulement si $f^{-1}(A') \in \mathcal{A}$ pour tout A' dans \mathcal{C}' .

Dans le cas d'une application à image dans \mathbb{R} muni de la tribu des boréliens, il suffit donc de vérifier, d'après ce qui précède, que les ensembles

$$\{x \in X, f(x) \leq b\}$$

appartiennent bien à la tribu \mathcal{A} de l'espace de départ, pour tout b réel.

Mesures

Sur un espace mesurable (X, \mathcal{A}) , une mesure¹³ est une application μ de \mathcal{A} dans $[0, +\infty]$, qui vérifie les deux propriétés suivantes :

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

(ii) pour toute collection (A_n) dénombrable d'éléments de \mathcal{A} disjoints deux à deux, la mesure de l'union de A_n est égale à la somme des mesures des A_n (définition B.3.1, page 200).

On dit que la mesure est σ -finie si X est réunion dénombrable d'éléments de \mathcal{A} de mesure finie.

On parle de mesure de probabilité si $\mu(X) = 1$.

On appelle le triplet (X, \mathcal{A}, μ) un espace mesuré (espace probabilisé s'il s'agit d'une mesure de probabilité).

On dit qu'une partie est négligeable si elle est incluse dans une partie de mesure nulle. Une union dénombrable de parties négligeables reste négligeable. On dit qu'une propriété est vérifiée presque partout (p.p. en abrégé) si elle est vérifiée en dehors d'un ensemble négligeable.

Toute mesure μ est monotone, c'est-à-dire que si $A \subset B$ est inclus dans $B \in \mathcal{A}$, alors la mesure de A est plus petite que celle de B . Si la mesure de A est finie, on a de plus $\mu(B \setminus A) = \mu(B) - \mu(A)$ (voir proposition B.3.3, page 201).

Exercice II.3.6. (Mesure grossière)

Soit \mathcal{A} une tribu sur X . Montrer que l'application qui à tout élément non vide de \mathcal{A} affecte la valeur $+\infty$, et 0 à l'ensemble vide, est une mesure.

Cela reste-t-il vrai si l'on remplace $+\infty$ par 1 ?

CORRECTION.

Pour toute collection (A_n) d'éléments de \mathcal{A} disjoints deux à deux, si au moins l'un d'eux est non vide alors l'union est non vide, et l'on a bien $+\infty \leq +\infty$, et s'ils sont tous vides l'union est vide, et l'on a bien $0 \leq 0$. On ne peut pas remplacer $+\infty$ par 1 en général, sauf dans le cas de la tribu grossière (qui s'identifie d'ailleurs à la tribu discrète si X est un singleton). Mais dès que la tribu contient un ensemble A non vide strictement inclus dans X , on a $\mu(X) = 1 < 2 = \mu(A) + \mu(X \setminus A)$, qui invalide (ii).

Exercice II.3.7. Décrire l'ensemble des mesures définies sur la tribu discrète d'un ensemble X fini.

CORRECTION.

Un telle mesure est entièrement déterminée par la masses des "atomes" (les singletons), qui est pour chacun un nombre de $[0, +\infty]$. On peut donc associer à toute mesure un et un seul élément de $[0, +\infty]^X$.

Exercice II.3.8. Donner un exemple de mesure finie sur \mathbb{Z} qui affecte une masse strictement positive à chaque singleton

CORRECTION.

Il suffit par exemple de considérer une suite α_n , de réels positifs tels que la série converge, et poser $\mu(\{j\}) = \alpha_{|j|}$ pour tout $j \in \mathbb{Z}$.

N.B. Noter qu'une telle mesure finie qui charge tous les points n'existe pas quand l'ensemble n'est pas dénombrable (voir exercice II.6.3, page 68).

Construction abstraite de mesures, application à la mesure de Lebesgue (• • •)

Une mesure extérieure μ^* (définition B.4.2, page 204) est une application de l'ensemble des parties d'un ensemble à valeurs dans \mathbb{R}_+ , qui

- (i) affecte 0 à l'ensemble vide ;
- (ii) possède la propriété de monotonie suivante :

$$\forall A, B \in \mathcal{P}(X), A \subset B, \mu^*(A) \leq \mu^*(B);$$

- (iii) vérifie la propriété de sous-additivité suivante : pour toute collection (A_n) au plus dénombrable de parties de X , alors la mesure extérieure de l'union est inférieure ou égale à la somme des mesures extérieures.

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

13. On peut voir cette définition comme une formalisation mathématique de la notion de variable extensive utilisée par les physiciens.

On peut associer à μ^* une notion de mesurabilité, qui exprime un bon comportement vis à vis de l'additivité : une partie A est dite *mesurable* pour μ^* si $\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c)$ pour toute partie B . On peut montrer en toute généralité que l'ensemble \mathcal{A} des parties μ^* -mesurables au sens ci-dessus constitue une *tribu*, et que la restriction de μ^* à \mathcal{A} est une mesure (théorème B.4.4, page 205).

Cette propriété abstraite permet de construire des mesures effectives sur des ensembles très généraux, selon la démarche décrite ici dans le cas de la construction de la mesure de Lebesgue sur \mathbb{R} (qui fait l'objet de la proposition B.4.6, page 207). On considère une famille de parties dont on connaît la valeur de la mesure que l'on souhaite leur affecter, en l'occurrence les intervalles ouverts de \mathbb{R} , auxquels on souhaite affecter leurs longueurs. On suppose que toute partie peut être recouverte par une réunion dénombrable de telles parties (ce qui est le cas ici). On définit alors $\lambda^*(A)$ l'infimum des sommes des longueurs sur l'ensemble des collections d'intervalles qui recouvrent A (équation B.4.4). On peut alors montrer que λ^* est une mesure extérieure, appelée mesure extérieure de Lebesgue. On appelle alors tribu de Lebesgue l'ensemble \mathcal{A} des parties mesurables pour λ^* . La propriété abstraite évoquée précédemment assure que la restriction de λ^* à \mathcal{A} est une mesure, appelée *mesure de Lebesgue*, qui affecte aux intervalles (ouverts ou fermés) leurs longueurs. On peut montrer que \mathcal{A} contient strictement la tribu des borélien, la mesure de Lebesgue λ se trouve ainsi définie sur \mathcal{A} et \mathcal{B} , on pourra être amené à considérer l'une ou l'autre de ces tribus selon le cas.

On peut ainsi construire sur \mathbb{R} une mesure λ , appelée *mesure de Lebesgue*, définie sur une tribu \mathcal{A} appelée tribu de Lebesgue, qui est telle que la mesure de tout intervalle est égale à sa longueur. La tribu sur laquelle elle est définie contient en particulier la tribu borélienne $\mathcal{B}(\mathbb{R})$ (proposition B.4.8, page 209). La mesure d'un singleton étant nulle, toute partie dénombrable (comme \mathbb{Q} ou \mathbb{D}) est de mesure nulle.

Parties non mesurables (•••)

Il est a priori impossible de construire une telle mesure sur l'ensemble des parties de \mathbb{R} , et en même temps il est impossible de décrire explicitement une partie qui ne serait pas dans la tribu de Lebesgue. La construction (dans un sens assez abstrait) de parties non mesurables nécessite l'axiome du choix (voir proposition B.5.1, page 210). On pourra consulter Villani¹⁴, section IV-5.3, page 171, pour des commentaires généraux sur ce problème d'existence de parties non mesurables, et l'attitude qu'il convient d'adopter devant le fait que la démonstration de l'existence de telles parties repose sur l'axiome du choix.

Exercice II.3.9. Donner des exemples de boréliens A de \mathbb{R} tels que

- (i) $\lambda(\partial A) = \lambda(A)$,
- (ii) $\lambda(\partial A) < \lambda(A)$,
- (iii) $\lambda(\partial A) > \lambda(A)$.

CORRECTION.

Pour (i) on peut prendre $A = \mathbb{R} \setminus \mathbb{Q}$, qui vérifie $\partial A = A$. Ou, si l'on veut des mesures finies, l'intersection de cet ensemble avec $]0, 1[$.

Pour (ii) on peut prendre $A =]0, 1[, dont la frontière est de mesure nulle.$

Pour (iii) on peut prendre $A =]0, 1[\cap \mathbb{Q}$, de mesure nulle, dont la frontière est $[0, 1]$.

Nous décrivons ci-dessous la démarche permet de construire une notion d'intégrale pour des fonctions définies d'un espace mesuré à valeur dans \mathbb{R} . Cette démarche s'appliquant à des fonctions mesurables, nous munissons l'espace d'arrivée \mathbb{R} de la tribu des boréliens¹⁵.

Intégrales de fonctions étagées

On appelle fonction étagée sur (X, \mathcal{A}) une fonction à valeurs dans \mathbb{R} mesurable (\mathbb{R} est ici muni de la tribu borélienne), qui prend un nombre fini de valeurs. On écrira une telle fonction

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

14. C. Villani, Intégration et Analyse de Fourier, https://www.cedricvillani.org/sites/dev/files/old_images/2013/03/IAF.pdf

15. Il est possible de munir \mathbb{R} de la tribu de Lebesgue, mais la notion de mesurabilité d'une application est alors plus exigeante. Noter que si l'espace de départ est aussi \mathbb{R} , on choisira de munir \mathbb{R} en tant qu'espace de départ de la tribu de Lebesgue. Ce double choix, qui n'aura pas d'incidence pratique, permet de relaxer au maximum la contrainte de mesurabilité, et donc d'avoir le maximum de fonctions mesurables.

où les A_i sont dans \mathcal{A} , et deux à deux disjoints (les α_i ne sont pas nécessairement distincts 2 à 2). On notera \mathcal{E}^+ l'ensemble des fonctions étagées positives. Si l'espace est muni d'une mesure μ , on définit l'intégrale de f comme

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

Intégrales de fonctions mesurables positives, intégrabilité

On définit ici l'intégrale de fonctions mesurables positives sur (X, \mathcal{A}, μ) à valeurs dans $[0, +\infty]$. Conformément à ce qui précède, une telle fonction est mesurable si l'image réciproque de tout $[-\infty, b]$ est dans \mathcal{A} . Si l'espace de départ est \mathbb{R} , on le considérera muni de la tribu de Lebesgue, qui contient la tribu borélienne.

On définit l'intégrale d'une fonction mesurable¹⁶ à valeurs positives comme

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

Intégrabilité

Toute fonction mesurable sur (X, \mathcal{A}, μ) à valeurs dans $[-\infty, +\infty]$, peut être écrite comme la différence de ses parties positive et négative :

$$f = f^+ - f^-, \quad f^+ = \max(f, 0) = \frac{1}{2} (f + |f|).$$

On dit que f est *intégrable* si $\int f^+$ et $\int f^-$ sont finies, et l'on note $\int f = \int f^+ + \int f^-$. Si au moins l'une des deux est infinie, on dira que la fonction est *non intégrable*. Si une seule des intégrales est finie, on dira simplement que l'*intégrale existe*, avec la valeur $\pm\infty$, selon que $\int f^+$ ou $\int f^-$ est finie¹⁷.

Une fonction mesurable f de (X, \mathcal{A}, μ) dans $\overline{\mathbb{R}}$ est intégrable si et seulement si $|f|$ l'est, et l'on a (voir proposition B.7.19, page 227).

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

Exercice II.3.10. La fonction $x \mapsto \sin(x)/x$ est-elle intégrable sur \mathbb{R} ?

CORRECTION.

L'intégrale de sa partie positive existe bien sûr, comme celle de toute fonction, mais elle est infinie. Cela suffit pour invalider l'intégrabilité. La partie négative est elle-même d'intégrale infinie, de telle sorte que l'on a une indétermination, l'intégrale n'existe donc pas : dans le cadre d'intégration de Lebesgue, il est interdit de parler de $\int \sin(x)/x$. On pourra se reporter à l'exercice B.7.3, page 227, pour plus de détail sur ces considérations.

Remarque II.3.6. (Existence d'une intégrale / intégrabilité)

On prendra garde au fait suivant : l'intégrale d'une fonction positive, telle que définie ci-dessus, peut prendre la valeur $+\infty$. Mais une telle fonction n'est dite intégrable que si la valeur est *finie*. Comme précisé ci-dessus, pour une fonction signée, les intégrales de f^+ et f^- peuvent exister et être toutes deux infinies, on aura alors une indétermination¹⁸. On pourra se reporter à l'exercice B.7.3, page 227, qui précise ces notions sur un exemple classique.

16. Noter que l'expression qui suit permet en fait de définir l'intégrale de n'importe quelle fonction positive, *fût-elle non mesurable*, puisqu'elle est simplement définie comme borne supérieure d'un ensemble de réels non vide (il contient 0, puisque la fonction nulle, étagée, minore toute fonction positive). Nous n'utiliserons pourtant cette définition que pour des fonctions mesurables.

17. On peut se demander ce qui pousse à refuser le statut formel d'*intégrable* à une fonction dont l'intégrale est bien définie. Les motivations de cette intransigeance apparaîtront plus loin. Comme nous l'avons évoqué précédemment, l'un des objectifs est de construire un espace vectoriel normé sur la base de cette intégrale ; une norme devant prendre des valeurs finies, les fonctions d'intégrale $\pm\infty$ ne seront pas intégrées à ces espaces.

Intégrale selon la mesure de Lebesgue, notation

Lorsque l'espace de départ est \mathbb{R} (muni de la mesure de Lebesgue), et que l'intégrale existe, on écrira cette intégrale, conformément à l'usage courant,

$$\int_{\mathbb{R}} f(x) dx.$$

Théorèmes fondamentaux

Nous exprimons ci-dessous les trois théorèmes fondamentaux portant sur des suites de fonctions mesurables. Ils sont exprimés sous leur forme abstraite dans l'annexe B (sur un espace mesuré général (X, \mathcal{A}, μ)), mais nous les énonçons ci-dessous dans le cadre d'usage courant, où l'espace de départ est \mathbb{R} muni de la mesure de Lebesgue.

Théorème de convergence monotone (th. B.7.23, page 228) : soit (f_n) une suite de fonctions mesurables de \mathbb{R} dans $[0, +\infty]$, qui converge presque partout vers f , et telle que $f_n(x)$ est croissant pour presque tout x . Alors f est mesurable et $\int f_n(x) dx$ converge vers $\int f(x) dx$.

Lemme de Fatou (lemme B.7.24, page 229) : soit (f_n) une suite de fonctions mesurables de \mathbb{R} dans $[0, +\infty]$. Alors

$$\int \liminf_n f_n(x) dx \leq \liminf_n \int f_n(x) dx.$$

Théorème de convergence dominée (th. B.7.25, page 229) : soit (f_n) une suite de fonctions mesurables de \mathbb{R} dans $[-\infty, +\infty]$, qui converge presque partout vers f , avec

$$|f_n(x)| \leq g(x) \quad \text{p.p.},$$

où g est une fonction intégrable sur \mathbb{R} . Alors les f_n et f sont intégrables, et $\int f_n$ converge vers $\int f$.

Exercice II.3.11. Pour les suites de fonctions de \mathbb{R}_+ dans \mathbb{R} définies ci-dessous, préciser la fonction qui est limite simple de la suite, la limite des intégrales, l'intégrale de la limite, et préciser si ces 3 suites (f_n^1) , (f_n^2) , et (f_n^3) , rentrent dans le cadre du théorème de convergence monotone ou du théorème de convergence dominée :

$$f_n^0(x) = \mathbf{1}_{]n, n+1[}, \quad f_n^1(x) = \frac{1}{|x|^2} \mathbf{1}_{]1/n, 1[}, \quad f_n^2(x) = \frac{1}{|x|} \mathbf{1}_{]n, 2n[}, \quad f_n^3(x) = -\frac{1}{n} \mathbf{1}_{]0, n[}.$$

CORRECTION.

La suite f^0 converge simplement vers la fonction nulle, d'intégrale nulle, alors que l'intégrale des termes de la suite est constante égale à $1 \neq 0$. De fait la convergence n'est ni monotone, ni dominée.

La suite de fonctions (f_n^1) converge simplement et de façon croissante vers la fonction $\mathbf{1}_{]0, 1[}/x^2$. La suite des intégrales converge vers l'intégrale de la limite, qui est $+\infty$, ce qui est conforme au TCM.

La suite de fonctions (f_n^2) converge simplement vers la fonction nulle. L'intégrale de la limite est 0, alors que la suite des intégrales est stationnaire égale à $\ln 2$, donc converge vers cette valeur non nulle. De fait, la convergence n'est ni monotone ni dominée.

La suite de fonctions (f_n^3) converge simplement vers la fonction nulle. L'intégrale de la limite est 0, alors que la suite des intégrales est stationnaire égale à -1 , donc converge vers cette valeur non nulle. De fait, la convergence n'est ni monotone ni dominée.

Exercice II.3.12. Préciser (sans faire de calcul) les limites quand n tend vers $+\infty$ de

$$u_n = \int_0^{+\infty} \frac{\cos(e^{-nx})}{1+x^2} dx, \quad v_n = \int_0^{+\infty} \frac{e^{-x}}{2+\cos(x)^n} dx$$

18. Dans le contexte de l'intégrale de Lebesgue, on ne cherchera pas à lever cette indétermination en invoquant des principes de compensation, qui permettent par exemple de définir dans le contexte de l'intégrale de Riemann l'intégrale (généralisée) de $x \mapsto \sin(x)/x$. Ces compensations font intervenir de façon essentielle l'ordre dans lequel on effectue les intégrations (de 0 vers $+\infty$ pour l'exemple cité). Cette directionnalité de l'espace sous jacent n'est pas du tout présente dans la construction de l'intégrale de Lebesgue, qui ne repose que sur une mesure définie sur une tribu de l'espace de départ.

CORRECTION.

Pour u_n , la suite des fonctions intégrées converge simplement sur $[0, +\infty[$ vers $1/(1+x^2)$. cette convergence est dominée par la fonction limite, qui est intégrable. La limite de l'intégrale est donc d'après le TCD l'intégrale de la limite, i.e. $\pi/2$.

Pour v_n , la suite de fonctions converge simplement vers e^{-x} sur \mathbb{R} en dehors des $k\pi$, donc presque partout. Noter qu'on n'a pas convergence sur les $(2k+1)\pi$. Cette convergence est dominée par e^{-x} , on a donc convergence vers l'intégrale de la limite simple, qui est $1/2$.

Tribus-produit, mesures-produit, intégrales multiples

Si (X_1, \mathcal{A}_1) et (X_2, \mathcal{A}_2) sont deux espaces mesurables, on appelle *tribu-produit*, et l'on note $\mathcal{A}_1 \otimes \mathcal{A}_2$, la tribu sur $X_1 \times X_2$ engendrée par les rectangles, i.e. les $A_1 \times A_2 \in \mathcal{A}_1 \times \mathcal{A}_2$.

Mesure-produit (théorème B.7.34) : si $(X_1, \mathcal{A}_1, \mu_1)$ et $(X_2, \mathcal{A}_2, \mu_2)$ sont deux espaces mesurés, où μ_1 et μ_2 sont σ -finies, il existe une unique mesure $\mu_1 \otimes \mu_2$ sur $\mathcal{A}_1 \otimes \mathcal{A}_2$, telle que

$$(\mu_1 \otimes \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

Ce théorème permet donc de construire une mesure sur le produit de deux espaces mesurés, mesure compatible avec le volume des “rectangles”.

Le résultat principal est le théorème de Fubini-Lebesgue (énoncé sous forme abstraite page 233, théorème B.7.36), que nous présentons ci-dessous dans le cas du produit $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, muni de la mesure de Lebesgue produit. Il énonce essentiellement que, si une fonction $(x_1, x_2) \mapsto f(x_1, x_2)$ est intégrable sur le produit $\mathbb{R} \times \mathbb{R}$, alors on peut intégrer d'abord par rapport à x_1 , puis x_2 , ou d'abord par rapport à x_2 , puis x_1 .

Théorème de Fubini-Lebesgue : soit f une fonction de \mathbb{R}^2 à valeurs dans $[-\infty, +\infty]$. On suppose que f est intégrable sur \mathbb{R}^2 . Alors

- (i) Pour presque tout x_1 , la fonction $f(x_1, \cdot)$ est intégrable sur \mathbb{R} , et pour presque tout x_2 , la fonction $f(\cdot, x_2)$ est intégrable sur \mathbb{R} .
- (ii) Les fonctions¹⁹ $x_1 \mapsto \int_{\mathbb{R}} f(x_1, x_2) dx_2$ et $x_2 \mapsto \int_{\mathbb{R}} f(x_1, x_2) dx_1$ sont intégrables sur \mathbb{R} .
- (iii) On a

$$\int_{\mathbb{R}^2} f(x_1, x_2) dx_1 dx_2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1, x_2) dx_2 \right) dx_1 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1, x_2) dx_1 \right) dx_2.$$

Exercice II.3.13. Soit $P(x, y)$ un polynôme des deux variables x et y , de degré quelconque. L'identité

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} P(x, y) e^{-(x^2+y^2)} dx \right) dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} P(x, y) e^{-(x^2+y^2)} dy \right) dx$$

est-elle vérifiée en général ?

CORRECTION.

Oui, c'est une conséquence du théorème de Fubini-Lebesgue. La fonction $(x, y) \mapsto P(x, y) e^{-(x^2+y^2)}$ est intégrable sur \mathbb{R}^2 (on peut par exemple majorer sa valeur absolue par une fonction du type $g(r) = C(1+r^p)e^{-r^2}$, avec $r = (x^2+y^2)^{1/2}$, qui est telle que $\int r g(r) dr < +\infty$). On peut donc intégrer dans un sens ou dans l'autre.

Exercice II.3.14. On considère la fonction

$$(x, y) \in]0, +\infty[\times]0, 1[\mapsto 2e^{-2xy} - e^{-xy}.$$

18. Appelées *section*, voir définition B.7.29.

19. on prend la valeur 0 lorsque l'intégrande n'est pas intégrable, ce qui peut arriver pour un ensemble négligeable.

Montrer que les deux quantités

$$\int_0^1 \left(\int_0^{+\infty} f(x, y) dx \right) dy \quad \text{et} \quad \int_0^{+\infty} \left(\int_0^1 f(x, y) dy \right) dx$$

sont bien définies, mais ont des valeurs différentes.

Que peut-on en déduire sur la fonction f ?

CORRECTION.

On a, pour tout $y > 0$,

$$\int_{\mathbb{R}_+} (2e^{-2xy} - e^{-xy}) dx = \left[\frac{1}{y} (-e^{-2xy} + e^{-xy}) \right]_0^{+\infty} = 0$$

d'où la nullité de la première expression. En intégrant d'abord en y , on a, pour tout $x > 0$

$$\int_0^1 (2e^{-2xy} - e^{-xy}) dy = \frac{1}{x} [-e^{-2x} + e^x]$$

qui est strictement positif pour tout $x > 0$, l'intégrale en x sur \mathbb{R}_+ va donc donner une valeur strictement positive.

Ceci semble en contradiction avec le théorème de Fubini-Lebesgue, cela prouve simplement que ses hypothèses ne sont pas vérifiées, c'est à dire que la fonction n'est pas intégrable sur la bande considérée.

Changement de variable

Si l'on considère une application f d'un espace mesurable (X, \mathcal{A}) vers un ensemble X' , la proposition B.2.10, page 196 définit la *tribu image* de \mathcal{A} comme

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

On peut définir de façon analogue la *mesure image* d'une mesure par une application :

Proposition II.3.7. (Mesure image)

Soit f une application d'un espace mesuré (X, \mathcal{A}, μ) dans un ensemble X' . Alors

$$\mu' : \mathcal{A}' = f_{\sharp}\mathcal{A} \longmapsto \mathbb{R}_+$$

définie par

$$\mu'(A') = \mu(f^{-1}(A')) ,$$

est une mesure sur la tribu image $f_{\sharp}\mathcal{A}$, appelée *mesure image* de μ par f . Ce transport conserve la masse totale.

Démonstration. On a bien $\mu'(\emptyset) = 0$, et

$$\mu' \left(\bigcup_{n \in \mathbb{N}} A'_n \right) = \mu \left(f^{-1} \left(\bigcup_{n \in \mathbb{N}} A'_n \right) \right) = \mu \left(\bigcup_{n \in \mathbb{N}} f^{-1}(A'_n) \right) = \sum_{n=0}^{+\infty} \mu(f^{-1}(A'_n)) .$$

On a par ailleurs $\mu'(X') = \mu(f^{-1}(X')) = \mu(X)$. □

Dans le cadre des deux propositions précédentes, on peut écrire une formule abstraite de changement de variable.

Proposition II.3.8. Soient (X, \mathcal{A}) et (X', \mathcal{A}') deux espaces mesurables, μ une mesure sur (X, \mathcal{A}) , et $T : X \rightarrow X'$ une application mesurable de X vers X' .

(i) Pour toute fonction f de X' dans $[0, +\infty]$, mesurable,

$$\int_{X'} f(y) d(T_\sharp \mu)(y) = \int_X f \circ T(x) d\mu(x)$$

(ii) Pour toute fonction f de X' dans $\overline{\mathbb{R}}$, mesurable, la fonction f est $T_\sharp \mu$ -intégrable si et seulement si la fonction $f \circ T$ est μ -intégrable, et la formule ci-dessus est alors valable.

Démonstration. Si f est une fonction étagée (voir définition B.7.5, page 221), (i) est une conséquence directe de la définition de la mesure image. Dans le cas général d'une fonction mesurable, on peut approcher f par une suite croissante (f_n) de fonctions étagées (proposition B.7.6, page 221). Alors $(f_n \circ T)$ est une famille croissante de fonctions étagées convergeant vers $f \circ T$, et on peut passer à la limite grâce au théorème de convergence monotone (théorème B.7.23 page 228).

Pour (ii), on applique simplement ce qui précède à f^+ et f^- . \square

Lorsqu'il s'agit d'une application régulière entre parties de \mathbb{R}^d , on dispose d'une formule de changement de variable plus explicite, qui fait intervenir la *différentielle* de l'application (voir proposition V.1.20, page 111, dans le chapitre V dédié au calcul différentiel).

II.4 Les espaces L^p

Nous introduisons dans cette section les espaces L^p qui jouent un rôle central dans un très grand nombre d'applications. Ce sont des espaces naturels pour décrire des champs de quantités physiques intensives sur des domaines, typiquement l'espace physique \mathbb{R}^d ou un ouvert Ω de cet espace. La construction pouvant se faire en toute généralité, nous la proposons sur un espace mesuré (X, \mathcal{A}, μ) quelconque, mais on pourra instancier cette construction abstraite en remplaçant (X, \mathcal{A}, μ) par $(\Omega, \mathcal{B}, \lambda)$, où Ω est un ouvert non vide de \mathbb{R}^d (on parle de *domaine*), \mathcal{B} la tribu des boréliens, et λ la mesure de Lebesgue.

Remarques préliminaires : espaces fonctionnels et modélisation

La construction décrite dans les sections précédentes permet d'intégrer des variables intensives contre la mesure sous-jacente, pour obtenir une variable extensive afférente au domaine sur lequel on a intégré. Prenons le cas de la mesure de Lebesgue qui, conformément à la terminologie employée au début de ce chapitre, correspond à une mesure de type "volume". Si l'on intègre sur un domaine une fonction correspondant à la densité locale d'une certaine substance, on obtient la masse contenue dans le domaine considéré. La mesure volume peut ainsi être vue comme une capacité à accueillir de la masse. Pour un système fermé, la conservation de la masse se traduira par la conservation d'une certaine norme, qui correspond au cas $p = 1$, de sorte que l'espace L^1 introduit constituera un cadre naturel à cette description. Dans un contexte thermique, on peut considérer cette mesure uniforme comme prenant une certaine valeur fixe de type capacité calorifique. Lorsque l'on intègre sur un domaine un champ de température contre cette mesure, on obtient la quantité de chaleur contenue dans le domaine. La mesure de départ peut ainsi être vue comme une capacité locale à emmagasiner de l'énergie thermique. Noter que si le milieu est hétérogène, cette capacité peut varier d'un endroit à l'autre. Dans ce contexte, si l'on considère un problème d'évolution pour un système fermé (adiabatique), le cas $p = 1$ sera également adapté pour décrire ces phénomènes.

Considérons maintenant une mesure de type "masse", toujours selon la terminologie employée au début du chapitre. Si l'on considère que la mesure correspond à la distribution dans l'espace d'une matière pesante en mouvement, on peut intégrer la quantité vectorielle *Vitesse* contre cette mesure, pour obtenir la quantité de mouvement. Là encore le cas $p = 1$ constituera un cadre naturel, la conservation de la quantité de mouvement assurant la préservation d'une quantité définie ci-après comme la norme L^1 associée à la distribution de masse en mouvement. Si l'on intègre une autre variable intensive, scalaire celle-là, égale à la moitié du module de la vitesse au carré, le résultat de l'intégration sur un domaine correspond à l'énergie cinétique emmagasinée par le fluide occupant le domaine en question. Dans ce contexte, c'est l'espace L^2 qui s'impose comme cadre naturel. On notera que les considérations précédentes permettent de concevoir la mesure sous-jacente

(distribution de masse dans l'espace) comme une *capacité* à accueillir de la quantité de mouvement, ou une capacité à accueillir de l'énergie cinétique.

II.4.1 L'espace $L^\infty(X)$

Définition II.4.1. Soit (X, \mathcal{A}, μ) un espace mesuré. On note²⁰ $\tilde{L}^\infty(X)$ l'ensemble des fonctions *essentiellement bornées*, c'est à dire des fonctions f qui sont \mathcal{A} -mesurables, et telles qu'il existe $C \in \mathbb{R}_+$ vérifiant

$$|f(x)| \leq C \quad \text{pour presque tout } x.$$

Pour une telle fonction, on définit

$$\|f\|_\infty = \inf \{C, |f| \leq C \text{ p.p.}\}, \quad (\text{II.4.1})$$

appelé supremum *essentiel* de la fonction f sur l'espace mesuré X . On définit l'espace $L^\infty(X)$ à partir de $\tilde{L}^\infty(X)$ en identifiant les fonctions égales presque partout, c'est-à-dire que $L^\infty(X)$ est l'espace des classes d'équivalence de $\tilde{L}^\infty(X)$ pour la relation d'équivalence

$$f \mathcal{R} g \iff f = g \text{ p.p.}$$

On vérifie immédiatement que la quantité $\|f\|_\infty$ (que l'on appellera *norme* de f) est bien définie pour une classe, puisque la valeur est la même pour deux fonctions de $\tilde{L}^\infty(X)$ égales presques partout.

Remarque II.4.2. On prendra garde au fait que, dans la pratique courante, il subsiste une certaine ambiguïté entre $\tilde{L}^\infty(X)$ et $L^\infty(X)$. En particulier, lorsque l'on écrit $f \in L^\infty(X)$, on considère parfois f comme une fonction au sens usuel, ce qui peut amener à écrire pour deux fonctions f et g de cet espace que f et g sont égales presque partout, ce qui n'a a priori pas de sens s'il s'agit de *classes* de fonctions (on devrait écrire simplement $f = g$ si l'on considérait les classes). En revanche lorsque l'on établit des propriétés de cet espace, il s'agit bien de l'espace des classes. Nous verrons en particulier que $\|\cdot\|_\infty$ définit une *norme* sur $L^\infty(X)$, ce qui n'est vrai que si l'on considère l'espace des classes d'équivalence. Cette quantité n'est en effet pas une norme sur $\tilde{L}^\infty(X)$: dès que \mathcal{A} admet des ensembles de mesure nulle, il existe des fonctions qui annulent $\|\cdot\|_\infty$ (toutes les fonctions nulles presque partout).

Lemme II.4.3. Pour tout $f \in L^\infty(X)$, on a

$$|f(x)| \leq \|f\|_\infty \quad \text{p.p.}$$

Démonstration. Il existe une suite (C_n) convergeant vers $\|f\|_\infty$ telle que

$$|f(x)| \leq C_n \quad \forall x \in X \setminus E_n,$$

avec $E_n \in \mathcal{A}$ négligeable. On note E l'union des E_n , qui est aussi négligeable, et l'on a, pour tout n ,

$$|f(x)| \leq C_n \quad \forall x \in X \setminus E,$$

d'où $|f(x)| \leq \|f\|_\infty$ pour tout x dans $X \setminus E$. □

Proposition II.4.4. L'ensemble $L^\infty(X)$ est un espace vectoriel, et $\|\cdot\|_\infty$ est une norme sur cet espace.

Démonstration. La structure vectorielle de $L^\infty(X)$ est immédiate d'après la définition. On a $\|f\|_\infty = 0$ si et seulement si f est nulle presque partout, et la 1 – homogénéité est immédiate. Pour tous f et g dans L^∞ (plus précisément des représentants de leurs classes respectives dans L^∞), on a

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty$$

presque partout. □

20. Nous omettrons dans la définition la référence explicite à la tribu \mathcal{A} et la mesure μ , pour alléger l'écriture.

Proposition II.4.5. L'espace $(L^\infty(X), \|\cdot\|_\infty)$ est un espace de Banach, c'est-à-dire un espace vectoriel normé complet.

Démonstration. Soit (f_n) une suite de Cauchy dans L^∞ . Pour tout $k \geq 1$, il existe N_k tel que,

$$\|f_m - f_n\|_\infty \leq \frac{1}{k} \quad \forall m, n \geq N_k,$$

ce qui signifie que, pour tous m, n plus grands que N_k , il existe $E_{m,n,k}$ négligeable tel que $|f_m(x) - f_n(x)| \leq 1/k$ sur $X \setminus E_{m,n,k}$. L'ensemble $E = \cup E_{m,n,k}$ est négligeable comme union dénombrable d'ensembles négligeables et, sur $X \setminus E$, la suite $(f_n(x))$ est de Cauchy, donc converge dans \mathbb{R} . En passant à la limite dans l'inégalité précédente, on obtient $|f_n(x) - f(x)| \leq 1/k$, d'où l'on déduit que $f \in L^\infty$, et que f_n converge vers f pour cette norme. \square

II.4.2 Les espaces $L^p(X)$, pour $p \in [1, +\infty[$

Définition II.4.6. Soit (X, \mathcal{A}, μ) un espace mesuré et $p \in [1, +\infty[$. On note $L^p(X)$ l'ensemble des fonctions f qui sont \mathcal{A} -mesurables et telles que $|f|^p$ est intégrable. On note

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

Comme pour L^∞ , on définit en fait cet ensemble comme l'espace des classes d'équivalences obtenu en identifiant les fonctions égales presque partout. Comme précédemment, on prendra garde au fait que dans la pratique, lorsque l'on écrit $f \in L^p(X)$, on considère en fait f comme une fonction (un représentant de sa propre classe), voir remarque II.4.2.

Proposition II.4.7. (Inégalité de Hölder)

Soit $p \in]1, +\infty[$, $f \in L^p$, et $g \in L^{p'}$, où p' est l'exposant conjugué de p , tel que $\frac{1}{p} + \frac{1}{p'} = 1$. On a alors $fg \in L^1$, avec

$$\|fg\|_{L^1} \leq \|f\|_{L^p} \|g\|_{L^{p'}}.$$

Démonstration. Soient $f \in L^q$ et $g \in L^{p'}$. D'après l'inégalité de Young (proposition A.1.43, page A.1.43) on a, pour tout $x \in X$,

$$|f(x)| |g(x)| \leq \frac{1}{p} |f(x)|^p + \frac{1}{p'} |f(x)|^{p'},$$

d'où $fg \in L^1$, et

$$\|fg\|_{L^1} \leq \frac{1}{p} \|f\|_{L^p} + \frac{1}{p'} \|f\|_{L^{p'}}.$$

On obtient, en remplaçant f par λf (avec $\lambda > 0$) dans l'inégalité ci-dessus,

$$\|fg\|_{L^1} \leq \frac{\lambda^{p-1}}{p} \|f\|_{L^p} + \frac{1}{\lambda p'} \|g\|_{L^{p'}}.$$

La fonction ci-dessus est convexe en λ , et tend vers $+\infty$ quand λ tend vers 0 et vers $+\infty$, elle est minimale pour $\lambda = \|f\|_{L^p}^{-1} \|f\|_{L^{p'}}^{p'/p}$, ce qui conclut la preuve. \square

Proposition II.4.8. L'ensemble $L^p(X)$ est un espace vectoriel, et $\|\cdot\|_p$ est une norme sur $L^p(X)$.

Démonstration. Pour tous f et g dans $L^p(X)$, tout $x \in X$, on a

$$|f(x) + g(x)|^p \leq (|f(x)| + |g(x)|)^p \leq (2 \max(|f(x)|, |g(x)|))^p \leq 2^p (|f(x)|^p + |g(x)|^p),$$

d'où $f + g \in L^p(X)$. On a par ailleurs $\lambda f \in L^p(X)$ pour tout $\lambda \in \mathbb{R}$.

L'inégalité triangulaire est une conséquence de l'inégalité de Hölder. En effet, pour tous f et g dans $L^p(X)$, on a

$$\int |f + g|^p = \int |f + g|^{p-1} |f + g| \leq \int |f + g|^{p-1} |f| + \int |f + g|^{p-1} |g|.$$

Or, si p' est l'exposant conjugué de p , on a $p'(p-1) = p$, d'où l'on déduit que la fonction $|f + g|^{p-1}$ est dans $L^{p'}(X)$. D'après l'inégalité de Hölder (proposition II.4.7), appliquée successivement aux deux intégrales du membre de droite ci-dessus, on a

$$\int |f + g|^p \leq \left(\int |f + g|^p \right)^{1/p'} \left(\int |f|^p \right)^{1/p} + \left(\int |f + g|^p \right)^{1/p'} \left(\int |g|^p \right)^{1/p}$$

d'où l'on déduit, en utilisant $1/p' = 1 - 1/p$, l'inégalité triangulaire (l'inégalité est trivialement vérifiée si $\int |f + g|^p = 0$). \square

Proposition II.4.9. Soit $p \in [1, +\infty[$. L'espace vectoriel normé $(L^p(X), \|\cdot\|_p)$ est complet.

Démonstration. Soit f_n une suite de Cauchy dans $L^p(X)$. Il suffit de montrer qu'il existe une sous-suite extraite convergente dans L^p , le caractère de Cauchy assurera la convergence de l'ensemble de la suite vers la même limite. La première étape consiste à extraire une sous-suite f_{n_k} telle que

$$\|f_{n_{k+1}} - f_{n_k}\|_{L^p} \leq \frac{1}{2^k}.$$

On procède de la façon suivante : il existe n_1 tel que $\|f_m - f_n\|_{L^p} \leq 1/2$ pour tous m et n plus grands que n_1 . Il existe ensuite un $n_2 \geq n_1$ tel que la même quantité est majorée par $1/4$, etc ... On construit ainsi une sous-suite qui vérifie l'inégalité ci-dessus. Pour simplifier l'écriture, nous écrivons (f_k) cette sous-suite, qui vérifie donc $\|f_{k+1} - f_k\|_{L^p} \leq 1/2^k$.

On introduit maintenant la fonction g_n définie par

$$g_n(x) = \sum_{k=1}^n |f_{k+1}(x) - f_k(x)|.$$

On a par construction $\|g_n\|_{L^p} \leq 1$. La suite $g_n(x)$ est croissante pour tout x , donc converge vers une limite

$$g(x) = \sum_{k=1}^{+\infty} |f_{k+1}(x) - f_k(x)| \in [0, +\infty].$$

La suite $g_n(x)^p$ est elle-même croissante, et converge simplement vers $g(x)^p$. L'intégrale de g^p est donc finie d'après le théorème de convergence monotone B.7.23, page 228. La fonction g est donc dans $L^p(X)$, et $g(x)$ est fini pour presque tout x (voir proposition B.7.22). On a par ailleurs, pour tous $m > n \geq 2$,

$$\begin{aligned} |f_m(x) - f_n(x)| &\leq |f_m(x) - f_{m-1}(x)| + \cdots + |f_{n+1}(x) - f_n(x)| \\ &\leq \sum_{k=n}^{+\infty} |f_{k+1}(x) - f_k(x)| = g(x) - g_{n-1}(x). \end{aligned}$$

Cette dernière quantité étant finie presque partout, et du fait que $g_n(x)$ converge vers $g(x)$, la suite $f_n(x)$ est de Cauchy pour presque tout x , donc converge vers une limite, que l'on note $f(x)$. On fait tendre m vers $+\infty$ dans l'inégalité précédente. On a ainsi, pour presque tout x et $n \geq 2$,

$$|f(x) - f_n(x)| \leq g(x) \implies |f(x)| \leq |f_n(x)| + g(x),$$

d'où $f \in L^p(X)$. Enfin, $|f_n(x) - f(x)|^p$ tend vers 0 presque partout, et $|f_n(x) - f(x)|^p \leq g(x)^p$, avec g^p intégrable. Le théorème de convergence dominée assure donc la convergence de f_n vers f en norme L^p . \square

Notation II.4.10. On note $L^p(X)^n$ l'espace des fonctions vectorielles, c'est-à-dire à valeurs dans \mathbb{R}^n , donc chaque composante est dans $L^p(X)$. Cet espace est complet pour la norme²¹

$$\|f\|_{L^p(X)^n} = \left(\int_X \|f(x)\|_2^p d\mu(x) \right)^{1/p}.$$

II.4.3 Les espaces $L^p(\mathbb{N}) = \ell^p$ et $L^p(\mathbb{R}^d)$

Espaces de suites

On considère dans un premier temps le cas $X = \mathbb{N}$, muni de la tribu discrète (ensemble des parties) et de la mesure de comptage canonique :

$$A \subset \mathbb{N} \longmapsto \mu(A) = \text{Card}(A).$$

(Noter que l'on peut écrire μ comme la somme des masses ponctuelles δ_n , pour n parcourant \mathbb{N} .)

Une “fonction” sur $X = \mathbb{N}$, c'est-à-dire une application de \mathbb{N} dans \mathbb{R} , peut s'écrire comme une suite (u_n) de $\mathbb{R}^{\mathbb{N}}$. Noter que, si les fonctions telles qu'on les a définies sont a priori autorisées à prendre la valeur $+\infty$, imposer l'appartenance à l'un des espaces L^p impose que toutes les valeurs soient finies. L'espace $L^p(\mathbb{N})$, pour $p \in [1, +\infty[$, s'identifie dans ce cas à l'espace des suites noté en général ℓ^p , défini par

$$\ell^p = \left\{ (u_n) \in \mathbb{R}^{\mathbb{N}}, \sum |u_n|^p < +\infty \right\}.$$

Nous avons montré précédemment qu'il s'agit d'un espace vectoriel normé complet pour la norme p , définie par

$$u = (u_n) \longmapsto \|u\|_{\ell^p} = \left(\sum |u_n|^p \right)^{1/p}.$$

L'espace ℓ^∞ des suites bornées est de la même manière un espace vectoriel normé pour

$$\|u\|_{\ell^\infty} = \sup_n |u_n|.$$

On notera qu'il s'agit ici d'un sup “traditionnel”, du fait que l'espace \mathbb{N} muni de la mesure de comptage ne contient aucun ensemble non vide qui soit négligeable.

Il peut être pertinent de construire de tels espaces de suites à partir d'une mesure non uniforme sur \mathbb{N} , pour représenter par exemple une collection de masses ponctuelles non identiques. On considère dans cet esprit une collection infinie de masses strictement positives $(m_i)_{i \in \mathbb{N}}$, et la mesure associée

$$A \subset \mathbb{N} \longmapsto m(A) = \sum_{i \in A} m_i \in [0, +\infty].$$

Si l'on se donne une fonction vectorielle sur X , à valeur dans \mathbb{R}^3 , qui correspond aux vitesses des masses, notée $v = (v_i)_{i \in \mathbb{N}}$, sa norme en tant qu'élément de l'espace $L^2(X, m)$ est définie par

$$\|v\|_{L^2}^2 = \sum_{i \in A} m_i |v_i|^2,$$

qui est (au facteur 1/2 près), l'énergie cinétique du système de masses. Le fait que cette énergie soit bornée n'empêche pas qu'il y ait des vitesses arbitrairement grandes. Dans ce qui précède nous avons considéré des particules labellisées, mais non situées dans l'espace. On peut construire un cadre fonctionnel permettant de suivre leurs positions dans l'espace en considérant la mesure discrète

$$\mu = \sum_{i=1}^{+\infty} m_i \delta_{x_i},$$

21. On notera l'utilisation de la norme euclidienne dans \mathbb{R}^n (indice 2 dans $\|f\|_2$). Ce choix est le plus couramment effectué, mais on pourrait munir \mathbb{R}^n d'une autre norme.

où les x_i sont des points de l'espace \mathbb{R}^3 (position des masses). La mesure μ , définie sur la tribu discrète (ensemble des parties de \mathbb{R}^3) est définie par

$$A \subset \mathbb{R}^3 \longmapsto \mu(A) = \sum_{i, x_i \in A} m_i.$$

Noter que si l'on cherche à modéliser de cette façon (dite eulérienne) un nuage de particules en mouvement, l'objet naturel est une mesure μ_t dépendant du temps, définie comme ci-dessus à partir des positions courantes des particules. L'espace naturel pour représenter la vitesse dépendra alors lui-même du temps (contrairement au cadre précédent, purement lagrangien), puisque la mesure de référence, définie comme mesure sur \mathbb{R}^3 , dépend du temps.

Espaces de fonctions sur \mathbb{R}^d

La construction de la mesure le Lebesgue n'était pas nécessaire pour construire les espaces de suites ci-dessus, elle l'est en revanche pour les espaces fonctionnels correspondant au cas $X = \mathbb{R}$, $X = \mathbb{R}^d$, ou $X = \Omega$, avec Ω ouvert de \mathbb{R}^d . Si l'on considère ainsi un ouvert Ω de \mathbb{R}^d muni canoniquement de la mesure de Lebesgue²², la construction précédente permet en particulier d'identifier, pour $p \in [1, +\infty[$, l'espace

$$L^p(\Omega) = \left\{ f \text{ mesurable sur } \Omega, \int_{\Omega} |f(x)|^p dx < +\infty \right\}$$

à un espace vectoriel normé complet pour la norme

$$\|f\|_{L^p} = \left(\int_{\Omega} |f(x)|^p dx \right)^{1/p}.$$

De la même manière l'espace L^∞ est un espace de Banach pour la norme

$$\|f\|_{L^\infty} = \sup_{x \in \Omega} |f(x)|,$$

étant entendu qu'il s'agit ici du supremum *essentiel*, tel que défini par (II.4.1).

Exercice II.4.1. Construire une isométrie entre ℓ^p et un sous-espace vectoriel de $L^p(\mathbb{R})$, pour $p \in [1, +\infty[$.

CORRECTION.

À tout tout $u = (u_n) \in \ell^p$ on associe la fonction

$$Tu = f = \sum_{n=0}^{+\infty} u_n \mathbf{1}_{]n, n+1[}.$$

Si l'on note f_n la somme partielle, $|f_n|^p$ est une suite croissante de fonctions mesurables, qui converge presque partout vers $|f|^p$. D'après le théorème de convergence monotone (théorème B.7.23, page 228), comme $\int |f_n|^p$ converge vers $\sum u_n^p < +\infty$, la fonction $|f|^p$ est intégrable, et d'intégrale $\|u\|_{\ell^p}^p$. L'application T réalise donc une isométrie entre ℓ^p et un sous-espace strict de $L^p(\mathbb{R})$.

Proposition II.4.11. L'espace $C_c(\mathbb{R})$ des fonctions continues à support compact est dense dans $L^p(\mathbb{R})$, pour tout $p \in [1, +\infty[$.

Démonstration. Nous démontrons tout d'abord cette propriété dans le cas $p = 1$. Toute fonction de $L^1(\mathbb{R})$ peut s'écrire comme somme d'une fonction positive et d'une fonction négative. Il suffit donc de montrer que l'on peut approcher une fonction positive par une fonction de C_c . D'après la proposition B.7.6 et le théorème de convergence monotone B.7.23, toute fonction positive peut être approchée avec une précision arbitraire en

22. On utilisera ici la notation dx (à la place de $d\lambda$) pour représenter le volume élémentaire d'intégration associé à la mesure de Lebesgue, conformément à l'usage.

norme L^1 par une fonction étagée. Toute fonction étagée s'écrit comme combinaison linéaire de fonctions de type $\mathbb{1}_A$, où A est dans la tribu de Lebesgue. On se ramène donc à la question de savoir si l'on peut approcher toute fonction de type $\mathbb{1}_A$ par une suite de fonctions continues à support compact. On utilise maintenant la régularité de la mesure de Lebesgue (théorème B.5.3, page 211), qui est en fait une conséquence directe de sa définition à partir de la mesure extérieure de Lebesgue. Il existe un ouvert U contenant A tel que $\lambda(U \setminus A) = \|\mathbb{1}_U - \mathbb{1}_A\|_{L^1}$ est arbitrairement petit, ce qui nous ramène à l'approximation d'une fonction de type $\mathbb{1}_U$, avec U ouvert. La suite de fonctions $\mathbb{1}_{]-n,n]}\mathbb{1}_U$ converge en norme L^1 vers $\mathbb{1}_U$ d'après le théorème de convergence monotone, on se ramène ainsi à l'approximation d'une fonction de type $\mathbb{1}_U$, avec U ouvert borné. On considère la fonction f_n définie par

$$f_n(x) = \min(nd(x, U^c), 1),$$

où $d(x, U^c)$ est la distance de x au complémentaire de U . Il s'agit d'une suite qui converge simplement vers $\mathbb{1}_U$, dominée par $\mathbb{1}_U$, on a donc, d'après le théorème de convergence dominée,

$$\|\mathbb{1}_U - f_n\|_{L^1} = \int (\mathbb{1}_U - f_n) = \int \mathbb{1}_U - \int f_n \rightarrow 0.$$

Les fonctions f_n étant continues et à support compact, nous avons ainsi montré que l'on peut approcher toute fonction de L^1 par une suite de telles fonctions.

Pour la convergence en norme L^p , on utilise le fait que toute fonction de L^p est approchable par une suite de fonctions bornées et à support compact (voir exercice II.6.15). On peut donc supposer la fonction f de L^p bornée par un certain M et à support compact. Elle est donc dans L^1 , et peut être approchée par une suite (f_n) de fonctions continues à support compact en norme L^1 . On a

$$|f(x) - f_n(x)|^p \leq M^{p-1} |f(x) - f_n(x)|.$$

L'intégrale de la fonction ci-dessus converge vers 0, d'où la convergence en norme L^p de f_n vers f . \square

Exercice II.4.2. Quelle est l'adhérence de $C_c(\mathbb{R})$ (fonctions continues à support compact) dans $L^\infty(\mathbb{R})$?

CORRECTION.

L'adhérence de $C_c(\mathbb{R})$ dans $L^\infty(\mathbb{R})$ est l'espace des fonctions continues qui tendent vers 0 en $\pm\infty$. Le fait que ce soit des fonctions continues est immédiat par convergence uniforme. Par ailleurs, si f est limite des f_n dans $C_c(\mathbb{R})$, pour tout $\varepsilon > 0$, il existe n tel que $\|f_n - f\|_\infty < \varepsilon$. Comme f_n est à support compact, elle est nulle en dehors de $[-M, M]$, on a donc $|f(x)| < \varepsilon$ pour tout x avec $|x| \geq M$.

Réiproquement, si f continue tend vers 0 en $\pm\infty$, on peut approcher f en norme uniforme par la suite de fonctions f_n continues à support compact, où f_n s'identifie à f sur $[-n, n]$ est affine sur les intervalles $[-n-1, -n]$ et $[n, n+1]$, et s'annule au delà.

II.5 Compléments

Théorème II.5.1. (Radon-Nikodym)

Soit μ et λ deux mesures σ -finies sur l'espace mesurable (X, \mathcal{A}) . On suppose que μ est absolument continue par rapport à λ (voir définition B.3.10, page 203). Alors il existe une fonction positive g mesurable telle que

$$\mu(A) = \int_A g d\lambda.$$

On dira que g est la *densité* de μ relativement à λ .

Définition II.5.2. (Points de Lebesgue)

Soit f une fonction intégrable sur \mathbb{R}^d . On dit que x est un point de Lebesgue de f si

$$\lim_{r \rightarrow 0} \frac{1}{\lambda(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| dy = 0.$$

Théorème II.5.3. (de différentiation de Lebesgue)

Soit f une fonction intégrable sur \mathbb{R}^d . Les points de Lebesgue forment un ensemble de mesure pleine, c'est à dire que l'ensemble des points qui ne sont pas des points de Lebesgue est négligeable.

II.6 Exercices

Exercice II.6.1. (Sommes de Darboux)

On note Λ l'ensemble des *subdivisions* de l'intervalle I , c'est à dire l'ensemble des

$$\sigma = (x_0, \dots, x_n), \quad a = x_0 < x_1 < \dots < x_n = b, \quad t = (t_1, \dots, t_n), \quad t_j \in [x_{j-1}, x_j] \quad \forall j = 1, \dots, n,$$

où $n \geq 1$ est un entier (qui dépend de σ). Soit f une fonction bornée sur $[a, b]$. On définit les sommes de Darboux inférieure et supérieure par

$$s_{[a,b]}(f, \sigma) = \sum_{i=1}^n m_i(x_i - x_{i-1}), \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x), \quad (\text{II.6.1})$$

$$S_{[a,b]}(f, \sigma) = \sum_{i=1}^n M_i(x_i - x_{i-1}), \quad M_i = \sup_{x \in [x_{i-1}, x_i]} f(x). \quad (\text{II.6.2})$$

On définit

$$s_{[a,b]} = \sup_{\sigma \in \Lambda} s_{[a,b]}(f, \sigma), \quad S_{[a,b]} = \inf_{\sigma \in \Lambda} S_{[a,b]}(f, \sigma).$$

Montrer que ces quantités sont bien définies, avec $s_{[a,b]} \leq S_{[a,b]}$. Si ces deux quantités sont égales, on dit que la fonction est intégrable au sens de Darboux.

Montrer qu'une fonction est intégrable au sens de Darboux si et seulement si elle est intégrable au sens de Riemann²³.

Exercice II.6.2. a) Soit (x_n) une suite de réels positifs telle que la série $\sum x_n$ converge. Montrer que la valeur de la somme ne dépend pas de l'ordre dans lequel on effectue les sommes.

b) Soit maintenant (x_n) une suite de réels telle que la série $\sum x_n$ soit convergente, mais sans que la série soit absolument convergente. On chercher à montrer que, selon l'ordre dans lequel on effectue la somme, on peut obtenir essentiellement *n'importe quoi*.

Plus précisément, montrer que, pour tout $\lambda \in [-\infty, +\infty]$, il existe une bijection φ sur \mathbb{N} telle que telle que la série $\sum x_{\varphi(n)}$ converge vers λ . Montrer qu'il existe aussi une bijection telle que l'ensemble des valeurs d'adhérence de la série soit \mathbb{R} tout entier.

CORRECTION.

a) Soit φ une bijection de \mathbb{N} . Pour tout $N \in \mathbb{N}$, il existe N' tel que $\{\varphi(n), n \leq N'\}$ contienne tous les indices de 0 à N , on a donc

$$\sum_0^N x_n \leq \sum_0^{N'} x_{\varphi(n)} \leq \sum_0^{+\infty} x_{\varphi(n)},$$

d'où

$$\sum_0^{+\infty} x_n \leq \sum_0^{+\infty} x_{\varphi(n)}.$$

Et l'on montre l'inégalité inverse de la même manière : pour tout N il existe N' tels que $\{n, n \leq N'\}$ contienne $\{\varphi(n), n \leq N\}$, etc ...

b) On déduit des hypothèses que x_n tend vers 0, et que x_n n'est pas nulle au-delà d'un certain rang. Par

23. Les sommes de Riemann (équation (II.3.1), page 49) sont définies pour des fonctions non nécessairement bornées, mais la proposition II.3.2, page 49, assure qu'une fonction Riemann-intégrable est bornée.

ailleurs, si l'on note I_+ l'ensemble des indices tels que $x_n \geq 0$, et I_- l'ensemble des indices tels que $x_n < 0$, ces deux ensembles sont infinis, et l'on a (du fait de la non convergence absolue)

$$\sum_{I_+} x_n = +\infty, \quad \sum_{I_-} x_n = -\infty.$$

Pour produire une limite infinie, on commence par prendre des indices dans I_+ jusqu'à ce que la série partielle dépasse 1, puis on prend le premier indice de I_i , puis on continue avec I_+ jusqu'à dépasser 2, puis le deuxième indice de I_- etc ... Le nombre d'étape est forcément infini du fait de la divergence de la série sur I_+ . On parcourt ainsi tous les indices de I_+ et I_- , et la série diverge vers $+\infty$ par construction (noter que les négatifs que l'on ajoute un par un à chaque étape tendent vers 0). On procède symétriquement pour $-\infty$. Pour λ , par exemple positif, on commence par parcourir les indices de I_+ jusqu'à dépasser λ , puis on passe à I_- jusqu'à passer en dessous de λ , puis I_+ , etc.... Comme la suite des x_n tend vers 0, on a bien convergence vers λ . Pour avoir un ensemble de valeurs prises dense dans \mathbb{R} , on monte jusqu'à dépasser 1 avec des indices de I_+ , puis on descend sous -2 , puis on remonte au-dessus de 3 , etc Comme x_n tend vers 0, les progressions de $-n$ à $n+1$ se font avec un pas de plus en plus petit, on finit donc par intersester n'importe quel intervalle ouvert, aussi petit soit-il.

Cet exercice montre que cela n'a en général aucun sens d'écrire

$$\sum_{n \in I} x_n$$

lorsque I est un ensemble infini dénombrable et que les x_n prennent des valeurs positives et négatives, du fait que l'ordre dans lequel on effectue les additions conditionne fortement le résultat obtenu.

Exercice II.6.3. Ce petit exercice illustre d'une certaine manière l'impossibilité de définir une mesure finie sur un ensemble infini non dénombrable (comme $[0, 1] \subset \mathbb{R}$) en affectant un point strictement positif à chaque point.

Soit $(a_i)_{i \in I}$ une famille infinie de réel positifs ou nuls, possédant la propriété suivante :

$$\forall J \subset I, \quad J \text{ dénombrable}, \quad \sum_{i \in J} a_i < +\infty.$$

Montrer que l'ensemble des i tels que $a_i > 0$ est nécessairement dénombrable.

CORRECTION.

Pour tout n , on pose

$$I_n = \left\{ i \in I, \quad a_i > \frac{1}{n} \right\}.$$

D'après l'hypothèse, cet ensemble est fini. L'ensemble

$$I_\infty = \{i \in I, \quad a_i > 0\} = \bigcup_{n \in \mathbb{N}} I_n$$

est donc dénombrable comme union dénombrable d'ensembles finis.

Exercice II.6.4. (Tribus sur l'ensemble à N éléments (••))

- a) Décrire l'ensemble des tribus sur $X_N = \llbracket 1, N \rrbracket$.
- b) Préciser le cardinal de chacune de ces tribus.
- c) (• • •) Préciser le nombre B_N total de tribus sur N , sous la forme d'une relation de récurrence, qui exprime B_{N+1} en fonction des B_n jusqu'à N .

CORRECTION.

a) Soit \mathcal{A} une tribu sur N . Soit $x \in X$. On considère A_x le plus petit élément de \mathcal{A} qui contient x (intersection

de tous ceux qui contiennent x). On considère maintenant y dans A_x . Si $x \notin A_y$, alors $A_x \setminus A_y$ est élément de A , est strictement inclus dans A , et contient x , ce qui est absurde. On a donc nécessairement $x \in A_y$. Alors A_y s'identifie nécessairement à A_x , car sinon l'intersection des deux est strictement incluse dans au moins l'un d'eux, par exemple A_x , ce qui est absurde car alors A_x ne serait plus le plus petit élément de A contenant x . La partie A_x contient donc les éléments tels que A_x soit le plus petit élément de A les contenant. On poursuit avec un point hors de A_x (s'il y en a un), et l'on continue jusqu'à avoir recouvert X_N par p parties non vides disjointes, qui constituent donc une partition P . La tribu engendrée par la partition P est contenue dans A . Pour tout élément A de \mathcal{A} , son intersection avec chacune des parties de la partition est la partie-elle même, sinon elle ne serait pas minimale. La tribu A ne contient donc que des unions de parties de P , donc elle est dans la tribu engendrée par P . On établit ainsi une correspondance univoque (pour un ensemble fini, rappelons-le) entre tribus et partitions. Toute tribu sur X_N est engendrée par une partition de X_N , elle est ainsi caractérisée par cette partition.

b) Le cardinal d'une tribu est défini par la granularité de la partition qui l'engendre. Si la partition contient p parties ($1 \leq p \leq N$), alors la tribu contient autant d'éléments que l'ensemble à p éléments contient de parties, c'est à dire 2^p . Une autre manière de voir les choses est de considérer que l'on peut coder un élément de la tribu par un mot de p bits, dont chaque bit indique si l'élément en question contient la partie correspondante.
c) Comme on l'a vu au a), compter les tribus revient à compter les partitions. On suppose connu le nombre B_k de tribus sur l'ensemble à k éléments, pour k entre 0 et N . On rajoute maintenant l'élément $x = n+1$ à X_N pour obtenir X_{N+1} . On se propose d'énumérer les partitions sur X_{N+1} en les classant selon le nombre k d'éléments qui ne sont pas dans la partie qui contient x . Pour $k = 0$, on a une seule partition, constituée de l'ensemble X_{N+1} entier. Pour $k = 1$, on a N possibilités pour choisir l'élément qui n'est pas dans la partie contenant x , et chaque choix correspond à une seule partition. Pour $k \geq 2$, on a C_N^k possibilités de choisir les éléments qui ne sont pas dans la partie contenant x . Pour chacun de ces choix, on a B_k partitions possibles. On a donc

$$B_{N+1} = \sum_{k=0}^N C_N^k B_k,$$

avec $B_0 = 1$ (on considère que l'ensemble vide admet une partition unique, qui est lui-même), et $B_1 = 1$ (tout singleton $\{1\}$ admet une partition unique). On appelle B_N le N -ième nombre de Bell, dont on peut montrer qu'il est égal à

$$B_N = \sum_{k=0}^{+\infty} \frac{k^N}{k!}.$$

Exercice II.6.5. Soit X un ensemble. On note \mathcal{A} l'ensemble des parties A telles que A ou A^c est dénombrable. Montrer que \mathcal{A} est une tribu. Quelle est cette tribu dans le cas où X est fini ou dénombrable ?

CORRECTION.

Les deux premières conditions sont trivialement vérifiées. Maintenant considérons une famille (A_n) d'éléments de \mathcal{A} . Si tous les A_n sont dénombrables, alors leur union l'est. Si l'un ne l'est pas, alors son complémentaire l'est nécessairement, et donc

$$\left(\bigcup A_n \right)^c = \left(\bigcap A_n^c \right)$$

est dénombrable, donc l'union est bien dans \mathcal{A} . Si l'ensemble est fini ou dénombrable, toute partie est dans \mathcal{A} , on retrouve donc la tribu discrète.

Exercice II.6.6. Soit X un ensemble infini. Décrire la tribu engendrée par la collection des singletons, selon que X soit dénombrable ou pas.

CORRECTION.

Si X est dénombrable, toute partie de X est dénombrable, donc réunion dénombrable de singletons. La tribu engendrée par les singletons est donc la tribu discrète $\mathcal{P}(X)$.

Si X n'est pas dénombrable, la tribu engendrée par les singletons contient au moins toutes les parties dénombrables, ainsi que celles de complémentaire dénombrable. Or la famille \mathcal{A} constituées de ces parties est une

tribu. En effet, la stabilité par complémentarité est immédiate. Pour la propriété sur l'union dénombrable, il s'agit de l'exercice II.6.5, dont nous reformulons la démonstration. Considérons une famille (A_n) de \mathcal{A} . Si tous les A_n sont dénombrables, alors leur réunion l'est, elle est donc dans \mathcal{A} . Si l'un d'eux, mettons A_1 n'est pas dénombrable, alors A_1^c l'est, donc l'intersection des complémentaires est dénombrable, et donc son complémentaire, qui est l'union des A_n , est dans \mathcal{A} . Il s'agit donc bien de la tribu engendrée par les singletons.

Exercice II.6.7. Montrer qu'il existe un ensemble dénombrable qui engendre la tribu borélienne de \mathbb{R} .

CORRECTION.

On sait que la tribu borélienne est engendrée par les intervalles du type $]-\infty, a]$. Chacun de ces intervalles peut s'écrire comme l'intersection d'intervalles $]-\infty, q_n]$, où q_n est une suite de nombres décimaux qui convergent par valeurs décroissantes vers a . La tribu engendrée part les $]-\infty, q]$, où q est décimal, contient donc les intervalles $]-\infty, a]$, et donc la tribu borélienne, et il s'agit d'un ensemble dénombrable car en bijection avec \mathbb{D} .

Exercice II.6.8. Donner des exemples de boréliens A de \mathbb{R} tels que

$$(i) \ 0 < \lambda(\mathring{A}) = \lambda(A), \ (ii) \ \lambda(\mathring{A}) = 0, \ \lambda(A) = +\infty, \ (iii) \ 0 < \lambda(\mathring{A}) < \lambda(A) < +\infty.$$

CORRECTION.

Pour (i) on peut prendre $A = [a, b]$,

Pour (ii) on peut prendre $A = \mathbb{R} \setminus \mathbb{Q}$. On a $\lambda(\mathring{A}) = 0$ et $\lambda(A) = +\infty$.

Pour (iii) on peut prendre $A =]0, 1] \cup]1, 2[\setminus \mathbb{Q}$. On a $\lambda(\mathring{A}) = 1$ et $\lambda(A) = 2$

Exercice II.6.9. Pour les suites de fonctions de \mathbb{R}_+ dans \mathbb{R} définies ci-dessous, préciser la fonction qui est limite simple de la suite, la limite des intégrales, l'intégrale de la limite, et préciser si ces 3 suites (f_n^1) , (f_n^2) , et (f_n^3) , rentrent dans le cadre du théorème de convergence monotone ou du théorème de convergence dominée :

$$f_n^1(x) = \frac{1}{x} \mathbf{1}_{]1/n, 1[}, \quad f_n^2(x) = \mathbf{1}_{]-n^2, n^2[}.$$

$$f_n^3(x) = -\frac{1}{n^2} \mathbf{1}_{]0, n[}.$$

CORRECTION.

La suite de fonctions (f_n^1) converge simplement et de façon croissante vers la fonction $\mathbf{1}_{]0, 1[}/x$. La suite des intégrales converge vers l'intégrale de la limite, qui est $+\infty$, ce qui est conforme au TCM.

La suite de fonctions (f_n^2) converge simplement et de façon croissante vers la fonction identiquement égale à 1. L'intégrale de la limite est $+\infty$, et la limite des intégrales ($\int f_n^2 = 2n^2$ converge vers $+\infty$, ce qui est conforme au TCM).

La suite de fonctions (f_n^3) converge simplement vers la fonction nulle. L'intégrale de la limite est 0, et la suite des intégrales converge vers 0. On a donc convergence de la suite des intégrales vers l'intégrale de la limite. La convergence est en effet dominée. En effet, comme $|f_n(x)| \leq 1/n^2$ sur $[n-1, n]$, elle est dominée par $1/x^2$ sur $[1, +\infty[$, et par 1 sur $[0, 1]$.

Exercice II.6.10. Les assertions suivantes sont-elles vraies ou fausses ?

(i) Soit f une fonction mesurable de \mathbb{R} dans $\overline{\mathbb{R}}$, telle que $\min(|f|, n)$ est intégrable sur \mathbb{R} , pour tout $n \in \mathbb{N}$. Alors la fonction f est elle-même intégrable sur \mathbb{R} .

(ii) Soit f une fonction mesurable de \mathbb{R} dans $\overline{\mathbb{R}}$, telle que $\min(|f|, n)$ est intégrable sur \mathbb{R} , d'intégrale inférieure à $C > 0$, pour tout $n \in \mathbb{N}$. Alors la fonction f est elle-même intégrable sur \mathbb{R} .

CORRECTION.

(i) Faux : la fonction f définie par $f(x) = 1/x^2$ sur $]0, +\infty[$, $f(x) = 0$ sur $]-\infty, 0]$, n'est pas intégrable sur \mathbb{R} , alors que $\min(|f|, n)$ est intégrable pour tout n .

(ii) Vrai. La fonction $\min(|f|, n)$ est d'intégrale $\leq C$, la suite des intégrales, croissante, converge donc vers une valeur $\leq C$. La suite $\min(|f|, n)$ est croissante et converge simplement vers $|f|$, on a donc convergence des intégrales d'après le TCM, d'où $\int |f| < \infty$, d'où l'intégrabilité de f .

Exercice II.6.11. Calculer les limites quand n tend vers $+\infty$ de

$$u_n = \int_0^1 \frac{1+nx^3}{(1+x^2)^n} dx, \quad v_n = \int_0^{+\infty} \frac{\sin(\pi x)}{1+x^n} dx$$

CORRECTION.

On note

$$f_n(x) = \frac{1+nx^3}{(1+x^2)^n}.$$

On a

$$0 \leq f_n(x) \leq \frac{1+nx^3}{1+nx^2} \leq \frac{1+nx^3}{1+nx^3} \leq 1,$$

et $f_n(x)$ converge vers 0 pour tout $x \in]0, 1]$. On a, d'après le théorème de convergence dominée, convergence des intégrales u_n vers $\int 0 = 0$. On note $g_n(x) = \sin(\pi x)/(1+x^n)$. Pour $n \geq 2$, $x \geq 1$, on a $|g_n(x)| \leq 1/(1+x^2)$, et $|g_n(x)| \leq 1$ sur $[0, 1]$, $|g_n(x)|$ est donc majoré par une fonction intégrable. On peut donc appliquer le théorème de convergence dominée qui assure la convergence des intégrales vers l'intégrale de la limite simple, qui est la fonction g définie par $g(x) = \sin(\pi x)$ sur $]0, 1[$ et 0 sur $]1, +\infty[$. On a donc convergence de v_n vers $2/\pi$.

Exercice II.6.12. Soit f une fonction de \mathbb{R} dans \mathbb{R} , intégrable. Décrire le comportement de la suite

$$u_n = \int_{\mathbb{R}} f(x) \cos(x)^n d\lambda.$$

CORRECTION.

La fonction définie par $f_n(x) = f(x) \cos(x)^n$ converge presque partout (sur le complémentaire des $k\pi x$) vers 0, et l'on a, pour tout x , $|f_n(x)| \leq |f(x)|$, avec $\int |f| < +\infty$. D'après le théorème de convergence dominée (théorème B.7.25, page 229), on a donc converge des intégrales vers l'intégrale de la limite, qui est 0.

Exercice II.6.13. Montrer que la boule unité fermée de $L^\infty(\mathbb{R})$ n'est pas compacte. Cette non compacité résulte-t-elle de la non compacité de \mathbb{R} ?

CORRECTION.

On considère la suite $f_n(x) = \mathbf{1}_{]0,n[}(x)$, qui est dans la boule unité fermée de L^∞ . Deux termes différents de cette suites sont toujours à distance 1, il est donc exclu que l'on puisse en extraire une sous-suite convergente. La non compacité ne vient pas de la non compacité de \mathbb{R} , on peut construire une suite analogue de fonctions à support compact, par exemple $f_n(x) = \mathbf{1}_{]0,1-1/n[}(x)$.

Exercice II.6.14. On se place sur l'intervalle borné $I =]a, b[$ muni de la mesure de Lebesgue.

- a) Montrer que $L^p(I) \subset L^1(I)$ pour tout $p \in]1, +\infty]$.
- b) Le sous-espace L^p est-il fermé dans L^1 ?
- c) Montrer que les inclusions du (a) sont invalidées si l'intervalle n'est pas borné (on pourra considérer le cas $I =]1, +\infty[$)

CORRECTION.

a) Soit $f \in L^p(I)$. On a, pour tout x dans I tel que $|f(x)| \geq 1$,

$$|f(x)| \leq |f(x)| |f(x)|^{p-1} = |f(x)|^p,$$

d'où l'on déduit que

$$|f(x)| \leq \max(1, |f(x)|^p),$$

et donc $|f(x)|$ est intégrable. b) Tout d'abord remarquons que l'inclusion ci-dessus est stricte, si l'on prend par exemple l'intervalle $I =]0, 1[$, la fonction définie par $f(x) = x^{-1/p}$ est dans L^1 , pas dans L^p . Par ailleurs $L^p(I)$ est dense dans L^1 , il contient en particulier les fonctions continues à support compact, qui sont denses dans L^1 . Comme cette densité des fonctions continues à support compact n'a pas été traitée en amphi, on peut aussi utiliser la troncature de l'exercice II.6.15 qui suit, si on l'a fait avant. Elle assure que l'on peut approcher toute fonction de L^1 par une suite de fonctions bornées, donc dans L^p (on est sur un intervalle borné). On a donc densité de $L^p(I)$ dans $L^1(I)$, mais inclusion stricte, L^p ne peut donc pas être fermé. c) Si l'intervalle n'est pas borné, par exemple si $I =]1, +\infty[$, la fonction $f(x) = 1/x$ est dans L^p pour tout $p > 1$, mais pas dans $L^1(I)$.

Exercice II.6.15. (Opérateur de troncature)

Soit $f \in L^p(\mathbb{R})$, avec $p \in [1, +\infty[$.

a) On définit

$$t \in \mathbb{R} \longmapsto T_n(t) = \begin{cases} t & \text{si } |t| \leq n \\ n \frac{t}{|t|} & \text{si } |t| > n \end{cases}$$

Montrer que $T_n \circ f$ tend vers f dans $L^p(\mathbb{R})$.

b) On note χ_n la fonction indicatrice de $] -n, n[$. Montrer que $\chi_n f$ tend vers f dans $L^p(\mathbb{R})$.

c) Montrer que $\chi_n T_n \circ f$ tend vers f dans L^p .

d) Que peut-on dire de $T_n \circ f$, $\chi_n f$, et $\chi_n T_n \circ f$, dans le cas $p = +\infty$?

CORRECTION.

a) On considère la suite de fonctions définie par

$$g_n(x) = |f(x) - T_n(x)|^p.$$

On a

$$|g_n(x)| \leq |f(x)|^p$$

qui est intégrable. Et $g_n(x)$ converge vers 0 pour presque tout x . On a donc convergence (dominée) de l'intégrale de g vers 0, d'où la convergence en norme L^p de $T_n \circ f$ vers f .

b) On peut considérer de la même manière

$$h_n(x) = |f(x) - \chi_n(x)f(x)|^p.$$

Le même raisonnement assure la convergence vers f de $\chi_n f$ vers f .

c) Le même raisonnement s'applique à $\chi_n T_n \circ f$.

d) Pour $p = +\infty$, $T_n \circ f$ est presque partout égal à f pour n assez grand, on a donc bien convergence en norme L^∞ . Pour $\chi_n f$ en revanche on n'a pas convergence. Pour $f \equiv 1$ par exemple, on a $\|f - \chi_n f\|$ identiquement égal à 1.

Exercice II.6.16. (Translations)

Soit $p \in [1, +\infty]$. On définit

$$\tau_h : f \in L^p \longmapsto \tau_h f, \quad \tau_a(f)(x) = f(x - a).$$

a) Montrer que, pour $p \in [1, +\infty[$, pour tout $f \in L^p(\mathbb{R})$, l'application qui à h associe $\tau_h f$ est continue de \mathbb{R} dans $L^p(\mathbb{R})$

b) Montrer que cette propriété n'est pas vraie pour $p = +\infty$.

CORRECTION.

a) On montre en premier lieu la propriété pour toute fonction g continue à support compact K . Une telle fonction est uniformément continue d'après le théorème de Heine (th. I.7.7, page 29). Pour tout $\varepsilon > 0$, il existe donc η tel que, pour tous x, y avec $|y - x| < \eta$, on ait $|g(y) - g(x)| < \varepsilon$. On a donc, pour tout $h < \eta$,

$$\|\tau_h g - g\|_p^p = \int_{\mathbb{R}} |g(x + h) - g(x)|^p dx \leq \text{diam}(K) \varepsilon^p,$$

d'où la propriété de convergence pour g . Considérons maintenant une fonction $f \in L^p$. Pour tout $\varepsilon > 0$, cette fonction peut être approchée à ε près par une fonction continue g à support compact (proposition II.4.11, page 65). Il existe alors η tel que, pour tout $h < \eta$, $\|\tau_h g - g\| < \varepsilon$. On a pour, un tel h ,

$$\|\tau_h f - f\|_p \leq \|\tau_h f - \tau_h g\|_p + \|\tau_h g - g\|_p + \|g - f\|_p < 3\varepsilon,$$

d'où la propriété.

b) La situation est différente pour $p = +\infty$. Considérons par exemple la fonction $f = \mathbf{1}_{]0,1[}$, toute translatée non triviale est à distance 1 d'elle-même.

Exercice II.6.17. Soit $I =]a, b[$ un intervalle borné, et $f \in L^\infty(I)$. Montrer que f appartient à tous les $L^p(I)$, et que

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

CORRECTION.

XXXX

Chapitre III

Espaces de Hilbert

Sommaire

III.1 Définitions, principales propriétés	75
III.2 Convergence faible	80
III.3 Sommes hilbertiennes, bases hilbertiennes	81
III.4 Théorie de Lax-Milgram	83
III.5 Opérateurs	85
III.6 Exercices	87

III.1 Définitions, principales propriétés

Définition III.1.1. (Produit scalaire)

Soit H un espace vectoriel sur \mathbb{R} . On appelle produit scalaire une forme bilinéaire $\langle u | v \rangle$ de $H \times H$ dans \mathbb{R} , symétrique, définie et positive :

$$\langle u | v \rangle = \langle v | u \rangle, \quad \langle u | u \rangle \geq 0 \quad \forall u \in H, \quad \text{et} \quad \langle u | u \rangle = 0 \iff u = 0.$$

Proposition III.1.2. (Inégalité de Cauchy-Schwarz)

Tout produit scalaire vérifie l'inégalité de Cauchy-Schwarz

$$|\langle u | v \rangle| \leq |u| |v| \quad \forall u, v \in H.$$

Démonstration. On écrit

$$\langle u + tv | u + tv \rangle \geq 0 \quad \forall t \in \mathbb{R}.$$

Le minimum de cette quantité est atteint en $t = -\langle u | v \rangle / |v|^2$, on a donc en particulier

$$|u|^2 - 2 \frac{\langle u | v \rangle^2}{|v|^2} + \frac{\langle u | v \rangle^2}{|v|^2} \geq 0,$$

d'où $|\langle u | v \rangle| \leq |u| |v|$. □

Proposition III.1.3. Un produit scalaire définit sur H une structure d'espace vectoriel normé pour la norme

$$u \mapsto |u| = \langle u | u \rangle^{1/2}.$$

Démonstration. La séparation est conséquence du caractère défini de $\langle \cdot | \cdot \rangle$, et l'homogénéité de sa bilinéarité. Pour l'inégalité triangulaire, on écrit

$$|u + v|^2 = |u|^2 + |v|^2 + 2\langle u | v \rangle \leq |u|^2 + |v|^2 + 2|u||v| = (|u| + |v|)^2.$$

□

Définition III.1.4. (Espace de Hilbert)

On appelle espace de Hilbert un espace vectoriel muni d'un produit scalaire, et qui est complet pour la norme associée.

Pour tout $d \geq 0$, \mathbb{R}^d muni de la norme ℓ^2 , est un espace de Hilbert (on parle plutôt d'espace euclidien lorsque la dimension est finie). Les espaces ℓ^2 et $L^2(\mathbb{R})$ (voir section II.4.3, cas $p = 2$), sont des espaces de Hilbert de dimension infinie¹.

Proposition III.1.5. Un espace de Hilbert est séparable (i.e. il contient une partie dénombrable dense, voir définition I.3.10) si et seulement s'il contient une famille dénombrable de vecteurs engendrant un sous-espace dense.

Démonstration. Si H est séparable, il admet une partie X dénombrable dense. L'ensemble des combinaisons linéaires (finies) à coefficient rationnels de ces éléments est elle-même dénombrable, et dense dans H puisqu'elle contient X . Réciproquement si H admet une famille (e_n) engendrant un sous-espace dense, on considère l'ensemble X de combinaisons linéaires *finies* des e_n à coefficients rationnels. Il s'agit d'un ensemble dénombrable, qui est dense dans H . □

Exercice III.1.1. Montrer que l'espace ℓ^2 muni du produit scalaire canonique est séparable.

CORRECTION.

On considère la "base canonique"² (e_n) , avec

$$e_n = (0, 0, \dots, 0, 1, 0, \dots).$$

Pour tout $u = (u_n) \in \ell^2$, la série des carrés étant convergente, la quantité

$$\sum_{n=0}^{+\infty} |u_n|^2$$

tend vers 0. On peut donc approcher en norme ℓ^2 , avec une précision arbitraire, u par la suite des $v^N = (u_0, u_1, \dots, u_{N-1}, \dots)$, combinaisons linéaires des e_n .

Exemple III.1.1. Tout espace de dimension finie munie d'un produit scalaire est un espace de Hilbert (espace Euclidien). En dimension infinie, l'exemple le plus simple d'espace de Hilbert de dimension infinie est l'espace ℓ^2 des suites de carré intégrable. On peut définir par extension une infinité de nouveaux espaces dits "à poids" en introduisant, pour $\gamma = (\gamma_n)$ une suite quelconque de réels strictement positifs,

$$\ell_\gamma^2 = \left\{ (u_n) \in \mathbb{R}^{\mathbb{N}}, \sum \gamma_n |u_n|^2 < +\infty \right\}.$$

Proposition III.1.6. (Identité du parallélogramme)

Toute norme issue d'un produit scalaire vérifie l'identité du parallélogramme

$$\left| \frac{u+v}{2} \right|^2 + \left| \frac{u-v}{2} \right|^2 = \frac{1}{2}(|u|^2 + |v|^2).$$

Démonstration. Il suffit de développer le membre de gauche. □

0. La démonstration est faite dans \mathbb{R}^d , mais on pourra vérifier qu'elle s'applique directement à n'importe quel produit scalaire.

1. Cela signifie simplement qu'il ne sont pas engendrés par une famille finie d'éléments. La fait qu'il existe une base algébrique (i.e. telle que tout élément s'écrive comme combinaison linéaire *finie* d'éléments de cette famille) est une question délicate (son existence repose sur l'axiome du choix). En tout cas une telle base, quand bien même elle existe, est complètement *inutilisable*.

2. On prendra garde au fait qu'il ne s'agit pas d'une base *algébrique* de ℓ^2 (i.e. telle que tout vecteur puisse s'écrire comme combinaison linéaire finie). On verra dans la section III.3 qu'il s'agit en revanche de ce que l'on appellera une *base hilbertienne*.

Proposition III.1.7. Tout sous-espace vectoriel fermé d'un espace de Hilbert est un espace de Hilbert (pour le même produit scalaire).

Démonstration. La propriété découle simplement du fait que la restriction d'un produit scalaire à un sous-espace est un produit scalaire, et qu'un sous-espace fermé d'un espace complet est complet (voir proposition I.5.4, page 23). \square

On sait que la distance d'un point à un fermé non vide de l'espace euclidien \mathbb{R}^d est atteinte (exercice I.9.9, page 36). Cette distance peut être atteinte en plusieurs points si le fermé n'est pas convexe. Pour un espace de Hilbert de dimension infinie, la distance peut ne pas être atteinte (voir exercice III.6.4, page 88). En revanche, si l'on suppose le fermé *convexe*, alors cette distance est atteinte, en un point unique, la convexité assurant à la fois l'existence et l'unicité. Ce résultat fondamental en optimisation fait l'objet du théorème qui suit.

Théorème III.1.8. (Projection sur un convexe fermé)

Soit H un espace de Hilbert et K un convexe fermé non vide de H . Pour tout $z \in H$, il existe un unique $u \in K$ (appelé projection de z sur K) tel que

$$|z - u| = \min_{v \in K} |z - v| = \text{dist}(z, K).$$

La projection u est caractérisée par la propriété

$$\left\{ \begin{array}{l} u \in K \\ \langle z - u | v - u \rangle \leq 0 \quad \forall v \in K. \end{array} \right. \quad (\text{III.1.1})$$

On notera $u = P_K z$.

Démonstration. On considère une suite minimisante (u_n)

$$u_n \in K, \quad |z - u_n| \longrightarrow d = \text{dist}(z, K).$$

Pour $p, q \in \mathbb{N}$, on applique l'identité du parallélogramme à $u_p - z$ et $u_q - z$:

$$\left| \frac{u_p + u_q}{2} - z \right|^2 + \left| \frac{u_p - u_q}{2} \right|^2 = \frac{1}{2}(|u_p - z|^2 + |u_q - z|^2).$$

Comme K est convexe $(u_p + u_q)/2 \in K$,

$$\left| \frac{u_p + u_q}{2} - z \right|^2 \geq d^2.$$

On a donc

$$\left| \frac{u_p - u_q}{2} \right|^2 \leq d^2 - d^2 + \varepsilon_p + \varepsilon_q = \varepsilon_p + \varepsilon_q,$$

avec $\varepsilon_n = |u_n - z|^2 - d^2 \longrightarrow 0$. La suite u_n est donc de Cauchy dans H complet, donc converge vers $u \in H$. Comme K est fermé, $u \in K$, et par continuité de la norme, $|u - z| = \text{dist}(z, K)$.

On écrit ensuite simplement que u réalise la distance si et seulement si, pour tout $v \in K$, l'inégalité $|z - w|^2 \geq |z - u|^2$ est vérifiée pour tout w du segment $[u, v]$. Cette propriété s'écrit

$$|z - (u + t(v - u))|^2 \geq |z - u|^2, \quad \text{i.e.} \quad -2t\langle z - u | v - u \rangle + t^2|v - u|^2 \geq 0 \quad \forall t \in [0, 1],$$

qui est équivalent à

$$\langle z - u | v - u \rangle \leq 0 \quad \forall v \in K,$$

soit l'égalité annoncée. \square

Remarque III.1.9. Si K est un sous-espace affine fermé de H , alors la caractérisation (III.1.1) prend la forme

$$\left\{ \begin{array}{l} u \in K \\ \langle z - u | v - u \rangle = 0 \quad \forall v \in K. \end{array} \right. \quad (\text{III.1.2})$$

Si K est un sous-espace vectoriel de H , on a

$$\left\{ \begin{array}{l} u \in K \\ \langle z - u | v \rangle = 0 \quad \forall v \in K. \end{array} \right. \quad (\text{III.1.3})$$

Remarque III.1.10. On prendra garde que la projection sur un sous-espace vectoriel n'est en général pas définie, car en dimension infinie les sous-espaces vectoriel peuvent ne pas être fermés (considérer par exemple le sous-espace de ℓ^2 des suites nulles au delà d'un certain rang).

Proposition III.1.11. L'application de projection P_K définie par le théorème précédent est 1-lipschitzienne, i.e.

$$|P_Kz - P_Kz'| \leq |z - z'| \quad \forall z, z' \in H.$$

Démonstration. Cette propriété est l'objet de l'exercice III.6.3, page 87). \square

Proposition III.1.12. (Caractérisation de la densité)

Soit H un espace de Hilbert et K un sous-espace de H tel que l'implication suivante soit vérifiée :

$$\langle h | w \rangle = 0 \quad \forall w \in K \implies h = 0.$$

Alors K est dense dans H

Démonstration: Si K n'est pas dense dans H , alors il existe $u \in H$, $u \notin \overline{K}$. On pose $h = u - P_{\overline{K}}u$. On a $\langle h | w \rangle = 0$ pour tout $w \in K$, et $h \neq 0$ car $u \notin \overline{K}$. \square

Théorème III.1.13. (Hahn-Banach)

Soit H un espace de Hilbert, $K \subset H$ un convexe fermé, et z un point de H qui n'appartient pas à K . Alors il existe un hyperplan fermé qui sépare K et z au sens strict, c'est-à-dire qu'il existe $h \in H$ et $\alpha \in \mathbb{R}$ tels que

$$\langle h | x \rangle \leq \alpha < \langle h | z \rangle \quad \forall x \in K.$$

Démonstration. On introduit la projection $u = P_Kz$ de z sur K , et l'on prend $h = z - u$ et $\alpha = \langle h | u \rangle$. Pour tout $x \in K$, on a

$$\langle h | x \rangle - \alpha = \langle h | x \rangle - \langle h | u \rangle = \langle z - u | x - u \rangle \leq 0.$$

et on a par ailleurs $\langle h | z \rangle - \alpha = \langle h | z \rangle - \langle h | u \rangle = |z - u|^2 > 0$. \square

Remarque III.1.14. Ce théorème est une version facile d'un théorème plus général et profond, appelé théorème de Hahn-Banach géométrique (théorème IV.1.8, page 93), qui exprime que tout convexe fermé d'un espace vectoriel normé peut être séparé de n'importe quel point qui lui est extérieur par un hyperplan fermé. La difficulté vient du fait que, en dehors du cas hilbertien, on ne peut pas en général définir de projection sur un convexe fermé (voir par exemple l'exercice II.2.1, page 48).

Définition III.1.15. (Orthogonal d'un ensemble)

Soit H un espace de Hilbert et K un sous-ensemble de H . On appelle orthogonal de K l'ensemble

$$K^\perp = \{v \in V, (v, u) = 0 \quad \forall u \in K\}.$$

On vérifie immédiatement que c'est un sous-espace vectoriel fermé.

Définition III.1.16. (Espace dual topologique)

Soit H un espace de Hilbert, on appelle *espace dual topologique* de H l'espace des formes linéaires (applications linéaires à valeurs dans \mathbb{R}) continues sur H . Pour $\varphi \in H'$, $u \in H$, on note

$$\langle \varphi, u \rangle \in \mathbb{R}$$

l'image de u par φ .

Tout espace de Hilbert peut s'identifier à son dual topologique, comme l'exprime le théorème suivant.

Théorème III.1.17. (Riesz-Fréchet)

Soit $\varphi \in H'$. Il existe $f \in H$ unique tel que

$$\langle \varphi, u \rangle = \langle f | u \rangle \quad \forall u \in H. \quad (\text{III.1.4})$$

De plus, on a

$$|f| = \|\varphi\|_{H'} = \sup_{v \in H, v \neq 0} \frac{\langle \varphi, v \rangle}{|v|}.$$

Démonstration. Si φ est la forme nulle, le résultat est immédiat. Dans le cas contraire, on introduit K le noyau de φ . C'est un hyperplan fermé de H . On construit ensuite un $h \in S_H \cap K^\perp$. Pour celà on considère $z \notin K$. D'après la caractérisation (III.1.3), on a $\langle z - P_K z | v \rangle = 0$ pour tout $v \in K$. Le vecteur

$$h = \frac{z - P_K z}{|z - P_K z|}$$

convient donc. Pour finir on remarque que tout $v \in H$ peut s'écrire

$$v = \underbrace{\frac{\langle \varphi, v \rangle}{\langle \varphi, h \rangle} h}_{\in K^\perp} + \underbrace{\left(v - \frac{\langle \varphi, v \rangle}{\langle \varphi, h \rangle} h \right)}_{\in K},$$

On a donc, pour tout $v \in H$ (on prend le produit scalaire de l'identité précédente avec h),

$$\langle \varphi, v \rangle = \langle \varphi, h \rangle \langle v | h \rangle$$

d'où l'identité (III.1.4) avec $f = \langle \varphi, h \rangle h$. L'unicité d'un tel f est immédiate.

On a enfin

$$\|\varphi\|_{H'} = \sup_{v \in H, v \neq 0} \frac{\langle \varphi, v \rangle}{|v|} = \sup_{v \in H, v \neq 0} \frac{\langle f | v \rangle}{|v|} = |f|.$$

□

On prendra garde au fait que cette identification dépend du produit scalaire choisi.

Proposition III.1.18. Soit H un espace de Hilbert de dimension infinie. La boule unité fermée de H n'est pas compacte.

Démonstration. Partant d'un élément unitaire quelconque, on construit une suite de vecteurs de norme 1 orthogonaux deux à deux selon le procédé d'orthonormalisation de Gram-Schmidt : Les vecteurs e_1, \dots, e_n étant construits, on prend un vecteur v qui n'est pas dans $F_n = \text{vect}(e_i)_{1 \leq i \leq n}$. On définit alors

$$e_{n+1} = \frac{v - P_{F_n}}{|v - P_{F_n}|},$$

qui est orthogonal à F_n (voir remarque III.1.9). On considère maintenant cette suite (e_n) de la boule unité fermée. On a $|e_p - e_q| = \sqrt{2}$ pour tous p, q distincts, il est donc impossible d'en extraire une sous-suite qui serait de Cauchy. □

Exercice III.1.2. Montrer que toute partie d'intérieur non vide d'un espace de Hilbert H de dimension infinie n'est pas compacte.

CORRECTION.

Si A est d'intérieur non vide, il contient une boule ouverte de rayon $r > 0$, donc une boule fermée de rayon $r/2$. Cette boule peut être mise en correspondance avec la boule unité fermée de H , par translation dilatation, on peut donc comme dans la proposition précédente construire une suite qui n'admette aucune sous-suite convergente.

III.2 Convergence faible

Comme précédemment H désigne un espace de Hilbert réel muni du produit scalaire (\cdot, \cdot) et de la norme associée $|\cdot|$.

Définition III.2.1. (Convergence faible)

Soit (u_n) une suite d'éléments de H . On dit que (u_n) converge faiblement vers u dans H , et on note $u_n \rightharpoonup u$, si

$$\langle u_n | v \rangle \rightarrow \langle u | v \rangle \quad \forall v \in H.$$

Proposition III.2.2. Si $u_n \rightharpoonup u$ et $|u_n| \rightarrow |u|$, alors la suite u_n converge fortement vers u .

Démonstration: On écrit

$$|u_n - u|^2 = |u_n|^2 - 2\langle u_n | u \rangle + |u|^2.$$

On a $(u_n, u) \rightarrow |u|^2$ d'où $|u_n - u|^2 \rightarrow 0$. □

Le résultat fondamental de cette section est le suivant.

Théorème III.2.3. Soit (u_n) une suite bornée dans un espace de Hilbert H . Alors on peut extraire une sous-suite convergeant faiblement vers u dans H .

Démonstration: On raisonne d'abord dans le cas où H est séparable. Il existe donc une famille dénombrable $\{x_k\}_{k \in \mathbb{N}}$ dense dans H . On se propose de suivre le procédé d'extraction diagonale de Cantor.

1. Comme $\langle u_n | x_1 \rangle$ est bornée dans \mathbb{R} on peut extraire une suite $u_{j_1(n)}$ telle que $\langle u_{j_1(n)} | x_1 \rangle$ converge.
2. Comme $\langle u_{j_1(n)} | x_2 \rangle$ est bornée dans \mathbb{R} on peut extraire de $u_{j_1(n)}$ une suite $u_{j_1 \circ j_2(n)}$ telle que $\langle u_{j_1 \circ j_2(n)} | x_2 \rangle$ converge.
3. Par récurrence, on construit une suite de sous-suites emboitées $u_{j_1 \circ j_2 \circ \dots \circ j_k(n)}$ telle que $\langle u_{j_1 \circ j_2 \circ \dots \circ j_k(n)} | x_k \rangle$ converge, pour tout k .
4. On utilise à présent le procédé d'extraction diagonale : on pose $\varphi(k) = j_1 \circ j_2 \circ \dots \circ j_k(k)$ (de telle sorte que φ est strictement croissante), et on considère $u_{\varphi(n)}$. Pour tout k , on remarque que $u_{\varphi(n)}$, à partir du rang k , est aussi une suite extraite de $(u_{j_1 \circ j_2 \circ \dots \circ j_k(n)})$, de telle sorte que $\langle u_{\varphi(n)} | x_k \rangle$ converge lorsque $n \rightarrow +\infty$.
5. On utilise ensuite la densité des x_k . Pour tout $x \in H$, on montre que $(u_{\varphi(n)}, x)$ est une suite de Cauchy : soit $\varepsilon > 0$, il existe (x_k) tel que $|x - x_k| < \varepsilon$. Comme $\langle u_{\varphi(n)} | x_k \rangle$ est de Cauchy, il existe un N au-delà duquel $|\langle u_{\varphi(p)} | x \rangle - \langle u_{\varphi(q)} | x \rangle| < \varepsilon$. Pour tous p, q supérieurs à N , on a donc

$$\begin{aligned} |\langle u_{\varphi(p)} | x \rangle - \langle u_{\varphi(q)} | x \rangle| &\leq |\langle u_{\varphi(p)} | x \rangle - \langle u_{\varphi(p)} | x_k \rangle| + |\langle u_{\varphi(p)} | x_k \rangle - \langle u_{\varphi(q)} | x_k \rangle| \\ &\quad + |\langle u_{\varphi(q)} | x_k \rangle - \langle u_{\varphi(q)} | x \rangle| \\ &\leq M\varepsilon + \varepsilon + M\varepsilon = (1 + 2M)\varepsilon, \end{aligned}$$

où M est un majorant de $|u_n|$.

On a donc démontré que, pour tout $x \in H$, $\langle u_{\varphi(n)} | x \rangle$ converge vers un élément de H que l'on note $h(x)$. L'application $x \mapsto h(x) \in \mathbb{R}$ est linéaire, et on a pour tout $x \in H$

$$|h(x)| = \lim_{n \rightarrow \infty} |\langle u_{\varphi(n)} | x \rangle| \leq M |x|,$$

d'où h continue³ sur H . D'après le théorème de Riesz-Fréchet, cette forme s'identifie à un élément u de H . On a donc convergence faible de la suite extraite vers u .

Dans le cas où le Hilbert n'est pas séparable, on se place dans l'adhérence de l'espace vectoriel engendré par les termes de la suite, qui est un espace de Hilbert séparable (pour le même produit scalaire) par construction. La convergence faible vers un u de ce sous-espace entraîne la convergence faible dans H .

III.3 Sommes hilbertiennes, bases hilbertiennes

Définition III.3.1. (Somme hilbertienne)

Soit $(E_n)_{n \in \mathbb{N}}$ une suite de sous-espaces fermés d'un espace de Hilbert H . On dit que H est somme Hilbertienne des E_n si

(i) Les E_n sont deux à deux orthogonaux, c'est-à-dire

$$\langle u, v \rangle = 0 \quad \forall u \in E_n, \quad \forall v \in E_m \quad \forall m, n \in \mathbb{N}, \quad m \neq n.$$

(ii) L'espace vectoriel engendré par les E_n est dense dans H .

Théorème III.3.2. On suppose que H est somme Hilbertienne des E_n . Pour $u \in H$, on note $u_n = P_{E_n} u$. On a

$$u = \sum_{i=1}^{\infty} u_n \text{ et } |u|^2 = \sum_{i=1}^{\infty} |u_n|^2.$$

Réciproquement, si l'on considère une suite (u_n) avec $u_n \in E_n$ pour tout n , et telle que $\sum |u_n|^2$ converge, alors la série $\sum u_n$ converge, et sa limite $u = \sum u_n$ est telle que $u_n = P_{E_n} u$.

Démonstration. On considère l'opérateur

$$S_k = \sum_{n=1}^k P_{E_n}.$$

On a $S_k \in \mathcal{L}(H)$, et $S_k u$ vérifie (les E_n sont orthogonaux deux à deux)

$$|S_k u|^2 = \sum_{n=1}^k |u_n|^2.$$

D'autre part on a, pour tout n

$$\langle u | u_n \rangle = \langle u_n + u - u_n | u_n \rangle = |u_n|^2,$$

d'où, en sommant de 1 à k ,

$$\langle u | S_k u \rangle = \sum_{n=1}^k |u_n|^2 = |S_k u|^2.$$

On a donc $|S_k u| \leq |u|$. On désigne par E l'espace vectoriel engendré par les E_n . Pour tout $\varepsilon > 0$, tout u dans H , il existe un $v \in E$ tel que $|v - u| < \varepsilon$. Pour k assez grand, on a $S_k v = v$, et ainsi

$$|S_k u - u| \leq |S_k(u - v)| + |v - u| \leq 2\varepsilon.$$

3. Remarquer qu'il n'est pas nécessaire ici d'utiliser le théorème de Banach–Steinhaus, du fait de l'hypothèse (u_n) bornée.

on a donc bien convergence de $S_k u$ vers u .

D'autre part l'égalité, pour tout k

$$|S_k u|^2 = \sum_{n=1}^k |u_n|^2,$$

entraîne, à la limite,

$$|u|^2 = \sum_{n=1}^{+\infty} |u_n|^2.$$

Pour la réciproque, on utilise le caractère de Cauchy de la suite $\sum_{n=1}^k u_n$, et la continuité des opérateurs de projection. \square

Le théorème précédent permet d'introduire la notion de base Hilbertienne :

Définition III.3.3. (Bases hilbertiennes)

Soit $(e_n)_{n \in \mathbb{N}}$ une famille de vecteurs d'un espace de Hilbert H . On dit que (e_n) est une base Hilbertienne si

- (i) $|e_n| = 1$ pour tout $n \in \mathbb{N}$, et $\langle e_m, e_n \rangle = 0$ pour tous m, n , avec $m \neq n$.
- (ii) L'espace vectoriel engendré par les (e_n) est dense dans H .

La définition dit exactement que H est somme hilbertienne des droites vectorielles $\mathbb{R}e_n$.

Théorème III.3.4. Tout espace de Hilbert séparable admet une base Hilbertienne.

Démonstration. Soit H un espace de Hilbert séparable⁴. On considère $(f_n)_{n \in \mathbb{N}}$ une famille dense dans H . On note F_k l'espace vectoriel engendré par les k premiers vecteurs. L'espace vectoriel engendré par les F_k est dense dans H . On peut construire la base Hilbertienne de la façon suivante : si f_1 est non nul, on prend $e_1 = f_1 / |f_1|$ comme premier vecteur. Une base orthonormale sur F_k étant construite, de dimension $n_k \leq k$, on complète, par procédé d'orthonormalisation de Gram-Schmidt : si $f_{k+1} \notin F_k$, on définit

$$e_{n_{k+1}} = \frac{f_{k+1} - P_{F_k} f_{k+1}}{|f_{k+1} - P_{F_k} f_{k+1}|}.$$

On construit ainsi une base Hilbertienne $(e_{n_k})_k$ de H . \square

Remarque III.3.5. L'existence d'une base hilbertienne sur un espace de Hilbert H séparable permet de construire une isométrie entre l'espace H et ℓ^2 (on associe simplement à tout élément u la suite de ses coefficients dans la base hilbertienne). La plupart des espaces fonctionnels que l'on utilise en pratique, en particulier dans l'étude des EDP, étant séparables, on pourrait penser que tous ces espaces se ramènent à l'espace *modèle* ℓ^2 . Cette isométrie est en effet très féconde dans certains cas, comme par exemple pour l'étude de $L^2(0, L)$, qui admet comme base hilbertienne la collection des

$$w_n : x \mapsto \sqrt{\frac{2}{L}} \sin\left(\frac{\pi n x}{L}\right),$$

cette approche spectrale (on parle de séries de Fourier dans ce contexte) est de fait très utile en traitement du signal, et dans l'étude de certaines EDP (équation de la chaleur, équation des ondes, ...), plus précisément pour l'étude de régularité des solutions, ou de leur comportement en temps long. Mais, plus généralement, la base n'est pas en général connue explicitement⁵, ce qui exclut toute possibilité de calcul effectif. Par ailleurs, même si cette base est connue explicitement comme dans le cas ci-dessus, elle peut être très inadaptée à la formalisation de certains problèmes. Par exemple décrire le cône des fonctions de $L^2(0, 1)$ qui sont positives ou nulles presque partout est très incommode si l'on représente les fonctions par leurs coefficients de Fourier.

4. C'est à dire qu'il existe un ensemble dénombrable et dense. C'est le cas pour l'essentiel des espaces de Hilbert que l'on rencontre dans la "nature", en particulier pour les espaces fonctionnels de type $L^2(\Omega)$ ou $H^m(\Omega)$.

La notion de base hilbertienne permet d'exprimer la convergence faible d'une suite composante par composante, sous réserve que la suite soit *bornée*.

La proposition ci-dessous propose une version faible de la propriété suivante des espaces euclidiens : si une suite de vecteurs de \mathbb{R}^d , exprimés dans la base canonique, est telle que chaque composante converge, alors la suite converge. Dans le cas d'un espace de Hilbert de dimension infinie, si une suite, exprimée comme combinaison infinie de vecteurs d'une base hilbertienne, est *bornée* et converge *composante par composante*, alors on a *convergence faible* de la suite.

Proposition III.3.6. (••) Soit H un espace de Hilbert séparable, et (e_n) une base hilbertienne de H . Soit (u^k) une suite d'éléments de H , bornée, telle que, pour tout n

$$\lim_{k \rightarrow +\infty} \langle u^k | e_n \rangle = u_n \in \mathbb{R}.$$

Alors la suite (u_n) est dans ℓ^2 , et la suite (u^k) converge faiblement vers

$$u = \sum_{n=0}^{+\infty} u_n e_n.$$

Démonstration. On écrit $u^k = \sum u_n^k e_n$. Comme u^k est bornée par un certain $M > 0$, on a, pour tout N ,

$$\sum_{n=0}^N |u_n^k|^2 \leq \sum_{n=0}^{+\infty} |u_n^k|^2 \leq M.$$

La convergence terme à terme dans la somme finie de droite assure que les u_n vérifient la même inégalité, pour tout N , d'où $u \in \ell^2$.

Montrons que u^k converge faiblement vers u . On considère d'abord le cas $u = 0$. Pour tout vecteur-test $v = \sum v_n e_n \in H$, on a, pour tout N ,

$$|\langle u^k | v \rangle| = \left| \sum_{n=0}^{+\infty} u_n^k v_n \right| \leq \left| \sum_{n=0}^{N-1} u_n^k v_n \right| + \left| \sum_{n=N}^{+\infty} u_n^k v_n \right|.$$

On a

$$\left| \sum_{n=N}^{+\infty} u_n^k v_n \right|^2 \leq \left(\sum_{n=N}^{+\infty} |u_n^k|^2 \right) \left(\sum_{n=N}^{+\infty} |v_n|^2 \right) \leq M \sum_{n=N}^{+\infty} |v_n|^2.$$

Pour tout $\varepsilon > 0$, le dernier terme peut être rendu inférieur à ε^2 pour N assez grand. Ce N étant maintenant fixé, la somme finie des $u_n^k v_n$ peut être rendue inférieure à ε , par convergence vers 0 des u_n^k quand k tend vers $+\infty$. On a donc bien convergence vers 0 de $\langle u^k | v \rangle$ pour tout v , qui exprime la convergence faible vers 0. Si u est non nul, on applique simplement ce qui précède à la suite $(u^k - u)$. \square

Remarque III.3.7. Le caractère borné de la suite est essentiel. On peut en fait montrer que l'on a équivalence entre convergence faible vers une limite, et caractère borné + convergence composante par composante. L'argument essentiel est le suivant : toute suite faiblement convergente dans un espace de Hilbert (la complétude est importante) est *nécessairement bornée*. C'est une conséquence du théorème (hors programme) de Banach-Steinhaus, théorème VI.7.5, page 165, détaillé en annexe).

III.4 Théorie de Lax-Milgram

Proposition III.4.1. (Continuité d'une forme bilinéaire)

Soit $a : H \times H \rightarrow \mathbb{R}$ une forme bilinéaire. Alors $a(\cdot, \cdot)$ est continue si et seulement s'il existe une constante $\|a\|$ telle que

$$|a(u, v)| \leq \|a\| |u| |v| \quad \forall u, v \in H.$$

5. Si l'on considère par exemple un domaine de \mathbb{R}^d de forme quelconque, on peut montrer que l'opérateur du laplacien avec conditions nulles au bord admet une collection de fonctions propres qui forment une base hilbertienne de L^2 . On peut approcher numériquement certaines de ces fonctions propres, et étudier leurs propriétés asymptotiques, mais elles ne sont en général pas connues sous forme analytique, contrairement au cas de la dimension 1.

Démonstration. On suppose a continue. La continuité en 0 assure l'existence d'un r tel que $|a(u, v)| \leq 1$ sur $\overline{B(0, r)} \times \overline{B(0, r)}$. On a donc, pour tous u, v , non nuls

$$\left| a\left(r \frac{u}{|u|}, r \frac{v}{|v|}\right) \right| \leq 1 \implies |a(u, v)| \leq \frac{1}{r^2} |u| |v|.$$

Réciproquement, le développement

$$a(u + h, v + k) = a(u, v) + a(h, v) + a(u, k) + a(h, k)$$

assure la continuité en tout $(u, v) \in H \times H$. \square

Définition III.4.2. (Coercivité d'une forme bilinéaire)

Soit $a : H \times H \rightarrow \mathbb{R}$ une forme bilinéaire. On dit que a est coercive s'il existe $\alpha > 0$ tel que

$$a(u, u) \geq \alpha |u|^2 \quad \forall u \in H.$$

Remarque III.4.3. En dimension finie, et dans le cas où la forme est symétrique ($a(u, v) = a(v, u)$), on retrouve la notion de forme symétrique définie positive. Le plus grand coefficient α est alors la plus petite valeur propre de la matrice associée, et la plus petite constante $\|a\|$ de la continuité sa plus grande valeur propre.

Exercice III.4.1. Soit $\alpha = (\alpha_n)$ une suite bornée de réels, et

$$a : (u, v) \in \ell^2 \times \ell^2 \mapsto \sum_{n=0}^{+\infty} \alpha_n u_n v_n.$$

- a) A quelles conditions sur α la forme bilinéaire $a(\cdot, \cdot)$ est-elle coercive dans ℓ^2 ?
- b) On suppose ici $\alpha = 1/2^n$. Donner un exemple d'espace de Hilbert dans lequel cette forme est bien définie, et coercive.

CORRECTION.

- a) La forme bilinéaire est coercive si et seulement si les α_n sont minorés par une constante α strictement positive. En effet, si c'est le cas, ou a de façon évidente $a(u, u) \geq \alpha |u|^2$. Si ça n'est pas le cas, et qu'il y a un coefficient k négatif ou nul, la coercivité est invalidée sur $e_k = (0, \dots, 0, 1, 0, \dots)$. Et s'il y a une sous-suite $u_{\varphi(k)}$ qui tend vers 0, alors la coercivité est invalidée sur la suite des éléments $e_{\varphi(k)}$.
- b) Si $\alpha = 1/2^n$, la forme n'est pas coercive sur ℓ^2 d'après ce qui précède. Mais si l'on considère l'espace ℓ^2 à poids

$$H = \left\{ u \in \mathbb{R}^{\mathbb{N}}, \sum \frac{|u_n|^2}{2^n} < \infty \right\},$$

muni du produit scalaire associé à $a(\cdot, \cdot)$, alors cette forme est le produit scalaire lui-même, elle est donc trivialement coercive.

Proposition III.4.4. Soit H un espace de Hilbert, et a une forme bilinéaire et continue sur l'espace produit $H \times H$. Pour tout $u \in H$, on note Au l'élément de H qui s'identifie à la forme linéaire $a(u, \cdot)$, définit par :

$$\langle Au | v \rangle = a(u, v) \quad \forall v \in H.$$

L'application $u \mapsto Au$ est linéaire et continue. De plus si $a(\cdot, \cdot)$ est coercive, alors l'application A est une bijection.

Démonstration: L'application A est évidemment linéaire, et

$$|Au| = \sup_{|v|=1} \langle Au | v \rangle = \sup_{|v|=1} a(u, v) \leq \|a\| |u|,$$

où $\|a\|$ est la constante de continuité de a .

Si $a(\cdot, \cdot)$ est coercive, on a $\langle Au | u \rangle = a(u, u) \geq \alpha |u|^2$, et donc $|Au| \geq \alpha |u|$ pour tout u dans H . L'application linéaire A est donc injective. On vérifie que l'image est fermée en considérant une suite (Au_n) qui converge vers un élément de l'image w . Comme (Au_n) converge, elle est de Cauchy, donc (u_n) est également de Cauchy d'après l'inégalité précédemment démontrée. Elle converge donc vers $u \in H$ qui vérifie $Au = w$ par continuité de A . On a de plus, pour tout $g \in H$,

$$\langle g | Au \rangle = 0 \quad \forall u \in H \implies \langle g | Ag \rangle = a(g, g) = 0$$

qui entraîne $g = 0$ par coercivité de $a(\cdot, \cdot)$. L'image de A est donc fermée et dense dans H : c'est l'espace H lui-même. L'injectivité est une conséquence immédiate de la coercivité. \square

Théorème III.4.5. (Lax-Milgram)

Soit H un espace de Hilbert, et a une forme bilinéaire continue et coercive sur $H \times H$. Pour tout $\varphi \in H'$, il existe un $u \in H$ unique tel que

$$a(u, v) = \langle \varphi, v \rangle \quad \forall v \in H. \quad (\text{III.4.1})$$

Si a est symétrique, u est l'unique élément de H qui réalise le minimum de la fonctionnelle

$$v \mapsto J(v) = \frac{1}{2}a(v, v) - \langle \varphi, v \rangle.$$

Démonstration. D'après le théorème de représentation de Riesz-Fréchet, il existe un unique $f \in H$ tel que

$$\langle f | v \rangle = \langle \varphi, v \rangle \quad \forall v \in H.$$

On introduit l'opérateur A associé à $a(\cdot, \cdot)$, qui est bijectif (voir proposition III.4.4). Il existe donc une unique solution u à l'équation $Au = f$.

On suppose maintenant $a(\cdot, \cdot)$ symétrique. On note toujours u la solution du problème variationnel (III.4.1). Pour tout $h \in H$, la fonction (de \mathbb{R} dans \mathbb{R})

$$t \mapsto \psi(t) = J(u + th) - J(u)$$

est convexe, nulle en 0, de dérivée nulle en 0. Elle est donc positive, et ainsi $J(u + h) \geq J(u)$ pour tout $h \in H$.

De la même manière, si w minimise J , on écrit que la dérivée de la fonction $J(w + th) - J(w)$ est nulle en 0, ce qui est exactement la formulation variationnelle (III.4.1). \square

III.5 Opérateurs

Proposition III.5.1. Soit T une application linéaire de H dans F , deux espaces de Hilbert. Alors F est continue si et seulement si elle est bornée sur la boule unité de H . On note $\mathcal{L}(H, F)$ l'espace vectoriel des applications linéaires continues de H dans F . C'est un e.v.n. pour la norme d'opérateur

$$\|T\| = \sup_{x \neq 0} \frac{|Tx|_F}{|x|_H}.$$

L'espace $\mathcal{L}(E, F)$ est un espace de Banach, i.e. un espace vectoriel normé complet.

Démonstration. Soit (T_n) une suite de Cauchy dans $\mathcal{L}(H, F)$. Pour tout $x \in H$, la suite $(T_n x)$ est de Cauchy dans H complet, elle converge donc vers un élément de F que l'on note Tx . On vérifie immédiatement T est linéaire. Pour la continuité, on utilise le caractère de Cauchy de la suite : pour tout $\varepsilon > 0$ il existe N tel que, pour tous $p, q \geq N$,

$$\|T_q - T_p\| < \varepsilon \text{ i.e. } |T_q x - T_p x| \leq \varepsilon |x| \quad \forall x \in H.$$

Pour x fixé on fait tendre q vers l'infini, et on prend $p = N$. On obtient

$$|Tx| \leq (\varepsilon + |T_N|) |x|.$$

d'où l'on déduit que T est continue. La convergence de T_n vers T pour la norme d'opérateur s'obtient à partir du critère de Cauchy qui précède, en faisant tendre comme précédemment q vers $+\infty$: pour tout $p \geq N$, tout $x \neq 0$

$$\frac{|T_p x - Tx|}{|x|} \leq \varepsilon,$$

qui termine la preuve. \square

Définition III.5.2. (Opérateur compact)

Soit T un opérateur linéaire continu de H vers F , deux espaces de Hilbert. On dit que T est compact si l'image de B_H , la boule unité fermée de H , est relativement compacte dans F , c'est à dire d'adhérence compacte.

Exercice III.5.1. Montrer que $T \in \mathcal{L}(H, F)$ est compact si et seulement si, pour toute suite bornée (x_n) dans H , on peut extraire une sous-suite $(x_{\varphi(n)})$ telle que $T(x_{\varphi(n)})$ converge dans F .

Proposition III.5.3. L'ensemble $\mathcal{K}(H, F)$ des applications linéaires compactes est un s.e.v. fermé de $\mathcal{L}(H, F)$

Démonstration. La somme de deux opérateurs compacts est compacte, ainsi que le produit d'un opérateur compact par un réel. Maintenant considérons une suite T_n d'opérateurs compacts qui converge vers un opérateur $T \in \mathcal{L}(H, F)$. Pour tout $\varepsilon > 0$, il existe n tel que $\|T_n - T\| < \varepsilon/2$. On peut recouvrir $T_n(B_H)$ par une union finie de boules de rayon $\varepsilon/2$. L'union des boules de mêmes centres et de rayon ε recouvre $T(B_H)$ par construction. L'image de la boule unité fermé par T est donc précompacte, donc relativement compacte d'après la proposition I.6.8, page 27. \square

Définition III.5.4. (Opérateur de rang fini)

Soit $T \in \mathcal{L}(H, F)$. On dit que T est de rang fini si son image $R(T)$ est de dimension finie.

Tout opérateur de rang fini étant compact (du fait que tout borné dans un espace de vectoriel de dimension finie est relativement compact), et du fait que $\mathcal{K}(H, F)$ est fermé (proposition III.5.3), toute limite d'une suite d'opérateur de rang fini est compacte. Il s'agit, dans le contexte bien spécifique des espaces de Hilbert, d'une équivalence, comme le précise la proposition suivante.

Proposition III.5.5. Un opérateur $T \in \mathcal{L}(H, F)$ est compact si et seulement s'il est limite dans $\mathcal{L}(H, F)$ d'opérateurs de rang fini.

Démonstration. On a vu que toute limite d'une suite (T_n) d'opérateurs de rang fini est compacte. Réciproquement, considérons un opérateur T compact. Comme $\overline{T(B_H)}$ est compact, pour tout $\varepsilon > 0$, on peut le recouvrir par une collection finie de boules de rayon ε :

$$\overline{T(B_H)} \subset \bigcup_{i=1}^n B(f_i, \varepsilon).$$

On considère maintenant l'espace vectoriel G engendré par les f_i . Pour tout $x \in B_F$, il existe i tel que

$$|Tx - f_i| < \varepsilon$$

d'où (d'après la proposition III.1.11, page 78)

$$|P_G(Tx) - f_i| = |P_G(Tx) - P_G f_i| \leq |Tx - f_i| < \varepsilon.$$

On a donc

$$|P_G \circ Tx - Tx| \leq |P_G \circ Tx - f_i| + |f_i - Tx| < 2\varepsilon \quad \forall x \in B_H,$$

d'où, par définition de la norme d'opérateur,

$$\|T_\varepsilon x - Tx\| < 2\varepsilon,$$

où $T_\varepsilon = P_G \circ T$ est de rang fini. \square

III.6 Exercices

Exercice III.6.1. Soit H un espace de Hilbert, K un convexe fermé non vide de H , $z \in H$. On note u la projection de z sur K . Montrer que

$$|v - u| \leq |v - z| \quad \forall v \in K.$$

CORRECTION.

On écrit simplement

$$|v - z|^2 = |v - u - (z - u)|^2 = |v - u|^2 + |z - u|^2 - (v - u, z - u) \geq |v - u|^2 + |z - u|^2,$$

d'où $|v - u|^2 \leq |v - z|^2 - |z - u|^2$, et par suite $|v - u| \leq |v - z|$.

Exercice III.6.2. La démonstration du théorème III.1.8 est basée sur le fait qu'une suite (v_n) minimisante de la distance de z à un convexe fermé K converge vers un point de K , que l'on définit comme la projection de z sur K . Cet exercice vise à estimer la vitesse de convergence de v_n vers $P_K z$ en fonction de la vitesse de convergence de $|v_n - z|$ vers $d(z, K)$.

Soit H un espace de Hilbert, K un convexe fermé non vide de H , $z \in H$. On note u la projection de z sur K . Pour tout $v \in K$, note $d_v = |v - z|$.

a) Montrer que

$$|u - v| \leq \sqrt{2} \sqrt{d_v - d} \sqrt{d_v}.$$

b) Montrer que cette estimation est optimale dès que la dimension de H est supérieure ou égale à 2.

CORRECTION.

a) On a

$$|z - v|^2 = |z - u + u - v|^2 = |z - u|^2 + |u - v|^2 + \underbrace{2\langle z - u | u - v \rangle}_{\geq 0} \geq |z - u|^2 + |u - v|^2.$$

On a donc

$$|u - v|^2 \leq d_v^2 - d^2 = (d_v - d)(d_v + d) \leq |d_v - d| |d_v + d| \leq 2 |d_v - d| |d_v + d|.$$

b) On se place dans \mathbb{R}^2 , avec K le demi-plan des points à ordonnées négatives ou nulles, $z = (0, 1)$, et $v \in K$ sur l'axe des x , tend vers $0 = P_K z$. La première inégalité utilisée est une égalité, et on a par ailleurs $d_v \sim d$ quand v tends vers 0, on ne peut donc avoir de meilleure inégalité que celle là.

Exercice III.6.3. (Caractère 1-Lipschitzien de la projection sur un convexe fermé)

Soit $K \subset H$ un convexe fermé. On note P_K l'application de projection définie par le théorème III.1.8, page 77. Montrer que, pour tous $f, g \in H$, on a

$$|P_K f - P_K g| \leq |f - g|.$$

CORRECTION.

On utilise la caractérisation de la projection (III.1.1) :

$$\begin{aligned} \langle f - P_K f | P_K g - P_K f \rangle &\leq 0, \\ \langle g - P_K g | P_K f - P_K g \rangle &\leq 0. \end{aligned}$$

En additionnant les deux inégalités, il vient

$$\langle g - f + P_K f - P_K g \mid P_K f - P_K g \rangle \leq 0,$$

d'où

$$|P_K f - P_K g|^2 \leq (f - g, P_K f - P_K g) \leq |f - g| |P_K f - P_K g|,$$

qui est l'inégalité annoncée.

Remarque III.6.1. Ne pas confondre le résultat précédent avec le caractère 1-lipschitzien de la fonction distance à un ensemble quelconque, dans tout espace métrique (voir exercice I.7.3, page 29).

Exercice III.6.4. a) Montrer que, si l'espace de Hilbert H est de dimension finie (on parle alors d'espace euclidien), alors pour tout fermé K non vide (sans hypothèse de convexité), la distance de tout point z à K est atteinte, mais que l'on peut perdre l'unicité.

b) (• • •) Montrer que, si H est de dimension infinie (on pourra se placer sur ℓ^2), alors la distance d'un point à un fermé peut ne pas être réalisée.

CORRECTION.

a) Soit K un fermé K non vide d'un espace euclidien H , et $z \notin K$. Soit v un point de K . La distance de z à K est la distance de z à $K \cap \overline{B}(z, |z - v|)$. Ce dernier ensemble est un fermé borné, il est donc compact. L'application $v \mapsto |z - v|$, continue sur le compact K , atteint donc son minimum : la distance est donc réalisée.

b) On se place sur $H = \ell^2$, et l'on considère l'ensemble K constitué des suites dont le premier terme (indice 0) vaut $1 - 1/n$, le n -ème terme vaut 1, et tous les autres sont nuls. Toute suite convergente de points de K est de Cauchy, donc stationnaire au delà d'un certain rang, l'ensemble est donc fermé. La distance du point $z = (1, 0, 0, \dots)$ à K est égale à 1, mais elle n'est pas atteinte.

Exercice III.6.5. Soient u, u_1, \dots, u_n , des éléments d'un espace de Hilbert H . Montrer l'équivalence suivante

$$\left(\bigcap_{i=1}^n u_i^\perp \right) \subset u^\perp \iff \exists \lambda_1, \dots, \lambda_n, u = \sum \lambda_i u_i.$$

CORRECTION.

La condition nécessaire est immédiate. Pour la condition suffisante, on note $u = u_0$, et l'on considère l'application

$$T : v \in H \mapsto (\langle u_0 \mid v \rangle, \langle u_1 \mid v \rangle, \dots, \langle u_n \mid v \rangle) \in \mathbb{R}^{n+1}.$$

Comme un vecteur orthogonal aux u_i , pour $i \geq 1$, est aussi orthogonal à u_0 par hypothèse, le vecteur $w = (1, 0, \dots, 0)$ n'est pas dans l'image de T . Cette dernière étant fermée dans \mathbb{R}^{n+1} comme espace vectoriel de dimension finie, et convexe, on peut séparer strictement w et l'image par un hyperplan d'après le théorème de Hahn Banach (théorème III.1.13, page 78) : il existe $h \in H$ et $\alpha \in \mathbb{R}$ tels que

$$\sum_{i=0}^n h_i \langle u_i \mid v \rangle \leq \alpha \leq h_0 \quad \forall v \in H.$$

On peut prendre $v = 0$ dans l'inégalité de droite, pour obtenir $\alpha \geq 0$. On a donc $h_0 > 0$. Par ailleurs le fait que la somme du membre de gauche, qui est linéaire par rapport à v , soit majorée, impose qu'elle est identiquement nulle. On a donc

$$\langle u_0 \mid v \rangle = -\frac{1}{h_0} \sum_{i=1}^n h_i \langle u_i \mid v \rangle \quad \forall v \in H,$$

et donc

$$u_0 = -\frac{1}{h_0} \sum_{i=1}^n h_i u_i,$$

d'où l'expression annoncée, avec $\lambda_i = -h_i/h_0$.

Exercice III.6.6. Soit (e_j) une base hilbertienne d'un espace de Hilbert H .

a) Montrer que la suite (e_n) tend faiblement vers 0.

Soit (a_j) une suite réelle bornée. On pose

$$u_n = \frac{1}{n} \sum_{j=1}^n a_j e_j.$$

b) Montrer que $|u_n|$ tend vers 0.

c) Montrer que $\sqrt{n} u_n \rightarrow 0$.

CORRECTION.

a) Tout élément v de H s'écrit $v = \sum v_j e_j$, avec $(v_j) \in \ell^2$, on a donc

$$\langle v | e_n \rangle = v_n \rightarrow 0.$$

b) On a

$$|u_n|^2 = \left| \frac{1}{n} \sum_{j=1}^n a_j e_j \right|^2 = \frac{1}{n^2} \sum_{j=1}^n |a_j|^2 \leq \frac{1}{n} \max_j |a_j|^2 \rightarrow 0.$$

c) La suite $(\sqrt{n} u_n)$ est bornée d'après ce qui précède. Par ailleurs, pour tout $m \in \mathbb{N}$, on a (pour $n \geq m$)

$$\langle \sqrt{n} u_n | e_m \rangle = \frac{1}{\sqrt{n}} a_m,$$

qui tend vers 0 quand m tend vers $+\infty$. La proposition III.3.6, page 83, assure donc la convergence de $\sqrt{n} u_n$ vers 0.

Exercice III.6.7. (Ondelettes de Haar)

On se place sur $H = L^2(0, 1)$, et l'on introduit la fonction de \mathbb{R} dans \mathbb{R}

$$W = \mathbf{1}_{]0, 1/2[} - \mathbf{1}_{]1/2, 1[}.$$

On considère maintenant les fonctions w_n^k définies par

$$w_n^k(x) = 2^{n/2} W(2^n(x - k/2^n)), \quad n \in \mathbb{N}, \quad 0 \leq k < 2^n.$$

Montrer que le système constitué de la fonction $w_0 \equiv 1$ et des fonctions précédemment définies constitue une base hilbertienne de H .

CORRECTION.

En premier lieu, le coefficient $2^{n/2}$ assure que ces fonctions sont de norme 1. En second lieu, si l'on considère deux fonctions distinctes du système, alors ou bien leurs supports sont disjoints, et le produit scalaire ou bien celle des deux associés à l'indice n le plus bas des 2 est constante sur le support de l'autre. Comme cette dernière est de moyenne nulle, le produit scalaire est aussi nul dans ce cas. Maintenant remarquons que l'espace vectoriel engendré par w_0 , w_0^0 , contient les fonctions caractéristiques de $]0, 1/2[$ et de $]1/2, 1[$. De la même manière, l'espace vectoriel engendré par w_0 et les w_n^k pour $n \leq N$ contient toutes les fonctions caractéristiques d'intervalles du type

$$\left] \frac{k}{2^{N+1}}, \frac{k+1}{2^{N+1}} \right[.$$

On utilise maintenant la densité des fonctions continues sur $[0, 1]$ dans $L^2(0, 1)$ (proposition II.4.11, page 65). Soit $f \in H$. Pour tout ε , on peut approcher (en norme L^2) à ε près f par une fonction g continue. Cette fonction continue est uniformément continue sur le compact $[0, 1]$ (théorème de Heine I.7.7, page 29), il existe donc $\eta > 0$ tel que $|g(x) - g(y)| < \varepsilon$ pour $|x - y| < \eta$. On choisit maintenant N tel que $1/2^{N+1} < \eta$.

On peut approcher g en norme ∞ par une combinaison de fonctions caractéristiques d'intervalles du type indiqué ci-dessus. Cette approximation reste inférieure à ε en norme L^2 . On a donc approché f par une combinaison linéaire de fonctions du système de Haar. L'espace vectoriel engendré par la famille est donc bien dense dans H .

Exercice III.6.8. Donner un exemple de forme linéaire non continue sur un espace vectoriel normé dont la norme est issue d'un produit scalaire. (*On pour considérer l'espace F des suites nulles au-delà d'un certain rang muni de la norme euclidienne*).

CORRECTION.

Noter que F est l'espace engendré par les vecteurs de la base hilbertienne canonique de ℓ^2 . Il suffit de considérer

$$\varphi : u = (u_n) \in F \mapsto \sum u_n$$

(somme finie). Si l'on note w^N la suite dont les N premiers termes sont égaux à 1, les autres à 0, on a

$$\frac{\langle \varphi, w_N \rangle}{|w_N|} = \sqrt{N},$$

qui tend vers $+\infty$ quand N tend vers $+\infty$.

Remarque : dans un espace de Hilbert, il est nécessaire d'utiliser l'axiome du choix pour montrer l'existence d'une forme linéaire non continue.

Exercice III.6.9. Soit H un espace de Hilbert, φ une forme linéaire sur H , et $K = \ker \varphi$ son noyau.

- a) Montrer que, si φ est continue et non nulle, alors K sépare H en deux composantes connexes par arcs.
- b) Montrer que, si φ n'est pas continue, alors $H \setminus K$ est connexe par arcs (!).

CORRECTION.

XXXXXX

Chapitre IV

Éléments d'analyse fonctionnelle

Sommaire

IV.1 Définitions	91
IV.2 Compléments sur la dualité	93

Ce chapitre propose une introduction aux espaces vectoriels normés de dimension infinie, i.e. qui ne sont pas engendrés par un nombre fini d'éléments. On se gardera pourtant de dire qu'ils sont engendrés par nombre infini d'éléments car les bases vectorielles (dont on ne peut en général montrer l'existence que par un argument fondé sur l'axiome du choix) sont complètement *inutilisables* en général. La définition d'un espace vectoriel est en revanche exactement la même qu'en dimension finie, ainsi que la définition d'une norme (voir définition I.2.7, page 13). Ce cadre abstrait est adapté aux espaces de fonctions (espaces de fonctions continues, différentiables, espaces L^p , espaces de Sobolev, ...), les “points” de ces espaces sont alors des fonctions, ce qui justifie l'appellation *Analyse Fonctionnelle*.

IV.1 Définitions

Définition IV.1.1. (Espace de Banach)

Soit E un espace vectoriel normé (e.v.n.). Si E est complet pour la distance associée à la norme, on dit que E est un espace de Banach.

On utilisera en général ce terme pour des espaces de dimension infinie, même si de fait tout e.v.n. de dimension finie est un espace de Banach (voir proposition I.5.5, page 23).

Exercice IV.1.1. Montrer que l'espace vectoriel X des suites nulles au-delà d'un certain rang, muni de la norme ∞ , n'est pas complet.

CORRECTION.

La suite (de suites) (u^n) définie par $u_k^n = 1/k$ si $k \leq n$, 0 sinon, est de Cauchy. Si elle convergeait vers (v_k) on aurait nécessairement $v_k = 1/k$ pour tout k , or cette suite n'est pas dans X .

Proposition IV.1.2. Soit T une application linéaire de E dans F , deux espaces vectoriels normés. Alors T est continue si et seulement si elle est bornée sur la boule unité de E . On note $\mathcal{L}(E, F)$ l'espace vectoriel des applications linéaires continues de E dans F . C'est un e.v.n. pour la norme d'opérateur

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|_F}{\|x\|_E}.$$

Si F est complet, alors $\mathcal{L}(E, F)$ est un espace de Banach.

Démonstration. La démonstration est identique à celle de la proposition III.5.1, page 85. Soit (T_n) une suite de Cauchy dans $\mathcal{L}(E, F)$. Pour tout $x \in E$, la suite $(T_n x)$ est de Cauchy dans F complet, elle converge donc vers un élément de F que l'on note Tx . On vérifie immédiatement T est linéaire. Pour la continuité, on utilise le caractère de Cauchy de la suite : pour tout $\varepsilon > 0$ il existe N tel que, pour tous $p, q \geq N$,

$$\|T_q - T_p\| < \varepsilon \text{ i.e. } \|T_q x - T_p x\| < \varepsilon \|x\| \quad \forall x \in E.$$

Pour x fixé on fait tendre q vers l'infini, et on prend $p = N$. On obtient

$$\|Tx\| \leq (\varepsilon + \|T_N\|) \|x\|.$$

d'où l'on déduit que T est continue. La convergence de T_n vers T pour la norme d'opérateur s'obtient à partir du critère de Cauchy qui précède, en faisant tendre comme précédemment q vers $+\infty$: pour tout $p \geq N$, tout $x \neq 0$

$$\frac{\|T_p x - Tx\|}{\|x\|} \leq \varepsilon.$$

□

Définition IV.1.3. (Dual topologique)

Soit E un espace vectoriel normé (e.v.n.). On appelle dual topologique de E , et l'on note E' , l'ensemble des formes linéaires (applications linéaires à valeurs dans \mathbb{R}) continues. C'est un e.v.n. pour la norme

$$\|\varphi\|_{E'} = \sup_{x \neq 0} \frac{|\langle \varphi, x \rangle|}{\|x\|_E} = \sup_{\|x\| \leq 1} \langle \varphi, x \rangle.$$

Remarque IV.1.4. Si E est de dimension finie, toutes les formes linéaires sont continues. Ca n'est pas le cas en dimension infinie. Considérer par exemple l'espace vectoriel X des suites nulles au-delà d'un certain rang, muni de la norme $\|\cdot\|_\infty$. La forme

$$\varphi : u = (u_n) \longmapsto \sum_{n=1}^{+\infty} u_n \in \mathbb{R}$$

(on gardera en tête qu'il s'agit en fait d'une somme *finie*, donc convergente) n'est pas bornée sur la boule unité, elle n'est donc pas continue¹.

Définition IV.1.5. (Convergences faible / faible- \star)

Soit E un e.v.n. On dit qu'une suite (x_n) de points de E converge faiblement vers x , ce qu'on écrira $x_n \rightharpoonup x$, si

$$\lim_{n \rightarrow +\infty} \langle \varphi, x_n \rangle = \langle \varphi, x \rangle \quad \forall \varphi \in E'.$$

On parle de convergence faible- \star pour une suite (φ_n) dans E' qui converge vers φ au sens suivant

$$\lim_{n \rightarrow +\infty} \langle \varphi_n, x \rangle = \langle \varphi, x \rangle \quad \forall x \in E.$$

On écrira alors $\varphi_n \xrightarrow{*} \varphi$.

Théorème IV.1.6. (Hahn-Banach, forme analytique)

Soit E un e.v.n. et $G \subset E$ un sous-espace vectoriel. On considère φ une forme linéaire sur G , continue, de norme

$$\|\varphi\|_{G'} = \sup_{x \in G, \|x\| \leq 1} \langle \varphi, x \rangle.$$

Il existe $f \in E'$ qui prolonge φ , telle que $\|f\|_{E'} = \|\varphi\|_{G'}$.

1. On notera que l'exemple choisi s'appuie sur un e.v.n. qui n'est pas complet. De fait, on s'épargnera de chercher à construire un exemple explicite d'application linéaire non continue sur un espace complet, une telle construction ne peut qu'être abstraite et nécessite l'axiome du choix.

Proposition IV.1.7. Soit E un e.v.n. et $x \in E$. Il existe $\varphi \in E'$ de norme 1 telle que $\langle \varphi, x \rangle = \|x\|$.

Démonstration. Soit $x \in E$. Tout y de $G = \mathbb{R}x$ s'écrit λx . On définit φ sur $\mathbb{R}x$ en posant $\langle \varphi, y \rangle = \lambda \|x\|$. On a $\|\varphi\|_{G'} = 1$ et $\langle \varphi, y \rangle = \|x\|$. Cette forme se prolonge en une forme de E' de norme 1 d'après le théorème IV.1.6. \square

Théorème IV.1.8. (Hahn-Banach, forme géométrique)

Soit E un e.v.n., $K \subset E$ un convexe fermé de E , et $z \notin K$. Il existe une forme linéaire continue $\varphi \in E'$ et un réel α tels que

$$\langle \varphi, x \rangle \leq \alpha < \langle \varphi, z \rangle \quad \forall x \in K.$$

Définition IV.1.9. (Injection canonique, espaces réflexifs)

On définit l'injection canonique $J : E \longrightarrow E''$ comme suit :

$$\forall x \in E, \quad \varphi \in E' \langle J(x), \varphi \rangle_{E'', E'} = \langle \varphi, x \rangle_{E', E}.$$

On a $\|J(x)\|_{E''} = 1$ d'après la proposition IV.1.7, il s'agit donc d'une *isométrie*. On dit que E est *réflexif* si J est surjective, i.e. $J(E) = E''$. On peut alors identifier E et son bidual E'' .

Définition IV.1.10. (Séparabilité)

Un e.v.n. E est dit séparable s'il existe une partie dénombrable dense dans E .

Comme pour la complétude, cette propriété est immédiatement vérifiée pour tout e.v.n. de dimension finie (\mathbb{Q}^n est dense dans \mathbb{R}^n), elle n'a donc de pertinence que pour les espaces de dimension infinie.

Exercice IV.1.2. a) Montrer que l'espace ℓ^∞ , espace des suites réelles bornées muni de la norme $\|u\|_\infty = \sup |u_n|$, n'est pas séparable

b) Montrer qu'en revanche le sous-espace $c_0 \subset \ell^\infty$ des suites qui tendent vers 0 est séparable.

c) Montrer que, pour $p \in [1, +\infty[$ l'espace

$$\ell^p = \left\{ u = (u_n) \in \mathbb{R}^{\mathbb{N}}, \sum u_n^p < +\infty \right\},$$

muni de la norme

$$\|u\| = \left(\sum_{n=0}^{+\infty} |u_n|^p \right)^{1/p},$$

est séparable.

CORRECTION.

a) On considère l'ensemble B des suites de 0 ou de 1. Il existe une surjection de B dans $[0, 1]$ (codage dyadique des réels), il est donc non dénombrable. On considère maintenant les boules ouvertes centrées en les points de B , de rayon $1/2$. Ces boules sont disjointes, et, si D est un ensemble dense, alors chaque boule contient au moins un élément de D , donc D n'est pas dénombrable.

b) Pour c_0 , on considère l'ensemble Q des suites dont les termes sont rationnels, et qui sont nulles au-delà d'un certain rang. Cet ensemble est dénombrable et dense dans c_0 .

c) Le même ensemble Q est dense dans ℓ^p .

IV.2 Compléments sur la dualité

Soient E et F deux e.v.n., et Ψ une forme bilinaire continue sur $E \times F$. On peut associer canoniquement à Ψ une application (linéaire et continue) de F dans E' , le dual topologique de E (espace des formes linéaires continues) :

$$y \in F \longmapsto Ty \in E', \quad \langle Ty, x \rangle = \Psi(x, y) \quad \forall x \in E. \quad (\text{IV.2.1})$$

Proposition IV.2.1. Soient E et F deux espaces vectoriels normés. Si E est séparable², alors de toute suite (y_n) bornée dans F on peut extraire une suite $(y_{n'})$ qui converge au sens suivant :

$$\exists \varphi \in E', \quad T y_{n'} \xrightarrow{*} \varphi,$$

où T est définie par (IV.2.1). Autrement dit, il existe $\varphi \in E'$ telle que

$$\psi(x, y_{n'}) \longrightarrow \langle \varphi, x \rangle \quad \forall x \in E.$$

Démonstration. Il existe une famille dénombrable $\{x_k\}_{k \in \mathbb{N}}$ dense dans E . On se propose de suivre le procédé d'extraction diagonale de Cantor.

1. Comme $\Psi(x_1, y_n)$ est bornée dans \mathbb{R} on peut extraire une suite $y_{j_1(n)}$ telle que $\Psi(x_1, y_{j_1(n)})$ converge.
2. Comme $\Psi(x_2, y_{j_1(n)})$ est bornée dans \mathbb{R} on peut extraire de $y_{j_1(n)}$ une suite $y_{j_1 \circ j_2(n)}$ telle que $\Psi(x_2, y_{j_1 \circ j_2(n)})$ converge.
3. Par récurrence, on construit une suite de sous-suites emboitées $y_{j_1 \circ j_2 \circ \dots \circ j_k(n)}$ telle que $\Psi(x_k, y_{j_1 \circ j_2 \circ \dots \circ j_k(n)})$ converge, pour tout k .
4. On utilise à présent le procédé d'extraction diagonale : on pose $j(k) = j_1 \circ j_2 \circ \dots \circ j_k(k)$ (de telle sorte que j est strictement croissante), et on considère $y_{j(n)}$. Pour tout k , on remarque que $y_{j(n)}$, à partir du rang k , est aussi une suite extraite de $(y_{j_1 \circ j_2 \circ \dots \circ j_k(n)})$, de telle sorte que $\Psi(x_k, y_{j(n)})$ converge lorsque $n \rightarrow +\infty$.
5. On utilise pour finir la densité des x_k pour montrer que, pour tout $x \in H$, $\Psi(x, y_{j(n)})$ est une suite de Cauchy. Soit $\varepsilon > 0$, il existe (x_k) tel que $|x - x_k| < \varepsilon$. Comme $\Psi(x_k, y_{j(n)})$ est de Cauchy, il existe un N au-delà duquel $|\Psi(x_k, y_{j(p)}) - \Psi(x_k, y_{j(q)})| < \varepsilon$. Pour tous p, q supérieurs à N , on a donc

$$\begin{aligned} & |\Psi(x, y_{j(p)}) - \Psi(x, y_{j(q)})| \\ & \leq |\Psi(x, y_{j(p)}) - \Psi(x_k, y_{j(p)})| + |\Psi(x_k, y_{j(p)}) - \Psi(x_k, y_{j(q)})| + |\Psi(x_k, y_{j(q)}) - \Psi(x, y_{j(q)})| \\ & \leq (1 + 2C \|\Psi\|) \varepsilon, \end{aligned}$$

où $\|\Psi\|$ est la constante de continuité de Ψ (telle que $|\Psi(x, y)| \leq \|\Psi\| \|x\| \|y\|$, et C un majorant de $\|y_n\|$).

La suite $(y_{j(n)})$ est donc telle que $\Psi(x, y_{j(n)})$ converge, pour tout x , vers un réel noté $h(x)$. Cette limite est linéaire par rapport à x , et de norme majorée par une constante fois la norme de x , il s'agit donc d'une forme linéaire continue sur F . \square

On notera l'importance de la séparabilité de E dans la démonstration ci-dessus. Par ailleurs, le procédé construit une limite qui n'est pas un élément de F , mais une forme linéaire sur E' , qui n'est pas nécessairement dans l'image de T .

La proposition précédente est très générale, et d'ailleurs très vide dans certains cas (prendre par exemple Ψ identiquement nulle, ou bien E de dimension finie alors que F est de dimension infinie). La propriété devient pertinente quand l'espace E et la forme Ψ sont tels que la dualité est *séparante*, c'est à dire (on privilégie ici l'espace E) que

$$\Psi(x, y) = 0 \quad \forall x \implies y = 0.$$

Cette propriété assure l'*injectivité* de l'application T définie ci-dessus.

La richesse de l'espace F peut être formalisée par la condition symétrique de dualité séparante :

$$\Psi(x, y) = 0 \quad \forall y \implies x = 0.$$

Si cette seconde condition est vérifiée, alors l'image de T est dense dans E' pour la topologie faible-* sur E' (i.e. en dualité avec E'). Dans le cas où E est réflexif, on aura bien densité de $T(F)$ dans E' . On prendra garde au fait que, si E n'est pas réflexif, on peut avoir E et F en dualité séparante sans que $T(F)$ ne soit dense

2. Il admet une famille dénombrable dense.

dans E' . Considérer par exemple $E = \ell^\infty$, $F = \ell^1$, et Ψ la dualité canonique entre ces deux espaces. Elle est évidemment (doublement) séparante, mais $T(\ell^1)$ n'est pas dense dans ℓ^∞ : la forme linéaire qui à une suite de ℓ^∞ convergente associe sa limite, prolongée sur ℓ^∞ (par le théorème de Hahn-Banach analytique IV.1.6, page 92), est à distance au moins 1 de $T(\ell^1)$.

Corollaire IV.2.2. Soit E un e.v.n. séparable. De toute suite bornée dans E' on peut extraire une sous-suite bornée qui converge pour la topologie faible- \star .

On fera bien la distinction entre le corollaire précédent et le théorème de Banach-Alaoglu-Bourbaki, qui établit la compacité de la boule unité de E' pour la topologie faible- \star , sans hypothèse de séparabilité. Dans le cas où E n'est pas séparable, on a bien compacité, mais la topologie n'est *pas métrisable*, de telle sorte que la compacité ne peut pas se traduire en termes de suites extraites convergentes³. Ainsi la boule unité de ℓ^1 est bien compacte pour $\sigma(\ell^\infty, \ell^1)$, mais on ne peut par exemple extraire aucune sous suite convergente (faible- \star) de la suite (e_n) .

Corollaire IV.2.3. Soit E un espace de Banach dont le dual est séparable. De toute suite bornée dans E on peut extraire une sous-suite qui converge⁴ dans E'' pour la topologie $\sigma(E', E'')$. Si E est réflexif, la sous-suite converge faiblement dans E .

Dans le cas Hilbertien on peut supprimer la condition de séparabilité.

Corollaire IV.2.4. Soit H un espace de Hilbert. De toute suite bornée dans H on peut extraire une sous-suite qui converge faiblement dans H

Démonstration. Il suffit de se placer dans l'adhérence V de l'espace vectoriel engendré par les termes de la suite, qui est séparable par construction. On vérifie ensuite que l'on a bien convergence faible sur $H = V + V^\perp$ de la suite extraite. \square

Espaces fonctionnels, mesures

On considère Ω un domaine de \mathbb{R}^d (qui peut être l'espace tout entier).

Le corollaire IV.2.3 permet d'extraire d'une suite bornée une sous-suite faiblement convergente dès que l'espace considéré est réflexif, donc en particulier dans les espaces $L^p(\Omega)$ pour $1 < p < +\infty$, ainsi que dans les espaces de Sobolev $W^{m,p}(\Omega)$, pour tout $m \in \mathbb{N}$, tout $p \in]1, +\infty[$.

Pour les espaces non réflexifs (comme $L^1(\Omega)$ ou $L^\infty(\Omega)$, ou les espaces de Sobolev associés), la propriété est fausse en général, comme l'illustrent les exemples suivants.

Dans $L^1(\mathbb{R})$: la suite $f_n = \mathbf{1}_{]n, n+1[}$ est sur la sphère unité. Si une sous-suite converge faiblement vers f , alors f s'annule contre toute fonction régulière à support compact, elle est donc nécessairement nulle. Mais par ailleurs $\langle 1, f_n \rangle$ est identiquement égale à 1, on doit donc avoir $\langle 1, f \rangle = 1$, ce qui est impossible.

Dans L^∞ , les choses sont un peu plus délicates, car le dual de cet espace n'est pas clairement identifié⁵. En particulier, le fait que l'on puisse (ou pas) extraire une sous-suite convergente de la suite définie précédemment n'est pas aisément à trancher. On peut néanmoins construire un contre-exemple analogue, en considérant par exemple la forme linéaire sur $L^\infty(\mathbb{R})$ qui à une fonction convergente en $+\infty$ associe sa limite, prolongée par le théorème de Hahn-Banach analytique en $\varphi \in (L^\infty(\Omega))'$. On considère alors la suite $f_n = \mathbf{1}_{]n, +\infty[}$. Si elle converge faiblement vers f , alors nécessairement f est nulle presque partout, donc tend vers 0 en $+\infty$, or on doit avoir $\langle \varphi, f \rangle = 1$, ce qui est absurde.

Convergence faible dans les cas non réflexifs L'espace $L^\infty(\Omega)$ s'identifie au dual de $L^1(\Omega)$, qui est séparable, on peut donc, d'une suite bornée dans L^∞ extraire une sous-suite qui converge (faible- \star) vers une limite de L^∞ .

3. Autant dire qu'elle n'est pas commode à *utiliser* dans la vie de tous les jours.

4. Plus précisément son image par la surjection canonique de E dans E'' .

5. Montrer que le dual de L^∞ contient des formes qui ne peuvent pas se représenter par des fonctions de L^1 nécessite l'utilisation du théorème de Hahn-Banach analytique IV.1.6, page 92, donc indirectement de l'axiome du choix.

L'espace $L^1(\Omega)$, dont le dual L^∞ n'est pas séparable, peut être mis en dualité avec des espaces de fonctions continues (munis de la norme ∞) : espace C_c des fonctions continues à support compact, espace C_0 des fonctions qui tendent vers 0 au bord de Ω , et l'espace C_b des fonctions bornées sur Ω . Noter que ces trois espaces s'identifient si l'on se place sur un compact. Dans le cas d'un domaine ouvert considéré ici, les 2 premiers espaces sont séparables, mais le troisième ne l'est pas. D'une suite bornée dans L^1 on pourra donc extraire une sous-suite qui converge vaguement (contre les fonctions de C_c) ou faiblement (contre les fonctions de C_0), mais la limite est définie comme une forme linéaire sur ces espaces, elle ne s'identifie pas forcément à une fonction de L^1 : il s'agit en toute généralité d'une mesure bornée. Par exemple la suite $f_n = n\mathbf{1}_{]0,1/n]}$ converge faiblement vers la masse de Dirac en 0. En l'occurrence, cette convergence est aussi étroite, mais on prendra garde au fait que l'on ne peut en général, d'une suite bornée de L^1 , extraire une sous-suite qui converge étroitement (du fait de la non séparabilité de $C_b(\Omega)$). Ainsi la suite $f_n = n\mathbf{1}_{]n,n+1/n]}$ converge vaguement ou faiblement vers 0, mais il n'en existe aucune sous-suite qui convergerait étroitement.

Exercice IV.2.1. On considère l'espace E des fonctions continues sur \mathbb{R}^d qui convergent vers une valeur finie lorsque $|x|$ tend vers $+\infty$. Montrer qu'il s'agit d'un espace complet (pour la norme ∞) séparable, et énoncer une propriété de compacité séquentielle faible- \star pour $L^1(\mathbb{R}^d)$ mis en dualité avec E . Que peut-on dire de la suite $f_n = n\mathbf{1}_{]n,n+1/n]}$ définie précédemment ?

Proposer une généralisation de cette approche à des fonctions pour lesquelles la limite en $+\infty$ dépend de la direction $x/|x|$. (On pourra commencer par le cas $d = 1$, avec simplement 2 limites différentes en $+\infty$ et $-\infty$.)

Chapitre V

Calcul Différentiel

Sommaire

V.1	Dérivées partielles, notion de différentielle	97
V.1.1	Définitions, premières propriétés	97
V.1.2	Compléments	104
V.1.3	Théorème fondamental de l'analyse	109
V.1.4	Formules de changement de variable pour les transformations régulières	111
V.2	Exercices	111
V.3	Théorèmes des fonctions implicites et d'inversion locale	119
V.4	Problème adjoint	124
V.5	Exercices	128
V.6	Dérivées d'ordre supérieur	134
V.6.1	Dérivées partielles d'ordre supérieur pour les fonctions scalaires	134
V.6.2	Différentielles d'ordre supérieur pour les fonctions de \mathbb{R}^n dans \mathbb{R}^m	137
V.7	Exercices	138

V.1 Dérivées partielles, notion de différentielle

V.1.1 Définitions, premières propriétés

On sait qu'une fonction f , définie d'un intervalle ouvert $I \subset \mathbb{R}$ à valeurs dans \mathbb{R} , est dérivable en $x \in I$ si le taux de variation admet une limite, notée alors $f'(x)$, lorsque h tend vers 0, c'est-à-dire si l'on peut écrire

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \varepsilon(h),$$

où $\varepsilon(h)$ tend vers 0 quand h tend vers 0.

On a alors

$$f(x+h) = f(x) + f'(x)h + \varepsilon(h)h = f(x) + f'(x)h + o(h), \quad (\text{V.1.1})$$

où $o(h)$ (que l'on peut aussi écrire $h\varepsilon(h)$) est une fonction définie au voisinage de 0, négligeable devant $|h|$. Ce développement exprime le fait que la fonction peut être approchée à l'ordre 1 au voisinage de x par une application affine.

Inversement, si une fonction f admet au voisinage de x un développement limité du type de (V.1.1) :

$$f(x+h) = f(x) + \gamma h + o(h),$$

alors la fonction est dérivable en x , et le coefficient du terme de premier ordre est $\gamma = f'(x)$, la dérivée de f en x .

Cette approche s'étend sans difficultés au cas où la fonction est à valeurs vectorielles. Considérons

$$f : x \in \mathbb{R} \longmapsto f(x) = (f_1(x), \dots, f_m(x)) \in \mathbb{R}^m.$$

Si la dérivée de chacune des composantes f_i par rapport à la variable réelle est définie en $x \in \mathbb{R}$, la dérivée $f'(x)$ s'écrit

$$f'(x) = (f'_1(x), f'_2(x), \dots, f'_m(x)),$$

et le développement limité est simplement écrit dans \mathbb{R}^m .

Nous allons nous intéresser maintenant la généralisation de ces notions au cas où l'*espace de départ* lui-même peut être de dimension strictement supérieure à 1, l'objet typique étudié à partir de maintenant sera donc une application

$$f : x = (x_1, \dots, x_n) \in \mathbb{R}^n \longmapsto f(x) = (f_1(x), \dots, f_m(x)) \in \mathbb{R}^m$$

où chacune des m composantes f_i est une application de \mathbb{R}^n dans \mathbb{R} .

Exemple V.1.1. (Champ de vecteurs, champs scalaire)

Un *champ de vecteurs* dans l'espace physique est une application qui à chaque point \mathbb{R}^3 associe un vecteur de \mathbb{R}^3 . On le note en général $u = (u_1, u_2, u_3)$ où chaque composante u_i est une fonction de $x = (x_1, x_2, x_3)$. Il peut encoder un champ de vitesses fluides à un instant donné, ou un champ de déformations infinitésimales au sein d'un objet élastique déformable. Un champ de vecteurs dans le plan (par exemple un champ de vitesses horizontales à la surface d'un lac) correspond au cas $n = m = 2$.

On parle d'un *champ scalaire* lorsque l'espace d'arrivée est \mathbb{R} (cas $n = 3$ et $m = 1$ pour un champ de l'espace physique). Cela correspond par exemple au champ de température dans une zone de l'espace à un instant donné.

Exemple V.1.2. On peut considérer les versions dynamiques des exemples ci-dessus en rajoutant une variable de temps dans l'espace de départ. Par exemple un champ de vitesse variable en temps correspond au cas $n = 4, m = 3$, il peut être considéré comme une application $\mathbb{R}^3 \times \mathbb{R}$ dans \mathbb{R}^3 , qui à chaque $(x, t) = (x_1, x_2, x_3, t)$ fait correspondre un vecteur $(u_1(x, t), u_2(x, t), u_3(x, t))$.

Si l'on cherche à écrire un développement limité du type (V.1.1), l'identité est à valeurs dans \mathbb{R}^m , et la variation h de la variable x de l'espace de départ vit dans \mathbb{R}^m . Le terme $f'(x)h$ doit être remplacé par un terme à valeurs dans \mathbb{R}^m , qui dépend linéairement du vecteur $h \in \mathbb{R}^n$, il s'écrit donc sous la forme d'une application linéaire (de \mathbb{R}^n dans \mathbb{R}^m) appliquée à la variation $h \in \mathbb{R}^n$. Cette section décrit la démarche permettant d'écrire dans ce contexte multidimensionnel le développement limité d'une fonction de n variables, à valeurs dans \mathbb{R}^m , c'est à dire d'approcher localement une fonction générale par une fonction *affine*.

La notion de *differentielle*, qui généralise la notion de dérivée d'une fonction de \mathbb{R} dans \mathbb{R} , peut se définir de façon abstraite, y compris pour des espaces de dimension infinie. Nous débutons néanmoins ce chapitre par la notion plus directement accessible et utilisable de dérivée partielle, pour des fonctions de \mathbb{R}^n dans \mathbb{R}^m .

Dérivées partielles

Définition V.1.1. (Dérivées partielles, matrice jacobienne)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m , et $x = (x_1, \dots, x_n) \in U$. On dit que f admet en x une dérivée partielle par rapport à la variable x_j si l'application de \mathbb{R} dans \mathbb{R}^m obtenue en figeant toutes les variables sauf la j -ième est dérivable en x_j . Plus formellement, si

$$y \longmapsto f(x_1, \dots, x_{j-1}, y, x_{j+1}, \dots, x_n) \in \mathbb{R}^m$$

définie d'un voisinage de x_j vers \mathbb{R}^m , est dérivable en $y = x_j$.

La dérivée de la i -ème composante de f par rapport à la variable x_j , telle que définie ci-dessus, est alors notée¹

$$\frac{\partial f_i}{\partial x_j}(x) \text{ ou } \partial_{x_j} f_i(x) = \lim_{s \rightarrow 0} \frac{f_i(x + se_j) - f_i(x)}{s},$$

où l'on a noté e_j le j -ème vecteur de la base canonique de \mathbb{R}^n . Si toutes les dérivées partielles des f_i par rapport aux x_j existent, on appelle *matrice Jacobienne* la matrice

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \in \mathcal{M}_{mn}(\mathbb{R}).$$

On désignera par $\partial f / \partial x_j \in \mathbb{R}^m$ le j -ième vecteur colonne de la matrice jacobienne.

Exercice V.1.1. Écrire la matrice jacobienne (en justifiant son existence) de l'application

$$f : (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \longmapsto \begin{pmatrix} 2x_1 + x_2 + 3x_3 \\ x_1 x_2^2 x_3^4 \end{pmatrix} \in \mathbb{R}^2.$$

CORRECTION.

Toutes les composantes sont polynomiales, donc dérivables sur \mathbb{R} , et la matrice jacobienne de f s'écrit

$$J_f = \begin{pmatrix} 2 & 1 & 3 & 0 \\ x_2^2 x_3^4 & 2x_1 x_2 x_3^4 & 4x_1 x_2^2 x_3^3 & 0 \end{pmatrix}.$$

La dernière colonne est nulle car rien ne dépend de x_4 .

Remarque V.1.2. La définition des dérivées partielles se base sur des variations, autour du point considéré, dans les directions des axes de coordonnées, et dans ces directions seulement. Il est possible que la fonction ait un comportement pathologique si l'on considère des variations dans d'autres directions. On peut par exemple imaginer une fonction qui ne varie pas lorsque l'on perturbe selon une direction de coordonnées (on aura alors existence de dérivées partielles nulles), mais qui a un comportement très singulier dans d'autres directions (voir exercice V.1.3 ci-après). L'existence d'une matrice jacobienne (et son expression le cas échéant), n'est donc pas une propriété intrinsèque, elle dépend du système de coordonnées choisi.

Nous allons à présent définir la notion plus intrinsèque de différentiabilité d'une application de \mathbb{R}^n dans \mathbb{R}^m , dont la définition ne repose pas sur un système de coordonnées particulier. La définition repose sur l'existence d'un développement limité *uniforme* vis-à-vis de la direction de variation. Lorsque $n = 1$, nous avons rappelé précédemment que l'existence d'un développement limité est équivalente à l'existence d'une dérivée. Comme nous le verrons, cette équivalence *ne se généralise pas* au cas où l'espace de départ est de dimension ≥ 2 : l'existence de dérivées partielles en un point ne garantit pas la différentiabilité.

Notation V.1.3. (Image par une application linéaire et produit matrice vecteur)

Nous adoptons dans ce qui suit une convention courante dans le contexte du calcul différentiel (et en particulier en mécanique des fluides), qui est de noter $F \cdot x$ l'image par une application linéaire F d'un vecteur x . De la même manière, si A est une matrice, écrira $A \cdot x$ le produit matrice vecteur. Cette notation est issue de ce que l'on appelle le calcul tensoriel², qui n'est pas abordé en tant que tel dans ce cours.

1. Nous commettons ici un abus de notation si courant qu'il nous paraît préférable de le commettre en connaissance de cause, plutôt que de le contourner. Dans ce qui suit x_j dans $\partial f_i / \partial x_j$ encode le fait que l'on dérive par rapport à la j -ème variable. Mais quand on écrit $x = (x_1, \dots, x_j, \dots, x_n)$, x_j désigne un réel, qui est la valeur particulière de la j -ième composante du point x considéré.

2. On pourra se reporter à
http://mms2.ensmp.fr/mmc_st_etienne_fort/calcul_tensoriel/polycop/tenseurs_poly.pdf
pour une présentation détaillée de ces notions.

Définition V.1.4. (Différentielle (\bullet))

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m . On dit que f est différentiable en $x \in U$ s'il existe une application linéaire de \mathbb{R}^n dans \mathbb{R}^m , notée $df(x)$, telle que

$$f(x + h) = f(x) + df(x) \cdot h + o(h) \quad (\text{V.1.2})$$

où $o(h)$ est une application de \mathbb{R}^n dans \mathbb{R}^m négligeable devant $\|h\|$, c'est à dire telle que $o(h)/\|h\|$ tend vers 0 quand h tend vers 0. On appelle $df(x)$ la *différentielle* de f en x .

Exercice V.1.2. a) Montrer que l'application suivante est différentiable, et préciser sa différentielle

$$f : (x_1, x_2, x_3) \in \mathbb{R}^3 \longmapsto x_1 x_2^2 x_3^3$$

b) Préciser le domaine de différentiabilité de l'application

$$x \in \mathbb{R}^n \longmapsto \|x\|$$

et exprimer sa différentielle lorsqu'elle existe.

CORRECTION.

a) On a

$$f(x + h) = (x_1 + h_1)(x_2 + h_2)^2(x_3 + h_3)^3 =$$

$$x_1 x_2^2 x_3^3 + x_2^2 x_3^3 h_1 + 2x_1 x_2 x_3^3 h_2 + 3x_1 x_2^2 x_3^2 h_3 + \mathcal{O}(\|h\|^2).$$

En effet, les termes qui restent sont tous au moins d'ordre total 2 vis-à-vis des h_i . L'application est donc différentiable en tout point, et sa différentielle est le terme d'ordre 1 :

$$df(x) \cdot h = x_2^2 x_3^3 h_1 + 2x_1 x_2 x_3^3 h_2 + 3x_1 x_2^2 x_3^2 h_3.$$

b) On a

$$\|x + h\| = \left(\|x + h\|^2 \right)^{1/2} = \left(\|x\|^2 + 2\langle x | h \rangle + o(h) \right)^{1/2}.$$

Pour $x \neq 0$, on peut écrire

$$\|x + h\| = \|x\| \left(1 + 2 \frac{\langle x | h \rangle}{\|x\|^2} + o(h) \right)^{1/2} = \|x\| \left(1 + \frac{\langle x | h \rangle}{\|x\|^2} + o(h) \right) = f(x) + \frac{\langle x | h \rangle}{\|x\|} + o(h).$$

L'application est donc différentiable, avec

$$df(x) \cdot h = \left\langle \frac{x}{\|x\|} \mid h \right\rangle$$

Proposition V.1.5. Toute application différentiable en un point est continue en ce point.

Démonstration. C'est une conséquence directe du développement limité (V.6.2), qui assure que $f(x + h)$ tend vers $f(x)$ quand h tend vers 0. \square

Définition V.1.6. (Continue différentiabilité (\bullet))

Une application f d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m est dite *continûment différentiable* sur U si elle est différentiable en tout $x \in U$, et si l'application $x \mapsto df(x)$ est continue sur U (l'espace d'arrivée est muni canoniquement de la norme d'opérateur subordonnée à la norme euclidienne, voir proposition I.8.6, page 32).

Lien entre différentielle et matrice jacobienne

Lorsque la différentielle existe, sa représentation dans les bases canoniques de \mathbb{R}^n et \mathbb{R}^m est la matrice jacobienne définie ci-dessus.

Proposition V.1.7. Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m . Si f est différentiable en $x \in U$, alors elle admet des dérivées partielles dans toutes les directions, et sa différentielle admet pour représentation matricielle la matrice jacobienne J définie ci-dessus : c'est-à-dire que f s'écrit

$$f(x + h) = f(x) + J(x) \cdot h + o(h).$$

Démonstration. Si f est différentiable en x , on peut écrire le développement limité composante par composante), en prenant la variation h de la forme se_j , où s est un réel et e_j un vecteur unitaire de la base canonique \mathbb{R}^n : pour tout $i = 1, \dots, m$, tout $j = 1, \dots, m$,

$$f_i(x + se_j) = f_i(x) + s(df(x) \cdot e_j)_i + o(s).$$

On a donc

$$(df(x) \cdot e_j)_i = \lim_{s \rightarrow 0} \frac{f_i(x + se_j) - f_i(x)}{s} = \frac{\partial f_i}{\partial x_j}$$

d'après la définition V.1.1), c'est-à-dire le coefficient (i, j) de la matrice jacobienne J . Ce coefficient s'identifie donc à la composante i de l'image du j -ème vecteur de la base canonique par la différentielle, ce qui termine la preuve. \square

Comme nous l'avons déjà évoqué, dès que la dimension n de l'espace de départ est strictement plus grande que 1, l'existence d'une matrice jacobienne (c'est à dire l'existence de toutes les dérivées partielles) *n'implique pas* la différentiabilité. Une application peut même admettre une matrice jacobienne en un point sans pour autant être continue en ce point (voir exercice V.1.3 ci-dessous). On verra néanmoins que, si une application admet sur un ouvert U des dérivées partielles qui sont toutes *continues* sur U , alors l'application est continûment différentiable sur U (voir proposition V.1.8 ci-après).

Exercice V.1.3. Montrer que la fonction

$$(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\} \mapsto \frac{xy}{x^2 + y^2}, \quad f(0, 0) = 0,$$

admet en $(0, 0)$ des dérivées partielles, mais n'est pas continue en ce point (et donc non différentiable d'après la proposition V.1.5).

CORRECTION.

La fonction est identiquement nulle sur les axes de coordonnées, elle admet donc des dérivées partielles nulles sur ces axes, et en particulier en 0. La fonction n'est par contre pas continue en 0. On a par exemple

$$\lim_{t \rightarrow 0} f(t, t) = \frac{1}{2}$$

qui n'est pas la valeur en 0. La fonction n'est donc a fortiori pas différentiable en 0.

Exercice V.1.4. (•) Montrer que toute application affine définie d'un ouvert de \mathbb{R}^n dans \mathbb{R}^m :

$$x \mapsto f(x) = A \cdot x + b, \quad A \in \mathcal{M}_{mn}(\mathbb{R}), \quad b \in \mathbb{R}^m,$$

est continûment différentiable sur cet ouvert, et préciser sa différentielle.

CORRECTION.

Lorsque l'application est affine, on a directement le développement limité

$$f(x + h) = f(x) + A \cdot h.$$

La différentielle est donc l'application qui à h associe $A \cdot h$, qui se représente donc matriciellement par la matrice A .

Exercice V.1.5. (•) a) Soient f et g deux applications différentiables sur un ouvert $U \subset \mathbb{R}^2$, à valeurs dans \mathbb{R} . Exprimer la différentielle de l'application produit

$$F : (x_1, x_2) \in U \mapsto f(x_1, x_2)g(x_1, x_2).$$

b) Soient f (respectivement g) une application différentiable sur un ouvert U (respectivement V) de \mathbb{R}^2 dans \mathbb{R} . Exprimer la différentielle de l'application

$$G : (x_1, x_2, x_3, x_4) \in U \times V \mapsto f(x_1, x_2)g(x_3, x_4),$$

et écrire sa jacobienne en fonctions de celles de f et g . On pourra utiliser la notation $x_{12} = (x_1, x_2) \in \mathbb{R}^2$, $h_{12} = (h_1, h_2) \in \mathbb{R}^2$, et de même pour les indices 3 et 4.

CORRECTION.

a) On a (avec $x = (x_1, x_2)$, $h = h_1, h_2$)

$$F(x + h) = F(x_1 + h_1, x_2 + h_2) = f(x_1 + h_1, x_2 + h_2)g(x_1 + h_1, x_2 + h_2)$$

$$= (f(x) + df(x) \cdot h + o(h))(g(x) + dg(x) \cdot h + o(h)) = f(x)g(x) + \underbrace{(g(x)df(x) + f(x)dg(x))}_{dF(x)} \cdot h + o(h).$$

b) On a maintenant (avec $x_{ij} = (x_i, x_j)$, $h_{ij} = (h_i, h_j)$)

$$\begin{aligned} G(x + h) &= f(x_{12} + h_{12})g(x_{34} + h_{34}) \\ &= (f(x_{12}) + df(x_{12}) \cdot h_{12} + o(h_{12}))(g(x_{34}) + dg(x_{34}) \cdot h_{34} + o(h_{34})) \\ &= \underbrace{f(x_{12})g(x_{34})}_{G(x)} + \underbrace{(g(x_{34})df(x_{12}) \cdot h_{12} + f(x_{12})dg(x_{34}) \cdot h_{34})}_{dG(x) \cdot h} + o(h). \end{aligned}$$

La matrice Jacobienne s'écrit par bloc de la façon suivante

$$J_G = \begin{pmatrix} g(x_{34})J_f(x_{12}) & 0 \\ 0 & f(x_{12})J_g(x_{34}) \end{pmatrix} \in \mathcal{M}_{2,4}(\mathbb{R}),$$

où J_f et J_g sont des matrices-lignes (à deux éléments).

D'après la proposition V.1.7, si une application est continûment différentiable sur un ouvert U , alors la matrice jacobienne est définie en tout point de cet ouvert, et la correspondance $x \mapsto J(x)$ est continue. La proposition suivante assure la réciproque de cette propriété.

Proposition V.1.8. (••) Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m . On suppose que la matrice jacobienne $J(x)$ est définie en chaque point x de U , et que l'application $x \mapsto J(x)$ est continue. Alors f est continûment différentiable sur U .

Démonstration. On écrit la démonstration pour le cas $n = 2$, et l'on suppose que $(0, 0) \in U$ pour simplifier les notations. Nous allons montrer la différentiabilité en $(0, 0)$, la démonstration pour les autres points étant essentiellement la même. Pour h_1, h_2 suffisamment petits, on a

$$f(h_1, h_2) = f(h_1, h_2) - f(h_1, 0) + f(h_1, 0) - f(0, 0) + f(0, 0).$$

On écrit

$$f(h_1, 0) - f(0, 0) = h_1 \int_0^1 \partial_1 f(th_1, 0) dt.$$

Comme $x \mapsto J(x)$ est continue, tous les coefficients de la matrice J sont des fonctions continues en x . On a

donc en particulier $\partial_1 f(th_1, 0) = \partial_1 f(0, 0) + \varepsilon(th_1)$, et ainsi³

$$f(h_1, 0) - f(0, 0) = h_1 \partial_1 f(0, 0) + h_1 \varepsilon(h_1).$$

On a par ailleurs

$$f(h_1, h_2) - f(h_1, 0) = h_2 \int_0^1 \partial_2 f(h_1, th_2) dt,$$

avec, par continuité des dérivées partielles,

$$\partial_2 f(h_1, th_2) = \partial_2 f(0, 0) + \varepsilon(h_1, th_2).$$

On a donc⁴

$$f(h_1, h_2) - f(h_1, 0) = h_2 \partial_2 f(0, 0) + h_2 \varepsilon(h).$$

On a donc finalement

$$f(h_1, h_2) = f(0, 0) + h_1 \partial_1 f(0, 0) + h_2 \partial_2 f(0, 0) + o(h),$$

qui exprime la différentiabilité de f en $(0, 0)$, et de la même manière en tout point de l'ouvert U . La différentielle peut s'exprimer matriciellement à partir de la jacobienne $J = [\partial_1 f, \partial_2 f]$ (écriture de la matrice en colonnes, chacune des dérivées partielles étant un vecteur de \mathbb{R}^m). Les dérivées partielles étant continues, la correspondance $x \mapsto J(x)$ est continue, l'application est donc continûment différentiable sur U .

La réciproque est une conséquence directe de la proposition V.1.7. \square

Proposition V.1.9. (Différentielle de la composée de deux applications)

Soient g une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^p , et f définie d'un ouvert $V \in \mathbb{R}^p$ dans \mathbb{R}^m . On suppose que $g(U) \subset V$. Si g est différentiable en $x \in U$ et f est différentiable en $g(x)$, alors $f \circ g$ est différentiable en x , et l'on a

$$d(f \circ g)(x) = df(g(x)) \circ dg(x).$$

Démonstration. On a

$$\begin{aligned} f \circ g(x + h) &= f(g(x + h)) = f(g(x) + dg(x) \cdot h + o(h)) \\ &= f(g(x)) + df(g(x)) \cdot (dg(x) \cdot h + o(h)) + o(dg(x) \cdot h + o(h)) \\ &= f \circ g(x) + (df(g(x)) \circ dg(x)) \cdot h + o(h), \end{aligned}$$

qui exprime la différentiabilité de $f \circ g$, avec l'expression annoncée de la différentielle. \square

Exercice V.1.6. a) Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$, $a \in \mathbb{R}^n$, et f une application différentiable de \mathbb{R}^n dans \mathbb{R}^m . Exprimer la différentielle et la matrice jacobienne de $F : x \mapsto f(b + Ax)$.

b) Soit f une fonction dérivable de \mathbb{R} dans \mathbb{R} . Déterminer la différentielle de

$$F : (x, y) \in \mathbb{R}^2 \mapsto f(x^2 + y^2),$$

et écrire le développement limité de F au voisinage d'un point $(x, y) : F(x + h_x, y + h_y) = \dots$

3. La fonction $s \mapsto \varepsilon(s)$ est une fonction définie d'un voisinage de 0 dans \mathbb{R} , à image dans \mathbb{R}^m , qui tend vers 0 quand s tend vers 0, on a donc : pour tout $\epsilon > 0$, il existe η tel que pour tout h_1 avec $|h_1| < \eta$, $\|\varepsilon(h_1)\| < \epsilon$. Pour un tel h_1 , pour tout $t \in [0, 1]$, on a aussi $|th_1| < \eta$, et donc $\|\varepsilon(th_1)\| < \epsilon$. Par suite

$$\left\| \int_0^1 \varepsilon(th_1) dt \right\| \leq \int_0^1 \|\varepsilon(th_1)\| dt \leq 1 \times \eta = \eta.$$

L'intégrale est donc un $\varepsilon(h_1)$, c'est-à-dire une fonction qui tend vers 0 quand h_1 tend vers 0.

4. Comme précédemment, l'intégration en t entre 0 et 1 du $\varepsilon(h_1, th_2)$ donne un $\varepsilon(h)$, qui n'est plus la même fonction ε que précédemment, conformément à l'usage, mais qui tend bien vers 0 quand h tend vers 0.

CORRECTION.

a) On a

$$dF(x) = df(b + Ax) \circ A,$$

et $J_F(x) = J_f A$ (produit matrice-vecteur).

b) On a

$$J_F(x, y) = f'(x^2 + y^2)(2x \ 2y).$$

$$F(x + h_x, y + h_y) = F(x, y) + f'(x^2 + y^2)(2xh_x + 2yh_y) + o(h).$$

Remarque V.1.10. (Et en pratique, on fait comment ?)

Les définitions ci-dessus suggèrent deux stratégies pour étudier la différentiabilité d'une application donnée, et préciser sa différentielle : calculer sa matrice jacobienne J_f , et identifier le domaine sur lequel elle est définie et continue, ou effectuer un développement limité $J(f + h)$ et en extraire la partie d'ordre 1 en h . Chacune de ses approches peut se révéler la plus pertinente dans certains cas. Ainsi pour l'application de l'exercice V.1.1, il est plus aisés d'écrire directement la matrice jacobienne que d'effectuer le développement limité. Pour la seconde application de l'exercice V.1.2 en revanche, le développement limité permet d'arriver de façon plus élégante à l'expression de la différentielle. Noter qu'en dimension infinie la notion de différentielle peut se définir comme nous l'avons fait en dimension finie, alors que la notion de matrice jacobienne n'a a priori pas de sens. L'approche par développement limité est de ce point de vue plus générale, et plus universelle puisqu'elle ne dépend pas des bases des espaces de départ ou d'arrivée.

V.1.2 Compléments

Proposition V.1.11. (Linéarité de la différentiation)

Soient f et g des applications définies d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m . Si f et g sont différentiables en $x \in U$, alors, pour tous λ, μ réels, l'application $\lambda f + \mu g$ est différentiable, et l'on a

$$d(\lambda f + \mu g) = \lambda df + \mu dg.$$

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

Corollaire V.1.12. L'ensemble $C^1(U, \mathbb{R}^m)$ des applications continûment différentiables sur un ouvert U de \mathbb{R}^n est un espace vectoriel.

Notion de gradient

Lorsqu'une fonction différentiable est à valeurs dans \mathbb{R} , la différentielle est une *forme linéaire*, c'est-à-dire une application linéaire de \mathbb{R}^n dans \mathbb{R} . Elle peut alors s'exprimer⁵ à l'aide du produit scalaire canonique sur \mathbb{R}^n , et s'identifie par ce biais à un vecteur de \mathbb{R}^n . C'est ce vecteur que l'on appelle gradient de f .

Définition V.1.13. (Gradient)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} . On suppose que f est différentiable en $x \in U$. Il existe alors un unique vecteur, noté $\nabla f(x)$, tel que

$$df(x) \cdot h = \langle \nabla f(x) | h \rangle \quad \forall h \in \mathbb{R}^n,$$

où $\langle \nabla f(x) | h \rangle$ représente le produit scalaire canonique sur \mathbb{R}^n .

Conformément à la proposition V.1.7, ce gradient s'écrit dans la base canonique

$$\nabla f(x) = (\partial_{x_1} f(x), \partial_{x_2} f(x), \dots, \partial_{x_n} f(x)) \in \mathbb{R}^n.$$

5. Lorsque l'on travaille sur \mathbb{R}^n , l'usage de la base orthonormée canonique et du produit scalaire canonique sont tellement naturels que l'on a tendance à identifier spontanément la différentielle et le gradient. On prendra cependant garde au fait que le gradient n'est pas défini de façon intrinsèque. Contrairement à la différentielle, qui est une application définie de façon non ambiguë par le développement limité, ce gradient dépend du produit scalaire choisi. Ce point est particulièrement sensible dans certaines situations, notamment en dimension infinie, où plusieurs produits scalaires "naturels" peuvent co-exister.

Exercice V.1.7. Reprendre l'exercice V.1.6 (en supposant $m = 1$ pour la première question), en précisant dans chaque cas le gradient de F en fonction de celui de f .

CORRECTION.

a) Le gradient est tel que, pour tout h ,

$$\langle \nabla F(x) | h \rangle = dF(x) \cdot h = df(b + Ax) \cdot Ah = \langle \nabla f(b + Ax) | Ah \rangle = \langle A^T \nabla f(b + Ax) | h \rangle,$$

d'où

$$\nabla F(x) = A^T \nabla f(b + Ax).$$

b) On a ici

$$\langle \nabla F(x | y), h \rangle = dF(x) \cdot h = f'(x^2 + y^2)(2xh_x + 2yh_y)$$

d'où

$$\nabla F = f'(x^2 + y^2) \begin{pmatrix} 2x \\ 2y \end{pmatrix}.$$

Il sera utile dans certaines applications d'utiliser la notion de *gradient partiel*. Cela consiste simplement à considérer une fonction de n variables comme une fonction d'une partie de ces variables, les autres étant gelées. La définition ci-dessous précise cette notion dans le cas général, que nous illustrons au préalable sur un cas particulier. Considérons une fonction de \mathbb{R}^3 dans \mathbb{R} , différentiable en un point $x = (x_1, x_2, x_3)$. Son gradient en x est un vecteur de \mathbb{R}^3 . Si on la considère maintenant comme une fonction de (x_1, x_2) , avec x_3 fixé à sa valeur correspondant à x , le gradient de cette nouvelle fonction est un vecteur de \mathbb{R}^2 , que l'on pourra noter $\nabla_{x_1 x_2} f$, ou $\nabla_{x_{12}} f$.

Définition V.1.14. (Gradient partiel)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} , différentiable en un point $x \in U$. On écrit $\mathbb{R}^n = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_p}$, de telle sorte que x peut s'écrire⁶ $x = (x_1, \dots, x_p)$, avec $x_j \in \mathbb{R}^{n_j}$. Pour i entre 1 et p , on considère la fonction partielle qui ne dépend que du vecteur x_j , les autres étant figées. Le gradient de cette fonction partielle est un vecteur de \mathbb{R}^{n_j} , on le note $\nabla_{x_j} f$.

Cette notion de gradient partiel est notamment très utile en pratique lorsque l'on définit un potentiel d'interaction sur un système de particules localisées en q_1, q_2, \dots, q_N , chacun des q_i étant un point de l'espace physique \mathbb{R}^3 . Si l'on définit un potentiel d'interaction sur le système comme une fonction Φ de $q = (q_1, \dots, q_N) \in \mathbb{R}^{3N}$, la force exercée sur la particule i dérivant de ce potentiel d'interaction est simplement $-\nabla_{q_i} \Phi \in \mathbb{R}^3$. On se reportera à l'exercice V.2.8, page 116, pour une étude plus approfondie de ces systèmes de particules en interaction, et l'utilisation dans ce cadre de la notion de gradient partiel.

Remarque V.1.15. Si l'on considère le gradient partiel d'une fonction de $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ vis-à-vis d'une unique variable x_i , on retrouve la notion de dérivée partielle par rapport à x_i déjà introduite.

Définition V.1.16. (Différentielle partielle)

On définit de façon tout à fait analogue une notion de différentielle partielle, pour des applications de \mathbb{R}^n vers \mathbb{R}^m . Si l'on écrit comme précédemment

$$x = (x_1, \dots, x_p) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_p},$$

la différentielle partielle par rapport à x_i , que l'on notera⁷ $\partial_{x_i} f$, est alors une application linéaire de \mathbb{R}^{n_i} dans \mathbb{R}^m .

6. Attention, x_j désigne dans ce qui suit non plus une variable scalaire, mais un groupe de n_j variables scalaires.

7. Cette notation est la plus couramment utilisée, et nous en recommandons l'usage, tout en reconnaissant qu'il aurait été assez naturel d'utiliser la notation d_{x_i} , puisqu'il s'agit d'une application linéaire, définie de façon intrinsèque comme associant à tout vecteur un vecteur, indépendamment du choix d'une base. La notation ‘ ∂ ’ a pour l'instant été utilisée pour représenter des dérivées partielles, qui reposent sur le choix d'un système de coordonnées. Dans le cas présent, on est un peu entre les deux : on souhaite représenter une application linéaire, mais définie sur un sous-espace dont la définition repose sur le choix d'un système de coordonnées.

Exercice V.1.8. Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive, de telle sorte que

$$(x, y) \mapsto \langle x | y \rangle_A = \langle Ax | y \rangle$$

puisse être utilisé comme produit scalaire sur \mathbb{R}^d . On note ∇_A le opérateur gradient pour ce produit scalaire. Pour J fonctionnelle continûment différentiable, exprimer $\nabla_A J$ en fonction de ∇J

Démonstration. On a

$$\begin{aligned} J(x + h) &= J(x) + dJ(x) \cdot h + o(h) = J(x) + \langle \nabla J(x) | h \rangle + o(h) = \\ &= J(x) + \langle AA^{-1} \nabla J(x) | h \rangle + o(h) = J(x) + \langle A^{-1} \nabla J(x) | h \rangle_A + o(h), \end{aligned}$$

d'où $\nabla_A J = A^{-1} \nabla J$. □

Exercice V.1.9. (Dépendance du gradient vis-à-vis du produit scalaire, système de particules)

Comme indiqué précédemment, le gradient dépend du produit scalaire sous-jacent. Dans le cadre de ce cours, nous utiliserons cette notion essentiellement en lien avec le produit scalaire canonique de \mathbb{R}^n , parfois sans re-préciser qu'il s'agit bien de ce produit scalaire. Cette exercice illustre le fait qu'il peut être naturel, dans certains contextes, de travailler avec d'autres produits scalaires, et permet de comprendre comment le gradient se voit modifié.

On se place dans $\mathbb{R}^{3N} = \mathbb{R}^3 \times \cdots \times \mathbb{R}^3$ pour représenter les vitesses dans l'espace physique de N particules, de masses m_1, m_2, \dots, m_N toutes strictement positives. On considère la fonction qui à un jeu de vitesses associe l'énergie cinétique

$$E(u) = E(u_1, \dots, u_N) = \frac{1}{2} \sum_{n=1}^N m_n |u_n|^2.$$

Montrer que E est différentiable, et calculer son gradient pour le produit scalaire canonique, puis pour le produit scalaire $\langle \cdot | \cdot \rangle_m$ pondéré par la collection de masses $m = (m_1, \dots, m_N)$, défini par

$$\langle u | v \rangle_m = \sum_{n=1}^N m_n \langle u_n | v_n \rangle,$$

où $\langle u_n | v_n \rangle$ représente le produit scalaire canonique sur \mathbb{R}^3 .

CORRECTION.

On a, pour $h \in \mathbb{R}^{3N}$,

$$E(u + h) = E(u) + \sum_{n=1}^N m_n \langle u_n | h_n \rangle + o(h),$$

d'où l'on déduit que E est différentiable, la différentielle étant l'application qui à h associe le deuxième terme du membre de droite ci-dessus. Ce terme s'écrit $\langle p | h \rangle$, avec

$$p = (p_1, \dots, p_N) \in \mathbb{R}^{3N}, \quad p_n = m_n u_n.$$

Le gradient de l'énergie cinétique par rapport aux vitesses est $\nabla E = p$, qui est la quantité de mouvement. Si l'on choisit de munir l'espace en vitesse du produit scalaire pondéré par les masses, le gradient est le vecteur vitesse $u \in \mathbb{R}^{3N}$.

Exercice V.1.10. Soient a_1 et a_2 deux réels strictement positifs. On considère la forme bilinéaire

$$(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto (x, x) = \frac{x_1 y_1}{a_1^2} + \frac{x_2 y_2}{a_2^2},$$

et l'on note $J(x) = a(x, x)/2$.

Montrer que cette forme $a(\cdot, \cdot)$ définit un produit scalaire sur \mathbb{R}^2 . Préciser la sphère unité pour ce produit scalaire. On considère un point $x = (x_1, x_2)$ de cette sphère unité S , n le gradient de J en x (selon le produit scalaire canonique), et n_a le gradient de J en x selon le produit scalaire défini par a . Décrire les droites qui passent par x et qui ont pour directions n et n_a .

Que ce passe-t-il en dimension $d > 2$?

CORRECTION.

La forme est bien bilinéaire, symétrique, positive et non dégénérée ($a(x, x) > 0 \iff x \neq 0$). Pour $x = (x_1, x_2) \in S$, on a

$$n = \frac{1}{2} \nabla \left(\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} \right) = \begin{pmatrix} \frac{x_1}{a_1^2} \\ \frac{x_2}{a_2^2} \end{pmatrix}$$

On obtient n_a en effectuant (voir exercice V.1.8 ci-dessus) le produit par la matrice $A^{-1} = \text{diag}(a_1^2, a_2^2)$. Le vecteur n_a est donc simplement le vecteur x . La droite associée au gradient canonique est simplement la droite passant par x , orthogonale à S au sens usuel. Elle ne passe pas par l'origine dès que $a_1 \neq a_2$ (faire un dessin). Pour le produit scalaire associé à $a(\cdot, \cdot)$ en revanche, cette droite est simplement le rayon qui passe par x .

La propriété est valable pour toute dimension, comme on peut le vérifier immédiatement.

N.B. : cette propriété est importante en optimisation. Si l'on cherche à minimiser la fonctionnelle J (c'est ici trivial, mais on peut imaginer que l'on rajoute à J une partie linéaire, aucun cas le problème n'est plus trivial), par une méthode de type gradient traditionnel, la direction de descente peut être très inefficace pour se rapprocher du minimum. Alors que si l'on connaît le gradient de la fonctionnelle pour le produit scalaire associé à $a(\cdot, \cdot)$, la direction de descente conduit directement au minimum. Pas de miracle : estimer ce gradient pour $a(\cdot, \cdot)$ est a priori de même difficulté que de résoudre le problème d'optimisation ; ce principe général est néanmoins à la base de la méthode dite du gradient conjugué.

Définition V.1.17. (Point critique / stationnaire)

Soit f une application continûment différentiable d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} . On appelle point *critique* ou point *stationnaire* tout point $x \in U$ en lequel le gradient s'annule.

Exercice V.1.11. (•) Justifier l'appellation *stationnaire* dans la définition précédente.

CORRECTION.

Ce terme vient du fait que la solution du système différentiel canoniquement associé à la fonction, appelé flot de gradient, et qui s'écrit $\dot{x} = -\nabla f(x)$, admet comme solution particulière $x \equiv x_s$ pour tout x_s qui annule le gradient. On parle aussi de point d'équilibre dans le contexte des systèmes dynamiques.

Calcul différentiel

Nous regroupons ici quelques considérations sur la pratique effective du calcul différentiel, et en particulier les notations dx_1 , $dx_1 + dx_2$, etc ...

Si l'on considère par exemple une fonction de \mathbb{R}^2 dans \mathbb{R} , définie par $f(x) = f(x_1, x_2) = x_1^2 + x_2^3 + x_1 x_2$, on écrira

$$df = d(x_1^2 + x_2^3 + x_1 x_2) = 2x_1 dx_1 + 3x_2 dx_2 + x_1 dx_2 + x_2 dx_1 = (2x_1 + x_2) dx_1 + (3x_2 + x_1) dx_2.$$

Dans ce qui précède, dx_1 représente par exemple la différentielle de la fonction $(x_1, x_2) \mapsto x_1$, qui est simplement l'application qui à (h_1, h_2) associe h_1 , qui peut se représenter matriciellement par $[1 \ 0]$. La différentielle de f en (x_1, x_2) est donc représentée dans la base canonique par

$$J(x) = [2x_1 + x_2 \quad 3x_2 + x_1].$$

De façon plus générale, on écrira⁸

$$df(x_1, x_2) = \frac{\partial f}{\partial x_1}(x_1, x_2) dx_1 + \frac{\partial f}{\partial x_2}(x_1, x_2) dx_2.$$

On prendra garde au fait que l'expression dx_1 dépend du contexte. La même expression peut correspondre par exemple à une application de \mathbb{R}^3 dans \mathbb{R} , auquel cas dx_1 est représentée matriciellement par $[1 \ 0 \ 0]$.

8. Nous commettons ici un abus de notation courant, auquel il convient de s'habituer car il est très répandu : le ' x_1 ' qui

Si l'application est à valeurs vectorielles, par exemple dans \mathbb{R}^2 :

$$f(x_1, x_2) = \begin{bmatrix} x_1^2 + x_2^3 + x_1 x_2 \\ x_1 \end{bmatrix},$$

on écrira de la même manière

$$df(x_1, x_2) = \begin{bmatrix} (2x_1 + x_2) dx_1 + (3x_2 + x_1) dx_2 \\ dx_1 \end{bmatrix},$$

qui peut se représenter matriciellement par

$$J(x_1, x_2) = \begin{bmatrix} 2x_1 + x_2 & 3x_2 + x_1 \\ 1 & 0 \end{bmatrix},$$

de telle sorte que, pour tout h dans \mathbb{R}^2 , on a le développement limité

$$f(x + h) = f(x) + J(x) \cdot h + o(h),$$

où $J(x) \cdot h$ représente le produit matrice vecteur, comme indiqué précédemment.

Exercice V.1.12. Calculer la différentielle de la forme de Minkovski

$$f : (x, y, z, t) \in \mathbb{R}^4 \longmapsto x^2 + y^2 + z^2 - c^2 t^2,$$

avec $c > 0$ (vitesse de la lumière).

CORRECTION.

On a

$$df = 2xdx + 2ydy + 2zdz - 2c^2tdt,$$

que l'on peut aussi représenter matriciellement par $(2x \ 2y \ 2z \ - 2c^2t)$.

Récapitulatif

Les développements ci-dessus décrivent des manières variées d'exprimer qu'une fonction peut être approchée localement par une fonction affine. Nous récapitulons ici ces différentes manières, en rappelant leur cadre d'utilisation et les liens entre elles. Dans ce qui suit f désigne, sauf mention contraire, une application de \mathbb{R}^n dans \mathbb{R}^m .

Comme on l'a vu, f est dite différentiable en $x \in \mathbb{R}^n$ s'il existe une application $df(x)$ telle que

$$f(x + h) = f(x) + df(x) \cdot h + o(h).$$

Le terme $df(x) \cdot h$ désigne l'image par $df(x)$ du vecteur h . Cette expression est intrinsèque, au sens où elle ne dépend pas du choix d'une base. En pratique, on assimile souvent une application linéaire et son écriture matricielle dans la base canonique, mais il est important de garder en tête la différence entre les deux. Cette approche permet notamment une extension immédiate de la définition en dimension infinie, dans un contexte où les bases sont inutilisables.

Si f est différentiable en x , alors (proposition V.1.7) la différentielle admet une représentation matricielle dans la base canonique qui est la matrice jacobienne $J = (\partial_{x_j} f_i)$. On a donc

$$f(x + h) = f(x) + J(x) \cdot h + o(h),$$

apparaît dans ∂_{x_1} et dans dx_1 représente une variable générique vis à vis de laquelle on dérive, alors que le ' x_1 ' de $df(x_1, x_2)$ est un nombre réel, première coordonnée du point en lequel on dérive la fonction. On devrait en toute rigueur distinguer ces deux acceptations en utilisant des noms différents, par exemple

$$df(a_1, a_2) = \partial_{x_1} f(a_1, a_2) dx_1 + \partial_{x_2} f(a_1, a_2) dx_2.$$

où $J(x) \cdot h$ est maintenant un produit matrice-vecteur. L'objet $J(x)$ dépend du choix de la base. L'expression ci-dessus peut être détaillée, de différentes manières. On peut l'écrire composante par composante

$$f_i(x + h) = f_i(x) + \sum_{j=1}^m \partial_{x_j} f_i(x) h_j + o(h),$$

ou de façon globale, avec e_i le i -ème vecteur de la base canonique de \mathbb{R}^m :

$$f(x + h) = f(x) + \sum_{i=1}^n \sum_{j=1}^m \partial_{x_j} f_i(x) h_j e_i + o(h),$$

Comme il a été précisé, l'existence de dérivées partielles en un point ne garantit pas la différentiabilité. En revanche (proposition V.1.8), si les dérivées partielles sont définies et continues sur un ouvert, alors la fonction est continûment différentiable sur cet ouvert.

Lorsque la fonction est à valeurs dans \mathbb{R} (cas $m = 1$), la matrice jacobienne est une matrice ligne, et l'application différentielle $df(x)$ est une forme linéaire. On peut alors écrire $df(x) \cdot h$ sous la forme d'un produit scalaire $\langle g | h \rangle$, où g est appelé *gradient* de f au point x , et noté $\nabla f(x)$. On a alors le développement

$$f(x + h) = f(x) + \langle \nabla f(x) | h \rangle + o(h).$$

Le vecteur $\nabla f(x)$ dépend du produit scalaire choisi. Lorsque ce choix n'est pas précisé, il s'agit du gradient associé au produit scalaire canonique sur l'espace euclidien \mathbb{R}^n . Lorsque l'on se place dans la base canonique de \mathbb{R}^n , que l'on considère muni du produit scalaire canonique, le vecteur $\nabla f(x)$ est représenté dans la base canonique par la matrice-ligne $J(x)$:

$$\nabla f(x) = (\partial_{x_1} f, \partial_{x_2} f, \dots, \partial_{x_n} f).$$

V.1.3 Théorème fondamental de l'analyse

Le théorème fondamental de l'analyse peut prendre plusieurs formes selon le sens que l'on donne à la notion d'intégrale. Le chapitre sur l'intégrale de Lebesgue montre que l'on peut définir cette intégrale pour des classes très générales de fonctions. Nous nous en limiterons ici à une définition plus classique de l'intégrale, en nous limitant à des fonctions continûment différentiables, de telle sorte que l'on n'aura besoin d'intégrer que des fonctions continues. On pourra donc s'en tenir à la notion d'intégrale de Riemann. L'objet de cette section est de généraliser au cas vectoriel la propriété portant sur les fonctions de \mathbb{R} dans \mathbb{R}^m : pour toute fonction f continûment dérivable sur $]a, b[$, à valeurs dans \mathbb{R}^m , pour tout x dans $]a, b[$, tout h tel que $x + h \in]a, b[$, on a

$$f(x + h) = f(x) + \int_x^{x+h} f'(s) ds.$$

Cette intégrale peut s'écrire différemment en introduisant la fonction $t \in [0, 1] \mapsto f(x + th)$, donc la dérivée en t est $f'(x + th)h$. On a

$$f(x + h) = f(x) + \int_0^1 f'(x + th)h dt.$$

Le théorème suivant généralise cette propriété aux fonctions de plusieurs variables.

Théorème V.1.18. Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m , continûment différentiable sur U , et h tel que le segment

$$[x, x + h] = \{x + \theta h, \theta \in [0, 1]\}$$

soit inclus dans U . On a alors

$$f(x + h) = f(x) + \int_0^1 df(x + th) \cdot h dt.$$

Démonstration. On introduit l'application Φ de $[0, 1]$ dans \mathbb{R}^m , définie par

$$\Phi : t \in [0, 1] \longmapsto \Phi(t) = f(x + th).$$

D'après la proposition V.1.9, cette application est continûment différentiable (on dira plus simplement *dérivable*, puisqu'il s'agit d'une fonction de \mathbb{R} dans \mathbb{R}^m), de dérivée

$$\Phi'(t) = df(x + th) \cdot h.$$

On a donc

$$f(x + h) - f(x) = \Phi(1) - \Phi(0) = \int_0^1 \Phi'(t) dt = \int_0^1 df(x + th) \cdot h dt,$$

qui est l'identité annoncée. \square

Ce théorème nous conduit naturellement au théorème des accroissements finis pour les fonctions de plusieurs variables, qui exprime un principe simple que l'on retrouve dans différents contextes⁹ : si l'on contrôle les variations d'une certaine quantité le long d'un chemin de longueur finie qui va de x vers $x + h$, alors on peut contrôler la différence des valeurs entre $x + h$ et x .

Théorème V.1.19. (des accroissements finis)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R}^m , continûment différentiable sur U , et h tel que le segment

$$[x, x + h] = \{x + \theta h, \theta \in [0, 1]\}$$

soit inclus dans U . Alors

$$\|f(x + h) - f(x)\| \leq \max_{t \in [0, 1]} \|df(x + th)\| \|h\|,$$

où $\|df(x + th)\|$ est la norme de l'application linéaire $df(x + th)$ de \mathbb{R}^n dans \mathbb{R}^m (norme subordonnée à la norme euclidienne, selon la définition I.8.6, page 32).

Démonstration. Notons en premier lieu que, la différentielle df étant continue sur le compact $[x, x + h]$, elle est bien bornée et atteint ses bornes, en particulier le max ci-dessus est bien défini comme un réel positif. On prend la norme de l'identité établie dans le théorème précédent : On a alors

$$\begin{aligned} \|f(x + h) - f(x)\| &= \left\| \int_0^1 df(x + th) \cdot h dt \right\| \leq \int_0^1 \|df(x + th) \cdot h\| dt \\ &\leq \int_0^1 \|df(x + th)\| \|h\| dt \leq \max_{t \in [0, 1]} \|df(x + th)\| \|h\|, \end{aligned}$$

qui est bien l'inégalité annoncée. \square

Exercice V.1.13. L'inégalité établie précédemment est-elle valide si l'on munit \mathbb{R}^n et \mathbb{R}^m d'autres normes que la norme euclidienne ?

CORRECTION.

Cette inégalité peut être étendue immédiatement à d'autres normes sur \mathbb{R}^n et \mathbb{R}^m , non nécessairement construites sur le même principe, sous réserve de munir l'espace des applications linéaires entre ces deux espaces de la norme subordonnée adaptée (pour le terme $\|df(x + th)\|$ dans l'inégalité des accroissements finis).

Exercice V.1.14. Soit f une application continûment différentiable sur un ouvert U , et $K \subset U$ un compact convexe. Montrer que f est Lipschitzienne sur K .

CORRECTION.

On applique le théorème V.1.19 (des accroissements finis), et l'on utilise le fait que la différentielle est bornée comme fonction continue sur un compact (proposition I.7.4, page 28).

9. On pourra penser à une version *Tour de France* de cette propriété très générale : si un coureur cycliste effectue un parcours de 10 km sur une route dont la pente n'excède pas 7 %, il sait qu'il n'aura pas monté en altitude de plus de $10 \text{ km} \times 0.07 = 700 \text{ m}$.

V.1.4 Formules de changement de variable pour les transformations régulières

Proposition V.1.20. Soient U un ouvert de \mathbb{R}^d et T un C^1 -difféomorphisme (définition V.3.7, page 123) entre U et $V \subset \mathbb{R}^d$.

(i) Pour tout borélien B de U

$$\lambda(T(B)) = \int_B |\det J_T| d\lambda,$$

où $|\det J_T|$ est la valeur absolue du déterminant de la matrice jacobienne de T .

(ii) Pour toute fonction f de V dans \mathbb{R} , mesurable, alors f est intégrable sur V si et seulement si $|\det J_T| f \circ T$ est intégrable sur U , et l'on a alors, sur tout borélien dans U ,

$$\int_{T(B)} f(y) d\lambda(y) = \int_B f(T(x)) |\det J_T| d\lambda.$$

V.2 Exercices

Exercice V.2.1. Soit f la fonction de \mathbb{R}^2 dans \mathbb{R} définie par

$$f(x_1, x_2) = \max(x_1, x_2).$$

a) Montrer que f est continue sur \mathbb{R}^2 .

b) Montrer que f est continûment différentiable sur

$$\{(x_1, x_2) \in \mathbb{R}^2, x_1 \neq x_2\},$$

et préciser sa différentielle et son gradient sur chaque composante de cet ensemble (de part et d'autre de la diagonale).

c) Montrer que f n'est pas différentiable sur $\{(x, x), x \in \mathbb{R}\}$.

CORRECTION.

a) On a

$$\max(x_1 + h_1, x_2 + h_2) \leq \max(x_1, x_2) + \max(|h_1|, |h_2|)$$

et

$$\max(x_1 + h_1, x_2 + h_2) \geq \max(x_1, x_2) - \max(|h_1|, |h_2|),$$

d'où

$$|\max(x_1 + h_1, x_2 + h_2) - \max(x_1, x_2)| \leq \max(|h_1|, |h_2|) \leq \|h\|_2,$$

d'où la continuité de f .

b) Pour $a = (x_1, x_2)$ tel que $x_1 < x_2$, cette inégalité stricte reste vérifiée dans un voisinage de x , la fonction f s'identifie donc dans ce voisinage à $x \mapsto x_2$, qui est différentiable, de différentielle dx_2 . Son gradient est le vecteur $(0, 1)$. De la même manière au voisinage d'un point tel que $x_2 < x_1$, le gradient est $(1, 0)$.

c) En un point du type (x, x) , pour tout $h > 0$, on a

$$f((x + h, x)) = f(x) + h,$$

et

$$f((x - h, x)) = f(x).$$

Il n'existe donc pas de développement limité à l'ordre 1 de f en tout point du type (x, x) .

Exercice V.2.2. (••) Soit $A \in \mathcal{M}(\mathbb{R}^{n \times n})$ une matrice carrée.

a) Montrer que la fonction

$$f : x \mapsto f(x) = \frac{1}{2} \langle Ax | x \rangle - \langle b | x \rangle \in \mathbb{R}, \quad b \in \mathbb{R}^n,$$

est différentiable, et préciser son gradient.

- b) Quelle forme prend ce gradient si A est symétrique ?
- c) Quels sont les points stationnaires de f ?

CORRECTION.

a) On a

$$f(x+h) = f(x) + \frac{1}{2} (\langle Ax | h \rangle + \langle Ah | x \rangle) - \langle b | h \rangle + o(h) = f(x) + \frac{1}{2} (\langle (A+A^T)x - b | h \rangle) + o(h),$$

d'où l'on déduit que f est différentiable, de différentielle

$$h \mapsto \frac{1}{2} (\langle (A+A^T)x - b | h \rangle),$$

et donc de gradient

$$\nabla f = \frac{1}{2}(A+A^T)x - b,$$

b) Le gradient devient $Ax - b$ pour une matrice symétrique.

c) Un point est critique pour f si et seulement s'il est solution du système linéaire $Ax = b$ (pour le cas symétrique). Cette propriété est utilisée pour approcher la solution de systèmes linéaires associés à des matrices symétriques définies positives. Pour résoudre le système on cherche un miniseur de la fonctionnelle quadratique f .

Exercice V.2.3. (Vecteur gaussien)

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice carrée, que l'on suppose *symétrique définie positive*, c'est-à-dire que $A^T = A$, et $\langle Ax | x \rangle > 0$ pour tout $x \neq 0$. On s'intéresse à la fonction qui représente (à constante de normalisation près) la loi d'un vecteur gaussien centré en a dans \mathbb{R}^n :

$$f(x) = \exp\left(-\frac{1}{2}\langle A \cdot (x-a) | x-a \rangle\right).$$

- a) Montrer que f est différentiable, et donner l'expression de son gradient.
- b) Quels sont les points stationnaires de f ?

CORRECTION.

a) On développe comme dans l'exercice précédent, pour obtenir (on prend $a = 0$ pour alléger l'écriture)

$$f(x+h) = f(x) + \exp\left(-\frac{1}{2}\langle A \cdot x | x \rangle\right) \langle Ax | h \rangle + o(h),$$

d'où

$$\nabla f = \exp\left(-\frac{1}{2}\langle A \cdot x | x \rangle\right) Ax.$$

b) Le seul point critique / stationnaire de f est donc 0 (ou plus généralement a si $a \neq 0$).

Exercice V.2.4. (Fonctions holomorphes)

Soit Ω un ouvert de \mathbb{C} . On dit qu'une fonction f de Ω dans \mathbb{C} est *holomorphe* sur Ω si, pour tout z de Ω , la quantité

$$\frac{f(z+h) - f(z)}{h}$$

admet une limite quand $h \in \mathbb{C}$ tend vers 0. On note alors $f'(z)$ cette valeur.

Montrer l'équivalence entre les deux assertions

- (i) La fonction f est dérivable sur Ω , et l'application $z \mapsto f(z)$ est continue.
- (ii) La fonction f vue comme application de \mathbb{R}^2 dans \mathbb{R}^2 est continûment différentiable, et l'on a sur Ω

$$\partial_x f + i \partial_y f = 0. \tag{V.2.1}$$

CORRECTION.

L'assertion (i) s'exprime

$$f(z+h) = f(x) + f'(z)h + o(h),$$

que l'on peut écrire (avec $z = x + iy$)

$$f(x+iy+h_x+ih_y) = f(x+iy) + f'(z)(h_x+ih_y) + o(h) = f(x+iy) + f'(z)h_x + if'(z)h_y + o(h).$$

L'application admet donc des dérivées partielles continues, qui s'expriment (on garde \mathbb{C} pour exprimer un vecteur de l'espace d'arrivée \mathbb{R}^2)

$$\partial_x f = f'(z), \quad \partial_y f = if'(z),$$

d'où $\partial_x f + i\partial_y f = f'(z) - f'(z) = 0$.

Réiproquement, si f est continûment différentiable, et vérifie la relation (V.2.1), les équations précédentes assurent que f admet un développement limité

$$f(z+h) = f(x) + f'(z)h + o(h),$$

en tout z de Ω , avec $f'(z) = \partial_x f = -i\partial_y f$.

Exercice V.2.5. (••) On se place sur \mathbb{R}^2 muni de la distance euclidienne. Pour un ensemble donné du plan $A \subset \mathbb{R}^2$, on considère la fonction définie par $f(x) = d(x, A)$ (distance du point x à l'ensemble A).

a) Préciser les zones de différentiabilité de f lorsque A est (i) un singleton, (ii) une paire de points distincts, (ii) un cercle, (iii) un disque, (iv) un rectangle, (v) une forme “quelconque”...

b) (*) On associe à chaque grande ville de France (pour fixer les idées on pourra imaginer les 20 plus grandes villes par exemple) un point (son barycentre), et l'on appelle A l'ensemble de ces points. Que peut-on dire des points de non différentiabilité de la fonction f définie ci-dessus ?

CORRECTION.

a) Pour un singleton $a \in \mathbb{R}^2$ la fonction est continûment différentiable sur l'ouvert $\mathbb{R}^2 \setminus \{a\}$.

Pour une paire de points, la fonction distance n'est pas différentiable sur la médiatrice, continûment différentiable sur son complémentaire.

Pour un cercle C de centre c , la fonction distance est différentiable sur l'ouvert $\mathbb{R}^2 \setminus (\{c\} \cup C)$.

Pour un disque, la fonction est différentiable sur le complémentaire du disque fermé, différentiable à l'intérieur du disque (car constante égale à 0), non différentiable sur le cercle frontière.

Pour un rectangle la fonction est continûment différentiable sur l'extérieur du rectangle, non différentiable sur la frontière, et différentiable à l'intérieur en dehors du “squelette” du rectangle : bissectrices des angles jusqu'à leur point de croisement, et segment reliant les deux points de croisement (en forme d'enveloppe postale).

b) L'ensemble des points de non différentiabilité est l'ensemble des points qui sont situés à équidistance de 2 centres urbains. Il s'agit de points qui ont une distance importante à la réunion des centres par rapport à la moyenne, on trouvera en particulier parmi ces points le point le plus “isolé” (le point le plus éloigné de toute ville). Il s'agit aussi de points pour lesquels on aura typiquement le choix entre deux hôpitaux de centre ville, plus généralement pour tout service restreint aux grandes agglomérations.

Exercice V.2.6. La figure V.2.1 représente les isovaleurs de la fonction altitude pour une certaine zone géographique, que l'on peut considérer comme une fonction de \mathbb{R}^2 dans \mathbb{R} .

a) Localiser des points critiques de cette fonction f (c'est à dire le point en lesquels le gradient s'annule), et décrire la forme de la fonction f au voisinage de ces points. Proposer des fonctions polynomiales qui vous paraissent de nature à reproduire la forme de la fonction au voisinage du point critique, dans les différents cas.

b) (*) Comment caractériser les zones correspondant aux lacs ?

c) (*) Comment peut-on caractériser le bassin d'attraction d'un lac, c'est à dire l'ensemble des x tels qu'une goutte d'eau tombée en x va alimenter le lac en question ?

d) (*) Le nombre de lacs peut-il augmenter ou diminuer en fonction de la pluviométrie ?

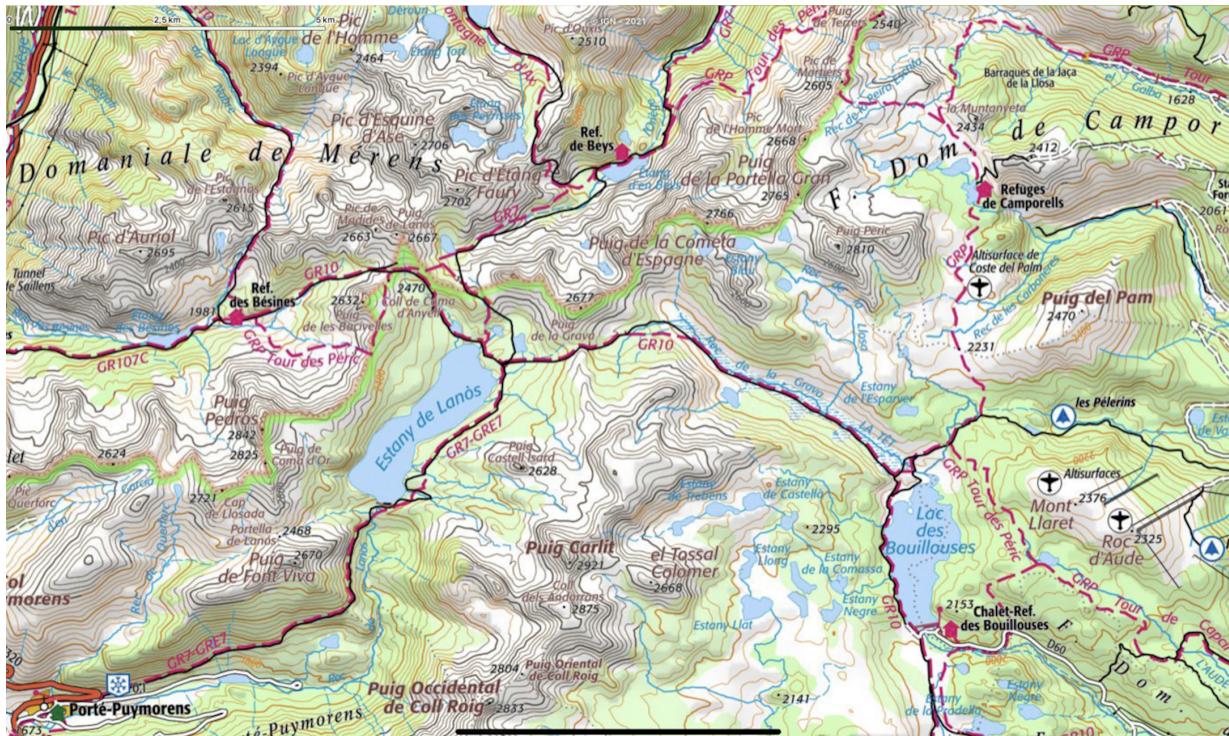


FIGURE V.2.1 – Isovaleurs de l'altitude sur une zone des Pyrénées

CORRECTION.

a) La plupart des points critiques visibles correspondent à des maxima locaux (sommets ou pointes locales). La plupart des minima locaux semblent "cachés" par les zones bleues représentant des lacs. On peut aussi identifier quelques points-cols (minimum dans une direction, maximum dans une autre), en particulier si l'on trace le chemin que l'on emprunterait pour relier deux maxima locaux en se descendant le moins possible possible (comme dans la zone en bas à gauche de l'image, au nord-est de l'indication "...morens"), on passe par un minimum local qui est un point col.

Les minima locaux, maxima locaux, et points-cols peuvent être approché par des fonctions de types respectifs

$$f(x, y) = \frac{x^2}{a^2} + \frac{y^2}{b^2}, \quad f(x, y) = -\frac{x^2}{a^2} - \frac{y^2}{b^2}, \quad f(x, y) = \frac{x^2}{a^2} - \frac{y^2}{b^2},$$

respectivement.

b) Pour qu'il y ait un lac il faut qu'il y ait un minimum local de l'altitude. Le contour du lac correspond à une isovaleur de l'altitude, qui doit être une courbe fermée (qui boucle sur elle-même) associée à une valeur supérieure au minimum, qui délimite une zone qui contient ce minimum. Pour un lac sans île, les valeurs d'altitude prises dans la zone sont inférieures à la valeur du contour.

c) Pour chaque point x_0 de la zone, on peut considérer l'équation différentielle (appelée *flot de gradient*) $dx/dt = -\nabla f$, qui suit la trajectoire selon la ligne de plus grande pente. Les x_0 tels que cette trajectoire arrive à un lac constituent le bassin d'attraction de ce lac. En tout généralité, ce bassin peut prendre des formes complexes, qui peuvent en particulier contenir des "trous", on pourrait par exemple imaginer une petite colline entourée d'un lac (que l'on peut voir comme une île sur ce lac), qui contiendrait elle-même un petit lac sur-élevé par rapport au lac périphérique. Il n'y en a pas sur la carte proposée, mais on trouve ce type de bizarrerie en Finlande par exemple (on trouve même une île sur un lac qui contient elle-même un lac, qui contient lui-même une île ...).

d) Le nombre de lacs peut varier en fonction de la pluviométrie. Le nombre maximal de lacs possibles est le nombre de minima locaux, au voisinage desquels émerge un lac en cas de première pluie sur un paysage sec. Mais le nombre de lacs peut diminuer alors que leur taille augmente. Le grand lac allongé estany de Lanòs,

sur la carte, est par exemple susceptible de contenir plusieurs minimina locaux, qui constitueront autant de petits lacs en cas de sécheresse.

Exercice V.2.7. (Taux de déformation (••))

On considère un champ de vitesses (on pourra penser à la vitesse instantanée d'un fluide occupant une partie de l'espace physique)

$$x = (x_1, x_2, x_3) \longmapsto u(x) = (u_1(x), u_2(x), u_3(x)).$$

On suppose u différentiable en un point x d'un ouvert U de \mathbb{R}^3 .

a) Montrer que l'on peut écrire le champ de vitesse au point $x + h$ voisin de x de la façon suivante

$$u(x + h) = u(x) + \omega \wedge h + D \cdot h + o(h), \quad (\text{V.2.2})$$

où ω est un vecteur de \mathbb{R}^3 , et $D \in \mathcal{M}_3(\mathbb{R})$ une matrice symétrique.

b) Justifier l'appellation *matrice des taux de déformation* utilisée pour désigner la matrice D .

c)(*) On dit qu'un écoulement est incompressible si $\partial_1 u_1 + \partial_2 u_2 + \partial_3 u_3 = 0$ (i.e. si ce qu'on appelle la *divergence* de u est nulle). Montrer que si l'écoulement est incompressible sur U , alors la somme des valeurs propres de la matrice D associée à tout point de U est égale à 0, et interprétez physiquement cette propriété.

d) Donner un exemple de champ de vitesses u non trivial défini sur \mathbb{R} tel qu'en tout point, la décomposition (V.2.2) soit telle que $D = 0$. (On pourra pour simplifier chercher un champ qui soit invariant par translation dans la direction verticale, de façon à se ramener à un champ bidimensionnel).

e) Donner un exemple de champ de vitesses u non trivial défini sur \mathbb{R} tel qu'en tout point, la décomposition (V.2.2) soit telle que $\omega = 0$ (champ irrotationnel).

f) (*) Que peut-on dire, au vu de ce qui précède, d'un champ qui dérive d'un potentiel, c'est-à-dire un champ qui s'écrit $u = -\nabla\Phi$, où Φ est une fonction scalaire suffisamment régulière pour que u soit différentiable ?

CORRECTION.

a) *On peut écrire*

$$u(x + h) = u(x) + du(x) \cdot h + o(h) = u(x) + J(x) \cdot h + o(h) = u(x) + \frac{1}{2}(J - J^T) \cdot h + \frac{1}{2}(J + J^T) \cdot h + o(h).$$

La matrice antisymétrique $J - J^T$ peut s'écrire

$$J - J^T = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}, \quad \omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}$$

de telle sorte que le produit matrice-vecteur $J \cdot u$ corresponde au produit vectoriel $\omega \wedge u$. Les ω_i dépendent des dérivées partielles des composantes de u , on a par exemple

$$\omega_2 = \partial_3 u_1 - \partial_1 u_3.$$

La matrice $D = (J + J^T)/2$ est symétrique par construction.

b) *La matrice $J = J(x)$ étant symétrique, elle est diagonalisable dans une base orthonormée. Si l'on se place dans cette base (e_1, e_2, e_3) , la matrice s'écrit $D = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$. Considérons le développement limité ci-dessus pour $y = x + he_1$. Le premier terme signifie que la vitesse en y est à l'ordre 0 la même que celle en x . Le second terme indique une rotation instantanée de y autour de x , donc sans changement de la distance de x à y . Le troisième terme s'écrit $hD \cdot e_1 = \lambda_1 he_1$. Si par exemple $\lambda_1 > 0$, cela signifie que le point matériel situé à l'instant considéré en y s'éloigne de x dans la direction e_1 . Il s'agit donc du terme qui modifie les distances entre les points, ce qui correspond à une déformation du milieu fluide, les deux premiers termes encodant un mouvement rigide instantané.*

La figure V.2.2 illustre en dimension 2 la décomposition local du champ de vitesse en ces trois composantes :

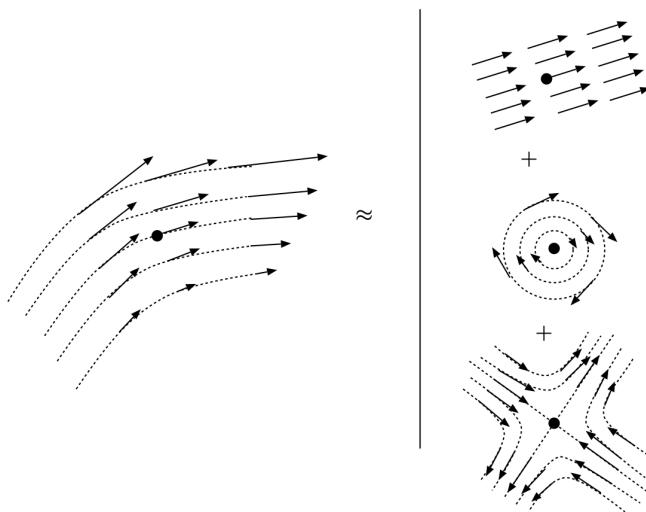


FIGURE V.2.2 – Décomposition locale d'un champ de vitesse

translation, rotation, et déformation

c) Si la divergence $\partial_1 u_1 + \partial_2 u_2 + \partial_3 u_3$ est nulle au point considéré, cela implique que la trace de la matrice D est nulle, donc que la somme des valeurs propres est nulle. Si elles sont toutes nulles, le mouvement instantané est rigide. Si ça n'est pas le cas, elle ne peuvent pas être toutes de même signe, il y en a donc au moins > 0 (étirement dans la direction correspondante), et au moins une < 0 (compression dans cette direction).

d) On peut par exemple considérer un mouvement (rigide) de rotation autour de l'axe vertical, caractérisé par le champ de vitesses

$$u = (u_1, u_2, u_3) = (-x_2, x_1, 0),$$

pour lequel on peut vérifier que la matrice D des taux de déformation est nulle.

e) Le champ suivant peut être vu comme la rencontre de deux masses de fluides qui "se rencontrent" sur le plan $x_1 = 0$ (on pourra faire un dessin dans le plan (x_1, x_2))

$$u = (u_1, u_2, u_3) = (x_1, -x_2, 0).$$

f) Si le champ dérive d'un potentiel, $u = -\nabla\Phi$ avec Φ régulier, alors les composantes du vecteur ω associé, qui sont du type $\partial_i u_j - \partial_j u_i$, sont toutes nulles.

Exercice V.2.8. (Potentiel d'interaction ($\bullet \bullet \bullet$))

On considère la fonction $D(\cdot)$ qui à $q = (q_1, q_2) \in \mathbb{R}^3 \times \mathbb{R}^3$ (attention, q_1 et q_2 désignent ici des points de \mathbb{R}^3) associe la distance entre les points q_1 et q_2 de \mathbb{R}^3 :

$$D(q) = D(q_1, q_2) = \|q_2 - q_1\|.$$

a) Montrer que $D(\cdot)$ est différentiable sur l'ouvert

$$U = \{ q = (q_1, q_2) \in \mathbb{R}^6, q_1 \neq q_2 \},$$

et exprimer son gradient (on pourra exprimer les gradients partiels ∇_{q_1} et ∇_{q_2}) en fonction du vecteur unitaire $e_{12} = (q_2 - q_1) / \|q_2 - q_1\|$.

b) On introduit un potentiel d'interaction sur le système de deux particules localisées en q_1 et q_2 sous la forme $V = V(q) = V(q_1, q_2) = \varphi(D(q_1, q_2))$, où φ est une fonction continûment dérivable de \mathbb{R}_+ dans \mathbb{R} . Montrer que V est différentiable sur U , et écrire son gradient.

c) Préciser les gradients partiels $\nabla_{q_1} V$ et $\nabla_{q_2} V$ si l'on prend pour φ le potentiel d'interaction gravitationnelle défini par $\varphi(D) = -1/D$.

d) On se replace dans le cas général d'un potentiel φ quelconque, et l'on considère maintenant un système de N particules dans \mathbb{R}^3 . On définit un potentiel d'interaction global de la façon suivante

$$V(q) = V(q_1, q_2, \dots, q_N) = \sum_{1 \leq i < j \leq N} \varphi(D(q_i, q_j)).$$

On s'intéresse au système résultant du principe fondamental de la dynamique, sous l'hypothèse de forces dérivant d'un potentiel (on prend des masses unitaires), c'est-à-dire

$$\frac{d^2 q}{dt^2} = -\nabla V(q). \quad (\text{V.2.3})$$

Écrire l'équation qui résulte de ce principe pour chacune des particules, qui s'écrit de façon abstraite

$$\frac{d^2 q_i}{dt^2} = -\nabla_{q_i} V(q).$$

e)(*) On se place dans le cadre des notations de la question précédente. On suppose que l'on connaît une solution $t \in [0, T[\mapsto q(t) \in U$ de l'équation d'évolution (V.2.3). Montrer que l'on a conservation de l'énergie totale, c'est à dire que la quantité

$$E(t) = \sum_{i=1}^N \frac{1}{2} \left\| \frac{dq_i}{dt}(t) \right\|^2 + V(q(t))$$

est constante sur $[0, T[$.

Dans le cas du potentiel gravitationnel $\varphi(D) = -1/D$, peut-on en déduire que les vitesses sont majorées sur $[0, T[$? Même question pour le cas du potentiel coulombien entre charges identiques $\varphi(D) = 1/D$.

CORRECTION.

a) La fonction D est régulière comme composée de fonctions régulières, en dehors des points qui annulent la norme, c'est à dire le complémentaire de la diagonale. Pour calculer son gradient, on considère $D(q_1, q_2)$ comme fonction de q_1 seulement. Si l'on se déplace de ε petit de 1 vers 2, la distance diminue de ε :

$$D(q_1 + h e_{12}, q_2) = D(q_1, q_2) - h,$$

pour h petit. Si l'on perturbe q_1 selon une direction orthogonale à e_{12} , la variation de la distance est d'ordre 2, la dérivée partielle est donc nulle. En procédant de même pour q_2 (en prenant garde au fait que si l'on se déplace de ε dans la direction e_{12} , la distance augmente de ε), on obtient

$$\nabla_{q_1} = -e_{12}, \quad \nabla_{q_2} = e_{12}.$$

b) Le potentiel est différentiable sur U comme composé de fonctions différentiables, et l'on a

$$\nabla_{q_1} V(q) = \varphi'(D) \nabla_{q_1} D = -\varphi'(D) e_{12},$$

et l'opposé pour $\nabla_{q_2} V(q)$.

c) Dans le cas du potentiel d'interaction gravitationnel, on obtient

$$\nabla_{q_1} V(q) = -\varphi'(D) e_{12} = -\frac{1}{D^2} e_{12},$$

et l'opposé pour $\nabla_{q_2} V(q)$. On retrouve bien le fait que, pour la force gravitationnelle qui dérive de ce potentiel (c'est à dire que la force est l'opposé du gradient du potentiel), on a une force d'attraction mutuelle proportionnelle

à l'inverse du carré de la distance.

d) L'équation s'écrit, pour chaque particule i ,

$$\frac{d^2q_i}{dt^2} = -\nabla_{q_i} V(q) = \sum_{j \neq i} \varphi'(D(q_i, q_j)) e_{ij}$$

e) On a

$$\frac{dE}{dt} = \sum_{i=1}^N \left\langle \frac{dq_i}{dt} \mid \frac{d^2q_i}{dt^2} \right\rangle + \left\langle \nabla V \mid \frac{dq}{dt} \right\rangle,$$

qui s'annule du fait que $d^2q/dt^2 = -\nabla V$.

Dans le cas du potentiel gravitationnel, qui n'est pas borné inférieurement, cela n'implique aucunement que la vitesse soit bornée¹⁰. En revanche si le potentiel est minoré, comme pour le potentiel électrostatique entre deux charges identiques, la conservation de l'énergie assure que l'énergie cinétique, et donc la norme de la vitesse, sont majorées. Noter aussi symétriquement que, l'énergie cinétique étant positive, la conservation de l'énergie assure que l'énergie potentielle est majorée, et qu'ainsi les masses ne peuvent pas se rapprocher en-dessous d'un certain seuil¹¹.

Exercice V.2.9. (Densité gaussienne)

a) Calculer

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} d\lambda(x, y) = \int_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy.$$

(On pourra utiliser les coordonnées polaires sur \mathbb{R}^2 .)

b) Montrer que l'intégrale ci-dessus peut s'exprimer simplement en fonction de $\int_{\mathbb{R}} e^{-x^2} dx$.

c) Pour $\sigma > 0$ fixé, on considère l'application

$$x = (x_1, \dots, x^d) \mapsto f(x) = e^{-\frac{x_1^2+\dots+x_d^2}{2\sigma^2}}.$$

Calculer le coefficient C assurant que la mesure associée à la fonction Cf , vue comme une densité, soit de masse totale égale à 1.

CORRECTION.

a) On considère

$$T : (r, \theta) \in U =]0, +\infty[\times]-\pi, \pi[\mapsto (r \cos \theta, r \sin \theta) \in V = \mathbb{R}^2 \setminus [0, +\infty[\times \{0\}.$$

Il s'agit d'un C^1 -difféomorphisme entre 2 ouverts de \mathbb{R}^2 , et l'intégrale de la fonction f sur V est la même que sur \mathbb{R}^2 , car ces ensemble ne diffèrent que par un ensemble de mesure nulle. La formule de changement de variable s'écrit ici

$$I = \int_V e^{-(x^2+y^2)} dx dy = \int_0^{+\infty} \int_{-\pi}^{\pi} e^{-((r \cos \theta)^2 + (r \sin \theta)^2)} |\det J_T| dr d\theta = \int_0^{+\infty} \int_{-\pi}^{\pi} e^{-r} |\det J_T| dr d\theta,$$

avec

$$\det J_T = \det \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & +r \cos \theta \end{pmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r,$$

d'où

$$I = 2\pi \int_0^{+\infty} r e^{-r^2} r dr = 2\pi [-e^{-r^2}/2]_0^{+\infty} = \pi.$$

10. Si la lune cessait brusquement de tourner autour de la terre, elle tomberait vers la terre avec une vitesse d'impact telle que l'énergie cinétique de la lune correspondrait à la variation d'énergie potentielle entre la position initiale et la configuration de contact.

11. Cette propriété permet de montrer l'existence d'une solution globale au système dynamique, qui est cantonnée à rester à une certaine distance des points de non-différentiabilité du second membre.

b) On a, d'après le théorème de Fubini

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-(x^2+y^2)} dy \right) dx = \int_{\mathbb{R}} e^{-x^2} \left(\int_{\mathbb{R}} e^{-y^2} dy \right) dx = \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2.$$

On en déduit

$$\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}.$$

c) En dimension d , on a

$$I = \int_{\mathbb{R}^d} f(x) dx = \int_{\mathbb{R}^d} e^{-\frac{x_1^2+\dots+x_d^2}{2\sigma^2}} dx$$

Le changement de variable homothétique $T : y \mapsto x = \sqrt{2}\sigma y$, de jacobien constant $2^{d/2} \sigma^d$, donne

$$I = 2^{d/2} \sigma^d \int_{\mathbb{R}^d} e^{-(y_1^2+\dots+y_d^2)} dy = 2^{d/2} \sigma^d \left(\int_{\mathbb{R}^d} e^{y^2} dy \right)^d = 2^{d/2} \sigma^d \pi^{d/2} = (2\sigma^2\pi)^{d/2}.$$

D'où la valeur de la constante de normalisation, et le fait que

$$\frac{1}{(2\sigma^2\pi)^{d/2}} e^{-\frac{x_1^2+\dots+x_d^2}{2\sigma^2}}$$

soit une mesure de probabilité (i.e. de masse totale égale à 1).

V.3 Théorèmes des fonctions implicites et d'inversion locale

Le résultat principal de cette section est le théorème dit *des fonctions implicites*, que l'on peut interpréter comme suit. On considère une équation portant sur $y \in \mathbb{R}^m$, équation qui dépend de paramètres x_1, \dots, x_n , et que l'on écrit

$$f(x, y) = 0.$$

Cette équation est à valeurs vectorielles. Pour se placer dans un contexte où l'équation, pour un jeu de paramètres x fixé, peut permettre de déterminer y , on s'intéresse au cas où il y a autant d'équations que d'inconnues, c'est à dire que f est à valeurs dans \mathbb{R}^m . L'inconnue y est donc définie de façon *implicite* par rapport aux paramètres x_1, \dots, x_n . On se place au voisinage d'une solution de cette équation : pour un jeu de paramètres $x = (x_1, \dots, x_n)$ donné, on suppose connue une solution $y = (y_1, \dots, y_m)$ de l'équation. Si l'on fait varier les paramètres de l'équation, on peut s'attendre à ce que, sous certaines conditions, la solution en y varie elle-même de façon régulière. Le théorème ci-dessous donne des conditions suffisantes pour que l'on puisse en effet exprimer y en fonction de x , de façon régulière, au voisinage d'un couple paramètres - solution (x_0, y_0) donné. La condition principale permettant cette explicitation de la dépendance apparaît clairement dans l'exemple-jouet suivant :

$$f : (x, y) \in \mathbb{R}^2 \mapsto ax + by + c.$$

On peut exprimer y fonction de x si et seulement si $b \neq 0$, où b quantifie la manière dont f varie vis-à-vis de y . Dans le cas le plus général ($y \in \mathbb{R}^m$, f à valeurs dans \mathbb{R}^m), cette dépendance sera encodée par la différentielle de f par rapport à y (qui est bien représentée dans la base canonique par une matrice carrée). L'hypothèse principale porte sur le caractère *inversible* de cette différentielle.

Théorème V.3.1. (••) Soit f une fonction définie sur un ouvert W de $\mathbb{R}^n \times \mathbb{R}^m$, à valeurs dans \mathbb{R}^m . On suppose f continûment différentiable sur W , et l'on suppose que la différentielle partielle de f par rapport à y , notée $\partial_y f(x, y)$, est inversible en tout point¹² de W . On considère un point $(x_0, y_0) \in W$ qui annule f :

$$f(x_0, y_0) = 0.$$

12. Comme précisé dans la remarque V.3.3 ci-après, il suffit de vérifier que la différentielle soit inversible en (x_0, y_0) pour qu'elle le soit dans un voisinage de ce point.

On peut alors exprimer y comme fonction de x au voisinage de (x_0, y_0) . Plus précisément : il existe des voisinages ouverts $U \in \mathbb{R}^n$ et $V \in \mathbb{R}^m$ de x_0 et y_0 , respectivement, et une fonction Ψ de U dans V , tels que

$$(x, y) \in U \times V, \quad f(x, y) = 0 \iff y = \Psi(x).$$

La fonction Ψ est continûment différentiable sur U , et sa différentielle s'exprime

$$d\Psi(x) = -(\partial_y f(x, y))^{-1} \circ \partial_x f(x, y), \quad \text{avec } y = \Psi(x).$$

Démonstration. La démarche, de nature constructive, est basée sur un processus itératif construit selon les principes suivants. On considère x proche de x_0 (dans un sens précisé plus loin), et l'on cherche y tel que $f(x, y) = 0$. On suppose que l'on dispose d'une première approximation y_k du y recherché, et on cherche un y_{k+1} qui en soit une meilleure approximation. On a

$$f(x, y_{k+1}) = f(x, y_k + (y_{k+1} - y_k)) \approx f(x, y_k) + \partial_y f(x, y_k) \cdot (y_{k+1} - y_k).$$

On souhaite annuler cette quantité, ce qui suggère de définir y_{k+1} comme

$$y_{k+1} = y_k - (\partial_y f(x, y_k))^{-1} \cdot f(x, y_k).$$

Il s'agit de la méthode dite de *Newton* pour trouver le zéro d'une fonction. Nous allons considérer ici une version modifiée de cette méthode, en remplaçant la différentielle partielle en y par sa valeur au point (x_0, y_0) . Partant de y_0 (en fait, on peut partir d'une valeur initiale différente de y_0 , mais nous le fixons comme point de départ pour simplifier), on construit donc la suite (y_k) par récurrence, selon la formule

$$y_{k+1} = y_k - Q^{-1} \cdot f(x, y_k), \quad \text{avec } Q = \partial_y f(x_0, y_0).$$

N.B. : On prendra garde au fait que, pour (x, y) donné, $\partial_y f(x, y)$ est une application linéaire de \mathbb{R}^m dans \mathbb{R}^m . Cette application dépend du point $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ où elle est prise, mais sans que la différentielle soit prise par rapport à la variable x . Cette différentielle partielle est définie par le développement limité suivant, où l'on ne perturbe que la variable y : pour $h \in \mathbb{R}^m$,

$$f(x, y + h) = f(x, y) + \partial_y f(x, y) \cdot h + o(h).$$

Il s'agit donc d'un champ d'applications linéaires, auquel on peut associer un champ de matrices carrées $m \times m$ (leurs représentations dans la base canonique de \mathbb{R}^m), qui vit sur un espace de dimension $n \times m$. L'application Q est simplement la valeur particulière de ce champ au point (x_0, y_0) .

Cette récurrence peut s'écrire $y_{k+1} = \Phi_x(y_k)$, où la fonction Φ_x est définie par

$$y \mapsto \Phi_x(y) = y - Q^{-1} \cdot f(x, y),$$

pour tout y tel que $(x, y) \in W$. Noter que y est point fixe de Φ_x si et seulement si $f(x, y) = 0$. Nous allons montrer que cette fonction admet bien un unique point fixe sur un voisinage de y_0 . Cette fonction est différentiable sur son domaine de définition, de différentielle

$$d\Phi_x(y) = I - Q^{-1} \circ \partial_y f(x, y).$$

En écrivant $I = Q^{-1}Q$ on obtient

$$\|d\Phi_x(y)\| = \|Q^{-1}(\partial_y f(x_0, y_0) - \partial_y f(x, y))\| \leq \|Q^{-1}\| \|\partial_y f(x_0, y_0) - \partial_y f(x, y)\|$$

Fixons $\kappa = 1/2$. La différentielle étant continue, il existe un $r > 0$ tel que, pour tout point $x \in \overline{B}(x_0, r)$, tout $y \in \overline{B}(y_0, r)$ (on prend r suffisamment petit pour que $\overline{B}(x_0, r) \times \overline{B}(y_0, r) \subset W$),

$$\|\partial_y f(x_0, y_0) - \partial_y f(x, y)\| \leq \kappa \|Q^{-1}\|^{-1},$$

de telle sorte que

$$\forall x \in \overline{B}(x_0, r), \quad \forall y \in \overline{B}(y_0, r), \quad \|d\Phi_x(y)\| \leq \kappa.$$

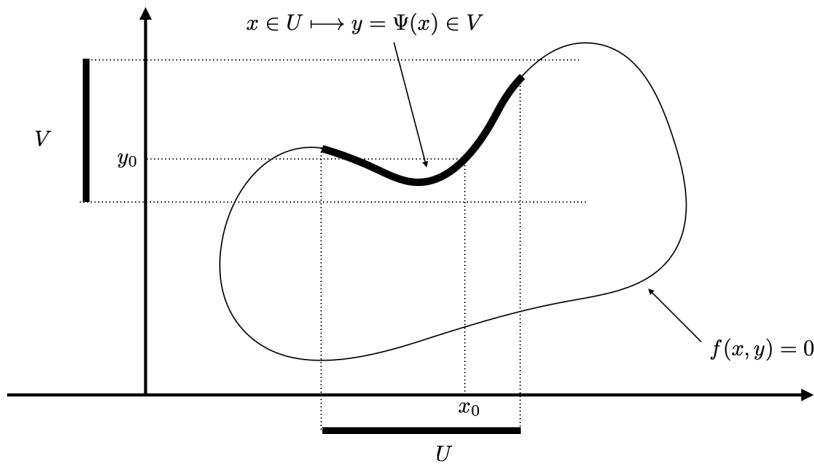


FIGURE V.3.1 – Théorème des fonctions implicites

On a donc, pour tous y, y' dans $\overline{B}(y_0, r)$,

$$\|\Phi_x(y) - \Phi_x(y')\| \leq \kappa \|y - y'\|$$

d'après le théorème des accroissements finis (théorème V.1.19, page 110), avec $\kappa = 1/2$. L'application Φ_x est donc contractante sur $\overline{B}(y_0, r)$. Montrons qu'elle laisse stable une boule autour de y_0 . Comme l'application

$$x \mapsto \Phi_x(y_0) = y_0 - Q^{-1} \cdot f(x, y_0)$$

est continue en x_0 , il existe un $r' < r$ tel que, pour tout $x \in \overline{B}(x_0, r')$, on ait

$$\|\Phi_x(y_0) - \Phi_{x_0}(y_0)\| \leq (1 - \kappa)r,$$

avec $\Phi_{x_0}(y_0) = y_0$ car $f(x_0, y_0) = 0$. On a alors, pour tout $x \in \overline{B}(x_0, r')$, tout $y \in \overline{B}(y_0, r)$,

$$\|\Phi_x(y) - y_0\| = \|\Phi_x(y) - \Phi_{x_0}(y_0)\| \leq \underbrace{\|\Phi_x(y) - \Phi_x(y_0)\|}_{\leq \kappa \|y - y_0\|} + \underbrace{\|\Phi_x(y_0) - y_0\|}_{\leq (1 - \kappa)r} \leq \kappa r + (1 - \kappa)r = r.$$

Pour tout $x \in \overline{B}(x_0, r')$, l'application Φ_x est donc bien définie de $\overline{B}(y_0, r)$ dans lui-même, et cet ensemble est complet comme fermé dans le complet \mathbb{R}^m . Elle par ailleurs contractante comme montré précédemment. D'après le théorème I.8.2, elle admet donc un unique point fixe sur $\overline{B}(y_0, r)$, c'est-à-dire qu'il existe un unique $y \in \overline{B}(y_0, r)$ tel que $f(x, y) = 0$. On note Ψ l'application qui à x associe cette unique solution en y de $f(x, y) = 0$.

Montrons maintenant la continuité de Ψ , et précisons le choix des voisinages U et V . Soient x_1 et x_2 deux points de $\overline{B}(x_0, r')$, et $y_1 = \Psi(x_1)$, $y_2 = \Psi(x_2)$. On a

$$\|y_2 - y_1\| = \|\Phi_{x_2}(y_2) - \Phi_{x_1}(y_1)\| \leq \|\Phi_{x_2}(y_2) - \Phi_{x_2}(y_1)\| + \|\Phi_{x_2}(y_1) - \Phi_{x_1}(y_1)\|.$$

Comme Φ_{x_2} est κ -contractante sur $\overline{B}(y_0, r)$, on a $\|\Phi_{x_2}(y_2) - \Phi_{x_2}(y_1)\| \leq \kappa \|y_2 - y_1\|$, d'où

$$\begin{aligned} \|y_2 - y_1\| &\leq \frac{1}{1 - \kappa} \|\Phi_{x_2}(y_1) - \Phi_{x_1}(y_1)\| = \frac{1}{1 - \kappa} \|Q^{-1} \cdot (f(x_2, y_1) - f(x_1, y_1))\| \\ &\leq \frac{1}{1 - \kappa} \|Q^{-1}\| \max_{\overline{B}(x_0, r') \times \overline{B}(y_0, r)} \|\partial_x f\| \|x_2 - x_1\| \end{aligned}$$

d'après le théorème des accroissements finis V.1.19 (f étant continûment différentiable sur le compact $\overline{B}(x_0, r') \times \overline{B}(y_0, r)$, sa différentielle partielle par rapport à x est bornée). Cette quantité tend en particulier vers 0 quand x_2 tend vers x_1 . L'application Ψ est donc continue sur $\overline{B}(x_0, r')$ à valeurs dans $\overline{B}(y_0, r)$, et même lipschitzienne : il existe $C > 0$ tel que

$$\|\Psi(x_2) - \Psi(x_1)\| \leq C \|x_2 - x_1\|.$$

Soit V voisinage ouvert de y_0 inclus dans $\overline{B}(y_0, r)$. Comme Ψ est continue, il existe un voisinage ouvert de x_0 , $U \subset \overline{B}(x_0, r')$, tel que $\Psi(U) \subset V$.

Il reste à montrer que Ψ est différentiable sur U . Soit $x \in U$, $y = \Psi(x) \in V$. On considère une variation h de x telle que $x + h \in U$. Il existe un unique g tel que $y + g \in V$ vérifie

$$f(x + h, y + g) = 0.$$

D'après ce qui précède il existe $C > 0$ tel que $\|g\| \leq C \|h\|$. La différentiabilité de f en (x, y) s'exprime

$$\underbrace{f(x + h, y + g)}_{=0} = \underbrace{f(x, y)}_{=0} + \partial_x f(x, y) \cdot h + \partial_y f(x, y) \cdot g + o(h, g).$$

On a donc

$$g = -\left((\partial_y f(x, y))^{-1} \circ \partial_x f(x, y)\right) \cdot h + o(h),$$

(le $o(h, g)$ s'est bien transformé en $o(h)$ du fait que la norme de h domine celle de g , comme indiqué précédemment). Ce g est, par construction, $\Psi(x + h) - \Psi(x)$, on a donc

$$\Psi(x + h) = \Psi(x) - \left((\partial_y f(x, y))^{-1} \circ \partial_x f(x, y)\right) \cdot h + o(h).$$

ce qui exprime que l'application $x \mapsto \Psi(x)$ est différentiable sur U , de différentielle

$$d\Psi(x) = (\partial_y f(x, \Psi(x)))^{-1} \circ \partial_x f(x, \Psi(x)).$$

Comme f est continûment différentiable, et que Ψ est continue, $x \mapsto d\Psi(x)$ est continue. \square

Remarque V.3.2. On notera que, par construction, Ψ est bien définie sur tout U mais, comme illustré par la figure V.3.1, elle n'est pas nécessairement surjective (cette remarque sera importante pour la démonstration du théorème des fonctions implicites, dans lequel il s'agira de construire deux ouverts en bijection). Par ailleurs, pour $x \in U$, il peut exister plusieurs y tels que $(f(x, y) = 0$, mais un seul qui soit dans V .

Remarque V.3.3. Pour vérifier l'applicabilité du théorème précédent en un point (x_0, y_0) qui annule f , et au voisinage duquel f est définie, il suffit de vérifier que la différentielle de f par rapport à y est inversible en (x_0, y_0) . En effet, si c'est le cas, l'application $(x, y) \mapsto \partial_y f(x, y)$ étant continue, et le déterminant étant une fonction continue, la différentielle reste inversible sur un ouvert de (x_0, y_0) , qui peut jouer le rôle du W dans les hypothèses du théorème précédent. On dira que le théorème des fonctions implicites s'applique *en* (x_0, y_0) , ou *au voisinage de* (x_0, y_0) .

Remarque V.3.4. Ce théorème, qui peut sembler assez abstrait et technique, peut être invoqué d'une manière *négative* pour qualifier la pertinence d'un modèle. Replaçons-nous dans le cadre de l'introduction, en interprétant $f(x, y)$ comme un *modèle* portant sur y , sous la forme d'un système d'équations dépendant de paramètres x_1, \dots, x_n . Le modèle a vocation à, pour un jeu de paramètres (qui peuvent être des températures, des pressions, des flux d'information, des prix, ...), déterminer la collection des inconnues y_1, \dots, y_m . Dans le cadre d'une utilisation de ce modèle dans la vie réelle, les paramètres ne sont en général connus qu'approximativement (erreurs de mesure, variabilité en temps de paramètres supposés statiques, ...). Si la solution y ne dépend *pas* de façon régulière des paramètres, cela signifie qu'une erreur petite sur les paramètres peut induire une variation très importante de la solution. On dira que le problème n'est pas *stable*¹³. Du fait de la non-différentiabilité de la correspondance paramètres \mapsto solution (même si le problème est bien posé au sens où la solution est définie de façon unique), il n'existera pas de constante c telle qu'une erreur relative ε sur les paramètres induise une erreur contrôlée par $c\varepsilon$. Un tel modèle est essentiellement *inutilisable* en situation réelle, ou tout du moins très délicat à exploiter.

13. On parle parfois de stabilité au sens de Hadamard, même si cette appellation fait plutôt référence à une dépendance *continue* de la solution par rapport aux données.

Remarque V.3.5. (Sensibilité vis à vis des paramètres)

Dans la continuité de la remarque précédente, mais de façon plus positive, lorsque l'on est bien dans le cadre du théorème des fonctions implicites, la différentielle de Ψ précise la dépendance de la solution vis-à-vis des paramètres. On écrira en général simplement $\Psi(x) = y(x)$, de telle sorte que la matrice jacobienne de Ψ contient les dérivées partielles $\partial y_i / \partial x_j$ (où y est maintenant considéré comme fonction de x), c'est-à-dire l'expression de la dépendance de la i -ième composante de y vis-à-vis du paramètre x_j (on parlera de *sensibilité*). Les paramètres les plus significatifs pour une composante y_i correspondent aux fortes valeurs de la dérivée, il sera important de bien en maîtriser la valeur, alors que les paramètres pour lesquels $\partial y_i / \partial x_j$ est petit pour tous les i peuvent être a priori estimés avec une précision médiocre, sans que cela n'influe de façon préjudiciable sur la solution.

Remarque V.3.6. (Identification de paramètres)

Il est courant de s'intéresser au *problème inverse*, qui peut se formuler comme suit. On fait confiance au modèle $f(x, y) = 0$, on dispose de mesures pour la solution y , et l'on cherche à *estimer les paramètres* correspondant à la solution mesurée. On est donc amené à considérer le problème dans l'autre sens, c'est-à-dire que l'on cherche à estimer x à partir de la connaissance de y . On ne peut espérer retrouver exactement les paramètres que si leur nombre est égal à celui des inconnues $n = m$. On notera que, pour ce nouveau problème, les paramètres les plus difficiles à identifier précisément sont ceux qui ont peu d'influence sur la solution, qui étaient considérés pour le problème direct comme peu significatifs, dont la connaissance précise n'était pas nécessaire. C'est précisément leur peu d'influence sur la solution qui rend difficile leur estimation à partir de la connaissance de cette solution¹⁴.

Définition V.3.7. Soit φ une application d'un ouvert $U \subset \mathbb{R}^n$ dans un ouvert $V = \varphi(U)$ dans \mathbb{R}^n . On dit que φ est un C^1 – difféomorphisme de U vers V si φ est bijective, et si φ et sa réciproque φ^{-1} sont continûment différentiables.

Proposition V.3.8. On se place dans les hypothèses de la définition précédente. La différentielle de φ est inversible en tout point de U , et son inverse est la différentielle de l'application réciproque φ^{-1} : pour tout $x \in U$, $y = \varphi(x) \in V$,

$$d\varphi^{-1}(y) = (d\varphi(x))^{-1}.$$

Démonstration. On a, pour tout $y \in V$,

$$\varphi \circ \varphi^{-1}(y) = y.$$

La règle de différentiation en chaîne implique donc (avec $x = \varphi^{-1}(y)$)

$$d\varphi(x) \circ d\varphi^{-1}(y) = \text{Id},$$

qui conclut la preuve. □

Théorème V.3.9. (Inversion locale)

Soit φ une application continûment différentiable d'un ouvert $W \subset \mathbb{R}^n$ dans \mathbb{R}^n . On suppose que $d\varphi(x)$ est inversible pour tout $x \in W$. Alors φ est un C^1 – difféomorphisme local : pour tout $x_0 \in W$, il existe un voisinage ouvert $U \subset W$ de x_0 et un voisinage ouvert V de $y_0 = \varphi(x_0)$ tel que $\varphi|_U$ soit un C^1 – difféomorphisme de U vers V .

Démonstration. On considère l'application (noter que l'on écrit (y, x) du fait qu'il va s'agir, contrairement à l'usage, d'exprimer x en fonction de y) :

$$g : (y, x) \in \mathbb{R}^n \times W \longmapsto g(y, x) = \varphi(x) - y.$$

14. Nous nous en tenons dans cette remarque à une vision un peu simpliste des choses, comme s'il était possible de séparer à la fois les paramètres et les composantes de la solution (ça n'est possible que si la différentielle est diagonale). En tout généralité, les études de sensibilité évoquées dans ces remarques passent par une étude plus complète de la matrice dans sa globalité, qui passe en particulier par une analyse spectrale.

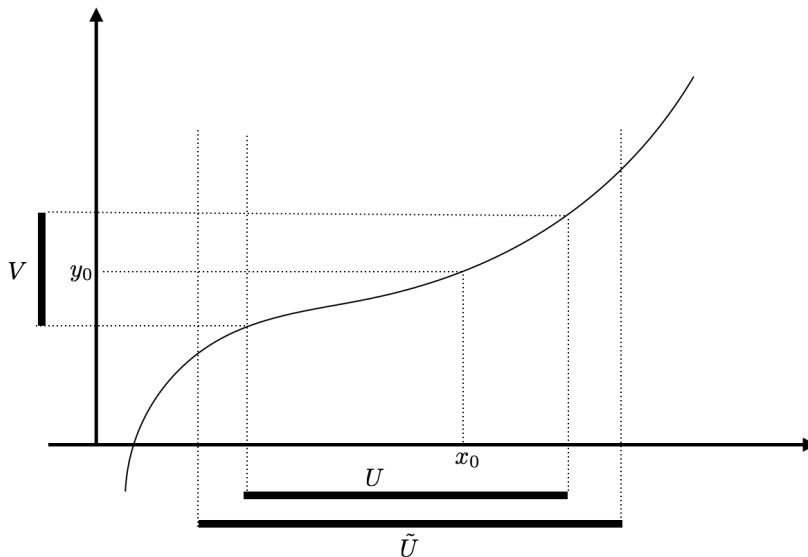


FIGURE V.3.2 – Théorème d'inversion locale

Cette application est différentiable sur $\mathbb{R}^n \times W$, de différentielle partielle par rapport à x

$$\partial_x g(y, x) = d\varphi(x).$$

Cette différentielle est inversible sur W par hypothèse. Soit $x_0 \in W$, et $y_0 = \varphi(x_0)$, d'où $g(y_0, x_0) = 0$. D'après le théorème des fonctions implicites, il existe un voisinage V de y_0 , un voisinage $\tilde{U} \subset W$ de x_0 , et Ψ une application continûment différentiable de V dans \tilde{U} , tels que

$$(y, x) \in V \times \tilde{U}, \quad g(y, x) = 0, \quad \text{i.e. } y = \varphi(x) \iff x = \Psi(y).$$

L'application Ψ est donc la réciproque de φ . Il reste à préciser les voisinages ouverts de x_0 et y_0 qui sont en bijection. Il faut prendre garde à une difficulté (annoncée dans la remarque V.3.2) : Ψ , qui est bien définie sur tout V , (cet ouvert V est noté U dans le théorème des fonctions implicites, du fait du renversement des rôles de x et y que nous avons effectué ici), n'est pas nécessairement *surjective* de V dans \tilde{U} . Pour garantir que les deux ouverts soient en bijection, on réduit l'ouvert \tilde{U} en introduisant

$$U = \tilde{U} \cap \Psi(V).$$

Comme, pour tout $y \in V$, l'équation $y = \varphi(x)$ n'a qu'une solution en $x \in \tilde{U}$, cet ensemble s'écrit aussi $U = \tilde{U} \cap \varphi^{-1}(V)$. Il s'agit donc bien d'un ouvert par continuité de V . \square

V.4 Problème adjoint

On s'intéresse ici à une méthode d'estimation du gradient d'une fonctionnelle $F(u)$ à valeurs réelles, définie comme $g(y_u)$, où y_u est solution d'un problème de type

$$f(u, y) = 0. \quad (\star)$$

considéré comme un problème en y , paramétré par u . On assimilera dans la suite la différentielle d'une application à la matrice jacobienne qui la représente dans les bases canoniques sur les espaces d'arrivée et de départ.

On considère une application f de $\mathbb{R}^n \times \mathbb{R}^m$ dans \mathbb{R}^m , continûment différentiable sur un ouvert W de $\mathbb{R}^n \times \mathbb{R}^m$. On notera $u \in \mathbb{R}^n$ la variable dite *de contrôle*, et $y \in \mathbb{R}^m$ la variable d'état.

On considère $(u_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ solution de $f(u_0, y_0) = 0$, et l'on suppose que $\partial_y f(u, y)$ est inversible dans un voisinage¹⁵ de (u_0, y_0) . D'après le théorème des fonctions implicites V.3.1, il existe un ouvert U contenant u_0 et un ouvert V contenant y_0 , et une application Ψ de U dans V telle que

$$\forall (u, y) \in U \times V, f(u, y) = 0 \iff y = \Psi(u),$$

avec Ψ continument différentiable sur U . La différentielle de Ψ en $u \in U$ s'exprime

$$d\Psi(u, y) = -(\partial_y f(u, y))^{-1} \circ \partial_u f(u, y).$$

Par la suite, pour désigner le y associé à u comme solution de l'équation (\star) , on utilisera indifféremment $\Psi(u)$ ou y_u .

On considère maintenant une application g de \mathbb{R}^m dans \mathbb{R} , et l'on définit la fonction F de $U \subset \mathbb{R}^n$ dans \mathbb{R} par

$$F(u) = g(y_u) = g \circ \Psi(u),$$

où y_u est la solution de (\star) associée à u .

Approche par inversion de matrice

La différentielle $dF(u)$ s'exprime .

$$dF(u) = -dg(y_u) \circ (\partial_y f(u, y_u))^{-1} \circ \partial_u f(u, y_u). \quad (\text{V.4.1})$$

Approche par différences finies

b) Si l'on souhaite estimer la différentielle de F en u_0 , sans avoir à inverser la matrice $\partial_y f(u, y_u)$, on peut approcher chacune des dérivées partielles par

$$\partial_{u_j} F(u) \approx \frac{F(u + he_j) - F(u)}{h},$$

pour un certain $h > 0$ petit, où les e_i sont les vecteurs de la base canonique de \mathbb{R}^n . Chaque estimation de F en un point nécessite une résolution d'un problème de type (\star) , et l'on doit estimer F en u_0 et en les $u + he_j$, ce qui nécessite $n + 1$ résolutions.

Exercice V.4.1. On se propose d'appliquer ce qui précède à la situation suivante : on suppose $n = 1$, $m \geq 1$, et l'on note $A(u)$ une matrice de $\mathcal{M}_m(\mathbb{R})$ dont les coefficients dépendent de la variable scalaire u :

$$A(u) = (a_{ij}(u))_{1 \leq i, j \leq n}.$$

On suppose que les $u \mapsto a_{ij}(u)$ sont des fonctions continument différentiables sur \mathbb{R} . Le problème définissant y fonction de u s'écrit sous la forme d'un système matriciel en y

$$A(u) \cdot y = b,$$

où b est un vecteur donné de \mathbb{R}^m . On suppose que $A(u_0)$ est inversible pour un certain $u_0 \in \mathbb{R}$. On définit g comme

$$y \in \mathbb{R}^m \longmapsto g(y) = \frac{1}{2} \|y - \bar{y}\|^2,$$

où $\bar{y} \in \mathbb{R}^m$ est donné.

a) On pose $f(u, y) = A(u) \cdot y - b$.

Préciser les expressions de $\partial_y f$ et $\partial_u f$ en fonction de la matrice A et de sa dérivée $A' = (a'_{ij})$.

N.B. : On pourra effectuer pour cela les développements limités de $f(u + \delta u, y)$ et $f(u, y + \delta y)$

15. Il suffit en fait de supposer que $\partial_y f(u_0, y_0)$ est inversible. La matrice qui la représente dans la base canonique de \mathbb{R}^m est de déterminant non nul. Comme le déterminant est fonction continue des coefficients, le déterminant reste non nul, et donc $\partial_y f(u, y)$ est inversible, dans un voisinage W de (u_0, y_0) .

CORRECTION.

L'application $y \mapsto f(u, y)$ étant linéaire, on a directement

$$\partial_y f(u, y) = A(u) \in \mathcal{M}_m(\mathbb{R}).$$

Pour la dépendance en u , on a

$$f(u + \delta u, y) = A(u + \delta u) \cdot y - b = (A(u) + A'(u)\delta u + o(\delta u)) \cdot y - b = f(u, y) + (A'(u) \cdot y)\delta u + o(\delta u).$$

La différentielle partielle est donc $\delta u \mapsto (A'(u) \cdot y)\delta u$ (où $A'(u) \cdot y$ est un vecteur de \mathbb{R}^m), que l'on peut représenter par la matrice colonne exprimant $A'(u) \cdot y$ dans la base canonique.

b) Montrer que ce problème rentre dans le cadre général décrit au début de cette section, en déduire que l'on peut exprimer y en fonction de u sur un voisinage U de u_0 , sous la forme $y = \Psi(u)$, et donner l'expression de la différentielle de Ψ .

N.B. : cette différentielle étant une application de \mathbb{R} dans \mathbb{R}^m , on pourra la représenter par une matrice colonne.

CORRECTION.

L'application linéaire $\partial_y f(u, y) = A(u)$ est inversible par hypothèse en u_0 . On peut donc exprimer y fonction de u sur un voisinage U de u_0 . La différentielle de Ψ s'exprime

$$d\Psi(u) = -A(u)^{-1} \circ (A'(u) \cdot y_u).$$

On peut voir cette expression comme le produit d'une matrice carrée par une matrice colonne, on peut donc l'écrire comme un produit matrice-vecteur $d\Psi(u) = -(A(u)^{-1} A'(u)) \cdot y_u$.

c) Préciser la différentielle de l'application $y \in \mathbb{R}^m \mapsto g(y) = \frac{1}{2} \|y - \bar{y}\|^2$, en expliquant pourquoi on peut l'identifier à une matrice ligne. Quel est le lien entre cette matrice ligne et le gradient de g (pour le produit scalaire canonique sur \mathbb{R}^m) ?

Expliquer pourquoi la différentielle de $u \mapsto F(u)$ peut être assimilée à un réel, et préciser l'expression de cette différentielle en u_0 .

CORRECTION.

Pour calculer la différentielle de g , on écrit

$$g(y + \delta y) = \frac{1}{2} \langle y + \delta y - \bar{y} | y + \delta y - \bar{y} \rangle = g(y) + \langle y - \bar{y} | \delta y \rangle + o(\delta y).$$

La différentielle est donc l'application qui à δy associe $\langle y - \bar{y}, \delta y \rangle$, que l'on peut voir comme un produit matrice vecteur en assimilant le vecteur $y - \bar{y}$ à la matrice ligne correspondante. Ce vecteur $y - \bar{y}$ est par définition le gradient de g .

L'application $u \mapsto F(u)$ est de \mathbb{R} dans \mathbb{R} , sa différentielle est donc donné par un réel qui est la dérivée de F au sens usuel. L'expression générale (V.4.1) s'écrit

$$dg(u) = -(y - \bar{y})(A(u)^{-1} A'(u)) \cdot y_u$$

qui est un produit de matrices (matrice ligne $y - \bar{y}$ par matrice colonne $(A(u)^{-1} A'(u)) \cdot y_u$), qui donne donc une matrice 1×1 que l'on peut identifier à un simple réel. Noter que le produit de matrices ci-dessus peut aussi s'exprimer comme un produit scalaire entre 2 vecteurs :

$$-\langle y - \bar{y} | (A(u)^{-1} A'(u)) \cdot y_u \rangle.$$

Approche par méthode de l'état adjoint

Nous présentons maintenant une méthode d'estimation de $dF(u)$ (appelée méthode de l'état adjoint) qui ne nécessite ni de calculer l'inverse d'une matrice (approche de la question 2a basée sur l'expression de la différentielle), ni de résoudre un grand nombre de problèmes de type (*) (approche de la question 2b basée sur l'approximation des dérivées partielles par des taux de variation).

On définit l'application L (appelée *Lagrangien*) de la façon suivante

$$L : (y, u, p) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \mapsto L(y, u, p) = g(y) + \langle f(u, y) | p \rangle.$$

Noter que, dans l'expression ci-dessus, y est une variable “libre” (il ne s’agit pas de y_u).

Par définition de y_u , quel que soit p , on $F(u) = L(y_u, u, p)$, avec $y_u = \Psi(u)$. Si l'on différencie cette identité par rapport à u , à p fixé quelconque, on obtient¹⁶

$$dF(u) = \partial_y L(y_u, u, p) \circ d\Psi(u) + \partial_u L(y_u, u, p) \quad \forall p \in \mathbb{R}^m.$$

On appelle *problème adjoint* le problème consistant à trouver p tel que

$$\partial_y L(y_{u_0}, u_0, p) = 0, \quad \text{avec } L(y_{u_0}, u_0, p) = g(y_{u_0}) + \langle f(u_0, y_{u_0}) | p \rangle.$$

Précisons la forme de ce problème adjoint. Le premier terme de $\partial_y L(y_{u_0}, u_0, p)$ s'écrit simplement $dg(y_0)$. Pour le second terme, on écrit

$$\langle f(u, y + \delta y) | p \rangle = \langle f(u, y) | p \rangle + \langle \partial_y f(u, y) \cdot \delta y | p \rangle = \langle \partial_y f(u, y)^T \cdot p | \delta y \rangle,$$

où l'on a assimilé l'application linéaire $\partial_y f(u, y)$ à sa matrice dans la base canonique, $\partial_y f(u, y)^T$ représente la matrice transposée, et le ‘·’ le produit matrice-vecteur. On assimilera la différentielle de $\langle f(u, y) | p \rangle$ à son gradient pour le produit scalaire canonique, on écrira donc $\partial_y f(u, y)^T \cdot p$ la différentielle en y de $\langle f(u, y) | p \rangle$.

On a donc

$$\partial_y L(y_{u_0}, u_0, p) = dg(y_0) + \partial_y f(u_0, y_{u_0})^T \cdot p.$$

Le problème adjoint s'écrit ainsi

$$\partial_y f(u_0, y_{u_0})^T \cdot p = -dg(y_0).$$

Il s'agit d'un système linéaire, qui fait intervenir la matrice carrée de $\partial_y f(u_0, y_{u_0})^T \in \mathcal{M}_m(\mathbb{R})$. Comme c'est la matrice transposée de $\partial_y f(u_0, y_{u_0})$, qui est inversible, elle est elle-même inversible, le problème admet donc une solution unique.

XXXX 6) Faire la synthèse des deux questions précédentes en décrivant une stratégie permettant d'ex-primer $dF(u_0)$ sans avoir à calculer $d\Psi(u_0)$.

CORRECTION.

On a montré

$$dF(u_0) = \partial_y L(y_{u_0}, u_0, p) \circ d\Psi(u_0) + \partial_u L(y_{u_0}, u_0, p) \quad \forall p \in \mathbb{R}^m.$$

Cette expression est en particulier vraie pour la solution p_0 du problème adjoint, qui est telle que

$$\partial_y f(u_0, y_{u_0})^T \cdot p_0 = -dg(y_0).$$

Pour ce p_0 particulier, on a $\partial_y L(y_{u_0}, u_0, p_0) = 0$, ce qui annule le premier terme de l'expression de $dF(u)$ ci-dessus, d'où

$$dF(u_0) = \underbrace{\partial_y L(y_{u_0}, u_0, p_0) \circ d\Psi(u_0)}_{=0} + \partial_u L(y_{u_0}, u_0, p_0) = \partial_u L(y_{u_0}, u_0, p_0),$$

ce qui permet bien de s'affranchir du calcul de $d\Psi(u_0)$.

16. On prendra garde au fait que $\partial_u L(y_u, u, p)$ désigne bien la différentielle de L par rapport à la seconde variable (u) uniquement, prise au point (y_u, u, p) .

V.5 Exercices

Exercice V.5.1. Soit f une fonction continûment différentiable sur $\mathbb{R}^n \times \mathbb{R}^m$, à valeurs dans \mathbb{R}^m .

Montrer que l'ensemble $F = \{(x, y) \mid f(x, y) = 0\}$ est un fermé de $\mathbb{R}^n \times \mathbb{R}^m$, et que l'ensemble des (x_0, y_0) au voisinage desquels on peut appliquer le théorème des fonctions implicites est un ouvert du fermé F (pour la topologie induite, dont les ouverts sont les intersections d'ouverts de $\mathbb{R}^n \times \mathbb{R}^m$ avec F).

CORRECTION.

L'ensemble F est fermé comme image réciproque du fermé $\{0\}$ par une application continue (car différentiable). Si l'on peut appliquer le TFI en un point (x_0, y_0) , alors il existe un voisinage ouvert W de (x_0, y_0) tel que $\partial_y f$ est inversible sur W . On peut donc appliquer le TFI au voisinage de tout $(x, y) \in F \cap W$.

Exercice V.5.2. Identifier dans les cas suivants l'ensemble F des solutions de $f(x, y) = 0$, ainsi que l'ensemble des points (x, y) au voisinage desquels on peut appliquer le théorème des fonctions implicites. Préciser, pour les points en lesquels les hypothèses ne sont pas vérifiées, si intervertir les rôles de x et y permet de les vérifier.

- a) $f(x, y) = x^2 + y^2 - r^2$.
- b) $f(x, y) = y - x^2$.
- c) $f(x, y) = (y - x^3)y$.

CORRECTION.

a) L'ensemble des zéros de f est le cercle de centre $(0, 0)$ et de rayon r . On peut appliquer le TFI (exprimer y fonction de x) en dehors des points situés à l'est et à l'ouest ($(r, 0)$ et $(-r, 0)$). En ces points, on peut appliquer le théorème dans l'autre sens, i.e. exprimer x fonction de y .

b) L'ensemble F est une parabole d'axe l'axe des y . On peut appliquer le TFI au voisinage de tout point. Pour exprimer localement x fonction de y , il faut exclure l'origine.

c) l'ensemble F est l'union de la courbe $y = x^3$ et de l'axe des x . On peut appliquer le TFI au voisinage de tout point de F en dehors de l'origine.

Exercice V.5.3. (Coordonnées polaires)

- a) Montrer que l'application

$$T : (r, \theta) \in U =]0, +\infty[\times]-\pi, \pi[\longrightarrow (r \cos \theta, r \sin \theta)$$

est un C^1 difféomorphisme entre U et $V = \mathbb{R}^2 \setminus (\{0\} \times \mathbb{R}_-)$.

b) On considère une fonction f de V dans \mathbb{R} , et l'on note g sa version polaire, i.e. $g(r, \theta) = f(r \cos \theta, r \sin \theta)$. Préciser le lien entre les différentielles de f et de g .

CORRECTION.

a) L'application T est bien une bijection entre U et V , continue, et de jacobienne

$$J = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$$

de déterminant $\Delta = r(\cos^2 \theta + \sin^2 \theta) = r > 0$ sur V . Il s'agit donc d'un C^1 difféomorphisme global entre U et V .

b) On a (on assimile les différentielles df et dg à leurs matrices-lignes associées)

$$dg(r, \theta) = df(r \cos \theta, r \sin \theta) \circ J,$$

et

$$df(x, y) = dg(r, \theta) \circ J^{-1},$$

avec

$$J = \frac{1}{r} \begin{pmatrix} r \cos \theta & r \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

Exercice V.5.4. On note $U = \mathbb{R}^2 \setminus \{(0,0)\}$. Montrer que l'application

$$\varphi : (x, y) \in U \longmapsto (x^2 - y^2, 2xy) \in \mathbb{R}^2$$

est un difféomorphisme local au voisinage de tout point de son domaine, mais pas global.

CORRECTION.

La jacobienne de φ est

$$J = \begin{pmatrix} 2x & -2y \\ 2y & -2x \end{pmatrix}$$

dont le déterminant est $4(x^2 + y^2)$, qui est bien non nul en tout point de U , on peut donc appliquer le théorème d'inversion locale au voisinage de tout point de U

Il ne s'agit en revanche pas d'un difféomorphisme global, du fait que l'application n'est pas injective (on a par exemple $\varphi(-x, -y) = \varphi(x, y)$).

Exercice V.5.5. (Dépendance d'une racine simple d'un polynôme réel vis-à-vis des coefficients)

A toute collection de coefficients $c = (c_0, c_1, \dots, c_N) \in \mathbb{R}^{N+1}$ on associe le polynôme

$$P_c(X) = c_0 + c_1 X + \cdots + c_N X^N.$$

On se donne \tilde{c} et \tilde{z} tels que $\tilde{z} \in \mathbb{R}$ est racine simple du polynôme $P_{\tilde{c}}$. Montrer qu'il existe une fonction différentiable Ψ des coefficients, définie dans un voisinage U de \tilde{c} , telle que $\tilde{z} = \Psi(\tilde{c})$ et telle que, pour tout $c \in U$, $z = \Psi(c)$ est racine du polynôme P_c .

Exprimer la différentielle de Ψ .

CORRECTION.

On définit la fonction f de la façon suivante

$$(c, z) \in \mathbb{R}^{N+1} \times \mathbb{R} \longmapsto f(c, z) = P_c(z) \in \mathbb{R}.$$

L'équation

$$f(c, z) = 0$$

exprime que $z \in \mathbb{R}$ est racine du polynôme P_c

On se propose d'appliquer le théorème des fonctions implicites à cette fonction, en vue d'exprimer z en fonction de x . En premier lieu on a bien $f(\tilde{c}, \tilde{z}) = 0$. La différentielle partielle $\partial_z f$ est obtenue en faisant un développement limité

$$\begin{aligned} f(c, z + h) = P_c(z + h) &= \sum_{n=0}^N c_n (z + h)^n \\ &= \sum_{n=0}^N c_n (z)^n + \sum_{n=1}^N c_n (z)^{n-1} h + o(h) \\ &= f(c, z) + P'(z) h + o(h). \end{aligned}$$

La dérivée partielle de f par rapport à z est donc $P'(z)$. Or $P'(\tilde{z}) \neq 0$ du fait que \tilde{z} est racine simple, et $P'(z)$ dépend continûment de z . La dérivée est donc non nulle dans un voisinage U de \tilde{z} , ce qui assure l'existence d'une fonction Ψ qui aux coefficient c associe une racine du polynôme. La différentielle de Ψ en c s'exprime

$$d\Psi(c) = -(\partial_z f)^{-1} \partial_c f, \text{ avec } z = \Psi(c),$$

et peut donc se représenter matriciellement par

$$\frac{1}{P'(z)} \begin{bmatrix} 1 & z & z^2 & \dots & z^N \end{bmatrix}.$$

Exercice V.5.6. (Dépendance d'une racine simple d'un polynôme complexe vis-à-vis des coefficients (version complexe de l'exercice V.5.5))

À toute collection de coefficients $c = (c_0, \dots, c_n) \in \mathbb{C}^n$ on associe le polynôme

$$P_c(X) = c_0 + c_1 X + \dots + c_N X^N.$$

On se donne \tilde{c} et \tilde{z} tels que $\tilde{z} \in \mathbb{C}$ est racine *simple* du polynôme $P_{\tilde{c}}$. Montrer qu'il existe une fonction différentiable Ψ des coefficients, définie dans un voisinage U de \tilde{c} , telle que $\tilde{z} = \Psi(\tilde{c})$ et telle que, pour tout $c \in U$, $z = \Psi(c)$ est racine du polynôme P_c .

CORRECTION.

On écrit, pour tout $n \leq N$, $c_n = a_n + ib_n$, avec a_n, b_n réels, ou plus globalement $c = a + ib$, avec a et b dans \mathbb{R}^{N+1} . On définit la fonction f de la façon suivante

$$(a, b, x, y) \in \mathbb{R}^{N+1} \times \mathbb{R}^{N+1} \times \mathbb{R}^2 \mapsto f(a, b, x, y) = P_{a+ib}(x + iy) \in \mathbb{R}^2$$

(on identifie le résultat complexe à un couple de réels, correspondant aux parties réelle et imaginaire). L'équation

$$f(a, b, x, y) = 0$$

exprime que $z = x + iy$ est racine du polynôme P_{a+ib}

On se propose d'appliquer le théorème des fonctions implicites à cette fonction, en vue d'exprimer (x, y) (ou $z = x + iy$) en fonction de (a, b) . En premier lieu on a bien $f(\tilde{a}, \tilde{b}, \tilde{x}, \tilde{y}) = 0 \in \mathbb{R}^2$. La différentielle partielle $\partial_{x,y} f$ est obtenue en faisant un développement limité (on utilise ci-après les notations $c = a + ib$, $z = x + iy$, $h = h_x + ih_y$)

$$\begin{aligned} f(a, b, x + h_x, y + h_y) &= P_c(z + h) = \sum_{n=0}^N c_n (z + h)^n \\ &= \sum_{n=0}^N c_n (z)^n + \sum_{n=1}^N c_n (z)^{n-1} h + o(h) \\ &= f(a, b, x, y) + P'(z)h + o(h). \end{aligned}$$

Si l'on écrit $P'(z) = \alpha + i\beta$ et $h = h_x + ih_y$, la différentielle $\partial_{x,y} f$ s'écrit donc matriciellement

$$\partial_{x,y} f = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$$

Le déterminant de cette matrice est $\alpha^2 + \beta^2$. Une matrice de cette forme est donc inversible si et seulement si elle non nulle. Or $P'(\tilde{z}) = \tilde{\alpha} + i\tilde{\beta}$ est non nul en \tilde{z} (car la racine est supposée simple), et $P'(z)$ dépend continûment de z . La différentielle est donc inversible dans un voisinage U de \tilde{z} , ce qui assure l'existence d'une fonction Ψ qui aux coefficient c associe une racine du polynôme.

Pour représenter la différentielle, on a intérêt à reprendre la notation complexe, pour écrire

$$d\Psi(c) = P'_c(z)^{-1} [1 \ z \ \dots \ z^N], \text{ avec } z = \Psi(c),$$

qui est l'application qui, à une variation des coefficients $h = (h_0, h_1, \dots, h_N)$, associe le complexe

$$P'(z)^{-1} \sum_{n=0}^N h_n z^n.$$

On peut retrouver une représentation matricielle réelle, comme application de $\mathbb{R}^{2(N+1)}$ (parties réelle et imaginaire de la variation de la collection c de coefficients) dans \mathbb{R}^2 (parties réelle et imaginaire de la racine), en introduisant explicitement les parties réelle et imaginaire de h .

Problème V.5.1. On se propose d'étudier un modèle simplifié de bilan radiatif de la terre, basé sur l'écriture d'un équilibre entre l'énergie solaire reçue par la terre et l'énergie ré-émise par rayonnement, supposé suivre la loi de Stefan-Boltzman. On note F le flux de rayonnement solaire reçu en moyenne par unité de surface sur terre, $F_0 \approx 341 \text{ W m}^{-2}$. On considère qu'une fraction $A \in [0, 1]$ de cette énergie est immédiatement réfléchie, où A_0 , appelé *albedo*, est autour de 0.3. L'énergie émise en moyenne par unité de surface par la terre s'écrit σT^4 , où T est la température moyenne (exprimée en Kelvin), et $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$. On considère qu'une fraction de cette énergie n'est pas rayonnée vers l'espace, du fait de l'effet de serre. On note $S \in [0, 1]$ la fraction d'énergie qui n'est pas évacuée vers l'espace. Ce paramètre sans dimension est estimé à $S = 0.4$. En supposant que l'on est à l'équilibre, on écrit le bilan entre les énergies reçue et émises :

$$\sigma T^4(1 - S) = (1 - A)F.$$

a) Estimer la température moyenne T_0 à la surface de la terre associée aux valeurs de référence S_0 , A_0 et F_0 selon ce modèle, et estimer la valeur qu'aurait cette température s'il n'y avait pas d'effet de serre (en supposant que le modèle reste valide¹⁷).

b) Montrer¹⁸ que, au voisinage du point d'équilibre considéré, on peut exprimer la température comme une fonction continûment différentiable des paramètres S , A , et F .

Exprimer la différentielle de cette fonction, et en déduire le coefficient de proportionnalité entre une variation de S autour de la valeur S_0 et la variation en degrés de la température. Quelle variation de S induit une augmentation de la température de 2°C ?

Si l'on mesure une petite variation de température de δT autour de T_0 , que peut-on dire (toujours dans l'hypothèse où l'on accorde une foi absolue au modèle) des variations δS , δA , et δF qui ont pu induire cette variation de température ?

c) On estime que le CO₂ est responsable de 60 % de l'effet de serre dû aux Gaz à Effets de Serre (GES), eux-mêmes responsables de 30 % de l'effet de serre global. Si l'on admet que l'effet de serre dû au CO₂ est proportionnel à sa concentration dans l'atmosphère, estimer l'augmentation du taux de CO₂ qui conduirait, selon ce modèle à une augmentation de la température de 2°C .

d) On considère que l'albedo dépend lui-même de la température : une augmentation de la température est susceptible d'induire une fonte des glaces, qui diminue la part de surface fortement réfléchissante, d'où une diminution de l'albedo. On écrit donc, pour encoder ce phénomène,

$$A = A_0 - \beta(T - T_0),$$

avec $\beta > 0$ (exprimé en K⁻¹). On considérera par ailleurs le terme F de flux radiatif fixé à sa valeur de référence F_0 . Faire l'étude de ce nouveau modèle au voisinage du point d'équilibre de référence (S_0, T_0) .

e) (*) L'effet de serre, qui dépend par exemple de la masse nuageuse présente en moyenne dans l'atmosphère, dépend lui-même de la température. Explorer la manière dont cette dépendance est susceptible d'affecter les considérations précédentes (on pourra écrire S comme la somme d'un terme dépendant du CO₂, et d'autres termes susceptibles de dépendre directement de la température).

CORRECTION.

a) Avec les valeurs de référence données, on trouve

$$T_0 \approx 289 \text{ K} = 16^\circ\text{C}.$$

Cette température en général estimée à 15°C la différence (qui correspond à une variation relative de 0.3 % sur la valeur en Kelvin), s'explique par le fait que les valeurs de S et A sont des estimations. La valeur correspondant à un effet de serre nul est de -18°C .

b) Le modèle peut s'écrire

$$f(S, A, F, T) = \sigma T^4(1 - S) - (1 - A)F = 0.$$

17. Vue la baisse de température importante induite par cette suppression virtuelle de l'effet de serre, une grande partie de l'eau liquide (peu réfléchissante) à la surface du globe se transformerait en glace (fort pouvoir réfléchissant), ce qui entraînerait une augmentation significative de l'albedo A , qui réduirait encore la température d'équilibre.

18. Même si cela n'est pas à strictement parler nécessaire ici, on s'efforcera de *jouer le jeu* en utilisant le théorème des fonctions implicites.

La fonction f est polynomiale, donc différentiable, et l'on a

$$\partial_T f = 4\sigma T^3(1 - S),$$

qui est non nul sur toute la plage $]0, +\infty[$ des températures physiques, et en particulier en T_0 . On peut donc appliquer le théorème des fonctions implicites qui assure l'existence d'un voisinage de T_0 sur lequel on peut exprimer T comme fonction des paramètres S , A , et F . La différentielle de cette application peut s'écrire, d'après le même théorème,

$$dT = -\frac{1}{4\sigma T^3(1 - S)} (-\sigma T^4 dS + F dA - (1 - A) dF).$$

On a donc en particulier

$$\frac{\partial T}{\partial S} = +\frac{T}{4(1 - S)},$$

dont la valeur aux conditions de référence est 120 K (S est sans unité). La variation de S induisant une augmentation de 2 °C de la température est donc $+2/120 \approx 0.016$.

c) La variation relative de la concentration en CO_2 susceptible d'induire une variation de S de 0.016 est donc, si l'on admet que l'effet de serre dû au CO_2 est proportionnel à cette concentration,

$$\frac{\delta c}{c} = \frac{1}{0.6 \times 0.3} 0.016 \approx 0.09.$$

Pour un taux actuel de 415 ppm (parties par million), cela correspond donc à une augmentation de l'ordre de 40 ppm.

d) Le modèle s'écrit maintenant

$$g(S, T) = \sigma T^4(1 - S) - (1 - A_0 + \beta(T - T_0))F_0.$$

La dérivée de g par rapport à T est

$$\partial_T g = 4\sigma T^3(1 - S) - \beta F_0.$$

Pour β petit, plus précisément inférieur à $4\sigma(1 - S_0)$, cette dérivée est strictement positive au voisinage de S_0 , on peut donc utiliser le théorème des fonctions implicites et exprimer T fonction de S , avec une sensibilité modifiée :

$$\partial_S T = \frac{\sigma T^4}{4\sigma T^3(1 - S) - \beta F_0} > \frac{T}{4(1 - S)}.$$

La sensibilité de T vis-à-vis de S est donc augmentée par cet effet (on parle de boucle de rétro-action positive, ou en langage commun de cercle vicieux, : quand la température augmente, elle induit un renforcement d'un des facteurs qui tendent à l'augmenter).

La valeur critique est $\beta_c = 4\sigma T^3(1 - S)/F_0 \approx 10^{-3} \text{ K}^{-1}$. Pour cette valeur de β , une augmentation de 10 °C diminue l'albedo de 0.01.

Lorsque β s'approche de cette valeur, $\partial_S T$ tend vers $+\infty$, et prend des valeurs négatives pour lorsque β_0 est dépassé. Le fait que les valeurs soit négative pourrait laisser penser que la situation s'est inversée : une augmentation de l'effet de serre entraînerait une diminution de la température. La situation est bien sûre tout autre, comme le suggère l'explosion en β_0 . En fait, nous allons vérifier que, pour $\beta \geq 0$, le point d'équilibre considéré n'en est plus un, ou plus précisément que c'est un point d'équilibre instable, qui n'a aucune chance d'être observé comme solution pérenne dans la réalité. Pour comprendre ce qui se passe, considérons un modèle dynamique d'évolution de la température, basée sur des hypothèses hautement simplificatrices. On considère que la chaleur (on exclut les autres formes d'énergie) emmagasinée par la terre à un instant donné s'écrit comme le produit d'une constante C (capacité calorifique du système terre-atmosphère) avec la température. On écrit que cette énergie évolue selon le bilan radiatif, que l'on ne suppose plus équilibré :

$$C \frac{dT}{dt} = (1 - A_0 + \beta(T - T_0)) - \sigma T^4(1 - S) = -f(T, S).$$

Pour $S = S_0$, la température T_0 est dite point d'équilibre de l'équation d'évolution, c'est à dire que $T \equiv T_0$ en est une solution triviale. Mais si l'on perturbe légèrement la température, tant que la température reste proche de T_0 , la variation u de cette température vérifie

$$C \frac{du}{dt} = C \frac{d(T_0 + u)}{dt} = -f(T_0 + u, S_0) \approx -f(T_0, S_0) - \partial_T f(T_0, S_0) u.$$

Si $\partial_T f(T_0, S_0) > 0$ (petites valeurs de β), l'évolution peut être décrite par une équation linéaire avec coefficient négatif, qui correspond à une relaxation exponentielle vers la température d'équilibre T_0 . Si $\beta > \beta_0$, on a $\partial_T f(T_0, S_0) < 0$, et l'évolution devient instable, avec une solution du type $u = u_0 \exp(\lambda t)$, $\lambda > 0$. La température a donc tendance à s'éloigner de la température d'équilibre. Noter que, dès que cette température est significativement différente de T_0 , le modèle linéaire n'a plus aucune légitimité, il faut mener une étude du modèle non linéaire pour étudier le comportement en temps long de la solution. En tout cas pour $\beta > \beta_0$, il apparaît que la température T_0 n'est pas un point d'équilibre stable, donc son étude en tant que point d'équilibre pertinent d'un système physique réel n'a pas de sens, ce qui peut expliquer le caractère paradoxal de la négativité de $\partial_S T$ dans ce régime.

Problème V.5.2. (Problème adjoint)

XXXXXXX

7) Appliquer la démarche décrite précédemment à la situation particulière de la question 3, où, nous le rappelons, u est une variable scalaire. On précisera en particulier le problème adjoint, dont on justifiera le caractère bien posé, et on donnera l'expression de $dF(u)$ qui fait intervenir la solution à ce problème adjoint.

CORRECTION.

Le problème adjoint s'écrit de façon abstraite

$$\partial_y f(u_0, y_{u_0})^T \cdot p = -dg(y_0).$$

Dans le cas présent, on a $f(u, y) = A(u) \cdot y - b$ et $g(y) = \frac{1}{2} \|y - \bar{y}\|^2$, d'où l'écriture du problème adjoint

$$A(u)^T \cdot p = -(y - \bar{y}).$$

Si l'on note p_0 la solution de ce problème, on a

$$dF(u_0) = \partial_u L(y_{u_0}, u_0, p_0),$$

avec

$$L(y, u, p) = g(y) + \langle A(u) \cdot y - b \mid p \rangle. \implies \partial_u L(y_{u_0}, u_0, p_0) = \langle A'(u_0) \cdot y \mid p_0 \rangle,$$

qui est donc l'expression du scalaire $dF(u)$.

8) (Méthode de “backpropagation” pour les réseaux de neurones)

On s'intéresse ici à un réseau de neurones, défini par la donnée de p applications (non linéaires a priori), qui sont les couches du réseau :

$$f_1(y_0, u_1) \in \mathbb{R}^{n_1}, f_2(y_1, u_2) \in \mathbb{R}^{n_2}, \dots, f_p(y_{p-1}, u_p) \in \mathbb{R}^{n_p}, y_i \in \mathbb{R}^{m_i}, u_i \in \mathbb{R}^{n_i}$$

Pour une collection $u = (u_1, \dots, u_p) \in \mathbb{R}^{n_p} \times \dots \mathbb{R}^{n_1}$ de paramètres (les “poids” des différentes couches du réseau de neurones), tout $x \in \mathbb{R}^{n_1}$ donné, on applique successivement les différentes couches : $y_0 = x$, $y_1 = f_1(y_0, u_1)$, ..., $y_p = f_p(y_{p-1}, u_p)$. On note $\Phi(u, x)$ le $y_p \in \mathbb{R}^{n_p}$ obtenu en fin de chaîne. On se donne un certain nombre de données d'entrée $\bar{x}_1, \dots, \bar{x}_K$ dans \mathbb{R}^{n_1} , associées à des données de sorties labellisées $\bar{y}_1, \dots, \bar{y}_K$, et l'on introduit

$$u \longmapsto F(u) = \frac{1}{2} \sum_{k=1}^K \|\Phi(u, \bar{x}_k) - \bar{y}_k\|^2.$$

La phase d'apprentissage d'un tel réseau consiste à trouver des poids u qui minimisent la fonction $F(u)$ (appelée fonction *loss* dans ce contexte) définie ci-dessus. Dans cette optique, il est précieux de pouvoir estimer efficacement le gradient de F . En vous inspirant de la démarche décrite précédemment, proposer une stratégie permettant d'estimer ce gradient en un point u donné. On pourra faire les hypothèses de régularité jugées nécessaires pour appliquer ce qui précède.

V.6 Dérivées d'ordre supérieur

Cette section porte sur les dérivées d'ordre supérieur. Nous nous focalisons au départ sur la différentielle seconde d'une fonction scalaire, et sur la notion de *matrice hessienne* qui permet de la représenter dans une base orthonormée, puis nous présentons un cadre plus abstrait permettant de généraliser ces notions à des applications à valeurs dans un espace multidimensionnel, et de définir une notion de dérivation à un ordre arbitraire.

V.6.1 Dérivées partielles d'ordre supérieur pour les fonctions scalaires

Définition V.6.1. (Dérivées partielles d'ordre 2)

Soit f une application d'un ouvert U de \mathbb{R}^n dans \mathbb{R} . On suppose que f admet des dérivées partielles $\partial f / \partial x_i$ continues sur U (f est donc continûment différentiable d'après la proposition V.1.8). Si chacune de ces dérivées partielles est dérivable en x par rapport à chacune des variables, on appelle dérivées partielles d'ordre 2 les quantités correspondantes, notées

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right).$$

Définition V.6.2. (Matrice hessienne)

Dans le cadre de la définition précédente, on appelle *matrice hessienne* en x , et l'on note $H_f(x)$ (ou plus simplement $H(x)$ s'il n'y a pas d'ambigüité) la matrice carrée dont les éléments sont les dérivées partielles d'ordre 2

$$H(x) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(x) \right).$$

La proposition qui suit, capitale, établit que, si les dérivées secondes sont définies au voisinage d'un point x , et sont continues en ce point, alors la matrice hessienne est symétrique.

Théorème V.6.3. (Schwarz)

Soit f une application d'un ouvert U de \mathbb{R}^n dans \mathbb{R} , et $x \in U$. On suppose que f admet des dérivées partielles d'ordre 2 dans un voisinage de x , et que ces dérivées partielles sont *continues* en x . Alors la matrice hessienne en x est *symétrique*, i.e.

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Démonstration. On considère une fonction de 2 variables seulement (on peut se ramener à ce cas-là en gelant $n - 2$ variables). L'idée est d'écrire de deux manières la quantité

$$f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2),$$

en suivant deux chemins différents entre (x_1, x_2) et $(x_1 + h_1, x_2 + h_2)$. On a en premier lieu

$$f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) = f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) + f(x_1 + h_1, x_2) - f(x_1, x_2).$$

Les 2 derniers termes s'écrivent

$$f(x_1 + h_1, x_2) - f(x_1, x_2) = h_1 \partial_1 f(x_1, x_2) + \frac{h_1^2}{2} \partial_{11} f(x_1 + \theta_1 h_1, x_2),$$

avec $\theta_1 \in]0, 1[$. La première différence du membre de droite s'écrit elle

$$f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) = h_2 \partial_2 f(x_1 + h_1, x_2) + \frac{h_2^2}{2} \partial_{22} f(x_1 + h_1, x_2 + \theta'_2 h_2)$$

Si l'on écrit maintenant la même quantité $f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2)$ de la façon suivante

$$f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) = f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2 + h_2) + f(x_1, x_2 + h_2) - f(x_1, x_2),$$

que l'on utilise des développement de Taylor-Lagrange comme précédemment, et que l'on identifie les deux écritures, on obtient

$$\begin{aligned} 0 &= h_2 h_1 \left(\frac{\partial_2 f(x_1 + h_1, x_2) - \partial_2 f(x_1, x_2)}{h_1} - \frac{\partial_2 f(x_1, x_2 + h_2) - \partial_2 f(x_1, x_2)}{h_2} \right) \\ &+ \frac{h_2^2}{2} (\partial_{22} f(x_1 + h_1, x_2 + \theta'_2 h_2) - \partial_{22} f(x_1, x_2 + \theta_2 h_2)) \\ &+ \frac{h_1^2}{2} (\partial_{11} f(x_1 + \theta'_1 h_1, x_2 + h_2) - \partial_{11} f(x_1 + \theta_1 h_1, x_2)). \end{aligned}$$

Si l'on prend maintenant h_1 et h_2 égaux à ε , et que l'on fait tendre ε vers 0, les deux derniers termes sont des $o(\varepsilon^2)$ par continuité de la dérivée seconde. Le premier terme doit donc lui même être un $o(\varepsilon^2)$, ce qui impose que la quantité entre parenthèse converge vers 0 avec ε , d'où le résultat. \square

Définition V.6.4. (Continue différentiabilité)

Soit f une application d'un ouvert U de \mathbb{R}^n dans \mathbb{R} . On dit que f est deux fois continûment différentiable sur U , et l'on écrit $f \in C^2(U)$, si toutes les dérivées partielles d'ordre 2 de f existent et sont continues sur U , ce qui est équivalent à dire que f admet une matrice hessienne $H(x)$ en tout point x de U , et que la correspondance $x \mapsto H(x)$ est continue.

Remarque V.6.5. En toute rigueur (voir à la fin de la section pour plus de détail), mais au prix de certaines définitions abstraites que nous avons choisi d'écarter, nous devrions définir la différentielle seconde comme l'application différentielle de la différentielle : $d^2 f = d(df)$, c'est à dire comme une application linéaire de \mathbb{R}^n dans l'espace des applications linéaires de \mathbb{R}^n dans \mathbb{R} . Et ensuite dire que l'application est C^2 si cette correspondance est continue, indépendamment des dérivées partielles premières ou secondes afférentes à une base particulière. On peut néanmoins montrer, dans l'esprit de la proposition V.1.8 pour les différentielles d'ordre 1, que la continuité de toutes les dérivées partielles secondes implique le caractère C^2 . Il est donc licite de fonder la définition précédente sur la caractérisation basée sur les dérivées partielles.

Proposition V.6.6. (Développement limité du gradient)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} , deux fois continûment différentiable sur U . On a alors (h est pris suffisamment petit pour que $x + h \in U$)

$$\nabla f(x + h) = \nabla f(x) + H(x) \cdot h + \varepsilon(h) \|h\|.$$

Démonstration. Pour tout $i = 1, \dots, N$, la fonction $y \mapsto \partial_i f(y)$ est continûment différentiable sur U , et l'on a

$$\begin{aligned} \partial_i f(x + h) &= \partial_i f(x) + \langle \nabla \partial_i f(x) | h \rangle + \varepsilon(h) \|h\| \\ &= \partial_i f(x) + \sum_{j=1}^N \partial_j \partial_i f(x) h_j + \varepsilon(h) \|h\| \\ &= \partial_i f(x) + H \cdot h + \varepsilon(h) \|h\|, \end{aligned}$$

qui est l'identité annoncée. \square

Proposition V.6.7. (Développement limité à l'ordre 2)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} , deux fois continûment différentiable sur U . On a alors (h est pris suffisamment petit pour que $x + h \in U$)

$$f(x + h) = f(x) + \langle \nabla f(x) | h \rangle + \frac{1}{2} \langle h | H(x) \cdot h \rangle + \varepsilon(h) \|h\|^2.$$

Démonstration. On introduit la fonction

$$h \longmapsto g(h) = f(x + h) - f(x) - \langle \nabla f(x) | h \rangle - \frac{1}{2} \langle h | H(x) \cdot h \rangle.$$

On a

$$\nabla g(h) = \nabla f(x + h) - \nabla f(x) - H(x) \cdot h = \|h\| \varepsilon(h)$$

d'après la proposition V.6.6. Pour tout $\epsilon > 0$, il existe donc $\eta > 0$ tel que, pour tout h tel que $\|h\| \leq \eta$,

$$\|\nabla g(h)\| \leq \epsilon \|h\|.$$

On applique à présent le théorème des accroissements finis V.1.19 :

$$\|g(h)\| = \|g(h) - g(0)\| \leq \sup_{h' \in [0, h]} \|\nabla g(x + h')\| \|h\| \leq \epsilon \|h\|^2,$$

avec $g(h) = f(x + h) - f(x) - \langle \nabla f(x) | h \rangle - \frac{1}{2} \langle h | H(x) \cdot h \rangle$. \square

Proposition V.6.8. (Développement de Taylor avec reste intégral)

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans \mathbb{R} , deux fois continûment différentiable sur U , et h tel que le segment $[x, x + h] = \{x + \theta h, \theta \in [0, 1]\}$ soit inclus dans U . On a alors

$$f(x + h) = f(x) + \langle \nabla f(x) | h \rangle + \int_0^1 \langle H(x + th) \cdot h | h \rangle (1-t) dt.$$

Démonstration. On pose $\Phi(t) = f(x + th)$. On a, par intégration par parties,

$$\begin{aligned} \Phi(1) &= \Phi(0) + \int_0^1 \Phi'(t) dt = \Phi(0) - [\Phi'(t)(1-t)]_0^1 + \int_0^1 \Phi''(t)(1-t) dt \\ &= \Phi(0) + \Phi'(0) + \int_0^1 \Phi''(t)(1-t) dt. \end{aligned} \tag{V.6.1}$$

On a

$$\Phi(t) = f(x + th),$$

$$\Phi(t + \varepsilon) = f(x + th + t\varepsilon) = f(x + th) + \langle \nabla f(x + th) | h \rangle \varepsilon + o(\varepsilon),$$

et donc

$$\Phi'(t) = \langle \nabla f(x + th) | h \rangle.$$

Enfin

$$\Phi'(t + \varepsilon) = \langle \nabla f(x + th + \varepsilon h) | h \rangle + \langle \nabla f(x + th) | h \rangle + \langle H(x + th) \cdot h | h \rangle \varepsilon + o(\varepsilon),$$

d'où

$$\Phi''(t) = \langle H(x + th) \cdot h | h \rangle.$$

On injecte ces expressions dans (V.6.1), ce qui donne la formule annoncée. \square

Définition V.6.9. (Laplacien)

Soit f une fonction définie sur un ouvert U de \mathbb{R}^n , à valeurs dans \mathbb{R} . On suppose que la matrice hessienne est définie en $x \in U$. On appelle laplacien de f en x , la quantité

$$\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} = \text{tr}(H)$$

(trace de la hessienne de f). Pour les fonctions telles que cette quantité est définie sur U , on appelle *laplacien* cet opérateur noté Δ .

Cet opérateur s'écrit aussi $\Delta = \nabla \cdot \nabla$, où ∇ est le gradient, et $\nabla \cdot$ l'opérateur de divergence : pour tout champ de vecteur $u = (u_1, \dots, u_n)$,

$$\nabla \cdot u = \sum_{i=1}^n \frac{\partial u_i}{\partial x_i} = \text{tr}J,$$

où J est la matrice jacobienne de u vu comme application de \mathbb{R}^n dans \mathbb{R}^n .

V.6.2 Différentielles d'ordre supérieur pour les fonctions de \mathbb{R}^n dans \mathbb{R}^m

La notion de différentielle seconde découle de celle de la différentielle. Comme pour les fonction de \mathbb{R} dans \mathbb{R} , la différentielle seconde sera simplement la différentielle de la différentielle. Pour une fonction de départ de \mathbb{R} dans \mathbb{R}^m , cette différentielle est une application de \mathbb{R}^n dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (définition V.1.4). Si nous souhaitons dériver cette différentielle, nous avons besoin d'une définition un peu plus générale, qui porte sur des applications à valeurs dans un espace vectoriel normé E .

Définition V.6.10. (Différentielle (\bullet))

Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans un espace vectoriel normé E . On dit que f est différentiable en $x \in U$ s'il existe une application linéaire de \mathbb{R}^n dans E , notée $df(x)$, telle que

$$f(x + h) = f(x) + df(x) \cdot h + \varepsilon(h) \|h\| \quad (\text{V.6.2})$$

où $\varepsilon(h)$ est une application de \mathbb{R}^n dans E , telle que $\|\varepsilon(h)\|$ tend vers 0 quand h tend vers 0.

Définition V.6.11. (Différentielle seconde ($\bullet\bullet\bullet$))

Soit f une application différentiable dans un voisinage U d'un point $x \in \mathbb{R}^n$, à valeurs dans \mathbb{R}^m . On dit que f est deux fois différentiable en $x \in U$ si l'application $x \mapsto df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (muni de la norme d'opérateur canonique) est différentiable en x . La différentielle de df en x , notée $d^2f(x)$, est une application linéaire de \mathbb{R}^n dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

On peut de la même manière, si df^2 est définie dans un voisinage de x , définir la différentielle d'ordre 3 par $d^3f = d(df^2)$, qui est une application de \mathbb{R}^n dans $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$, et les différentielles d'ordre $k = 4, 5, \dots$

V.7 Exercices

Exercice V.7.1. Calculer les matrices hessiennes des applications suivantes

$$f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2), \quad f(x, y) = x_1^p x_2^q.$$

CORRECTION.

XXX

Exercice V.7.2. Soit A une matrice carrée d'ordre n , et f l'application quadratique de \mathbb{R}^n dans \mathbb{R} définie par

$$x = (x_1, \dots, x_n) \mapsto f(x) = \langle A \cdot x | x \rangle.$$

- a) Calculer la matrice Hessienne de f .
- b) Dans quel cas cette matrice hessienne est-elle nulle ?

CORRECTION.

a) On a

$$f(x + h) = \langle A \cdot x | x \rangle + \langle A \cdot x | h \rangle + \langle A \cdot h | x \rangle + o(h) = f(x) + \langle (A + A^T) \cdot x | h \rangle + o(h),$$

d'où $\nabla f = (A + A^T) \cdot x$, et par suite $H(x) = A + A^T$.

b) Cette matrice est identiquement nulle si et seulement si la matrice est anti-symétrique.

Exercice V.7.3. Soit f une fonction deux fois continûment différentiable au voisinage d'un point $x \in \mathbb{R}^n$, et h un vecteur de \mathbb{R}^n .

- a) Quelle est la limite de

$$\frac{f(x - \varepsilon h) - 2f(x) + f(x + \varepsilon h)}{\varepsilon^2}$$

quand ε tend vers 0 ?

- b) Soit f une fonction deux fois continûment différentiable sur un ouvert convexe U . On suppose de plus f convexe, c'est-à-dire telle que

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) \quad \forall x, y \in U, \forall \theta \in]0, 1[.$$

Montrer que pour tout x de U , la matrice H est positive, c'est à dire que

$$\langle H(x) \cdot h | h \rangle \geq 0 \quad \forall h \in \mathbb{R}^n.$$

- c) Soit f une fonction deux fois continûment différentiable sur un ouvert convexe U . On suppose que f est λ -convexe, c'est-à-dire telle que

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) - \frac{\lambda}{2}\theta(1 - \theta)\|y - x\|^2 \quad \forall x, y \in U, \forall \theta \in]0, 1[,$$

avec $\lambda \in \mathbb{R}$. Que peut-on en déduire sur la matrice $H(x)$, pour $x \in U$?

- d) (Condition suffisante d'optimalité locale)

On considère f une fonction deux fois continûment différentiable sur un ouvert U . On considère un point $x \in U$ en lequel le gradient de f s'annule, et tel que les valeurs propres de $H(x)$ sont toutes strictement positives. Que peut-on dire de x vis-à-vis de f ?

- e) (Condition suffisante d'optimalité globale)

On suppose maintenant l'ouvert U convexe, et f convexe sur U . Montrer que x minimise f sur U , c'est à-dire-que

$$f(y) \geq f(x) \quad \forall y \in U.$$

- f) (Condition nécessaire d'optimalité)

On considère pour finir f une fonction deux fois continûment différentiable sur un ouvert U . On suppose que $x \in U$ est un minimiseur local de f . Montrer que $\nabla f(x) = 0$, et que $H(x)$ est une matrice positive.

CORRECTION.

a) On fait un développement limité à l'ordre 2 (proposition V.6.7), pour obtenir

$$\frac{f(x - \varepsilon h) - 2f(x) + f(x + \varepsilon h)}{\varepsilon^2} = \langle h | H(x) \cdot h \rangle + o(1),$$

qui converge donc vers $\langle h | H(x) \cdot h \rangle$ quand ε tend vers 0.

b) On écrit l'inégalité de convexité au milieu du segment $[x - \varepsilon h, x + \varepsilon h]$ (pour ε suffisamment petit pour que le segment soit dans l'ouvert U), qui implique la positivité de la quantité $f(x - \varepsilon h) - 2f(x) + f(x + \varepsilon h)$, d'où, en faisant tend ε vers 0,

$$\langle h | H(x) \cdot h \rangle \geq 0 \quad \forall h.$$

c) La même démarche conduit à l'inégalité

$$\langle h | H(x) \cdot h \rangle \geq \lambda \|h\|^2 \quad \forall h.$$

On en déduit donc que, en tout x , la plus petite valeur propre de $H(x)$ (qui est bien réelle car H est symétrique) est minorée par λ . On retrouve bien la positivité de H pour les fonctions convexes (i.e. 0-convexe).

d) Si $\nabla f(x) = 0$ et $\langle H(x) \cdot h | h \rangle > 0$ pour tout h non nul, le développement limité en x assure que x est un minimum local strict pour f , c'est à dire qu'il existe $\eta > 0$ tel que $f(y) > f(x)$ pour tout $y \neq x$ à distance de x inférieure à η .

e) On écrit le développement de Taylor avec reste intégral (proposition V.6.8) entre x et y :

$$f(y) = f(x) + \underbrace{\langle \nabla f(x) | y - x \rangle}_{=0} + \int_0^1 \underbrace{\langle H(x + t(y - x)) \cdot (y - x) | y - x \rangle}_{=0} (1-t) dt \geq f(x).$$

f) D'après le a), on a, pour tout h , $\langle h | H(x) \cdot h \rangle \geq 0$ pour tout h

Exercice V.7.4. Soit f une fonction deux fois continûment différentiable sur un ouvert U de \mathbb{R}^n , et telle que ∇f est de norme constante égale à un sur U . Montrer que

$$H(x) \cdot \nabla f(x) = 0.$$

pour tout x dans U .

CORRECTION.

On a $\|\nabla f\|^2 / 2 \equiv 1$, d'où, pour tout j

$$0 = \partial_i \left(\|\nabla f\|^2 / 2 \right) = \frac{1}{2} \partial_i \left(\sum_{j=1}^n (\partial_j f)^2 \right) = \sum_{j=1}^n \partial_{ij} f \partial_j f$$

qui exprime précisément que la i -ième composante de $H \cdot \nabla f$ est nulle.

Exercice V.7.5. On cherche ici à exprimer le fait que le laplacien quantifie l'écart entre la valeur ponctuelle d'une fonction et la moyenne des valeurs de la fonction au voisinage de ce point. En dimension 1, une telle propriété est donnée par le a) de l'exercice V.7.3. En dimension 2, cette propriété prend la forme exprimée ci-dessous.

Soit f une fonction à valeurs réelles deux fois continûment différentiable au voisinage d'un point $x \in \mathbb{R}^2$. On note e_θ le vecteur unitaire $(\cos \theta, \sin \theta)$. Montrer que

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \frac{1}{2\pi} \int_0^{2\pi} (f(x + \varepsilon e_\theta) - f(x)) d\theta = \frac{1}{4} \Delta f(x).$$

CORRECTION.

On écrit le développement limité à l'ordre 2 de f en x :

$$f(x + \varepsilon e_\theta) = f(x) + \varepsilon \nabla f(x) \cdot e_\theta + \frac{\varepsilon^2}{2} \langle H \cdot e_\theta | e_\theta \rangle + o(\varepsilon^2).$$

Si l'on intègre $f(x + \varepsilon e_\theta) - f(x)$ entre 0 et 2π , le premier terme (avec le gradient), donne

$$\int_0^{2\pi} \nabla f(x) \cdot e_\theta d\theta = \nabla f(x) \cdot \int_0^{2\pi} e_\theta d\theta = 0.$$

Le terme d'ordre 2 est la somme de 4 contributions. La première s'écrit

$$\frac{1}{2} \int_0^{2\pi} \partial_{11} f(x) \cos(\theta)^2 = \frac{\pi}{2} \partial_{11} f(x).$$

L'autre contribution diagonale (en $\sin(\theta)^2$), vaut de la même manière $\partial_{22} f \pi/2$. Les contributions extra-diagonales sont multiples de l'intégrale de $\sin(\theta) \cos(\theta) = \sin(2\theta)/2$, dont l'intégrale sur $[0, 2\pi]$ vaut 0. On a donc finalement

$$\frac{1}{\varepsilon^2} \frac{1}{2\pi} \int_0^{2\pi} (f(x + \varepsilon e_\theta) - f(x)) \longrightarrow \frac{1}{4} (\partial_{11} f(x) + \partial_{22} f(x)) = \frac{1}{4} \Delta f(x).$$

Exercice V.7.6. (Fonctions holomorphes / harmoniques)

On se place dans le cadre de l'exercice V.2.4, page 112. On écrit f , fonction de l'ouvert $\Omega \subset \mathbb{C}$ dans \mathbb{C} , sous la forme

$$f(x + iy) = P(x, y) + iQ(x, y)$$

où P et Q sont à valeurs réelles. Montrer que, si f est holomorphe, alors P et Q vérifient

$$\partial_x P - \partial_y Q = \partial_y P + \partial_x Q = 0.$$

En déduire que P et Q sont harmoniques, i.e. de laplacien nul :

$$\Delta Q = \Delta P = \left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial x^2} \right) P = 0.$$

N.B. on pourra admettre que P et Q sont deux fois continûment différentiables (il s'agit d'une conséquence du caractère holomorphe de f).

CORRECTION.

On a

$$f(x + iy) = P(x, y) + iQ(x, y),$$

d'où

$$\partial_x f = \partial_x P + i\partial_x Q, \quad \partial_y f = i\partial_y P - \partial_y Q.$$

La relation (V.2.1) s'écrit ainsi

$$0 = \partial_x f + i\partial_y f = \partial_x P - \partial_y Q - i(\partial_y P + \partial_x Q),$$

d'où les relations

$$\partial_x P - \partial_y Q = \partial_y P + \partial_x Q = 0.$$

On a donc

$$\partial_{xx} P = \partial_{xy} Q, \quad \partial_{yy} P = -\partial_{xy} Q,$$

d'où l'harmonicité de P , et de la même manière celle de Q .

Chapitre VI

Développements

Sommaire

VI.1 Entiers p -adiques, espaces ultramétriques	141
VI.2 Dendrogrammes	149
VI.3 Introduction au transport optimal	151
VI.3.1 Problème d'affectation et problème de Monge Kantorovich discret	151
VI.3.2 Transport optimal, cas général	153
VI.4 Distance de Gromov-Wasserstein	154
VI.5 Propagation d'opinion et flot de gradient	155
VI.6 Modèles macroscopiques de trafic routier	161
VI.7 Autour de la notion de complétude, théorème de Banach-Steinhaus	164
VI.7.1 Lemme de Baire	164
VI.7.2 Théorème de Banach Steinhaus	165
VI.7.3 Théorème de l'application ouverte, théorème du graphe fermé	166

VI.1 Entiers p -adiques, espaces ultramétriques

Cette section est consacrée à la construction du complété de \mathbb{N} pour une certaine métrique, appelée p -adique. Nous détaillons cette construction pour le cas $p = 2$, qui aboutit à l'espace dit des entiers 2-adiques, noté \mathbb{Z}_2 , dont on présente ci-dessous certaines propriétés.

Distance 2-adique sur \mathbb{N}

Définition VI.1.1. (Valuation et valeur absolue 2-adiques)

Tout $a \in \mathbb{Z}$ non nul peut s'écrire de façon unique $a'2^k$, où $a' \in \mathbb{Z}$ est impair. On appelle $v_2(a) = k$ la *valuation* 2-adique de a , et l'on définit la valeur absolue 2-adique

$$|a|_2 = 2^{-v_2(a)}.$$

On pose $v_2(0) = +\infty$, et $|0|_2 = 0$.

On se place maintenant sur $X = \mathbb{N}$ l'ensemble des entiers naturels, et l'on définit

$$(a, b) \in \mathbb{N}^2 \longmapsto d(a, b) = |b - a|_2.$$

On peut donner une expression équivalente de cette définition, en considérant que tout entier naturel a s'écrit (écriture dyadique)

$$a = \sum_{n=0}^{+\infty} a_n 2^n, \quad a_n \in \{0, 1\}, \quad a_n \text{ nul au delà d'un certain rang.} \quad (\text{VI.1.1})$$

Deux nombres a et b peuvent ainsi être écrits en base 2 sous la forme de suites finies (a_n) et (b_n) de 0 et de 1, et l'on a $d(a, b) = 2^{-n_{ab}}$, où

$$n_{ab} = \min \{n, a_n \neq b_n\}, \quad (\text{VI.1.2})$$

pris égal à $+\infty$ si les bits de a et b s'identifient (i.e. si $a = b$). En effet, par définition de n_{ab} , on a

$$a - b = 2^{n_{ab}} c,$$

où $c \in \mathbb{Z}$ est impair.

Proposition VI.1.2. L'application $d(\cdot, \cdot)$ est une distance sur \mathbb{N} , qui est *ultramétrique*, c'est-à-dire qu'elle vérifie l'inégalité triangulaire renforcée :

$$d(a, b) \leq \max(d(a, c), d(b, c)) \quad \forall a, b, c \in X.$$

Démonstration. La séparation est immédiate par hypothèse, et l'on a bien $d(a, b) = d(b, a)$. Pour l'inégalité triangulaire renforcée, on utilise la formulation (VI.1.2). Pour tous entiers a, b, c , on considère les écritures dyadiques associées. Les $n_{ab} - 1$ premiers bits de a et b s'identifient par définition, ainsi que les $n_{bc} - 1$ premiers bits de b et c . Les premiers bits de a et c s'identifient donc au moins sur les $\min(n_{ab}, n_{bc}) - 1$ indices, d'où

$$n_{ac} \geq \min(n_{ab}, n_{bc}),$$

et ainsi

$$d(a, c) = 2^{-n_{ac}} \leq \max(2^{-n_{ab}}, 2^{-n_{bc}}) = \max(d(a, b), d(b, c)),$$

qui est l'inégalité ultramétrique annoncée, qui entraîne l'inégalité triangulaire usuelle. \square

Cette distance munit \mathbb{N} d'une distance, appelée *ultramétrique*, aux propriétés particulières (voir proposition VI.1.6 et les suivantes).

Le diamètre de (\mathbb{N}, d) est 1, et ce diamètre est atteint en de multiples couples : tout entier impair est en particulier diamétralement opposé à 0, ainsi qu'à tout entier pair. La sphère unité (sphère de centre 0 et de rayon 1) est l'ensemble des nombres impairs. De façon générale, la sphère de centre 0 et de rayon 2^{-k} est l'ensemble des nombres du type $2^k b$, avec b impair. Ces sphères de centre 0 et de rayons 1, $1/2$, $1/4$, ..., et 0, réalisent une partition de \mathbb{N} (comme un emboîtement infini de poupées russes).

La suite (2^n) tend vers 0, ainsi que toute suite telle que l'exposant de 2 dans la décomposition en facteurs premiers des termes tend vers $+\infty$.

Nous terminons cette section par une propriété négative sur (\mathbb{N}, d) , qui justifie la démarche de complémentation qui suit.

Proposition VI.1.3. L'espace ultramétrique (\mathbb{N}, d) n'est pas complet.

Démonstration. Considérons la suite $(2^n - 1)$. Cette suite est de Cauchy pour $d(\cdot, \cdot)$. Si elle converge vers $a \in \mathbb{N}$, alors 2^n converge vers $a + 1$, d'où $a + 1 = 0$, équation qui n'admet pas de solution dans \mathbb{N} . \square

Remarque VI.1.4. Noter que l'on aurait pu choisir comme espace de départ $X = \mathbb{Z}$ au lieu de \mathbb{N} , auquel cas le contre-exemple ci-dessus n'en est plus un, puisque $a + 1$ admet une solution dans \mathbb{Z} . Pour se convaincre qu'il manque plus à (\mathbb{N}, d) que les nombres négatifs pour être complet, on peut considérer par exemple la suite

$$a_N = \sum_{n=0}^{2N} a_n 2^n,$$

avec $a_n = n + 1 \bmod 2$ (alternance périodique de 1 et de 0). On a

$$a_N = \sum_{n=0}^N 2^{2n} = \frac{1 - 4^{N+1}}{1 - 4} = \frac{4^{N+1} - 1}{3}.$$

Cette suite est de Cauchy (comme toute série partielle de ce type). S'il existe $a \in \mathbb{N}$ tel que a_N converge vers a , alors on a

$$\left| \frac{4^{n+1} - 1}{3} - a \right|_2 \rightarrow 0 \Rightarrow |4^{n+1} - 1 - 3a|_2 = 0,$$

d'où $3a + 1 = 0$, équation qui n'admet pas de solution dans \mathbb{N} , ni dans \mathbb{Z} .

Complété de \mathbb{N}

Définition VI.1.5. Le complété de \mathbb{N} (voir théorème A.2.4) pour la distance définie ci-dessus est appelé ensemble des entiers 2-adiques. Il est noté \mathbb{Z}_2 .

L'écriture (VI.1.1), qui permet de représenter tout entier comme une suite finie de 0 ou de 1 (écriture en base 2), permet de se faire une meilleure idée de \mathbb{Z}_2 , et de préciser à quoi correspondent certains de ses éléments. Considérons une suite (a^k) dans \mathbb{N} , de Cauchy pour $d(\cdot, \cdot)$. On peut écrire les termes

$$a^k = (a_0^k, a_1^k, a_2^k, \dots)$$

Si une suite est de Cauchy alors, comme dans le cas des réels en écriture décimale (voir la démonstration de la proposition A.1.39), chacune des suites $(a_n^k)_k$ finit par se stabiliser à une valeur a_n . On peut donc représenter la limite par une suite (infinie) de 0 et de 1. On écrira le nombre correspondant somme somme d'une série, ou en écriture flottante, avec la convention de placer les bits *avant* la virgule :

$$a = \sum_{n=0}^{+\infty} a_n 2^n, \quad \text{ou } a = \dots a_3 a_2 a_1 a_0, 0.$$

Considérons par exemple le nombre $\dots 111, 0$. On a

$$\dots 111, 0 = \lim_{N \rightarrow \infty} \left(\sum_{n=0}^N 2^n \right) = \lim_{N \rightarrow \infty} \left(\frac{1 - 2^{N+1}}{1 - 2} \right) = -1 - \lim_{N \rightarrow \infty} 2^{N+1}$$

qui tend donc selon les règles de calcul usuel vers un nombre qui n'est pas dans l'espace de départ, et qu'il est tentant de noter -1 .

La métrique induite sur \mathbb{Z}_2 est définie essentiellement de la même manière que sur \mathbb{N} . Pour deux éléments a et b de \mathbb{Z}_2 , si l'on note n_{ab} le plus petit indice pour lequel les bits diffèrent ($n_{ab} = +\infty$ si $a = b$), la distance entre a et b est simplement définie par $d(a, b) = 2^{-n_{ab}}$.

La distance sur \mathbb{Z}_2 définie ci-dessus hérite des propriétés ultramétriques de la distance de départ sur \mathbb{N} . L'espace \mathbb{Z}_2 possède donc des propriétés propres aux espaces ultramétriques, telles que celles énoncées ci-après.

Espaces ultramétriques généraux

Proposition VI.1.6. Soit (X, d) un espace ultramétrique. Tout point d'une boule (ouverte ou fermée) est centre de cette boule.

Démonstration. Soient $x \in X$ et $r \in \mathbb{R}_+$. On considère $x' \in B_f(x, r)$, i.e. tel que $d(x, x') \leq r$. Pour tout $y \in B_f(x', r)(x, r)$, on a

$$d(y, x') \leq \max(d(y, x), d(x, x')) \leq r, \quad \text{d'où } y \in B_f(x', r).$$

On a donc $B_f(x, r) \subset B_f(x', r)$, et on a l'inclusion inverse en intervertissant les rôles de x et x' . La démonstration est identique pour une boule ouverte. \square

Corollaire VI.1.7. Soit (X, d) un espace ultramétrique. L'intersection entre deux boules est soit vide soit l'une des deux boules. Si les rayons sont les mêmes, deux boules sont donc soit disjointes, soit identiques.

Démonstration: Soit $B(x, r)$ et $B(x', r')$. Si y appartient aux deux boules, il est aussi centre de ces deux boules d'après ce qui précède, la plus petite est donc incluse dans la plus grande, avec égalité si les rayons sont les mêmes.

Proposition VI.1.8. Soit (X, d) un espace ultramétrique. Toute boule ouverte est un fermé.

Démonstration. Si $r = 0$ (alors $B(x, r) = \emptyset$) ou si $B(x, r) = X$, c'est immédiat. Sinon, pour tout $y \in B(x, r)^c$, $B(x, r) \cap B(y, r) = \emptyset$ d'après la proposition précédente, d'où $B(y, x) \subset B(x, r)^c$. Le complémentaire de $B(x, r)$ est donc un ouvert, la boule ouverte elle-même est donc un fermé. \square

Proposition VI.1.9. Un espace ultramétrique est totalement discontinu, au sens où chaque point s'identifie à sa propre composante connexe (définition A.2.17).

Démonstration. Soient x et $y \neq x$ dans X , et $0 < r < d(x, y)$. La boule ouverte $B(x, r)$ et son complémentaire (qui est aussi un ouvert) réalisent une partition de X en 2 ouverts, x et y ne peuvent donc appartenir à la même composante connexe. \square

Proposition VI.1.10. Soit (X, d) un espace ultramétrique. Tout triangle dans X est isocèle (avec deux grands côtés et un petit côté).

Proposition VI.1.11. On considère trois points dans X , et l'on note ℓ_1 , ℓ_2 , et ℓ_3 les longueurs des côtés. On a $\ell_1 \leq \max(\ell_2, \ell_3)$, et les mêmes propriétés obtenues en permutant les indices. Si les longueurs sont distinctes deux à deux, on a par exemple $\ell_1 < \ell_2 < \ell_3$, qui invalide $\ell_3 \leq \max(\ell_1, \ell_2)$. Deux des longueurs au moins sont donc identiques, par exemple $\ell_1 = \ell_2$. Le troisième côté est de longueur $\ell_3 \leq \max(\ell_1, \ell_2) = \ell_1$.

Proposition VI.1.12. Soit (X, d) un espace ultramétrique. Une suite (x_n) est de Cauchy si et seulement si

$$\lim_{n \rightarrow +\infty} d(x_{n+1}, x_n) = 0.$$

Exemple VI.1.1. On se reportera à la section VI.2 pour la description d'un procédé qui permet de construire un espace ultramétrique à partir d'un espace métrique fini, par une démarche de *clustering* très utilisée en analyse de données.

Retour sur l'espace des entiers 2-adiques

Addition sur \mathbb{Z}_2 . On peut définir une addition sur \mathbb{Z}_2 , qui est “l'addition de l'écolier” en partant de la droite dans l'écriture 2-adique ci-dessus. On considère deux entiers dyadiques a et b , et l'on cherche à construire la somme $c = a + b$, qui étende la somme sur \mathbb{N} . Si $a_0 + b_0 = 0$ ou 1 , on affecte à c_0 cette valeur, et on passe au rang suivant. Si la somme vaut 2 on pose $c_0 = 0$, et l'on garde une retenue de 1 pour le rang suivant. Au rang 1 on se retrouve dans la même situation, avec éventuellement une retenue de 1 en plus. La somme peut donc maintenant valoir 3 . Si c'est le cas on pose $c_1 = 1$ et on garde 1 de retenue pour le rang suivant, etc ...

De façon assez frappante, le fait de prendre le complété de \mathbb{N} pour la distance choisie, qui est une démarche purement topologique, conduit à espace qui possède une structure algébrique que l'espace de départ n'avait pas.

Proposition VI.1.13. L'espace \mathbb{Z}_2 muni de l'addition définie ci-dessus est un groupe additif.

Proposition VI.1.14. L'espace \mathbb{Z}_2 n'est pas dénombrable.

Démonstration. Nous avons vu que tout élément de \mathbb{Z}_2 peut se représenter de façon unique par une suite infinie de 0 ou de 1 . On peut donc identifier (d'un point de vue ensembliste) \mathbb{Z}_2 à l'ensemble des parties de

\mathbb{N} , en considérant la suite (a_n) comme la fonction indicatrice d'une partie de \mathbb{N} . L'ensemble \mathbb{Z}_2 n'est donc pas dénombrable d'après le théorème A.1.10. \square

Définition VI.1.15. (Ordre lexicographique)

On peut définir sur \mathbb{Z}_2 un ordre *lexicographique*, en considérant, pour deux éléments différents $\dots a_2 a_1 a_0, 0$ et $\dots b_2 b_1 b_0, 0$ le plus petit indice n pour lequel les deux diffèrent, et poser $a < b$ si $a_n < b_n$.

Noter que l'ordre défini ci-dessus est très différent de l'ordre usuel. Comme $\mathbb{Z} \subset \mathbb{Z}_2$, on peut comparer de ce point de vue deux éléments de \mathbb{Z} , on retrouve certaines propriétés usuelles du type $1 < 3$, $1 < 5$, mais aussi des choses plus déroutantes, comme $5 < 3$ et, plus globalement,

$$0 \leq a \leq -1 \quad \forall a \in \mathbb{Z},$$

ce qui permet d'écrire

$$\mathbb{Z} = [0, -1].$$

Proposition VI.1.16. L'espace \mathbb{Z}_2 est compact.

Démonstration. Considérons une suite (a^k) dans \mathbb{Z}_2 . On considère la suite $(a_0^k)_k$ de 0 et de 1. Cette suite visite une infinité de fois 0 ou 1 (ou les deux). On prend a_0 égal à une valeur visitée une infinité de fois, et l'on extrait la sous-suite $(a^{\varphi_0(k)})$ correspondante. On procède de même avec $(a_1^{\varphi_0(k)})$, pour extraire une suite $(a^{\varphi_0 \circ \varphi_1(k)})$. On extrait ainsi des sous-suites emboitées les unes dans les autres. On définit maintenant (selon le processus d'extraction diagonale dit de Cantor)

$$\varphi(k) = \varphi_0 \circ \varphi_1 \circ \dots \circ \varphi_k(k).$$

La suite ainsi construite est telle que $a_n^{\varphi(k)}$ a une valeur constante $a_n \in \{0, 1\}$ au-delà d'un certain rang (pour tout $k \geq n$), on a donc convergence de cette suite extraite vers $a = \dots a_3 a_2 a_1 a_0, 0$. \square

Système projectif, limite projective (••••)

Un *système projectif* désigne une famille (X_n) d'ensemble muni d'une famille d'applications $(f_n^m)_{n \leq m}$, avec $f_n^m : X_m \rightarrow E_n$, qui vérifient les propriétés suivantes

- (1) L'application f_n^n est l'identité sur X_n
- (2) Pour tous $n \leq m \leq q$, on a $f_n^m \circ f_m^q = f_n^q$.

On appelle limite projective l'ensemble des éléments du produit infini $X_0 \times X_1 \times \dots$ dont les projections sont compatibles avec les f_i^j au sens suivant :

$$\varprojlim \left((X_n), (f_i^j) \right) = \{ x = (x_0, x_1, x_2, \dots) \in X_0 \times X_1 \times \dots, f_n^m(x_m) = x_n \quad \forall i \leq j \}.$$

Pour tous $n \leq m$, on peut définir canoniquement¹ une surjection de $\mathbb{Z}/2^m\mathbb{Z}$ dans $\mathbb{Z}/2^n\mathbb{Z}$. On note f_n^m cette surjection. On peut identifier ainsi \mathbb{Z}_2 à la limite projective $((\mathbb{Z}/2^n\mathbb{Z}), (f_n^m))$.

Représentation des arbres dyadiques, application au poumon humain

Le poumon humain se présente comme un arbre constitué de bronches (appelée bronchioles pour les plus petites), structuré de façon dyadique : la trachée se divise en deux, chacune des branches filles se divise elle-même en deux, etc... Pour l'arbre respiratoire d'un adulte, le nombre de bifurcations est de l'ordre de 23, soit autour de $2^{23} \approx 8 \times 10^8$ bronchioles terminales, dont on appelle *feuilles* les extrémités libres. Pour diverses raisons (construction d'un modèle homogénéisé du parenchyme, construction d'un opérateur de la ventilation, qui à un champ de pression aux feuilles associe un champ de flux, etc ...), il peut être intéressant

1. On peut considérer la relation d'équivalence sur $\mathbb{Z}/2^m\mathbb{Z}$ définie par $z \mathcal{R} z' \iff z \equiv z' [2^n]$. L'espace quotient s'identifie à $\mathbb{Z}/2^n\mathbb{Z}$.

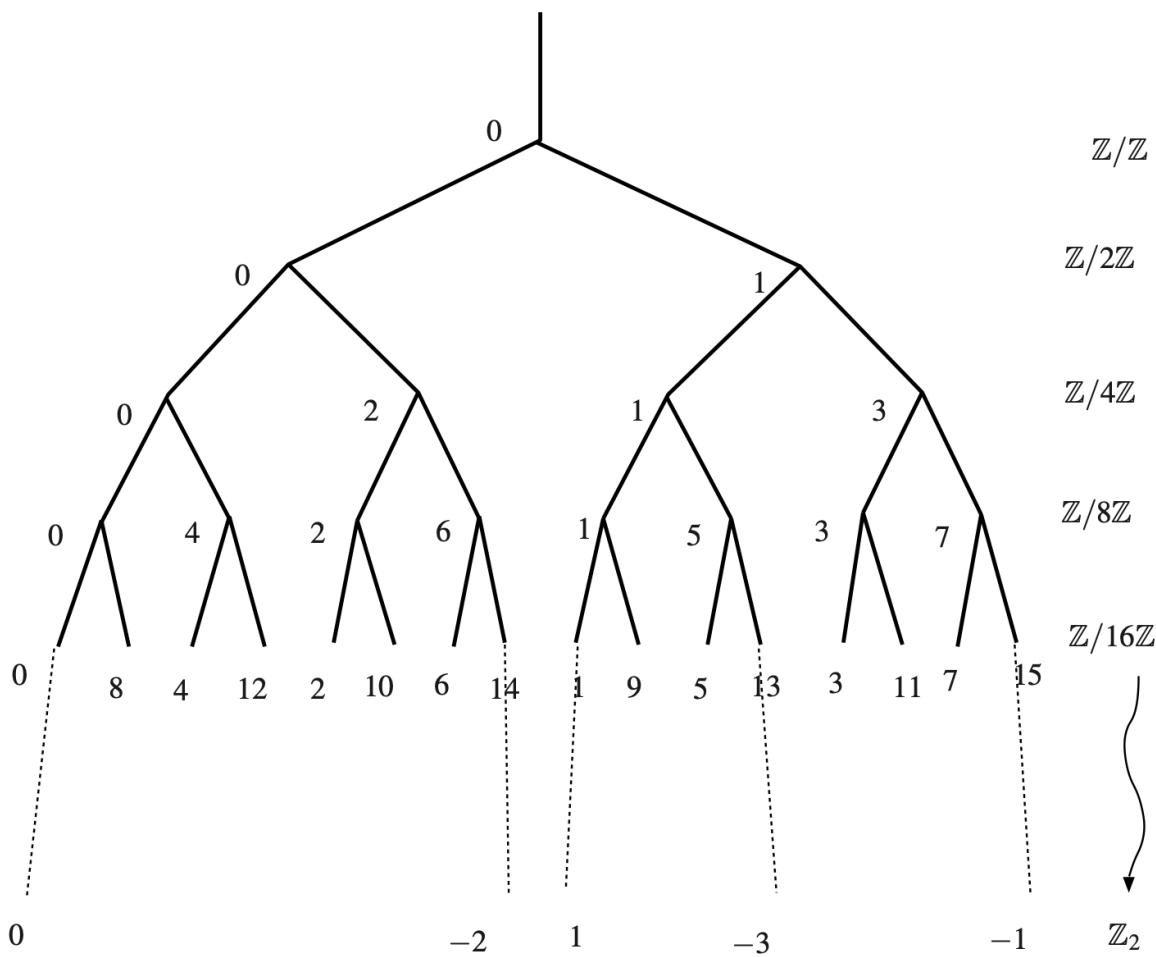


FIGURE VI.1.1 – Numérotation 2-adique

d'extrapoler cet arbre vers un nombre de générations infinis. La notion de feuille est alors remplacée par celle de *bout*, on parle de l'*espace des bouts*, chacun de ces bouts correspondant à un chemin centrifuge s'éloignant de la racine (entrée de la trachée). Il est naturel de coder chacun de ces chemins par une suite infinie de 0 et de 1, par exemple selon la convention par exemple (on se figurera une représentation avec la racine en haut, comme sur la figure VI.1.1) : partant de la racine, si l'on part à gauche, on prend $a_0 = 0$, et $a_0 = 1$ si c'est à droite. On encode le "choix" à chaque étape par un nombre entre 0 et 1.

La figure VI.1.1 représente la numérotation obtenue à chaque génération, où l'on a représenté chaque mot $(a_N \dots, a_0)$ par sa représentation entière $a = \sum a_n$.

On notera que cette numérotation ne correspond pas du tout à l'énumération linéaire 0, 1, Cette numérotation présente un énorme avantage par rapport à l'énumération linéaire, en celà qu'elle respect la structure de l'arbre. Plus précisément, considérons deux feuilles de la 4-ième génération, d'indices a et b . Leur proximité *vis à vis de l'arbre*, qui correspondrait à un degré de parenté s'il s'agissait d'un arbre généalogique, ne dépend que de $|a - b|$, contrairement à ce qui se passerait pour la numérotation linéaire, comme on peut s'en convaincre facilement. Par exemple deux feuilles dont la différence d'indices est impaire appartiennent nécessairement à des lobes différents. Si la différence est divisible par 2, et pas par 4, les deux point appartiennent nécessairement tous deux à l'un des 4 sous-arbres issus de la génération 2, etc...

Plus précisément, si l'on considère l'arbre à 4 générations représenté sur la figure, qui présente $2^4 = 16$ feuilles, on peut identifier la métrique dyadique sur l'ensemble des feuilles (identifié à $\llbracket 0, 15 \rrbracket$ ou $\mathbb{Z}/2^4\mathbb{Z}$) avec la métrique du plus court chemin au travers de l'arbre considéré comme un graphe pondéré. Pour retrouver

exactement la distance 2-adique, on peut vérifier qu'il suffit de considérer que les arêtes des deux dernières générations (3 et 4) ont pour longueur 1/16, longueur 1/8 pour la génération 2 (4 branches), et 1/4 pour la génération 1.

Si l'on considère maintenant le poumon infini, on voit apparaître des indices déjà identifiés pour certains bouts. Par exemple le bout le plus à gauche est clairement 0. Pour le bout correspondant au chemin le plus à droite, ou retrouve notre $\dots 1111, 0 = -1$. Noter que -1 est le plus "grand" élément de \mathbb{Z}_2 pour la relation d'ordre lexicographique (voir définition VI.1.15), et 0 le plus petit. De façon générale, la représentation de la figure VI.1.1 correspond à cet ordre (croissant de la gauche vers la droite).

On notera que, pour tout sous-arbre infini, le bout le plus à gauche est un entier positif, et le bout le plus à droite un entier négatif.

Nombres entiers p -adiques

On peut généraliser cette approche en remplaçant 2 par un nombre p premier quelconque : tout $a \in \mathbb{Z}$ non nul peut s'écrire de façon unique $a'p^k$, où $a' \in \mathbb{Z}$ n'est pas divisible par p . On appelle $v_p(a) = k$ la *valuation p -adique* de a , et l'on définit la valeur absolue p -adique

$$|a|_p = p^{-v_p(a)}.$$

On pose $v_p(0) = +\infty$, et $|0|_p = 0$. Cette valeur absolue vérifie $|ab|_p = |a|_p |b|_p$.

On peut définir de la même manière que précédemment une distance d_p sur \mathbb{N} , et considérer le complété de \mathbb{N} pour cette distance, que l'on note \mathbb{Z}_p .

L'identification des éléments de \mathbb{Z}_p peut se faire à partir de la décomposition d'un entier en base p :

$$a = \sum_{n=0}^{+\infty} a_n p^n, \quad a_n \in \llbracket 0, p-1 \rrbracket \quad \forall n, \quad a_n = 0 \text{ au-delà d'un certain rang.}$$

Un élément de \mathbb{Z}_p peut ainsi s'écrire comme une série infinie du type de celle ci-dessus, ou simplement codée par ses coefficients (on garde la convention d'une infinité de chiffres *avant* la virgule)

$$a = \dots a_2 a_1 a_0, 0.$$

Noter que cette complétion et l'identification à des nombres écrits en base p avec une infinité de chiffres avant la virgule ne nécessite pas que p soit premier. Si p n'est pas premier, on perd la propriété $|ab|_p = |a|_p |b|_p$, par exemple $|2 \times 5|_{10} = 10^{-1}$ alors que $|2|_{10} = |5|_{10} = 1$. Mais tant que l'on s'en tient à des aspects métriques, et que l'on se contente d'additionner les nombres entre eux, la construction est valide. On peut en particulier construire l'objet \mathbb{Z}_{10} des entiers 10-adiques, qui présente une forte analogie apparente (le codage semble le même, à symétrie près) avec l'ensemble des réels de l'intervalle $[0, 1]$, mais qui présente des propriétés très différentes. Pour l'anecdote, si l'on considère la relation d'ordre lexicographique sur \mathbb{Z}_{10} , son plus grand nombre est (on notera que l'infinité de 9 avant la virgule n'est pas ici pathologique, le codage de \mathbb{Z}_{10} est tout à fait injectif, contrairement au codage décimal des réels)

$$\dots 9999, 0 = \sum_{n=0}^{+\infty} 9 \times 10^n = 9 \frac{1}{1-10} = -1,$$

de telle sorte que l'on peut écrire $\mathbb{Z}_{10} = [0, -1]$ (*sic*).

On se restreint néanmoins en général aux nombres premiers, qui permettent des développements très féconds dans un cadre algébrique. Si l'on se restreint ainsi aux nombres premiers, on a une formule à la fois spectaculaire et très simple à établir, qui relie toutes les valeurs absolues p -adiques entre elles. Plus précisément, si l'on note $|\cdot|_\infty$ la valeur absolue usuelle sur \mathbb{Z} , on a

$$\forall a \in \mathbb{Z}, \quad |a|_\infty \prod_{p \text{ premier}} |a|_p = 1.$$

On notera que pour p assez grand (notamment plus grand que $|a|$), on a $|a|_p = 1$, le produit ci-dessus peut donc en fait se ramener à un produit fini, dont la définition ne nécessite donc pas d'arguments topologiques.

Mesure sur \mathbb{Z}_2 , sur \mathbb{Z}_p

Nous décrivons dans cette section la démarche permettant de construire une mesure sur \mathbb{Z}_2 . Le rôle joué par les intervalles dans le cas réel est ici joué par les boules. Nous privilégierons les boules fermées $B_f(a, 2^{-k})$, en gardant en tête que ce sont aussi des boules ouvertes. En effet les valeurs prises par la distance étant quantifiées (ce sont les 2^{-k}), on a $B(a, 2^{-k}) = B_f(a, 2^{-(k-1)})$. On notera que la boule fermée de centre a et de rayon 2^{-k} s'écrit aussi

$$a + 2^k \mathbb{Z}_2 = \{a + 2^k z, z \in \mathbb{Z}_2\} = \{\dots a_{k-1} a_{k-2} \dots a_0\}$$

c'est-à-dire l'ensemble des éléments de \mathbb{Z}_2 dont l'écriture 2-adique commence comme celle de a (au moins jusqu'au rang $k-1$).

Nous noterons $\mathcal{B} = \mathcal{B}(\mathbb{Z}_2)$ la tribu des boréliens sur \mathbb{Z}_2 , i.e. la tribu engendrée par les ouverts de \mathbb{Z}_2 .

Proposition VI.1.17. La tribu \mathcal{B} des boréliens est engendrée par les $a + 2^k \mathbb{Z}_2$, avec a et k parcourant \mathbb{N} .

Démonstration. Montrons en premier lieu que tout ouvert est réunion dénombrable de boules (fermées!). Soit un ouvert U de \mathbb{Z}_2 , et $a \in U$. Il existe k tel que $a + 2^k \mathbb{Z}_2 \subset U$. Notons $\bar{a} \in \mathbb{N}$ la troncature de a au rang k , i.e.

$$\bar{a} = \sum_{n=0}^{k-1} a_n 2^n.$$

On a $a + 2^k \mathbb{Z}_2 = \bar{a} + 2^k \mathbb{Z}_2$. L'ouvert U est donc union de boules du type $a + 2^{-k} \mathbb{Z}_2$, il s'agit donc d'une union dénombrable². Cette propriété implique que la tribu engendrée par les $a + 2^k \mathbb{Z}_2$, contient la tribu des boréliens, et donc s'identifie à elle du fait qu'il s'agit d'ouverts. \square

On cherche maintenant à définir une mesure μ sur \mathbb{Z}_2 , qui affecte une masse 1 à \mathbb{Z}_2 . Nous verrons qu'il est possible de définir une telle mesure sur la tribu borélienne. Du fait que \mathbb{Z}_2 est union disjointe de 2^k boules fermées de rayon 2^{-k} , on affecte le volume 2^{-k} à toute boule fermée de rayon 2^{-k} . Il peut sembler étonnant d'affecter un volume égal au *rayon*, mais on se souviendra que dans ce contexte ultramétrique, le rayon est égal au diamètre.

On se propose maintenant de construire, à partir de cette définition du volume sur le π -système des $a + 2^k \mathbb{Z}_2$, une mesure extérieure sur \mathbb{Z}_2 , en suivant la démarche décrite sur \mathbb{R} dans la proposition B.4.6, page 207.

Proposition VI.1.18. Pour tout $A \subset \mathbb{Z}_2$, on note C_A l'ensemble des suites de boules fermées dont l'union recouvre A :

$$C_A = \left\{ (a_i + 2^{k_i} \mathbb{Z}_2)_{i \in \mathbb{N}}, A \subset \bigcup_{\mathbb{N}} (a_i + 2^{k_i} \mathbb{Z}_2) \right\}.$$

On autorise les rayons à être nul (i.e. $k = +\infty$), ce qui autorise à considérer des collections avec un nombre fini de boules de volume non nul. On définit alors $\mu^* : \mathcal{P}(\mathbb{Z}_2) \rightarrow [0, +\infty]$ par

$$\lambda^*(A) = \inf_{C_A} \left(\sum_i 2^{-k_i} \right). \quad (\text{VI.1.3})$$

Cette application est une mesure extérieure, et elle attribue à toute boule fermée de rayon 2^{-k} la valeur 2^{-k} .

Démonstration. Cette démonstration est parfaitement analogue³ à celle de la proposition B.4.6, page 207. En premier lieu $\mu^*(\emptyset) = 0$. Pour la monotonie,

2. Noter, même si ça n'est pas nécessaire dans la démonstration, qu'on peut ne garder qu'une seule boule par $a \in \mathbb{N}$ impliqué, du fait que deux boules sont soit disjointes soit concentriques (l'une dans l'autre). On peut donc "coder" un ouvert de \mathbb{Z}_2 par une suite $(a_i, k_i)_i$ dans \mathbb{N}^2 .

3. Elle s'applique d'ailleurs à toute construction d'une mesure extérieure par l'extérieur précisément, basée sur ce principe de recouvrement par des objets donc on a fixé la taille.

Considérons maintenant une collection (A_n) de parties de \mathbb{Z}_2 . On peut trouver pour chaque partie une collection de boules qui la recouvre, et telle que la somme des volumes approche $\mu^*(A_n)$ à $\varepsilon/2^n$ près. Comme dans la démonstration de la proposition B.4.6, page 207, cela permet d'établir que

$$\mu^*\left(\bigcup A_n\right) \leq \sum \mu^*(A_n) + 2\varepsilon,$$

et ce pour tout ε . On en déduit la sous-additivité.

Montrons maintenant que cette mesure extérieure affecte 2^{-k} aux boules fermées de rayon 2^{-k} . Soit $B = a + 2^k \mathbb{Z}_2$ une telle boule. En premier lieu, la boule est recouverte par elle-même, d'où $\mu^*(B) \leq 2^{-k}$. Maintenant considérons un recouvrement de B par des boules. Ces boules fermées étant des ouverts, et B étant compacte, on peut en extraire un recouvrement fini, on peut enlever les boules qui sont incluses dans une autre du recouvrement, pour obtenir (d'après le corollaire VI.1.7) une partition finie, telle que la somme des volumes est exactement 2^{-k} . \square

On considère maintenant \mathcal{A} la tribu des parties mesurables pour μ^* (selon la définition B.4.2, page 204), et l'on note μ la mesure définie sur \mathcal{A} issue de μ^* , selon le théorème B.4.4, page 205. Il s'agit maintenant de montrer que la tribu \mathcal{A} contient la tribu des boréliens. Comme dans le cas réel (voir proposition B.4.8, page 209), il suffit de vérifier que les boules fermées, qui engendrent \mathcal{A} , sont mesurables.

Proposition VI.1.19. La tribu \mathcal{A} des parties mesurables pour μ^* contient les boules fermées, donc les boréliens. La mesure μ introduite ci-dessus est ainsi définie sur la tribu des boréliens.

Démonstration. Soit $B = a + 2^k \mathbb{Z}_2$ une boule fermée. Il s'agit de montrer que pour toute partie $A \subset \mathbb{Z}_2$,

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

On considère un recouvrement de A par des boules fermées $A_n = a_n + 2^{k_n} \mathbb{Z}_2$, avec

$$\sum 2^{-k_n} \leq \mu^*(A) + \varepsilon.$$

Il s'agit de distribuer au mieux chacune de ces boules entre B et B^c . Là encore, la preuve est plus simple que dans le cas réel. Soit une telle boule A_n . Si elle ne rencontre pas B , alors elle est incluse dans B^c , auquel cas on l'affecte à B^c . Si elle rencontre B , alors soit elle est incluse dans B , auquel cas on "l'affecte à B " soit, du fait de l'ultramétricité, elle contient B . Son rayon est 2^{-k_n} , avec $k_n < k$. La boule A_n peut alors s'écrire, comme toute boule fermée, comme réunion disjointe de 2^{k-k_n} boules fermées de rayon 2^{-k_n} (l'une d'elles étant B). On atomise donc la boule B_n en 2^{k-k_n} boules plus petites, on affecte la boule qui s'identifie à B à B , et l'on affecte les $2^{k-k_n} - 1$ autres à B^c , ce qui se fait sans augmenter la masse totale. On construit ainsi à partir du recouvrement de A deux recouvrements de $A \cap B$ et $A \cap B^c$, respectivement, sans changer la masse totale, d'où l'on déduit que

$$\mu^*(A \cap B) + \mu^*(A \cap B^c) \leq \sum 2^{-k_n} \leq \mu^*(A) + \varepsilon$$

pour tout ε , d'où la propriété. \square

VI.2 Dendrogrammes

On décrit ici un procédé de construction d'un arbre à partir d'un espace métrique fini, qui induit une nouvelle métrique sur cet espace, de nature ultramétrique, et permet de visualiser d'une certaine manière la structure de l'espace. Il s'agit d'un procédé constructif et graphique, dont nous donnons ici une définition abstraite

Définition VI.2.1. (Dendrogramme)

On considère un espace métrique fini (X, d) , de cardinal N . On appelle dendrogramme de X une suite finie $(X^k)_{0 \leq k \leq N}$ de partitions de X :

$$X^k = \{X_1^k, \dots, X_{N-k}^k\}, \quad X_i^k \in \mathcal{P}(X), \quad X_i^k \cap X_j^k = \emptyset \quad \forall i \neq j, \quad X = \bigcup_{j=1}^{N-k} X_j^k,$$

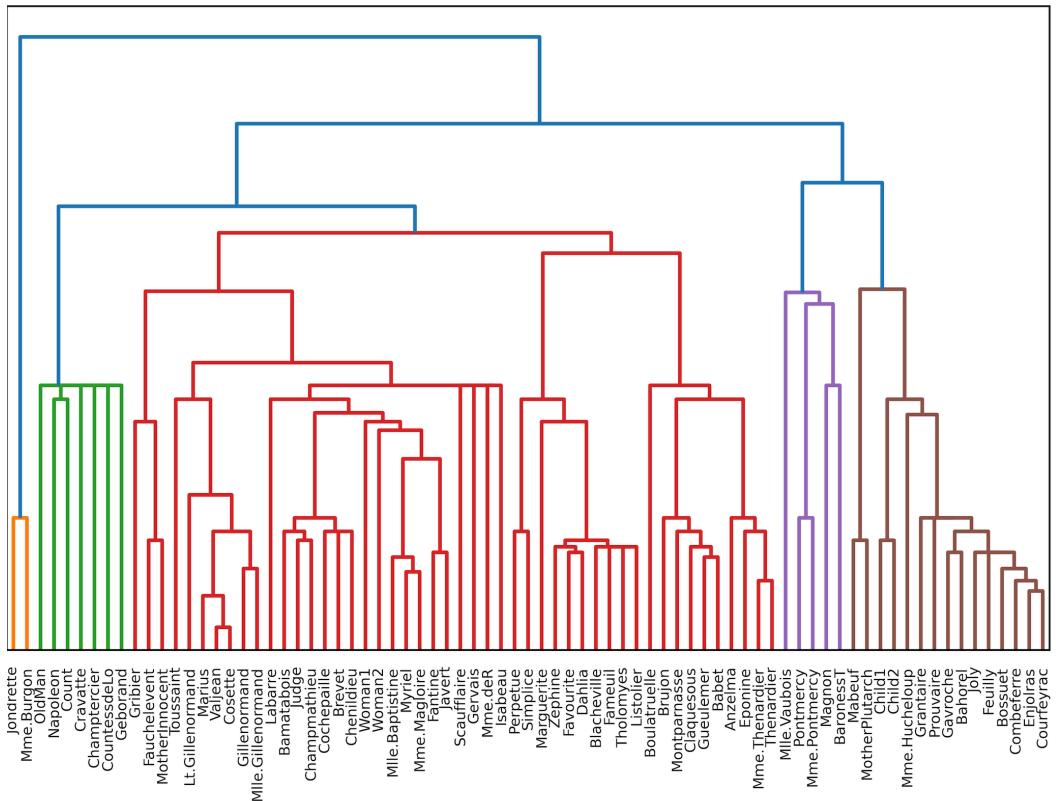


FIGURE VI.2.1 – Dendrogramme des Misérables

ainsi qu'une suite réelle $(D^k)_{1 \leq k \leq N}$, construits selon le processus d'agrégation suivant. La première partition X_0 est la plus fine, de cardinal N , et $D^0 = 0$. Supposons X^k connue, de cardinal $N - k$, avec $k \geq 0$. On construit la partition suivante en agrégeant 2 éléments de la partition, selon le principe suivant. On définit

$$D_{ij}^{k+1} = d(X_i^k, X_j^k) = \inf_{x \in X_i^k, y \in X_j^k} d(x, y),$$

et l'on choisit un couple (i, j) qui minimise cette quantité. On note D^{k+1} le minimiseur, et l'on définit la partition X^k en conservant les éléments autres que i et j , auxquels on adjoint la réunion de X_i^k et X_j^k . Le cardinal de la partition est donc $N - k - 1$. La suite D^k est croissante par construction, avec $D^1 > 0$, et X^{N-1} est la partition triviale ($X^{N-1} = \{X\}$).

Proposition VI.2.2. (Métrique ultramétrique associée à un dendrogramme)

On définit

$$\delta : (x, y) \in X \times X \mapsto \delta(x, y) \in \mathbb{R}_+$$

de la façon suivante : pour x et y dans X donné, on définit k_{xy} comme le plus petit entier tel que x et y sont dans une même sous-partie de X^k , et l'on fixe $\delta(x, y) = D^k$. L'application $\delta(\cdot, \cdot)$ ainsi définie est une distance, qui est ultra-métrique, c'est à dire qu'elle vérifie l'inégalité triangulaire renforcée

$$\delta(x, y) \leq \max(\delta(x, z), \delta(z, y)) \quad \forall x, y, z \in X$$

Démonstration. L'application δ prend bien des valeurs positives, on a $\delta(x, y) = 0$ si et seulement si $k_{xy} = 0$, si et seulement si $x = y$. Elle est symétrique par construction. Soient maintenant x, y, z , et les k_{xy} et k_{yz} associés. Pour $k \geq \max(k_{xy}, k_{yz})$, x et z sont dans la même composante, d'où

$$k_{xz} \leq \max(k_{xy}, k_{yz}),$$

d'où l'inégalité ultramétrique par croissance de la suite (D_k) . \square

À titre d'illustration, la figure VI.2.1 représente graphiquement le dendrogramme obtenu à partir de l'ensemble X des personnages des Misérables, muni d'une métrique qui prend en compte la proximité dans le roman des personnages (la distance est d'autant plus petite que les 2 personnages partagent un nombre important de scènes).

VI.3 Introduction au transport optimal

VI.3.1 Problème d'affectation et problème de Monge Kantorovich discret

Le problème d'affectation se formule comme suit :

Problème VI.3.1. On considère 2 ensembles de même cardinal $N \in \mathbb{N}$, tous deux identifiés à $\{1, \dots, N\}$, et l'on se donne une collection de coûts $c_{ij} \in \mathbb{R}$. Le problème consiste à trouver une bijection φ qui minimise la quantité

$$\sum_{i=1}^N c_{i\varphi(i)}.$$

Le problème ci-dessus ne présente pas d'intérêt théorique particulier : l'ensemble des bijections (groupe symétrique S_N) est fini, le problème admet bien (au moins) une solution. Mais la recherche effective de ce minimum peut extrêmement laborieuse, car le cardinal de l'ensemble des candidats croît comme $N!$.

Nous allons considérer une version relaxée du problème ci-dessus⁴, qui peut se formuler intuitivement de la façon suivante, dans un contexte de transport : on considère le premier ensemble comme contenant des positions dans un certain espace (il n'est pas nécessaire de préciser lequel ici), et le second ensemble aussi comme une collection de positions dans un espace (éventuellement le même, mais pas forcément). On note c_{ij} ce que cela coûte de transporter une quantité unitaire de matière de x_i vers y_j . Le problème précédent consistait à considérer que l'on avait une même quantité de matière en chaque point (par exemple $1/N$), et que l'on cherchait à transporter cette matière vers le second ensemble en envoyant toute la matière de chaque point vers une destination unique. Nous allons considérer maintenant qu'il est possible de distribuer la matière venant d'un point vers plusieurs destinations. Cette relaxation du problème permet de lever la contrainte d'avoir le même nombre de points au départ et à l'arrivée. Dans ce qui suit on notera γ_{ij} la quantité de matière allant de i vers j . On appellera $\gamma = (\gamma_{ij})$ un *plan de transport*.

Problème de Monge-Kantorovich discret

On considère 2 ensembles⁵ finis X et Y , de cardinaux respectifs N et $M \in \mathbb{N}$ et l'on se donne une collection de coûts $c_{ij} \in \mathbb{R}$. On se donne deux mesures de probabilités discrètes μ et ν sur X et Y , respectivement (μ_i est la masse portée par i , avec $\sum \mu_i = 1$, de même pour ν). On supposera tous les poids strictement positifs⁶. Le problème s'écrit

$$\min_{\Pi_{\mu\nu}} C(\gamma) \tag{VI.3.1}$$

avec

$$C(\gamma) = \sum_{i,j} c_{ij} \gamma_{ij}, \quad \Pi_{\mu,\nu} = \left\{ \gamma \in \mathbb{R}_+^{N \times M}, \quad \sum_j \gamma_{ij} = \mu_i \quad \forall i, \quad \sum_i \gamma_{ij} = \nu_j \quad \forall j \right\}$$

4. Cette approche a été proposée par L.V. Kantorovich en 1942. On trouvera une traduction du papier original sur <http://www.math.toronto.edu/mccann/assignments/477/Kantorovich42.pdf>

5. Il n'y a pas lieu de préciser ici les points d'arrivée et points de départ. Nous nous intéresserons plus loin au transport entre points d'un espace euclidien, mais ici on peut tout aussi bien concevoir le transport d'une essoreuse vers le *concept de néant* chez Sartre.

6. On peut toujours se ramener à cette situation en supprimant de X et Y les points non chargés.

Remarque VI.3.1. On peut formuler ce problème en termes probabilistes, en considérant γ comme une loi de probabilité sur l'espace produit $X \times Y$, dont les mesures images par les projections sur X et Y sont respectivement μ et ν . Parmi de telles lois, on cherche celle(s) qui minimise(nt) l'espérance de la “fonction” $c = (c_{ij})$ sur $X \times Y$.

Remarque VI.3.2. L'ensemble admissible est non vide, il contient en particulier le plan correspondant à une loi de probabilité sur $X \times Y$ pour deux variables indépendantes, qui s'écrit

$$\gamma_{ij} = \mu_i \nu_j.$$

Proposition VI.3.3. Le problème VI.3.1 admet un minimiseur.

Démonstration. Les γ_{ij} sont positifs, et chacun d'eux est majoré par le max des μ_i , l'ensemble Π est donc borné, il est évidemment fermé donc compact : la fonction continue (car linéaire) $C(\cdot)$ admet donc un minimiseur sur Π . \square

Remarque VI.3.4. Dans le cas d'un coût du type $c_{ij} = a_i + b_j$, le problème est fortement dégénéré, puisque tout transport de μ vers ν réalise le même coût. Par ailleurs, pour deux ensembles de même cardinal N , avec μ et ν lois uniformes sur X et Y , si l'on se donne une bijection φ de S_n , on peut construire une famille de coûts telle que le plan associé à la bijection⁷ soit l'unique minimiseur, en prenant par exemple $c_{i\varphi(i)} = -1$, et $c_{ij} = 0$ si $j \neq \varphi(i)$.

Question VI.3.1. Étant donnée une collection de coût (c_{ij}) , existe-t-il des ensembles X et Y de points de \mathbb{R}^d tels que $c_{ij} = |y_j - x_i|$? (on pourra aussi considérer $c_{ij} = |y_j - x_i|^p$, $c_{ij} = \psi(|y_j - x_i|)$ avec ψ croissante et nulle en 0.)

Question VI.3.2. Le problème VI.3.1 admet-il une solution unique “en général”? (on s'attachera à exprimer précisément ce que l'on entend par unicité générique.)

Lien avec le problème d'affectation

Dans le cas où les cardinaux sont les mêmes, et les mesures équidistribuées, on peut préciser le lien entre le modèle relaxé basé sur les plans de transports et le problème d'affectation. Pour simplifier les notations, on considère ici la situation où chaque point porte une masse unitaire, de telle sorte que la masse totale des mesures considérées est égale au nombre de points. Il ne s'agit donc plus de mesure de probabilité, mais on peut s'y ramener en divisant la mesure par le nombre de points.

Proposition VI.3.5. On se place dans le cas $N = M$ (même nombre de points de part et d'autre, et $\mu_i = \nu_j \equiv 1$), et l'on note Π_S l'ensemble des plans de transport associés à une affectation, i.e. $\gamma_{ij} = \delta_{i\varphi(i)}$, où φ est une permutation du groupe symétrique. L'ensemble des points extrémaux⁸ de Π s'identifie à Π_S .

Démonstration. Tout point de Π_S est de façon évidente extrémal pour Π . Réciproquement, considérons un plan générique (i.e. qui n'est pas associé à une bijection) γ . On considère dans un premier temps les indices i pour lesquels γ_{ij} est nul pour tous les indices j sauf un (qui vaut donc 1). Cette sous-famille des points de départ est en bijection avec les points d'arrivées j correspondants, pour lesquels, symétriquement, γ_{ij} est nul pour tous les i sauf 1. On note I (resp. J) l'ensemble des indices non concernés dans l'espace de départ (resp. d'arrivée). Les ensemble I et J sont de même cardinal, et non vides par hypothèse. La restriction du plan γ à $X_I \times Y_J$ est diffuse, au sens que pour tout $i \in I$, $\gamma_{ij} \in]0, 1[$ pour au moins 2 indices $j \in J$, et pour tout $j \in J$, on a $\gamma_{ij} \in]0, 1[$ pour au moins 2 indices $i \in I$. On part d'un indice $i_0 \in I$, et l'on choisit j_0 tel que $\gamma_{i_0 j_0} > 0$. On choisit ensuite $i_1 \neq i_0$ tel que $\gamma_{i_1 j_0} > 0$, puis $j_1 \neq j_0$ tel que $\gamma_{i_1 j_1} > 0$. On construit ainsi une

7. C'est à dire : $\gamma_{i\varphi(i)} = 1/N$, et $\gamma_{ij} = 0$ si $j \neq \varphi(i)$.

8. On dit que $\gamma \in \Pi \subset \mathbb{R}^{d^2}$ est point extrémal de Π si $\gamma = (\gamma^1 + \gamma^2)/2$, avec $\gamma^1, \gamma^2 \in \Pi$, implique $\gamma^1 = \gamma^2 = \gamma$.

suite d'indices

$$i_0, j_0, i_1, \dots, i_{n-1}, i_n,$$

que l'on peut voir comme un chemin dans le graphe sur $I \cup J$ associé au plan γ , chemin qui ne contient pas d'aller-retour. L'ensemble des indices étant fini, il existe forcément un n tel que i_n correspond à un indice $i_\ell \neq i_{n-1}$ déjà visité. On considère alors la variation

$$h = \sum_{k=\ell}^{n-1} (\pi_{i_k, j_k} - \pi_{i_{k+1}, j_k}),$$

avec $i_n = i_\ell$, et où $\pi_{i,j}$ est l'élément de \mathbb{R}^{NM} qui vaut 1 sur la composante (i, j) , et qui est nul pour les autres couples. Pour η suffisamment petit, $\gamma \pm \eta h$ est positif, et par construction $\gamma \pm \eta h$ vérifie les contraintes de marginales, les deux perturbations sont donc dans $\Pi_{\mu, \nu}$, et γ est moyenne non triviale de ces deux plans de transport, il ne s'agit donc pas d'un point extrémal.

Les seuls points extrêmaux correspondent donc aux permutations. \square

Corollaire VI.3.6. L'ensemble Π des plans de transport admissibles est l'enveloppe convexe de Π_S .

Démonstration. Il s'agit d'une conséquence du théorème de Krein-Milman en dimension finie, qui assure que tout convexe compact d'un espace affine de dimension finie est l'enveloppe convexe de ses points extrêmaux. \square

Proposition VI.3.7. On se place comme précédemment dans la situation de mesures équidistribuées sur des ensembles de même cardinal. Le problème de Monge Kantorovich discret VI.3.1 admet au moins une solution dans Π_S , i.e. une solution optimale du type permutation.

Démonstration. D'après la proposition VI.3.3, le problème VI.3.1 admet un minimiseur γ . D'après la proposition VI.3.5, ce minimiseur s'écrit comme combinaison convexe de plans associés à des permutations $\varphi_1, \dots, \varphi_K$:

$$\gamma = \sum \theta_k \gamma^k$$

(on ne garde dans la somme ci-dessus que les termes non triviaux, de telle sorte que $\theta_k > 0$ pour tout k). Le coût étant linéaire, on a

$$C(\gamma) = \sum \theta_k C(\gamma^k).$$

Comme chaque $C(\gamma^k)$ est supérieur ou égal à $C(\gamma)$, et que $\sum \theta_k = 1$ avec $\theta_k > 0$ pour tout k , la combinaison convexe ci-dessus implique que $C(\gamma^k)$ est égal à $C(\gamma)$ pour tout k . Chaque permutation impliquée dans la combinaison réalise donc le minimum. \square

VI.3.2 Transport optimal, cas général

On considère ici deux espaces mesurables (X, \mathcal{A}) et (X', \mathcal{A}') . On se donne deux mesures de probabilité μ et μ' sur ces espaces respectifs. On introduit l'ensemble $\Lambda_{\mu, \mu'}$ des applications T mesurables qui “envoient” μ sur μ' , c'est à dire telles que $T_\sharp \mu = \mu'$. On définit une fonction de *coût* mesurable pour la tribu produit $\mathcal{A} \otimes \mathcal{A}'$

$$c : X \times X' \longrightarrow [0, +\infty]$$

La fonction $x \mapsto c(x, T(x))$ est \mathcal{A} mesurable. En effet, pour tout $b \in \mathbb{R}$, l'ensemble

$$B = \{(x, x') \in X \times X' \mid c(x, x') \leq b\} \subset X \times X',$$

est dans $\mathcal{A} \otimes \mathcal{A}'$ par mesurabilité de c . L'image réciproque de $] -\infty, b]$ par $x \mapsto c(x, T(x))$ est donc l'image réciproque de B par l'application $x \mapsto (x, T(x))$, il suffit donc de vérifier que cette dernière application est mesurable. Il suffit pour cela de vérifier que l'image réciproque de tout rectangle $A \times A'$, avec $A \in \mathcal{A}$ et $A' \in \mathcal{A}'$, est dans \mathcal{A} . Or cette image réciproque est

$$A \cap T^{-1}(A'),$$

qui est mesurable par mesurabilité de T .

On définit maintenant le coût de transport associé à T comme

$$C(T) = \int_X c(x, T(x)) d\mu(x),$$

qui est bien défini d'après ce qui précède. Le problème abstrait de Monge consiste à minimiser $C(T)$ parmi les transports admissibles, il s'écrit

$$\min_{T \in \Lambda_{\mu, \mu'}} C(T).$$

VI.4 Distance de Gromov-Wasserstein

Définition VI.4.1. Soient (X, d, μ) et (X', d', μ') deux espaces métriques finis probabilisés (i.e. munis d'une mesure définie sur la tribu discrète, de masse totale 1). On dit que ces espaces sont isomorphes s'il existe une bijection de X vers X' qui préserve les structures de distance et de mesure, i.e. s'il existe une isométrie T telle que

$$T_\sharp \mu = \mu' \quad \text{i.e.} \quad \mu'(A') = \mu(T^{-1}(A')) \quad \forall A \in \mathcal{P}(X').$$

Soient (X, d, μ) et (X', d', μ') deux espaces métriques probabilisés et finis, de cardinaux respectifs N et N' , munis de leurs tribus discrètes respectives. On suppose que μ et μ' sont toutes deux de masse 1 (mesures de probabilité). On note Π l'ensemble des plans de transport entre μ et μ' :

$$\Pi = \left\{ \gamma = (\gamma_{xx'}) \in \mathbb{R}_+^{N \times N'}, \sum_x \gamma_{xx'} = \mu'_{x'}, \sum_{x'} \gamma_{xx'} = \mu_x \right\}.$$

Définition VI.4.2. Pour $p \in [1, +\infty[$, on définit

$$d_{GWp}(X, X') = \inf_{\gamma \in \Pi} \left(\sum_{xx'} \sum_{yy'} |d(x, y) - d(x', y')|^p \gamma_{xx'} \gamma_{yy'} \right)^{1/p}.$$

Lemme VI.4.3. L'infimum de la définition précédente est atteint.

Démonstration. L'ensemble admissible Π est compact, et l'application

$$\gamma \mapsto \left(\sum_{xx'} \sum_{yy'} |d(x, y) - d(x', y')|^p \gamma_{xx'} \gamma_{yy'} \right)^{1/p}$$

est continue. \square

Proposition VI.4.4. La quantité définie ci-dessus est une distance sur l'ensemble des espaces métriques probabilisés finis (quotienté par les isomorphismes au sens de la définition VI.4.1)

Démonstration. Si l'on a $d_{GWp}(X, X') = 0$ alors, pour tous x, x', y, y' tels que $\gamma_{xx'} \neq 0$ et $\gamma_{yy'} \neq 0$, on a $d(x, y) = d(x', y')$. Soit maintenant x, x', y', y , tels que x envoie de la masse à x' et y' . On a $d(x', y') = d(x, x') = 0$, d'où $x' = y'$. Pour chaque x il existe donc unique x' tel que $\gamma_{xx'} \neq 0$, et l'on a donc $\gamma_{xx'} = \mu_x$. On peut mener le même raisonnement dans l'autre sens : pour chaque x' il existe un unique x tel que $\gamma_{xx'} \neq 0$, et l'on a donc $\gamma_{xx'} = \mu_{x'}$. Le plan de transport γ correspond donc à une bijection T entre X et X' , et l'on a

$$0 = d_{GWp}(X, X') = \left(\sum_{xx'} \sum_{yy'} |d(x, y) - d(x', y')|^p \gamma_{xx'} \gamma_{yy'} \right)^{1/p}$$

$$= \left(\sum_x \sum_y |d(x, y) - d(T(x), T(y))|^p \right)^{1/p}$$

La symétrie est immédiate d'après la définition.

Pour l'inégalité triangulaire, on considère 3 espaces métriques probabilisés X , X' , et X'' , et des plans de transport $(\gamma_{xx'})$ et $(\gamma_{x'x''})$ qui réalisent les distance entre X et X' et entre X' et X'' , respectivement. On construit à partir de ces plans un plan entre X et X'' (non nécessairement optimal, mais qui suffira pour l'inégalité triangulaire) en “collant” dans un premier temps les plans, puis en condensant l'espace X' intermédiaire. Plus précisément, on introduit

$$\gamma_{xx'x''} = \frac{\gamma_{xx'} \gamma_{x'x''}}{\mu_{x'}},$$

et l'on définit

$$\gamma_{xx''} = \sum_{x' \in X'} \gamma_{xx'x''}.$$

On a

$$\sum_{xx''} \sum_{yy''} |d(x, y) - d(x'', y'')| \gamma_{xx''} \gamma_{yy''} = \sum_{xx''} \sum_{yy''} |d(x, y) - d(x'', y'')| \sum_{x'} \frac{\gamma_{xx'} \gamma_{x'x''}}{\mu_{x'}} \sum_{y'} \frac{\gamma_{yy'} \gamma_{y'y''}}{\mu_{y'}}.$$

On écrit $|d(x, y) - d(x'', y'')| \leq |d(x, y) - d(x', y')| + |d(x', y') - d(x'', y'')|$, ce qui permet de majorer la quantité de départ par deux termes. le premier s'écrit

$$\begin{aligned} & \sum_{xx''} \sum_{yy''} |d(x, y) - d(x', y')| \sum_{x'} \frac{\gamma_{xx'} \gamma_{x'x''}}{\mu_{x'}} \sum_{y'} \frac{\gamma_{yy'} \gamma_{y'y''}}{\mu_{y'}} \\ &= \sum_x \sum_y |d(x, y) - d(x', y')| \sum_{x'} \frac{\gamma_{xx'}}{\mu_{x'}} \underbrace{\sum_{x''} \gamma_{x'x''}}_{\mu_{x'}} \sum_{y'} \frac{\gamma_{yy'}}{\mu_{y'}} \underbrace{\sum_{y''} \gamma_{y'y''}}_{\mu_{y'}} \\ &= \sum_x \sum_y |d(x, y) - d(x', y')| \sum_{x'} \gamma_{xx'} \gamma_{yy'} = d_{GW}(X, X'). \end{aligned}$$

Le second terme s'identifie de la même manière à $d_{GW}(X'X'')$

□

VI.5 Propagation d'opinion et flot de gradient

On s'intéresse ici à un modèle simple de propagation d'opinion sur un réseau. Les noeuds de ce réseaux sont des personnes, ou “agents”, auxquels on affecte un nombre représentant une opinion sur un certain sujet à un instant donné. Ce nombre peut par exemple représenter la tendance qu'a un individu à voter pour tel ou tel candidat au second tour d'une élection présidentielle, ou l'idée que l'on peut se faire de la probabilité de gain d'une équipe nationale à une finale de coupe du monde. Dans un autre contexte, on pourra penser à la valeur d'une quantité qui fait l'objet d'un débat public, comme l'augmentation de la température moyenne sur la planète dans 20 ans.

On considère un ensemble V de N individus, on note u_x^k l'opinion de l'individu x à l'instant k , et par $u^k = (u_x^k)_{x \in V}$ la collection des opinions. L'influence de $y \in V$ sur x est quantifiée par un coefficient $K_{xy} \in [0, 1]$, et l'on suppose :

$$\sum_{y \in V} K_{xy} = 1 \quad \forall x \in V.$$

La collection de l'ensemble des coefficient est donc encodée par une matrice (sans choix de numérotation des sommets) $(K_{xy})_{(x,y) \in V^2}$ stochastique.

On notera que toute l'information sur le graphe est dans la collection des influences : $E \subset V \times V$ est défini par

$$E = \text{supp}(K_{xy}) = \{(x, y) \in V \times V, K_{xy} > 0\}.$$

On conservera néanmoins la notation (redondante) (V, E, K) pour désigner le réseau.

On écrira $x \rightarrow y$ si $K_{xy} > 0$, qui signifie que x écoute y , ou x suit y , ou plus généralement x est influencé par y . Avec cette convention, l'opinion / influence remonte le sens des flèches.

Le coefficient diagonal K_{xx} peut être non nul (présence de boucles dans le réseau)), ce qui correspond à une certaine inertie de x , ou résistance de x à modifier son opinion sous l'effet d'influences extérieures, jusqu'à éventuellement ne plus se préoccuper de l'opinion des autres (cas extrême $K_{xx} = 1$) On note $\Gamma \subset V$ l'ensemble des sommets qui ne pointent que vers eux-mêmes

$$\Gamma = \{x \in V, K_{xx} = 1\}, \quad (\text{VI.5.1})$$

et par $\mathring{V} = V \setminus \Gamma$ l'ensemble des sommets intérieurs. La frontière Γ correspond aux individus qui n'écoutent qu'eux mêmes, et étant éventuellement suivis par d'autres. Si l'on considère que ces agents affichent une opinion dans le dessein de modifier l'opinion d'autres agents du réseau, on peut voir ces individus comme des *influenceurs*⁹, ou plus simplement, s'ils ne nourrissent aucun dessein particulier, de personnes non influençables, ou *têtues*.

Modèle discret d'évolution

On se donne une collection d'opinions initiales $(u_x^0)_{x \in V}$, et l'on considère que l'opinion évolue d'un jour à l'autre selon la relation

$$u_x^{k+1} = \sum_{x \rightarrow y} K_{xy} u_y^k \quad \forall x \in V. \quad (\text{VI.5.2})$$

qui peut s'écrire aussi matriciellement. On notera que l'indication ' $x \rightarrow y$ ' n'est pas à strictement parler obligatoire, du fait que $K_{xy} = 0$ dès que $(x, y) \notin E$. Elle sera parfois omise, et l'on écrira alors simplement $\sum K_{xy} u_y^k$.

Le fait que cela prenne un certain temps pour x d'absorber l'influence de ses voisins peut être modélisé en introduisant un paramètre d'inertie $\theta \in [0, 1]$, et en écrivant le modèle relaxé

$$u_x^{k+1} = (1 - \theta)u_x^k + \theta \sum_{x \rightarrow y} K_{xy} u_y^k. \quad (\text{VI.5.3})$$

Pour $\theta = 1$, on retrouve le problème discret. Noter que ce nouveau problème rentre dans le cadre discret précédent, en introduisant les paramètres modifiés

$$K'_{xx} = (1 - \theta) + \theta K_{xx}, \quad K'_{xy} = \theta K_{xy} \quad \text{for } y \neq x.$$

Équation différentielle

On considère que θ s'écrit ε/η , où η est un temps de relaxation fixe (temps typique de propagation de l'influence), et ε un petit paramètre (également homogène à un temps). On a

$$\frac{u_x^{k+1} - u_x^k}{\varepsilon} = \frac{1}{\eta} \left(\sum_{x \rightarrow y} K_{xy} (u_y^k - u_x^k) \right).$$

L'évolution prend la forme de la discréttisation en temps d'une système d'équations différentielles ordinaires pour des quantités $t \mapsto u_x^t \in \mathbb{R}$ qui varient continûment en temps, et vérifient le système d'équations

$$\frac{d}{dt} u_x^t = \frac{1}{\eta} \left(\sum_{x \rightarrow y} K_{xy} (u_y^t - u_x^t) \right).$$

9. Ces influenceurs peuvent aussi être vus comme des agents influencés par une entité extérieure (entreprise, groupe de pression, ...) qui les contrôle.

Le problème continu en temps s'écrit donc

$$\frac{d}{dt} u^t = -\frac{1}{\eta} A u^t \quad \text{avec} \quad A = I - K. \quad (\text{VI.5.4})$$

On se propose de caractériser ici les cas où le problème d'évolution (VI.5.4) a une structure de flot de gradient pour un certain produit scalaire.

Le cas où l'on se restreint au produit scalaire euclidien canonique est immédiat :

Proposition VI.5.1. Le problème VI.5.4 a une structure de flot de gradient pour le produit scalaire canonique, i.e. il existe une fonctionnelle Ψ deux fois continûment différentiable telle que $Au = \nabla\Psi(u)$, si et seulement si A est symétrique.

Démonstration. Si A est symétrique on a $Au = \nabla\Psi(u)$ avec

$$\Psi(u) = \frac{1}{2}\langle u | u \rangle.$$

Si $Au = \nabla\Psi(u)$, alors $a_{ij} = \partial^2\Psi/\partial x_i \partial x_j = \partial^2\Psi/\partial x_j \partial x_i = a_{ji}$. \square

Plus généralement, si l'on considère une matrice M symétrique définie positive, on note $\langle \cdot | \cdot \rangle_M$ le produit scalaire associé, i.e.

$$\langle u | v \rangle_M = \langle Mu | v \rangle.$$

Pour toute fonctionnelle $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}$ continûment différentiable, on note $\nabla_M \Phi(u) \in \mathbb{R}^N$ son gradient en u selon le produit scalaire associé à M , c'est à dire le vecteur tel que

$$\Phi(u + h) = \Phi(u) + \langle \nabla_M \Phi | h \rangle_M + o(h) = \Phi(u) + \langle M \nabla_M \Phi | h \rangle + o(h).$$

On a donc $\nabla_M \Phi = M^{-1} \nabla \Phi$, ce qui permet d'énoncer une première caractérisation des flots de gradient.

Proposition VI.5.2. Le problème VI.5.4 a une structure de flot de gradient pour le produit scalaire associé à la matrice s.d.p. M , i.e. il existe une fonctionnelle Ψ deux fois continûment différentiable telle que $Au = \nabla_M \Psi(u)$, si et seulement A s'écrit $M^{-1}B$, où B est une matrice symétrique.

Démonstration. Si $A = M^{-1}B$, alors $Au = \nabla_M \Psi(u)$ avec $\Psi(u) = \frac{1}{2}\langle Bu | u \rangle$. Si $Au = \nabla_M \Psi(u)$ alors, comme précédemment, MA est nécessairement symétrique. \square

Dans la suite, nous aborderons un cas particulier de systèmes présentant une structure de gradient, il s'agit que réseaux que nous appellerons *charismatiques* (voir la définition VI.5.4). Pour de tels réseaux, on aura $K = M^{-1}B$, où M est une matrice *diagonale*. Les coefficients diagonaux de M correspondent aux charismes $(m_x)_{x \in V}$ des agents. D'un point de vue probabiliste, cette situation correspond au cas d'une chaîne de Markov *réversible*. Nous proposons maintenant une caractérisation plus exploitable des matrices A pour lesquelles l'équation (VI.5.4) a une structure de flot gradient (pour matrice s.d.p. M quelconque).

Proposition VI.5.3. Le problème (VI.5.4) a une structure de flot de gradient pour un certain produit scalaire si et seulement si A (ou, de façon équivalente, K) est diagonalisable, et ses valeurs propres sont réelles.

Démonstration. Si Au est le gradient d'un fonctionnelle quadratique en u pour un produit scalaire $\langle \cdot | \cdot \rangle_M$, il existe une matrice symétrique B telle que $A = M^{-1}B$. Comme M est s.d.p., elle s'écrit $M = UDU^{-1}$, où U est une matrice orthogonale et D est diagonale. On définit alors $M^{1/2}$ comme $UD^{1/2}U^{-1}$. La matrice A est semblable à

$$M^{1/2}AM^{-1/2} = M^{1/2}M^{-1}BM^{-1/2} = M^{-1/2}BM^{-1/2},$$

qui est symétrique car $M^{-1/2}$ et B le sont. La matrice A est donc semblable à une matrice symétrique, elle est donc diagonalisable de valeurs propres réelles.

On suppose maintenant que A est diagonalisable, de valeurs propres réelles : $A = PDP^{-1}$ où D est diagonale réelle, et P est une matrice inversible. On écrit

$$A = PDP^{-1} = PP^TP^{-T}DP^{-1} = M^{-1}B,$$

où $M = P^{-T}P^{-1}$ est une matrice symétrique définie positive¹⁰, et $B = P^{-T}DP^{-1}$ est symétrique réelle. \square

Réseaux charismatiques

Nous nous intéressons ici à des réseaux qui encodent des interactions d'une nature symétrique, qui correspondent comme nous verrons à la situation d'un modèle d'évolution de type flot de gradient, pour une matrice M diagonale. Plus précisément, les réseaux que nous considérons ici sont basés sur l'existence d'un paramètre afférent à chaque individus, un poids que nous appellerons *charisme* dans ce contexte, qui conditionne l'influence qu'il exerce sur les autres. Comme nous allons le voir, cette hypothèse rapprochera les réseaux qui la vérifie des réseaux résistifs ou, dans un contexte stochastique, des chaînes de Markov *réversibles*.

Définition VI.5.4. (Réseaux charismatiques)

On dit que le réseau (VE, K) est *charismatique* s'il existe un champ $m = (m_x) \in]0, +\infty[^V$ tel que, pour tous $x, y \in V$,

$$m_x K_{xy} = m_y K_{yx}. \quad (\text{VI.5.5})$$

Remarque VI.5.5. On vérifie immédiatement que les réseaux charismatiques sont un cas particulier de flots de gradient. En effet, si l'on introduit la matrice $C = (C_{xy})$, avec $C_{xy} = m_x K_{xy}$, la matrice C est symétrique, et l'on a, avec $M = \text{diag}(m_x)$,

$$C = MK \implies A = I - K = I - M^{-1}C = M^{-1}(M - C) = M^{-1}B,$$

où M est s.d.p. et $B = M - C$ est symétrique, on est donc bien dans le cas d'un flot de gradient pour la métrique induite par M , dans le cas d'une matrice M diagonale. On a de plus

$$m_x K_{xy} = C_{xy} \implies \sum_{y \sim x} m_x K_{xy} = m_x = \sum_{y \sim x} C_{xy}.$$

Remarque VI.5.6. Si nous considérons K comme la matrice de transition d'une chaîne de Markov sur l'ensemble V , où K_{xy} est la probabilité de transition de x à y , alors la définition VI.5.4 correspond à celle d'une chaîne *reversible*, et $m = (m_x)$ joue le rôle d'une mesure invariante.

Remarque VI.5.7. Si le réseau est charismatique, $K_{xy} > 0$ si et seulement si $K_{yx} > 0$. En conséquence $x \rightarrow y$ si et seulement si $y \rightarrow x$. On considérera néanmoins le graphe sous jacent comme orienté, étant entendu que $(x, y) \in E \iff (y, x) \in E$, mais l'on ne fait pas l'identification entre les deux arêtes. Par ailleurs, tout influenceur $x \in \Gamma$ est isolé, et ne joue donc aucun rôle dans la dynamique d'opinion. Si l'on se restreint à des réseaux connexes, il ne peut donc pas y avoir d'influenceurs.

Remarque VI.5.8. L'identité (VI.5.5) peut s'écrire

$$K_{xy} = \frac{m_y}{m_x} K_{yx}.$$

L'influence que y exerce sur x dépend donc du rapport des charismes de x et de y , et de l'influence que x exerce sur y . Plus le charisme de y est grand comparé à celui de x , plus y influence x comparé à l'influence de x sur y . C'est ce qui justifie l'appellation *charisma* : plus le charisme est grand, plus l'influence exercée sur les autres est grande.

10. Ce produit scalaire fait de la famille des vecteurs colonnes de P une base orthonormée.

Proposition VI.5.9. Soit (V, E, K) un réseau charismatique. Si le réseau est connexe, alors le charisme est défini de façon unique à constante multiplicative près. En particulier il admet un unique charisme $m = (m_x)$ qui est une loi de probabilité sur V , i.e. tel que

$$\sum_V m_x = 1.$$

Démonstration. Notons en premier lieu que, d'après la remarque VI.5.7, la connexité entraîne la forte connexité. Considérons m et m' deux champs de charisme sur V . Soit $x \in V$ arbitraire. On pose $\lambda = m'_x/m_x > 0$. Pour tout $y \sim x$, on a

$$m'_y = m'_x \frac{K_{xy}}{K_{yx}} = m'_x \frac{m_y}{m_x} = \lambda m_y.$$

Cette relation de proportionnalité se propage de proche en proche, donc en tous les sommets par connexité du graphe, on a donc $m' = \lambda m$. Il existe donc en particulier un unique champ de charisme de masse totale unitaire. \square

Remarque VI.5.10. A cardinal de V fixé, on peut établir une relation de bijection entre les réseaux charismatiques et l'ensemble des matrices symétriques à coefficient positifs ou nul, à constante positive multiplicative près. En effet, si K et m satisfont les relations (VI.5.5) alors la matrice C définie par $C_{xy} = m_x K_{xy}$, est symétrique. Si l'on note M la matrice $\text{diag}(m_x)$, on peut écrire $K = M^{-1}C$. Réciproquement, si C est une matrice symétrique dont les éléments sont positifs, et que l'on souhaite lui associer une matrice $K = M^{-1}C$ encodant les influences d'un réseau charismatique, le seul choix possible est, étant donnée la contrainte de normalisation des lignes de K ,

$$m_x = \sum_y C_{xy}.$$

Si l'on prend pour M la matrice diagonale de coefficients $(m_x)_{x \in V}$, alors $K = M^{-1}C$ correspond à un réseau charismatique.

Les réseaux charismatiques présentent une propriété de conservation particulière. Nous avons noté (voir remarque ??) que l'opinion totale n'est en général pas conservée. Dans le cas des réseaux charismatiques, une certaine quantité est pourtant conservée, il s'agit d'une certaine moyenne de l'opinion, plus précisément de l'espérance de l'opinion relativement à la mesure (m_x) .

Proposition VI.5.11. (Propriété de conservation)

Soit (V, E, K, m) un réseau charismatique, et (u^k) la suite des opinions associées au modèle (VI.5.2). L'opinion moyenne relativement à la mesure m , définie par

$$\bar{u}^k = \sum_{x \in V} m_x u_x^k,$$

se conserve au cours des itérations.

Démonstration. On a

$$\begin{aligned} \bar{u}^{k+1} &= \sum_{x \in V} m_x u_x^{k+1} = \sum_{x \in V} \sum_{y \leftarrow x} m_x K_{xy} u_y^k = \sum_{x \in V} \sum_{y \leftarrow x} m_y K_{yx} u_y^k \\ &= \sum_{y \in V} m_y u_y^k \sum_{x \leftarrow y} K_{yx} = \sum_{y \in V} m_y u_y^k = \bar{u}^k, \end{aligned} \tag{VI.5.6}$$

qui établit la propriété de conservation annoncée. \square

Cette propriété permet de caractériser les limites possibles du problème d'évolution.

Proposition VI.5.12. Soit (V, E, K, m) un réseau charismatique. On suppose que la suite des itérés du modèle discret converge vers un consensus associé à la valeur u^∞ . Alors cette valeur u^∞ correspond à la moyenne des opinions initiales relativement au charisme m normalisé à 1 selon la proposition VI.5.9 :

$$u^\infty = \sum_{x \in V} m_x u_x^0.$$

Démonstration. Si toutes les opinions u_x^k convergent vers u^∞ , on a, d'après la proposition VI.5.11,

$$\sum_{x \in V} m_x u_x^0 = \sum_{x \in V} m_x u_x^k \rightarrow \sum_{x \in V} m_x u^\infty = u^\infty.$$

qui montre que l'opinion limite commune est bien la combinaison barycentrique des opinions initiales, pondérée par les charismes des agents. \square

Point de vue variationnel, flot de gradient & réseaux résistifs

Une autre particularité du cadre charismatique est que le problème présente une structure variationnelle. Considérons un réseau charismatique (V, E, K) , de charisme normalisé m . Comme décrit dans la section ?? (voir aussi la remarque VI.5.5), nous sommes dans la situation où A s'écrit $M^{-1}B$, avec $M = \text{diag}(m_x)$, et $B = M - C$. Le problème d'évolution continu en temps est donc, d'après la proposition VI.5.2, un flot de gradient pour la fonctionnelle

$$\Psi(u) = \frac{1}{2} \langle (M - C)u | u \rangle = \frac{1}{2} \sum_x u_x \sum_{y \sim x} C_{xy} (u_x - u_y) = \frac{1}{2} \sum_{e \in E} C_{xy} (u_x - u_y)^2, \quad (\text{VI.5.7})$$

pour le produit scalaire défini par

$$\langle u | v \rangle_M = \sum_{x \in V} m_x u_x v_x.$$

En conséquence, l'énergie Ψ décroît au cours du temps, et le modèle exprime un principe d'évolution selon la ligne de plus grande pente vis à vis de Ψ , pour la métrique définie par M .

Cette énergie permet de faire un lien avec les réseaux résistifs. On peut penser u_x comme un potentiel en x , la quantité $C_{xy} = m_x K_{xy}$ (qui est symétrique en x, y) jouant le rôle d'une *conductance* (inverse d'une résistance) de l'arête (symétrique selon ce point de vue) joignant x et y . La quantité $\Phi(u)$ correspond dans cette analogie électrique à (la moitié de) l'énergie dissipée au sein du réseau aux conductances $m_x K_{xy}$ et aux potentiels u_x . L'évolution tend donc à minimiser cette énergie dissipée, que l'on peut voir comme une estimation de l'écart à l'équilibre en termes d'opinion. Dans cette optique, les paramètres C_{xy} peuvent être interprétés comme des *coefficients de friction*, et les u_x comme des vitesses¹¹.

En poursuivant cette analogie avec les systèmes mécaniques, concevoir l'opinion d'un sommet x comme une quantité scalaire de type vitesse, la quantité obtenue par multiplication par la la “masse” m_x , donne une quantité de mouvement-opinion. On a bien un principe de Newton pour ce système mécanique : d'après la proposition VI.5.11, la quantité de mouvement-opinion globale pour ce système libre (non forcé de l'extérieur) se conserve. Le carré de la norme naturellement associée au modèle correspond à une énergie cinétique. On prendra garde en revanche que l'énergie globale Φ dont dérive l'équation d'évolution, quadratique en les vitesses, n'a rien d'une énergie cinétique, il s'agit plutôt comme indiqué ci-dessus d'une somme de termes de nature frictionnelle, qui quantifieraient des puissances dissipées au niveau de chaque arête (relation entre deux individus), d'autant plus que les opinion divergent. Cette interprétation est étayée par un pseudo-bilan énergétique que l'on peut obtenir à partir de l'équation de conservation de la quantité de mouvement, on effectue le produit scalaire de

$$M \frac{du}{dt} = -\nabla \Psi(u)$$

11. Comme deux objets en contact, allant à des vitesses différentes, sont soumis à une force d'interaction de type friction proportionnelle à leur vitesse relative, dissipant ainsi une énergie proportionnelle au carré de cette vitesse relative.

avec la “vitesse” u , pour obtenir

$$\frac{d}{dt} \frac{1}{2} \langle Mu | u \rangle = -\langle \nabla \Psi(u) | u \rangle = \sum_{e \in E} C_{xy} |u_x - u_y|^2,$$

qui peut se lire : la dérivée en temps de l’énergie cinétique est égale à la puissance dissipée par friction entre opinions différentes. S’il s’agissait d’un système mécanique standard, cette énergie serait dissipée sous forme de chaleur au sein du système ou vers le monde extérieur (l’ajout d’un modèle thermique permettrait de préciser le devenir de cette énergie thermique).

VI.6 Modèles macroscopiques de trafic routier

On considère l’évolution d’une population de piétons ou de véhicules sur une voie rectiligne, population représentée par une densité linéique $\rho(x, t)$. On considère que la vitesse des entités est fonction de la densité : $v = v(\rho)$. La manière la plus simple de prendre en compte le fait que la vitesse est d’autant plus faible que la densité est importante est $v(\rho) = U(1 - \rho/\rho_{\max})$. La conservation de la masse s’écrit alors

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} (\rho v(\rho)) = 0,$$

qui a la forme d’une équation de conservation que l’on peut écrire sous forme générale

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} f(\rho) = 0, \quad (\text{VI.6.1})$$

où f est le *flux*.

Propagation des perturbations

Si l’on considère une solution stationnaire ρ_{eq} de l’équation, et une solution perturbée $\rho_{eq} + h$, on obtient formellement une équation de transport sur la perturbation :

$$\partial_t h + f'(\rho_{eq}) \partial_x h = 0 \quad (\text{VI.6.2})$$

qui exprime que les perturbations sont transportées à la vitesse $f'(\rho_{eq})$.

Supposons que $\rho(x, t)$ est une solution régulière de cette équation. On appelle courbe caractéristique une courbe $t \mapsto x(t)$ telle que

$$\dot{x}(t) = f'(\rho(x(t), t)).$$

On vérifie immédiatement que ρ est constant le long de telles courbes :

$$\frac{d}{dt} \rho(x(t), t) = \partial_t \rho(x(t), t) + \dot{x}(t) \partial_x \rho(x(t), t) = \partial_t \rho(x(t), t) + f'(\rho(x(t), t)) \partial_x \rho(x(t), t) = 0.$$

Comme ρ est constant le long de la caractéristique, la célérité (fonction de cette seule densité) elle-même est constante, et l’on a

$$t \mapsto x + t f'(\rho_0(x)).$$

Si l’on se donne une densité initiale ρ_0 , on peut ainsi construire la solution associée en reportant la valeur de densité initiale le long des caractéristiques. Cette démarche n’est évidemment possible que tant que les caractéristiques ne se croisent pas.

Pour une densité initiale donnée, supposée lisse (continûment différentiable), on peut considérer le flot associé aux caractéristiques

$$\Phi_t : x \mapsto x + f'(\rho(x_0, 0))t.$$

Si l’on suppose que la fonction f est C^2 , on peut calculer le jacobien de la transformation

$$J(t, x) = 1 + t f''(\rho_0(x)) \rho'_0(x).$$

Ce Jacobien reste > 0 (la transformation est un difféomorphisme, i.e. les trajectoires ne se croisent pas) pour tout t si $f''(\rho_0(x)) \rho'_0(x) \geq 0$. Si en revanche cette dernière quantité est négative, alors l'application ne sera régulière que pour

$$t < -\frac{1}{f''(\rho_0(x)) \rho'_0(x)}.$$

Le temps de vie de la solution lisse sera donc

$$T = \frac{1}{\max |(f''(\rho_0(x)) \rho'_0(x))_-|}$$

(inverse du max de la partie négative de $f''(\rho_0(x)) \rho'_0(x)$).

Si l'on considère le flux indiqué précédemment $f(\rho) = U\rho(1 - \rho/\rho_{\max})$, on a $f''(\rho) = -2U/\rho_{\max} < 0$. On aura donc existence de solution lisse si ρ_0 est décroissante, et croisement de caractéristique en temps fini si en revanche ρ_0 est croissante.

Remarque VI.6.1. On prendra garde au fait que, bien que l'on ait considéré le Jacobien de l'application Φ_t , ce qui suggère un transport de mesure, n'est aucunement associée à un quelconque transport conservatif de masse.

Solutions faibles

Les considérations précédentes indiquent qu'il ne peut, en général, exister de solution lisse globale. Pour donner un sens aux solutions non lisses qui sont susceptibles d'apparaître spontanément, on définit la notion de solution faible :

Définition VI.6.2. On dit que $\rho \in L^1_{loc}(\mathbb{R} \times]0, T[)$ est une solution faible de (VI.6.1) sur $\mathbb{R} \times]0, T[$ si $f(\rho) \in L^1_{loc}(\mathbb{R} \times]0, T[)$ et si, pour tout φ , fonction C^1 à support compact dans $\mathbb{R} \times]0, T[$, on a

$$\int_{\mathbb{R}} \int_0^T \partial_t \varphi \rho(x, t) dx dt + \int_{\mathbb{R}} \int_0^T \partial_x \varphi f(\rho(x, t)) dx dt = 0.$$

On peut intégrer une condition initiale à cette définition. On dira que ρ est solution faible associée à la condition initiale $\rho|_{t=0} = \rho^0 \in L^1_{loc}(\mathbb{R})$ si

$$\int_{\mathbb{R}} \int_0^T \partial_t \varphi \rho(x, t) dx dt + \int_{\mathbb{R}} \int_0^T \partial_x \varphi f(\rho(x, t)) dx dt + \int_{\mathbb{R}} \varphi(x, 0) \rho^0(x) dx = 0$$

pour toute fonction φ régulière à support compact dans $\mathbb{R} \times [0, T[$

On vérifie immédiatement que toute solution régulière est solution faible. Mais cette définition peut s'appliquer à des solutions qui ne sont pas régulières. Considérons par exemple deux densités qui réalisent le même flux : $F = f(\rho_-) = f(\rho_+)$. La densité

$$\rho = \rho_- \mathbf{1}_{]-\infty, 0[} + \rho_+ \mathbf{1}_{]0, +\infty[}$$

est solution faible stationnaire de (VI.6.1), de même que la densité obtenue en intervertissant ρ_- et ρ_+ . On peut construire des solutions non stationnaires de la façon suivante : on se donne deux densités ρ_L et ρ_R , et l'on cherche une solution ρ constante de part et d'autre d'un point de discontinuité $s(t)$ variable en temps. On vérifie qu'une telle densité est solution faible dès que s vérifie une condition dite de *Rankine-Hugoniot*, comme l'exprime la

Proposition VI.6.3. (Relation de Rankine-Hugoniot)

On suppose la fonction flux f continue sur son intervalle de définition, et ρ_L et ρ_R deux valeurs sur cet intervalle. La densité

$$\rho = \rho_L \mathbf{1}_{]-\infty, s(t)[} + \rho_R \mathbf{1}_{]s(t), +\infty[}$$

est solution faible de (VI.6.1) si et seulement si la discontinuité s progresse à la vitesse constante

$$\dot{s} = \frac{f(\rho_L) - f(\rho_R)}{\rho_L - \rho_R}. \tag{VI.6.3}$$

Démonstration. On utilise la définition d'une solution faible, en écrivant la première intégrale double

$$\int_{\mathbb{R}} \int_0^{+\infty} \partial_t \varphi \rho = \int_0^{+\infty} \left(\rho_L \int_{-\infty}^{s(t)} \partial_t \varphi + \rho_R \int_{s(t)}^{+\infty} \partial_t \varphi \right),$$

avec

$$\int_{-\infty}^{s(t)} \partial_t \varphi = \frac{d}{dt} \left(\int_{-\infty}^{s(t)} \varphi \right) - \dot{s}(t) \varphi(s(t), t), \quad \int_{s(t)}^{+\infty} \partial_t \varphi = \frac{d}{dt} \left(\int_{s(t)}^{+\infty} \varphi \right) + \dot{s}(t) \varphi(s(t), t).$$

La seconde intégrale double (avec la dérivée en espace sur la fonction test s'écrit

$$\begin{aligned} \int_{\mathbb{R}} \int_0^{+\infty} \partial_x \varphi f(\rho(x, t)) &= \int_0^{+\infty} \left(f(\rho_L) \int_{-\infty}^{s(t)} \partial_x \varphi + f(\rho_R) \int_{s(t)}^{+\infty} \partial_x \varphi \right) \\ &= \int_0^{+\infty} \varphi(s(t), t) (f(\rho_L) - f(\rho_R)). \end{aligned}$$

On obtient donc finalement

$$\int_0^{+\infty} \varphi(s(t), t) (-\dot{s}(t)(\rho_L - \rho_R) + f(\rho_L) - f(\rho_R)),$$

qui est identiquement nul pour toute fonction test φ si et seulement si la condition (VI.6.3) est identiquement vérifiée. \square

Remarque VI.6.4. On peut retrouver la relation (VI.6.3) en écrivant simplement un bilan de masse au voisinage de la discontinuité.

Remarque VI.6.5. On peut voir cette formule comme la généralisation de la formule donnant la vitesse de propagation de perturbations au voisinage d'une densité uniforme, en prenant $\rho_R = \rho_L + \varepsilon$, ce qui donne $\dot{s} \approx f'(\rho_L)$.

On peut vérifier que, sous sa forme faible, l'équation n'est pas bien posée, au sens où elle admet en général plusieurs solutions. La théorie complète de telles équation dépasse le cadre de ce cours sous sa forme actuelle, disons simplement ici qu'il est possible d'imposer à la solution considérer de vérifier un critère supplémentaire, dit *d'entropie*, qui permet de sélectionner *la* solution physique¹² parmi les nombreuses possibles. Ce critère n'est pertinent que pour discriminer des solutions qui présentent des discontinuités, on peut montrer que ces solutions acceptables sont telles que, lorsque la solution présente une discontinuité, les courbes caractéristiques doivent arriver vers la discontinuité, et non pas en partir. Le développement précédent donnant la vitesse de propagation de la discontinuité en fonction des états à gauche et à droite, on peut exprimer le fait que les caractéristiques vont vers la discontinuité de la façon suivante :

Définition VI.6.6. Soit $\rho(x, t)$ une solution faible de l'équation de conservation (VI.6.1), avec $f(\cdot)$ une fonction C^1 , au sens de la définition VI.6.2. On suppose que ρ présente localement (au voisinage d'un point de l'espace temps) une discontinuité entre les valeurs ρ_L et ρ_R . On dit que cette discontinuité vérifie la condition d'entropie de Lax si

$$f'(\rho_L) > \frac{f(\rho_R) - f(\rho_L)}{\rho_R - \rho_L} > f'(\rho_R).$$

On notera que, dans le cas où f est convexe (ou f concave), la condition ci-dessus peut se limiter à l'inégalité entre les bornes.

12. Ce type de critère a été élaboré dans le cadre de la dynamique des gaz. Précisons que, dans le cadre du transport d'entités vivantes, sa légitimité est moins nette

VI.7 Autour de la notion de complétude, théorème de Banach-Steinhaus

VI.7.1 Lemme de Baire

Définition VI.7.1. (Parties maigres)

Soit X un espace topologique, et A une partie de X . On dit que A est maigre si A est contenue dans une réunion dénombrable de fermés d'intérieur vide.

Définition VI.7.2. (Espaces de Baire)

Un espace topologique est appelé espace de Baire si toute partie maigre est d'intérieur vide.

Exercice VI.7.1. Vérifier que l'espace des suites qui s'annulent au-delà d'un certain rang, muni de la norme ℓ^∞ , n'est pas un espace de Baire.

Le théorème suivant traduit le fait que tout espace métrique complet est de Baire.

Théorème VI.7.3. (Lemme de Baire)

Soit X un espace métrique complet, et $(X_n)_{n \in \mathbb{N}}$ une suite de fermés de X . On suppose que

$$\text{Int}(X_n) = \emptyset \quad \forall n \in \mathbb{N}.$$

On a alors

$$\text{Int}\left(\bigcup_{n=0}^{+\infty} X_n\right) = \emptyset.$$

On utilise pour la démonstration une formulation équivalente du théorème : soit X un espace métrique complet, et $(U_n)_{n \in \mathbb{N}}$ une suite d'ouverts de X denses dans X . Alors l'intersection des U_n est dense dans X .

Démonstration. On introduit

$$U = \bigcap_{n \in \mathbb{N}} U_n,$$

et on se donne $x \in X$. Pour toute boule $B(x, r)$, on va construire une suite a_n qui converge vers une limite a dans $B(x, r) \cap U$, ce qui établira la densité de U . Comme U_o est un ouvert dense, il existe $a_o \in U_o$ et $r_o > 0$, avec $r_o \leq r/2$, tel que

$$\overline{B}(a_o, r_o) \subset B(x, r).$$

On construit par récurrence la suite (a_n) de la façon suivante : supposons a_k et r_k construits pour $k \leq n$, la densité de l'ouvert U_{n+1} assure l'existence d'une boule fermée $\overline{B}(a_{n+1}, r_{n+1})$ incluse dans $U_{n+1} \cap B(a_n, r_n)$, et telle que $r_{n+1} \leq r_n/2$. Les suites $(a_n)_{n \in \mathbb{N}}$ et $(r_n)_{n \in \mathbb{N}}$ ainsi construites, on vérifie immédiatement que

$$d(a_n, a_{n+1}) < r_n \leq \frac{r}{2^n},$$

d'où l'on déduit que $(a_n)_{n \in \mathbb{N}}$ est de Cauchy, donc qu'elle converge vers une limite $a \in X$. Par construction, a est dans $B(x, r)$ et dans chacune des boules fermées $\overline{B}(a_n, r_n)$, il est donc dans U , ce qui termine la démonstration. \square

Exercice VI.7.2. Soit E un e.v.n. de dimension infinie. Montrer que si E est complet, alors sa dimension est non dénombrable, i.e. il n'existe pas de famille génératrice dénombrable.

VI.7.2 Théorème de Banach Steinhaus

Définition VI.7.4. Soient E et F deux espaces vectoriels normés. On note $\mathcal{L}(E, F)$ l'espace des applications linéaires **continues** de E dans F . C'est un e.v.n. muni de la norme

$$\|f\|_{\mathcal{L}(E, F)} = \sup_{x \neq 0} \frac{\|f(x)\|_F}{\|x\|_E}.$$

De plus, si F est complet, alors $\mathcal{L}(E, F)$ est un espace de Banach.

Théorème VI.7.5. (Banach-Steinhaus)

Soient E et F deux espaces de Banach et $(T_\alpha)_{\alpha \in A}$ une famille (non nécessairement dénombrable) d'opérateurs de $\mathcal{L}(E, F)$. On suppose

$$\sup_{\alpha \in A} \|T_\alpha x\|_F < +\infty \quad \forall x \in E.$$

On a alors

$$\sup_{\alpha \in A} \|T_\alpha\|_{\mathcal{L}(E, F)} < +\infty.$$

Ce théorème est parfois appelé “principe de la borne uniforme”. En effet, on déduit de majorations ponctuelles une majoration uniforme : ce théorème assure l'existence d'une constante $c > 0$ telle que

$$\|T_\alpha x\|_F \leq c \|x\|_E \quad \forall x \in E \quad \forall \alpha \in A.$$

Démonstration. On introduit les ensembles

$$E_n = \left\{ x \in E, \sup_{\alpha \in A} \|T_\alpha x\|_F \leq n \right\}.$$

Les E_n sont des fermés de E comme intersection de fermés. D'autre part leur réunion est E tout entier, d'après l'hypothèse. L'un des E_n est donc d'intérieur non vide. Soit n_o tel que $\text{Int}(E_{n_o}) \neq \emptyset$. Il existe $a \in E$ et $\rho > 0$ tel que $\overline{B}_\rho(a) \subset E_{n_o}$, d'où

$$\|T_\alpha(a + \rho u)\| \leq n_o \quad \forall u \in B_E.$$

On a donc, pour tout $\alpha \in A$,

$$\|T_\alpha\|_{\mathcal{L}(E, F)} \leq \frac{1}{\rho} \left(n_o + \sup_{\alpha \in A} \|T_\alpha(a)\|_F \right),$$

ce qui conclut la démonstration. \square

Le théorème de Banach-Steinhaus est souvent utilisé sous la forme suivante :

Corollaire VI.7.6. Soient E et F deux Banach, et $(T_n)_{n \in \mathbb{N}}$ une suite d'opérateurs de $\mathcal{L}(E, F)$ telle que, pour tout $x \in E$, $T_n x$ converge vers un élément de F , que l'on note Tx . On a alors

1. $\sup_{n \in \mathbb{N}} \|T_n\| < +\infty$,
2. $T \in \mathcal{L}(E, F)$,
3. $\|T\|_{\mathcal{L}(E, F)} \leq \liminf \|T_n\|_{\mathcal{L}(E, F)}$.

La dernière inégalité du corollaire précédent peut être stricte. Considérer par exemple $E = \ell^p$ avec $p \in [1, +\infty[$, et la suite des formes linéaires

$$T_k : x = (x_n)_{n \in \mathbb{N}} \mapsto x_k \in \mathbb{R}.$$

Corollaire VI.7.7. Soit G un espace vectoriel normé. Un sous-ensemble B de G est borné si et seulement si

$$f(B) = \{\langle f, x \rangle, x \in B\} \text{ est borné} \quad \forall f \in G'.$$

Démonstration. On applique le théorème VI.7.5 avec $E = G'$, $F = \mathbb{R}$, et la famille d'applications, indexée par B lui-même, $(T_x)_{x \in B}$:

$$f \in G' \longmapsto \langle f, x \rangle.$$

Le théorème assure l'existence d'une constante c telle que

$$|\langle f, x \rangle| \leq c \|f\| \quad \forall x \in B \quad \forall f \in G'.$$

Comme $\|x\| = \sup_{f \in B_{G'}} \langle f, x \rangle$ (où $B_{G'}$ est la boule unité fermée de G'), $\|x\| \leq c$ sur B . \square

Remarque VI.7.8. Le corollaire précédent ne nécessite pas l'hypothèse G complet. L'espace qui joue le rôle de l'espace de départ dans le théorème de Banach–Steinhaus est le dual topologique de G , qui est toujours complet.

VI.7.3 Théorème de l'application ouverte, théorème du graphe fermé

Théorème VI.7.9. (Application ouverte)

Soient E et F deux espaces de Banach et soit $T \in \mathcal{L}(E, F)$ surjectif. Alors il existe une constante c telle

$$B(0, c) \subset T(B_E).$$

La conclusion signifie que T transforme tout ouvert en un ouvert.

Démonstration: La démonstration s'effectue en deux étapes. On montre dans un premier temps que l'adhérence de $T(B_E)$ contient une boule ouverte centrée en 0. On note $C = \overline{T(B_E)}$ cette adhérence. Les $C_n = nC$ sont des fermés de F par construction, et leur union est F tout entier (car T est surjective). L'espace d'arrivée F étant complet, il en existe donc un d'intérieur non vide (d'après le lemme de Baire), donc C lui-même est d'intérieur non vide (les C_n sont homothétiques à C) : C contient une boule ouverte $B = B(a, 2c)$. Par symétrie de C , $-B$ est également dans C , et par convexité (C est l'adhérence de l'image d'un convexe par une application linéaire),

$$\begin{aligned} \frac{1}{2}B + \frac{1}{2}(-B) &= \left\{ \frac{1}{2}a - \frac{1}{2}a + \frac{1}{2}h_1 - \frac{1}{2}h_2, h_1, h_2 \in B(0, 2c) \right\} \\ &= \{0 + h, h \in B(0, 2c)\} \\ &= B(0, 2c) \subset C = \overline{T(B_E)}. \end{aligned}$$

On va maintenant montrer, en utilisant cette fois la complétude de l'espace de départ E , que C contient $B(0, c)$. On se donne $y \in B(0, c)$, et on cherche à construire un antécédent x dans B_E . D'après (VI.7.1), pour tout $\varepsilon > 0$, il existe z dans $1/2B_E$ tel que $\|y - Tz\| < \varepsilon$. On construit ainsi z_1 tel que

$$\|y - Tz_1\| < \frac{c}{2}, \quad \|z_1\| \leq \frac{1}{2}.$$

On construit de la même manière z_2

$$\|(y - Tz_1) - Tz_2\| < \frac{c}{4}, \quad \|z_2\| \leq \frac{1}{4},$$

puis par récurrence les termes de la suite (z_n) tels que

$$\|(y - Tz_1 - \cdots - Tz_{n-1}) - Tz_n\| < \frac{c}{2^n}, \quad \|z_n\| \leq \frac{1}{2^n}.$$

Par construction la suite $x_n = z_1 + \cdots + z_n$ est de Cauchy, donc converge vers un certain x de norme ≤ 1 , et on a bien $y = Tx$ par continuité de T . \square

On en déduit le

Corollaire VI.7.10. Soient E et F deux Banach et soit $T \in \mathcal{L}(E, F)$ bijectif. Alors T^{-1} est continu de F dans E .

Démonstration: Comme T est surjectif, on a, d'après le théorème de l'application ouverte, $\|Tx\| \geq c$ pour tout x de norme 1. On en déduit, par homogénéité,

$$\|Tx\| \geq c \|x\| \quad \forall x \in E,$$

d'où

$$\|T^{-1}y\| \leq \frac{1}{c} \|y\| \quad \forall y \in F.$$

□

Dans le cas où T n'est pas surjectif, on peut néanmoins utiliser le théorème de l'application ouverte pour la même application vue comme un opérateur de E dans $T(E)$, sous réserve que $T(E) \subset F$ soit un espace de Banach :

Corollaire VI.7.11. Soient E et F deux espaces de Banach, et $T \in \mathcal{L}(E, F)$. On suppose que l'image de T est fermée. Alors il existe $\alpha > 0$ tel que

$$\forall y \in T(E), \exists x \in E, \|x\| \leq \alpha \|y\|, y = Tx. \quad (\text{VI.7.1})$$

Démonstration: L'espace $T(E)$ est fermé dans le complet F , il est donc complet. On note T_1 l'application T vue comme surjection de E dans $T(E)$. On peut appliquer le théorème de l'application ouverte à T_1 :

$$\exists c > 0, B_{T(E)}(0, c) \subset T_1(B_E).$$

Pour tout $y \in T(E)$, on a $cy/2\|y\| \in B_{T(E)}(0, c)$. Il existe donc $x \in B_E$ tel que

$$T_1x = Tx = \frac{cy}{2\|y\|},$$

d'où

$$T\left(\frac{2\|y\|}{c}x\right) = y, \left\|\frac{2\|y\|}{c}x\right\| \leq \frac{2}{c}\|y\|,$$

ce qui termine la démonstration ($\alpha = 2/c$). On remarque que, si T est de plus injectif, ce corollaire exprime simplement la continuité de l'application réciproque T^{-1} définie sur $T(E)$, conformément au corollaire précédent. □

Définition VI.7.12. (Graphe)

Soient E et F deux e.v.n et T un opérateur linéaire de E vers F . On appelle graphe de T l'ensemble

$$G(T) = \{(x, Tx) \mid x \in E\} \subset E \times F.$$

Théorème VI.7.13. (Graphe fermé)

Soient E et F deux Banach, et T un opérateur linéaire de E vers F . On a

$$G(T) \text{ fermé} \iff T \in \mathcal{L}(E, F).$$

Démonstration: La condition suffisante est immédiate. Supposons maintenant $G(T)$ fermé. On considère la nouvelle norme sur E :

$$\|x\|_1 = \|x\|_E + \|Tx\|_F.$$

Montrons que E est complet pour la norme $\|\cdot\|_1$. Toute suite de Cauchy $(x_n)_{n \in \mathbb{N}}$ pour $\|\cdot\|_1$ est de Cauchy pour $\|\cdot\|_E$, donc converge vers x dans E . De la même manière, (Tx_n) converge vers $y \in F$. Comme $G(T)$ est fermé, on a $y = Tx$, et $(x_n)_{n \in \mathbb{N}}$ converge bien vers x pour $\|\cdot\|_1$. Comme $(E, \|\cdot\|_1)$ est complet, d'après le corollaire VI.7.10, l'identité $(E, \|\cdot\|_1) \rightarrow (E, \|\cdot\|)$ est bicontinue. La norme $\|\cdot\|_1$ est donc équivalente à la norme $\|\cdot\|_E$, d'où la continuité de T . □

Exercice VI.7.3. Soit E un espace de Banach, et T une application linéaire de E dans E' telle que

$$(Tx, x) \geq 0 \quad \forall x \in E.$$

Montrer que T est continue. (On pourra utiliser le théorème du graphe fermé : on considère $(x_n, T(x_n)) \rightarrow (x, f)$, et on cherche à montrer que $(x, f) \in G(T)$. On pourra montrer dans un premier temps que $(f - Tz, x - z) \geq 0$ pour tout $z \in E$, puis prendre z de la forme $z = x + \lambda h$.

Annexe A

Fondamentaux et compléments

Sommaire

A.1	Fondamentaux	169
A.1.1	Éléments de théorie des ensembles	169
A.1.2	Structures fondamentales : relations et structures algébriques	170
A.1.3	Cardinalité	172
A.1.4	L'ensemble des réels : construction et structures afférentes	178
A.1.5	Inégalités fondamentales	183
A.2	Pour aller plus loin (****)	185
A.2.1	Théorie des ensembles, cardinalité	185
A.2.2	Complété d'un espace métrique (****)	185
A.2.3	Topologie générale (****)	186

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 :

$$x \in A \implies x \in B.$$

On a¹ $\emptyset \subset B$ 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 *graph* 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 note³ $f^{-1}(\{y\})$ (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

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

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

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

On parle de *relation d'équivalence* si elle vérifie 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 néanmoins 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 $x, y \in X$, $x \mathcal{R} y \iff y \mathcal{R} x$.
- (iii) (*transitivité*) Pour tous $x, y, z \in X$,

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

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

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

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

L'application qui à $x \in X$ associe sa classe \bar{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 \bar{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 \bar{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, notées $\bar{0}$ et $\bar{1}$. On peut définir sur l'espace quotient une addition $\bar{x} + \bar{y} = \overline{x + y}$ (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

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

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 $x, y, z \in X$,

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

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 $X \simeq Y$ ou⁴ $\text{Card}(X) = \text{Card}(Y)$.

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

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

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

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 . \square

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 grande 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 fini⁵, 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 $\text{Card}(X) = \text{Card}(Y)$ 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" ($\text{Card}(X)$ et $\text{Card}(Y)$) qui n'ont pas été définies.

5. Certains auteurs considèrent que l'attribut dénombrable est restreint aux ensembles 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} .

	9							
X_2	5	8						
X_1	2	4	7					
X_0	0	1	3	6				

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 : $X_n = \{x_n^k, k \in \mathbb{N}\}$. On peut énumérer les éléments de la réunion de la façon suivante :

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

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. \square

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

Démonstration. On considère N ensembles dénombrables X_1, \dots, X_N , pour lesquels on se donne une énumération, et l'on note $P_k \subset X_1 \times \dots \times X_N$ les éléments du produit 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). \square

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 ?

CORRECTION.

Un magma est déterminé par le choix (sans contrainte) d'un élément de X associé à tout couple (x, y) . Le nombre de ces couples est N^2 , le nombre de choix possible est N , le nombre de magmas est donc N^{N^2} . Si l'on impose à la loi d'être commutative, ce nombre est $N^2(N+1)/2$.

Définition A.1.18. (Monoïde)

Un monoïde est un ensemble muni d'une loi de composition interne associative, qui admet un élément neutre.

Exercice A.1.4. On se place sur $X = [0, 1]$. Pour toute paire de lois de Bernoulli de paramètres p et q , on considère l'expression de la probabilité que parmi deux tirages de v.a. suivant ces lois, l'une au moins des valeurs soient 1, qui s'exprime

$$p \oplus q = 1 - (1 - p)(1 - q) = p + q - pq.$$

Montrer que X muni de \oplus est un monoïde abélien. Quel est le lien entre \oplus et l'addition de deux réels?

CORRECTION.

Il s'agit bien d'une loi de composition interne commutative, et on a

$$(p_1 \oplus p_2) \oplus p_3 = p_1 \oplus p_2 + p_3 - (p_1 \oplus p_2)p_3 = p_1 + p_2 - p_1 p_2 + p_3 - (p_1 + p_2)p_3 + p_1 p_2 p_3 = p_1 + p_2 + p_3 - p_1 p_2 - p_2 p_3 - p_1 p_3 + p_1 p_2 p_3,$$

qui est invariant par permutation des indices. On a donc, du fait de la commutativité, associativité.

Pour p et q petits, on a

$$p \oplus q = p + q + O(pq),$$

qui vaut $p + q$ au premier ordre.

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 :

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

- (iii) Tout élément admet un inverse :

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

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.1.5. 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

CORRECTION.

On suppose qu'il existe e tel que $e \star x = x$, et que pour tout x , il existe y tel que $y \star x = e$. Montrons que les conditions (ii) et (iii) sont alors vérifiées. Pour tout $x \in G$, il existe y tel que $y \star x = e$. Il existe aussi z tel que $z \star y = e$. On a alors

$$x \star y = e \star (x \star y) = (z \star y) \star (x \star y) = z \star (y \star x) \star y = z \star y = e.$$

Montrons maintenant que e est aussi un inverse à droite. Pour tout $x \in G$, on sait maintenant qu'il existe y tel que $y \star x = x \star y = e$. On a alors

$$x \star e = x \star (y \star x) = (x \star y) \star x = e \star x = x.$$

Pour l'unicité de l'inverse, on considère y et y' deux inverses de x , on a

$$y = y \star e = y \star (x \star y') = (y \star x) \star y' = y'.$$

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.

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

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 .

Définition A.1.23. (Partie génératrice)

Soit (G, \star) un groupe. On a qu'une partie S de G est *généatrice* 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 aucunes relations autres que celles imposées 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

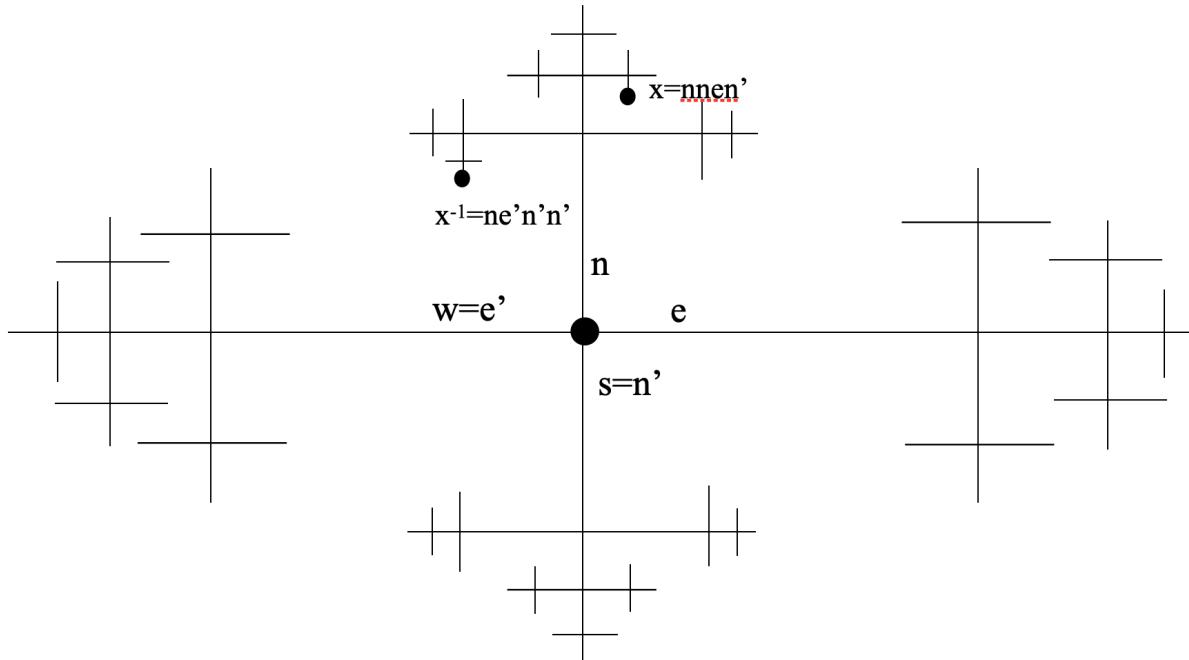


FIGURE A.1.2 – Groupre libre engendré par 2 éléments

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 chaînes 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.1.6. Décrire le groupe libre engendré par un singleton.

CORRECTION.

On note x l'unique élément de S , x' l'unique élément de S' . Avec la règle indiquée, tout mot de F_S peut s'écrire comme une suite finie de x , ou une suite finie de x' . Si l'on associe à $xx \dots x$ son nombre de lettres, et à $x'x' \dots x'$ l'opposé de son nombre de lettres, on construit un isomorphisme entre F_S et \mathbb{Z} .

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.

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

(iii) La loi \star admet un élément neutre⁶

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 :

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

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

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

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 minimisliste⁷, constitué des *combinaisons linéaires formelles* d'éléments de 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

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

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 $(X^n)_{n \in \mathbb{N}}$.

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 développements 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,

6. Cette condition n'est pas demandée par certains auteurs. Pour éviter toute ambiguïté, 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.

d'avoir conscience des structures que l'on utilise, qui représentent en quelque sorte l'essence des objets que l'on manipule. Il nous paraît en particulier fécond, lorsque l'on manipule un ensemble multiplement structuré⁸ 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 produits) 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* S_N , ensemble des bijections d'un ensemble à N éléments muni de la loi de composition. Un autre exemple fécond correspond au 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 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 développements 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 '×' 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 $\{+, -\} \times \mathbb{N} \times \llbracket 0, 9 \rrbracket^{\mathbb{N}^*}$, c'est-à-dire de l'ensemble des

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

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'écriture⁹, qui concerne les nombres décimaux. Tout nombre décimal peut

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, ...).

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 $0.111\dots + 0.888\dots$.

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'infini 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

$$\begin{aligned} r_1 &= 0, \mathbf{a}_1^1 a_1^2 a_1^3 \dots \\ r_2 &= 0, a_2^1 \mathbf{a}_2^2 a_2^3 \dots \\ r_3 &= 0, a_3^1 a_3^2 \mathbf{a}_3^3 \dots \\ &\vdots = 0, \vdots \end{aligned}$$

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

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 $[a, b]$, $]a, b[$, $[a, b[\dots$ \square

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

$$m_1 = \max_{a \in A_0} \{P_1(a)\} \in \llbracket 0, 9 \rrbracket.$$

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. \square

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 maximale¹⁰ $c_k c_{k+1} \dots c_\ell = 99 \dots 9$, 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 .

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 $] -1, 0[$, $a = -0, a_1 a_2 \dots$ (écriture propre), auquel on ajoute +1. Si a est décimal, $a = -0, a_1 a_2 \dots a_p$, la somme est définie comme

$$1 + a = 0, c_1 c_2 \dots c_p, \quad c_i = 9 - a_i \text{ pour } i = 1, \dots, p-1, \quad c_p = 10 - a_p.$$

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

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)

$$a + b = a_0 + 0, a_1 a_2 \dots - b_0 = a_0 - b_0 + 0, a_1 a_2 \dots$$

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

$$a + b = -c_0, c_1 \dots \text{ avec } c_0 = |a_0 - b_0| - 1 \text{ et } 0, c_1 c_2 \dots = 1 - 0, a_1 a_2 \dots$$

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

$$a + b = a_0 + 0, a_1 a_2 \dots - b_0 - 1 + (1 - 0, b_1 \dots) = a_0 - b_0 - 1 + 0, a_1 a_2 \dots + (1 - 0, b_1 b_2 \dots).$$

Le terme entre parenthèses a été défini précédemment comme un réel de l'intervalle $]0, 1[$. On sait effectuer sa somme avec $0, a_1 a_2 \dots$ (tous deux positifs). Si cette somme est dans l'intervalle $]0, 1[$, on se retrouve dans la situation précédente. Sinon, on l'écrit $1 + 0, d_1 d_2 \dots$, 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 $0, d_1 d_2 \dots$, 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 $x_n \leq \ell$ 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 . \square

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 \limsup et \liminf .

Définition A.1.36. (Limite inférieure, limite supérieure (••))

Soit x_n une suite réelle, on définit

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

Exercice A.1.7. (Limites supérieure et inférieure (•))

a) Donner les \limsup et \liminf de la suite (x_n) dans les cas suivants

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

b) Que peut-on dire d'une suite dont la \liminf et la \limsup ont la même valeur finie ?

CORRECTION.

XXX

Exercice A.1.8. (•) Soient (x_n) et (y_n) deux suites réelles. Montrer que

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

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 (\bullet))

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 $(x, y) \in \mathbb{R} \mapsto d(x, y) = |x - y|$ 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. \square

Proposition A.1.39. L'espace métrique (\mathbb{R}, d) est complet (voir définition I.5.3, page 23).

Démonstration. On considère une suite de Cauchy dans \mathbb{R} (défini par A.1.27), notée¹¹ (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

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

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). \square

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 $a = a_0, a_1 a_2 \dots$, 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 $a = +a_0, a_1 a_2 \dots$ et $b = +b_0, b_1 b_2 \dots$ 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, \dots, x_n des réels positifs ou nuls, et $(\alpha_n) \in]0, +\infty[^n$ une famille de poids. On a

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

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. \square

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 $x_2 = b^q$. \square

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 $\sum \theta_i |x_i|^p = \sum \theta_i |y_i|^q = 1$, 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. \square

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

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

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

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

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

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 élément 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

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

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

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

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 $\bar{d}(\bar{x}, \bar{x}')$ cette limite.

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

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

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

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

Montrons maintenant que \bar{X} muni de \bar{d} est un espace métrique complet. On considère une suite (\bar{x}^n) de Cauchy dans \bar{X} . Pour tout n , comme on l'a vu précédemment, la classe \bar{x}^n peut être approchée par la classe d'une suite constante écrite \bar{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

$$d(u_p, u_q) = \bar{d}(\bar{u}_p, \bar{u}_q) \leq \bar{d}(\bar{u}_p, \bar{x}^p) + \bar{d}(\bar{x}^p, \bar{x}^q) + \bar{d}(\bar{x}^q, \bar{u}_q) \leq \frac{1}{p} + \bar{d}(\bar{x}^p, \bar{x}^q) + \frac{1}{q}.$$

On note \bar{u} sa classe. On a

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

qui tend vers 0 par construction. \square

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 topologies 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 voisinages 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 voisinages 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.

CORRECTION.

Soit f une application continue de X dans X' . Soit F' un fermé de X' . On a

$$f^{-1}(F')^c = f^{-1}(F^c),$$

qui est ouvert comme image réciproque d'un ouvert. On démontre la réciproque de la même manière.

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' .

CORRECTION.

a) L'image réciproque de X' est X qui est ouvert, et l'image réciproque de \emptyset est \emptyset qui est aussi ouvert. Toute application est donc continue dans ce cas.

b) L'image réciproque de tout ouvert de X' est une partie de X , qui est ouverte comme toutes les parties. Toute application est donc continue dans ce cas également.

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 partition¹² 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. \square

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 classes d'équivalence de cette relation sont appelées *composantes connexes* de X .

Remarque A.2.18. Un espace connexe est un espace qui ne possède qu'une seule composante connexe, qui est l'espace entier. Pour la topologie discrète, tout ensemble X est *totalelement 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 $n \geq N$, $x_n \in U$.

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

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.

CORRECTION.

Si la topologie est discrète, alors $\{x\}$ est un ouvert, la définition assure donc que la suite (x_n) convergeant vers x est stationnaire en x au delà d'un certain rang.

Si la topologie est grossière, alors le seul ouvert contenant x est X , la convergence vers n'importe quel x est donc assurée pour n'importe quelle suite.

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 (\bullet))

Soit (X, \mathcal{T}) un espace topologique, et K une partie de X (qui peut être X lui-même). On dit que K est *compact* s'il vérifie la propriété de *Borel-Lebesgue* : de tout recouvrement de K par des ouverts on peut extraire un recouvrement fini :

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

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

CORRECTION.

Pour la topologie discrète, tout singleton est ouvert. Pour toute partie K on peut donc considérer les recouvrements par les singletons. La compacité impose donc que l'ensemble soit fini, et cette condition est bien sûr suffisante.

À l'opposé, si la topologie est grossière, le seul recouvrement possible d'une partie non vide est X lui-même. Toute partie, y compris l'espace lui-même, est donc compacte.

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 possibles¹³ 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	189
B.2	Tribus, espaces mesurables	193
B.2.1	Tribus	193
B.2.2	Applications mesurables	197
B.2.3	Classes monotones	198
B.3	Mesures	200
B.4	Mesures extérieures	204
B.4.1	Définitions, premières propriétés	204
B.4.2	D'une mesure extérieure à une mesure	205
B.4.3	Mesure de Lebesgue	207
B.5	Compléments	210
B.6	Exercices	213
B.7	Fonctions mesurables, intégrale de Lebesgue	219
B.7.1	Fonctions mesurables	219
B.7.2	Intégrale de fonctions étagées	222
B.7.3	Intégrale de fonctions mesurables	225
B.7.4	Théorèmes fondamentaux	228
B.7.5	Intégrales multiples	230
B.8	Exercices	234

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 connaît 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 $X = \llbracket 1, N \rrbracket$ 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}_+ :

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

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 dualité² 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 considérera 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 $I \subset \llbracket 1, N \rrbracket$, 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

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

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 $X = \llbracket 1, N \rrbracket$ 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 daucune 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 σ -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'intervalle³ $]0, 1[$ 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 ou 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 $\mu^*(A)$ 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

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

n'est *pas vérifiée*. Plus précisément $\mu^*(A)$ 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 explicitement⁴. 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 famille 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 météorite 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 météorite 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) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$,
- (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. \square

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).

CORRECTION.

On a

- (i) $\emptyset = \emptyset \cap F \in \mathcal{A}_F$.
- (ii) Pour tout $A \in \mathcal{A}_F$, son complémentaire dans F s'écrit $A^c \cap F$ avec $A^c \in \mathcal{A}$ d'où son appartenance à \mathcal{A}_F , par définition.
- (iii) Si (A_n) est une collection dénombrable d'éléments de \mathcal{A}_F , chaque A_n s'écrit $B_n \cap F$ avec $B_n \in \mathcal{A}$, d'où

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} (B_n \cap F) = F \cap \underbrace{\bigcup_{n \in \mathbb{N}} B_n}_{\in \mathcal{A}}$$

qui est donc dans \mathcal{A}_F par définition.

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. \square

Comme pour toute propriété stable par intersection⁵, on peut définir la notion de plus petite tribu contenant une collection de parties de X .

Définition B.2.5. (Tribu engendrée (\bullet))

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.

CORRECTION.

Si $A = \emptyset$ ou $A = X$, la tribu est $\{\emptyset, X\}$, de cardinal 2, et si A est non vide et strictement inclus dans X , la tribu engendrée est $\{\emptyset, X, A, A^c\}$, de cardinal 4. Noter que si X est de cardinal 1, seul le premier cas est possible.

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 20). 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.

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.

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

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

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 $A' \in \mathcal{A}'$,

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

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}' . \square

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

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

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

CORRECTION.

On suppose $f(x) = x'$ pour tout x (le même x'). Soit $A' \subset X'$. Si $x' \notin A'$, alors $f^{-1}(A') = \emptyset \in \mathcal{A}$, et si $x' \in A'$, alors $f^{-1}(A') = X \in \mathcal{A}$. Il s'agit donc de la tribu discrète, quelle que soit la tribu \mathcal{A} de départ.

Si maintenant \mathcal{A}' est une tribu sur X' , alors l'image réciproque de tous ses membres qui contiennent x' (il en existe au moins un, puisque $X' \in \mathcal{A}'$) est X , et l'image réciproque de tous ses membres qui ne pas contiennent x' (il en existe au moins un, puisque $\emptyset \in \mathcal{A}'$) est \emptyset . Il s'agit donc de la tribu grossière, quelle que soit la tribu \mathcal{A}' d'arrivée.

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 mesurable⁶ 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 $f_{\sharp}(\mathcal{A})$.

CORRECTION.

Si f est $\mathcal{A} - \mathcal{A}'$ mesurable, alors \mathcal{A} doit contenir tous les $f^{-1}(A')$ pour $A' \in \mathcal{A}'$, elle contient donc en particulier $f^{-1}(\mathcal{A}')$.

Si f est $\mathcal{A} - \mathcal{A}'$ mesurable, $f^{-1}(A') \in \mathcal{A}$ pour $A' \in \mathcal{A}'$, tout $A' \in \mathcal{A}'$ est donc 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.

CORRECTION.

Soit f une application constante : $f(x) = a' \in X'$. On a

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

d'où $f^{-1}(A') \in \mathcal{A}$, puisque toute tribu contient \emptyset et X , même la plus grossière d'entre elles.

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 ?

CORRECTION.

D'après la définition, l'identité est mesurable si et seulement si tout élément de \mathcal{A}' est dans \mathcal{A} , c'est à dire si $\mathcal{A}' \subset \mathcal{A}$ (\mathcal{A} est plus fine que \mathcal{A}').

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. La condition nécessaire est immédiate. Pour la condition suffisante, on introduit

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

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}' . \square

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' .

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 ?

CORRECTION.

Si \mathcal{A} est la tribu discrète, toute application est mesurable. Si \mathcal{A} est la tribu grossière, alors dès qu'il existe A' tel que $f^{-1}(A')$ est non vide et non identifié à X , l'application f n'est pas mesurable. Les seules applications mesurables sont donc les applications qui sont d'une certaine manière constantes, mais dans un sens un peu particulier : elles sont constantes autant que A' soit en mesure de distinguer les valeurs. Plus précisément, il s'agit d'applications telles que, si $f(x) \neq f(y)$, alors nécessairement les éléments de A' qui contiennent x sont exactement ceux qui contiennent y . En effet, si ça n'est pas le cas, alors il existe $A' \in \mathcal{A}'$ qui contient $f(x)$ et pas $f(y)$, et donc l'image réciproque de A' contient x et pas y , il ne s'agit donc ni de X ni de la partie vide.

Si par exemple la tribu \mathcal{A}' est suffisamment fine pour distinguer tous les éléments, c'est-à-dire que pour tous $x' \in X'$ il existe $A' \in \mathcal{A}'$ tel que $x' \in A'$, alors les seules applications mesurables sont les applications constantes $f(x) = a' \in X'$. On peut avoir des situations intermédiaires du type (cas d'une tribu "assez grossière") : si \mathcal{A}' est la tribu engendrée par une partie A' de X' non triviale (ni vide ni totale), une tribu qui contient donc 4 membres (\emptyset, X', A' et A'^c), alors si \mathcal{A} est la tribu grossière une application est mesurable si et seulement si elle envoie tout le monde dans A , ou tout le monde dans A^c . À l'extrême, si \mathcal{A}' est elle-même la tribu grossière, alors toutes les applications sont mesurables.

La propriété énoncée précédemment dans le cas où \mathcal{A} est la tribu grossière est vraie a fortiori pour tout \mathcal{A} : si \mathcal{A}' est la tribu grossière, toute application est mesurable (quelle que soit \mathcal{A}). Si \mathcal{A}' est la tribu discrète, une application est mesurable si et seulement si l'image réciproque de toute partie est dans \mathcal{A} . Si \mathcal{A} est la tribu discrète, tout est mesurable. S'il existe des ensembles non mesurables sur X , dont on note la collection \mathcal{A}^c (complémentaire de \mathcal{A} dans $\mathcal{P}(X)$), alors les applications non mesurables sont celles pour lesquelles il existe un $B \in \mathcal{A}^c$ qui soit image réciproque d'une partie A' de $\mathcal{P}(X')$. Or une partie de X est l'image réciproque d'une partie de X' si et seulement si $f(B) \cap f(B^c) = \emptyset$ (sinon l'image réciproque de l'image de B contient des éléments qui ne sont pas dans B). L'application sera donc mesurable si et seulement s'il n'existe aucune partie B de X à l'extérieur de la tribu \mathcal{A} telle que $f(B) \cap f(B^c) = \emptyset$, ce qui peut être délicat à vérifier en pratique ...

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) $A, B \in \mathcal{D}, A \subset B \implies B \setminus A \in \mathcal{D}$,
- (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} . \square

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

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

Définition B.2.16. (Classe monotone engendrée par un ensemble de parties (\bullet))

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 (\bullet))

On appelle π -système sur un ensemble X un sous-ensemble \mathcal{C} non vide de parties de X stable par intersection finie :

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

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 .
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 π -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 $B_n \cup A_{n+1} = A_0 \cup A_1 \cup \dots \cup A_{n+1}$ (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} .

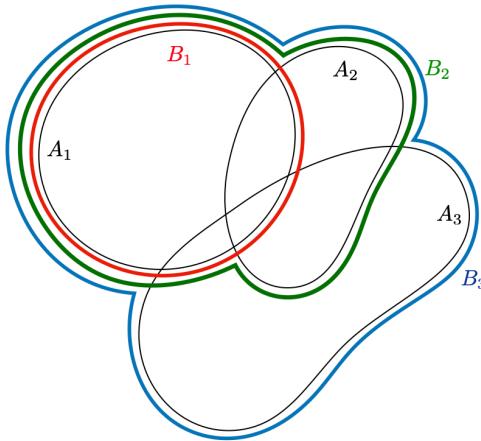
La suite (B_n) d'éléments de \mathcal{D} , est croissante par construction, d'où $\cup B_n \in \mathcal{D}$, et cette union s'identifie par construction à l'union des A_n . La famille \mathcal{D} est donc bien une tribu. \square

Proposition B.2.20. (Lemme de classe monotone)

($\bullet\bullet$) Soit \mathcal{C} un π -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

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

FIGURE B.2.1 – Construction de la suite croissante B_1, B_2, \dots

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. \square

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

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

Le triplet (X, \mathcal{A}, μ) (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 σ -*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 217). 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

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

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 addition⁷ primitive basée sur :

$$\begin{aligned} \text{rien} + \text{rien} &= \text{rien}, & \text{rien} + \text{quelque chose} &= \text{quelque chose}, \text{ et} \\ \text{quelque chose} + \text{quelque chose} &= \text{quelque chose}. \end{aligned}$$

Proposition B.3.3. (Monotonie (•))

Soit (X, \mathcal{A}, μ) 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 :

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

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 σ -finie. Il existe une partition (B_n) de X constituée d'ensembles mesurables et de mesures finies.

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*.

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 $A_n \setminus (A_{n-1} \cup \dots \cup A_0)$. 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 partition⁸ de X . \square

Proposition B.3.5. (Sous-additivité (\bullet))

Soit (X, \mathcal{A}, μ) un espace mesuré, et (A_n) une suite dans \mathcal{A} . On a

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

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

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

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

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

qui est l'inégalité annoncée. \square

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

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

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

$$\begin{aligned} \mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) &= \mu \left(\bigcup_{n \in \mathbb{N}} B_n \right) = \sum_{j=0}^{+\infty} \mu(B_j) = \lim_{n \rightarrow +\infty} \sum_{j=0}^n \mu(B_j) \\ &= \lim_{n \rightarrow +\infty} \mu \left(\bigcup_{j=0}^n B_j \right) = \lim_{n \rightarrow +\infty} \mu(A_n). \end{aligned}$$

qui termine la preuve. \square

Définition B.3.7. (Partie négligeable, mesure complète (\bullet))

Soit (X, \mathcal{A}, μ) 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} .

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 $B_n = \emptyset$.

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}, μ) 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 introducives 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 accueillir de la matière). Cette mesure sera en général *statische*, 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

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

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

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

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. \square

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

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

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 . \square

Noter que les hypothèses de la proposition précédente imposent que μ et ν soient σ -finies, mais l'on demande en outre que la famille de parties de mesures finies qui recouvre X puisse être composée d'éléments du π -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 μ^* de $\mathcal{P}(X)$ dans $[0, +\infty]$ telle que

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

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

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 $\mu^*(\emptyset) = 0$ et $\mu^*(A) = 1$ 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 $\mu^*(A) = 0$ si A est dénombrable, et $\mu^*(A) = 1$ 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 μ^* -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 μ^* une mesure extérieure sur X . On dit que A est μ^* -mesurable si, pour tout $B \subset X$, on a

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c). \quad (\text{B.4.1})$$

Du fait de la sous-additivité de μ^* , l'inégalité

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

pour tout B suffit pour assurer la μ^* -mesurabilité de A .

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

CORRECTION.

Les inégalités à vérifier pour que μ^* soit une mesure extérieure sont toutes des tautologies du type $0 \leq 0$, $0 \leq 1$, $1 \leq 1$, $1 \leq +\infty$.

Si X est un singleton, μ^* est une mesure. En revanche si X contient au moins 2 éléments a et b , on a $\mu(\{a\} \cup \{b\}) = 1 < 2 = \mu(\{a\}) + \mu(\{b\})$, il ne s'agit donc pas d'une mesure.

Si X est réduit à un singleton, toutes les parties (il y en a deux au total : X et \emptyset) sont mesurables. Si ça n'est pas un singleton, alors pour toute partie B de X non vide et non égale à X , on a

$$\mu^*(X \cap B) + \mu^*(X \cap B^c) = 1 + 1 > 1 = \mu^*(X),$$

la partie B n'est donc pas mesurable. Les seules parties mesurables pour μ^* sont donc \emptyset et X .

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 μ^* une mesure extérieure sur X (définition B.4.2). Tout $B \subset X$ tel que $\mu^*(B) = 0$ ou $\mu^*(B^c) = 0$ est μ^* -mesurable.

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

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 μ^* une mesure extérieure sur X (définition B.4.2). On note \mathcal{A}_{μ^*} la collection des parties μ^* -mesurables (selon la définition B.4.2).

Alors \mathcal{A}_{μ^*} est une tribu, et la restriction de μ^* à \mathcal{A}_{μ^*} est une mesure.

Démonstration. (•••) Montrons dans un premier temps que \mathcal{A}_{μ^*} 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}_{μ^*} . Par ailleurs l'identité (B.4.1) est inchangée si l'on échange les rôles de B et B^c , \mathcal{A}_{μ^*} est donc stable par complémentarité.

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

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

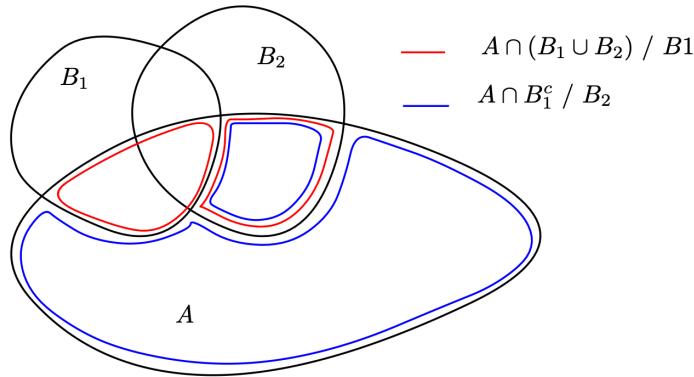


FIGURE B.4.1 – Mesurabilité de $B_1 \cup B_2$. La notation C / D de la légende indique que l'on écrit C comme l'union de $C \cap D$ et de $C \cap D^c$, où D est un ensemble mesurable.

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

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

qui est égal à $\mu^*(A)$ du fait que $B_1 \in \mathcal{A}_{\mu^*}$. L'ensemble \mathcal{A}_{μ^*} 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}_{μ^*} , 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)

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

respecte μ^* , au sens où, pour tout A

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

Nous allons en fait démontrer par récurrence l'identité équivalente

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

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

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

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 $+\infty$,

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

d'où, par σ -sous-additivité de μ^* ,

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

qui est supérieur ou égal à $\mu^*(A)$ par sous-additivité. On a donc identité entre $\mu^*(A)$ et l'expression ci-dessus, pour tout A , ce qui prouve que l'union des B_k est dans \mathcal{A}_{μ^*} . 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}_{μ^*} peut s'écrire comme union d'ensembles disjoints

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

Nous avons ainsi démontré que \mathcal{A}_{μ^*} est une tribu.

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

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

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

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 $\lambda^* : \mathcal{P}(X) \rightarrow [0, +\infty]$ par

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

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

Démonstration. (•••) Montrons que λ^* 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 $\lambda^*(\emptyset) = 0$.
- (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 $\lambda^*(A)$, porte donc sur un ensemble plus grand que celui associé à B , d'où $\lambda^*(A) \leq \lambda^*(B)$.

(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.

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

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

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

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

Il reste à montrer que λ^* 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 $\lambda^*(A)$ est donc majorée par une quantité arbitrairement proche de $b - a$, on a donc $\lambda^*([a, b]) \leq b - a$. 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

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

qui est inférieur ou égal à la longueur totale de la collection d'intervalles initiale (car on en a enlevé certains). On a donc $b - a \leq \lambda^*(A)$. Pour finir si l'on considère $]a, b[$, 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 $\lambda^*([a, b]) = b - a$. Le raisonnement est analogue pour les intervalles de type $[a, b[$ et $]a, b]$. \square

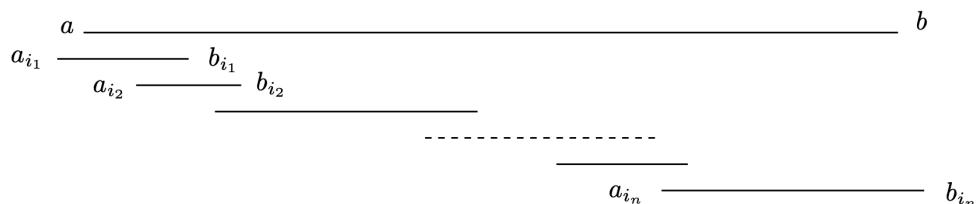


FIGURE B.4.2 – Recouvrement d'un intervalle

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

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

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 λ^* définie ci-dessus.

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

$$\lambda^*(A) \geq \lambda^*(A \cap B) + \lambda^*(A \cap B^c). \quad (\text{B.4.5})$$

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

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

Il s'agit de construire, à partir de ce recouvrement, deux collections d'intervalles qui recouvrent respectivement $A \cap B$ et $A \cap B^c$. Dans le cas présent, seul le point c pose problème, nous commençons par le gérer en considérant un intervalle $]c - \varepsilon, c + \varepsilon[$, que nous prenons comme premier intervalle du recouvrement de $A \cap B$. Ensuite pour chaque intervalle $]a_n, b_n[$, si $b_n < c$, nous l'affectons à $A \cap B$, si $a_n > c$, nous l'affectons à $A \cap B^c$, et s'il contient c , on le sépare en deux en affectant $]a_n, c[$ à $A \cap B$, et $]c, b_n[$ à $A \cap B^c$. On perd ce faisant le point c , mais il a été recouvert par l'intervalle $]c - \varepsilon, c + \varepsilon[$ introduit au début.

On obtient ainsi deux collections d'intervalles qui recouvrent respectivement $A \cap B$ et $A \cap B^c$, dont les sommes des longueurs est égal à la somme des longueurs des $]a_n, b_n[$ augmentée de 2ε (longueur de $]c - \varepsilon, c + \varepsilon[$). Pour résumer, si l'on note $\{]a'_n, b'_n[\}$ le recouvrement de gauche, et $\{]a''_n, b''_n[\}$ celui de droite, on a

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

et ainsi

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

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

D'après le théorème B.4.4, la famille des parties mesurables pour λ^* est une tribu, qui contient les $] -\infty, c]$ 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. \square

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 λ^* définie par la proposition B.4.6. Cette mesure est définie sur la tribu des parties mesurables pour la mesure extérieure λ^* . 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.

CORRECTION.

Toute réunion d'intervalles ouverts dont la somme des longueurs est finie ne pouvant recouvrir \mathbb{R} la mesure extérieure de Lebesgue de \mathbb{R} , qui s'identifie à $\lambda(\mathbb{R})$, est infinie. On peut en revanche écrire \mathbb{R} comme réunion des $] -n, n[$, qui sont tous de mesure 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 $\lambda^*(A + c)$ est égale à $\lambda^*(A)$. 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 λ^* 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 λ^* 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 λ^* , 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 choisit⁹ 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 \bar{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.

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 185), 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 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

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

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. \square

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 ($\bullet\bullet\bullet$))

La mesure de Lebesgue sur \mathbb{R}^d est dite *régulière* au sens où, pour tout A mesurable,

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

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

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

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 fermées de rayon j . La mesure de A est la limite des $\lambda(A_j)$ d'après la proposition B.3.6, page 202. 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 $+\infty$. \square

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

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

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}, μ) un espace mesuré, (X', \mathcal{A}') un espace mesurable¹⁰, et f une application mesurable de X vers X' . On définit l'application

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

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'un espace mesurable (X, \mathcal{A}) dans un espace mesuré (X', \mathcal{A}', ν) , 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).

(*) 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) (*) 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) (*) 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.

CORRECTION.

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

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

d'où la stabilité par passage au complémentaire. On a enfin, si (A'_n) est une collection d'éléments de \mathcal{A}' ,

$$f^{-1} \left(\bigcup_n A'_n \right) = \bigcup_n f^{-1}(A'_n) \in \mathcal{A}.$$

b) Si \mathcal{A}'' est une tribu sur X' qui rend f mesurable, pour tout A'' dans \mathcal{A}'' , on a $f^{-1}(A'') \in \mathcal{A}$, d'où $A'' \in \mathcal{A}'$.

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

On a donc $\mathcal{A}'' \subset \mathcal{A}'$.

c) On a $\nu(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset)$. Par ailleurs, pour toute collection (A'_n) d'éléments disjoints de \mathcal{A}' , on a

$$\nu\left(\bigcup_n A'_n\right) = \mu\left(f^{-1}\left(\bigcup_n A'_n\right)\right) = \sum_n \mu(f^{-1}(A'_n)),$$

car les $f^{-1}(A'_n)$ sont disjoints. Il s'agit donc bien d'une mesure qui fait de (X', \mathcal{A}', ν) un espace mesuré.

On a bien sûr conservation de la masse totale, du fait que $\nu(X') = \mu(f^{-1}(X')) = \mu(X)$.

Si la mesure μ est la loi de probabilité d'une variable aléatoire X , alors ν est simplement la loi de la variable aléatoire $f(X)$.

c bis) Remarquons en premier lieu, même si ça ne répond pas encore à la question, que s'il existe un singleton $\{x'\}$ dans \mathcal{A}' de mesure nulle, alors l'application constante $f(x) \equiv x'$ affecte une masse nulle à tout $A \in \mathcal{A}$. Il s'agit bien dans ce cas d'une mesure, mais elle est identiquement nulle, la masse n'est donc pas conservée. On se place maintenant pour simplifier dans le cas où $X' = X$, $\mathcal{A}' = \mathcal{A}$, et considérons le cas où il existe un singleton $\{x'\}$ de mesure non nulle. Alors l'application constante considérée précédemment, donne une même valeur à tous les éléments de \mathcal{A} , ce qui invalide l'additivité (sauf dans des situations très particulières), par exemple dès qu'il existe dans \mathcal{A} deux ensemble disjoints de mesures non nulles.

d) (i) $f(x) = x + c$. Comme vu dans le cours, les translations laissent invariante la mesure de Lebesgue, on a donc $\nu = \lambda$.

(ii) $f(x) = \alpha x$, avec $\alpha \neq 0$. Toute partie A mesurable est dilatée par l'homothétie de rapport α . La mesure d'un objet A dans l'espace d'arrivée est donc $\lambda(A)/\alpha$.

(iii) $f(x) = E(x)$. Cette application balaie en quelque sorte tous les intervalles $[n, n+1[$ et les concentre en $\{n\}$. La mesure image est donc la mesure de comptage sur \mathbb{N} :

$$\nu(A) = \text{Card}\{n \in \mathbb{N}, n \in A\} = \text{Card}(A \cap \mathbb{N}).$$

(iv) $f(x) = 0$. Cette application envoie toute la masse en 0. La mesure image est donc très singulière, il s'agit d'une masse ponctuelle (infinie) en 0 : $\nu(A) = 0$ si $0 \notin A$, $\nu(A) = +\infty$ si $0 \in A$,

(v) En premier lieu, notons que $\nu(A') = 0$ dès que $A' \cap]0, +\infty[= \emptyset$, on a donc

$$\nu(A') = \nu(A' \cap]0, +\infty[).$$

Pour tous a, b , avec $0 < a < b$, la mesure de $]a, b[$ est $\log(b) - \log(a) = \log(b/a)$. On a donc, pour tout $y > 0$, $dy > 0$,

$$\nu(]y, y+dy[) = \log\left(\frac{y+dy}{y}\right) \sim \frac{dy}{y}.$$

La mesure ν a donc sur $]0, +\infty[$ une densité par rapport à la mesure de Lebesgue égale à $1/y$.

e) Soit f et g sont égale μ – presque partout, alors $\mu(f^{-1}(A')) = \mu(g^{-1}(A'))$ pour tout $A' \in \mathcal{B}$.

f) Si μ est une masse ponctuelle, alors $f_\sharp\mu$ est aussi une masse ponctuelle. Si ν ne l'est pas, alors $\Lambda_{\mu, \nu} = \emptyset$.

On a un singleton quand les deux sont des masses ponctuelles. Si μ et ν sont toutes deux sommes de n masses ponctuelles de même poids,

$$\mu = \sum_{i=1}^n \delta_{x_i}, \quad \nu = \sum_{j=1}^n \delta_{y_j},$$

où les x_i sont distincts deux à deux, tout comme les y_j , alors $\Lambda_{\mu, \nu}$ contient $n!$ éléments. En effet, les éléments de $\Lambda_{\mu, \nu}$ sont du type $f(x_i) = y_{\varphi(i)}$, où φ est une bijection sur $\llbracket 1, N \rrbracket$ ($\varphi \in S_n$).

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 $\mu_0(A) = \text{Card}(A \cap \mathbb{Z})$ 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).

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

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

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 $A \in \mathcal{B}$.

CORRECTION.

a) On a bien $\mu_0(\emptyset) = 0$ et, pour toute collection disjointe de A_i mesurables, on a

$$\mu_0(\cup A_i) = \text{Card}((\cup A_i) \cap \mathbb{Z}) = \sum_i \text{Card}(A_i \cap \mathbb{Z}),$$

il s'agit donc bien d'une mesure, infinie car $\mathbb{Z} \subset \mathbb{R}$ est infini, mais σ -finie car \mathbb{R} s'écrit comme réunion des intervalles $[-n, n]$.

Les ensembles de mesure pleines pour μ_0 sont simplement les ensembles qui contiennent \mathbb{Z} .

Deux fonctions sont égales μ_0 -presque partout si et seulement si elles prennent les mêmes valeurs sur les entiers, ce qui laisse évidemment de la marge pour les choisir très différentes hors des entiers. Par exemple la fonction $\sin(\pi x)$ est nulle μ_0 -presque partout.

On peut construire une version finie de cette mesure en introduisant des poids, par exemple

$$\tilde{\mu}_0 = \sum_{k \in \mathbb{Z}} \frac{1}{2^{|k|}} \delta_k,$$

qui donne à \mathbb{R} (ou a n'importe quelle partie qui contient les entiers), une masse égale à 3.

b) la mesure μ_n est du même type que μ_0 , elle en a les mêmes propriétés. On n'a pas absolue continuité relativement à λ car les singletons $\{k/2^n\}$ sont de masse nulle pour λ , non nulle pour μ_n . Inversement les intervalles $]k/2^n, (k+1)/2^n[$ ont une masse non nulle pour λ , nulle pour μ_n .

Pour tout intervalle $]a, b[$ le nombre d'entiers de la forme $k/2^n$ qu'il contient est équivalent à $2^n(b-a)$, ce qui assure la convergence demandée.

c) Considérons par exemple l'ensemble A des nombres de l'intervalle $]0, 1[$ qui ne sont pas de la forme $k/2^n$. C'est ensemble est de mesure pleine sur $]0, 1[$, car on a enlevé un ensemble dénombrable, donc de mesure nulle. On a ainsi $\lambda(A) = 1$, et pourtant $\mu_n(A) = 0$ pour tout n .

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} .

CORRECTION.

a) On a

$$A \cup B = (A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B) \text{ (union disjointe)},$$

d'où

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

b) On écrit

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

On développe $\mu(A \cup B)$ d'après ce qui précède, et l'on écrit

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

c) On peut montrer par récurrence la formule générale

$$\begin{aligned} \mu\left(\bigcup_{n=A}^N A_n\right) &= \sum_{1 \leq n \leq N} \mu(A_n) - \sum_{1 \leq i_1 < i_2 \leq N} \mu(A_{i_1} \cap A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq N} \mu(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ &\quad + (-1)^N \mu(A_1 \cap A_2 \cap \cdots \cap A_N). \end{aligned}$$

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 $\lambda \ll \mu$).

CORRECTION.

a) Il s'agit d'une mesure de comptage, dont on vérifie immédiatement que c'est bien une mesure. Tout ouvert non vide contient un intervalle ouvert, donc une infinité de rationnels, sa mesure est donc infinie. Et la mesure de l'ouvert \emptyset est 0.

On note A_n l'ensemble des rationnels qui s'écrivent $\pm a/b$, avec a et b des entiers naturels entre 0 et n . on a $\mu(A_n \cap [-n, n]) < +\infty$, et l'union des $A_n \cap [-n, n]$ recouvre \mathbb{R} .

b) On a $\mu([0, 1] \cap \mathbb{Q}) = +\infty$ et $\lambda([0, 1] \cap \mathbb{Q}) = 0$ ce qui invalide $\mu \ll \lambda$, et dans l'autre sens $\mu([0, 1] \cap \mathbb{Q}^c) = 0$ et $\lambda([0, 1] \cap \mathbb{Q}^c) = 1$, ce qui invalide $\lambda \ll \mu$.

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

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

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

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

CORRECTION.

a) L'ensemble A est la collection des intervalles $[k \times 10^{-n+1}, (k+1/2) \times 10^{-n+1}]$. L'intersection d'un intervalle de longueur 10^{-n+1} avec A est donc de longueur totale $10^{-n+1}/2$, d'où le résultat pour les intervalles dont la longueur est un multiple entier de cette quantité.

Soit maintenant $I =]a, b[$ un intervalle. Pour tout N , on a

$$b - a = k_N \times 10^{-N} + \varepsilon_N$$

avec ($E(x)$ est la partie entière de x)

$$k_N = E((b - a) \times 10^N), \quad \varepsilon_N = b - a - k_N \times 10^{-N} \in [0, 10^{-N}].$$

l'intervalle s'écrit donc comme la réunion d'un intervalle dont la longueur est multiple de 10^{-N} , avec un intervalle de longueur $\varepsilon_N < 10^{-N}$. On a donc

$$\lambda(]a, b[\cap A_{N+1}) \geq \frac{b - a - \varepsilon_N}{2} \geq \frac{b - a}{2} - \frac{1}{2}10^{-N},$$

et

$$\lambda(]a, b[\cap A_{N+1}) \leq \frac{b - a - \varepsilon_N}{2} + \varepsilon_N \leq \frac{b - a}{2} + 10^{-N},$$

d'où la convergence de $\lambda(]a, b[\cap A_{N+1})$ vers $(b - a)/2$.

b) On se restreint à l'intervalle $]0, 1[$, en considérant l'ensemble $A' = A \cap]0, 1[$. D'après les hypothèses on a $\lambda(A') = 1/2$. L'ensemble A' étant mesurable, sa mesure s'identifie à sa mesure extérieure

$$\lambda(A') = \inf_{C_{A'}} \left(\sum_i (b_i - a_i) \right),$$

avec

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

Il existe donc une collection d'intervalles ouverts dont l'union contient A' telle que

$$\sum_i (b_i - a_i) \leq 3/4.$$

On note U la réunion ci-dessus. Comme $A' \subset U$, on a

$$\frac{1}{2} = \lambda(A') = \lambda(A' \cap U) \leq \sum_i \lambda(A' \cap]a_i, b_i[) = \frac{1}{2} \sum_n (b_i - a_i) = 3/8,$$

ce qui est absurde.

Exercice B.6.6. (Ensemble de Cantor)

On pose $K_0 = [0, 1]$, et l'on définit $K_1 = K_0 \setminus]1/3, 2/3[$, 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, \dots, 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 équivalent à \mathbb{R} .

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

CORRECTION.

a) L'ensemble K est une intersection de fermés (comme unions finies d'intervalles fermés), il s'agit donc d'un fermé, qui est borné par construction, donc d'un compact. Si K contient un intervalle ouvert, cet intervalle

est dans chacune des K_n , réunion d'intervalles de longueur $1/3^n$, sa longueur est donc inférieure à tout $1/3^n$, donc nécessairement nulle.

b) À toute suite $a = (a_n)_{n \geq 1}$ dans $\{0, 1\}$, on peut associer

$$x_a = \sum_{n=0}^{+\infty} 2 \frac{a_n}{3^n}.$$

Le réel x_a appartient à K . En effet, la série partielle définit x_a^n qui est l'extrémité gauche de l'un des intervalles de K_n . La suite x_a^n est donc dans K , et elle est de Cauchy par construction, elle converge donc dans \mathbb{R} , donc dans K car K est fermé. Cette application $\{0, 1\}^{\mathbb{N}^*} \rightarrow K$ est injective. Or l'ensemble de départ s'identifie à l'ensemble des parties de \mathbb{N}^* , qui est non dénombrable.

c) On a $\lambda(K_n) = 2^n/3^n$, et $K \subset K_n$ pour tout n , d'où $\lambda(K) \leq 2^n/3^n \rightarrow 0$.

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 195).

Dans le cas où l'espace d'arrivée est $\bar{\mathbb{R}} = [-\infty, \infty]$, 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 195).

On parlera donc simplement de fonction mesurable de (X, \mathcal{A}, μ) à valeurs dans \mathbb{R} ou dans $\bar{\mathbb{R}}$, 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 Lebesgue¹¹ (voir définition B.4.9, page 209).

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 $\bar{\mathbb{R}}$, 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 197, 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. \square

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

CORRECTION.

Considérons la fonction f par exemple croissante. Soit $c \in \mathbb{R}$. Pour tout x dans l'ensemble $f^{-1}(] -\infty, c])$, tout $z \leq x$ est dans ce même ensemble d'après la croissance de f . L'ensemble est donc du type $] -\infty, a]$ ou $] -\infty, a[$ (avec éventuellement $a = +\infty$), qui sont des boréliens dans les deux cas, donc a fortiori des membres de la tribu de Lebesgue. La démonstration est semblable pour une fonction décroissante (avec des intervalles de type $[a, +\infty[$ ou $]a, +\infty[$).

Proposition B.7.2. Soit (X, \mathcal{A}, μ) 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

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

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.

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

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

d'où

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

qui est mesurable comme intersection de mesurables.

On a par ailleurs

$$f_{\inf}(x) = \inf(f_n(x)) \leq c \iff \forall N, \exists n, f_n(x) \leq c + 1/N,$$

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

$$f_{\limsup}(x) = \limsup f_n(x) = \lim_{n \rightarrow +\infty} \sup_{k \geq n} f_k(x).$$

On introduit $g_n = \sup_{k \geq n} f_k$ 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 $h_n = \inf_{k \geq n} f_k$ 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. \square

Proposition B.7.3. Soit (X, \mathcal{A}, μ) 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 $\alpha > 0$

$$(\alpha f)^{-1}([-\infty, c]) = \{x, f(x) \leq c/\alpha\},$$

qui est dans \mathcal{A} . Si $\alpha < 0$, on a

$$\begin{aligned} (\alpha f)^{-1}([-\infty, c]) &= \{x, \alpha f(x) \leq c\} = \{f(x) \geq c/\alpha\} = \\ &\{x, f(x) < c/\alpha\}^c = \left(\bigcup_n \{x, f(x) \leq c/\alpha - 1/2^n\} \right)^c \end{aligned}$$

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

$$\{x, f(x) + g(x) \leq c\} = \bigcap_{n=0}^{+\infty} \{x, f(x) + g(x) < c + 1/2^n\}.$$

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

$$\{x, f(x) + g(x) \leq c\}$$

est donc mesurable comme intersection dénombrable d'ensembles mesurables. \square

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 \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). \square

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 $\alpha_1, \dots, \alpha_n$.

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

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

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 $\beta_0 \mathbf{1}_{A_0}$, avec $\beta_0 = 0$, et où A_0 est le complémentaire de l'union des A_i , de telle sorte que les A_0, \dots, A_N forment une partition de X . On fait de même pour g . La somme est une fonction étagée¹²

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

L'ensemble des fonctions étagées 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}, μ) 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

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

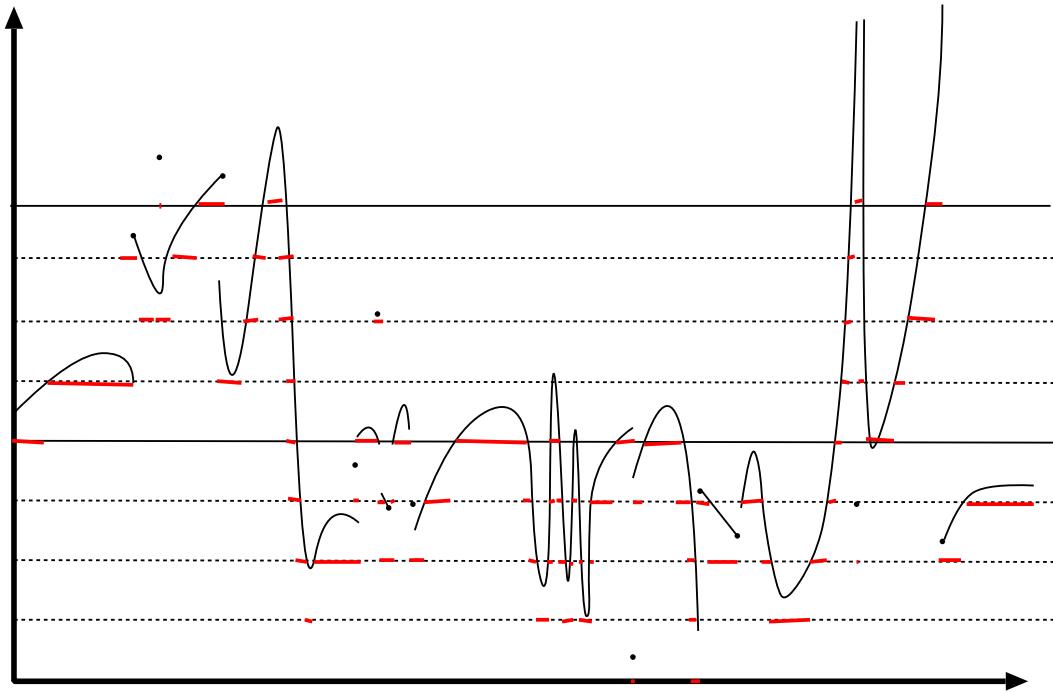


FIGURE B.7.1 – Approximation inférieure d'une fonction f (en noir) par une suite croissante de fonctions étagées.

plus grande. Pour tout entier $n \geq 1$, tout $k = 1, \dots, n2^n$, 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 . \square

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}, μ) 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 positifs¹³. On définit¹⁴ l'intégrale

13. Cette stricte positivité n'était pas imposée dans l'écriture B.7.1 pour anticiper la situation où la somme de deux fonctions 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.

de f sur X comme la quantité

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

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

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

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

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

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}, μ) un espace mesuré, f et g deux fonctions de \mathcal{E}^+ , et $\alpha \geq 0$. On a

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

et

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

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

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 $A_i \cap A'_j$,

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

des A_i , de telle sorte que les A_0, \dots, A_N forment une partition de X . On fait de même pour g . On a¹⁵

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

d'où

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

par additivité de la mesure. \square

Proposition B.7.11. (•) Soit (X, \mathcal{A}, μ) 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

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

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

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

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 $i = 1, \dots, N$, 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 202,

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

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. \square

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

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}, μ) 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}, μ) un espace mesuré, f et g deux fonctions mesurables de X dans $[0, +\infty]$. On a

$$f \leq g \implies \int_X f d\mu \leq \int_X g d\mu.$$

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. \square

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.

CORRECTION.

Soit g une fonction étagée positive

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

Si g est dominée par $I_{\mathbb{Q}}$, alors pour tout i tels que $\alpha_i > 0$, nécessairement $A_i \subset \mathbb{Q}$, d'où $\lambda(A_i) = 0$. On a donc $\int g = 0$ d'où, d'après la définition, $\int_{\mathbb{R}} \mathbf{1}_{\mathbb{Q}}(x) d\lambda = 0$.

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

Proposition B.7.14. Soit (X, \mathcal{A}, μ) 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

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

On a

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

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}^+ inférieur 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. \square

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}, μ) 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

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

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

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

Démonstration. On a de façon évidente

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

d'où l'on déduit que la limite de $\int f_n d\mu$ 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. \square

Proposition B.7.16. Soit (X, \mathcal{A}, μ) 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. \square

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}, μ) 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 ?

CORRECTION.

a) *On note*

$$u_n = \int_{n\pi}^{(n+1)\pi} \frac{\sin(x)}{x} dx.$$

La série $\sum u_n$ est alternée, avec

$$|u_n| \leq \frac{1}{n\pi} \int_0^\pi \sin(x) dx = \frac{2}{n\pi},$$

qui tend vers 0 quand n tend vers $+\infty$. La série est donc convergente, l'intégrale généralisée l'est donc aussi.

b) *On a, pour n pair,*

$$u_n \geq \frac{1}{(n+1)\pi} \int_0^\pi \sin(x) dx = \frac{2}{(n+1)\pi}.$$

L'intégrale de g^+ (qui existe bien) est donc infinie, la fonction n'est donc pas intégrable. L'intégrale de g^- est de la même manière infinie, on n'a donc pas existence de l'intégrale de g (comme on l'a dit, la notion de compensation est complètement étrangère au cadre de l'intégrale de Lebesgue).

Proposition B.7.19. Soit f une fonction mesurable de (X, \mathcal{A}, μ) 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}, μ) 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}, μ) 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 . \square

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

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

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

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

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

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}, μ) 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 :

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

Démonstration. La mesurabilité de f est une conséquence de la proposition B.7.2, page 219. On suppose dans un premier temps que les propriétés de monotonie et de convergence ponctuelle sont vérifiées pour tout x dans X . La monotonie de l'intégrale (proposition B.7.16) assure que

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

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 221). 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 222). 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 $f_n \mathbf{1}_A$ (qui met à 0 toutes les valeurs sur A^c) vérifie les hypothèses vis-à-vis de la fonction cible $f \mathbf{1}_A$. On a donc, d'après ce qui précède, convergence de la suite des intégrales vers l'intégrale de f . Or, comme $f_n \mathbf{1}_A$ s'identifie à f_n presque partout, de même pour $f \mathbf{1}_A$ et f , les intégrales sont les mêmes (d'après la proposition B.7.14), ce qui conclut la preuve. \square

Lemme B.7.24. (Fatou)

Soit (X, \mathcal{A}, μ) 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. \square

Théorème B.7.25. (Convergence dominée)

Soit (X, \mathcal{A}, μ) 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 nulle¹⁶, 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 \liminf en \limsup 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. \square

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.

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 $E^{x_2} = \{x_1 \in X_1, (x_1, x_2) \in E\}$.

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 $\mathcal{A}_1 \otimes \mathcal{A}_2$. 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 $x_1 \mapsto f^{x_2}(x_1) = f(x_1, x_2)$.

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. \square

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

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

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 $x_1 \mapsto \mu_2(E_{x_1})$ 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

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

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

$$\mu_2\left(\left(\bigcup E_n\right)_{x_1}\right) = \lim \mu_2((E_n)_{x_1}) = \sup \mu_2((E_n)_{x_1}),$$

qui est mesurable d'après la proposition B.7.2, page 219. 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 $x_1 \mapsto \mu_2(E_{x_1})$ est \mathcal{A}_1 – mesurable. Cette propriété est donc vraie pour tout $E \in \mathcal{A}$. On montre symétriquement que $x_2 \mapsto \mu_1(E^{x_2})$ 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 $x_1 \mapsto \mu_2^n(E_{x_1})$ est μ_1 – mesurable, d'où

$$x_1 \mapsto \mu_2(E_{x_1}) = \sum_{n=0}^{+\infty} \mu_2^n(E_{x_1})$$

est mesurable. On démontre de la même manière la propriété symétrique. \square

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

$$(\mu_1 \otimes \mu_2)_1(E) = \int_{X_1} \mu_2(E_{x_1}) d\mu_1(x_1), \quad (\mu_1 \otimes \mu_2)_2(E) = \int_{X_2} \mu_1(E^{x_2}) d\mu_2(x_2).$$

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 203, qui assure que ces mesures s'identifient sur la tribu engendrée par \mathcal{R} , qui est par définition la tribu-produit $\mathcal{A}_1 \otimes \mathcal{A}_2$. \square

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

$$F : (x_1, x_2) \mapsto (f_1(x_1), f_2(x_2))$$

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}, μ) vers $(X'_1, \mathcal{A}'_1, \mu'_1)$ et $(X'_2, \mathcal{A}'_2, \mu'_2)$, respectivement. Montrer que l'application

$$G : x \in X \mapsto (f_1(x), f_2(x))$$

est mesurable pour les tribus produits sur les espaces d'arrivée et de départ.

CORRECTION.

a) Pour tout rectangle $A_1 \times A_2 \in \mathcal{A}_1 \otimes \mathcal{A}_2$, on a

$$F^{-1}(A_1 \times A_2) = f_1^{-1}(A_1) \times f_2^{-1}(A_2) \in \mathcal{A}_1 \otimes \mathcal{A}_2,$$

d'où l'on déduit que F est mesurable d'après la proposition B.2.12, page 197, qui assure qu'il suffit de vérifier la mesurabilité sur un sous ensemble de parties de l'espace d'arrivée qui engendre la tribu sur cet espace.

b) Pour tout rectangle $A_1 \times A_2 \in \mathcal{A}_1 \otimes \mathcal{A}_2$, on a

$$F^{-1}(A_1 \times A_2) = f_1^{-1}(A_1) \cap f_2^{-1}(A_2) \in \mathcal{A},$$

d'où l'on déduit que G est mesurable, toujours d'après la proposition B.2.12.

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 :

$$f_{x_1}(x_2) = f(x_1, x_2) = \mathbf{1}_E(x_1, x_2) = \mathbf{1}_{E_{x_1}}(x_2), \quad f^{x_2}(x_1) = \mathbf{1}_{E^{x_2}}(x_1).$$

elles sont donc respectivement \mathcal{A}_2 – mesurable et \mathcal{A}_1 – mesurable d'après la proposition B.7.33, et l'on a

$$\int_{X_2} f_{x_1}(x_2) d\mu_2(x_2) = \mu_2(E_{x_1}) \text{ et } \int_{X_1} f^{x_2} d\mu_1(x_1) = \mu_1(E^{x_2}).$$

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 221). Pour toute fonction étagée g sur $\mathcal{A}_1 \otimes \mathcal{A}_2$, la section g_{x_1} est également étagée :

$$g(x_1, x_2) = \sum \alpha_i \mathbb{1}_{C_i}(x_1, x_2), \quad g_{x_1}(x_2) = \sum \alpha_i \mathbb{1}_{(C_i)_{x_1}}(x_2).$$

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 219, assure la mesurabilité des sections. Le théorème B.7.23 de convergence monotone assure la convergence des intégrales, ce qui conclut la preuve. \square

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

$$x_1 \in X_1 \longmapsto I_f^1(x_1) = \begin{cases} \int_{X_2} f_{x_1}(x_2) d\mu_2 & \text{si } f_{x_1} \text{ est } \mu_2\text{-intégrable} \\ 0 & \text{sinon} \end{cases}$$

et

$$x_2 \in X_2 \longmapsto I_f^2(x_2) = \begin{cases} \int_{X_1} f^{x_2}(x_1) d\mu_1 & \text{si } f^{x_2} \text{ est } \mu_1\text{-intégrable} \\ 0 & \text{sinon} \end{cases}$$

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

$$x_1 \longmapsto \int_{X_2} (f^+)_x_1 d\mu_2 \quad \text{et} \quad x_1 \longmapsto \int_{X_2} (f^-)_x_1 d\mu_2$$

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

$$N = \left(\bigcap_n \left\{ x_1, \int_{X_2} (f^+)_x d\mu_2 > n \right\} \right) \cup \left(\bigcap_n \left\{ x_1, \int_{X_2} (f^-)_x d\mu_2 > n \right\} \right).$$

la fonction I_f^1 vaut 0 sur N , et prend la valeur

$$\int_{X_2} (f^+)_x d\mu_2 - \int_{X_2} (f^-)_x d\mu_2$$

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 227,

$$\begin{aligned} \int_{X_1 \times X_2} f d(\mu_1 \otimes \mu_2) &= \int_{X_1 \times X_2} f^+ d(\mu_1 \otimes \mu_2) - \int_{X_1 \times X_2} f^- d(\mu_1 \otimes \mu_2) \\ &= \int_{X_1} \int_{X_2} (f^+)_x d\mu_2 - \int_{X_1} \int_{X_2} (f^-)_x d\mu_2 = \int_{X_1} I_f^1 d\mu_1. \end{aligned}$$

La même démarche appliquée aux sections $(f^+)_x$ et $(f^-)_x$ permet de conclure. \square

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.

CORRECTION.

La fonction f étant dérivable, elle est continue, et donc mesurable (proposition B.7.4, page 221). Pour tout $n \geq 1$, la fonction g_n définie par

$$g_n(x) = \frac{f(x + 1/n) - f(x)}{1/n}$$

est elle-même continue, donc mesurable. La fonction f' s'exprime

$$f'(x) = \lim_{n \rightarrow +\infty} g_n(x) = \limsup_{n \rightarrow +\infty} g_n(x).$$

L'application f' est donc mesurable d'après la proposition B.7.2, page 219.

Exercice B.8.2. Soit f une fonction mesurable de (X, \mathcal{A}, μ) dans \mathbb{R} , avec μ mesure finie. Pour tout $n \in \mathbb{N}$, on introduit

$$A_n = \{x \in X, |f(x)| \geq n\}, \quad B_n = \{x \in X, |f(x)| \in [n, n+1[\}.$$

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) \iff (ii) et (ii) \iff (iii).)

CORRECTION.

Les B_n constituant une partition de X , on a

$$\int_X |f| = \sum_n \int_{B_n} |f|,$$

avec, pour tout n

$$n\mu(B_n) \leq \int_{B_n} |f| \leq (n+1)\mu(B_n).$$

On a donc

$$\int_X |f| < +\infty \iff \sum_n \int_{B_n} |f| < \infty \iff \sum_n n\mu(B_n) < \infty.$$

Pour montrer $(ii) \iff (iii)$, on écrit

$$\mu(B_n) = \mu(A_n) - \mu(A_{n+1}).$$

En multipliant l'identité par n , et en sommant les termes de 1 à N , on obtient

$$\sum_1^N n\mu(B_n) = \sum_1^N \mu(A_n) - N\mu(A_{N+1}).$$

Si $\sum \mu(A_n)$ converge, alors $\sum n\mu(B_n)$ est bornée donc converge aussi. Réciproquement, si $\sum n\mu(B_n)$ converge, alors la fonction f est intégrable d'après ce qui précède, et on a (proposition B.7.21, page 227)

$$\mu(A_N) = \mu(\{x \in X, |f(x)| \geq N\}) \leq \frac{1}{N} \int |f|,$$

d'où l'on déduit que $N\mu(A_{N+1}) \leq (N+1)\mu(A_{N+1})$ est borné, d'où la convergence de la série des $\mu(A_n)$.

Exercice B.8.3. Soit (X, \mathcal{A}, μ) 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

$$\int_X f(x) d\mu \leq M.$$

CORRECTION.

Il s'agit d'une application du lemme de Fatou (lemme B.7.24, page 229), qui assure que

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

Comme on a convergence simple de f_n vers f , on a $\liminf_n f_n = f$, d'où le résultat.

Exercice B.8.4. Soit f une fonction mesurable de (X, \mathcal{A}, μ) 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

$$\lim_{n \rightarrow +\infty} \int_{A_n} |f(x)| d\mu = 0$$

b) Montrer que, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que

$$\forall A \in \mathcal{A} \text{ tel que } \mu(A) \leq \delta, \text{ on a } \int_A |f(x)| d\mu < \varepsilon.$$

(On pourra utiliser la décomposition d'une partie A en $A \cap A_n$ et $A \cap A_n^c$.)

CORRECTION.

a) On introduit la fonction $f_n = \mathbb{1}_{A_n^c} f$. La suite $|f_n|$ converge simplement vers $|f|$, avec $|f_n(x)| \leq |f(x)|$ pour tout x , et $|f|$ intégrable. D'après le théorème de convergence dominée (théorème B.7.25, page 229), on a convergence de l'intégrale $|f_n|$ vers l'intégrale de $|f|$, d'où

$$\int_{A_n} |f(x)| d\mu = \int_X |f(x)| d\mu - \int_X |f_n(x)| d\mu \longrightarrow 0.$$

b) D'après la question précédente, il existe n tel que

$$\int_{A_n} |f(x)| d\mu < \frac{\varepsilon}{2}.$$

Sur le complémentaire de A_n , on a $|f(x)| < n$. On prend $\delta = \varepsilon/(2n)$. Pour tout $A \in \mathcal{A}$ tel que $\mu(A) < \delta$, on a

$$\int_A |f(x)| d\mu = \int_{A \cap A_n} |f(x)| d\mu + \int_{A \cap A_n^c} |f(x)| d\mu < \frac{\varepsilon}{2} + n \frac{\varepsilon}{2n} = \varepsilon.$$

Exercice B.8.5. (Intégrale dépendant d'un paramètre)

Soit (X, \mathcal{A}, μ) 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 $x \mapsto f(x, t)$ est mesurable.
- (ii) Pour tout $x \in X$, la fonction $t \mapsto f(x, t)$ 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

$$F : t \in I \mapsto \int_X f(x, t) d\mu(x)$$

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 $t \mapsto f(x, t)$ 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

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

CORRECTION.

a) Le fait que $f(\cdot, t)$ soit mesurable, et la condition (ii) assurent que l'intégrale est bien définie. Il s'agit maintenant de montrer la continuité. Soit t_n une suite de réels qui tend vers t , on pose $f_n(x) = f(x, t_n)$. La continuité de f par rapport à t assure la convergence simple de $f(\cdot, t_n)$ vers $f(\cdot, t)$. La condition (iii) permet d'appliquer le théorème de convergence dominée, qui assure la convergence des intégrales :

$$\int_X f(x, t_n) d\mu(x) \longrightarrow \int_X f(x, t) d\mu(x).$$

b) Soit $t \in I$ et $\varepsilon > 0$ tel que $t \pm \varepsilon \in I$. On définit (la fonction Φ introduite ci-dessous est définie pour ce t particulier)

$$\Phi : (x, s) \in X \times]-\varepsilon, \varepsilon[\mapsto \Phi(x, s) = \begin{cases} \frac{f(x, t+s) - f(x, t)}{s} & \text{si } s \neq 0 \\ \frac{\partial f}{\partial t}(x, t) & \text{si } s = 0 \end{cases}$$

La fonction $\Phi(\cdot, s)$ est mesurable pour tout $s \neq 0$, comme $\Phi(x, 0)$ est la limite de $\Phi(x, s)$ quand s tend vers 0, la fonction $\Phi(\cdot, 0)$ est également mesurable. La fonction $s \mapsto \Phi(x, s)$ est continue pour tout x d'après l'hypothèse de différentiabilité de f . On a par ailleurs $|\Phi(x, s)| \leq h(x)$ d'après le théorème des accroissements finis. On peut donc appliquer la première question, qui assure

$$\frac{F(t+s) - F(t)}{s} = \int_X \Phi(x, s) d\mu(x) \longrightarrow \int_X \Phi(x, 0) d\mu(x) = \int_X \frac{\partial f}{\partial t}(x, t) d\mu(x).$$

qui exprime la dérivabilité de F .

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 $X_1 = X_2 = \llbracket 1, N \rrbracket$, 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 Π_{μ_1, μ_2} 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) (*) Imaginer une situation de la vie réelle (dans un contexte de logistique de transport), où les x_i correspondent 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) (*) 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.

CORRECTION.

a) On a

$$\mu_1(A_1) = \gamma(\pi_1^{-1}(A_1)) = \gamma(A_1 \times X_2),$$

et de même $\mu_2(A_2) = \gamma(X_1 \times A_2)$. On a en particulier $\mu_1(X_1) = \gamma(X_1 \times X_2) = \mu_2(X_2)$.
On n'a pas en général $\gamma = \mu_1 \otimes \mu_2$. Si l'on considère par exemple $X_1 = X_2 = \{0, 1\}$, et

$$\gamma(\{0, 0\}) = \gamma(\{1, 1\}) = 1/2, \quad \gamma(\{1, 0\}) = \gamma(\{0, 1\}) = 1/2.$$

On a $\mu_1(\{0\}) = \mu_1(\{1\}) = \mu_2(\{0\}) = \mu_2(\{1\}) = 1/2$, de telle sorte que $\gamma \neq \mu_1 \otimes \mu_2$.

Si γ est la loi d'une variable aléatoire $(Y_1, Y_2) \in X_1 \times X_2$, μ_i est simplement la loi de Y_i . La situation $\gamma = \mu_1 \otimes \mu_2$ correspond au cas de variables aléatoires indépendantes.

b) Dans le cas où $X_1 = X_2 = \llbracket 1, N \rrbracket$, les singletons sont des rectangles, et la mesure γ est entièrement déterminée par sa valeurs sur les singletons. La mesure s'identifie donc à une matrice $\gamma = (\gamma_{ij})$, avec $\gamma_{ij} = \gamma(\{(i, j)\})$. La première marginale μ_1 correspond à la somme des éléments des lignes :

$$(\mu_1)_i = \sum_{j=1}^N \gamma_{ij},$$

et μ_2 la somme des éléments des colonnes successives.

c) On peut toujours prendre pour γ la mesure produit $\mu_1 \otimes \mu_2$, qui correspondrait à des variables aléatoires indépendantes dans le cas de mesures de probabilité).

d) On peut identifier les plans de transport à des matrices $\gamma = (\gamma_{ij})$, et on a

$$\Pi_{\mu_1, \mu_2} = \left\{ \gamma = (\gamma_{ij}) \in \mathbb{R}_+^{n \times m}, \sum_{i=1}^n \gamma_{ij} = \mu_2^j, \sum_{j=1}^m \gamma_{ij} = \mu_1^i \right\}.$$

Il s'agit donc d'un ensemble de matrices dont la somme des éléments de chaque ligne, et de chaque colonne, est fixée. C'est un convexe (intersection d'un sous-espace vectoriel avec le convexe $\mathbb{R}_+^{n \times m}$). C'est un singleton dès que l'une des mesures (arrivée ou départ) est concentrée en un point unique. Dès qu'on a deux points chargés de part et d'autre, il existe plusieurs plans de transport.

e) L'application considérée est linéaire donc continue, elle est donc minorée sur le compact Π_{μ_1, μ_2} , et atteint sa borne inférieure en au moins un point. L'application n'est pas strictement convexe, il n'y a pas de raison que ce point soit unique. De fait il ne l'est pas en général, c'est clair dans des cas dégénérés (par exemple si tous les coûts sont égaux à une même valeur), mais aussi par exemple dans le cas de points (au moins 2 points à l'arrivée et au départ) appartenant à une même droite, avec $c_{ij} = \|y_j - x_i\|$. On pourra considérer par exemple (on se place sur la droite réelle) $x_1 = 0, x_2 = 1, y_1 = 1, y_2 = 2$, avec des masses égales. On peut laisser la masse en x_2 sur place (en y_1), et envoyer la masse en x_1 en y_2 , on translater de 1 la mesure. Noter que si le coût est quadratique en la distance, alors on a une solution unique (qui correspond à la translation).

f) Dans ce cas de figure, le problème consistant à trouver une solution du problème de minimisation revient à trouver un protocole d'envoi de la masse portée par les points de production vers les points de consommation de façon à minimiser le coût total de transport. Par exemple si c_{ij} correspond à la longueur du parcours routier entre x_i et y_j , cela revient à minimiser la longueur totale du parcours, et donc en première approximation l'essence consommée¹⁷.

g) On peut simplement définir l'opposé de la valeur ajoutée

$$A = - \sum_{ij} u_{ij} \gamma_{ij}.$$

Minimiser A revient à maximiser la valeur ajoutée, et le problème obtenu est du type des précédents, avec des coûts c_{ij} égaux aux opposés des productivités $i - j$.

Exercice B.8.7. Soit f une fonction d'un espace mesuré (X, \mathcal{A}, μ) 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.

17. Le problème réel est en général plus complexe, dans la mesure où le coût s'écrit, pour de petites masses à transporter, comme un coût fixe, dû au fait que l'on fait partir un camion. L'approche proposée correspond au cas où l'on a des grandes masses de produit à transporter, et donc un nombre de camions important à mobiliser (que l'on peut alors considérer comme une variable réelle).

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".

CORRECTION.

a) La fonction qui à (x, t) associe $f(x)$ est mesurable (car l'image réciproque de $]-\infty, c]$ est $f^{-1}(]-\infty, c]) \times \mathbb{R}_+$). La fonction qui à (x, t) associe t est mesurable, car l'image réciproque de $]-\infty, c]$ est $X \times [0, c]$ (qui est l'ensemble vide si $c < 0$).

L'application $F : (x, t) \mapsto f(x) - t$ est donc mesurable d'après la proposition B.7.3, page 220.

b) L'image réciproque de $]-\infty, c]$ par $\mathbb{1}_{\{(x,t), f(x) \geq t\}}$ est $X \times \mathbb{R}_+$ si $c \geq 1$, l'ensemble vide si $c < 0$ et, pour $c \in [0, 1[$, elle s'écrit

$$\{(x, t), t > f(x)\} = F^{-1}(]-\infty, 0[),$$

qui est mesurable d'après la mesurabilité de F (d'après la question précédente).

c) D'après ce qui précède, la fonction $\mathbb{1}_{\{(x,t), f(x) \geq t\}}$ rentre dans le cadre du théorème de Fubini-Tonelli (théorème B.7.35, page 232), on a donc

$$\begin{aligned} \int_0^{+\infty} \mu(\{x, f(x) \geq t\}) dt &= \int_0^{+\infty} \left(\int_X \mathbb{1}_{\{(y,s), f(y) \geq s\}}(x, t) d\mu \right) dt \\ &= \int_{X \times [0, +\infty[} \mathbb{1}_{\{(y,s), f(y) \geq s\}}(x, t) d\mu dt = \int_X \left(\int_0^{+\infty} \mathbb{1}_{\{(y,s), f(y) \geq s\}}(x, t) dt \right) d\mu, \\ &= \int_X f(x) d\mu, \end{aligned}$$

car $\int_0^{+\infty} \mathbb{1}_{\{(y,s), f(y) \geq s\}}(x, t) dt = f(x)$ pour tout x .

d) Le dessin est représenté sur la figure B.8.1. L'aire de la zone hachurée vaut

$$\lambda(\{x, f(x) \leq t\} \times \delta t),$$

et l'aire sous la courbe est approchée par la somme de ces aires.

Exercice B.8.8. Soit f une fonction intégrable de (X, \mathcal{A}, μ) 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

$$\mu(\{x, f(x) = t\}) = 0.$$

(On pourra s'inspirer de la démarche proposée à l'exercice B.8.7).

CORRECTION.

Comme dans l'exercice B.8.7, la fonction $(x, t) \mapsto f(x) - t$ est mesurable, on en déduit que $\mathbb{1}_{\{(x,t), f(x) \geq t\}}$ est également mesurable sur $X \times \mathbb{R}_+$, et elle est à valeurs dans \mathbb{R}_+ . Elle rentre donc dans le cadre du théorème de Fubini-Tonelli ((théorème B.7.35, page 232)), et l'on a

$$\begin{aligned} \int_{-\infty}^{+\infty} \mu(\{x, f(x) = t\}) dt &= \int_{-\infty}^{+\infty} \left(\int_X \mathbb{1}_{\{(y,s), f(y) = s\}}(x, t) d\mu \right) dt \\ &= \int_{X \times \mathbb{R}} \mathbb{1}_{\{(y,s), f(y) = s\}}(x, t) d\mu dt = \int_X \left(\int_{\mathbb{R}} \mathbb{1}_{\{(y,s), f(y) = s\}}(x, t) d\mu \right) dt \end{aligned}$$

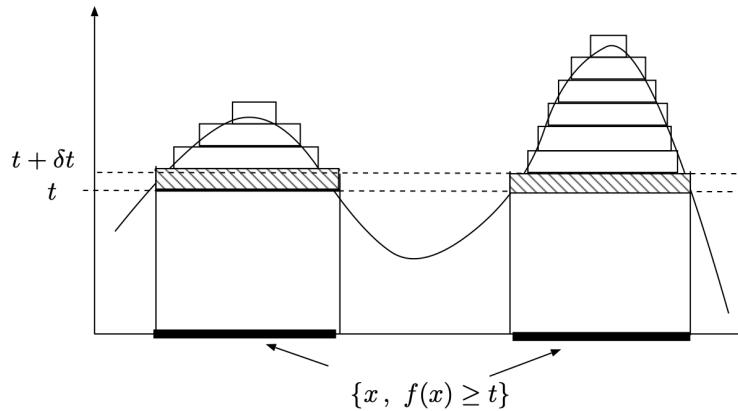


FIGURE B.8.1 – Quadrature par rectangles horizontaux

$$= \int_X \lambda(\{t, t = f(x)\}) d\mu(x),$$

or pour tout x l'ensemble $\{t, t = f(x)\}$ est le singleton $\{f(x)\}$, il est donc de mesure nulle. L'intégrale est donc nulle, d'où l'on déduit que la fonction positive $t \mapsto \mu(\{x, f(x) = t\})$ est d'intégrale nulle, elle est donc nulle presque partout (c'est une conséquence de la proposition B.7.21, page 227).

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 $\{(x, y), \varepsilon < \sqrt{x^2 + y^2} < 1\}$.)

CORRECTION.

a) On a

$$\int_{-1}^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} dx = \left[-\frac{x}{(x^2 + y^2)^2} \right]_{-1}^1 = -\frac{2}{1 + y^2}.$$

L'intégrale de cette fonction en y vaut

$$-2 [\arctan y]_{-1}^1 = -\pi$$

En intégrant dans l'autre sens (d'abord en y , puis en x , on obtient π).

b) On n'a pas de contradiction avec le théorème de Fubini-Lebesgue car la fonction n'est pas intégrable au voisinage de 0. On peut s'en convaincre en calculant l'intégrale sur la couronne $\{(x, y), \varepsilon < \sqrt{x^2 + y^2} < 1\}$. En passant en coordonnées polaires, on a

$$\int_{\{(x, y), \varepsilon < x^2 + y^2 < 1\}} |f(x, y)| dx dy = \int_{\varepsilon}^1 \int_{-\pi}^{+\pi} \frac{r^2 |\cos^2 \theta - \sin^2 \theta|}{r^4} r dr d\theta$$

$$= 8 \int_{\varepsilon}^1 \frac{1}{r} \int_0^{\pi/4} \cos(2\theta) dr d\theta = -4 \ln(\varepsilon),$$

qui tend vers $+\infty$ quand ε tend vers 0.

Index

- Équipotence, 172
- Équivalence
 - de normes, 31
- Équivalence (relation), 170
- Étagée (fonction), 221
- Liminf*, 181
- Limsup, 181
- Absolue continuité, 203
- Accroissements finis (théorème), 110
- Achevée (droite réelle), 20
- Adhérence, 18, 22
- Adjoint (méthode), 126
- adjoint (problème), 127
- Albedo, 131
- Anneau, 176
- Application
 - bijective, 170
 - continue, 27
 - contractante, 29
 - deux fois différentiable, 137
 - differentiable, 100
 - lipschitzienne, 29
 - mesurable, 197
 - surjective, 170
- Axiome du choix, 185
- Banach (espace de), 91
- Banach (espace), 47, 62
- Banach (théorème de point fixe, 29
- Bell (nombre de), 69
- Bijection, 170
- Borel – Lebesgue (propriété), 24, 188
- Borel – Lebesgue (théorème), 26
- Borne
 - inférieure, 171
 - supérieure, 171
- Borélienne (tribu), 195
- Bouts (espace des), 146
- Cantor
 - Processus d'extraction diagonale, 145
 - procédé d'extraction diagonale, 179
- Cartésien (produit), 170
- Cauchy (suite de), 22
- Cauchy (suite), 50
- Cellules de Voronoï, 36
- Champ
 - de vecteur, 98
 - scalaire, 98
- Charismatique (réseau), 158
- Classe
 - monotone, 198
- Clustering, 144
- Compacité, 24, 188
- Compacité
 - relative, 27
- Compact (espace métrique), 24
- Compact (espace topologique), 188
- Compact (opérateur), 86
- Complet (espace métrique), 23
- Complète (mesure), 202
- Complémentaire, 169
- Complété (d'un espace métrique), 185
- Composantes connexes, 187
- Conditions
 - aux limites, 19
 - d'interface, 19
- Conjugué (gradient), 107
- Connexe
 - composante, 187
 - définition (topologie générale), 187
- Connexité, 187
- Continuité, 27
- Continuité (topologie générale), 186
- Contractante (application), 29
- Convergence
 - simple, 46
 - uniforme, 46
- Convergence (d'une suite), 21
- Corps, 176
- Critique (point), 107
- Densité, 20
- Densité (d'une mesure relativement à une autre), 66
- Différence ensembliste, 170
- Différentielle
 - d'ordre 2, 137
 - de la composée de deux applications, 103
 - définition, 100
- Discret, 10
- Discret (ensemble), 12

- Discreté
 - distance, 15, 16
- Distance
 - de Hamming, 34
 - de hausdorff, 37
 - de Manhattan, 15
 - discrète, 15, 16
 - triviale, 15, 16
 - ultramétrique, 15, 34, 142
- Divergence, 115
- Droite réelle achevée, 20
- Dual (espace), 79
- Dual topologique, 92
- Ensemble
 - de mesure pleine, 203
 - de Vitali, 210
- Ensembles
 - équipotents, 172
- Espace
 - topologique, 186
 - $L^\infty(X)$, 61
 - $L^p(X)$, 62
 - complet, 23
 - de Banach, 47, 62, 91
 - de Hilbert, 3
 - des bouts, 146
 - dual topologique, 79
 - mesuré, 54, 201
 - métrique compact, 24
 - séparé, 186
 - topologique compact, 188
 - vectoriel, 177
 - vectoriel normé, 13
- Essentiel (sup), 65
- Extensive (variable), 14, 52
- Fatou (lemme), 229
- Fermé, 17
- Finesse (tribus), 194
- Fixe (point), 29
- Flot de gradient, 110
- Fonction
 - holomorphe, 112
 - indicatrice, 170
 - simple, 221
 - étagée, 221
- Fonctions implicites (théorème), 119
- France (tour de), 110
- Frontière, 18
- Gradient
 - conjugué (méthode), 107
- Graphé
 - d'une application, 170
- définition, 172
- non orienté, 172
- orienté, 172
- Grossière (topologie), 186
- Groupe, 174
- Groupe
 - isomorphisme, 175
 - libre, 175
 - morphisme, 175
 - symétrique, 178
- Hahn-Banach (théorème de), forme géométrique, cas hilbertien, 78
- Hahn-Banach (théorème), 92
- Hamming (distance de), 34
- Hausdorff (distance), 37
- Heine (théorème), 29
- Heine – Borel (théorème), 26
- Hilbert (espace), 3
- Hilbertienne (somme), 81
- histogramme, 223
- Holomorphe (fonction), 112
- Hölder (inégalité), 184
- Identité du parallélogramme , 76
- Image
 - mesure, 213
 - tribu, 213
- Image (mesure), 59
- Image (tribu), 196
- Image réciproque, 170
- Indicatrice (fonction), 170
- Infimum, 180
- Intensive (variable), 14
- Intersection, 169
- Intégrabilité, 56
- Intégrale
 - de Lebesgue, 227
 - de Riemann, 48, 109
- Intérieur, 18, 22
- Inversion locale (théorème), 123
- Inégalité
 - de Cauchy-Schwarz, 75
 - de Hölder, 184
 - de Minkovski, 184
- Isomorphisme de groupe, 175
- Jacobienne (matrice), 99
- Lagrangien, 126
- Landau (notation), 100
- landau (notation), 100
- Lax-Milgram (théorème de), 85
- Lebesgue, 3
- Lebesgue

- Points de , 66
- Lebesgue (théorème de différentiation), 67
- Lemme
 - de classe monotone, 199
- lemme de Fatou, 229
- Limite inférieure, 181
- Limite supérieure, 181
- Lipschitzienne (Application), 29
- Loi de composition
 - externe, 174
 - interne, 173
- Magma, 174
- Majorant, 171
- Manhattan (distance, 15)
- Marquée (Subdivision), 49
- Matrice
 - jacobienne, 99
 - symétrique définie positive, 112
- Mesurable (application), 197
- mesure, 30
- Mesure
 - complète, 202
 - σ – finie, 54, 201
 - définition, 200
 - extérieure, 204
 - finie, 201
 - image, 59, 213
 - pleine (pour les ensembles), 203
 - produit, 231
- Mesuré (espace), 201
- Minkovski (inégalité), 184
- Minorant, 171
- Monotone (classe), 198
- Monotone (théorème de convergence - , 228)
- Monotonie (mesure), 54, 201
- Monoïde, 174
- Méthode
 - de l'état adjoint, 126
- Nombre
 - Bell, 69
- Nombres
 - décimaux, 179
 - réels (construction), 179
- Norme
 - $\|\cdot\|_p$, 13
 - définition, 13
 - subordonnée, 32
- Norme équivalente, 31
- Négligeable (partie), 54, 202
- Opérateur
 - compact, 86
 - de rang fini, 86
- Ordre
 - partiel, 171
 - relation, 171
 - total, 171
- Ouvert, 17
- Parallélogramme (identité), 76
- Partie
 - entière, 179
 - dense, 20
 - négligeable, 202
- Partie génératrice, 175
- Partition, 170
- Point
 - critique, 107
 - d'équilibre, 132
 - stationnaire, 107
- Point fixe d'une application, 29
- Poupée russe, 142
- Productivité, 237
- Produit
 - cartésien (de deux ensembles), 170
- Produit (mesure), 231
- Projection, 77
- Propriété de Borel – Lebesgue, 24, 188
- Précompacte (partie), 27
- Radon-Nikodym (théorème), 66
- Rang fini (opérateur), 86
- Relation, 170
- Relation
 - d'ordre, 171
 - d'équivalence, 170
- Riemann (intégrale), 48
- Riemann (somme), 48
- Riesz-Fréchet (théorème de représentation de), 79
- Russe (poupée), 142
- Réiproque (image), 170
- Régularité (d'une mesure), 211
- Réseau
 - charismatique, 158
- Schwarz (théorème), 134
- Section, 230
- Simple (convergence), 46
- Simple (fonction), 221
- Somme
 - de Darboux, 49, 67
 - de Riemann, 48
 - hilbertienne, 81
- Sous-groupe, 175
- Subdivision marquée, 49
- Subordonnée (norme), 32
- Suite
 - convergente, 21

- de Cauchy, 22
- maximisante, 180
- minimisante, 180
- Suite (de Cauchy), 50
- Supremum, 180
- Supremum
 - essentiel, 61, 65
- Surjection, 170
- Surjection
 - canonique, 171
 - définition, 170
- Séparable
 - espace de Hilbert, 76
 - espace métrique, 20
- Séparé
 - espace topologique, 186
- Théorème
 - d'inversion locale, 123
 - de Banach (point fixe), 29
 - de convergence dominée, 229
 - de convergence monotone, 228
 - de différentiation de Lebesgue, 67
 - de Hahn-Banach géométrique (cas hilbertien), 78
 - de Hahn-Banach, forme analytique, 92
 - de Hahn-Banach, forme géométrique, 93
 - de Heine, 29
 - de Heine – Borel ou Borel – Lebesgue, 26
 - de Lax-Milgram, 85
 - de Radon-Nikodym, 66
 - de représentation de Riesz-Fréchet, 79
 - de Schwarz, 134
 - des accroissements finis, 110
 - des fonctions implicites, 119
- Topologie
 - discrète, 186
 - grossière, 186
 - générale, 186
- Totallement discontinu, 187
- Tour (de France), 110
- Tribu, 3, 52
- Tribu
 - borélienne, 53, 195
 - de Lebesgue, 209
 - définition, 193
 - engendrée par un ensemble de parties, 195
 - image, 196, 213
 - produit, 230
- Triviale
 - distance, 15, 16
- Troncature entière, 179
- Ultramétrique (distance), 15, 142
- Uniforme (convergence), 46
- Union, 170
- Valuation 2-adique, 141
- Variable
 - extensive, 14, 52
 - intensive, 14
- Vitali (ensemble), 210
- Voronoi (cellules), 36