

REMISE À NIVEAU EN ANALYSE
MASTER MATHÉMATIQUES POUR LES
SCIENCES DU VIVANT

Pierre J. A. Roux

Sommaire

I. Espaces de Hilbert	11
1. Rappels généraux	11
a. Espaces métriques	12
b. Espaces vectoriels normés	19
c. Espaces de Lebesgue	21
d. Produit scalaire	24
2. Structure hilbertienne	25
a. Théorèmes généraux dans le cadre hilbertien	26
b. Bases Hilbertiennes	31
3. Synthèse et pratique	34
a. Synthèse de la section	34
b. Exercices	34
II. Théorie spectrale	37
1. Spectre d'un opérateur	39
2. Opérateur compact	41
3. Adjoint d'un opérateur	42
4. Synthèse et pratique	44
a. Synthèse de la section	44
b. Exercices	44
III. Régularité et espaces de Sobolev	47
1. Rappels sur des outils usuels	48
a. Multi-indices	48
b. Fonctions tests et fonctions plateaux	48
c. Ouverts à bords réguliers	49
d. Mesure superficielle et formule de Green	50
2. Produit de convolution et régularisation	53
3. Transformée de Fourier	55
4. Espaces de Sobolev	58
a. Dérivées faibles	58
b. Les espaces de Sobolev	60
c. Injections et inégalités	61
d. Traces et formule de Green	63
5. Valeurs propres du laplacien	64

6. Synthèse et pratique	66
a. Synthèse de la section	66
b. Exercices	67
IV. Équations différentielles ordinaires	69
1. Cylindres de sécurité, maximalité des solutions	69
2. Existence d'une solution	71
3. Résultats généraux	75
a. Unicité, finitude et longévité de l'existence	75
4. Principes généraux de l'analyse qualitative	80
a. Description qualitative de la biologie	81
b. Flot d'un champ de vecteur	87
c. Comportement asymptotique	89
5. Synthèse et pratique	94
a. Synthèse de la section	94
b. Exercices théoriques et de modélisation	94
V. Équations aux dérivées partielles	97
1. Les équations de transport	98
a. Transport à vitesse constante	98
b. Transport à vitesse non constante	99
c. Solutions faibles pour des données irrégulières	100
2. L'équation de la chaleur	103
a. Origine physique de l'équation	103
b. Étude mathématique	105
c. Simulation numérique par les différences finies	107
3. L'équation de Laplace	113
a. Solutions classiques	114
b. Simulation numérique par les éléments finis	117
4. Modélisation plus avancée : la chimiotaxie	122
5. Synthèse et pratique	129
a. Synthèse de la section	129
b. Exercices théoriques et de modélisation	130
Annexes	130
A. Compléments en analyse fonctionnelle et en géométrie	131
1. Notions de topologie générale	131
2. Quelques résultats importants	135
a. Prolongement d'applications	135
b. Conséquences de la propriété de Baire	135
3. Dualité	137
a. Topologie faible	138
b. Topologie faible*	139
4. Topologie des espaces de fonctions tests	141
a. Les espaces de fonctions différentiables	141

b. Les espaces de fonctions différentiables à support compact	142
5. Éléments de géométrie différentielle	143
a. Les variétés différentielles	143
b. Espaces tangents	146
c. Fibrations	148
6. Dimension fractale	149
7. Synthèse et pratique	151
a. Synthèse de la section	151
b. Exercices	151
B. Introduction à la théorie des distributions	153
1. Distributions de Schwartz	153
2. Extension de la notion de convolution	155
a. Convolution d'une fonction avec une distribution	155
b. Convolution de deux distributions	156
3. Transformée de Fourier	157
a. Distributions tempérées	158
b. Transformée de Fourier d'une distribution	158
4. Masse de Dirac et solutions fondamentales	160
C. Les semi-groupes d'opérateurs	163
1. Semi-groupes et générateur infinitésimal	163
2. Le théorème de Hille-Yosida	164
3. Applications	164
D. Approche générale des systèmes dynamiques	165
1. Systèmes continus sur une variété différentielle	165
a. Champs de vecteurs	165
b. Flot d'un champ de vecteurs	166
2. Systèmes dynamiques discrets	168
3. Systèmes topologiquement transitifs	169
4. Moyenne temporelle d'un système dynamique	170
a. Cas d'un système périodique	170
b. Cas général	175
Bibliographie	177

Introduction

"J'ai profondément regretté de n'avoir pas été assez loin pour au moins comprendre un petit peu des grands principes fondamentaux des Mathématiques : car les hommes qui les ont acquis ont alors acquis un sens supplémentaire - un sixième sens..."

Charles Darwin, *Autobiographie*

CE POLYCOPIE EST ENCORE UNE VERSION TRES PRELIMINAIRE EN PLEINS TRAVAUX ET EST VOUE A EVOLUER. LES EXERCICES SONT A PRENDRE AVEC PRECAUTIONS, ILS SONT MAL CALIBRES, PEU TESTES ET PEU NOMBREUX.

Ce document est conçu pour servir de support à un cours de remise à niveau en analyse et analyse fonctionnelle dans le cadre du master Mathématiques pour les sciences du vivant à l'université Paris-Sud d'Orsay, membre de l'université Paris-Saclay.

Rien dans ce document n'est une production originale de la part de l'auteur. Ce dernier n'a fait que condenser et organiser les savoirs qui lui ont été enseignés par d'excellents professeurs au sein de l'université Paris-Saclay (en particulier à l'université Paris-sud d'Orsay) ou qu'il a sauvagement plagié dans les ouvrages de référence cités dans la bibliographie.

L'objectif est de donner aussi rapidement que possible les bases nécessaires à la compréhension d'un cours de master recherche en équations différentielles ou en équations aux dérivées partielles. L'esprit du cours est de commencer par revoir les résultats généraux sur les espaces de Hilbert (partie **I.**) et leurs opérateurs linéaires (partie **II.**), ainsi que les outils principaux de l'analyse (convolution, Fourier, espaces de Sobolev,...) (partie **III.**). Cela étant fait, nous aborderons les bases de la théorie des équations différentielles ordinaires (partie **IV.**), puis nous étudierons plusieurs familles d'équations aux dérivées partielles à travers des exemples concrets (partie **V.**).

Les quatre annexes sont destinées à des publics maîtrisant déjà largement le texte principal (LES ANNEXES SONT ENCORE EN PLEIN CHANTIER ET

COMPORTENT DE NOMBREUSES ERREURS, MANQUEMENTS, IMPRECISIONS,...). Il est parfaitement possible de réussir le master mathématiques pour les sciences du vivant sans avoir connaissance de leur contenu et nous ne les aborderons pas en cours magistral (sauf demande explicite), mais elle peuvent être profitables pour comprendre certaines thématiques de recherche et donnent un éclairage intéressant sur de nombreux résultats (on pensera au lien qu'entretiennent les semi-groupes d'opérateurs avec les processus stochastiques). L'annexe A. présente des bases techniques sur la topologie générale, l'analyse fonctionnelle, la dualité, l'intégration en dimension infinie, et les variétés différentielles. L'annexe B. est une introduction à la théorie des distributions et à ses applications en équations aux dérivées partielles. L'annexe C. pose les fondements de la théorie des semi-groupes d'opérateurs et illustre son intérêt pour résoudre des équations aux dérivées partielles. Enfin, l'annexe D. est une introduction à une étude plus large des systèmes dynamiques discrets et continus.

Certaines preuves sont faites, d'autres pas. Il n'y avait ni la place ni le temps de tout faire et des choix étaient nécessaires. Lorsqu'une preuve n'est pas faite, cela ne signifie pas qu'elle est nécessairement difficile ou longue. Certaines preuves non détaillées peuvent éventuellement être très longues et difficiles.

Ce document contient des notes historiques, des éléments biographiques et diverses digressions autour des mathématiques, de la physique et de la biologie. Elles sont aussi importantes que les définitions et les théorèmes pour quiconque se destine à faire des mathématiques appliquées car elle permettent d'avoir du recul sur les liens qu'entretiennent les objets *via* leur élaboration lente et circonstanciée à travers les siècles. Trop de mathématiciens sont formés dans une logique purement technique, prétendument faute du temps et des moyens adéquats. Le temps consacré à ces aspects périthématiques est loin d'être perdu, que ce soit pour l'ingénieur, le mathématicien ou l'individu. Puissent les convaincus et les inconvertibles excuser cette prêche intempestive.

Il est absolument certain que malgré les efforts de l'auteur de nombreuses fautes mathématiques, historiques, syntaxiques, grammaticales et (surtout) orthographiques se sont glissées çà et là. Nous nous en excusons patement et invitons les lectrices¹ à signaler à l'auteur toute erreur découverte par mail à l'adresse pierre.roux@math.u-psud.fr ou par l'usage d'ondes acoustiques émises par des cordes vocales si une proximité spatiale, fortuite ou contrainte, le permet.

1. Suivant l'initiative de Michèle Audin, mathématicienne spécialiste de géométrie symplectique, écrivaine et membre de l'Oulipo, nous utiliserons parfois le féminin pour le lecteur de sexe indéterminé. Que ces messieurs ne se privent pas de faire remonter les erreurs à l'auteur par crainte d'être mal reçus.

Nous tenons à remercier chaleureusement pour leur contribution à la relecture et à l'écriture de ce polycopié quelques personnes : Céline Bonnet, François Delgove et tout particulièrement Juliette Chevallier et Léo Zaradzki.

Notations

Afin d'éviter toute ambiguïté, nous précisons ici le sens de certaines notations utilisées dans ce document. Ce paragraphe peut être sauté en première lecture et consulté chaque fois qu'advient un doute. D'autres notations seront précisées au fur et à mesure de l'introduction des concepts.

Nous noterons $\mathbb{N}$ l'ensemble des entiers naturels, $\mathbb{Z}$ l'ensemble des entiers relatifs, $\mathbb{Q}$ le corps des rationnels, $\mathbb{R}$ le corps des réels et $\mathbb{C}$ le corps des complexes. On notera également $\bar{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$, $\bar{\mathbb{Z}} = \{-\infty\} \cup \mathbb{Z} \cup \{+\infty\}$, $\bar{\mathbb{Q}} = \{-\infty\} \cup \mathbb{Q} \cup \{+\infty\}$ et surtout $\bar{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ la droite réelle achevée. On écrira $\mathbb{R}_+^* =]0, +\infty[$ et $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$; on prendra garde de ne pas confondre cette notation avec celle des espaces duaux introduite par la suite.

Si K est un corps (souvent $\mathbb{R}$ ou $\mathbb{C}$ dans ce document), n, m des entiers naturels non nuls, E et F des espaces vectoriels sur K , nous noterons $\mathcal{M}_n(K)$ (respectivement $\mathcal{M}_{n,m}(K)$) l'ensemble des matrices de taille $n \times n$ (respectivement $n \times m$) à coefficients dans K , $\mathcal{L}(E)$ l'ensemble des endomorphismes continus de E et $\mathcal{L}(E, F)$ l'ensemble des morphismes d'espaces vectoriels continus de E dans F , ces quatre ensembles étant vus comme munis de leur structure naturelle d'algèbre. On notera $GL(E)$ l'ensemble des automorphismes de E et $GL_n(K)$ l'ensemble des matrices de taille $n \times n$ inversibles à coefficients dans K , tous deux munis de leur structure naturelle de groupe. Pour des raisons de lisibilité, on notera parfois MN le produit matriciel des matrices M et N . Pour une matrice M de taille $n \times n$, on notera M^T la transposée de M et $Tr(M)$ la trace de M . L'opérateur trace sera indifféremment utilisé sur les endomorphismes.

Les produit scalaires d'éléments x et y seront notés $\langle x | y \rangle$ ou $x \cdot y$ selon le besoin ou l'humeur de l'auteur.

Pour une fonction $f : \mathbb{R} \rightarrow E$, E espace vectoriel normé, dérivable, nous écrirons f' ou $\frac{df}{dt}$ la fonction dérivée de f . Pour E, F des espaces vectoriels normés et $f : E \rightarrow F$ une fonction différentiable, nous noterons $d_x f : E \rightarrow F$ la différentielle de f au point x et, pour $h \in \mathbb{R}^n$, $d_x f(h)$ ou $d_x f \cdot h$ la différentielle de f au point x appliquée au vecteur h . En dimension finie, nous écrirons $\nabla f(x)$ le gradient de f en x . Rappelons que, pour f différentiable sur l'ouvert U de $\mathbb{R}^n$,

$$\forall x \in U, \forall h \in \mathbb{R}^n, d_x f(h) = \nabla f(x) \cdot h.$$

Les différentielles secondes, troisièmes, n -ièmes de f au point x seront notées respectivement $d_x^2 f, d_x^3 f, d_x^n f$. Les dérivées partielles de f par rapport à la variable x_i ou par rapport aux variables $x_{i_1}, \dots, x_{i_k}$ s'écriront $\partial_{x_i} f$ et $\partial_{x_{i_1}, \dots, x_{i_k}} f$.

ou $\frac{\partial f}{\partial x_i}$ et $\frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}$. On notera $\mathcal{C}^k(E, F)$ l'algèbre des fonctions de classe $\mathcal{C}^k$ de E dans F et $\mathcal{C}^\infty(E, F)$ l'algèbre des fonctions de classe $\mathcal{C}^k$ pour tout $k \in \mathbb{N}$.

Nous utiliserons les opérateurs différentiels classiques laplacien, divergent et rotationnel dont nous noterons l'application à une fonction f adéquate Δf , $\nabla \cdot f$ et $\nabla \wedge f$ respectivement.

S'il y a lieu, nous écrirons $f * g$ le produit de convolution des fonctions f et g et $\hat{f}$ la transformée de Fourier de f . On notera $\mathcal{S}(\mathbb{R}^n)$ l'espace de Schwartz. L'ensemble des fonctions de l'ouvert Ω de $\mathbb{R}^n$ à valeurs réelles (ou complexes si précisé) de classe $\mathcal{C}^k$, $k \in \bar{\mathbb{N}}$ à support compact sera écrit $\mathcal{C}_0^k(\Omega)$. Nous écrirons parfois $\mathcal{D}(\Omega) = \mathcal{C}_0^\infty(\Omega)$.

I. ESPACES DE HILBERT

« [...] parce que, [les Anciens] s'étant élevés jusqu'à un certain degré où ils nous ont portés, le moindre effort nous fait monter plus haut, et avec moins de peine et moins de gloire nous nous trouvons au-dessus d'eux. C'est de là que nous pouvons découvrir des choses qu'il leur était impossible d'apercevoir. Notre vue a plus d'étendue, et, quoiqu'ils connussent aussi bien que nous tout ce qu'ils pouvaient remarquer de la nature, ils n'en connaissaient pas tant néanmoins, et nous voyons plus qu'eux. »

Blaise Pascal, *Préface au traité du vide* (1647)

Comme Zack Weinersmith le dessine sur son site internet <http://www.smbc-comics.com/comic/how-math-works>, cette citation de Pascal revient à dire "Comment pouvez-vous ne pas comprendre ce concept ? Nous en avons parlé pendant au moins une heure hier !". La théorie "élémentaire" des espaces de Hilbert a mis des siècles à voir le jour, bien qu'elle fût en germe dans les travaux de nombreux "grands" mathématiciens (Fourier, Cauchy, Poincaré...), et plusieurs décennies à prendre sa forme finale (Hilbert, Schmidt, Riesz, Banach, Steinhaus, Fréchet, von Neumann...). Sa formulation actuelle a l'avantage d'être tout à la fois simple et puissante. Les espaces de Hilbert permettent dans une large mesure de travailler avec des fonctions comme si elles étaient des points dans l'espace et par conséquent de montrer des résultats très généraux à partir d'intuitions géométriques (projection, représentation, orthogonalité, théorème de Pythagore, base d'un espace,...).

Nous allons dans cette partie rappeler quelques définitions relatives aux espaces métriques, de Banach et de Hilbert ainsi que quelques théorèmes fondateurs. Ces résultats nous seront utiles dans toutes les sections suivantes. Le contenu qui est présenté ici est tiré des livres [8], [5], du cours (et du polycopié associé) de Frédéric Paulin à l'université Paris-Sud et du site internet Wikipédia.

1. Rappels généraux

Avant d'étudier les espaces fonctionnels qui paveront notre chemin au royaume des équations différentielles, il convient d'être au point sur les différentes structures de travail que l'on peut mettre sur les ensembles : métrique, norme, pro-

duit scalaire. Tous ces espaces s'inscrivent plus généralement dans le cadre de la topologie (en grec $\eta \tau ο π ο λ ο γ ι α$, l'étude du lieu).

1.a. Espaces métriques

Distances sur un ensembles :

Définition 1 (Distance). *Une distance (ou métrique) sur un ensemble E est une application positive*

$$d : E \times E \rightarrow \mathbb{R}_+,$$

vérifiant

- $\forall x, y \in E, d(x, y) = d(y, x)$; (symétrie)
- $\forall x, y \in E, \left(d(x, y) = 0 \iff x = y \right)$; (séparation)
- $\forall x, y, z \in E, d(x, z) \leq d(x, y) + d(y, z)$. (inégalité triangulaire)

Définition 2 (Espace métrique). *On appelle espace métrique un ensemble muni d'une distance.*

De nombreux espaces de fonctions possèdent des structures métriques utiles. La structure d'espace métrique est très générale et certains exemple vont contre l'intuition de ce que peut être un espace qui "autorise les mesures". Il y a des monstres comme la distance triviale qui vérifie $d(x, x) = 0$ pour tout $x \in E$ et $d(x, y) = 1$ si $x \neq y$, mais aussi des distances à la fois étranges et utiles. On pensera par exemple aux distances ultramétriques.

Définition 3 (Espace ultramétrique). *Un espace métrique (E, d) est dit ultramétrique si sa distance vérifie*

$$\forall x, y, z \in E, d(x, z) \leq \max \left(d(x, y), d(y, z) \right).$$

On remarquera que grâce à la positivité de la distance, la propriété nécessaire pour être ultramétrique est une version plus exigeante de l'inégalité triangulaire. La distance triviale décrite ci-dessus est un cas particulier de distance ultramétrique.

Les distances ultramétriques vérifient des propriétés curieuses. Par exemple dans une boule ultramétrique, tous les points sont au centre. Si deux boules ont une intersection non vide, l'une est incluse dans l'autre. Enfin, tous les triangles sont isocèles dans un espace ultramétrique. Pour des énoncés plus détaillés, on pourra regarder les question de l'exercice 67. Elles peuvent être utiles pour l'études d'objets ayant une structure d'arbre binaire, comme les bronches des mammifères.

Outils de topologie métrique : Afin de pouvoir travailler dans les espaces métriques et plus tard dans les espaces vectoriels normés, de Banach et de Hilbert, il nous faut des outils élémentaires. Nous allons présenter au mieux notre caisse à outils.

Pour commencer, nous voulons définir les voisinages des points d'un espace métrique de manière générale.

Définition 4 (Boule). *Soit (E, d) un espace métrique. Soit $x_0 \in E$ et $r \in \mathbb{R}_+$. On appelle boule ouverte centrée en x_0 et de rayon r , et on note $B(x_0, r)$ l'ensemble*

$$B(x_0, r) = \{x \in E \mid d(x_0, x) < r\}.$$

On appelle boule fermée de centre x_0 et de rayon r , et on note $B^f(x_0, r)$ l'ensemble

$$B^f(x_0, r) = \{x \in E \mid d(x_0, x) \leq r\}.$$

Ces objets de bases permettent de repérer des parties particulières d'un espace métrique. Elles donnent aussi l'occasion de définir proprement la convergence dans ce cadre.

Définition 5 (Convergence). *Soit $(x_n)_{n \in \mathbb{N}}$ une suite à valeur dans un espace métrique (E, d) . On dit que la suite converge vers l'élément $x \in E$ si*

$$\forall \varepsilon \in \mathbb{R}_+^*, \exists n_0 \in \mathbb{N}, \forall n \in \mathbb{N}, n \geq n_0 \implies d(x_n, x) < \varepsilon.$$

La même définition fonctionne pour les applications et permet de définir la continuité.

Définition 6 (Continuité). *Soit (E, d_E) et (F, d_F) deux espaces métriques. On dit que l'application $f : E \rightarrow F$ converge vers $y_F \in F$ quand x tend vers y_E si*

$$\forall \varepsilon \in \mathbb{R}_+^*, \exists \eta \in \mathbb{R}_+^*, \forall x \in E, d(x, y_E) < \eta \implies d(f(x), y_F) < \varepsilon.$$

L'application f est continue en a si elle converge vers $f(a)$ quand x tend vers a . Elle est dite continue sur E ou simplement continue si elle est continue en tout point de E .

Définition 7 (Ouvert, Fermé). *Soit (E, d) un espace métrique. Une partie $O \subset E$ est dite ouverte si pour tout $x \in O$, il existe un rayon strictement positif $r \in \mathbb{R}_+^*$ tel que*

$$B(x, r) \subset O.$$

Une partie $F \subset E$ est dite fermée si pour toute suite $(x_n)_{n \in \mathbb{N}}$ à valeur dans F qui converge vers $x \in E$, x est un élément de F .

Dans une partie ouverte, il y a de l'espace autour de chaque point. Les parties fermées contiennent toutes les limites de leurs suites convergentes. Dans l'espace métrique $(\mathbb{R}, |\cdot|)$, pour $a, b \in \mathbb{R}$, les intervalles ouverts $]a, b[$ sont ouverts, les intervalles fermés $[a, b]$ sont fermés et les intervalles $]a, b]$ ne sont ni ouverts, ni

fermés. Tout espace métrique E , considéré comme une partie de lui-même, est ouvert et fermé à la fois. L'ensemble vide $\emptyset$, considéré comme une partie de E , est ouvert et fermé à la fois.

Propriété 8 (No shit, Sherlock!). *Dans un espace métrique, toute boule ouverte est un ouvert et toute boule fermée est un fermé.*

Démonstration : Soit $B(x, r)$ une boule ouverte. Soit $y \in B(x, r)$, on a

$$B(y, \frac{r - d(x, y)}{2}) \subset B(x, r).$$

Donc $B(x, r)$ est un ouvert.

Soit $B^f(x, r)$ une boule fermée. Soit $(y_n)_{n \in \mathbb{N}}$ une suite à valeurs dans cette boule fermée qui converge vers y . Si $y \notin B^f(x, r)$, alors $d(x, y) > r$. Donc on a

$$B(y, \frac{d(x, y) - r}{2}) \cap B^f(x, r) = \emptyset.$$

Cela contredit la définition de la limite. ■

Théorème 9 (*Intersection, union*)

Dans un espace métrique, l'intersection d'un ensemble **quelconque** de fermés est un fermé. L'union d'un nombre **fini** de fermés est un fermé, l'intersection d'un nombre **fini** d'ouverts est un ouvert, l'union d'un ensemble **quelconque** d'ouverts est un ouvert.

Les ouverts et les fermés sont liés de façon cruciale à la continuité (comme on le verra de plus près dans l'annexe A.).

Théorème 10

Une application $f : E \rightarrow F$ entre deux espaces métriques E et F est continue si et seulement si pour tout ouvert de F l'image réciproque de cet ouvert par f est un ouvert de E .

Remarque 11. *Le théorème fonctionne aussi avec fermé à la place d'ouvert. Cela nous donne un moyen simple de montrer qu'une partie est ouverte ou fermée, en l'écrivant comme image réciproque d'un ouvert ou d'un fermé par une application continue.*

Certaines parties ne sont ni ouvertes ni fermées, mais il est possible de trouver un ouvert maximal dans chaque partie et un fermé minimal autour de chaque partie, comme l'illustre la définition suivante.

Définition 12 (Intérieur, Adhérence). *Soit (E, d) un espace métrique. Soit une partie $A \subset E$. On appelle adhérence (ou fermeture) de A et on note $\overline{A}$ le plus petit fermé qui contient A , c'est-à-dire*

$$\overline{A} = \bigcap_{F \text{ fermé, } A \subset F} F.$$

On appelle intérieur de A et on note A° le plus grand ouvert contenu dans A , c'est-à-dire

$$A^\circ = \bigcup_{O \text{ ouvert, } O \subset A} O.$$

Remarque 13. On peut, de manière équivalente, définir l'adhérence de A comme l'ensemble de toutes les limites des suites convergentes de A .

En effet nous obtiendrons alors par construction une partie fermée qui contient A et nous avons le fermé et on vérifie facilement que c'est le plus petit.

Exemple 14. Dans l'espace métrique $(\mathbb{R}, |\cdot|)$, la partie $\mathbb{Q}$ contenant les nombres rationnels vérifie $\overline{\mathbb{Q}} = \mathbb{R}$ et $\mathbb{Q}^\circ = \emptyset$; la partie $]a, b[$, $a, b \in \mathbb{R}$, vérifie $\overline{]a, b[} = [a, b]$ et $]a, b[^\circ =]a, b[$. Tout fermé est sa propre adhérence, tout ouvert est son propre intérieur.

Définition 15 (Densité). Soit (E, d) un espace métrique. Une partie A est dite dense dans E si $\overline{A} = E$.

Remarque 16. De manière équivalente, d'après la remarque 13 une partie A d'un espace métrique E est dense si pour tout élément $x \in E$, il existe une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de A qui converge vers x .

Définition 17 (Connexité). Soit (E, d) un espace métrique. Il est dit connexe si les seuls ensembles à la fois ouverts et fermés sont E et $\emptyset$.

Remarque 18. Même si ce n'est pas évident d'un premier abord, cette définition revient à dire que E est "en un seul morceau". En effet, si $E = [0, 1] \cup [2, 3]$ muni de la distance usuelle de $\mathbb{R}$ (la valeur absolue), la partie $[0, 1]$ est fermée (évident), mais aussi ouverte (c'est contre-intuitif mais c'est vrai). Cela tient au fait que la boule ouverte de E centrée en 1 et de rayon $1/2$ est en fait l'ensemble $[\frac{1}{2}, 1]$ (vérifier avec la définition). L'espace E est en deux morceaux, et chacun de ces morceaux est ouvert et fermé à la fois. Chacun de ces morceaux est par contre lui-même connexe, d'un seul tenant².

Complétude :

Définition 19 (Suite de Cauchy). Soit (E, d) un espace métrique. On appelle suite de Cauchy de E toute suite $(x_n)_{n \in \mathbb{N}}$ telle que

$$\forall \varepsilon \in \mathbb{R}_+^*, \exists n_0 \in \mathbb{N}, \forall p, q \in \mathbb{N}, (p, q > n_0 \implies d(x_p, x_q) < \varepsilon).$$

2. Cette vision abstraite n'est que la face immergée d'un iceberg cauchemardesque. On peut définir plein d'autres manières d'être un brave ensemble bien solide : la connexité par arc, la locale connexité, la locale connexité par arc, la simple connexité, la simple connexité semi-locale par arc... Et, évidemment, le diable étant dans les détails (le diable aime les maths, en témoigne l'escalier du diable de Cantor), il existe des espaces connexes qui ne sont pas connexes par arc, et même des espaces connexes et connexes par arc qui ne sont ni localement connexes, ni localement connexes par arc. Voilà, j'aurais dû fabriquer de la méta-amphétamine comme dans Breaking bad, crouler sous la thune, me demander qu'est-ce j'veis faire de tout cet oseille... (en prévision des lettres de menace, je sais qu'oseille est féminin ; je retranscris simplement la faute de la chanson de Booba dont je fais l'érudite paraphrase).

Définition 20 (Complétude). *Un espace métrique est dit complet si toute suite de Cauchy converge vers un élément de cet espace.*

Le cadre des espaces métriques permet l'énoncé d'un résultat très général qui appartient à la famille des théorèmes de point fixe.

Théorème 21 (*Théorème du point fixe de Picard*)

Soit (E, d) un espace métrique complet non vide et $f : E \rightarrow E$ une application. Si f a une itérée contractante, au sens où il existe $k \in]0, 1[$ et $q \in \mathbb{N}$ tels que

$$\forall x, y \in E, \quad d(f^q(x), f^q(y)) \leq kd(x, y),$$

où l'on a noté $f^q = f \circ \dots \circ f$ avec q compositions, alors f admet un unique point fixe dans E , au sens où il existe un unique point $x^* \in E$ tel que $f(x^*) = x^*$. De plus, toute suite vérifiant la relation de récurrence $x_{n+1} = f(x_n)$ converge vers x^* .

Ce théorème est parfois aussi appelé théorème du point fixe de Banach.

Compacité : Une autre notion importante dans les espaces métriques est la compacité. Un espace métrique compact est comme un intervalle $[a, b]$, fermé et borné dans $\mathbb{R}$: tout s'y passe plutôt bien et il est plus facile d'y travailler que dans un domaine ouvert et/ou infini sous certains aspects.

Définition 22 (Compacité). *Un espace métrique (E, d) est dite compact si pour toute suite $(x_n)_{n \in \mathbb{N}}$ à valeurs dans E , il existe une suite extraite $(x_{\sigma(n)})_{n \in \mathbb{N}}$ qui converge vers un élément $x \in E$ pour la distance d . On dit qu'une partie F de E est compacte l'espace métrique $(F, d|_F)$ est compact, où $d|_F$ est la restriction à F de la distance.*

Exemple 23. *L'ensemble des nombres réels $\mathbb{R}$ muni de la distance*

$$d : (x, y) \mapsto |\arctan(x) - \arctan(y)|,$$

est un espace borné.

Mieux encore, l'ensemble $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ muni de cette même distance et des conventions

$$\arctan(+\infty) = \frac{\pi}{2}, \quad \arctan(-\infty) = -\frac{\pi}{2},$$

est un espace métrique compact. On l'appelle droite réelle achevée. Par contre, l'addition n'est plus définie correctement dans $\bar{\mathbb{R}}$ car on ne peut pas sommer $-\infty$ et $+\infty$.

Théorème 24

Soit K un espace métrique compact non vide et $f : K \rightarrow \mathbb{R}$ continue

sur K . Alors f admet un minimum et un maximum sur K .

Démonstration : Supposons que la fonction f n'admette pas de minimum sur K .

- Si $f(K)$ n'est pas minorée dans $\mathbb{R}$, il existe une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de K telle que $\lim_{n \rightarrow +\infty} f(x_n) = -\infty$. Or K est compact, il existe donc une suite $(x_{\sigma(n)})_{n \in \mathbb{N}}$ extraite de $(x_n)_{n \in \mathbb{N}}$ qui converge vers un élément $x \in K$. La fonction f est continue, donc par caractérisation séquentielle de la continuité $\lim_{n \rightarrow +\infty} f(x_{\sigma(n)}) = f(x)$. Or, $(f(x_{\sigma(n)}))_{n \in \mathbb{N}}$ étant extraite de $(f(x_n))_{n \in \mathbb{N}}$, elle converge également vers $-\infty$. Contradiction.

- Si $f(K)$ est minorée dans $\mathbb{R}$, $f(K)$ admet une borne inférieure y^* dans $\mathbb{R}$. Il existe donc une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de K telle que $(f(x_n))_{n \in \mathbb{N}}$ converge vers y^* .

L'ensemble K est compact, il existe donc une suite $(x_{\sigma(n)})_{n \in \mathbb{N}}$ extraite de $(x_n)_{n \in \mathbb{N}}$ qui converge vers un élément $x^* \in K$.

De plus, la fonction f est continue, donc $\lim_{n \rightarrow +\infty} f(x_{\sigma(n)}) = f(x^*)$. Or, $(f(x_{\sigma(n)}))_{n \in \mathbb{N}}$ étant extraite de $(f(x_n))_{n \in \mathbb{N}}$, elle converge également vers y^* . Donc $y^* = f(x^*)$ et par suite $y^* \in f(K)$. Comme y^* minore $f(K)$ dans $\mathbb{R}$, y^* est le minimum de $f(K)$, donc le minimum de f sur K . Contradiction.

Donc f admet un minimum sur K . La démonstration pour le maximum se fait de la même manière. ■

La partie la plus importante du théorème est que ce sont des maxima et minima, ils sont par définition atteints. Nous pouvions montrer le théorème avec moins d'efforts en disant que l'image d'un compact par une application continue est un compact dans le cadre plus abstrait de la topologie générale mais, outre le fait que cela escamotait les arguments en s'appuyant sur une autre théorie (voir Annexe A.), cela nous privait surtout du plaisir enfantin d'extraire des suites.

La vision équivalente plus topologique de la compacité est d'une grande utilité dans de nombreux cas, en particulier pour passer du local au global sur de nombreuses propriétés. Le théorème suivant exprime cette notion.

Théorème 25 (Propriété de Borel-Lebesgue)

Un espace métrique (E, d) est compact si et seulement si pour tout recouvrement ouvert de E (c'est-à-dire une famille $(O_i)_{i \in I}$ d'ouverts telle que $E \subset \cup_{i \in I} O_i$) il existe un sous-recouvrement fini de E (Une partie finie $O_{i_1}, \dots, O_{i_n}$ du recouvrement précédent telle que $E \subset \cup_{k=1}^n O_{i_k}$).

Démonstration : • Supposons la propriété du théorème de Borel-Lebesgue pour E . Soit $(x_n)_{n \in \mathbb{N}}$ une suite de E . Si elle prend un nombre fini de valeurs, alors elle admet une suite extraite constante donc convergente. Sinon, supposons par l'absurde qu'elle prend un nombre infini de valeurs différentes et n'admet pas de suite extraite convergente.

Notons $A = \{x_n \mid n \in \mathbb{N}\}$. Notre supposition implique que pour tout $x \in E$, il existe $r_x \in \mathbb{R}_+^*$ tel que l'ensemble $B(x, r_x) \cap A$ est fini.

La famille $(B(x, r_x))_{x \in E}$ est un recouvrement ouvert de E , on peut donc en extraire un sous-recouvrement ouvert fini $B(y_1, r_1), \dots, B(y_l, r_l)$. L'ensemble

$$\bigcup_{k=1}^l (B(y_k, r_k) \cap A)$$

est fini comme union finie d'ensembles finis, mais on a aussi

$$A = \bigcup_{k=1}^l (B(y_k, r_k) \cap A),$$

ce qui constitue une contradiction puisque A est un ensemble infini.

Donc $(x_n)_{n \in \mathbb{N}}$ admet une suite extraite convergente. Nous avons montré que E est compact.

• Supposons que (E, d) est compact. Soit $(O_i)_{i \in I}$ un recouvrement ouvert de E . Nous allons trouver un nombre fini de ces ouverts qui recouvrent l'espace métrique E en utilisant l'hypothèse de compacité sur une suite adaptée et le fait que voici :

$$\exists r \in \mathbb{R}_+^*, \forall x \in E, \exists i \in I, B(x, r) \subset O_i. \quad (1)$$

Supposons que c'est vrai, nous le démontrerons à la fin ; cela revient à dire que pour r fixé assez petit, toute boule de rayon r dans E est contenue dans l'un des ouverts du recouvrement. Fixons le réel strictement positif r qui vérifie cette propriété.

Pour tout élément $x_0 \in E$, il existe un ouvert O_{i_0} dans le recouvrement tel que $x_0 \in O_{i_0}$ et $B(x_0, r) \subset O_{i_0}$. Si $E \subset O_{i_0}$, la preuve est terminée. Sinon, il existe $x_1 \in E$ tel que $x_1 \notin O_{i_0}$. Il existe alors dans le recouvrement un deuxième ouvert O_{i_1} tel que $x_1 \in O_{i_1}$ et $B(x_1, r) \subset O_{i_1}$.

On choisit, tant que c'est possible, des éléments $x_2, \dots, x_n, \dots$ et des ouverts $O_{i_2}, \dots, O_{i_n}, \dots$ tels que

$$x_n \in O_{i_n} \quad \text{et} \quad B(x_n, r) \subset O_{i_n} \quad \text{et} \quad x_n \notin \bigcup_{k=0}^{n-1} O_{i_k}.^3$$

Si on ne peut répéter cette opération qu'un nombre n fini de fois, alors par construction

$$E \subset \bigcup_{k=0}^n O_{i_k},$$

et le résultat est démontré. Sinon, on obtient par ce procédé une suite $(x_n)_{n \in \mathbb{N}}$. Par compacité, elle admet une sous-suite extraite $(x_{\sigma(n)})_{n \in \mathbb{N}}$ qui converge vers un élément x^* .

En particulier,

$$\lim_{n \rightarrow +\infty} d(x_{\sigma(n+1)}, x_{\sigma(n)}) = 0.$$

3. I can see you coming in my nightmares, infuriated crowd of logicians, with pitchforks and torches, willing to burn me for my pleonasms. My work shall survive your narrow-mindedness.

Par conséquent, il existe $n_0 \in \mathbb{N}$ tel que $d(x_{\sigma(n_0+1)}, x_{\sigma(n_0)}) < r$. On a donc $x_{\sigma(n_0+1)} \in B(x_{\sigma(n_0)}, r)$, et par suite $x_{\sigma(n_0+1)} \in O_{i_{\sigma(n_0)}}$, ce qui constitue une contradiction.

Donc on ne pouvait répéter l'opération qu'un nombre fini de fois et il existe un sous-recouvrement fini de E . Il ne reste plus qu'à prouver (1). Supposons que ce résultat est faux, c'est-à-dire

$$\forall r \in \mathbb{R}_+, \exists x \in E, \forall i \in I, B(x, r) \not\subset O_i,$$

où le symbole $\not\subset$ signifie "n'est pas inclus dans". Donc,

$$\forall n \in \mathbb{N}^*, \exists x_n \in E, \forall i \in I, B(x_n, \frac{1}{n}) \not\subset O_i.$$

Par compacité, la suite $(x_n)_{n \in \mathbb{N}}$ admet une sous-suite $(x_{\sigma(n)})_{n \in \mathbb{N}}$ qui converge vers $x^* \in E$. Par définition d'un recouvrement, il existe $i^* \in I$ tel que $x^* \in O_{i^*}$ et par définition d'un ouvert, il existe $\varepsilon \in \mathbb{R}_+$ tel que $B(x^*, \varepsilon) \subset O_{i^*}$.

Par définition de la convergence on peut trouver n_0 tel que

$$\frac{1}{\sigma(n_0)} < \frac{\varepsilon}{3} \quad \text{et} \quad d(x_{\sigma(n_0)}, x^*) < \frac{\varepsilon}{3}.$$

Il vient,

$$B(x_{\sigma(n_0)}, \frac{1}{\sigma(n_0)}) \subset B(x^*, \varepsilon) \subset O_{i^*},$$

ce qui constitue une contradiction et achève la preuve du théorème. ■

1.b. Espaces vectoriels normés

Définition 26 (Norme). Soit E un espace vectoriel sur $\mathbb{K}$. Une norme $\|\cdot\|$ sur E est une application

$$\|\cdot\| : E \rightarrow \mathbb{R}_+$$

vérifiant

- $\forall x \in E, \forall \lambda \in \mathbb{K}, \|\lambda x\| = |\lambda| \|x\|$; (absolue homogénéité)
- $\forall x, y \in E, \|x + y\| \leq \|x\| + \|y\|$; (inégalité triangulaire)
- $\forall x \in E, (\|x\| = 0 \iff x = 0)$. (séparation)

Définition 27 (Espace vectoriel normé). On appelle espace vectoriel normé un espace vectoriel muni d'une norme.

Remarque 28. Tout espace vectoriel normé possède une structure naturelle d'espace métrique avec la distance

$$(x, y) \mapsto \|x - y\|.$$

Toutes les distances sur des espaces vectoriels ne sont par contre pas associées à des normes, comme le montre l'exemple de la distance arctan sur $\mathbb{R}$ que nous avons vu dans la sous-section précédente.

Toutes les notions que nous avons vues pour les espaces métriques et tous les

théorèmes qui leurs sont consacrés sont valables dans un espace vectoriel normé : boules, limites, continuité, ouverts, fermés, adhérence, intérieur, densité, théorème du point fixe de Picard, suites de Cauchy, complétude, compacité, propriété de Borel-Lebesgue,...

Définition 29 (Normes équivalentes). Deux normes $\|\cdot\|_1$ et $\|\cdot\|_2$ sont dites équivalentes s'il existe deux constantes $k_1, k_2 \in \mathbb{R}_+^*$ telles que

$$\forall x \in E, \quad k_1 \|x\|_1 \leq \|x\|_2 \leq k_2 \|x\|_1.$$

Deux normes équivalentes définissent la même topologie sur l'espace E au sens où elles définissent les mêmes ouverts et les mêmes fermés. Cela revient aussi à dire que l'application identité entre $(E, \|\cdot\|_1)$ et $(E, \|\cdot\|_2)$ est un isomorphisme d'espaces vectoriels normés, c'est-à-dire une bijection continue et linéaire entre deux espaces vectoriels normés dont la bijection réciproque est elle aussi continue.

Propriété 30. Soit E un espace vectoriel de dimension finie. Toutes les normes sur E sont équivalentes.

Ce n'est pas le cas en dimension infinie et le choix des normes devient alors très important. Certaines normes sont plus précises que d'autres. On dit qu'elles sont plus fines :

Définition 31 (Norme plus fine). Soit E un espace vectoriel muni de deux normes $\|\cdot\|_1$ et $\|\cdot\|_2$. On dit que $\|\cdot\|_2$ est plus fine que $\|\cdot\|_1$ s'il existe une constante $k \in \mathbb{R}_+^*$ telle que

$$\|x\|_1 \leq k \|x\|_2.$$

Cela équivaut à dire que l'application $id_E : (E, \|\cdot\|_2) \rightarrow (E, \|\cdot\|_1)$ est continue (attention à l'ordre des numéros). Intuitivement, les boules sont plus petites pour $\|\cdot\|_2$. Donc $\|\cdot\|_2$ sépare mieux les points de l'espace.

Exemple 32. Soit $E = \mathcal{C}([0, 1])$ l'ensemble des fonctions continues de $[0, 1]$ dans $\mathbb{R}$. Soient les normes

$$\|\cdot\|_1 : f \mapsto \int_0^1 f(x) dx, \quad \text{et} \quad \|\cdot\|_\infty : f \mapsto \sup_{x \in [0, 1]} |f(x)|,$$

sur E . Elles ne sont pas équivalentes, car on peut trouver (facilement) une suite de fonctions qui converge vers 0 pour l'une et pas pour l'autre. La norme $\|\cdot\|_\infty$ est plus fine que la norme $\|\cdot\|_1$ car

$$\forall f \in \mathcal{C}([0, 1]), \quad \|f\|_1 \leq \|f\|_\infty.$$

Nous insistons ici sur l'équivalent des notions de complétude et de compacité vues pour les espaces métriques.

Définition 33 (Suite de Cauchy). Soit E un espace vectoriel normé de norme $\|\cdot\|$. Une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de E est dite de Cauchy si

$$\forall \varepsilon \in \mathbb{R}_+^*, \exists n_0 \in \mathbb{N}, \forall p, q \in \mathbb{N}, \left(p, q \geq n_0 \implies \|x_p - x_q\| \leq \varepsilon \right).$$

Définition 34 (Espace de Banach). *On appelle espace de Banach un espace vectoriel normé complet, c'est-à-dire dans lequel toute suite de Cauchy est convergente.*

Exemple 35. *L'espace vectoriel $\mathbb{R}$ muni de la norme valeur absolue $|\cdot|$ est un espace de Banach.*

Définition 36 (Compact). *Une partie A d'un espace vectoriel normé E est dite compacte si pour toute suite $(x_n)_{n \in \mathbb{N}}$ à valeurs dans A , il existe une suite extraite $(x_{k_n})_{n \in \mathbb{N}}$ qui converge vers un élément $x \in A$.*

Remarque 37 (IMPORTANT). *Toute partie compacte d'un espace vectoriel normé est complète (Les suites de Cauchy à valeurs dans la partie convergent et leurs limites sont dans la partie). Dans un espace vectoriel de dimension finie, une partie est compacte si et seulement si elle est fermée et bornée. Les espaces vectoriels de dimension finie sont complets pour n'importe quelle norme. Certains espaces de dimension infinie très usuels ne sont pas complets ; par exemple l'ensemble des polynômes à coefficients réels $\mathbb{R}[X]$ pour la norme $\|P\| = \sup_{x \in [0,1]} |P(x)|$.*

On remarquera aussi qu'un espace vectoriel normé tout entier, non réduit à $\{0\}$, ne sera jamais compact parce qu'il n'est pas borné alors qu'un espace métrique tout entier peut l'être (même $\mathbb{R}$ peut-être borné pour certaines distances, comme nous l'avons vu avec la distance arctan). Il n'est pas vrai en dimension infinie qu'une partie fermée et bornée est compacte. En fait, les compacts des espaces vectoriels normés de dimension infinie sont "maigres". Ils sont d'intérieur vide et leur cardinal est au plus celui de $\mathbb{R}$. On a en particulier le critère suivant :

Théorème 38 (Théorème de Riesz, 1918)

Soit E un espace vectoriel normé. La boule unité fermée de E est compacte si et seulement si E est de dimension finie.

1.c. Espaces de Lebesgue

Dans cette sous-section nous notons $\mathbb{K}$ pour $\mathbb{R}$ ou $\mathbb{C}$. Les espaces de Lebesgue sont parmi les espaces vectoriels normés les plus utilisés en analyse. Ils sont fondés sur la théorie de la mesure et la théorie de l'intégration de Lebesgue que nous n'avons pas le temps de rappeler dans ces notes⁴. Disons simplement que sur l'ensemble $\mathbb{R}^n$ muni de sa tribu borélienne $\mathcal{B}$ (la plus petite tribu qui contient tous les ouverts), il existe une unique mesure σ —finie, invariante par toutes les isométries et de valeur 1 sur tout cube engendré par des vecteurs d'une base orthonormée. C'est la mesure de Lebesgue sur $\mathbb{R}^n$, que nous noterons λ . Cette accumulation de titres revenait juste à dire qu'il n'y a qu'une mesure naturelle

4. Est-ce le temps de l'auteur ou celui des lecteurs ? Les deux mon capitaine !

sur la tribu borélienne au sens où elle mesure les pavés au sens intuitif et que si on bouge le pavé sans le déformer, cela ne change pas sa taille⁵.

Évidemment, la mesure de Lebesgue ne peut mesurer que les éléments de $\mathcal{B}$ et l'intégrale qui lui est associée ne permet d'intégrer que les fonctions à la fois mesurables pour la tribu borélienne et intégrables. On rappelle qu'un ensemble $B \in \mathcal{B}$ de la tribu borélienne est dit négligeable si $\lambda(B) = 0$. Une propriété est dite vraie presque partout si elle est vérifiée en dehors d'un ensemble (de la tribu borélienne) de mesure nulle.

La mesure de Lebesgue et la notion d'intégrale peuvent être restreintes à tout ouvert de $\mathbb{R}^n$, mais nous allons définir les objets dans $\mathbb{R}^n$ entier pour plus de clarté.

On peut donc parler pour tout $p \in [1, +\infty[$ de l'espace des fonctions mesurables et intégrables à la puissance p , ou encore

$$\mathcal{L}^p(\mathbb{R}^n) = \mathcal{L}^p(\mathbb{R}^n, \mathcal{B}, \lambda) = \{f : \mathbb{R}^n \rightarrow \mathbb{K} \mid f \text{ mesurable et } \int_{\mathbb{R}^n} |f(x)|^p dx < +\infty\}.$$

L'ensemble $\mathcal{L}^p(\mathbb{R}^n)$ est un espace vectoriel que l'on peut munir de la semi-norme (que nous notons volontairement comme une norme)

$$\|f\|_p = \left(\int_{\mathbb{R}^n} |f(x)|^p dx \right)^{\frac{1}{p}}.$$

Elle vérifie l'inégalité triangulaire (ce n'est pas immédiat) et l'homogénéité (facile) mais n'est pas une norme car une fonction égale à 0 presque partout (par exemple partout sauf en un point) a une intégrale nulle en valeur absolue à la puissance p , ce qui contredit la séparation.

Nous souhaitons donc "réparer" l'espace $\mathcal{L}^p(\mathbb{R}^n)$. Il suffit pour cela d'éliminer brutalement le problème en identifiant les fonctions qui sont non nulles mais d'intégrale nulle en valeur absolue et à la puissance p . Pour cela, nous identifions les fonctions égales presque partout en les mettant ensemble dans le même objet ; rigoureusement on dit que l'on quotiente l'espace $\mathcal{L}^p(\mathbb{R}^n)$ par l'ensemble $\mathcal{N}$ des fonctions nulles presque partout pour la mesure de Lebesgue. En gros, on définit la classe d'équivalence de f , notée $\tilde{f}$, puis de nouveau f par la suite, par

$$\tilde{f} = \{g \in \mathcal{L}^p(\mathbb{R}^n) \mid f = g \text{ presque partout}\}.$$

L'intégrale de toutes les fonctions d'une classe d'équivalence est la même, aussi ne distinguerons-nous presque plus à l'avenir les fonctions et leurs classes d'équivalence dans un contexte d'intégration (il convient tout de même de se méfier par moments).

5. La mesure de Lebesgue est donc adaptée à la mesure des robustes pavés de Paris que lancèrent les étudiants lors de mai 68. C'est pour cela que l'on dit souvent que les travaux de Lebesgue étaient révolutionnaires.

Nous notons donc, pour tout $p \in [1, +\infty[$, l'ensemble de ces classes d'équivalence

$$L^p(\mathbb{R}^n)$$

qui est un espace vectoriel normé (un vrai !) pour la norme

$$\|f\|_p = \left(\int_{\mathbb{R}^n} |f(x)|^p dx \right)^{\frac{1}{p}}.$$

De la même manière, on peut construire l'espace des fonctions essentiellement bornées en définissant,

$$\mathcal{L}^\infty(\mathbb{R}^n) = \mathcal{L}^\infty(\mathbb{R}^n, \mathcal{B}, \lambda) = \{f : \mathbb{R}^n \rightarrow \mathbb{K} \mid f \text{ mesurable et } \exists M \in \mathbb{R}_+, |f(x)| \leq M \text{ pour presque tout } x\}.$$

On quotiente encore une fois par l'ensemble des fonctions nulles presque partout pour obtenir l'espace vectoriel

$$L^\infty(\mathbb{R}^n),$$

auquel on associe la norme

$$\|f\|_\infty = \supess(f) = \inf(\{M \in \mathbb{R}_+ \mid f(x) \leq M \text{ presque partout}\}).$$

Tous ces espaces sont appelés espaces de Lebesgue et ils ont le bon goût d'être des espaces de Banach

Théorème 39 (*Riesz-Fisher*)

Pour tout $p \in [1, +\infty]$, $L^p(\mathbb{R}^n)$ muni de la norme $\|\cdot\|_p$ est un espace de Banach.

C'est évidemment vrai pour des espaces de Lebesgue définis sur n'importe quel ouvert de $\mathbb{R}^n$.

Le résultat suivant n'est pas surprenant mais souvent utile :

Théorème 40

Soit Ω un ouvert de $\mathbb{R}^n$, soit $f \in L^1(\Omega)$. Si pour toute fonction ϕ continue sur Ω et nulle en dehors d'un compact de Ω (on dit que ϕ est à support compact) on a

$$\int_{\Omega} f(x)\phi(x) dx = 0,$$

alors $f = 0$ presque partout dans Ω et est donc nulle vue comme élément de $L^1(\Omega)$.

1.d. Produit scalaire

Définition 41 (Produit scalaire (réel)). Soit E un espace vectoriel sur $\mathbb{R}$. On appelle produit scalaire sur E une application

$$\langle \cdot | \cdot \rangle : E \times E \rightarrow \mathbb{R},$$

qui vérifie

- $\forall x, y \in E, \langle x | y \rangle = \langle y | x \rangle$; (symétrie)
- $\forall x, y, z \in E, \forall \lambda \in \mathbb{R}, \langle x + \lambda y | z \rangle = \langle x | z \rangle + \lambda \langle y | z \rangle$; (linéarité à gauche)
- $\forall x \in E, \langle x | x \rangle \geq 0$; (positivité)
- $\forall x \in E, (\langle x | x \rangle = 0 \iff x = 0)$. (définie positivité)

Autrement dit, un produit scalaire réel est une forme bilinéaire symétrique, positive et définie. Remarquons que la symétrie et la linéarité à gauche impliquent la linéarité à droite et donc la bilinéarité. On peut aussi définir le produit scalaire dans le cas complexe, mais la définition est plus délicate à manipuler au début (avec l'habitude, c'est pareil ou presque)

Définition 42 (Produit scalaire hermitien). Soit E un espace vectoriel sur $\mathbb{C}$. On appelle produit scalaire hermitien sur E une application

$$\langle \cdot | \cdot \rangle : E \times E \rightarrow \mathbb{C},$$

qui vérifie

- $\forall x, y \in E, \langle x | y \rangle = \overline{\langle y | x \rangle}$; (symétrie hermitienne)
- $\forall x, y, z \in E, \forall \lambda \in \mathbb{R}, \langle x + \lambda y | z \rangle = \langle x | z \rangle + \lambda \langle y | z \rangle$ et $\langle x | y + \lambda z \rangle = \langle x | y \rangle + \overline{\lambda} \langle x | z \rangle$; (sesquilinearité)
- $\forall x \in E, \langle x | x \rangle \geq 0$; (positivité)
- $\forall x \in E, (\langle x | x \rangle = 0 \iff x = 0)$. (définie positivité)

Exemple 43. On peut mettre sur $\mathbb{C}^n$ le produit scalaire

$$\langle x | y \rangle = \sum_{i=1}^n x_i \overline{y_i}.$$

On peut mettre sur l'ensemble des fonctions continues de $[0, 1]$ dans $\mathbb{C}$ le produit scalaire

$$\langle f | g \rangle = \int_0^1 f(x) \overline{g(x)} dx.$$

Remarque 44. Un produit scalaire réel ou hermitien définit de manière naturelle une norme sur l'espace vectoriel :

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

Lorsque l'on se donne un produit scalaire, on se donne une norme de manière implicite.

Propriété 45 (Identité du parallélogramme). *Soit E un espace vectoriel muni d'un produit scalaire $\langle \cdot | \cdot \rangle$ dont la norme associée est notée $\|\cdot\|$. On a*

$$\forall x, y \in E, \|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

L'identité du parallélogramme est un bon moyen de savoir si la norme d'un espace vectoriel est issue d'un produit scalaire. C'est même une condition nécessaire et suffisante. Dans l'espace $\mathcal{C}([0, 1])$ des applications continues de $[0, 1]$ dans $\mathbb{R}$, la norme de la convergence uniforme $\|f\|_\infty = \sup_{x \in [0, 1]} |f(x)|$ ne vérifie pas l'identité du parallélogramme : pour $f(x) = x$ et $g(x) = 1$,

$$\|f + g\|^2 + \|f - g\|^2 = 2^2 + 1^2 = 5 \neq 4 = 2(1^2 + 1^2) = 2(\|f\|^2 + \|g\|^2).$$

Propriété 46 (Inégalité de Cauchy-Schwarz). *Soit E un espace vectoriel muni d'un produit scalaire et de sa norme naturelle. On a*

$$\forall x, y \in E, |\langle x | y \rangle| \leq \|x\| \|y\|,$$

avec égalité si et seulement si x et y sont linéairement dépendants.

On en déduit que le produit scalaire est continu sur $E \times E$ comme application de deux variables.

Une autre notion importante des espaces munis de produits scalaires est l'orthogonalité.

Définition 47 (Orthogonalité). *Soit E un espace vectoriel muni d'un produit scalaire $\langle \cdot | \cdot \rangle$. On dit que deux éléments x et y de E sont orthogonaux (relativement à ce produit scalaire), et on note $x \perp y$, lorsque*

$$\langle x | y \rangle = 0.$$

Soit F une partie de E , on appelle orthogonal de F , et on note $F^\perp$, la partie de E

$$F^\perp = \{y \in E \mid x \perp y\}.$$

L'orthogonal d'une partie est toujours un sous-espace vectoriel fermé dans l'espace vectoriel normé qui la contient.

2. Structure hilbertienne

Définition 48 (Espace de Hilbert). *On appelle espace de Hilbert un espace vectoriel muni d'un produit scalaire et complet pour la norme associée à ce produit scalaire.*

Si l'espace vectoriel est muni d'un produit scalaire et qu'il n'est pas complet, on dit juste qu'il est préhilbertien. Nous allons voir par la suite à quel point cette structure donne un cadre puissant pour faire de l'analyse et de la géométrie. En particulier, de nombreux espaces de fonctions sont utilisés en raison de leur structure hilbertienne naturelle.

Exemple 49 (Important). *L'espace de Lebesgue $L^2(\mathbb{R}^n)$ est un espace de Hilbert pour le produit scalaire*

$$\langle f | g \rangle = \int_{\mathbb{R}^n} f(x) \overline{g(x)} dx.$$

Le conjugué disparaît de lui-même dans le cas réel. Ce produit scalaire est associée à la norme $\|\cdot\|_2$ naturelle, ce qui rend la complétude évidente avec le théorème 39. De même pour tout ouvert Ω de $\mathbb{R}^n$, $L^2(\Omega)$ muni du produit scalaire précédent (intégrale sur Ω) est un espace de Hilbert.

Les espaces de dimension finie étant toujours complets, ce sont des espaces de Hilbert pour tout produit scalaire. La structure hilbertienne n'est par contre pas acquise pour les espaces vectoriels de dimension infinie.

2.a. Théorèmes généraux dans le cadre hilbertien

Définition 50 (Convexe). *Soit E un espace vectoriel sur $\mathbb{R}$ ou $\mathbb{C}$. Une partie C de E est dite convexe si*

$$\forall x, y \in C, \forall t \in [0, 1], (1-t)x + ty \in C$$

Définition 51 (Distance à une partie). *Soit E un espace métrique et F une partie de E . On appelle distance à la partie F l'application*

$$\begin{aligned} d_F : E &\rightarrow \mathbb{R}_+ \\ x &\mapsto \inf_{y \in F} d(x, y). \end{aligned}$$

Cette application n'est pas une distance au sens que nous avons vu plus haut mais elle remplit le même rôle : décrire l'éloignement d'un point à une partie. Nous sommes armés pour écrire l'un des résultats les plus centraux de l'analyse hilbertienne : l'idée qu'il est toujours possible de projeter les points sur les convexes fermés. Informellement, on peut dire qu'il est possible dans un espace de Hilbert de projeter les points sur les patates.

Théorème 52 (Projection sur un convexe fermé)

Soit H un espace de Hilbert et C une partie **convexe**, **fermée** et **non vide** de H . Alors, pour tout $x \in H$, il existe un *unique* $y \in C$, appelé projeté de x sur C , tel que :

$$\|x - y\| = d_C(x).$$

C'est l'unique point qui vérifie

$$\forall z \in C, \operatorname{Re}(\langle x - y | z - y \rangle) \leq 0.$$

Démonstration : Nous allons commencer par montrer l'existence et l'unicité du projeté en utilisant l'identité du parallélogramme puis nous montrerons la

deuxième caractérisation. Remarquons que nous travaillons ci-dessous avec des quantités au carré : tout se passe mieux ainsi ; les normes hilbertiennes *aiment* être au carré.

• Existence : Soit $x \in H$. Notons $d = \text{dist}(x, C)$. Si $d = 0$, alors il existe une suite de C qui converge vers x . Comme C est fermé, $x \in C$. Donc le projeté existe et il s'agit de x .

Sinon, soit $(x_n)_{n \in \mathbb{N}^*}$ une suite d'éléments de C telle que pour tout $n \in \mathbb{N}^*$,

$$\|x - x_n\|^2 \leq d^2 + \frac{1}{n}.$$

Soit $p, q \in \mathbb{N}^*$. On applique l'identité du parallélogramme à $u = x - x_p$ et $v = x - x_q$, ce qui donne

$$\|2x - x_p - x_q\|^2 + \|x_q - x_p\|^2 = 2\left(\|x - x_p\|^2 + \|x - x_q\|^2\right).$$

En réécrivant à gauche et en majorant à droite, on obtient

$$4\left\|x - \frac{x_p + x_q}{2}\right\|^2 + \|x_p - x_q\|^2 \leq 2\left(2d^2 + \frac{1}{p} + \frac{1}{q}\right).$$

Or, par convexité de C , $(x_p + x_q)/2 \in C$ et par conséquent

$$\left\|x - \frac{x_p + x_q}{2}\right\|^2 \geq d^2,$$

ce qui permet d'obtenir

$$\|x_p - x_q\|^2 \leq \frac{2}{p} + \frac{2}{q}.$$

La suite $(x_n)_{n \in \mathbb{N}^*}$ est donc de Cauchy. Comme H est complet, c'est une suite convergente. Comme C est fermée, la limite appartient à C .

Par passage à la limite sur l'inégalité vérifiée par la suite, la limite y vérifie $d(x, y) \leq d$, donc y satisfait

$$\|x - y\| = \text{dist}(x, C).$$

• Unicité : Supposons que deux points $y_1, y_2 \in C$ réalisent la distance. On a $\|x - y_1\| = \|x - y_2\| = d$. On applique l'identité du parallélogramme à $u = x - y_1$ et $v = x - y_2$:

$$4\left\|x - \frac{y_1 + y_2}{2}\right\|^2 + \|y_1 - y_2\|^2 = 2\left(\|x - y_1\|^2 + \|x - y_2\|^2\right),$$

comme précédemment, $(y_1 + y_2)/2 \in C$ par convexité et donc

$$\|y_1 - y_2\|^2 \leq 2(d^2 + d^2) - 4d^2 = 0.$$

Par séparation de la norme, $y_1 = y_2$.

• Caractérisation : Soit $z \in C$, par convexité, pour tout $t \in]0, 1]$, $(1-t)y + tz \in C$. Donc,

$$\|x - y\|^2 \leq \left\|x - \left((1-t)y + tz\right)\right\|^2 = \|x - y + t(y - z)\|^2.$$

De plus,

$$\|x - y + t(y - z)\|^2 = \|x - y\|^2 + \|t(y - z)\|^2 + 2\operatorname{Re}(\langle x - y | t(y - z) \rangle).$$

On a alors

$$t^2 \|y - z\|^2 + 2t\operatorname{Re}(\langle x - y | y - z \rangle) \geq 0.$$

Comme $t \neq 0$, on peut diviser par t puis faire tendre t vers 0, ce qui donne $\operatorname{Re}(\langle x - y | y - z \rangle) \geq 0$ et enfin en changeant le signe de $y - z$,

$$\operatorname{Re}(\langle x - y | z - y \rangle) \leq 0.$$

Réciproquement, supposons que $y \in C$ est tel que, pour tout $z \in C$,

$$\operatorname{Re}(\langle x - y | z - y \rangle) \leq 0.$$

Alors, pour tout $z \in C$,

$$\|x - z\|^2 = \|x - y + y - z\|^2 = \|x - y\|^2 + \|y - z\|^2 - 2\operatorname{Re}(\langle x - y | z - y \rangle) \geq \|x - y\|^2.$$

Donc par unicité du point qui minimise la distance à x dans C , y minimise la distance (z est quelconque). ■

On note $y = p_C(x)$ le projeté de x et p_C l'application projection. L'application projection est continue, et même lipschitzienne de constante 1 :

$$\forall x_1, x_2 \in H, \|p_C(x_1) - p_C(x_2)\| \leq \|x_1 - x_2\|.$$

Si on projette sur un sous-espace vectoriel, ce qui est très fréquent, on obtient de meilleures propriétés.

Corollaire 53. *Soit H un espace de Hilbert, F un sous-espace vectoriel **fermé** de H . Alors l'application projection p_F est linéaire, continue et pour tout $x \in H$, $y = p_F(x)$ est l'unique point qui vérifie*

$$y \in F \text{ et } x - y \in F^\perp.$$

Démonstration : Nous avons déjà la continuité de l'application projection.

Soit $x \in H$. Si un point $y \in F$ vérifie $x - y \in F^\perp$, alors $d_F(x)^2 = \inf_{z \in F} \|x - z\|^2$.

Or, comme pour tout $z \in F$, $y - z \in F$, par orthogonalité de $x - y$ à F on a

$$\|x - z\|^2 = \|x - y + y - z\|^2 = \|x - y\|^2 + \|y - z\|^2. \text{ }^6$$

Donc

$$\inf_{z \in F} \|x - z\|^2 = \|x - y\|^2,$$

et par conséquent $y = p_F(x)$. Réciproquement si on note $y = p_F(x)$, on a d'après le théorème précédent que pour tout $z \in F$,

$$\operatorname{Re}(\langle x - y | z - y \rangle) \leq 0.$$

6. Ce n'est rien d'autre que ce bon vieux théorème de Pythagore avec lequel on nous embête depuis le collège, mais généralisé à la dimension supérieure.

Si H est réel, comme F est un sous-espace vectoriel de H et $y \in F$, on peut écrire les éléments $z \in F$ sous la forme $z = y + \lambda\omega$ où λ parcourt $\mathbb{R}$ et ω parcourt F . Cela donne

$$\lambda \langle x - y \mid \omega \rangle \leq 0,$$

ce qui n'est possible que si $\langle x - y \mid \omega \rangle = 0$ pour tout $\omega \in F$, c'est-à-dire $x - y \in F^\perp$.

Si H est complexe, on peut écrire les éléments de F sous les deux formes $z = y + \lambda\omega$ et $z = y + i\lambda\omega$, ce qui donne par la même méthode, pour tout $\omega \in F$,

$$\operatorname{Re}(\langle x - y \mid \omega \rangle) = 0 \quad \text{et} \quad \operatorname{Im}(\langle x - y \mid \omega \rangle) = 0,$$

et finalement $\langle x - y \mid \omega \rangle = 0$, c'est-à-dire $x - y \in F^\perp$.

Enfin, la linéarité découle facilement de ce qui précède : si $y_1 = p_F(x_1)$ et $y_2 = p_F(x_2)$, alors $x_1 - y_1 \in F^\perp$ et $x_2 - y_2 \in F^\perp$. Donc pour tout $\lambda \in \mathbb{K}$, comme $F^\perp$ est un espace vectoriel, $(x_1 + \lambda x_2) - (y_1 + \lambda y_2) \in F^\perp$. De plus, $y_1 + \lambda y_2 \in F$, donc par la caractérisation que nous avons montrée, $p_F(x_1 + \lambda x_2) = y_1 + \lambda y_2$. ■

Il y a même équivalence : on peut prouver que l'application projection n'est linéaire que lorsque le convexe fermé sur lequel on projette est un sous-espace vectoriel.

Un deuxième résultat central, très utile en analyse, est le théorème de représentation. Nous allons nous intéresser aux formes linéaires, c'est-à-dire les applications linéaires à valeurs dans $\mathbb{K}$.

Définition 54 (Dual topologique). *Soit E un espace vectoriel normé sur $\mathbb{K}$. On appelle dual (topologique) de E et on note E^* l'ensemble*

$$E^* = \{\phi : H \rightarrow \mathbb{K} \mid \phi \text{ est linéaire et continue}\}.$$

Théorème 55 (Théorème de représentation de Riesz-Fréchet)

Soit H un espace de Hilbert. Pour toute forme linéaire continue $\phi \in H^*$, il existe un unique $y \in H$ tel que

$$\forall x \in H, \quad \phi(x) = \langle x \mid y \rangle.$$

Démonstration : Soit $\phi \in H^*$. L'application linéaire ϕ est continue, donc son noyau est fermé dans H . Il est donc en somme directe avec son supplémentaire orthogonal :

$$H = \operatorname{Ker}(\phi) \oplus \operatorname{Ker}(\phi)^\perp.$$

La codimension du noyau d'une forme linéaire est 1, donc la dimension de $\operatorname{Ker}(\phi)^\perp$ est 1. Soit u un vecteur unitaire (de norme 1) de $\operatorname{Ker}(\phi)^\perp$.

Posons $y = \overline{\phi(u)}u$.

Pour tout $x \in \operatorname{Ker}(\phi)$, $\langle x \mid y \rangle = 0 = \phi(x)$. Pour tout $x \in \operatorname{Ker}(\phi)^\perp$, x s'écrit λu pour un certain $\lambda \in \mathbb{R}$.

$$\langle x \mid y \rangle = \langle \lambda u \mid \overline{\phi(u)}u \rangle = \lambda \phi(u) \langle u \mid u \rangle = \lambda \phi(u) = \phi(\lambda u) = \phi(x).$$

Par somme directe, on a

$$\forall x \in H, \phi(x) = \langle x \mid y \rangle.$$

Supposons que y n'est pas unique et qu'il existe $z \in H$ qui vérifie la même propriété, alors pour tout $x \in H$,

$$\langle x \mid y - z \rangle = \phi(x) - \phi(x) = 0.$$

C'est en particulier vrai pour $x = y - z$, donc $\|y - z\|^2 = 0$ et enfin $y = z$ par séparation de la norme. ■

Ce théorème revient aussi à dire que l'application

$$\begin{aligned} J : H &\rightarrow H^* \\ y &\mapsto (x \mapsto \langle x \mid y \rangle), \end{aligned}$$

est une bijection.

Dans les cas "concrets" cela se comprend mieux. Par exemple, on peut se placer dans l'espace de Hilbert $L^2(\mathbb{R})$, choisir $\langle f \mid g \rangle = \int_{\mathbb{R}} f(x)g(x) dx$ et $\phi(g) = \int_{\mathbb{R}} g(x) e^{-x^2} dx$. Le théorème nous dit qu'il existe un unique f qui vérifie

$$\forall g \in H, \int_{\mathbb{R}} f(x)g(x) dx = \int_{\mathbb{R}} g(x) e^{-x^2} dx.$$

Dans ce cas, f est facile à trouver, c'est $x \mapsto e^{-x^2}$, mais outre le fait que le théorème nous dit que ce f est unique, cela fonctionne pour toute forme linéaire ϕ .

En équations aux dérivées partielles, on a parfois besoin d'un résultat similaire (en particulier pour fonder la méthode d'approximation dite des éléments finis).

Théorème 56 (Théorème de Lax-Milgram)

Soit H un espace de Hilbert réel (resp. complexe) muni d'un produit scalaire réel (resp. hermitien) $\langle \cdot \mid \cdot \rangle$ et de sa norme associée $\|\cdot\|$. Soit $a(\cdot, \cdot)$ une forme bilinéaire sur H (resp. sesquilinéaire) et soit $l \in H^*$ une forme linéaire continue sur H . Supposons que a est

- continue, c'est-à-dire

$$\exists C \in \mathbb{R}_+^*, \forall u, v \in H, |a(u, v)| \leq C \|u\| \|v\|;$$

- coercive, c'est-à-dire

$$\exists \alpha \in \mathbb{R}_+^*, \forall u \in H, a(u, u) \geq \alpha \|u\|^2.$$

Alors il existe un unique $u \in H$ tel que

$$\forall v \in H, a(u, v) = l(v).$$

Si de plus a est symétrique, alors

$$u = \min_{v \in H} \left(\frac{1}{2} a(v, v) - l(v) \right).$$

Ce théorème très long revient simplement à dire que dans un cadre adapté il y a une unique solution à l'équation fonctionnelle de la variable u

$$a(u, \cdot) = l(\cdot).$$

Sa démonstration n'est pas insurmontable mais demande un peu de temps et d'énergie. Une manière de faire est de montrer un théorème plus général, le théorème de Stampacchia. On trouvera la démonstration du théorème de Stampacchia dans l'excellent polycopié d'introduction à la théorie spectrale et à l'analyse harmonique de Frédéric Paulin, trouvable sur sa page ou sur le lien https://www.math.u-psud.fr/~paulin/notescours/cours_magistere2.pdf. On remarquera que a n'a pas besoin d'être symétrique. Si elle l'est, on retrouve le théorème de représentation car a définit alors un nouveau produit scalaire qui conserve le caractère hilbertien de l'espace grâce à la continuité.

2.b. Bases Hilbertiennes

Définition 57 (Famille orthogonale, famille orthonormée). *Soit E un espace vectoriel muni d'un produit scalaire $\langle \cdot | \cdot \rangle$. Une famille $(x_i)_{i \in I}$ d'éléments de E indexée par un ensemble I quelconque est dite orthogonale si*

$$\forall i, j \in I, x_i \perp x_j.$$

La famille est dite orthonormée si de plus

$$\forall i \in I, \|x_i\| = 1.$$

Propriété 58. *Toute famille orthonormée est libre.*

Définition 59 (Famille totale). *Soit E un espace vectoriel normé et $(x_i)_{i \in I}$ une famille d'éléments de E . La famille est dite totale dans E si l'ensemble*

$$F = \left\{ \sum_{j=1}^n \lambda_j x_{k_j} \mid n \in \mathbb{N}, k_1, \dots, k_n \in I, \lambda_1, \dots, \lambda_n \in \mathbb{K} \right\},$$

des combinaisons linéaires finies d'éléments de la famille $(x_i)_{i \in I}$ est une partie dense de E .

Définition 60 (Base orthonormée). *Une famille $(x_n)_{n \in \mathbb{N}}$ d'un espace de Hilbert H est une base orthonormée de H si c'est une famille orthonormée et totale.*

On prendra garde à ne pas confondre base d'un espace vectoriel (tout vecteur est combinaison linéaire d'un nombre fini d'élément de la base) et base orthonormée au sens hilbertien (définition précédente).

Théorème 61 (Décomposition dans une base Hilbertienne)

Soit H un espace de Hilbert et $(x_n)_{n \in \mathbb{N}}$ une base orthonormée de H , s'il en existe. Alors pour tout $x \in H$,

$$x = \sum_{n=0}^{+\infty} \xi_n x_n, \quad \text{avec} \quad \xi_n = \langle x | x_n \rangle,$$

où la somme infini est par définition une limite des sommes partielles pour la norme de H .

Théorème 62 (Formule de Parseval⁷)

Soit H un espace de Hilbert et $(x_n)_{n \in \mathbb{N}}$ une base orthonormée de H . Alors pour tout $x \in H$,

$$\|x\|^2 = \sum_{n=0}^{+\infty} |\langle x | x_n \rangle|^2;$$

et pour tout $y \in H$

$$\langle x | y \rangle = \sum_{n=0}^{+\infty} \langle x | x_n \rangle \overline{\langle y | x_n \rangle},$$

où la série converge absolument.

Dans la deuxième partie, le conjugué n'est utile que pour le cas d'un espace complexe.

Définition 63 (Espace séparable). *Un espace vectoriel normé est dit séparable s'il possède une partie dénombrable dense.*

L'espace $\mathbb{R}$ est séparable parce qu'il contient la partie dénombrable et dense $\mathbb{Q}$. Tous les espaces de dimension finie de dimension n sont séparables parce qu'il contiennent la partie dense dénombrable $\mathbb{Q}^n$. Il existe des espaces de dimension infinie qui ne sont pas séparables. Selon la norme que l'on met sur un espace de dimension infinie, il peut être séparable ou non.

Théorème 64

Tout espace de Hilbert séparable possède des bases orthonormées.

Corollaire 65. *Tout espace de Hilbert séparable de dimension infinie est isomorphe à l'espace l^2 des suites de carré sommable muni de la norme définie par*

$$\|(x_n)_{n \in \mathbb{N}}\|_2 = \sqrt{\sum_{n=0}^{+\infty} x_n^2}.$$

En particulier tous les espaces de Hilbert séparables de dimension infinie sont isomorphes entre eux. Étudier différents espaces de Hilbert séparables (des suites, des fonctions, etc) revient donc à étudier le même objet sous différents points de vues, mais c'est au fond ce que l'on fait tout le temps en mathématiques.

Exemple 66 (Important). *Pour tout $n \in \mathbb{Z}$, soit la fonction*

$$e_n : t \mapsto e^{2i\pi nt}.$$

La famille $(e_n)_{n \in \mathbb{Z}}$ est une base orthonormée de l'espace de Hilbert $L^2(]0, 1[)$. Si on note, pour toute fonction $f \in L^2(]0, 1[)$,

$$\hat{f}(n) = \langle f \mid e_n \rangle = \int_0^1 f(t) e^{2i\pi nt} dt,$$

*les **coefficients de Fourier** de f , on a alors convergence de la série de Fourier de f vers f pour la norme $\|\cdot\|_2$ associée au produit scalaire de $L^2(]0, 1[)$. Plus explicitement, cela signifie que*

$$\lim_{n \rightarrow +\infty} \left\| f - \sum_{k=0}^n \hat{f}(k) e_k \right\|_2 = \lim_{n \rightarrow +\infty} \sqrt{\int_0^1 \left(f(t) - \sum_{k=0}^n \hat{f}(k) e^{2i\pi kt} \right)^2 dt} = 0.$$

Le résultat est également vrai pour les fonctions définies sur $\mathbb{R}$ et 1-périodiques de carré intégrable sur une période car cet espace est isomorphe à $L^2(]0, 1[)$ de manière naturelle. Il est très important de noter que la convergence est en norme et pas ponctuelle. Un résultat classique de théorie de la mesure nous dit que pour toute fonction $f \in L^2(]0, 1[)$, il existe une suite $(j_k)_{k \in \mathbb{N}}$, telle que pour presque tout $t \in]0, 1[$,

$$\lim_{n \rightarrow +\infty} \sum_{k=0}^n \hat{f}(j_k) e^{2i\pi j_k t} = f(t).$$

Ce que l'on souhaiterait avoir, c'est que pour presque tout $t \in]0, 1[$,

$$\lim_{n \rightarrow +\infty} \sum_{k=0}^n \hat{f}(k) e^{2i\pi kt} = f(t).$$

Cette question a été ouverte par Nikolaï Lusin⁸ en 1913 et Lennart Carleson a répondu par l'affirmative en 1966. C'est une des démonstrations les plus difficiles et les plus arides de l'histoire de l'analyse. Elle a valu à Carleson le prix Abel (une des plus hautes distinctions en mathématiques).

8. Parfois orthographié Nikolaï Nikolaïevitch Louzine (1883-1950). C'était un spécialiste d'analyse, de topologie et de théorie des ensembles. Outre son importante contribution mathématique, il a participé à la formation de grands mathématiciens russes, dont Andreï Kolmogorov.

3. Synthèse et pratique

3.a. Synthèse de la section

L'idée à retenir de cette accumulation de définitions et de résultats abstraits est qu'il est possible de mettre trois types majeurs de structures (outre la structure d'espace vectoriel) sur un ensemble avant de l'utiliser pour faire de l'analyse : la métrique, la norme et le produit scalaire. Dans les trois cas, on a besoin de la propriété de complétude pour travailler plus facilement, d'où les notions d'espaces de Banach et de Hilbert.

En dimension infinie, tout est plus difficile parce que la boule unité n'est pas compacte. Il y a "beaucoup d'espace" partout et certaines intuitions sont fausses. Il faut faire très attention à ce qu'on dit et se forger de nouvelles intuitions par le travail.

Les espaces de Lebesgue sont très intéressants pour faire de l'analyse et pour résoudre des équations aux dérivées partielles parce qu'ils ont une structure naturelle définie à partir de l'intégrale qui en fait des espaces de Banach, et même de Hilbert pour $p = 2$. C'est pourquoi ils sont énormément utilisés.

Dans les espaces de Hilbert séparables, il est possible de construire un équivalent des bases en dimension finie : les bases hilbertiennes. Elles permettent de décomposer les vecteurs et donc les problèmes. C'est le cas notamment de la décomposition en série de Fourier qui est connu comme un outil puissant pour attaquer certains problèmes d'analyse et d'ingénierie. On retiendra les *grands* théorèmes des espaces de Hilbert et leurs interprétations :

- Il est possible de projeter sur les ensembles convexes fermés et de faire des projections orthogonales sur les sous-espaces vectoriels fermés. L'intuition de la dimension finie est valide tant qu'on est **fermé**.
- Les éléments du dual (formes linéaires) sont représentés par les éléments de l'espace de Hilbert. L'ensemble des fonctions continues de l'espace de Hilbert dans $\mathbb{R}$ (ou $\mathbb{C}$) est semblable à l'espace de Hilbert et s'écrit de manière simple avec le produit scalaire ; toutes ces fonctions s'écrivent $\langle x | \cdot \rangle$.
- Le théorème de Lax-Milgram permet de dire des choses dans certains cas plus techniques liés aux EDP. On peut représenter les formes linéaires à partir d'autres produits scalaire ou même de "presque produits scalaires" : les applications bilinéaires coercives continues.

3.b. Exercices

Exercice 67 (Espaces ultramétriques). Soit (E, d) un espace ultramétrique au sens de la définition 3. On note

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

la boule de centre x et de rayon r pour la distance d dans l'espace E .

1. Soit un rayon $r \in \mathbb{R}_+^*$ et un élément $x \in E$. Montrer que tout point d'une boule est son centre ; c'est-à-dire

$$\forall y \in B(x, r), \quad B(y, r) = B(x, r).$$

2. Soit deux rayons $r_1, r_2 \in \mathbb{R}_+^*$ tels que $r_1 \leq r_2$ et deux éléments $x_1, x_2 \in E$. Montrer que si deux boules ont une intersection non-vidue alors l'une est contenue dans l'autre ; c'est-à-dire

$$B(x_1, r_1) \cap B(x_2, r_2) \neq \emptyset \implies B(x_1, r_1) \subset B(x_2, r_2).$$

3. Montrer que tous les triangles sont isocèles, au sens où pour tous points $x, y, z \in E$,

$$d(x, y) = d(x, z) \quad \text{ou} \quad d(x, y) = d(y, z) \quad \text{ou} \quad d(x, z) = d(y, z).$$

Exercice 68 (Représentation dans les espaces pré-Hilbertiens). Soit l'ensemble $\mathcal{C}^0([0, 1], \mathbb{R})$ des applications continues de $[0, 1]$ dans $\mathbb{R}$. Munissons-le du produit scalaire

$$\langle f \mid g \rangle = \int_0^1 f(x)g(x) \, dx,$$

et de la norme associée. Soit $p \in \mathbb{R}_+^*$ et $a \in]0, 1[$, et soit la forme linéaire sur $\mathcal{C}^0([0, 1], \mathbb{R})$ définie par

$$l(f) = \int_0^a x^p f(x) \, dx.$$

1. Montrer que l est continue et calculer sa norme.
2. Montrer qu'il n'existe pas d'élément $g \in \mathcal{C}^0([0, 1], \mathbb{R})$ tel que $\langle g \mid \cdot \rangle = l(\cdot)$. En déduire que $\mathcal{C}^0([0, 1], \mathbb{R})$ n'est pas un espace de Hilbert mais seulement un espace préhilbertien.

II. THÉORIE SPECTRALE

« HAMLET. - Fais-le-moi vite connaître, pour qu'avec des ailes rapides comme l'idée ou les pensées d'amour, je vole à la vengeance !
LE SPECTRE. - Tu es prêt, je le vois. Tu serais plus inerte que la ronce qui s'engraisse et pourrit à l'aise sur la rive du Léthé, si tu n'étais pas excité par ceci. Maintenant, Hamlet, écoute ! »

William Shakespeare, *Hamlet*⁹

Dans toute cette partie, nous allons nous intéresser aux applications linéaires continues d'un espace de Banach dans lui-même, que nous appellerons **opérateurs**. Soit E un espace de Banach, on notera

$$\mathcal{L}(E) = \{T : E \rightarrow E \mid T \text{ linéaire et continue}\}$$

l'ensemble de ses opérateurs et

$$GL(E) = \{T \in \mathcal{L}(E) \mid T \text{ inversible}\}$$

l'ensemble de ses opérateurs inversibles. On rappelle qu'un opérateur T est inversible si et seulement si il existe un opérateur $T^{-1} \in \mathcal{L}(E)$ tel que

$$T \circ T^{-1} = T^{-1} \circ T = I_E,$$

où $I_E : x \mapsto x$ est l'opérateur identité de E . Comme E est un espace de Banach, d'après le théorème d'isomorphisme de Banach (voir Annexe A.), si T est bijectif et continu, alors il est inversible au sens de la définition précédente et réciproquement. Inversible et bijectif seront donc dans notre cadre la même notion.

Le composé de deux opérateurs $T, U \in \mathcal{L}(E)$ sera noté

$$T \circ U \quad \text{ou pour simplifier} \quad TU,$$

et on écrira souvent Tx à la place de $T(x)$ s'il n'y a pas d'ambiguïté. On écrira aussi T^n pour désigner $T \circ \dots \circ T$, où l'on compose n fois.

À partir de la norme $\|\cdot\|_E$ de E , on définit de manière naturelle une norme sur $\mathcal{L}(E)$, appelée norme d'opérateurs, par

$$\|T\| = \sup_{\|x\|_E \leq 1} \|Tx\|_E.$$

9. L'idée de citer Hamlet avant de parler du spectre des opérateurs linéaire a été ici plagée sur l'excellent) polycopié de théorie des graphes d'Olivier Fouquet, professeur à l'université Paris-Sud.

Muni de ces opérations (addition, multiplication par les scalaires du corps de base de E , composition) et de la norme d'opérateur, $\mathcal{L}(E)$ est un anneau, une algèbre et un espace de Banach. De plus, cette norme vérifie

$$\forall T, U \in \mathcal{L}(E), \|TU\| \leq \|T\| \|U\|,$$

propriété que l'on appelle sous-multiplicativité. On dit en résumé que $\mathcal{L}(E)$ est une algèbre de Banach.

Dans toute la suite, $\mathbb{K}$ désigne $\mathbb{R}$ ou $\mathbb{C}$.

Nous énonçons et démontrons également avant de commencer un résultat (plutôt simple en réalité) dû à Carl Neumann¹⁰ qui est souvent utile pour travailler sur les opérateurs.

Théorème 69 (Théorème de perturbation de Neumann)

Soit $\|\cdot\|$ une norme sous-multiplicative sur $\mathcal{M}_n(\mathbb{R})$. Soit $A \in \mathcal{M}_n(\mathbb{R})$ vérifiant $\|A\| < 1$, alors $I_n - A$ est inversible. De plus, si $B \in GL_n(\mathbb{R})$ vérifie

$$\|A\| < \frac{1}{\|B^{-1}\|},$$

alors $B - A$ est inversible et

$$(B - A)^{-1} = \sum_{k=0}^{+\infty} (B^{-1}A)^k B^{-1}$$

Démonstration : • Montrons d'abord le résultat pour $B = I_n$: l'hypothèse est $\|A\| < 1$. Comme nous avons supposé la norme sous-multiplicative, $\|A^k\| \leq \|A\|^k$, ce qui implique que la série

$$\sum_{k \in \mathbb{N}} A^k,$$

est normalement convergente. Or, on peut vérifier en développant l'égalité, pour tout $j \in \mathbb{N}$,

$$I_n - A^{j+1} = (I_n - A) \sum_{k=0}^j A^k$$

Comme la série de terme général A^k est normalement convergente, A^j tend vers 0 pour la norme $\|\cdot\|$ quand j tend vers $+\infty$, ce dont on déduit par passage à la limite dans l'égalité précédente

$$I_n = (I_n - A) \sum_{k=0}^{+\infty} A^k.$$

10. Carl Neumann (1832 - 1925), à ne pas confondre avec d'autres Neumann, est un mathématicien allemand. C'était l'un des pionniers dans le domaine des équations intégrales. Il a également donné son nom aux conditions de Neumann pour les équations aux dérivées partielles.

Donc $I_n - A$ est inversible et son inverse est la limite de la série de terme général A^k .

- Considérons à présent $B \in GL_n(\mathbb{R})$: B inversible, donc

$$B - A = B(I_n - B^{-1}A)$$

Or,

$$\|B^{-1}A\| \leq \|B^{-1}\| \|A\| < 1,$$

par hypothèse. Donc, d'après ce qui précède, $I_n - B^{-1}A$ est inversible, et grâce à la structure de groupe de $GL_n(\mathbb{R})$, $B - A$ l'est aussi. De plus, en appliquant la formule de l'inverse trouvée précédemment, on obtient :

$$(B - A)^{-1} = (I_n - B^{-1}A)^{-1} B^{-1} = \left(\sum_{k=0}^{+\infty} (B^{-1}A)^k \right) B^{-1}. \quad \blacksquare$$

1. Spectre d'un opérateur

Définition 70 (Spectre, valeur spectrale). *Soit E un espace de Banach sur $\mathbb{K}$ et $T \in \mathcal{L}(E)$, on appelle spectre de E l'ensemble*

$$Sp(T) = \{\lambda \in \mathbb{K} \mid T - \lambda I_E \text{ n'est pas inversible}\}.$$

Les éléments du spectre sont appelés valeurs spectrales.

Certaines valeurs spectrales particulières sont intéressantes dans les applications : celles pour lesquelles $T - \lambda I_E$ n'est pas injectif, c'est-à-dire :

Définition 71 (Valeur propre). *Soit E un espace de Banach sur $\mathbb{R}$ (resp. $\mathbb{C}$) et $T \in \mathcal{L}(E)$, on appelle valeur propre de T un nombre réel (resp. complexe) λ tel que $T - \lambda I_E$ n'est pas injectif, c'est-à-dire*

$$\exists x \in E, x \neq 0, Tx = \lambda x.$$

En dimension finie toutes les valeurs spectrales sont des valeurs propres et il y en a un nombre fini inférieur ou égal à la dimension de l'espace (égal dans le cas complexe en comptant plusieurs fois les valeurs propres multiples). En dimension infinie, on peut avoir des valeurs spectrales mais sans aucune valeur propre, un nombre fini de valeurs propres, une infinité de valeurs propres... On peut par contre dire des choses sur la structure du spectre.

Théorème 72

Soit E un espace de Banach sur $\mathbb{K}$ et $T \in \mathcal{L}(E)$. Le spectre de T est une partie compacte de $\mathbb{K}$ et

$$\forall \lambda \in Sp(T), |\lambda| \leq \|T\|.$$

Démonstration : On va montrer que le spectre est une partie fermée et bornée de $\mathbb{K}$, ce qui prouve qu'il est compact.

Pour montrer que le spectre est fermé, on va montrer que son complémentaire est ouvert. Le complémentaire du spectre est souvent appelé *ensemble résolvant*. Soit $\lambda_0 \in \mathbb{K} \setminus Sp(T)$. Soit $\lambda \in \mathbb{K}$; si

$$|\lambda - \lambda_0| < \frac{1}{\|(T - \lambda_0 I)^{-1}\|},$$

alors

$$\|(T - \lambda I_E) - (T - \lambda_0 I_E)\| = |\lambda_0 - \lambda| < \frac{1}{\|(T - \lambda_0 I)^{-1}\|}.$$

Donc, d'après le théorème de perturbation de Neumann, $T - \lambda I_E$ est inversible et par conséquent $\lambda \in \mathbb{K} \setminus Sp(T)$. Donc le complémentaire du spectre est ouvert et le spectre est fermé.

Ensuite, si $|\lambda| > \|T\|$,

$$\left\| I_E + \frac{1}{\lambda}(T - \lambda I_E) \right\| = \frac{1}{|\lambda|} \|T\| < 1,$$

donc d'après le théorème de perturbation de Neumann, $-\frac{1}{\lambda}(T - \lambda I_E)$ est inversible, donc $T - \lambda I_E$ aussi. Donc λ n'est pas dans le spectre. Le spectre est donc borné. ■

Théorème 73 (Stone)

Soit E un espace de Banach **sur** $\mathbb{C}$ et $T \in \mathcal{L}(E)$. Le spectre de T est non-vide.

On peut démontrer ce théorème sans efforts excessifs en utilisant l'analyse complexe (en particulier le théorème de Liouville qui dit que toute fonction holomorphe entière et bornée est constante). Le résultat est faux pour un espace de Banach réel.

Définition 74 (Rayon spectral). Soit E un espace de Banach et $T \in \mathcal{L}(E)$ un opérateur de spectre non vide. On appelle rayon spectral de T le nombre réel

$$r(T) = \sup_{\lambda \in Sp(T)} |\lambda|.$$

On a une caractérisation simple du rayon spectral dans le cas d'un espace complexe :

Théorème 75 (Théorème de Gelfand)

Soit E un espace de Banach **sur** $\mathbb{C}$ et $T \in \mathcal{L}(E)$. Alors,

$$r(T) = \lim_{n \rightarrow +\infty} \|T^n\|^{\frac{1}{n}}.$$

Rappelons encore, au besoin, que T^n désigne $T \circ \dots \circ T$, où l'on compose n fois. Dans le cas réel, la limite existe encore mais le rayon spectral lui est inférieur et non nécessairement égal. Encore une fois, le théorème peut être démontré sans trop d'effort *via* l'utilisation de l'analyse complexe.

2. Opérateur compact

Nous allons à présent considérer des opérateurs un peu plus généraux. Soit E et F deux espaces de Banach. On appelle toujours opérateur une application linéaire continue de E dans F . On note $\mathcal{L}(E, F)$ l'ensemble des opérateurs de E dans F . C'est un espace de Banach pour la norme d'opérateur

$$\|T\| = \sup_{\|x\|_E \leq 1} \|Tx\|_F.$$

Avant d'attaquer le cœur du sujet, nous rappelons une définition de topologie.

Définition 76 (Partie relativement compacte). *Soit E un espace métrique. Une partie A de E est dite relativement compacte si son adhérence $\overline{A}$ (le plus petit fermé qui la contient) est compacte.*

En dimension finie tous les ensembles bornés sont relativement compacts. Si l'espace métrique est complet, une partie est relativement compacte si et seulement si pour tout $\varepsilon \in \mathbb{R}_+^*$ il est possible de la recouvrir par un nombre fini de boules de rayon ε .

Définition 77 (Opérateur compact). *Soit E et F deux espaces de Banach. Un opérateur $T : E \rightarrow F$ est dit compact si l'image par T de la boule unité fermée de E est relativement compacte dans F .*

Si l'on note $\overline{B}_E = \{x \in E \mid \|x\|_E \leq 1\}$, cela revient à demander que $\overline{T(\overline{B}_E)}$ soit un compact de F . Si F est de dimension finie, tous les opérateurs de E dans F sont compacts. Si $T(E)$ est un sous-espace vectoriel de dimension finie de F , T est compact et on dit que T est un **opérateur de rang fini**. De plus, toute limite (en norme d'opérateur) d'une suite d'opérateurs de rang fini est un opérateur compact (mais pas nécessairement de rang fini). C'est un corollaire direct du théorème suivant.

Théorème 78

Soit E et F deux espaces de Banach. L'ensemble des opérateurs compacts de E dans F est un fermé de $\mathcal{L}(E, F)$.

La réciproque est une question plus difficile : tout opérateur compact est-il la limite d'une suite d'opérateurs de rang fini ? Banach avait ouvert la question en 1932. Stanisław Mieczysław Mazur l'avait reposée dans le *Cahier Écossais*. Mazur promettait une oie vivante à quiconque répondrait à la question.

Le mathématicien Per Enflo répond en 1972 : le résultat est vrai dans les espaces de Hilbert et dans les espaces du type $L^p(]0, 1])$ mais faux dans certains espaces de Banach. Mazur a remis l'oie à Enflo lors d'un congrès à Varsovie lors d'une cérémonie diffusée dans toute la Pologne (voir figure 1).

Les opérateurs compacts ont un spectre particulier.

Théorème 79 (Spectre des opérateurs compacts ; F. Riesz, 1918)

Soit E un espace de Banach de dimension infinie et $T \in \mathcal{L}(E)$ un



FIGURE 1 – L’oie vivante remise à Enflo par Mazur en 1972, on peut observer dans son œil un air de profond désarroi face à l’absurdité de la situation.

opérateur compact. Alors soit le spectre de T s’écrit

$$Sp(T) = \{0\} \cup \{\lambda_n \mid n \in \mathbb{N}\},$$

où $(\lambda_n)_{n \in \mathbb{N}}$ est une suite de valeurs propres qui tend vers 0, soit il est constitué de la valeur spectrale 0 (peut être propre) et d’un nombre fini de valeurs propres. Dans les deux cas, les sous-espaces propres $E_\lambda = \text{Ker}(T - \lambda I_E)$ associés aux valeurs propres non-nulles sont de dimension finie.

En particulier, le spectre d’un opérateur compact est au plus dénombrable. Le nombre 0 est toujours une valeur spectrale d’un opérateur compact en dimension infinie (et c’est parfois aussi une valeur propre). C’est faux en dimension finie : tous les opérateurs sont compacts mais ils n’admettent pas tous 0 comme valeur spectrale.

3. Adjoint d’un opérateur

Dans le cadre plus riche d’un espace de Hilbert, on peut définir la notion d’opérateur adjoint qui permet parfois d’avoir des informations sur le spectre. L’existence de l’adjoint est une conséquence élémentaire du théorème de Riesz-Fréchet que nous avons énoncé plus tôt.

Propriété 80. *Soit H un espace de Hilbert et $T \in \mathcal{L}(H)$ un opérateur. Il existe*

un unique opérateur $T^* \in \mathcal{L}(H)$ vérifiant

$$\forall x, y \in H, \langle Tx \mid y \rangle = \langle x \mid T^*y \rangle.$$

Cet unique opérateur est appelé adjoint de T dans H . En dimension finie, l'adjoint d'un opérateur T associé à une matrice A est l'opérateur associé à la matrice transposée de A si l'espace est réel et à la matrice transconjuguée de A si l'espace est complexe.

Propriété 81. Soit H un espace de Hilbert et $T, U \in \mathcal{L}(H)$ deux opérateurs. On a

- $(T^*)^* = T$;
- $(TU)^* = U^*T^*$;
- $\|T^*\| = \|T\|$;
- $\text{Ker}(T) = \text{Im}(T^*)^\perp$;
- $\overline{\text{Im}(T)} = \text{Ker}(T^*)^\perp$.

Définition 82 (Opérateur auto-adjoint). Soit H un espace de Hilbert. L'opérateur $T \in \mathcal{L}(H)$ est dit auto-adjoint si $T^* = T$.

Les opérateurs auto-adjoints sont une généralisation en dimension infinie et éventuellement dans un espace complexe de la notion très utile de matrice symétrique en dimension finie. Dans le cas complexe on parle parfois d'opérateurs hermitiens. Ces opérateurs possèdent de nombreuses propriétés agréables.

Propriété 83. Soit H un espace de Hilbert et $T, U \in \mathcal{L}(H)$ deux opérateurs auto-adjoints.

- $\forall x \in H, \langle Tx \mid x \rangle \in \mathbb{R}$;
- $\|T\| = \sup_{\|x\|=1} |\langle Tx \mid x \rangle|$;
- $\|T^*\| = \|T\|$;
- $\text{Sp}(T) \subset [m, M]$ (donc le spectre est réel¹¹), avec

$$m = \inf_{\|x\|=1} |\langle Tx \mid x \rangle| \quad \text{et} \quad M = \sup_{\|x\|=1} |\langle Tx \mid x \rangle|.$$

On a également $m \in \text{Sp}(T)$ et $M \in \text{Sp}(T)$.

Corollaire 84. Le spectre d'un opérateur auto-adjoint d'un espace de Hilbert non réduit à $\{0\}$ n'est jamais vide, possède toujours au moins une valeur propre et son rayon spectral est égal à sa norme. Les sous-espace propres correspondants à des valeurs propres distinctes sont deux à deux orthogonaux.

Tous ces résultats permettent d'obtenir la décomposition suivante, très utile dans de nombreux cas pratiques d'analyse des équations différentielles et intégrales-différentielles.

11. S'il est réel, espérons qu'il n'est pas aussi vindicatif que celui qui visite Hamlet...

Théorème 85 (Décomposition des opérateurs auto-adjoints compacts)

Soit H un espace de Hilbert séparable de dimension infinie. Pour tout opérateur **compact** et **auto-adjoint** $T \in \mathcal{L}(H)$, il existe une base orthonormée $(e_n)_{n \in \mathbb{N}}$ dont les éléments sont des vecteurs propres de T et l'on a

$$\forall x \in H, Tx = \sum_{n=0}^{+\infty} \lambda_n \langle x | e_n \rangle e_n,$$

où λ_n est la valeur propre associée à e_n .

4. Synthèse et pratique

4.a. Synthèse de la section

Les opérateurs, c'est-à-dire les applications linéaires continues en dimension infinie ont un comportement différent des opérateurs en dimension finie : ils peuvent avoir une infinité de valeurs propres et ils ont des valeurs dites "spectrales" qui ne sont pas forcément des valeurs propres. Comme en dimension finie, l'étude du spectre permet de dire beaucoup de choses sur les opérateurs mais c'est plus difficile en dimension infinie pour les raisons que nous venons d'évoquer.

Afin de pouvoir étudier plus avant les opérateurs en dimension infinie, nous avons parfois besoin de faire des hypothèses simplificatrices. L'une d'elles est la compacité. Les opérateurs compacts sont un peu *comme des opérateurs en dimension finie*. Il ont un spectre particulier, comme l'exprime le théorème sur la forme de leur spectre. Ils ont au plus une valeur spectrale qui n'est pas une valeur propre.

On peut supposer aussi que les opérateurs sont auto-adjoints, en quelque sorte l'équivalent des matrices symétriques en dimension finie. Cela permet de trouver des bases orthonormées (au sens hilbertien, évidemment), ce qui s'avère d'une grande utilité dans certaines applications (voir la section sur les valeurs propres du laplacien).

4.b. Exercices

Exercice 86. Soit H un espace de Hilbert réel et $T \in \mathcal{L}(H)$ un opérateur positif; c'est-à-dire

$$\forall x \in E, \langle Tx | x \rangle \geq 0.$$

Montrer que ses valeurs propres (réelles) éventuelles sont positives.

Exercice 87. Soit H un espace de Hilbert complexe séparable de dimension infinie, soit $(e_n)_{n \in \mathbb{N}}$ une base hilbertienne de H et u l'opérateur défini par la

formule

$$u : x = \sum_{n=0}^{+\infty} e_n \mapsto \sum_{n=0}^{+\infty} e_{n+1},$$

appelé opérateur de décalage.

Montrer que u est bien défini, qu'il est continu, qu'il n'a aucune valeur propre et que son spectre est le disque unité fermé :

$$Sp(u) = \{z \in \mathbb{C}, |z| \leq 1\}.$$

Exercice 88. Soit H un espace de Hilbert complexe séparable de dimension infinie, soit $(e_n)_{n \in \mathbb{N}}$ une base hilbertienne de H , C un compact de $\mathbb{C}$ et $(\lambda_n)_{n \in \mathbb{N}}$ une suite de nombres complexes contenus dans C telle que l'ensemble $\{\lambda_n \mid n \in \mathbb{N}\}$ est dense dans C .

Montrer qu'il existe un unique opérateur $u \in \mathcal{L}(H)$ de H tel que

$$\forall n \in \mathbb{N}, u(e_n) = \lambda_n e_n,$$

qu'il vérifie

$$Sp(u) = C,$$

que ses valeurs propres sont les λ_n . Trouver ses espaces propres.

III. RÉGULARITÉ ET ESPACES DE SOBOLEV

« La logique parfois engendre des monstres. On vit surgir toute une foule de fonctions bizarres qui semblaient s'efforcer de ressembler aussi peu que possible aux honnêtes fonctions qui servent à quelque chose. Plus de continuité, ou bien de la continuité, mais pas de dérivées[...] Autrefois, quand on inventait une fonction nouvelle, c'était en vue de quelque but pratique ; aujourd'hui, on les invente tout exprès pour mettre en défaut les raisonnements de nos pères, et on n'en tirera jamais que cela. »

Henri Poincaré, *Science et méthodes* (1908)

Contrairement à ce qu'en pensait l'illustre Poincaré, on peut tirer beaucoup d'enseignements des contre-exemples monstrueux. Les fonctions continues et nulle-part dérivables ont connu leur heure de gloire par l'intermédiaire de Benoît Mandelbrot¹² et Felix Hausdorff qui ont grandement contribué à la caractérisation et à la compréhension des fractales. Cela a agrandi considérablement le champ des possibles tant en mathématiques qu'en physique. D'une manière plus subtile, le relâchement des hypothèses de régularité traditionnellement associées aux équations différentielles a profondément transformé la recherche en équations aux dérivées partielles et en systèmes dynamiques. On s'est mit à accepter toutes sortes de ces fonctions affreuses comme solutions (espaces de Sobolev, distributions de Schwartz,...), comme éléments constitutifs des problèmes (attracteurs fractals pour des systèmes dynamiques continus et discrets, chaos, entropie,...) ou comme réponses à des problèmes aléatoires (mouvement Brownien, processus càdlàg,...). Nous allons dans cette partie étudier quelques uns des nombreux outils qu'à produits l'analyse pour caractériser, exploiter ou provoquer la régularité.

12. Ce dernier a d'ailleurs écrit dans "Les dossiers de la recherche" la chose suivante à propos des fractales : « *Les fractales promettent une nouvelle science – celle de la rugosité . Toutes les sensations que nous avons ont à tour de rôle été domptées par la science : le poids par la mécanique, le chaud et le froid par la thermodynamique, le brillant par l'optique, le son par l'acoustique...* »

Mais il n'y avait pas de mesure numérique de la rugosité perçue par l'œil et la main, jusque ce que j'en publie une...

La nature n'est pas lisse, elle est rugueuse... Les nuages ne sont pas des sphères, les montagnes ne sont pas des cônes...

J'ai trouvé dans les fractales l'objet fondamental de la rugosité, comme la sinusoïde est l'objet fondamental de la lumière et du son. »

On admettra que la modestie n'était pas la principale qualité de Mandelbrot !

1. Rappels sur des outils usuels

1.a. Multi-indices

Définition 89 (Multi-indice). *On appelle multi-indice tout élément de l'ensemble*

$$\bigcup_{k \in \mathbb{N}} \mathbb{N}^k,$$

c'est-à-dire tout k -uplet $\alpha = (\alpha_1, \dots, \alpha_k) \in \mathbb{N}^k$ d'entiers naturels pour tout $k \in \mathbb{N}$.

Les multi-indices permettent d'écrire rapidement certaines choses sur les fonctions de plusieurs variables, comme l'exprime la définition suivante.

Définition 90 (Dérivée suivant un multi-indice). *On appelle dérivée suivant le multi-indice $\alpha = (\alpha_1, \dots, \alpha_n)$ et on note ∂^α l'opérateur*

$$\partial_{x_1}^{\alpha_1} \dots \partial_{x_n}^{\alpha_n}$$

agissant sur les fonctions de $\mathbb{R}^n$ pour lesquelles il est défini.

Par exemple, dans $\mathbb{R}^2$, dériver $f : (x, y) \mapsto f(x, y)$ suivant le multi-indice $\alpha = (1, 3)$ revient à calculer

$$\frac{\partial}{\partial x} \frac{\partial^3}{\partial y^3} f.$$

Pour un multi-indice α , on définit le module et la factorielle comme suit :

$$|\alpha| = \alpha_1 + \dots + \alpha_n \quad \text{et} \quad \alpha! = \alpha_1! \dots \alpha_n!,$$

on notera alors, pour la variable $x = (x_1, \dots, x_n) \in \mathbb{R}^n$,

$$x^\alpha := x_1^{\alpha_1} \dots x_n^{\alpha_n}.$$

Lorsque concernant un calcul dans $\mathbb{R}^n$ on écrit

$$\sum_{|\alpha| \leq k},$$

cela signifie que l'on fait la somme sur tous les multi-indices $\alpha = (\alpha_1, \dots, \alpha_n)$ à n éléments tels que

$$|\alpha| \leq k.$$

1.b. Fonctions tests et fonctions plateaux

Définition 91 (Support). *Soit Ω un ouvert de $\mathbb{R}^n$. Soit $\phi : \Omega \rightarrow \mathbb{C}$ une fonction. On appelle support de ϕ l'ensemble*

$$\text{supp}(\phi) = \overline{\{x \in \Omega \mid \phi(x) \neq 0\}}.$$

La fonction ϕ est dite à support compact s'il existe une partie compacte K de Ω telle que $\text{supp}(\phi) \subset K$.

On note $\mathcal{C}_0^\infty(\Omega)$ l'espace des applications $\mathcal{C}^\infty$ à support compact définies sur l'ouvert Ω . C'est un espace sur lequel il est difficile de faire de la topologie, mais il est possible de le munir de semi-normes qui le dotent d'une topologie non normable. Pour plus de détails sur ces aspects théoriques, on pourra consulter l'annexe A.

Ces fonctions sont utiles pour de nombreuses raisons. Le résultat suivant est très puissant en l'analyse, il permet de décomposer un problème sur des morceaux d'un ouvert.

Théorème 92 (*Partitions de l'unité*)

Soit $(\Omega_j)_{j \in J}$ une famille d'ouverts de Ω telle que $\cup_{j \in J} \Omega_j = \Omega$. Il existe une famille $(\phi_j)_{j \in J}$ de fonctions telle que,

- $\forall j \in J, \phi_j \in \mathcal{C}(\Omega)$ et $\text{supp}(\phi_j) \subset \Omega_j$ et $0 \leq \phi_j \leq 1$,
- Pour tout compact K de Ω , l'ensemble $\{j \in J \mid K \cap \text{supp}(\phi_j) \neq \emptyset\}$ est fini,
- $\forall x \in \Omega, \sum_{j \in J} \phi_j(x) = 1$.

Pour une partition de deux ouverts, on peut écrire un résultat plus précis, on parle alors de fonctions plateaux.

Théorème 93 (*Fonctions plateaux*)

Soit K un ensemble compact de Ω . Soit, pour $\varepsilon \in \mathbb{R}_+^*$, l'ensemble

$$K_\varepsilon = \{x \in \mathbb{R}^n \mid \text{dist}(x, K) < \varepsilon\}.$$

L'ensemble K_ε est un ouvert et il existe une fonction $\phi_\varepsilon \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ telle que

$$\forall x \in K, \phi_\varepsilon(x) = 1 \quad \text{et} \quad \forall x \in K_\varepsilon^c, \phi_\varepsilon(x) = 0,$$

où K_ε^c est le complémentaire de K_ε dans $\mathbb{R}^n$.

Ces résultats ne sont pas difficiles mais plutôt techniques à démontrer, et surtout ils ont le désavantage d'utiliser ce qui suit (la convolution notamment) dans leur démonstration. Cependant, il est plus clair de les énoncer ici et nous assurons au lecteur qu'il est possible de tout bien construire sans que le serpent ne se morde la queue.

1.c. Ouverts à bords réguliers

Travailler rigoureusement avec des ouverts à bords irréguliers est très difficile, aussi nous limiterons-nous à des ouverts de classe $\mathcal{C}^k$, $k \in \mathbb{N} \cup \{+\infty\}$, avec souvent $k > 0$, dans ce qui suit. Nous *arnaquerons* éventuellement en considérant certains ouverts polyédraux (carrés, rectangles,...) comme réguliers même s'ils ne le sont pas aux angles, et ce afin d'éviter de lourdes difficultés techniques.

Définir un ouvert régulier est assez simple :

Définition 94 (Bord, Extérieur, Ouvert de classe $\mathcal{C}^k$). Soit $k \in \mathbb{N} \cup \{+\infty\}$, $k > 0$, et Ω un ouvert de $\mathbb{R}^n$. On dit que Ω est de classe $\mathcal{C}^k$ s'il existe une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe $\mathcal{C}^k$ telle que

$$\Omega = \{x \in \mathbb{R}^n \mid f(x) < 0\},$$

et dont le gradient ∇f ne s'annule en aucun point où f s'annule. On appelle alors bord de l'ouvert Ω et on note $\partial\Omega$ l'ensemble

$$\partial\Omega = \{x \in \mathbb{R}^n \mid f(x) = 0\}.$$

On appelle extérieur de l'ouvert Ω l'ensemble

$$E(\Omega) = \{x \in \mathbb{R}^n \mid f(x) \geq 0\}.$$

L'hypothèse que le gradient ne s'annule pas là où f s'annule est très importante. En effet, les points où f s'annule sont ceux du bord de l'ouvert, cette hypothèse empêche donc que le bord soit trop épais : si f s'annule sur un ensemble contenant une boule, aussi petite soit-elle, le gradient de f est nul à l'intérieur de la boule. Cette condition revient à dire que la frontière se comporte comme une ligne, qu'il y a un intérieur et un extérieur et qu'on a un vecteur, le gradient de f , qui pointe depuis la frontière vers l'extérieur (car f est négative à l'intérieur puis positive à l'extérieur). On peut montrer que cette notion de bord coïncide avec la notion topologique de bord d'un ensemble ; pour de tels ouverts réguliers, nous avons bien

$$\partial\Omega = \overline{\Omega} \setminus \Omega^\circ.$$

Définition 95 (Normale unitaire extérieure (sortante)). Soit Ω un ouvert de classe $\mathcal{C}^k$, $k > 0$, défini par une fonction f . Pour tout point $x \in \partial\Omega$, on définit la normale unitaire extérieure (sortante) en x par

$$\nu(x) = \frac{\nabla f(x)}{\|\nabla f(x)\|}.$$

1.d. Mesure superficielle et formule de Green

On aimerait pouvoir intégrer des fonctions sur le bord des ouverts. Le problème est que pour la mesure de Lebesgue, le bord est de mesure nulle : la plupart des intégrales vont être nulles. Il convient alors de trouver une autre notion d'intégration.

Définition 96 (Intégrer sur le bord). Soit Ω un ouvert de classe $\mathcal{C}^k$, $k > 0$, défini par une fonction f . Soit $u : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction mesurable. On appelle intégrale sur le bord de Ω pour la mesure superficielle (ou juste intégrale sur le bord) de la fonction u la quantité, si elle existe,

$$\int_{\partial\Omega} u(x) \, d\sigma(x) = \lim_{\varepsilon \rightarrow 0} \int_{|f(x)| < \varepsilon} u(x) \frac{\|\nabla f(x)\|}{2\varepsilon} \, dx.$$

La formulation peut paraître obscure d'un premier abord mais l'idée est en fait intuitive : on intègre sur une bande de largeur 2ε autour du bord. Nous divisons donc par la largeur de cette bande afin de compenser. Lorsque la bande tend vers 0, nous obtenons une valeur de l'intégrale sur le bord. Le facteur en $\|\nabla f(x)\|$ provient du théorème de changement de variable.

Avec cette notion d'intégrale sur le bord, nous pouvons écrire deux résultats très importants :

Théorème 97

Soit Ω un ouvert de classe $\mathcal{C}^k$, $k > 0$, borné. Soit $u : \overline{\Omega} \rightarrow \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. Alors, pour tout $i \in \{1, \dots, n\}$,

$$\int_{\Omega} \frac{\partial u}{\partial x_i}(x) dx = \int_{\partial\Omega} u(x) \nu_i(x) d\sigma(x),$$

où $\nu(x) = (\nu_1(x), \dots, \nu_n(x))$ est la normale extérieure unitaire.

Démonstration : Soit f une fonction qui définit l'ouvert Ω . On peut prolonger u comme une fonction de classe $\mathcal{C}^1$ et à support compact sur $\mathbb{R}^n$, que l'on appelle encore u .

Soit la fonction g définie par

$$g(x) = \begin{cases} 1 & \text{si } x < -1, \\ \frac{1-x}{2} & \text{si } -1 \leq x \leq 1, \\ 0 & \text{si } x > 1. \end{cases}$$

On a

$$\int_{\Omega} \frac{\partial u}{\partial x_i}(x) dx = \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}^n} \frac{\partial u}{\partial x_i}(x) g\left(\frac{f(x)}{\varepsilon}\right) dx.$$

On écrit le taux d'accroissement pour u et on modifie l'expression pour faire apparaître un accroissement sur f : si $(e_1, \dots, e_n)$ est la base canonique de $\mathbb{R}^n$, pour tout $h \in \mathbb{R}^n$, on a par le changement de variable $y = x + he_i$,

$$\begin{aligned} & \int_{\mathbb{R}^n} \frac{u(x + he_i) - u(x)}{h} g\left(\frac{f(x)}{\varepsilon}\right) dx \\ &= \int_{\mathbb{R}^n} \frac{u(x + he_i)}{h} g\left(\frac{f(x)}{\varepsilon}\right) dx + \int_{\mathbb{R}^n} \frac{u(x)}{h} g\left(\frac{f(x)}{\varepsilon}\right) dx \\ &= \int_{\mathbb{R}^n} \frac{u(y)}{h} g\left(\frac{f(y - he_i)}{\varepsilon}\right) dy + \int_{\mathbb{R}^n} \frac{u(x)}{h} g\left(\frac{f(x)}{\varepsilon}\right) dx \\ &= \int_{\mathbb{R}^n} u(x) \frac{1}{h} \left(g\left(\frac{f(x - e_i)}{\varepsilon}\right) - g\left(\frac{f(x)}{\varepsilon}\right) \right) dx. \end{aligned}$$

La fonction $x \mapsto g\left(\frac{f(x)}{\varepsilon}\right)$ est de classe $\mathcal{C}^1$ en tout point où $\left|\frac{f(x)}{\varepsilon}\right| \neq 1$. De plus, si $\left|\frac{f(x)}{\varepsilon}\right| < 1$, on a

$$\frac{\partial}{\partial x_i} g\left(\frac{f(x)}{\varepsilon}\right) = -\frac{1}{2\varepsilon} \frac{\partial f}{\partial x_i}(x),$$

et si $|\frac{f(x)}{\varepsilon}| > 1$, on a

$$\frac{\partial}{\partial x_i} g\left(\frac{f(x)}{\varepsilon}\right) = 0.$$

Pour ε assez petit, les ensembles $\{x \in \mathbb{R}^n \mid \frac{f(x)}{\varepsilon} = 1\}$ et $\{x \in \mathbb{R}^n \mid \frac{f(x)}{\varepsilon} = -1\}$ sont de mesure nulle¹³ et on peut ne pas les considérer dans l'intégrale précédente. On prend alors la limite $h \rightarrow 0$ dans l'égalité obtenue, ce qui donne

$$\int_{\mathbb{R}^n} \frac{\partial u}{\partial x_i}(x) g\left(\frac{f(x)}{\varepsilon}\right) dx = \int_{|f(x)| < \varepsilon} u(x) \frac{1}{2\varepsilon} \frac{\partial f}{\partial x_i}(x) dx.$$

Or,

$$\int_{|f(x)| < \varepsilon} u(x) \frac{1}{2\varepsilon} \frac{\partial f}{\partial x_i}(x) dx = \int_{|f(x)| < \varepsilon} u(x) \nu_i(x) \frac{\|\nabla f(x)\|}{2\varepsilon} dx.$$

On en déduit par passage à la limite sur ε en 0, par définition de l'intégrale sur le bord, que

$$\int_{\Omega} \frac{\partial u}{\partial x_i}(x) dx = \int_{\partial\Omega} u(x) \nu_i(x) d\sigma(x). \quad \blacksquare$$

On peut étendre par des calculs spécifiques ou des vérifications simples ce théorème à quelques ouverts qui ne sont pas de classe $\mathcal{C}^1$ mais se comportent comme des ensembles civilisés : les carrés, les rectangles, les polygones réguliers, etc. De manière générale, ces notions d'intégrales surfaciques se généralisent bien à des ouverts dont le bord est localement le graphe d'une fonction lipschitzienne, ce qu'on appelle un ouvert à bord lipschitzien, mais c'est parfois très difficile à montrer en toute généralité.

À présent, nous pouvons facilement énoncer et démontrer un théorème classique et fondamental.

Théorème 98 (Théorème de la divergence)

Soit Ω un ouvert de classe $\mathcal{C}^k$, $k > 0$, borné défini par une fonction f . Soit $u : \bar{\Omega} \rightarrow \mathbb{R}^n$, $u = (u_1, \dots, u_n)$, une fonction de classe $\mathcal{C}^1$. Alors,

$$\int_{\Omega} \nabla \cdot u(x) dx = \int_{\partial\Omega} \langle u(x) \mid \nu(x) \rangle d\sigma(x),$$

où $\nabla \cdot u(x) = \frac{\partial u_1}{\partial x_1}(x) + \dots + \frac{\partial u_n}{\partial x_n}(x)$ est la divergence de u .

Démonstration : On utilise le théorème précédent pour chaque coordonnée et on somme :

$$\begin{aligned} \int_{\Omega} \nabla \cdot u(x) dx &= \sum_{i=1}^n \int_{\Omega} \frac{\partial u_i}{\partial x_i}(x) dx \\ &= \sum_{i=1}^n \int_{\partial\Omega} u_i(x) \nu_i(x) d\sigma(x) \\ &= \int_{\partial\Omega} \langle u(x) \mid \nu(x) \rangle d\sigma(x). \end{aligned}$$

13. C'est difficile à justifier proprement, alors nous l'admettons.

■

On trouve souvent ce résultat sous le nom de théorème de Green-Ostrogradski, ou juste théorème d'Ostrogradski (en physique surtout). L'intégrale à droite est appelée le flux de u à travers la surface. Ce théorème a un sens physique fort : si l'on intègre la divergence d'un champ de vecteurs dans un domaine ouvert, cette intégrale est égale au flux sortant à travers la frontière. C'est une forme de loi de conservation dont l'usage est central en électromagnétisme, en mécanique des fluides et dans bien d'autres domaines de la physique mathématique.

Une variante très utilisée de cette formule est la formule dite de Green pour le laplacien.

Théorème 99 (Formule de Green)

Soit Ω un ouvert de classe $\mathcal{C}^k$, $k > 0$, borné, défini par une fonction f . Soit $u, v : \bar{\Omega} \rightarrow \mathbb{R}$, deux fonctions, u de classe $\mathcal{C}^2$ et v de classe $\mathcal{C}^1$. Alors, on a

$$\int_{\Omega} \Delta u(x) v(x) dx = - \int_{\Omega} \langle \nabla u(x) \mid \nabla v(x) \rangle dx + \int_{\partial\Omega} \frac{\partial u}{\partial \nu} v(x) d\sigma(x),$$

où on a noté $\frac{\partial u}{\partial \nu}(x) = \langle \nabla u(x) \mid \nu(x) \rangle$ ce qui peut s'appeler dérivée de u sur la frontière.

Démonstration : On applique le théorème de la divergence pour la fonction de classe $\mathcal{C}^1$, à valeurs dans $\mathbb{R}^n$, $w = v\nabla u$:

$$\int_{\Omega} \nabla \cdot (v\nabla u)(x) dx = \int_{\partial\Omega} \langle v(x)\nabla u(x) \mid \nu(x) \rangle d\sigma(x) = \int_{\partial\Omega} v(x) \langle \nabla u(x) \mid \nu(x) \rangle d\sigma(x).$$

Or, pour tout $x \in \Omega$,

$$\nabla \cdot (v\nabla u)(x) = v(x)\Delta u(x) + \langle \nabla u(x) \mid \nabla v(x) \rangle,$$

d'où le résultat en posant $\frac{\partial u}{\partial \nu}(x) = \langle \nabla u(x) \mid \nu(x) \rangle$. ■

2. Produit de convolution et régularisation

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et soit $\phi \in \mathcal{C}_0^\infty(\Omega)$. Pour tout $y \in \mathbb{R}^n$, la fonction $x \mapsto f(x)\phi(y-x)$ est à support compact, donc intégrable sur $\mathbb{R}^n$. On peut alors écrire la définition suivante.

Définition 100 (Produit de convolution). *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et soit $\phi \in \mathcal{C}_0^\infty(\Omega)$, on appelle produit de convolution de f et ϕ et on note $f * \phi$ la fonction,*

$$f * \phi : y \mapsto \int_{\mathbb{R}^n} f(x)\phi(y-x) dx.$$

Propriété 101. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et soit $\phi \in \mathcal{C}_0^\infty(\Omega)$. Pour $\psi \in \mathcal{C}_0^\infty(\Omega)$ on a $\psi * \phi = \phi * \psi$, $\phi * \psi \in \mathcal{C}_0^\infty(\Omega)$ et

$$(f * \phi) * \psi = f * (\phi * \psi).$$

Plus généralement, on appelle produit de convolution de deux fonctions, s'il est défini, l'opération qui à deux fonctions (ou classes d'équivalence de fonctions) f et g de $\mathbb{R}^n$ dans $\mathbb{R}$ associe la fonction (ou classe d'équivalence de fonctions)

$$f * g : y \mapsto \int_{\mathbb{R}^n} f(x)g(y-x) dx.$$

Il y a de nombreux cas où le produit de convolution de deux fonctions est bien défini, comme l'exprime le théorème suivant.

Théorème 102 (Existence de la convolution)

- Soit $f \in L^1(\mathbb{R}^n)$ et $g \in L^\infty(\mathbb{R}^n)$, le produit de convolution $f * g$ existe presque partout, est une (classe d'équivalence de) fonction(s) appartenant à $L^\infty(\mathbb{R}^n)$ et,

$$\|f * g\|_\infty \leq \|f\|_1 \|g\|_\infty.$$

- Soit $f, g \in L^1(\mathbb{R}^n)$, le produit de convolution $f * g$ existe presque partout, est une (classe d'équivalence de) fonction(s) appartenant à $L^1(\mathbb{R}^n)$ et,

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

- Soit $f \in L^p(\mathbb{R}^n)$ et $g \in L^q(\mathbb{R}^n)$, où $p, q \in]1, +\infty[$ vérifient $\frac{1}{p} + \frac{1}{q} = 1$, alors le produit de convolution $f * g$ existe presque partout, est une (classe d'équivalence de) fonction(s) appartenant à $L^\infty(\mathbb{R}^n)$ et,

$$\|f * g\|_\infty \leq \|f\|_p \|g\|_q.$$

- Soit $f \in L^p(\mathbb{R}^n)$ et $g \in L^q(\mathbb{R}^n)$, où $p, q, r \in]1, +\infty[$ vérifient $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, alors le produit de convolution $f * g$ existe presque partout, est une (classe d'équivalence de) fonction(s) appartenant à $L^1(\mathbb{R}^n)$ et,

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

On pouvait écrire uniquement le dernier cas en autorisant $p, q, r \in [1, +\infty]$ à l'exception du cas $p = q = +\infty$, mais c'eût été moins clair et moins pédagogique. De plus, la démonstration distingue les différents cas. On retiendra que la convolution marche très bien pour les fonctions des espaces L^p . Le cas $p = q = 1$ a un côté un peu magique, mais c'est juste un avatar du théorème de Fubini.

L'un des principaux (et nombreux) intérêts du produit de convolution est qu'il a un effet régularisant : le produit de convolution d'une fonction irrégulière avec une fonction régulière est régulier. Mieux encore, le produit de convolution de deux fonctions irrégulières peut être irrégulier.

Théorème 103

- Si $f \in L^1(\mathbb{R}^n)$ et $g \in L^1(\mathbb{R}^n)$ alors $f * g \in \mathcal{C}^0(\mathbb{R}^n)$ et $f * g = g * f$.
- Si $f \in L^p(\mathbb{R}^n)$ et $g \in L^q(\mathbb{R}^n)$ avec $p, q > 1$ et $\frac{1}{p} + \frac{1}{q} = 1$, alors $f * g \in \mathcal{C}^0(\mathbb{R}^n)$ et tend vers 0 à l'infini.
- Si $f \in L^1(\mathbb{R}^n)$ et $g \in L^\infty(\mathbb{R}^n)$ alors $f * g \in \mathcal{C}^0(\mathbb{R}^n)$ et est bornée.
- Si $f \in L^p(\mathbb{R}^n)$ avec $p \in [1, +\infty]$ et $g \in \mathcal{C}_0^k(\mathbb{R}^n)$, alors $f * g \in \mathcal{C}_0^k(\mathbb{R}^n)$ et pour tout multi-indice α de module inférieur à k ,

$$\partial^\alpha (f * g) = f * \partial^\alpha g.$$

La convolution permet aussi d'approcher des fonctions irrégulières par des fonctions régulières, ce qui fournit de nombreux théorèmes de densité.

Théorème 104

Soit $\phi \in \mathcal{C}_0^\infty(\Omega)$ positive et d'intégrale 1 (une densité de probabilité en somme). On pose pour tout $\varepsilon \in \mathbb{R}_+$,

$$\phi_\varepsilon(x) = \varepsilon^{-n} \phi\left(\frac{x}{\varepsilon}\right).$$

Alors, pour tout $f \in L^p(\mathbb{R}^n)$, $p \in [1, +\infty[$,

$$\lim_{\varepsilon \rightarrow 0^+} \|f * \phi_\varepsilon - f\|_p = 0$$

Corollaire 105. Pour tout $p \in [1, +\infty[$, pour tout $k \in \mathbb{R}_+$, le sous-ensemble $\mathcal{C}_0^k(\mathbb{R}^n)$ est dense dans $L^p(\mathbb{R}^n)$ muni de sa norme naturelle.

3. Transformée de Fourier

La transformée de Fourier est un outil puissant pour faire de l'analyse. Elle permet de simplifier des problèmes différentiels ou de trouver des solutions à certaines équations aux dérivées partielles. Pour gagner en lisibilité, nous noterons parfois $x \cdot y$ le produit scalaire standard de x et y dans $\mathbb{R}^n$ au cours de cette sous-section.

Définition 106 (Espace de Schwartz). On note $\mathcal{S}(\mathbb{R}^n)$ l'ensemble des fonctions $f: \mathbb{R}^n \rightarrow \mathbb{C}$ de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^n$ et telles que

$$\forall k \in \mathbb{N}, \forall \alpha, \beta \in \mathbb{N}^k, \quad \sup_{x \in \mathbb{R}^n} |x^\alpha \partial^\beta f(x)| < +\infty.$$

On dit que $\mathcal{S}(\mathbb{R}^n)$ est l'espace de Schwartz des fonctions indéfiniment dérivables à décroissance rapide.

On munit l'espace $\mathcal{S}(\mathbb{R}^n)$ des semi-normes ainsi définies lorsque α et β parcourent $\mathbb{N}^k$, pour $k \in \mathbb{N}$:

$$\forall f \in \mathcal{S}(\mathbb{R}^n), \quad N_{\alpha,\beta}(f) := \sup_{x \in \mathbb{R}^n} |x^\alpha \partial^\beta f(x)|$$

Définition 107 (Transformée de Fourier). *Soit $f \in \mathcal{S}(\mathbb{R}^n)$. On appelle transformée de Fourier de f l'application $\hat{f}$ définie par*

$$\forall \xi \in \mathbb{R}^n, \quad \hat{f}(\xi) := \int_{\mathbb{R}^n} e^{-ix \cdot \xi} f(x) dx,$$

où $\cdot$ désigne le produit scalaire usuel dans $\mathbb{R}^n$ et transformation de Fourier l'application (linéaire) suivante :

$$\begin{array}{ccc} \mathcal{F} : \mathcal{S}(\mathbb{R}^n) & \longrightarrow & \mathcal{S}(\mathbb{R}^n) \\ f & \longmapsto & \hat{f}. \end{array}$$

On peut montrer que pour toutes fonctions f et g de l'espace de Schwartz $\mathcal{S}(\mathbb{R})$ et tout multi-indice $\alpha \in \mathbb{N}^k$, $\partial^\alpha f \in \mathcal{S}(\mathbb{R})$ et $x^\alpha f \in \mathcal{S}(\mathbb{R})$ d'une part et $f \times g \in \mathcal{S}(\mathbb{R})$ et $f * g \in \mathcal{S}(\mathbb{R})$ d'autre part, et on a les propriétés fondamentales suivantes :

Propriété 108. *L'espace de Schwartz vérifie :*

- i. $\mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$ et $f \in \mathcal{S}(\mathbb{R}) \implies \mathcal{F}f \in \mathcal{S}(\mathbb{R})$;
- ii. $\mathcal{F}$ réalise une bijection de $\mathcal{S}(\mathbb{R})$ sur lui-même ;
- iii. Pour tout $f, g \in \mathcal{S}(\mathbb{R})$,

$$\mathcal{F}(f * g) = (\mathcal{F}f) \times (\mathcal{F}g) \quad \text{et} \quad \mathcal{F}(f \times g) = (\mathcal{F}f) * (\mathcal{F}g) ;$$

- iv. Pour tout $f \in \mathcal{S}(\mathbb{R})$, $j \in \llbracket 1, n \rrbracket$ et $\xi \in \mathbb{R}^n$,

$$\widehat{\partial_j f}(\xi) = i \xi_j \hat{f}(\xi) \quad \text{et} \quad \widehat{x_j f}(\xi) = i \partial_j \hat{f}(\xi),$$

et en notant pour tout $x, y \in \mathbb{R}^n$, $\tau_y f(x) := f(x + y)$,

$$\widehat{\tau_y f}(\xi) = e^{iy \cdot \xi} \hat{f}(\xi) \quad \text{et} \quad \widehat{e^{-x \cdot y} f}(\xi) = \tau_y \hat{f}(\xi).$$

Enfin, pour les semi-normes définies ci-dessus, on le résultat suivant :

Propriété 109. *la transformation de Fourier est continue sur $\mathcal{S}(\mathbb{R})$,*

où la continuité doit être comprise au sens de la topologie décrite par les semi-normes (voir la définition topologique de la continuité et l'exemple des semi-normes pour les fonctions lisses à support compact dans l'Annexe A.).

On a même plus fort :

Théorème 110

La transformation de Fourier est continue et bijective de $\mathcal{S}(\mathbb{R})$ dans lui-même, de bijection réciproque vérifiant

$$\forall \xi \in \mathbb{R}^n, \quad \mathcal{F}^{-1} f(\xi) = \frac{1}{(2\pi)^n} \mathcal{F} f(-\xi) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\xi \cdot x} f(x) dx.$$

Il est possible de définir la transformée de Fourier pour toute fonction mesurable et intégrable, c'est-à-dire sur l'espace $L^1(\mathbb{R}^n)$. Malheureusement, ce n'est pas une bijection de $L^1(\mathbb{R}^n)$ dans lui-même.

Il convient alors de remarquer que l'espace

$$L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$$

est dense dans $L^2(\mathbb{R}^n)$ pour la norme naturelle. On peut donc définir par densité une transformée de Fourier de L^2 dans L^2 . Cette fois, on tient une bijection comme dans le cas de l'espace de Schwartz, et il est beaucoup plus facile de définir des notions de continuité pour l'application transformée de Fourier car $L^2(\mathbb{R}^n)$ est un espace de Hilbert. Énonçons proprement quelques théorèmes et remarquons quelques faits, mais avec une convention différente pour la transformée de Fourier :

Définition 111. On appelle transformée de Fourier d'une fonction $f \in L^1(\mathbb{R}^n)$ la fonction définie par

$$\hat{f}(x) = \int_{\mathbb{R}^n} e^{-2i\pi x \cdot y} f(y) dy.$$

Cette nouvelle convention, avec le $2i\pi$ dans l'exponentielle, rend certains théorèmes plus lisibles dans le cas qui nous intéresse : les fonctions de carré intégrable.

Théorème 112

Soit $f \in L^1(\mathbb{R}^n)$. La fonction $\hat{f}$ est continue sur $\mathbb{R}^n$ et tend vers 0 en l'infini.

Théorème 113

Soit $f, g \in L^1(\mathbb{R}^n)$. On a $\widehat{f * g} = \hat{f} \hat{g}$.

Théorème 114

La transformée de Fourier est linéaire et continue de $L^1(\mathbb{R}^n)$ dans l'ensemble des fonctions continues et bornées muni de la norme de la convergence uniforme $\|\cdot\|_\infty$. Elle est également injective.

Remarque 115. Elle n'est pas surjective dans les fonctions continues qui tendent vers 0 en l'infini. Son image est par contre dense dans cet espace.

Et maintenant sur L^2 , nous allons voir que la notion est moins intuitive mais beaucoup plus puissante.

Théorème 116

La transformation de Fourier sur $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ se prolonge de manière unique en un isomorphisme d'espaces de Hilbert (continue, bijective, bijection continue, respecte le produit scalaire) de $L^2(\mathbb{R}^n)$ dans lui-même. On a en particulier, pour toute fonction $f \in L^2(\mathbb{R}^n)$,

$$\forall x \in \mathbb{R}^n, \quad \mathcal{F}^{-1}f(x) = \mathcal{F}f(-x) = \int_{\mathbb{R}^n} e^{2i\pi x \cdot y} f(y) dy$$

Remarque 117. Respecter le produit scalaire signifie que $\langle f | g \rangle = \langle \hat{f} | \hat{g} \rangle$. C'est donc en particulier une isométrie sur cet espace : $\|f\|_2 = \|\hat{f}\|_2$.

4. Espaces de Sobolev

4.a. Dérivées faibles

Comme nous le verrons dans les sections suivantes, certains problèmes ne sont pas adaptés à la recherche de solutions régulières. Il est donc intéressant de définir l'opération de dérivation d'une manière moins rigide. Remarquons que si une fonction f est de classe $\mathcal{C}^1$ sur $\mathbb{R}$, alors pour toute fonction régulière à support compact $\phi \in \mathcal{C}_0^\infty(\mathbb{R})$, la formule d'intégration par parties sur un intervalle $[-R, R]$ nous donne, après passage à la limite $R \rightarrow +\infty$,

$$\int_{\mathbb{R}} f'(x)\phi(x) dx = - \int_{\mathbb{R}} f(x)\phi'(x) dx.$$

Soit la fonction $H = 1_{\mathbb{R}_+}$ la fonction indicatrice de $\mathbb{R}_+$, aussi appelée fonction de Heaviside, qui vaut 0 pour $x < 0$ et 1 pour $x \geq 0$. Soit la fonction $F : x \mapsto xH(x)$, qui est continue sur $\mathbb{R}$, dérivable sur $\mathbb{R}_+^*$, dérivable sur $\mathbb{R}_-^*$, mais pas dérivable en 0. Pour toute fonction test $\phi \in \mathcal{C}_0^\infty(\mathbb{R})$, on calcule,

$$\int_{\mathbb{R}} F(x)\phi'(x) dx = \int_{\mathbb{R}_+^*} x\phi'(x) dx = - \int_{\mathbb{R}_+^*} \phi(x) dx = - \int_{\mathbb{R}} H(x)\phi(x) dx,$$

où l'on a utilisé l'intégration par partie sur $\mathbb{R}_+^*$ pour les fonctions de classe $\mathcal{C}^1$ $x \mapsto x$ et ϕ . On voit donc que la fonction H se comporte un peu comme une dérivée de F vis-à-vis des fonctions test. Nous allons faire de cela une définition qui généralise la notion de dérivée à des fonctions qui ne sont pas dérivables.

Définition 118 (Dérivée faible). *Soit $F \in L^1(\mathbb{R})$. On dit que f est une dérivée faible de F si pour toute fonction $\phi \in \mathcal{C}_0^\infty(\mathbb{R})$,*

$$\int_{\mathbb{R}} f(x)\phi(x) \, dx = - \int_{\mathbb{R}} F(x)\phi'(x) \, dx.$$

Par contraste, on appelle parfois la dérivée standard *dérivée forte*.

On peut généraliser encore : comme on intègre contre des fonctions à support compact, on n'a besoin que d'être localement intégrable, donc seulement dans $L^1_{\text{loc}}(\mathbb{R})$. Il est aussi possible de définir la notion pour les dérivées partielles en dimension supérieure. Il est également naturel de chercher des dérivées faibles pour des fonctions définies seulement sur un ouvert Ω de $\mathbb{R}^n$. Cela nous amène à la définition suivante.

Définition 119 (Dérivée faible selon un vecteur de base). *Soit Ω un ouvert de $\mathbb{R}^n$ et $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$. Soit $F \in L^1_{\text{loc}}(\Omega)$. On dit que $f_i \in L^1_{\text{loc}}(\Omega)$ est une dérivée faible de F suivant le vecteur e_i si pour toute fonction $\phi \in \mathcal{C}_0^\infty(\Omega)$,*

$$\int_{\Omega} f_i(x)\phi(x) \, dx = - \int_{\Omega} F(x) \frac{\partial \phi}{\partial x_i}(x) \, dx.$$

Propriété 120. *Soit Ω un ouvert de $\mathbb{R}^n$ et $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$. Soit $F \in L^1_{\text{loc}}(\Omega)$. Si il existe une dérivée faible suivant un vecteur e_i , elle est unique. Si la fonction F est de classe $\mathcal{C}^1$, la dérivée forte suivant le vecteur e_i coïncide avec la dérivée faible presque partout.*

Notons que les fonctions F et f sont des classes d'équivalences de fonctions pour la relation d'égalité presque partout et que l'unicité est à comprendre en ce sens. On peut bien évidemment changer la valeur de la dérivée faible en un nombre fini (et même infini dénombrable¹⁴) et obtenir une nouvelle dérivée faible de F . Cependant, leur classe d'équivalence est la même, et on peut donc dire en ce sens que la dérivée faible est unique.

Définition 121 (Dérivée faible selon un multi-indice). *Soit Ω un ouvert de $\mathbb{R}^n$ et $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$. Soit $F \in L^1_{\text{loc}}(\Omega)$. Soit $\alpha = (\alpha_1, \dots, \alpha_n)$ un multi-indice. Si cela existe, on appellera dérivée faible suivant α de f et on notera $\partial^\alpha f$ la dérivée faible de f suivant le vecteur e_1 appliquée α_1 fois, puis selon le vecteur e_2 appliquée α_2 fois, etc.*

La propriété précédente implique de manière évidente que les dérivées faibles suivant les multi-indices sont uniques si elles existent.

Remarque 122 (Abus de notation). *Afin de ne pas alourdir les théorèmes par des expressions du type "si la dérivée faible de f suivant le multi-indice α existe et appartient à A ", on considérera implicitement que la formule $\partial_\alpha f \in A$ signifie la même chose.*

Ce n'est pas vraiment un abus puisque la dérivée faible existe en fait toujours dans un espace appelé espace des distributions sur Ω .

14. Ou pire, l'ensemble de Cantor est infini indénombrable et de mesure nulle...

4.b. Les espaces de Sobolev

Avec cette notion de dérivée faible, on s'intéresse naturellement aux ensembles de fonctions qui admettent des dérivées faibles utilisables dans les calculs. Par exemple, si une fonction est dans L^1 , on a envie que sa dérivée faible y soit aussi, ce qui conduit très naturellement à poser la définition suivante.

Définition 123 (TRÈS IMPORTANTE – Espace de Sobolev). *Soit $p \in [1, +\infty]$. On appelle espace de Sobolev et on note $W^{1,p}(\mathbb{R})$ l'ensemble*

$$W^{1,p} = \{f \in L^p(\mathbb{R}) \mid f' \in L^p(\mathbb{R})\},$$

où f' est ici la dérivée faible. Dans le cas $p = 2$, on adoptera la notation particulière

$$H^1 = W^{1,2}.$$

On peut munir $W^{1,p}$ de la norme $\|f\| = \|f\|_p + \|f'\|_p$ ou de la norme $\|f\| = \left(\|f\|_p^p + \|f'\|_p^p\right)^{\frac{1}{p}}$ qui sont équivalentes, cela en fait un espace de Banach. Pour $p = 2$, on a le produit scalaire sur $H^1(\mathbb{R})$

$$\langle f \mid g \rangle = \int_{\mathbb{R}} (f(x)g(x) + f'(x)g'(x)) \, dx,$$

qui en fait un espace de Hilbert.

On peut généraliser cette notion à une dimension quelconque et à des fonctions définies seulement sur un ouvert de $\mathbb{R}^n$.

Définition 124 (TRÈS IMPORTANTE, contient 123 – Espace de Sobolev). *Soit $(m, p) \in \mathbb{N} \times [1, +\infty]$ et Ω un ouvert de $\mathbb{R}^n$. On appelle espace de Sobolev et on note $W^{m,p}(\Omega)$ l'ensemble*

$$W^{m,p} = \{f \in L^p(\Omega) \mid \forall |\alpha| \leq m, \partial^\alpha f \in L^p(\Omega)\}.$$

Dans le cas $p = 2$, on adoptera la notation particulière

$$H^m = W^{m,2}.$$

La structure d'espace de Banach se généralise aussi :

Théorème 125

Soit $(m, p) \in \mathbb{N} \times [1, +\infty[$. Si l'on munit l'espace $W^{m,p}(\Omega)$ de la norme

$$\|\cdot\|_{W^{m,p}} = \sum_{|\alpha| \leq m} \|\partial^\alpha u\|_{L^p},$$

ou de la norme équivalente

$$\|\cdot\|_{W^{m,p}} = \left(\sum_{|\alpha| \leq m} \|\partial^\alpha u\|_{L^p}^p \right)^{\frac{1}{p}},$$

on obtient un espace de Banach séparable.

Si l'on munit l'espace $H^m(\Omega)$ du produit scalaire défini par

$$\langle f | g \rangle = \sum_{|\alpha| \leq m} \langle \partial^\alpha f | \partial^\alpha g \rangle_{L^2},$$

on obtient un espace de Hilbert séparable dont la norme est équivalente à la norme que nous venons de mettre sur $W^{m,2}$.

On peut évidemment également munir l'espace d'une norme qui en fait un espace de Banach dans le cas $p = \infty$, en prenant la somme des normes infinies. Il nous reste à introduire un dernier petit espace, utile dans la résolution d'équations aux dérivées partielles et dans la théorie de l'approximation numérique par les éléments finis des solutions d'équations paraboliques ou elliptiques.

Définition 126 (Adhérence des fonctions test (dans H^1)). *Soit Ω un ouvert de $\mathbb{R}^n$. On définit l'espace $H_0^1(\Omega)$ comme l'adhérence de l'espace $\mathcal{C}_0^\infty(\Omega)$ des fonctions de classe $\mathcal{C}^\infty$ à support compact dans l'espace de Sobolev $H^1(\Omega)$, où l'adhérence est prise pour la norme naturelle de $H^1(\Omega)$, c'est-à-dire,*

$$H_0^1(\Omega) = \overline{\mathcal{C}_0^\infty(\Omega)}^{H^1}.$$

Remarque 127. Dans le cas particulier $\Omega \subset \mathbb{R}^n$, on a en fait

$$H_0^1(\mathbb{R}^n) = H^1(\mathbb{R}^n).$$

L'espace $H_0^1(\Omega)$ est donc un outil prévu pour l'étude de problèmes dans des ouverts bornés ou possédant un bord non vide, avec une condition de Dirichlet au bord (les solutions s'annulent sur le bord de Ω).

On définit de même

Définition 128 (Adhérence des fonctions test (dans $W^{1,p}$)). *Soit Ω un ouvert de $\mathbb{R}^n$. On définit l'espace $W_0^{1,p}(\Omega)$ comme l'adhérence de l'espace $\mathcal{C}_0^\infty(\Omega)$ dans l'espace de Sobolev $W^{1,p}(\Omega)$, où l'adhérence est prise pour la norme naturelle de $W^{1,p}(\Omega)$, c'est-à-dire,*

$$W_0^{1,p}(\Omega) = \overline{\mathcal{C}_0^\infty(\Omega)}^{W^{1,p}}.$$

4.c. Injections et inégalités

Dans le cadre des espaces de Sobolev, on s'intéresse beaucoup aux inclusions des espaces les uns dans les autres. Par exemple toute fonction de H^1 est une

fonction de L^2 , on dit que H^1 s'injecte dans L^2 et on note $H^1 \hookrightarrow L^2$. Nous allons nous intéresser à des cas particuliers utiles et à des inégalités entre normes, fonctions, dérivées. Ce sera surtout utile pour travailler en équations aux dérivées partielles.

En voici un exemple dont l'usage est répandu :

Théorème 129 (Théorème de Rellich-Kondrakov)

Pour tout ouvert Ω de $\mathbb{R}^n$ borné et de classe $\mathcal{C}^1$, l'injection naturelle $i : H^1(\Omega) \hookrightarrow L^2(\Omega)$ est un opérateur compact.

En particulier, cela implique que toute suite bornée dans $H^1(\Omega)$ ¹⁵ possède une sous-suite qui converge fortement (pour la norme $\|\cdot\|_2$) dans $L^2(\Omega)$.

Il est utile de pouvoir borner une fonction à partir de ses dérivées, c'est ce que permet la très classique

Théorème 130 (Inégalité de Poincaré)

Pour tout ouvert Ω de $\mathbb{R}^n$ borné et de classe $\mathcal{C}^1$, pour tout $p \in [1, +\infty[$, il existe une constante C ne dépendant que de p et Ω telle que pour toute fonction $f \in W_0^{1,p}(\Omega)$, on a

$$\|f\|_{L^p} \leq C \|\nabla f\|_{L^p},$$

où $\|\nabla f\|_{L^p} = (\|\partial_{x_1} f\|_{L^p}^p + \cdots + \|\partial_{x_n} f\|_{L^p}^p)^{\frac{1}{p}}$.

Ce résultat peut s'étendre au cas d'un ouvert non borné mais borné dans une direction, comme un bande infinie. Par contre, la régularité du bord est importante. Des extensions existent pour des ouverts à frontière lipschitzienne, mais pour des ouverts à géométrie fractale, les choses se compliquent. Le résultat est évidemment faux pour les fonctions de $H^1(\Omega)$ ou de $W^{1,p}(\Omega)$, comme on peut s'en convaincre en considérant la fonction identiquement égale à 1 sur Ω .

On peut légitimement se demander si les fonctions des espaces de Sobolev ne sont pas parfois plus régulières que simplement mesurables et intégrables à la puissance p . C'est en fait le cas. Toutes les fonctions de $W^{1,p}(\mathbb{R})$ sont continues. Par contre, toutes les fonctions de $W^{1,p}(\mathbb{R}^n)$ ne sont pas continues. En fait, on a l'injection suivante.

Théorème 131

Si $\Omega \subset \mathbb{R}^n$ est borné, si $p \in [2, +\infty[$ et si $m \geq k + \frac{n}{p}$,

$$W^{m,p}(\Omega) \hookrightarrow \mathcal{C}^k(\Omega).$$

15. c'est-à-dire faiblement convergente, voir dualité, réflexivité, faible-* compacité des boules dans l'annexe A..

On peut montrer ainsi que

$$\bigcap_{m \in \mathbb{N}} W^{m,p}(\Omega) \subset \mathcal{C}^\infty(\Omega).$$

4.d. Traces et formule de Green

Puisque l'on peut dériver au sens faibles les fonctions des espaces de Sobolev, on a envie d'appliquer la formule de Green. Le problème, c'est que la construction de l'intégrale de surface que nous avons faite précédemment ne fonctionne que pour des fonctions régulières. Pire encore, la notion de restriction d'une fonction de $W^{1,p}(\Omega)$ sur le bord de Ω n'a aucun sens puisque le bord est de mesure nulle et que les fonctions des espaces de Sobolev sont définies, comme toutes les fonctions de L^p , à un ensemble de mesure nulle près.

La solution à ce problème n'est pas difficile mais très technique et nous ne la présentons ici que grossièrement : il s'agit de l'opérateur trace. On va définir un opérateur qui envoie une fonction de $H^1(\Omega)$ sur une fonction de $L^2(\partial\Omega)$ qui représente sa restriction sur le bord. On admet qu'il est possible de définir de façon satisfaisante une intégrale de surface sur le bord et donc un espace $L^2(\partial\Omega)$ des fonctions sur le bord de carré intégrable.

Théorème 132

Soit Ω un ouvert borné de $\mathbb{R}^n$ de classe $\mathcal{C}^k$, $k > 0$, alors $\mathcal{C}^\infty(\overline{\Omega})$ est dense dans $H^1(\Omega)$ et l'application qui à $u \in \mathcal{C}^\infty(\overline{\Omega})$ associe la restriction sur le bord $\gamma_0(u) = u|_{\partial\Omega}$ se prolonge en un opérateur linéaire et continu de $H^1(\Omega)$ dans $L^2(\partial\Omega)$, encore noté γ_0 , appelé *opérateur de trace*.

Le théorème s'étend à des ouverts qui ne sont pas de classe $\mathcal{C}^1$, mais le résultat devient plus technique à énoncer. On retiendra, pour changer, que dans le cas d'un carré ça ne se passe pas trop mal.

L'intérêt de cette formule des traces est que l'on retrouve nos formules préférées.

Théorème 133 (Intégration par partie dans H^1)

Soit Ω un ouvert borné de $\mathbb{R}^n$ de classe $\mathcal{C}^k$, $k > 0$. Soit γ_0 l'opérateur trace de $H^1(\Omega)$ dans $L^2(\partial\Omega)$. Pour tout $u, v \in H^1(\Omega)$,

$$\int_{\Omega} u(x) \frac{\partial v}{\partial x_i}(x) dx = - \int_{\Omega} \frac{\partial u}{\partial x_i}(x) v(x) dx + \int_{\partial\Omega} \gamma_0(u)(x) \gamma_0(v)(x) \nu_i d\sigma(x),$$

où $\nu = (\nu_1, \dots, \nu_n)$ est la normale unitaire à $\partial\Omega$ dirigée vers l'extérieur de Ω .

Comme $H^1 \subset H^2$, on peut appliquer l'opérateur trace dans H^2 et généraliser la formule de Green pour le laplacien.

Théorème 134 (Formule de Green pour le laplacien dans H^2)

Soit Ω un ouvert borné de $\mathbb{R}^n$ de classe $\mathcal{C}^k$, $k > 0$. Soit γ_0 l'opérateur trace de $H^1(\Omega)$ dans $L^2(\partial\Omega)$. Pour tous $u \in H^2(\Omega)$, $v \in H^1(\Omega)$,

$$\int_{\Omega} \Delta u(x) v(x) \, dx = - \int_{\Omega} \langle \nabla u(x) \mid \nabla v(x) \rangle \, dx + \int_{\partial\Omega} \frac{\partial u}{\partial \nu}(x) v(x) \, d\sigma(x),$$

où on a noté $\frac{\partial u}{\partial \nu}(x) = \langle \gamma_0(\nabla u(x)) \mid \nu(x) \rangle$, où on a écrit par abus de notation v à la place de $\gamma_0(v)$.

En général, on n'écrit pas l'application trace dans les formules d'intégration afin de gagner en lisibilité. Il faut prendre garde d'interpréter les intégrales de surface au sens d'une intégrale de la trace chaque fois que c'est nécessaire.

5. Valeurs propres du laplacien

Soit Ω un ouvert **borné** de $\mathbb{R}^n$, de classe $\mathcal{C}^1$. Posons $E = \{u \in H_0^1(\Omega) \mid \Delta u \in L^2(\Omega)\}$, muni de la norme $\|\cdot\|_E = \|\cdot\|_{H_0^1} + \|\cdot\|_{L^2}$. De manière évidente, l'application

$$-\Delta : E \rightarrow L^2(\Omega),$$

est linéaire et continue pour les normes choisies. On peut montrer (c'est un peu long), le théorème suivant :

Théorème 135

L'opérateur $-\Delta : E \rightarrow L^2$ est un isomorphisme d'espaces vectoriels.

C'est en particulier une bijection. Notons alors $T = (-\Delta)^{-1}$. C'est une application de L^2 dans E linéaire et continue. E s'injecte dans $L^2(\Omega)$ donc T peut être vu comme une application de $L^2(\Omega)$ dans lui-même. On a le résultat suivant.

Théorème 136

L'application $T = (-\Delta)^{-1} : L^2(\Omega) \rightarrow L^2(\Omega)$ est un opérateur (donc continu) compact, auto-adjoint et défini positif.

Démonstration : T est un opérateur de L^2 dans E . Or l'application $u \mapsto u$ de E dans $H_0^1(\Omega)$ est continue par définition de la norme de E et l'application $u \mapsto u$ de $H_0^1(\Omega)$ dans $L^2(\Omega)$ est compacte d'après le théorème de Rellich-Kondrachov (car $H_0^1(\Omega)$ s'injecte continûment dans $H^1(\Omega)$ et que la composée d'un opérateur avec un opérateur compact est un opérateur compact).

Comme la composée d'une application linéaire continue avec une application linéaire continue compacte est une application linéaire continue compacte (propriété d'idéal pour la compacité), on en déduit que T est linéaire continue et compacte de $L^2(\Omega)$ dans lui-même.

Il reste à voir que T est auto-adjoint.

Soit $f, g \in L^2$. On note $u = Tf$ et $v = Tg$, ce qui donne par définition de T , $-\Delta u = f$ et $-\Delta v = g$. On a alors

$$\langle Tf | g \rangle_{L^2} = \langle u | -\Delta v \rangle_{L^2} = - \langle u | \Delta v \rangle_{L^2}.$$

Montrons que

$$- \langle u | \Delta v \rangle_{L^2} = \sum_{i=1}^n \left\langle \frac{\partial u}{\partial x_i} \left| \frac{\partial v}{\partial x_i} \right. \right\rangle_{L^2}.$$

Si v est dans E et $u \in \mathcal{C}_0^\infty(\Omega)$, on a (attention de distinguer les crochets de produit scalaire des crochets d'application d'une distribution naturellement associée à une fonction)

$$\begin{aligned} \left\langle u \left| \frac{\partial^2 v}{\partial x_i^2} \right. \right\rangle_{L^2} &= \left\langle \frac{\partial^2 v}{\partial x_i^2} \left| \bar{u} \right. \right\rangle_{L^2} \\ &= \left\langle \frac{\partial^2 \bar{v}}{\partial x_i^2} \left| \bar{u} \right. \right\rangle_{L^2} \\ &= \left\langle \frac{\partial^2 \bar{v}}{\partial x_i^2} \left| \bar{u} \right. \right\rangle_{\mathcal{D}'(\Omega)} \\ &= - \left\langle \frac{\partial \bar{v}}{\partial x_i} \left| \frac{\partial \bar{u}}{\partial x_i} \right. \right\rangle_{\mathcal{D}'(\Omega)} \\ &= - \left\langle \frac{\partial \bar{v}}{\partial x_i} \left| \frac{\partial \bar{u}}{\partial x_i} \right. \right\rangle_{L^2(\Omega)} \\ &= - \left\langle \frac{\partial u}{\partial x_i} \left| \frac{\partial v}{\partial x_i} \right. \right\rangle_{L^2}. \end{aligned}$$

donc la formule est vérifiée. Si u est dans $H_0^1(\Omega)$, il existe une suite $(u_j)_{j \in \mathbb{N}}$ de $\mathcal{C}_0^\infty(\Omega)$ qui converge vers u dans $H_0^1(\Omega)$. Par continuité du laplacien sur $L^2(\Omega)$ et du produit scalaire, on a

$$\lim_{j \rightarrow +\infty} \langle u_j | \Delta v \rangle = \langle u | \Delta v \rangle,$$

et pour tout i entre 1 et n ,

$$\lim_{j \rightarrow +\infty} \langle \partial_{x_i} u_j | \partial_{x_i} v \rangle = \langle \partial_{x_i} u | \partial_{x_i} v \rangle.$$

La formule est donc vraie dans ce cas aussi.

On a par conséquent

$$\langle Tf | g \rangle_{L^2} = - \langle u | \Delta v \rangle_{L^2} = \sum_{i=1}^n \left\langle \frac{\partial u}{\partial x_i} \left| \frac{\partial v}{\partial x_i} \right. \right\rangle_{L^2} = \langle -\Delta u | v \rangle = \langle f | Tg \rangle.$$

L'opérateur T est donc auto-adjoint.

L'opérateur T est positif car

$$\langle Tf | f \rangle = \langle u | -\Delta u \rangle = \sum_{i=1}^n \left\| \frac{\partial u}{\partial x_i} \right\|_{L^2}^2 \geq 0.$$

L'opérateur T est défini car il est injectif. ■

Grâce à ce résultat, on peut trouver les "valeurs propres" du Laplacien.

Théorème 137

Soit $\mu \in \mathbb{R}$. Le problème

$$-\Delta u = \mu u$$

admet une solution non uniformément nulle $u \in L^2(\Omega)$ (elle sera nécessairement dans $H_0^1(\Omega)$ avec notre construction) si et seulement si μ est l'un des éléments d'une famille $(\mu_n)_{n \in \mathbb{N}}$ de nombres réels strictement positifs qui tendent vers $+\infty$.

Démonstration : L'opérateur $(-\Delta)^{-1}$ est auto-adjoint compact et défini positif, son spectre est constitué de 0 et d'une suite $(\lambda_n)_{n \in \mathbb{N}}$ de valeurs propres réelles et strictement positives qui tendent vers 0.

Or, on a

$$Tu = \lambda u,$$

si et seulement si on a

$$\frac{1}{\lambda} u = -\Delta u,$$

d'où le résultat. ■

Remarquons que par hypoellipticité du laplacien (confère l'Annexe B. sur les distributions), les vecteurs propres sont des fonctions de classe $\mathcal{C}^\infty$ et que par le théorème de diagonalisation il existe une base orthonormale de vecteurs propres pour l'espace de Hilbert L^2 . Remarquons aussi que la manière dont nous avons construit les solutions intègre implicitement une condition de Dirichlet au bord de l'ouvert.

6. Synthèse et pratique

6.a. Synthèse de la section

Cette section présente des outils techniques relevant de la notation (multi-indices, ouvert régulier, intégrale de surface) et d'autres relevant de théories plus profondes (théorème de divergence, convolution, transformée de Fourier, espaces de Sobolev).

Il faut retenir que les fonctions indéfiniment dérivables à support compact sont idéales pour servir de fonction test, c'est-à-dire régulariser les autres fonctions ou manger les dérivées à la place des fonctions qui n'en sont pas capables.

Il faut retenir aussi que la convolution est un formidable outil de régularisation et que la transformée de Fourier se comporte plutôt bien vis-à-vis de la convolution, outre plusieurs autres propriétés plaisantes qui lui sont propres. L'espace naturel pour travailler avec la transformée de Fourier est l'espace de Schwarz, mais il existe des extensions naturelles à L^1 et L^2 (ainsi qu'aux distributions comme on peut le voir dans l'annexe B.).

On peut dériver au sens faible les fonctions intégrables et s'intéresser aux fonctions intégrables à la puissance p dont les dérivées faibles sont intégrables à la puissance p . Les espaces associés sont des espaces de Banach, parfois de Hilbert. Cela permet de mélanger les forces théoriques des espaces de Banach et de Hilbert et les dérivées qui permettent de parler d'équation différentielle.

Il existe des injections et des contrôles de normes dans les espaces de Sobolev qui sont utiles pour contrôler les différentes quantités qui apparaissent naturellement lors du travail analytique ou numérique sur une équation aux dérivées partielles.

6.b. Exercices

IV. ÉQUATIONS DIFFÉRENTIELLES ORDINAIRES

« L'équation différentielle est une façon de regarder. »

Antoine de Saint-Exupéry, *Carnets*

Une équation différentielle ordinaire est une équation du type

$$y'(t) = f(t, y(t)), \quad (t, y(t)) \in U \quad (\ominus)$$

où U est un ouvert de $\mathbb{R} \times \mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une fonction continue sur U et y une fonction inconnue. Résoudre l'équation différentielle $(\ominus)$ revient à trouver toutes les fonctions $y : I \rightarrow \mathbb{R}^n$ définies sur des intervalles I de $\mathbb{R}$ et de classe $\mathcal{C}^1$ sur I qui vérifient les deux conditions

- $\forall t \in I, (t, y(t)) \in U$
- $\forall t \in I, y'(t) = f(t, y(t))$

Cette partie rappelle les résultats fondamentaux de l'étude de ce type d'équations différentielles. Elle se base majoritairement sur le livre [2] ainsi que sur des cours de systèmes dynamiques donnés à l'université d'Orsay par Dominique Hulin, Frédéric Bourgeois, Hans-Henrik Rugh, Alexandre Vidal et Béatrice Laroche.

1. Cylindres de sécurité, maximalité des solutions

La première préoccupation à avoir pour étudier notre problème est celle du contrôle spatial d'une éventuelle solution sur un petit intervalle de temps. À ce titre, nous allons par la suite considérer des cylindres de la forme

$$C = [t_0 - T, t_0 + T] \times \bar{B}(y_0, r_0) \subset \mathbb{R} \times \mathbb{R}^n.$$

où $\bar{B}(y_0, r_0)$ est la boule fermée, dans $\mathbb{R}^n$, de centre y_0 et de rayon r_0 . On voudrait que les éventuelles solutions de $(\ominus)$ vérifiant $y(t_0) = y_0$ soient contenues dans un tel cylindre pour un court temps au moins. On considérera implicitement par la suite que T et r_0 sont choisis assez petits pour que C soit contenu dans l'ouvert U , dont il est alors un compact.

Définition 138. *Le cylindre $C = [t_0 - T, t_0 + T] \times \bar{B}(y_0, r_0) \subset U$ est un cylindre de sécurité pour l'équation $(\ominus)$ et la condition initiale $(t_0, y_0) \in U$ si pour toute solution $y : I \rightarrow \mathbb{R}^n$ vérifiant $y(t_0) = y_0$, on a $y(I \cap [t_0 - T, t_0 + T]) \subset \bar{B}(y_0, r_0)$.*

Cela revient simplement à demander qu'entre les temps $t_0 - T$ et $t_0 + T$ les solutions ne soient pas trop loin de y_0 .

Théorème 139

Soit $T_0, r_0 \in \mathbb{R}_+^*$ tels que $C_0 = [t_0 - T_0, t_0 + T_0] \times \bar{B}(y_0, r_0) \subset U$. Il existe $M \in \mathbb{R}_+^*$ tel que pour tout $T \in \mathbb{R}_+^*$ vérifiant $T \leq \min(T_0, \frac{r_0}{M})$,

$$C = [t_0 - T, t_0 + T] \times \bar{B}(y_0, r_0)$$

est un cylindre de sécurité pour l'équation (⊙) et la condition initiale (t_0, y_0) .

Démonstration : Soit $T_0, r_0 \in \mathbb{R}_+^*$ tel que le cylindre $C_0 = [t_0 - T_0, t_0 + T_0] \times \bar{B}(y_0, r_0)$ est inclus dans U . C_0 est compact, donc il existe $M \in \mathbb{R}_+^*$ tel que

$$\forall (t, x) \in C_0, \|f(t, x)\| \leq M.$$

Soit $T \in \mathbb{R}_+^*$ tel que $T \leq \min(T_0, \frac{r_0}{M})$ et soit $C = [t_0 - T, t_0 + T] \times \bar{B}(y_0, r_0)$. On a $C \subset C_0$, ce qui entraîne

$$\forall (t, x) \in C, \|f(t, x)\| \leq M.$$

Supposons par l'absurde qu'il existe une solution y et un temps $t_1 \in [t_0, t_0 + T]$ (la preuve est identique si $t_1 \in [t_0 - T, t_0]$) tel que $\|y(t_1) - y_0\| > r_0$. Posons alors

$$\tau = \inf\{t \in [t_0, t_0 + T] \mid \|y(t) - y_0\| > r_0\},$$

où l'infimum est pris sur un ensemble non vide et a une valeur strictement inférieure à T .

Les valeurs de y sont donc dans $\bar{B}(y_0, r_0)$ sur $[t_0, t_0 + \tau]$, ce qui entraîne, par continuité de y , que $\|y(\tau) - y_0\| = r_0$. On a alors, grâce à la classe $\mathcal{C}^1$ de y ,

$$r_0 = \|y(\tau) - y_0\| = \left\| \int_{t_0}^{\tau} y'(t) dt \right\| \leq M(\tau - t_0),$$

d'où

$$\frac{r_0}{M} \leq \tau - t_0,$$

ce qui constitue une contradiction avec $\tau - t_0 < T \leq \frac{r_0}{M}$.

Donc C est un cylindre de sécurité. ■

À présent que nous savons contrôler localement les solutions, on voudrait avoir des outils pour parler de leur temps de vie. Comme une solution peut jusqu'à preuve du contraire se prolonger en plusieurs solutions différentes définies sur un intervalle plus large, le bon concept est celui de la maximalité.

Définition 140. Une solution $y : J \rightarrow \mathbb{R}^n$ de (⊙) est dite maximale si pour toute solution $\tilde{y} : \tilde{J} \rightarrow \mathbb{R}^n$ qui prolonge y , on a $\tilde{J} = J$.

Théorème 141

Toute solution y de (⊙) se prolonge en une solution maximale $\tilde{y}$.

Démonstration : Supposons que y est définie sur l'intervalle $|a, b|$ (où $|$ désigne ici $]$ ou $[$, $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R} \cup \{+\infty\}$), et montrons qu'elle se prolonge en une solution maximale à droite, le résultat final devenant alors évident. Soit

$$c_1 = \sup\{c \in \mathbb{R} \mid y \text{ se prolonge sur } |a, c| \} \in \mathbb{R} \cup \{+\infty\}.$$

Par définition de la borne supérieure, si $c_1 < +\infty$, il existe b_1 tel que

$$b \leq b_1 \text{ et } c_1 - 1 < b_1 \leq c_1.$$

Si $c_1 = +\infty$, on pose $b_1 = \max(1, b)$.

Par récurrence, on peut construire $y_k : |a, b_k| \rightarrow \mathbb{R}^n$ une suite de prolongements successifs en posant

$$c_k = \sup\{c \in \mathbb{R} \mid y_{k-1} \text{ se prolonge sur } |a, c| \},$$

et en demandant, si $c_k < +\infty$,

$$b_{k-1} \leq b_k \text{ et } c_k - \frac{1}{k} < b_k \leq c_k,$$

ou, si $c_k = +\infty$, $b_k = \max(k, b_{k-1})$.

La suite c_k est décroissante car l'ensemble des prolongements de y_{k-1} contient l'ensemble de ceux de y_k . Si il existe k_0 tel que c_k est fini, alors c_k est fini pour tout $k > k_0$ et les suites $(b_k)_{k \in \mathbb{N}}$ et $(c_k)_{k \in \mathbb{N}}$ sont adjacentes et convergent par conséquent vers une même limite $\tilde{b}$.

Si pour tout $k \in \mathbb{N}$, $c_k = +\infty$, on pose $\tilde{b} = +\infty$.

Soit $\tilde{y}$ la solution définie sur $|a, \tilde{b}|$ définie sur tout intervalle $|a, b_k|$ par les valeurs de y_k et éventuellement prolongée en $\tilde{b}$ si c'est possible. Cette fonction est bien définie puisque $\cup_{k \in \mathbb{N}} |a, b_k| = |a, \tilde{b}|$.

Par construction, $\tilde{y}$ est maximale à droite. ■

Le prolongement de y en une solution maximale n'a aucune nécessité d'être unique.

2. Existence d'une solution

Afin de montrer un résultat d'existence général, nous avons besoin d'une forme affaiblie d'un résultat topologique classique : le théorème d'Arzela-Ascoli. Soient (E, d_E) et (F, d_F) deux espaces métriques. Pour toutes fonctions $\phi : E \rightarrow F$ et $\psi : E \rightarrow F$, on définit la quantité non nécessairement finie

$$d(\phi, \psi) = \sup_{x \in E} d_F(\phi(x), \psi(x)).$$

On notera $\mathcal{Lip}_k(E, F)$ l'ensemble des applications k -lipschitziennes de E dans F .

Théorème 142 (Arzela-Ascoli, forme faible)

Si E et F sont compacts, alors $(\mathcal{Lip}_k(E, F), d)$ est un espace métrique compact.

Démonstration : • Par continuité de la distance d_F et des applications k -lipschitziennes ϕ et ψ , l'application $x \mapsto d_F(\phi(x), \psi(x))$ est continue du compact E dans $\mathbb{R}$ muni de sa distance usuelle, ce dont on déduit qu'elle est bornée et atteint ses bornes. La quantité $d(\phi, \psi)$ est donc toujours finie. De plus, elle vérifie de manière évidente les axiomes d'une distance.

L'espace $(\mathcal{Lip}_k(E, F), d)$ est donc bien un espace métrique.

• Soit une suite (ϕ_p) de $\mathcal{Lip}_k(E, F)$, montrons qu'elle admet une suite extraite convergente pour d vers un élément $\phi \in \mathcal{Lip}_k(E, F)$.

Posons $S_0 = \mathbb{N}$. E et F sont compacts, donc il existe des recouvrements ouverts finis $(A_i)_{i \in I_1}$ et $(B_j)_{j \in J_1}$ de E et F par des boules de rayon $1 = \frac{1}{1}$.

Soit l'application $f_0 : S_0 \rightarrow J_1^{card(I_1)}$ telle que si $(x_i)_{i \in I_1}$ sont les centres des A_i , on a pour tout $p \in S_0$,

$$(f_0(p) = (j_i)_{i \in I_1}) \Rightarrow (\forall i \in I_1, \phi_p(x_i) \in B_{j_i}).$$

La partie S_0 est infinie et $J_1^{card(I_1)}$ est fini, donc il existe un élément $(l_i)_{i \in I_1} \in J_1^{card(I_1)}$ tel que $f_0^{-1}((l_i)_{i \in I_1})$ est une partie infinie de S_0 . On pose alors $S_1 = f_0^{-1}((l_i)_{i \in I_1})$.

• De même, on peut construire par récurrence, pour tout $n \in \mathbb{N}^*$, une partie S_n en prenant des recouvrement $(A_i)_{i \in I_n}$ et $(B_j)_{j \in J_n}$ de E et F par des boules de rayon $\frac{1}{n}$ et en définissant comme précédemment

$$f_{n-1} : S_{n-1} \rightarrow J_n^{card(I_n)},$$

et

$$S_n = f_{n-1}^{-1}((l_i)_{i \in I_n}).$$

• On a alors

$$\cdots \subset S_n \subset S_{n-1} \subset \cdots \subset S_1 \subset S_0$$

En outre, pour tout $n \in \mathbb{N}^*$, pour tout $i \in I_n$ et pour tous $p, q \in S_n$, on a par construction

$$d_F(\phi_p(x_i), \phi_q(x_i)) \leq \frac{2}{n}.$$

Or, ϕ_p et ϕ_q sont k -lipschitziennes, ce qui donne pour tout $x \in A_i$,

$$d_F(\phi_p(x_i), \phi_p(x)) \leq kd(x_i, x) \leq \frac{2k}{n},$$

et de même pour ϕ_q , d'où par double application de l'inégalité triangulaire,

$$d_F(\phi_p(x), \phi_q(x)) \leq \frac{2 + 4k}{n},$$

et le résultat vaut pour tout $x \in E$ puisque les A_i recouvrent E . Construisons la suite extraite $(\phi_{\sigma(p)})_{p \in \mathbb{N}}$ où pour tout $p \in \mathbb{N}$, $\sigma(p) \in S_p$. On a alors, pour tous $(p, q) \in \mathbb{N}$ tel que $p < q$ et tout $x \in E$,

$$d_F(\phi_{\sigma(p)}(x), \phi_{\sigma(q)}(x)) \leq \frac{2 + 4k}{p}, \quad (\mathfrak{D})$$

ce dont on déduit que la suite $\left(\phi_{\sigma(p)}(x)\right)_{p \in \mathbb{N}}$ est une suite de Cauchy de F qui est compact donc complet. Cette suite admet donc une limite que nous noterons $\phi(x)$.

Cette limite, considérée pour tout $x \in E$, induit une application $\phi : E \rightarrow F$. Cette dernière vérifie, pour tout $x, y \in E$, pour tout $p \in \mathbb{N}$ et par inégalité triangulaire,

$$d_F(\phi(x), \phi(y)) \leq d_F(\phi(x), \phi_{\sigma(p)}(x)) + d_F(\phi_{\sigma(p)}(x), \phi_{\sigma(p)}(y)) + d_F(\phi_{\sigma(p)}(y), \phi(y))$$

ce dont on déduit, sachant que $\phi_{\sigma(p)}$ est k -lipschitzienne,

$$d_F(\phi(x), \phi(y)) \leq d_F(\phi(x), \phi_{\sigma(p)}(x)) + kd(x, y) + d_F(\phi_{\sigma(p)}(y), \phi(y)),$$

et on obtient par passage à la limite sur p ,

$$d_F(\phi(x), \phi(y)) \leq kd(x, y).$$

L'application ϕ est donc k -lipschitzienne. Enfin, en passant à la limite sur q puis en prenant le supremum sur $x \in E$ dans l'inégalité (2), on peut écrire

$$d(\phi_{\sigma(p)}, \phi) \leq \frac{2 + 4k}{p},$$

ce dont on déduit que $(\phi_{\sigma(p)})_{p \in \mathbb{N}}$ converge vers $\phi \in \mathcal{Lip}_k(E, F)$ pour la distance d .

En conclusion, $(\mathcal{Lip}_k(E, F), d)$ est un espace métrique compact. ■

Théorème 143 (*Cauchy-Peano-Arzela*)

Pour tout couple $(t_0, y_0) \in U$, il existe une solution maximale $y : I \rightarrow \mathbb{R}^n$ à l'équation (1) telle que $y(t_0) = y_0$.

Démonstration : Soit $C = [t_0 - T, t_0 + T] \times \bar{B}(y_0, r_0) \subset U$ un cylindre de sécurité tel que sur C , $\|f\| \leq M$ et $T \leq \min(T_0, \frac{r_0}{M})$.

• Commençons par construire par la méthode d'Euler des solutions ε -approchées au sens où,

$$\forall t \in [t_0 - T, t_0 + T], \quad \|y'(t) - f(t, y(t))\| \leq \varepsilon.$$

On travaille sur $[t_0, t_0 + T]$, la construction étant la même sur $[t_0 - T, t_0]$. On se donne une subdivision

$$t_0 < t_1 < \dots < t_N = t_0 + T,$$

dont les pas successifs sont notés $h_n = t_{n+1} - t_n$. On pose $h_{\max} = \max(h_0, \dots, h_N)$. Soit $(y_k)_{k \in \{0, \dots, N\}}$ la suite finie définie par récurrence en posant

$$y_{k+1} = y_k + h_k f(t_k, y_k),$$

et soit y la fonction affine par morceaux qui passe par tous les points $y_0, \dots, y_N$. On a $y([t_0, t_1]) \subset C$, et on vérifie aisément, en raisonnant par récurrence, que $y([t_0, t_0 + T]) \subset C$. On la construit de même sur $[t_0 - T, t_0]$ et on appelle toujours y la fonctions affine par morceaux sur $[t_0 - T, t_0 + T]$. On a encore $y([t_0 - T, t_0 + T]) \subset C$.

Posons l'application

$$\omega_f(x) = \max\{\|f(t_\alpha, y_\alpha) - f(t_\beta, y_\beta)\| \mid (t_\alpha, y_\alpha), (t_\beta, y_\beta) \in C \text{ et } |t_\alpha - t_\beta| + \|y_\alpha - y_\beta\| \leq x\},$$

où l'existence du maximum provient de la compacité de C . La limite de ω_f en 0^+ vaut 0 par continuité, donc uniforme continuité, de f sur le compact C .

Pour tout $t \in [t_0, t_0 + T]$, tel que $t_k < t < t_{k+1}$, on a

$$\|y'(t) - f(t, y(t))\| = \|f(t_k, y_k) - f(t, y(t))\| \leq \omega_f(Mh_k + h_k),$$

car $\|y(t) - y_k\| \leq (t - t_k) \|f(t_k, y_k)\| \leq Mh_k$. On a donc,

$$\|y'(t) - f(t, y(t))\| \leq \omega_f((M + 1)h_{max}).$$

On peut donc construire pour tout $\varepsilon \in \mathbb{R}_+$, en choisissant h_{max} assez petit, des solutions y_ε dites ε -approchées, au sens où $y'_\varepsilon(t)$ ne s'éloigne pas de plus de ε de $f(t, y'_\varepsilon(t))$.

• Montrons ensuite qu'une suite de solutions de plus en plus approchées qui converge uniformément a pour limite une solution exacte de l'équation différentielle.

Soit $(y_p)_p \in \mathbb{N}$ une suite de solutions $\frac{1}{p}$ -approchées définies sur $[t_0 - T, t_0 + T]$ et à valeurs dans C , qui existe d'après ce qui précède. Supposons en outre que $(y_p)_p \in \mathbb{N}$ converge uniformément vers une fonction y sur $[t_0 - T, t_0 + T]$.

Par majoration, on a

$$\begin{aligned} \left\| y_p(t) - y_0 - \int_{t_0}^t f(s, y_p(s)) \, ds \right\| &= \left\| \int_{t_0}^t (y'_p(s) - f(s, y_p(s))) \, ds \right\| \\ &\leq \int_{t_0}^t \|y'_p(s) - f(s, y_p(s))\| \, ds \\ &\leq \frac{t - t_0}{p} \end{aligned}$$

Par convergence uniforme, y est continue et

$$\lim_{p \rightarrow +\infty} \left\| \int_{t_0}^t f(s, y_p(s)) \, ds - \int_{t_0}^t f(s, y(s)) \, ds \right\| = 0,$$

d'où par passage à la limite dans ce qui précède :

$$y(t) - \left(y_0 - \int_{t_0}^t f(s, y(s)) \, ds \right) = 0,$$

ce qui, par continuité de y , implique que y est solution de (⊙).

• Concluons enfin avec le théorème d'Arzela-Ascoli.

Pour tout $p \in \mathbb{N}$, pour tous $t_1, t_2 \in [t_0, t_0 + T]$ (le raisonnement se généralise sans mal à $[t_0 - T, t_0 + T]$), tels que $t_1 < t_2$, il existe k_1, k_2 entiers naturels tels que $k_1 \leq k_2$ et,

$$t_{k_1} \leq t_1 \leq t_{k_1+1} \quad \text{et} \quad t_{k_2} \leq t_2 \leq t_{k_2+1},$$

et on a alors, par définition des solutions approchées,

$$\begin{aligned} \|y_p(t_2) - y_p(t_1)\| &= \|y_{k_2} + (t_2 - t_{k_2})f(t_{k_2}, y_{k_2}) - (y_{k_1} + (t_1 - t_{k_1})f(t_{k_1}, y_{k_1}))\| \\ &= \left\| \sum_{k_1}^{k_2-1} (t_{i+1} - t_i)f(t_i, y_i) + (t_2 - t_{k_2})f(t_{k_2}, y_{k_2}) - (t_1 - t_{k_1})f(t_{k_1}, y_{k_1}) \right\| \\ &\leq \sum_{k_1}^{k_2-1} (t_{i+1} - t_i)M + (t_2 - t_{k_2})M - (t_1 - t_{k_1})M \\ &\leq (t_2 - t_1)M. \end{aligned}$$

Donc, pour tout $p \in \mathbb{N}$, y_p est M -lipschitzienne. D'après le théorème d'Arzela-Ascoli, il existe une sous-suite de $(y_p)_p \in \mathbb{N}$ qui converge uniformément sur $[t_0 - T, t_0 + T]$ vers une fonction M -lipschitzienne y . D'après ce qui précède, cette solution est une solution exacte.

D'après le théorème 141, il existe une solution maximale de (O) qui prolonge y , donc vérifie la condition initiale, ce qui achève la preuve. ■

Remarquons que l'ensemble de définition de la solution maximale est nécessairement ouvert, car s'il a un bord, par exemple à droite en b , on peut construire par les méthodes précédentes une solution centrée en b qui prolonge à droite. Remarquons encore qu'on n'a pas nécessairement unicité de la solution maximale : il existe des exemples pour lesquels l'unicité n'est pas vérifiée.

Pendant la preuve, nous avons construit des approximations de la solution qui peuvent être calculées par un ordinateur. C'est la méthode d'Euler, une des formes les plus simples de l'analyse numérique des équations différentielles.

3. Résultats généraux

3.a. Unicité, finitude et longévité de l'existence

Sous une hypothèse plus forte que la seule continuité de f , on a non seulement existence mais aussi unicité de la solution de (O). La preuve que nous proposons du théorème suivant utilise le théorème du point-fixe de Picard que nous avons déjà vu au chapitre 1 (voir théorème 21). Définissons d'abord l'hypothèse pertinente.

Définition 144. Soit U un ouvert de $\mathbb{R} \times \mathbb{R}^n$ et $f : U \rightarrow \mathbb{R}^n$ une fonction. f est dite localement lipschitzienne par rapport à la seconde variable si pour tout ensemble compact $S \subset U$, il existe une constante $k_S \in \mathbb{R}_+$ telle que

$$\forall (t, x_1), (t, x_2) \in S, \quad \|f(t, x_1) - f(t, x_2)\| \leq k_S \|x_1 - x_2\|.$$

Remarquons que l'on pouvait dire aussi qu'autour de tout point de U il y a un voisinage sur lequel f est lipschitzienne selon la seconde variable, ce qui était équivalent par argument de compacité (propriété de Borel-Lebesgue et un peu de cogitation).

Théorème 145 (Cauchy-Lipschitz)

Si f est localement lipschitzienne par rapport à la seconde variable, alors le problème de Cauchy $y' = f(t, y)$ avec condition initiale $y(t_0) = y_0$ admet une unique solution maximale $y : I \rightarrow \mathbb{R}^n$ définie sur un intervalle I .

Démonstration : Soit $C = [t_0 - T, t_0 + T] \times \bar{B}(y_0, r_0)$ un cylindre de sécurité contenant (t_0, y_0) et vérifiant $T \leq \min(T_0, \frac{r_0}{M})$. Par hypothèse, il existe $k \in \mathbb{R}_+$ tel

que f soit k -lipschitzienne par rapport à la seconde variable sur C .
Notons $J = [t_0 - T, t_0 + T]$. Soit $E = \mathcal{C}^0(J, \bar{B}(y_0, r_0))$ l'ensemble des applications continues de J dans $\bar{B}(y_0, r_0)$ que l'on munit de la distance d de la convergence uniforme :

$$d(x, y) = \sup_{t \in J} \|x(t) - y(t)\|.$$

L'espace métrique (E, d) est complet.

On définit la fonctionnelle $\Phi : E \rightarrow E$ par

$$\Phi(y)(t) = y_0 + \int_{t_0}^t f(s, y(s)) \, ds.$$

y est une solution de (⊙) si et seulement si y est un point fixe de Φ .

Φ est bien définie car, pour $y \in E$ et $t \in I$,

$$\|\Phi(y)(t) - y_0\| = \left\| \int_{t_0}^t f(s, y(s)) \, ds \right\| \leq \int_{t_0}^t \|f(s, y(s))\| \, ds \leq M |t - t_0| \leq MT \leq r_0$$

et donc $\Phi(y) \in E$.

On peut donc, en particulier, itérer la fonctionnelle Φ . Soit $x, y \in E$. Alors, $\forall t \in J$,

$$\begin{aligned} \|\Phi(x)(t) - \Phi(y)(t)\| &= \left\| \int_{t_0}^t (f(s, x(s)) - f(s, y(s))) \, ds \right\| \\ &\leq \int_{t_0}^t \|f(s, x(s)) - f(s, y(s))\| \, ds \\ &\leq \int_{t_0}^t k \|x(s) - y(s)\| \, ds \leq k |t - t_0| d(x, y) \end{aligned}$$

De même, $\forall t \in J$,

$$\begin{aligned} \|\Phi^2(x)(t) - \Phi^2(y)(t)\| &\leq \int_{t_0}^t k \|\Phi(x)(s) - \Phi(y)(s)\| \, ds \\ &\leq \int_{t_0}^t k^2 |s - t_0| d(x, y) \, ds = k^2 \frac{|t - t_0|^2}{2} d(x, y). \end{aligned}$$

D'où, par récurrence, $\forall i \in \mathbb{N}$, $\forall t \in J$,

$$\|\Phi^i(x)(t) - \Phi^i(y)(t)\| \leq k^i \frac{|t - t_0|^i}{i!} d(x, y),$$

et, en passant à la borne supérieure, $\forall i \in \mathbb{N}$,

$$d(\Phi^i(x), \Phi^i(y)) \leq k^i \frac{T^i}{i!} d(x, y).$$

Φ^i est donc $\frac{(kT)^i}{i!}$ -lipschitzienne sur E .

Or, comme $\lim_{i \rightarrow +\infty} \frac{(kT)^i}{i!} = 0$, il existe $i_0 \in \mathbb{N}$ tel que pour tout $i \geq i_0$, $\frac{(kT)^i}{i!} < 1$.

Pour une telle valeur de i , Φ^i est alors contractante et le théorème du point fixe de Picard nous donne l'existence d'un unique point fixe pour Φ , c'est-à-dire d'une unique solution continue définie sur $[t_0 - T, t_0 + T]$.

Soit deux solutions $y_1 : I_1 \rightarrow \mathbb{R}^n$ et $y_2 : I_2 \rightarrow \mathbb{R}^n$ vérifiant $y_j(t_0) = y_0$, $j = 1, 2$. Considérons l'ensemble

$$\mathcal{I} = \{t \in I_1 \cap I_2 \mid y_1(t) = y_2(t)\}.$$

$\mathcal{I}$ est connexe comme intersection d'intervalle et non-vide car il contient t_0 . C'est un fermé de $I_1 \cap I_2$ car c'est l'image réciproque du fermé $\{0\}$ par l'application $t \mapsto y_1(t) - y_2(t)$. C'est un ouvert de $I_1 \cap I_2$ par application en tout point de $\mathcal{I}$ du résultat d'unicité locale que nous avons démontré. Par connexité, on a $\mathcal{I} = I_1 \cap I_2$.

On en déduit que la solution maximale, qui existe d'après le théorème 141, est unique. ■

Remarque 146. *Toutes les fonctions f de classe $\mathcal{C}^1$ vérifient les hypothèses du théorème de Cauchy-Lipschitz. Le théorème local que nous avons obtenu au cours de la preuve nous assure l'existence d'un intervalle minimal d'existence $[t_0 - T, t_0 + T]$ qui dépend de f .*

En fait, lorsque l'on se déplace légèrement dans l'ensemble des conditions initiales, on ne perturbe *pas trop* le domaine d'existence minimal des solutions, comme l'exprime le lemme suivant.

Lemme 147 (Uniformité locale du temps de vie). *Supposons f localement lipschitzienne par rapport à la seconde variable. Soit $(t_0, y_0) \in U$. Il existe un ouvert $V \subset U$ contenant (t_0, y_0) et un réel strictement positif ε tel que pour tout $(t_1, y_1) \in V$, la solution maximale de (⊙) de condition initiale (t_1, y_1) est définie au moins sur $]t_1 - \varepsilon, t_1 + \varepsilon[$.*

Démonstration : Soit $a, R \in \mathbb{R}_+^*$ tels que

$$C = [t_0 - 2a, t_0 + 2a] \times \bar{B}(y_0, 2R) \subset U$$

Soit M le maximum de la norme de f sur le compact C . Posons également

$$V =]t_0 - a, t_0 + a[\times B(y_0, R).$$

Pour tout $(t_1, y_1) \in V$, on a

$$[t_1 - a, t_1 + a] \times \bar{B}(y_1, R) \subset C,$$

et par conséquent la norme de f est inférieure ou égale à R sur ce cylindre centré en (t_1, y_1) . Donc, si l'on pose

$$\varepsilon = \min(a, \frac{R}{M}),$$

le cylindre

$$[t_1 - \varepsilon, t_1 + \varepsilon] \times \bar{B}(y_1, R),$$

est un cylindre de sécurité, ce qui implique par le sous-résultat obtenu dans la démonstration de Cauchy-Lipschitz que la solution maximale existe au moins sur $[t_1 - \varepsilon, t_1 + \varepsilon]$, et ε ne dépend pas de (t_1, y_1) . ■

Remarque 148. On en déduit par un développement usuel utilisant le théorème de Borel-Lebesgue que pour tout compact de U , il existe $\alpha \in \mathbb{R}_+^*$ tel que toutes les solutions de conditions initiales (t_1, y_1) dans ce compact vivent au moins sur $[t_1 - \alpha, t_1 + \alpha]$.

On peut légitimement se demander ce qui arrive lorsque le temps de vie maximal est limité. En fait, il ne peut arriver qu'une chose en cas de durée de vie limitée : l'explosion en temps fini. Le théorème suivant, aussi appelé lemme de sortie de tout compact, précise cette intuition.

Théorème 149 (Lemme des bouts)

Supposons f localement lipschitzienne par rapport à la seconde variable. Une solution $y :]a, b[\rightarrow \mathbb{R}^n$ de $(\odot)$ est maximale si et seulement si elle s'échappe définitivement de tout compact de U lorsque t tend vers a ou b .

Démonstration : On ne démontrera que la maximalité à droite, celle à gauche fonctionnant de la même façon.

Supposons que la solution maximale revient une infinité de fois dans un compact K de U au voisinage de b , au sens où il existe une suite $\left((t_n, y(t_n))\right)_{n \in \mathbb{N}^*}$ qui est à valeurs dans K et telle que t_n converge vers b . Par compacité, elle admet une suite extraite $\left((t_{\sigma(n)}, y(t_{\sigma(n)}))\right)_{n \in \mathbb{N}}$ qui converge dans K vers une valeur $(b, y_+) \in K$.

D'après le lemme précédent, il y a uniformité locale du temps de vie autour du point (b, y_+) : toutes les solutions dans un voisinage V vivent au moins sur un intervalle de longueur 2ε .

Or, la suite $(t_{\sigma(n)}, y(t_{\sigma(n)}))$ s'approche à moins de ε de (b, y_+) et entre définitivement dans le voisinage V , donc la solution vit sur un intervalle plus grand que $]a, b[$.

Contradiction.

Donc la solution maximale s'échappe définitivement de tout compact.

Réciproquement, si y n'est pas maximale à droite, alors elle se prolonge à droite en $\tilde{y}$ qui est définie sur un intervalle ouvert $]a, d[$, $b < d$, donc au moins sur un intervalle compact $[a + \delta, c]$ avec $b < c < d$ et $\delta \in \mathbb{R}_+^*$. Au voisinage de b , la solution y est contenue dans le compact $\tilde{y}([a + \delta, c])$.

Par contraposée, toute solution qui sort définitivement de tout compact de U est maximale. ■

Il nous manque encore une condition suffisante pour que les solutions soient définies sur le plus grand intervalle possible.

Théorème 150 (Lemme de Grönwall)

Soit I un intervalle ouvert de $\mathbb{R}$ et $y : I \rightarrow \mathbb{R}^n$ une fonction de classe $\mathcal{C}^1$. Supposons qu'il existe $a \in \mathbb{R}_+^*$, $b \in \mathbb{R}$ tels que

$$\forall t \in I, \|y'(t)\| \leq a \|y(t)\| + b.$$

Alors,

$$\forall t_0 \in I, \forall t \in I, \|y(t)\| \leq \|y(t_0)\| e^{a|t-t_0|} + b \frac{e^{a|t-t_0|} - 1}{a}.$$

Démonstration : La preuve qui va suivre ne dépendant pas de la valeur de t_0 , on suppose sans perte de généralité $0 \in I$ et $t_0 = 0$ pour simplifier les calculs. Soit $t \in \mathbb{R}_+^*$, on a par inégalité triangulaire

$$\|y(t)\| \leq \|y(0)\| + \int_0^t \|y'(s)\| \, ds,$$

d'où par hypothèse,

$$\|y(t)\| \leq \|y(0)\| + tb + a \int_0^t \|y(s)\| \, ds. \quad (2)$$

Posons

$$v : t \mapsto \int_0^t \|y(s)\| \, ds.$$

v est de classe $\mathcal{C}^1$, vérifie $v(0) = 0$ et $v'(t) = \|y(t)\|$. Donc l'inégalité ci-dessus se réécrit

$$v'(t) - av(t) \leq \|y(0)\| + tb,$$

ou encore

$$\left(v'(t) - av(t)\right) e^{-at} \leq \left(\|y(0)\| + tb\right) e^{-at}.$$

On peut aussi réécrire cette inégalité sous la forme

$$\frac{d}{dt} \left(v(t) e^{-at}\right) \leq \left(\|y(0)\| + tb\right) e^{-at},$$

ce qui donne par croissance de l'intégrale

$$v(t) e^{-at} - v(0) \leq \|y(0)\| \frac{1 - e^{-at}}{a} + b \frac{1 - (1 + at) e^{-at}}{a}.$$

Sachant que $v(0) = 0$, on peut réécrire (2) sous la forme

$$\|y(t)\| \leq \|y(0)\| + bt + \|y(0)\| (e^{at} - 1) + b \frac{e^{at} - (1 + at)}{a},$$

dont on tire enfin,

$$\|y(t)\| \leq \|y(0)\| e^{at} + b \frac{e^{at} - 1}{a}.$$

On montre le même résultat avec des valeurs absolues autour de t lorsque $t < 0$ en posant $z(t) = y(-t)$, d'où le résultat final. ■

Ce théorème a pour conséquence un résultat très utile pour étudier les équations différentielles :

Théorème 151 (Corollaire de Grönwall)

Supposons que $U = I \times \mathbb{R}^n$, I intervalle ouvert, que f est lipschitzienne par rapport à la seconde variable et qu'il existe deux fonctions continues

$$\alpha : I \rightarrow [0, +\infty[\quad \text{et} \quad \beta : I \rightarrow [0, +\infty[,$$

telles que

$$\forall (t, x) \in I \times \mathbb{R}^n, \quad \|f(t, x)\| \leq \alpha(t) \|x\| + \beta(t).$$

Alors les solutions maximales de (⊙) sont globales, c'est-à-dire définies sur I .

Démonstration : Supposons qu'il existe une solution maximale $y :]a, b[\rightarrow \mathbb{R}^n$ telle que $b < \sup I$.

Le compact $[a, b]$ est inclus dans I , donc α et β sont majorées sur $[a, b]$ par a et b respectivement. On a donc,

$$\|y'(t)\| = \|f(t, y(t))\| \leq a \|y(t)\| + b,$$

donc d'après le lemme de Grönwall, pour $t_0 \in]a, b[$,

$$\forall t \in [t_0, b[, \quad \|y(t)\| \leq \|y(t_0)\| e^{a|t-t_0|} + b \frac{e^{a|t-t_0|} - 1}{a}.$$

La fonction y reste donc bornée sur $[t_0, b[$, ce qui constitue une contradiction avec le lemme des bouts.

Le même raisonnement fonctionne dans le passé en posant $z(t) = y(-t)$.

Donc y est définie sur I . ■

Remarque 152. Les équations différentielles linéaires de la forme

$$Y' = A(t)Y + B(t),$$

où $A(t)$ est une matrice pour tout temps, vérifient les conditions du théorème et ont donc des solutions définies sur $\mathbb{R}$.

4. Principes généraux de l'analyse qualitative

Il est souvent très difficile, parfois impossible, de donner une expression explicite des solutions d'une équation différentielle non-linéaire. Une première solution est de calculer, à la main ou avec un ordinateur, des solutions approchées par diverses méthodes (schéma d'Euler implicite ou explicite, schéma multi-pas, méthode de Runge-Kutta,...). Une autre approche est l'analyse qualitative qui consiste à étudier les propriétés des solutions à partir de la forme de l'équation : on parle alors de l'étude des systèmes dynamiques.

Toute équation différentielle ordinaire est équivalente à une équation autonome, c'est-à-dire de la forme

$$y'(t) = f(y(t)), \quad y(t) \in U \quad (\delta)$$

où U est un ouvert de $\mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une fonction continue, appelée ici champ de vecteurs, et y une fonction inconnue.

En effet, étant donné une équation de la forme $u'(t) = g(t, u(t))$ définie sur $V \subset \mathbb{R} \times \mathbb{R}^n$, on peut définir la nouvelle variable $y(t) = (u(t), t)$ et on a alors $y'(t) = f(y(t))$ sur l'ouvert $U \subset \mathbb{R}^{n+1}$ avec $f : U \rightarrow U$ et $f(y(t)) = (g(t, u(t)), 1)$.

4.a. Description qualitative de la biologie

Avant de parler des mathématiques permettant l'analyse qualitative, nous allons motiver la question en expliquant comment les modèles dynamiques permettent de décrire des propriétés importantes des systèmes biologiques.

Les définitions mathématiques de certaines notions introduites dans ce paragraphe – telles que celle de portrait de phase, etc. – sont par ailleurs reportées au paragraphe suivant.

De la physique à la biologie : En caricaturant beaucoup, on peut dire que les systèmes dynamiques ont émergé d'une erreur d'Henri Poincaré. Ce dernier avait envoyé un manuscrit à l'académie royale de Suède pour un concours portant sur la stabilité du système solaire. Il gagna haut la main le concours et le prix associé. Cependant, après qu'on lui eut demandé des précisions, il se relut et fit la découverte dans son travail d'une grave erreur que personne n'avait remarquée ; les lectrices intéressées par l'erreur en question pourront consulter l'excellent texte tiré d'une conférence de Cédric Villani, *La meilleure et la pire des erreurs de Poincaré*, sur la page http://www.apmep.fr/IMG/pdf/503_Villani-Metz.pdf. Plutôt que de la taire, Poincaré fit rapatrier à ses frais tous les articles faux, pour un montant supérieur au prix gagné dans le concours, et se pencha sur ce qui allait renaître dans les années 1970 sous le nom de systèmes dynamiques et de théorie du chaos.

Henri Poincaré, en plus de sa connaissance hors du commun de l'édifice mathématique de son époque, avait une connaissance approfondie, sinon exhaustive, de la physique et de la philosophie, aussi a-t-il développé les notions fondatrices des systèmes dynamiques selon des intuitions venant de phénomènes physiques, d'où un vocabulaire reconnaissable que nous allons détailler (orbites, attracteurs, flot,...). Cependant, les générations suivantes de mathématiciens et de physiciens (Kolmogorov, Turing, Schrödinger,...) exploiteront ces notions pour essayer de comprendre les dynamiques présentes dans la nature. Des décennies plus tard, avec la mathématisation progressive (mais tardive par bien des aspects) de la biologie, une forte communauté de mathématiciens s'organisa autour de l'étude de problèmes issus des sciences du vivant.

Dynamique des populations : Les systèmes dynamiques décrivent très naturellement les phénomènes de dynamiques de population, par exemple avec des équations de la forme

$$y'(t) = f(y(t)) - g(y(t)),$$

où f et g sont des fonctions positives et continûment dérivables ; $y(t)$ représente la densité de population au temps t , $f(y)$ représente le taux de naissance de la population en fonction de la densité y , g joue le même rôle mais représente un taux de mort. Le cas le plus simple, dit malthusien, décrit une croissance ou décroissance exponentielle de la population

$$y'(t) = \alpha y(t) - \beta y(t),$$

où $\alpha, \beta \in \mathbb{R}_+$. La solution, évidente, s'écrit $\exp((\alpha - \beta)t)$. On peut ici prédire très facilement le comportement en temps long, donc l'avenir de la population. Pour des fonctions f et g plus générales, nous n'avons parfois pas de moyens d'exprimer simplement la solution et une théorie qui prédit autrement l'avenir de la population peut s'avérer utile pour prévenir l'invasion par une espèce ou son extinction.

Écosystèmes et réseaux : On peut étudier l'interaction d'un nombre n fini d'espèces vivantes de la même manière, ou la communication entre des entités biologiques connectées en réseau. Si $x_1, \dots, x_n$ représentent, par exemple, la concentration de n espèces dans un milieu, l'évolution de l'écosystème ainsi constitué peut être décrite par une dynamique de la forme

$$\begin{cases} x_1'(t) &= f_1(x_1, \dots, x_n) \\ \vdots & \vdots \\ x_n'(t) &= f_n(x_1, \dots, x_n). \end{cases}$$

Les fonctions f_i peuvent, selon les choix, représenter des dynamiques compétitives, collaboratives ou symbiotiques au sein de la population. Ici, nous ne souhaitons pas seulement savoir si les espèces vont s'éteindre, mais également connaître les états d'équilibre possibles pour l'écosystème (ou l'état d'équilibre pour l'activité d'un réseau de régulation, d'un réseau de neurones,...), et savoir si ces équilibres sont stables, si pour tout état initial on va se diriger vers l'équilibre ou au contraire s'en éloigner.

Des comportements complexes peuvent apparaître, comme dans le célèbre système proie-prédateur de Lotka-Volterra ¹⁶ : si y décrit la concentration d'un prédateur dans un milieu et x la concentration de son unique proie dans le même milieu, l'évolution des deux populations peut être décrite en première approche par le système

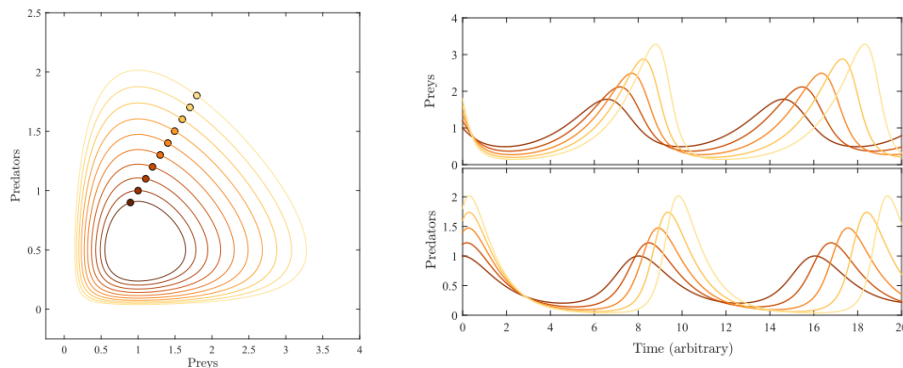
$$\begin{cases} x'(t) &= x(t)(a - bx(t)) \\ y'(t) &= y(t)(-c + dx(t)), \end{cases}$$

16. Alfred Lotka et Vito Volterra ont proposés ce modèle indépendamment en 1925 et 1926.

où $a, b, c, d \in \mathbb{R}_+^*$. Sans proie, les prédateurs meurent, sans prédateurs, la population de proie croît exponentiellement. Le système proie-prédateur admet deux équilibres, les points

$$(0, 0) \quad \text{et} \quad \left(\frac{a}{b}, \frac{c}{d}\right),$$

le premier représentant l'extinction pour les deux espèces et le second la coexistence dans un état d'équilibre. On peut montrer que pour la plupart des conditions initiales, les populations de proies et de prédateurs oscillent autour du point d'équilibre, comme on peut le voir sur la figure 2.



(a) Trajectoire dans l'espace de phase proie-prédateur pour quelques conditions initiales; les conditions initiales sont indiquées par des cercles.

(b) Solutions des équations de Lotka-Volterra obtenues par un algorithme de Runge-Kutta d'ordre 4 pour différentes conditions initiales.

FIGURE 2 – Le système de Lotka-Volterra pour les valeurs de paramètres $a = \frac{2}{3}$, $b = \frac{4}{3}$, $c = 1$ et $d = 1$: pour la plupart des conditions initiales, les populations de proies et de prédateurs oscillent autour du point d'équilibre du système.
Source : Wikipédia.

Modélisation des neurosciences : Les neurones ont une structure particulière, fortement liée à leur fonction ; ils se divisent en une partie centrale et des prolongements, les *neurites*. Plus précisément, on distingue :

- Le *corps cellulaire* (également appelé *soma* ou *pérıkaryon*, littéralement autour du noyau), relativement sphérique et d'une taille souvent aux alentours de 20 μm , contient le *noyau* de la cellule et les divers organites que l'on trouve dans la plupart des cellules humaines (mitochondries, appareil de Golgi, réticulum endoplasmique, etc).
- L'*axone* est un neurite qui prend naissance sur le corps cellulaire, sa base étant appelée *cône axonique*. L'axone peut mesurer de moins d'un millimètre à plus d'un mètre de longueur et son diamètre varie entre 1 et 25

μm . La composition protéique de l'axone est très différente de celle du corps cellulaire.

Les axones se divisent en branches multiples, qui partent fréquemment de façon orthogonale, appelées les *colatérales axoniques*. À leurs bouts, les axones se ramifient en des *terminaisons axoniques*, aussi appelées *boutons terminaux*, qui font avec d'autres neurones des points de contact appelés *synapses*.

- Les *dendrites* (du grec *dendron* qui signifie arbre) sont des neurites qui prennent naissance sur le corps cellulaire et se ramifient considérablement. L'ensemble des dendrites est appelé *arborisation dendritique*.

Cette structure (figure 3) permet aux neurones de se connecter les uns aux autres (des axones aux dendrites) et de communiquer par les potentiels d'action.

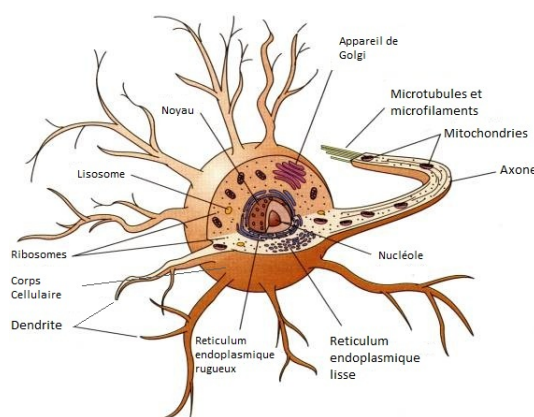


FIGURE 3 – Structure du neurone

Si les axones des neurones rappellent des fils électriques, il n'ont pas le même fonctionnement. En effet, le cuivre des fils est isolé par l'air ou la gaine qui les entoure et transmet des électrons à la moitié de la vitesse de la lumière. Une telle prouesse est impossible pour des axones dont la conduction électrique se base sur les ions, lourds et patauds, et qui sont de surcroît entourés d'un liquide électriquement chargé. Les neurones utilisent pour transmettre l'information une méthode beaucoup plus élaborée qu'un transfert passif d'électrons.

Deux des éléments fondamentaux du fonctionnement neuronal sont les ions et les canaux ioniques. Des ions sont naturellement présents de part et d'autre de la membrane, ils déterminent une différence de potentiel entre les milieux extracellulaire et intracellulaire : le *potentiel membranaire*. Les canaux ioniques sont des assemblages de protéines qui forment une ouverture dans la membrane capable de laisser passer les ions en réaction à certains stimuli. Chaque canal ionique est spécifique à certains ions. Les principaux ions en cause dans le fonctionnement du cerveau – les ions sodium (Na^+), potassium (K^+), calcium

(Ca^{2+}) et chlore (Cl^{-}) – ont donc leurs canaux correspondants sur la membrane des neurones : canaux sodiques, potassiques, calciques, chloriques.

Lorsque les canaux ioniques sont ouverts, les ions traversent la membrane par diffusion suivant le gradient de concentration et par l'effet du potentiel électrique. La diffusion des ions à travers les canaux par gradient de concentration entraîne une différence de potentiel qui freine ensuite icelle par une force électromotrice, et ce jusqu'à atteindre un état d'équilibre. Au repos, il y a donc une différence de potentiel entre l'intérieur et l'extérieur de la membrane du neurone : le potentiel à l'intérieur est négatif par rapport à l'extérieur, une différence de 65 mV, soit un potentiel membranaire $V_m = -65mV$.

L'existence du gradient de concentration à travers la membrane est assurée par des pompes ioniques, situées elles aussi sur la membrane, qui utilisent de l'énergie, sous forme d'ATP, pour faire passer des ions à travers la membrane contre le gradient de concentration. Les pompes sodium-potassium et calcium sont particulièrement importantes dans le fonctionnement des neurones ; on estime que 70% de l'ATP utilisée dans le cerveau sert à faire fonctionner la pompe sodium-potassium.

Nous pouvons enfin décrire le mécanisme qu'utilisent les neurones pour communiquer : le potentiel d'action. C'est un évènement bref, de l'ordre de la milliseconde, qui se décompose en plusieurs phases :

- *phase de seuil* : un évènement extérieur provoque l'ouverture des canaux sodiques et le gradient de concentration élevé, maintenu par les pompes, cause une entrée d'ions Na^{+} dans le neurone, ce qui change le potentiel membranaire ; la membrane se retrouve dépolarisée ; le niveau critique de dépolarisation qui cause le potentiel d'action s'appelle le *seuil* ;
- *Dépolarisation* : une fois le seuil atteint, une force électromotrice provoque une entrée massive d'ions Na^{+} ; le potentiel membranaire augmente brutalement ;
- *Repolarisation* : les canaux sodiques se referment et les canaux potassiques s'ouvrent ; les ions K^{+} sortent massivement du neurone et le potentiel chute très vite, plus bas encore qu'au repos ; la cellule est hyperpolarisée (le potentiel de la membrane redevient négatif) ;
- *post-hyperpolarisation* : les concentrations d'ions reviennent à l'équilibre et le potentiel revient "lentement" vers le potentiel de repos ;

Après avoir émis un potentiel d'action, le neurone entre dans une *période réfractaire* durant laquelle les canaux sodiques sont inactivés jusqu'au retour à un potentiel de membrane suffisamment négatif. Les neurones sont comme fatigués après l'émission d'un potentiel d'action. Les potentiels d'actions peuvent parfois être émis en salve avant que le neurone n'entre dans sa période réfractaire.

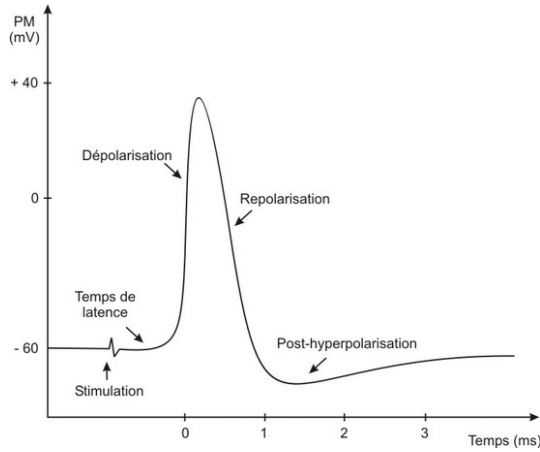


FIGURE 4 – Le potentiel d'action

Le potentiel d'action a toujours la même forme, à peu près la même durée et une amplitude d'environ 100 mV. Il code donc l'information par sa fréquence : c'est un phénomène binaire, 1 ou 0.

De nombreux modèles dynamiques permettent de décrire la décharge des neurones en prenant pour variable le potentiel transmembranaire.

Historiquement, le premier modèle électrophysiologique du neurone est le modèle *Intègre-et-tire* (*Integrate-and-fire* en anglais) développé au début du XXème siècle par Louis Lapicque. Il n'était pas possible à cette époque de mesurer directement l'activité électrique d'un seul neurone, aussi le modèle ne contient-il aucun paramètre physiologique. C'est un modèle phénoménologique qui décrit qualitativement la décharge du neurone sous l'effet d'un courant électrique extérieur. On modélise le neurone comme un circuit contenant une résistance R et un condensateur C et soumis à un courant extérieur $I(t)$. La loi des mailles s'écrit, en termes de potentiel et eu égard aux propriétés du condensateur et de la résistance,

$$C \frac{dV}{dt}(t) = I(t) - \frac{1}{R}V(t).$$

On décide que le neurone a déchargé dès que $V \geq V_{th}$ et on réinitialise alors le potentiel V . Posons $I_{th} = \frac{V_{th}}{R}$. Un courant $I(t)$ faible relativement à I_{th} ne fera pas décharger le neurone, un courant fort le fera décharger, parfois même en série.

À partir des années 1930, des méthodes de mesures du potentiel transmembranaire ont été mises au point et des modèles plus précis et physiologiques ont pu voir le jour. C'est le cas du fameux modèle de Hodgkin-Huxley en 1952, qui

prend en compte les mouvements transmembranaires des ions sodium (Na^+) et potassium (K^+). Si on note I le courant total de la membrane, I_{Na} le courant sodique, I_K le courant potassique, I_f un courant de fuite, C la capacité de la membrane et V la différence de potentiel de la membrane par rapport au potentiel de repos, on a

$$C \frac{dV}{dt} = I - I_{Na} - I_K - I_f,$$

puis on exprime les courants I_{Na} , I_K , I_f en fonction de V , des potentiels d'équilibre V_{Na} , V_K , V_f associés et des conductances de la membranes pour les ions et le terme de fuite $g_{Na}(V)$, $g_K(V)$, $g_f(V)$ qui dépendent de V :

$$I_{Na} = g_{Na}(V)(V - V_{Na}), \quad I_K = g_K(V)(V - V_K) \quad I_f = g_f(V)(V - V_f).$$

Il est ensuite possible à partir de données expérimentales de trouver des équations différentielles pour les conductances pour compléter le système.

De nombreux modèles ont été, et sont encore, dérivés de celui-ci par simplification ou modification, comme le célèbre modèle de FitzHugh-Nagumo dont la clarté mathématique et les propriétés abstraites ont fait le succès :

$$\begin{cases} \varepsilon x'(t) &= -y(t) + 4x(t) - x^3(t) + I \\ y'(t) &= a_0x(t) + a_1y(t) + a_2, \end{cases}$$

où $a_0, a_1, \varepsilon \in \mathbb{R}_+^*$, et $a_2 \in \mathbb{R}$ sont des paramètres. Si on les choisit de sorte que la cubique $y = 4x - x^3 + I$ a trois points d'intersection avec la droite $a_0x + a_1y + a_2 = 0$, qui donnent trois points singuliers pour le système, la dynamique présente des phénomènes intéressants tant mathématiquement que du point de vue de la modélisation : pour I sous un certain seuil I_{th} , il y a convergence en temps long vers un point singulier pour les solutions, si I est au-dessus du seuil, toutes les orbites convergent (au sens des ensembles ω -limites que nous verrons plus loin) vers la même orbite périodique. Avec un peu de travail, on peut même voir apparaître un canard (si, un canard) dans cette équation, comme expliqué sur la page internet <http://images.math.cnrs.fr/Des-canards-dans-mes-neurones.html>.

4.b. Flot d'un champ de vecteur

Un champ de vecteurs autonome, dans $\mathbb{R}^2$ ou $\mathbb{R}^3$ par exemple, définit des *trajectoires* ou *courbes intégrales* locales. Si le champ de vecteurs est localement lipschitzien, le théorème de Cauchy-Lipschitz nous assure qu'il existe pour chaque point de U une unique trajectoire passant par x et de longueur maximale. Nous souhaitons développer une vision plus globale, par exemple en observant les points se déplacer collectivement sous l'effet du champ de vecteurs : c'est le flot du système dynamique.

Définition 153 (Flot, Espace des phases). *Soient U un ouvert de $\mathbb{R}^n$ et $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne. Soit $\mathcal{D}(f) \subset \mathbb{R} \times \mathbb{R}^n$ l'ensemble*

des points (t, x) tels que si J_x est l'intervalle de définition maximal de la solution de (8) qui vaut x en 0, alors $t \in J_x$. On appelle *flot* de l'équation (8) (ou encore *flot* de f) l'application

$$\phi : (t, x) \mapsto \phi_t(x),$$

définie sur $\mathcal{D}(f)$ et à valeurs dans $\mathbb{R}^n$ telle que pour tout $(t, x) \in \mathcal{D}(f)$, la fonction $J_x \times \{x\} \subset \mathcal{D}(f)$ et la fonction $t \mapsto \phi_t(x)$, définie sur J_x est la solution maximale de l'équation (8) qui vaut x pour $t = 0$.

Dans la définition précédente, l'ouvert U est appelé *espace des phases* du champ de vecteurs (ou du système différentiel associé) et l'application ϕ_t est appelée le *flot* au temps t du champ f (ou du système différentiel associé).

De plus, si l'application ϕ est définie pour tout temps $t \in \mathbb{R}$ et toute condition initiale $x \in U$, i.e. si $\mathcal{D}(f) = \mathbb{R} \times \mathbb{R}^n$, alors le flot est dit *complet*.

Remarque 154. On remarquera que $\{0\} \times U \subset \mathcal{D}(f)$ et que le flot vérifie trivialement

$$\forall x \in U, \quad \phi_0(x) = x.$$

Théorème 155

Soit U un ouvert de $\mathbb{R}^n$ et $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne. L'ensemble $\mathcal{D}(f)$ est un ouvert et le flot ϕ est une application continue de $\mathcal{D}(f)$ dans $\mathbb{R}^n$.

Ce résultat, très important, signifie que ce ne sont pas seulement les trajectoires qui sont continues mais aussi le mouvement d'ensemble des points. Deux points qui sont proches au temps 0 resteront arbitrairement proches si on les regarde pendant un temps suffisamment court. En particulier, la solution du système dynamique dépend continument de la condition initiale et d'éventuels paramètres du système (il suffit pour le voir d'appliquer le théorème précédent en prenant ledit paramètre comme variable). Or, comme nous l'avons vu au paragraphe précédent, dans le contexte de la modélisation, il est commun de considérer une dynamique dépendant d'un ou de plusieurs paramètres.

Le flot permet de parler de manière efficace et élégante de la plupart des notions relatives aux équations différentielles.

Définition 156 (Orbite). Soit U un ouvert de $\mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne et ϕ le flot de f . Soit $x \in U$, on appelle *trajectoire* de x , ou *orbite* de x et on note $\mathcal{O}(x)$ l'ensemble

$$\mathcal{O}(x) = \{\phi_t(x) \mid t \in J_x\}.$$

Cette courbe est orientée par le sens de variation de t . En chacun de ses points $x(t) = \phi_t(x)$, sa tangente est la droite passant par $x(t)$ et dirigée par le vecteur $f(x(t))$. On distingue éventuellement l'orbite positive $\mathcal{O}_+(x)$ et l'orbite négative $\mathcal{O}_-(x) = \text{de } x$:

$$\mathcal{O}_+(x) = \{\phi_t(x) \mid t \in J_x \cap [0, +\infty[\} ; \quad \mathcal{O}_-(x) = \{\phi_t(x) \mid t \in J_x \cap]-\infty, 0] \}.$$

De plus, le flot a une structure particulière, semblable à celle des semi-groupes d'opérateurs (voir Annexe C.).

Théorème 157

Soient U un ouvert de $\mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne et ϕ le flot de f .

$$\forall (t_1, x), (t_2, x) \in \mathcal{D}(f), \quad \phi_{t_1+t_2}(x) = \phi_{t_1}(\phi_{t_2}(x)) = \phi_{t_2}(\phi_{t_1}(x)).$$

Par unicité dans le théorème de Cauchy-Lipschitz l'ensemble des orbites forme naturellement une partition de l'espace des phases U .

Définition 158 (Portrait de phase). *On appelle portrait de phase la partition de U formée par (le tracé des) les orbites dans l'ouvert U .*

Ce n'est pas vraiment une définition mathématique mais plutôt une précieuse aide visuelle à la compréhension des systèmes dynamiques, en particulier en dimension 2. On peut en effet ramener l'étude de la dynamique d'un système à l'étude de la géométrie du portrait de phase. Par exemple, une orbite fermée dans le portrait de phase (cf le système de Lotka-Volterra, figure 2a) traduit une solution périodique (figure 2b).

Dans le portrait de phase, certains points sont plus important que d'autre : les points stationnaires.

Définition 159 (Point stationnaire, point régulier). *Soient U un ouvert de $\mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne. On appelle point stationnaire, ou point singulier, du champ de vecteurs f les points $x \in U$ tels que $f(x) = 0$. On appelle points réguliers les autres points.*

Remarque 160. *Pour un point stationnaire x , l'orbite $\mathcal{O}(x)$ associée est réduite à $\{x\}$. En effet, $t \mapsto x$ est solution sur $\mathbb{R}$, donc maximale, donc unique. D'où le nom de ces points. Ces points sont tout désignés pour attirer dans le passé ou dans le futur les autres orbites, comme nous le verrons dans la section suivante.*

On peut montrer qu'en dehors des points singuliers, il est possible de redresser le flot au sens où les orbites sont localement rectilignes et parallèles. Lors de l'analyse qualitative des systèmes dynamiques, on se concentre donc sur l'étude du portrait de phase au voisinage des points singuliers.

4.c. Comportement asymptotique

Définition 161 (Ensemble ω -limite, Ensemble α -limite). *Soient U un ouvert de $\mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne et ϕ le flot de f . Soit $x \in U$. On appelle ensemble ω -limite de x l'ensemble*

$$\omega(x) = \bigcap_{t \in \mathbb{R}} \overline{\{\phi_s(x) \mid s > t\}}.$$

De même, on appelle ensemble α -limite de x l'ensemble

$$\alpha(x) = \bigcap_{t \in \mathbb{R}} \overline{\{\phi_s(x) \mid s < t\}}.$$

On prendra garde que ces ensembles peuvent être vides. Cette définition abstraite sert à poser un cadre mathématique sur la notion intuitive de comportement limite dans le passé ou le futur. Pour l'ensemble ω -limite, par exemple, on considère l'ensemble $\{\phi_s(x) \mid s > t\}$ qui n'est rien d'autre que la trajectoire de x après le temps t sous l'effet du champ de vecteur ; on en prend l'adhérence pour avoir tous les points qui sont infiniment proches de la trajectoire dans le futur ; puis on ne garde que ce qui reste dans toutes les fins de trajectoires, aussi courtes soient-elles. Un point dont la trajectoire est périodique de période T sur $\mathbb{R}$ a pour ensemble ω -limite et α -limite l'ensemble

$$\{\phi_t(x) \mid t \in [0, T[),$$

où l'on a pu retirer l'adhérence parce que la trajectoire est compacte (image d'un compact par une application continue) donc fermée. Une trajectoire qui vérifie $\lim_{t \rightarrow -\infty} \phi_t(x) = a$ et $\lim_{t \rightarrow +\infty} \phi_t(x) = b$ a pour ensemble ω -limite $\{b\}$ et pour ensemble α -limite $\{a\}$.

Théorème 162

Soient U un ouvert de $\mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne et ϕ le flot de f . Soit $K \subset U$ un compact et $x \in U$ tel que $\mathcal{O}(x) \subset K$. Alors $\omega(x)$ et $\alpha(x)$ sont des ensembles compacts, connexes et non-vides.

Le cas particulier d'un ensemble ω -limite ou α -limite ponctuel nous renvoie à la question des points stationnaires. En effet, nous avons le théorème suivant.

Théorème 163

Soient U un ouvert de $\mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne et ϕ le flot de f . Soit $x, x^* \in U$ tel que $\omega(x) = \{x^*\}$. Le point x^* est alors stationnaire.

En regardant d'un seul tenant tous les ensembles ω -limites, on arrive naturellement à l'idée d'ensemble attracteur dans le futur, que voici :

Définition 164 (Attracteur). Soient U un ouvert de $\mathbb{R}^n$, $f : U \rightarrow \mathbb{R}^n$ une application localement lipschitzienne et ϕ le flot de f . On appelle attracteur futur (respectivement passé) le plus petit ensemble qui contient $\omega(x)$ (respectivement $\alpha(x)$) pour tout $x \in U$.

L'attracteur, comme son nom l'indique, attire sur le long terme toutes les trajectoires. De manière plus générale, on peut caractériser les ensembles qui capturent une partie des trajectoires.

Définition 165 (Bassin d'attraction). Soit $A \subset U$ un ensemble fermé. On appelle bassin d'attraction dans le futur de A l'ensemble

$$B_\omega(A) = \{x \in U \mid \omega(x) \subset A\}.$$

On appelle bassin d'attraction dans le passé de A l'ensemble

$$B_\alpha(A) = \{x \in U \mid \alpha(x) \subset A\}.$$

Déterminer le passé ou l'avenir des points, ou décrire les différentes choses qui peuvent leur arriver en temps infiniment long sans avoir à résoudre l'équation (ni analytiquement ni numériquement) est un des objectifs principaux de l'étude des systèmes dynamiques. Dans le cas particulier de la dimension 2, il est possible de trouver tous les cas possibles, comme l'exprime le très beau théorème suivant.

Théorème 166 (Théorème de Poincaré-Bendixson)

Soient U un ouvert de $\mathbb{R}^2$, $f : U \rightarrow \mathbb{R}^2$ une application localement lipschitzienne et ϕ le flot de f . Soit $x \in U$, si $\omega(x)$ est non-vide, compact et ne contient aucun point fixe du champ de vecteur f , alors $\omega(x)$ est une orbite périodique de f .

Corollaire 167. Soit U un ouvert de $\mathbb{R}^2$, $f : U \rightarrow \mathbb{R}^2$ une application localement lipschitzienne et ϕ le flot de f . Soit $x \in U$. Si l'orbite de x est contenue dans un ensemble compact inclus dans U , alors

- soit $\omega(x)$ consiste en des points fixes de f ;
- soit $\omega(x)$ consiste en une orbite périodique ;
- soit $\omega(x)$ consiste en des points fixes de f et en des trajectoires les reliant.

Il est par contre (heureusement pour la richesse mathématique de l'étude des systèmes dynamiques) impossible d'étendre ce résultat aux dimensions supérieures. Dès la dimension 3, certains systèmes dynamiques présentent des attracteurs dits *étranges*, dont la géométrie est fractale. C'est le cas du fameux système dynamique de Lorenz¹⁷. Ce système, qui s'écrit

$$\begin{cases} x'(t) &= \sigma(y(t) - x(t)) \\ y'(t) &= \rho x(t) - y(t) - x(t)z(t) \\ z'(t) &= x(t)y(t) - \beta z(t), \end{cases}$$

où σ, ρ, β sont des paramètres. C'est un modèle simple pour décrire certains phénomènes de convection liés au couplage de l'atmosphère terrestre avec l'océan, obtenu par simplification d'un système d'équations aux dérivées partielles de type Navier-Stokes très complexe à étudier (et impossible à traiter numériquement du temps de Lorenz, d'où son intérêt pour le système différentiel plus simple). Les variables x , y et z décrivent respectivement l'intensité du mouvement de convection, la différence de température entre les courants ascendants et descendants et l'écart du profil de température vertical par rapport à un profil linéaire. Pour les choix de paramètres $\sigma = 10$, $\beta = \frac{8}{3}$ et $\rho = 28$, toutes les trajectoires adoptent un comportement chaotique après une trentaine de seconde et elle convergent toutes vers l'attracteur fractal appelé *attracteur de Lorenz*, vu à travers la trajectoire d'un point sur la figure 5.

17. Edward Norton Lorenz (1917 - 2008), météorologue américain au MIT, découvre par hasard en 1963 par la simulation numérique des comportements chaotiques pour un système très simple (le système de Lorenz). La découverte révolutionne l'idée de chaos et de complexité. Quoique Henri Poincaré en ait eu l'intuition des décennies plus tôt, c'était la première preuve concrète d'un phénomène chaotique issue de règles simples.

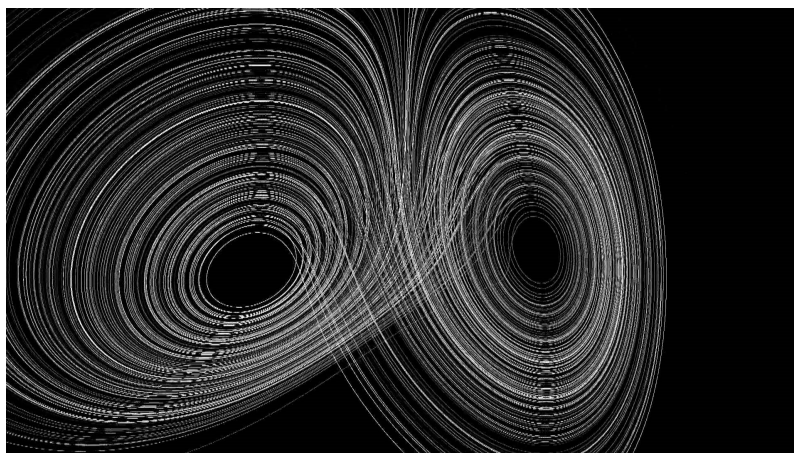


FIGURE 5 – Trajectoire dans $\mathbb{R}^3$ d'un point du système de Lorenz, trois millions d'itérations avec un pas de temps de 0.0001, trouvé sur la chaîne YouTube de Geoff Milburn, <https://www.youtube.com/watch?v=G3h1QTe9XIE>.

L'attracteur de Lorenz n'est pas seulement une abstraction mathématique, il caractérise une tendance globale chaotique dans un phénomène météorologique réel. Le côté racoleur de l'appellation *Théorie du chaos* a été décrié par une partie de la communauté mathématique lors de son apparition, pourtant, ce fut un puissant moteur de vocation scientifique et de vulgarisation pour le grand public. En 1972 Lorenz posera devant l'Association américaine pour le progrès des Sciences la célèbre question "Le battement d'aile d'un papillon au Brésil peut-il déclencher une tornade au Texas"¹⁸ qui inspira une somme colossale d'œuvres scientifiques, littéraires et cinématographiques¹⁹. Insistons sur le

18. La conférence était intitulée *Predictability : Does the Flap of a Butterfly's Wings in Brazil Set off a Tornado in Texas?*. Conscient de l'hostilité potentielle de la communauté scientifique vis-à-vis de cette manière de décrire le problème, Lorenz ajoutera lors de cette conférence : " De crainte que le seul fait de demander, suivant le titre de cet article, « Un battement d'ailes de papillon au Brésil peut-il déclencher une tornade au Texas? », fasse douter de mon sérieux, sans même parler d'une réponse affirmative, je mettrai cette question en perspective [...]". L'idée du papillon a probablement pour origine la nouvelle *A sound of thunder* de Ray Bradbury, publiée en 1952

19. On citera le film *l'Effet papillon* d'Eric Bress et J. Mackye Gruber (2004) qui représente un tournant idéologique majeur de la science-fiction après le caractère extrêmement fataliste du classique des classiques dans l'approche littéraire du voyage temporel : *La machine à explorer le temps* (*The Time Machine : An Invention* en langue originale) d'H.G Wells (nouvelle en 1895! Film en 1960 par George Pal et en 2002 par Simon Wells, arrière-petit-fils d'H.G. Wells). Mais plus que dans le rapport au temps, le concept s'est inscrit dans un changement de paradigme plus global de la science-fiction : les conséquences néfastes des actions humaines ne sont plus une punition divine, et donc transcendantes, mais sont issues de l'imprédictibilité de l'avenir liée au chaos, et sont donc immanentes. Dans le roman *Jurassic Park* de Michael Crichton (1990), puis le film éponyme de Steven Spielberg (1993), le personnage Ian Malcolm, mathématicien spécialiste de la théorie du Chaos, tente d'expliquer par les mathématiques qu'il est impossible de garder le Parc jurassique sous contrôle. La large audience du film (et

fait que l'idée du chaos dans les systèmes complexes est très ancienne et que la révolution apportée par Lorenz et son attracteur étrange est d'avoir prouvé que certains systèmes simples sont intrinsèquement sources de chaos, et donc que la complexité du réel ne provient pas seulement de la dépendance en un grand nombre de variables mais parfois simplement de l'interdépendance de quelques variables. Nous citons aussi la lutte, souvent oubliée, de Jean-Christophe Yoccoz, médaille Fields 1994, pour la reconnaissance de l'utilité en diffusion des sciences d'une théorie du chaos rigoureuse au sensationnalisme assumé.

De nombreux systèmes biologiques, chimiques et physiques plus ou moins complexes présentent de tels comportements, par exemple la dispersion des gènes dans les populations bactériennes due à la forte influence des transferts horizontaux. La biologie est par excellence très sensible au problème de la sensibilité aux conditions initiales qui est au cœur de la théorie du chaos. Qu'on ne pense pas pour autant que la reconnaissance du chaos comme phénomène réel et mathématique est un aveu d'échec face à la prédiction ou la compréhension des systèmes dynamiques et des réalités qu'ils recouvrent. Quoique la prédiction à long terme des trajectoires individuelles soit impossible tant analytiquement que numériquement, les divers outils d'analyse qualitative que nous avons développés permettent de dire beaucoup de choses sur le comportement global du système. Ce que nous décrivons alors n'est plus le comportement des individus mais celui du système composé de l'ensemble des individus : on parle en termes globaux ou statistiques. En philosophie, on décrit cette notion sous le terme *Émergence* : une propriété d'un ensemble d'objets est émergente si elle survient des propriétés des objets élémentaires, n'est pas une propriété de ces objets pris séparément, est distincte de toute propriété structurelle des objets pris indépendamment et possède une influence causale descendante sur les objets constituant l'ensemble sujet à la propriété émergente.

Cette idée très générale est la fondation de la compréhension de l'idée d'écosystème, de réseau trophique, de rétro-contrôle et de la réflexion sur la stabilité et l'évolution de ces derniers. La question de l'influence causale descendante est capitale en biologie, car contrairement à la plupart des phénomènes physiques – diffusion de la chaleur, comportement des plasmas, hydrodynamique,... – pour lesquels l'influence causale descendante est faible au sens où la dynamique du tout est entièrement déterminée par la dynamique causale des parties, de nombreux phénomènes biologiques – génétique des populations, sélection naturelle, groupements d'animaux, insectes sociaux, neurosciences,... – présentent une influence causale descendante forte, au sens où les propriétés émergentes du tout ont une influence causale directe sur les parties élémentaires – gènes, cellules, neurones, individus d'une population,...

le fait que cet aspect du livre ait été préservé!) ont grandement contribué à diffuser dans la culture populaire une idée mathématique nouvelle et par certains aspects très subtile.

5. Synthèse et pratique

5.a. Synthèse de la section

Nous avons étudié les équations différentielles de la forme $y' = f(t, y)$.

Si la fonction f est continue, nous avons prouvé l'existence de solutions. Si elle est de plus localement lipschitzienne par rapport à la variable y , nous avons prouvé l'unicité de la solution maximale.

Nous avons ensuite montré que lorsque les solutions n'existent pas pour tous temps, c'est parce qu'elle s'échappent vers l'infini ou vers la frontière du domaine spatial. La seule raison d'une non-existence en temps long pour ces équations est l'explosion en temps fini.

Avec le lemme de Grönwall, nous avons obtenu un critère utile pour prouver que les solutions existent pour tout temps, qui fonctionne en particulier pour toutes les équations linéaires en espace.

Lorsqu'il n'est pas possible d'exprimer les solutions explicitement, de nombreux outils permettent de prédire des comportements d'ensemble, même dans des cas où la dynamique est chaotique, c'est-à-dire très sensible aux conditions initiales. On peut pour cela considérer le portrait de phase, le flot, les attracteurs et bassins d'attraction, et démontrer des résultats qualitatifs très généraux comme le théorème de Poincaré-Bendixson.

5.b. Exercices théoriques et de modélisation

Exercice 168 (Modèle logistique, dit aussi de Verhulst). *Pour représenter la croissance d'une population d'êtres vivants qui se reproduisent dans un environnement qui ne peut nourrir qu'une densité finie d'individus en même temps, on peut utiliser en première approche le modèle suivant*

$$n'(t) = \alpha n(t) \left(1 - \frac{n(t)}{k}\right), \quad t \in \mathbb{R},$$

avec une condition initiale $n(0) = n_0$, où n est une fonction représentant la densité d'individus au temps t , n_0 la densité initiale d'individus au temps 0. Les nombres α et k sont deux paramètres strictement positifs.

1. Que modélisent les paramètres α et k ?
2. Montrer que ce problème de Cauchy admet des solutions globales et en donner une expression explicite en fonction des paramètres et de la condition initiale.

Exercice 169 (Modélisation d'un réseau de neurone). *Considérons n neurones connectés les uns aux autres. On représente leur activité par une fonction $v : \mathbb{R} \rightarrow \mathbb{R}^n$, où $v(t) = (v_1(t), \dots, v_n(t))$ représente l'activité au temps t de chacun*

des neurones exprimée comme une valeur réelle (positive ou négative, ce n'est pas un problème si on parle en terme de potentiel). On représente les connexions du réseau de neurones par une matrice $W \in \mathcal{M}_n(\mathbb{R})$. Le coefficient w_{ij} de W représente la connexion qui part du neurone j et va au neurone i . On note $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ la fonction définie par

$$S(x_1, \dots, x_n) = \left(\frac{2}{1 + e^{-x_1}} - 1, \dots, \frac{2}{1 + e^{-x_n}} - 1 \right),$$

et on modélise l'évolution de l'activité du réseau de neurones par l'équation différentielle

$$v'(t) = -v(t) + WS(v(t)).$$

On représente un stimulus extérieur par une fonction $u \in \mathcal{C}^\infty(\mathbb{R}_+, \mathbb{R}^n)$ de période τ , et on fait l'hypothèse supplémentaire que la fonction u est solution du système dynamique

$$u' = \xi(u), \quad (*)$$

où $\xi \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}^n)$, ce qui revient à dire que le stimulus est produit par un phénomène déterministe.

L'objectif est de faire "apprendre" au réseau de n neurones la dynamique définie par le champ de vecteurs lisse ξ , en un sens à préciser. Nous considérons que le réseau de neurones a "appris" le phénomène extérieur si il évolue comme lui lorsque l'état de départ est le même. Nous allons essayer de trouver des matrices de connexions W qui permettent cela. Pour cela, posons la fonctionnelle

$$H(u, W) = \frac{1}{2} \int_0^\tau \left\| -u(t) + W.S(u(t)) - \xi(u(t)) \right\|^2 dt,$$

où la norme considérée est la norme euclidienne sur $\mathbb{R}^n$.

1. Que modélise la fonction S ?
2. Montrer que pour toute condition initiale les solutions v de l'activité nerveuse existent sur $\mathbb{R}$.
3. Pourquoi, à u fixé, les matrices W qui minimisent $H(u, \cdot)$ donnent-elles une activité nerveuse proche du phénomène extérieur u ?
4. La fonctionnelle est convexe en W et de classe $\mathcal{C}^\infty$. Cela signifie qu'elle admet un unique minimum si et seulement si son gradient s'annule en un point. On pose la notation

$$x.y^\tau := \left(\int_0^\tau x_i(t)y_j(t) dt \right)_{1 \leq i, j \leq n}.$$

Montrer que si $S(u).S(u)^\tau$ est inversible, la fonctionnelle admet un unique minimum qui s'écrit

$$W^* = \left(\xi(u).S(u)^\tau + u.S(u)^\tau \right) \left(S(u).S(u)^\tau \right)^{-1}.$$

V. ÉQUATIONS AUX DÉRIVÉES PARTIELLES

« Toute loi physique doit être empreinte de beauté Mathématique. »

Paul Dirac

Le physicien Paul Dirac (1902-1984) était connu pour son génie et son laconisme. Il usait de fort peu de mots et ne disait jamais rien d'inutile²⁰. L'obsession de son existence était la description des phénomènes physiques la plus simple et mathématiquement élégante possible. Une de ses plus grandes contributions, l'équation de Dirac qui a permis plus tard d'envisager l'existence des anti-particules, est une équation aux dérivées partielles. De manière générale, les équations aux dérivées partielles sont des outils puissants pour décrire des phénomènes physiques et biologiques qui dépendent de plusieurs variables (souvent l'espace et le temps, ou plusieurs coordonnées d'espace).

Cependant, il serait présomptueux de vouloir étudier le concept général d'équation aux dérivées partielles comme nous l'avons fait pour les équations différentielles ordinaires²¹. Deux équations aux dérivées partielles semblables du point de vue de la graphie peuvent avoir des propriétés complètement différentes pour ce qui est de la régularité des solutions, de la conservation de certaines grandeurs ou des méthodes nécessaires à leur étude. On distingue en général trois grandes familles d'équations aux dérivées partielles : les équations paraboliques, elliptiques et hyperboliques ; cette distinction a un rapport avec les courbes du plan décrites par ces mots, comme nous l'expliquerons plus loin. Plutôt que de citer des théorèmes généraux impossible à démontrer en un temps raisonnable ou de donner une liste de méthodes abscones, nous allons nous pencher sur des équations particulières pour illustrer les différences de comportement entre les différents types d'équations. Les bourbakistes acharnés trouveront leur bonheur

20. Un peu l'inverse de l'auteur de ces notes qui est prolixe et dont le génie laisse à désirer !

21. Mais l'être humain n'est-il pas présomptueux pas nature ? Quoique je n'eusse pas trouvé la source, je tombai un jour, sur Facebook (si, si...), sur une photo d'une définition très générale du concept d'EDP que voici aussi fidèle que possible : *"Let H an Hilbert space and Σ a V -manifold, and let $Y \in H/\Sigma$ be an object in the slice over Σ thought as a bundle. A generalized partial differential equation on the space of sections of Y is an object $\mathcal{E} \in H/\Sigma$ together with a monomorphism into the infinite jet bundle of $Y : \mathcal{E} \rightarrow J_{\Sigma}^{\infty} Y$."* J'offre une bière à quiconque me trouve la référence et un resto (prix raisonnable) à quiconque m'explique en détail ce concept !

dans le contenu (beaucoup) plus abstrait de l'annexe C. sur les semi-groupes d'opérateurs. Toutefois, nous insistons sur le fait qu'il est absurde d'avoir du mépris pour les bestiaires d'équations aux dérivées partielles, car derrière leur aspect zoologique se cachent des concepts aussi riches et profonds que le contenu des autres théories mathématiques.

Cette partie se base majoritairement sur les livres [12], [3] ainsi que sur des cours d'équations aux dérivées partielles donnés à l'université d'Orsay par Frédéric Lagoutière, Nicolas Burq, Patrick Gérard, Bertrand Maury et Martina Hoffmanova (cours doctoral). L'auteur met également ici à profit l'expérience en EDP transmise par ses directrices de thèse Danielle Hilhorst et Delphine Salort.

1. Les équations de transport

1.a. Transport à vitesse constante

L'équation de transport à vitesse constante est une des équations aux dérivées partielles les plus simples. Il s'agit simplement, étant données une vitesse constante c et une fonction u_0 de classe $\mathcal{C}^1$ sur $\mathbb{R}$, de trouver une solution u à l'équation

$$\begin{cases} \frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0 & x \in \mathbb{R}, t \in \mathbb{R} \\ u(x, 0) = u_0(x) & x \in \mathbb{R}, \end{cases} \quad (3)$$

où u est une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$.

Comme souvent en équations aux dérivées partielles, l'analyse qualitative et la construction effective de solutions se font *via* l'exploitation des spécificités structurelles de l'équation étudiée. Plus prosaïquement, si la vie vous donne des courbes caractéristiques, il faut les dessiner (voir figure 6) et s'en servir.

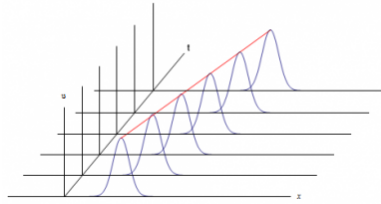


FIGURE 6 – Courbe caractéristique (en rouge) et transport de la condition initiale le long de celle-ci au cours du temps

Lemme 170. *Les solutions u de classe $\mathcal{C}^1$ de l'équation (3) sont constantes sur tout ensemble de la forme*

$$C_a = \{(a + ct, t) \mid t \in \mathbb{R}\},$$

où a est un nombre réel.

Démonstration : Pour toute solution u , la fonction $t \mapsto u(a + ct, t)$ est dérivable et, d'après la règle de la chaîne, sa dérivée vaut en tout point t de $\mathbb{R}$

$$\frac{\partial u}{\partial t}(a + ct, t) \times 1 + \frac{\partial u}{\partial x}(a + ct, t) \times c = 0.$$

Donc la fonction en question est constante et égale à $u(a, 0) = u_0(a)$. ■

Ce résultat assez trivial est la base de la méthode générale dite des caractéristiques. Les caractéristiques peuvent en effet être définies ainsi

$$C_a = \{(x(t), t) \in \mathbb{R}^2 \mid t \in \mathbb{R}, x'(t) = c, x(0) = a\},$$

et cette définition a toujours un sens lorsque c dépend du temps et/ou de l'espace. Dans ce cas particulier, les caractéristiques sont des droites et il est possible de les utiliser pour résoudre complètement le problème de Cauchy (3), comme l'explique le théorème suivant, lui aussi relativement trivial.

Théorème 171

Il existe une unique solution u dans l'espace $\mathcal{C}^1(\mathbb{R}^2)$ qui vérifie l'équation (3) et elle s'écrit

$$\forall (x, t) \in \mathbb{R}^2, \quad u(x, t) = u_0(x - ct).$$

Démonstration : La fonction $u : (x, t) \mapsto u_0(x - ct)$ est dans l'espace $\mathcal{C}^1(\mathbb{R}^2)$ et vérifie l'équation (3). Soit une solution quelconque $v \in \mathcal{C}^1(\mathbb{R}^2)$ de (3). Elle coïncide avec u sur $\mathbb{R} \times \{0\}$, donc d'après le lemme 170 elle coïncide avec u sur chaque courbe caractéristique donc sur $\mathbb{R}^2$ tout entier. Il vient $u = v$ d'où l'unicité de la solution. ■

1.b. Transport à vitesse non constante

Dans certains cas, on peut utiliser la méthode des caractéristiques pour résoudre des problèmes de transport avec une vitesse qui dépend du temps et de l'espace. Il faut tout refaire pour chaque nouvelle fonction vitesse. Étudions par exemple l'équation suivante : soit une fonction u_0 de classe $\mathcal{C}^1$ sur $\mathbb{R}$, on cherche une fonction u de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$ vérifiant

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{x}{1+t^2} \frac{\partial u}{\partial x} = 0 & x \in \mathbb{R}, t \in \mathbb{R} \\ u(x, 0) = u_0(x) & x \in \mathbb{R}, \end{cases} \quad (4)$$

Ici encore, le problème se ramène à étudier les courbes caractéristiques qui s'écrivent cette fois, pour $a \in \mathbb{R}$,

$$C_a = \{(x(t), t) \mid t \in \mathbb{R}, x'(t) = \frac{x(t)}{1+t^2}, x(0) = a\}.$$

Théorème 172

Il existe une unique solution u dans l'espace $\mathcal{C}^1(\mathbb{R}^2)$ au problème (4) et elle s'écrit

$$\forall (x, t) \in \mathbb{R}^2, \quad u(x, t) = u_0(x e^{-\arctan(t)}).$$

Démonstration : Soit $a \in \mathbb{R}$, le problème de Cauchy

$$\begin{cases} x'(t) = \frac{1}{1+t^2} x(t) & t \in \mathbb{R} \\ x(0) = a \end{cases}$$

admet une unique solution globale d'après le corollaire de Grönwall (théorème 151). On vérifie en dérivant que cette unique solution est la fonction

$$t \mapsto a e^{\arctan(t)}.$$

De plus, on vérifie comme précédemment que toute solution régulière u de (4) est constante le long des courbes caractéristiques

$$C_a = \{(a e^{\arctan(t)}, t) \mid t \in \mathbb{R}\},$$

en dérivant la fonction $t \mapsto u(a e^{\arctan(t)}, t)$. ■

1.c. Solutions faibles pour des données irrégulières

Dans les applications, il est souvent utile de considérer des condition initiales irrégulières. Si l'on souhaite regarder le transport d'une masse compacte (mais pas solide) de matière organique dans un milieu liquide, on doit partir d'une donnée à support compact et discontinue : $u_0 = 1$ sur K compact, $u_0 = 0$ sur le complémentaire de K . Il faut alors changer la notion de solution avec laquelle nous travaillons.

En effet, si à présent la condition initiale u_0 n'est pas de classe $\mathcal{C}^1$ mais seulement continue, il est absolument vain de chercher des solutions dans $\mathcal{C}^1(\mathbb{R}^2)$. Il faut chercher les solutions dans l'espace $\mathcal{C}^0(\mathbb{R}^2)$ des fonctions continues sur $\mathbb{R}^2$. Le problème est que, dans ce cas, les dérivées partielles de la solution ne sont pas définies. On va donc chercher des solutions au sens faible. Nous allons définir une nouvelle notion de solution qui est utilisable même pour des fonctions bien plus désagréables que les fonctions seulement continues : les fonctions localement essentiellement bornées, dont nous notons $\mathbb{L}_{\text{loc}}^\infty(\mathbb{R}^n)$ l'ensemble.

On peut définir $\mathbb{L}_{\text{loc}}^\infty(\mathbb{R}^n)$ comme l'ensemble des (classes d'équivalence de) fonctions u mesurables de $\mathbb{R}^n$ dans $\mathbb{R}$ telles que pour tout $x \in \mathbb{R}^n$ il existe un voisinage ouvert U qui contient x et une constante M tels que pour presque tout y de U , $\|u(y)\| \leq M$. L'espace $\mathbb{L}_{\text{loc}}^\infty(\mathbb{R}^n)$ contient par exemple des (classes d'équivalence de) fonctions nulle-part continues ou continues et nulle-part dérivables. Nous pouvons pourtant sans mal donner un sens à une équations aux

dérivées partielles pour ces fonctions très irrégulières avec la notion de dérivée faible que nous avons vu dans la section précédente.

Pour une condition initiale $u_0 \in \mathbb{L}_{\text{loc}}^\infty(\mathbb{R})$, on dit que u est solution faible de l'équation

$$\begin{cases} \frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0 & x \in \mathbb{R}, t \in \mathbb{R}_+ \\ u(x, 0) = u_0(x) & x \in \mathbb{R}, \end{cases} \quad (5)$$

si pour toute fonction test $\phi \in \mathcal{C}_0^1(\mathbb{R}^2)$ on a

$$\int_{\mathbb{R}_+} \int \mathbb{R} u(x, t) \left(\frac{\partial \phi}{\partial t}(x, t) + c \frac{\partial \phi}{\partial x}(x, t) \right) dx dt + \int_{\mathbb{R}} u_0(x) \phi(x, 0) dx = 0. \quad (6)$$

On reporte le manque de régularité de u sur une fonction test régulière. Tout est donc bien défini. Remarquons tout de même que la fonction test n'est pas prise à support compact dans $\mathbb{R} \times \mathbb{R}_+$ car sinon le terme contenant la condition initiale disparaîtrait. Remarquons également, et surtout, que nous sommes obligés de mettre la donnée initiale dans l'équation au lieu de simplement demander $u(x, 0) = u_0(x)$. En effet, la fonction u est définie à un ensemble de mesure nulle près. Regarder ses valeurs sur le sous-ensemble $\mathbb{R} \times \{0\}$ de mesure nulle n'a pas de sens. Par contre, u_0 est définie à un ensemble de mesure nulle près, mais dans $\mathbb{R}$ (un ensemble de mesure nulle pour la mesure de Lebesgue dans $\mathbb{R}$, par dans $\mathbb{R}^2$), donc l'intégrale sur $\mathbb{R}$ de $u_0(\cdot)\phi(\cdot, 0)$ a un sens.

On vérifie sans mal par intégration par partie que si u est de classe $\mathcal{C}^1$ les solutions faibles et les solutions fortes sont les mêmes. Ce résultat est un corollaire du théorème suivant, car la solution est unique dans un espace plus gros que $\mathcal{C}^1(\mathbb{R}^2)$ (*i.e.* qui le contient).

Théorème 173

Soit $u_0 \in \mathbb{L}_{\text{loc}}^\infty(\mathbb{R})$. Il existe une unique solution faible dans $\mathbb{L}_{\text{loc}}^\infty(\mathbb{R} \times \mathbb{R}_+)$ au problème (5) et elle s'écrit

$$\forall (x, t) \in \mathbb{R}^2, \quad u(x, t) = u_0(x - ct).$$

Démonstration : La fonction $u : (x, t) \mapsto u_0(x - ct)$ définit bien un élément de

$\mathbb{L}_{\text{loc}}^\infty(\mathbb{R} \times \mathbb{R}_+)$ ²². On a par les théorèmes de changement de variable et de Fubini

$$\begin{aligned} & \int_{\mathbb{R}} \int_{\mathbb{R}_+} u_0(x - ct) \left(\frac{\partial \phi}{\partial t}(x, t) + c \frac{\partial \phi}{\partial x}(x, t) \right) dx dt \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}_+} u_0(y) \left(\frac{\partial \phi}{\partial t}(y + ct, t) + c \frac{\partial \phi}{\partial x}(y + ct, t) \right) dy dt \\ &= \int_{\mathbb{R}} u_0(y) \left(\int_{\mathbb{R}_+} \left(\frac{\partial \phi}{\partial t}(y + ct, t) + c \frac{\partial \phi}{\partial x}(y + ct, t) \right) dt \right) dy. \quad (7) \end{aligned}$$

Or, à y fixé, la fonction de classe $\mathcal{C}^1$ à support compact $t \mapsto \phi(y + ct, t)$ est une primitive de la fonction continue $t \mapsto \frac{\partial \phi}{\partial t}(y + ct, t) + c \frac{\partial \phi}{\partial x}(y + ct, t)$. On a donc

$$\int_{\mathbb{R}_+} \left(\frac{\partial \phi}{\partial t}(y + ct, t) + c \frac{\partial \phi}{\partial x}(y + ct, t) \right) dt = \lim_{R \rightarrow +\infty} \left[\phi(y + ct, t) \right]_{t=0}^{t=R} = -\phi(y, 0).$$

En injectant dans 7 on obtient

$$\int_{\mathbb{R}} \int_{\mathbb{R}_+} u_0(x - ct) \left(\frac{\partial \phi}{\partial t}(x, t) + c \frac{\partial \phi}{\partial x}(x, t) \right) dx dt = - \int_{\mathbb{R}} u_0(x) \phi(x, 0) dx.$$

ce qui achève de prouver que u est une solution faible.

Pour montrer l'unicité, supposons que $v \in \mathbb{L}_{\text{loc}}^\infty(\mathbb{R} \times \mathbb{R}_+)$ est une solution faible. La fonction $w = u - v \in \mathbb{L}_{\text{loc}}^\infty(\mathbb{R} \times \mathbb{R}_+)$ vérifie pour toute fonction test $\phi \in \mathcal{C}_0^1(\mathbb{R}^2)$,

$$\int_{\mathbb{R}} \int_{\mathbb{R}_+} w(x, t) \left(\frac{\partial \phi}{\partial t}(x, t) + c \frac{\partial \phi}{\partial x}(x, t) \right) dx dt = 0,$$

Supposons pour l'instant que pour toute fonction $f \in \mathcal{C}_0^1(\mathbb{R} \times \mathbb{R}_+^*)$ il existe $\psi \in \mathcal{C}_0^1(\mathbb{R}^2)$ telle que

$$f = \frac{\partial \psi}{\partial t} + c \frac{\partial \psi}{\partial x}.$$

Alors, nous avons pour tout $f \in \mathcal{C}_0^1(\mathbb{R} \times \mathbb{R}_+^*)$,

$$\int_{\mathbb{R}} \int_{\mathbb{R}_+} w(x, t) f(x, t) dx dt = 0,$$

ce qui implique que $w = 0$ presque partout et donc que $w = 0$ comme élément de $\mathbb{L}_{\text{loc}}^\infty(\mathbb{R} \times \mathbb{R}_+)$. La solution est donc unique.

Il reste à démontrer ce que nous avons supposé pour finir la preuve. Pour cela, remarquons que pour toute $f \in \mathcal{C}_0^1(\mathbb{R} \times \mathbb{R}_+^*)$, il existe un temps $T \in \mathbb{R}_+^*$ tel

22. Il est intéressant de remarquer qu'en étant parfaitement honnête, il n'est pas immédiat (mais ça reste relativement trivial dans son cadre théorique) qu'un représentant de la classe d'équivalence presque partout associée à u_0 définit bien par cette formule une fonction essentiellement bornée de $\mathbb{R} \times \mathbb{R}_+$. Cela provient de la définition de la mesure produit. Ce genre de choses font partie des abus de réflexion tolérés en EDP car nécessaires à la fluidité des raisonnements et à la lisibilité des preuves. Nous passerons des choses bien pire sous silence par la suite.

que si $t \leq T$, alors pour tout $x \in \mathbb{R}$, $f(x, t) = 0$. On il suffit de vérifier que le problème

$$\begin{cases} \frac{\partial \psi}{\partial t} + c \frac{\partial \psi}{\partial x} = f & \text{dans } \mathbb{R} \times]0, T] \\ \psi(x, T) = 0 & x \in \mathbb{R} \end{cases}$$

admet une solution.

La fonction définie par

$$\psi : (x, t) \mapsto \int_t^T f(x - c(s - t), s) \, ds$$

est une solution, ce qui conclut la preuve. ■

Le problème auxiliaire avec une condition finale en temps est appelé équation rétrograde. Les équations rétrogrades n'admettent pas toujours des solutions et sont un sujet difficile dans l'absolu.

2. L'équation de la chaleur

2.a. Origine physique de l'équation

Nous allons faire un peu de physique afin d'établir soigneusement l'équation en dimension 1 d'espace. Nous l'étudierons mathématiquement en dimension quelconque ensuite.

La *conduction thermique* est un transfert thermique dans la matière sans mouvement macroscopique (contrairement à des phénomènes comme la convection). Elle est la conséquence d'une inhomogénéité du champ de température. Au niveau microscopique, la conduction thermique correspond dans un métal à l'accroissement de proche en proche de l'énergie microscopique de vibration du réseau d'atomes le constituant, *via* les électrons libres²³ et dans un fluide à l'accroissement de l'énergie cinétique d'agitation désordonnée des molécules.

Afin de pouvoir travailler, il est important de se placer sous l'hypothèse d'un *équilibre thermodynamique local* afin que la notion de température en un point et à un instant donnés ait un sens. La grandeur température est définie de manière homogène sur une portion *mésoscopique* du système, cette portion est dite en équilibre thermodynamique local. Cette manière de faire ne convient pas pour modéliser une explosion ou le déplacement aléatoire d'un faible nombre de particules.

Considérons le cas simple d'un barreau métallique fixe, de constitution homogène, de section S et de masse volumique μ , assimilé à un segment $[a, b]$. On suppose que la conduction s'effectue uniquement de manière longitudinale

23. C'est pourquoi les bons conducteurs électriques sont de bons conducteurs thermiques, et réciproquement.

(de manière orthogonale à toute section latérale du barreau). Pour $t \in \mathbb{R}_+$ et $x \in [a, b]$, on note $T(x, t)$ la température au point x et au temps t .

Soient $x, t \in [a, b] \times \mathbb{R}_+$, et $dx, dt \in \mathbb{R}_+$ tels que les grandeurs étudiées soient constantes sur $[x, x + dx] \times [t, t + dt]$. On applique le premier principe de la thermodynamique au système $[x, x + dx]$ dont on note U l'énergie interne. Comme il n'y a ni travail d'une force ni énergie cinétique macroscopique, on obtient

$$dU = \delta Q.$$

Or $dU = \frac{\partial U}{\partial T} dT + \frac{\partial U}{\partial V} dV$, où V est le volume du barreau. Comme une barre métallique est une phase condensée, la transformation est isochore : $dV = 0$. En posant $\mu S dx C(T) = \frac{\partial U}{\partial T}$, on a $dU = S dx \mu C(T) dT$. La quantité $C(T)$ est la capacité thermique massique pour la température T .

Appelons $j(x, t)$ le flux thermique au point x et au temps t . On peut écrire :

$$S dx \mu C(T) dT = j(x, t) S dt - j(x + dx, t) S dt$$

$$\mu C(T) dT = - \frac{j(x + dx, t) - j(x, t)}{dx}$$

et pour dx et dt "assez petits", on obtient

$$\mu C(T) \frac{\partial T}{\partial t} = - \frac{\partial j}{\partial x}$$

On utilise alors la loi de Fourier, qui est une loi phénoménologique :

$$j = -\lambda \frac{\partial T}{\partial x}$$

où $\lambda \in \mathbb{R}_+$ est le coefficient de *conductivité thermique*²⁴, exprimé en $W.m^{-1}.K^{-1}$.

En injectant la loi de Fourier dans notre système, avec la supposition $C(T) \neq 0$, nous trouvons :

$$\frac{\partial T}{\partial t} = \frac{\lambda}{\mu C(T)} \frac{\partial^2 T}{\partial x^2} \quad (8)$$

L'équation (8) est appelée équation de la chaleur. C'est une équation aux dérivées partielles linéaire d'ordre 2.

On peut généraliser cette construction à la dimension 3 d'espace où à des dimensions supérieures, ce qui donne l'équation

$$\frac{\partial T}{\partial t}(x, t) - \kappa \Delta T(x, t) = 0.$$

L'équation de la chaleur est particulièrement intéressante en biologie parce qu'elle représente tous les phénomènes diffusifs. Une population d'animaux qui

²⁴. En toute rigueur, λ peut dépendre de la température, mais c'est négligeable aux températures usuelles.

se diffuse de manière homogène dans un milieu, un composé chimique qui diffuse dans l'eau ou le sang, l'oxygène qui se diffuse dans une alvéole pulmonaire,...

L'ajout d'un second membre ou d'un terme non-linéaire d'ordre 0 permet d'étudier des phénomènes de dynamique de population très complexes.

Par exemple, dans l'exercice 168 sur le modèle de Verhulst, on peut déplorer que l'équation ne prenne pas en compte la dimension spatiale de la conquête d'un milieu par des êtres vivants. C'est un modèle assez médiocre pour décrire par exemple la croissance d'une population d'animaux se déplaçant sur un large territoire ou la présence dans une population d'un allèle avantageux sur un gène. Pour ce genre de phénomènes, on peut étudier l'équation dite de Fisher-KPP. Ce modèle s'écrit

$$\begin{cases} \frac{\partial u}{\partial t}(x, t) = d\Delta u(x, t) + \alpha u(x, t) \left(1 - \frac{u(x, t)}{k}\right) & x \in \mathbb{R}^n, t \in \mathbb{R}_+ \\ u(x, 0) = u_0(x), & x \in \mathbb{R}^n \end{cases}$$

où $u(x, t)$ est la densité d'individu (ou la proportion de l'allèle avantageux) au point x et au temps t et d, α, k sont des paramètres strictement positifs. u_0 est une condition initiale qui décrit la population en tout point de l'espace au temps 0. On remarquera que si u_0 est constante, on retombe sur le modèle de Verhulst et on sait résoudre complètement le problème. Ce problème d'apparence simple a fait l'objet d'intenses recherches. L'exercice 188 lui est consacré.

De manière générale, les équations de la forme

$$\frac{\partial u}{\partial t}(x, t) = d\Delta u(x, t) + f(u(x, t)),$$

appelées équations de réaction-diffusion, sont d'une grande utilité pour l'étude mathématique de l'écologie, mais aussi de la médecine, de la chimie, de l'économie... Des fonctions f simples conduisent à des problèmes très difficiles et encore largement ouverts de nos jours.

De telles équations peuvent être couplées pour représenter des phénomènes de compétition ou de prédation comme nous l'avons vu pour les équations différentielles ordinaires.

2.b. Étude mathématique

On étudie dans un domaine ouvert Ω de $\mathbb{R}^n$ l'équation d'évolution

$$\begin{cases} \frac{\partial T}{\partial t}(x, t) - \kappa \Delta T(x, t) = 0 \\ T(x, 0) = T_0(x) \end{cases} \quad (9)$$

avec $x \in \Omega$, $t \in \mathbb{R}_+$ et $T_0 \in \mathbb{L}^\infty(\mathbb{R}^n) \cap \mathcal{C}^0(\mathbb{R}^n)$.

La donnée initiale est continue, donc définie partout et pas seulement presque partout. On pose l'équation au sens fort. Il est possible d'avoir des données initiales moins régulières et soit de poser l'équation au sens faible, comme nous l'avons fait pour l'équation de transport, soit, si on sait déjà que les solutions de cette équation ont tendance à être très régulières, en posant l'équation au sens fort avec une condition initiale vérifiée au sens d'une limite pour la norme appropriée. Le problème est qu'une limite uniforme de fonctions continues est continue, on ne peut donc pas choisir la norme de $\mathbb{L}^\infty$. Si on étudie le problème sur $]0, 1[$ en espace, les fonctions de $\mathbb{L}^\infty$ sont aussi dans $\mathbb{L}^2$, on peut donc définir la condition initiale en demandant

$$\lim_{t \rightarrow 0} \|T(\cdot, t) - T_0(\cdot)\|_2 = \lim_{t \rightarrow 0} \sqrt{\int_0^1 (T(x, t) - T_0(x))^2 dx} = 0,$$

où $T(\cdot, t)$ a un sens si la solution est une fonction régulière définie partout et pas seulement une fonction de $\mathbb{L}^p$ pour laquelle la restriction à une bande de $\mathbb{R}^2$ de mesure nulle n'a pas de sens *a priori* (nous insistons sur le *a priori* ; il est possible de donner un sens à cette restriction *via* la notion de *traces*).

Théorème 174

Soit $T_0 \in \mathbb{L}^2(\mathbb{R}^n)$, la fonction T , de carré intégrable (en espace) et de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^n \times \mathbb{R}_+^*$ définie par

$$T(x, t) = \frac{1}{(4\pi\kappa t)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-\frac{\|y\|^2}{4\kappa t}} T_0(x - y) dy,$$

est solution du problème (9).

On peut aussi montrer que cette solution est unique en prouvant d'abord un résultat appelé *principe du maximum* et qui traduit l'idée que la température ne peut se focaliser en un point : les maxima et minima de températures sont atteints sur le bord du domaine spatiotemporel.

La solution ci-dessus fonctionne aussi pour des données initiales qui ne sont pas de carré intégrable, par exemple continues et bornées, ou même mesurables et bornées mais définies partout.

Remarque 175. Si l'on pose

$$G(x, t) = \frac{1}{(4\pi\kappa t)^{\frac{n}{2}}} e^{-\frac{\|x\|^2}{4\kappa t}},$$

on voit que la solution s'écrit en fait

$$T(x, t) = G * T_0(x, t).$$

On dit que G est le noyau de la chaleur, ou encore que c'est une solution fondamentale, ou une fonction de Green, de l'équation de la chaleur. On pourra

consulter l'Annexe B. pour plus de détails sur les solutions fondamentales. La convolution donne un caractère non-local à l'équation : la condition initiale sur l'espace entier contribue en tout point pour des temps aussi petits que l'on veuille considérer. La régularité du noyau donne la régularité de la solution par convolution.

Théorème 176

Soit $T_0 \in \mathbb{L}^\infty \cap \mathbb{L}^2$ une donnée initiale et T une solution régulière de (9). On a pour tout temps $t \in \mathbb{R}_+^*$,

$$\|T(\cdot, t)\|_{\mathbb{L}^\infty(\Omega)} \leq \|T_0\|_{\mathbb{L}^\infty(\Omega)},$$

et

$$\|T(\cdot, t)\|_{\mathbb{L}^2(\Omega)} \leq \|T_0\|_{\mathbb{L}^2(\Omega)}.$$

2.c. Simulation numérique par les différences finies

Pour éviter d'alourdir les explications par des difficultés techniques dues à la dimension, nous allons étudier la résolution numérique de l'équation de la chaleur dans le segment $[0, 1]$ en espace, et sur l'intervalle $[0, t_{max}]$ en temps :

$$\begin{cases} \frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial x^2}, & (x, t) \in]0, 1[\times]0, t_{max}] \\ T(0, t) = 0, & t \in]0, t_{max}] \\ T(1, t) = 0, & t \in]0, t_{max}] \\ T(x, 0) = T_0(x), & x \in]0, 1[. \end{cases}$$

Nous allons développer une méthode de simulation numérique qui marche plus généralement pour de nombreuses équations aux dérivées partielles. C'est en vérité la même idée que celle du schéma d'Euler pour les équations différentielles ordinaires que nous avons utilisée pour montrer le théorème de Cauchy-Peano-Arzela.

L'idée principale est de se donner un maillage spatio-temporel discret, les ordinateurs étant par nature assez réticents à l'idée du continu. On représente le segment $[0, 1]$ par un maillage régulier ($0 = x_0, \dots, x_{n+1} = 1$) de pas d_x et le segment temporel $[0, t_{max}]$ par un maillage régulier ($0 = t_0, \dots, t_{m+1} = t_{max}$) de pas d_t . C'est-à-dire que $x_{i+1} - x_i = dx$ et $t_{j+1} - t_j = dt$, pour tous i, j inférieurs ou égaux à n, m . Cela revient aussi à dire que l'on a $x_i = idx$ et $t_j = jdt$.

On note T_i^j la valeur de notre approximation numérique au point (x_i, t_j) ; à ne pas confondre avec la valeur $T(x_i, t_j)$ de la véritable solution en ce même point du maillage. L'objectif est de rendre les deux aussi proches que possible. Les trois conditions aux limites du problèmes reviennent alors à imposer

$$\begin{cases} T_0^j = 0 & 0 \leq j \leq m+1 \\ T_{n+1}^j = 0 & 0 \leq j \leq m+1 \\ T_i^0 = T_0(x_i) & 0 \leq i \leq n+1 \end{cases}$$

en supposant pour simplifier que $T_0(0) = T_0(1) = 0$.

On va ensuite remplacer les dérivées partielles par des taux d'accroissement, et nous justifierons plus tard mathématiquement que cela permet d'avoir une bonne approximation de la solution :

$$\frac{\partial T}{\partial x}(x_i, t_j) \simeq \frac{T(x_i + \frac{dx}{2}, t_j) - T(x_i - \frac{dx}{2}, t_j)}{dx},$$

puis ensuite

$$\frac{\partial^2 T}{\partial x^2}(x_i, t_j) \simeq \frac{\frac{\partial T}{\partial x}(x_i + \frac{dx}{2}, t_j) - \frac{\partial T}{\partial x}(x_i - \frac{dx}{2}, t_j)}{dx},$$

soit encore

$$\frac{\partial^2 T}{\partial x^2}(x_i, t_j) \simeq \frac{\frac{T(x_i + dx, t_j) - T(x_i, t_j)}{dx} - \frac{T(x_i, t_j) - T(x_i - dx, t_j)}{dx}}{dx},$$

et enfin :

$$\frac{\partial^2 T}{\partial x^2}(x_i, t_j) \simeq \frac{T(x_{i+1}, t_j) - 2T(x_i, t_j) + T(x_{i-1}, t_j)}{dx^2}.$$

Pour la dérivée en temps, plusieurs choix sont possibles qui ne donnent pas les mêmes performances numériques. Le plus simple à utiliser (mais pas toujours le plus efficace numériquement) est de prendre

$$\frac{\partial T}{\partial t}(x_i, t_j) = \frac{T(x_i, t_{j+1}) - T(x_i, t_j)}{dt},$$

dit schéma d'Euler explicite (comme pour les EDO en gros).

On considère donc sur notre grille les formules

$$\frac{T_i^{j+1} - T_i^j}{dt} - \kappa \frac{T_{i+1}^j - 2T_i^j + T_{i-1}^j}{dx^2} = 0.$$

Comme on connaît la ligne spatiale pour le temps 0 grâce à la troisième condition initiale, on va déterminer les lignes spatiales pour les temps supérieurs à partir des lignes spatiales pour les temps inférieurs, et l'on voit que l'on peut le faire explicitement (d'où le nom de schéma explicite), ce que l'on peut voir en mettant l'équation sous la forme

$$T_i^{j+1} = T_i^j + \frac{\kappa dt}{dx^2} (T_{i+1}^j - 2T_i^j + T_{i-1}^j).$$

Introduisons un premier outil pour déterminer la qualité de l'approximation de l'équation de la chaleur par la méthode (ou schéma) d'Euler explicite :

Définition 177 (Erreur de consistance). *On appelle erreur de consistance d'un schéma numérique (en un point (x_i, t_j) du maillage) l'erreur commise en injectant la solution exacte dans l'équation approchée, c'est-à-dire*

$$\varepsilon_i^j(T) = \frac{T(x_i, t_{j+1}) - T(x_i, t_j)}{dt} - \kappa \frac{T(x_{i+1}, t_j) - 2T(x_i, t_j) + T(x_{i-1}, t_j)}{dx^2},$$

où T est solution exacte du problème.

Ces termes d'erreurs décrivent donc simplement l'erreur que l'on commet avec notre équation approchée sur une solution exacte. Notons pour simplifier

$$\varepsilon^j(T) = \left(\varepsilon_i^j \right)_{i \in \{0, \dots, n+1\}}$$

le vecteur des erreurs de consistance au temps j .

On veut donc montrer que quand les pas de temps et d'espace dt et dx se rapprochent de 0, les erreurs de consistance diminuent, ce qu'exprime mathématiquement le théorème suivant :

Théorème 178

Supposons que la solution exacte T du problème est de classe $\mathcal{C}^2$ en temps et $\mathcal{C}^4$ en espace sur $[0, 1] \times [0, t_{max}]$. Alors il existe $B \in \mathbb{R}_+$ tel que

$$\max_{j \in \{0, \dots, m+1\}} \|\varepsilon^j(T)\|_\infty \leq B(dt + dx^2)$$

où $\|\cdot\|_\infty$ est la norme infinie (le max des valeurs absolues des coefficients).

Démonstration : Commençons par comparer la qualité d'approximation des dérivées spatiale et temporelle séparément, pour ensuite tout réunir.

- Dérivée en temps :

T est de classe $\mathcal{C}^2$ en temps, donc d'après le théorème de Taylor-Lagrange appliqué pour x_i fixé à la fonction $t \mapsto T(x_i, t)$, il existe $\theta \in [t_j, t_{j+1}]$ tel que

$$T(x_i, t_{j+1}) = T(x_i, t_j) + dt \frac{\partial T}{\partial t}(x_i, t_j) + \frac{dt^2}{2} \frac{\partial^2 T}{\partial t^2}(x_i, \theta),$$

Soit en réarrangeant les termes et en divisant par dt ,

$$\frac{T(x_i, t_{j+1}) - T(x_i, t_j)}{dt} - \frac{\partial T}{\partial t}(x_i, t_j) = \frac{dt}{2} \frac{\partial^2 T}{\partial t^2}(x_i, \theta).$$

- Dérivée en espace : T est de classe $\mathcal{C}^4$ en temps, donc d'après le théorème de Taylor-Lagrange appliqué pour t_j fixé à la fonction $x \mapsto T(x, t_j)$, il existe $\eta_1 \in [x_i, x_{i+1}]$ tel que

$$\begin{aligned} T(x_{i+1}, t_j) &= T(x_i, t_j) + dx \frac{\partial T}{\partial x}(x_i, t_j) + \frac{dx^2}{2} \frac{\partial^2 T}{\partial x^2}(x_i, t_j) \\ &\quad + \frac{dx^3}{6} \frac{\partial^3 T}{\partial x^3}(x_i, t_j) + \frac{dx^4}{24} \frac{\partial^4 T}{\partial x^4}(\eta_1, t_j), \end{aligned} \quad (10)$$

En appliquant une deuxième fois Taylor Lagrange à la même fonction, on trouve qu'il existe $\eta_2 \in [x_{j-1}, x_j]$ tel que

$$T(x_{i-1}, t_j) = T(x_i, t_j) - dx \frac{\partial T}{\partial x}(x_i, t_j) + \frac{dx^2}{2} \frac{\partial^2 T}{\partial x^2}(x_i, t_j) - \frac{dx^3}{6} \frac{\partial^3 T}{\partial x^3}(x_i, t_j) + \frac{dx^4}{24} \frac{\partial^4 T}{\partial x^4}(\eta_2, t_j), \quad (11)$$

Ce qui donne en combinant (10) et (11) et en divisant par dx^2 :

$$\begin{aligned} \frac{T(x_{i+1}, t_j) - 2T(x_i, t_j) + T(x_{i-1}, t_j)}{dx^2} &= \frac{\partial^2 T}{\partial x^2}(x_i, t_j) \\ &= \frac{dx^2}{24} \left(\frac{\partial^4 T}{\partial x^4}(\eta_1, t_j) + \frac{\partial^4 T}{\partial x^4}(\eta_2, t_j) \right). \end{aligned} \quad (12)$$

• On remarque enfin que comme T est solution exacte, et que

$$\begin{aligned} \frac{T(x_i, t_{j+1}) - T(x_i, t_j)}{dt} - \frac{\partial T}{\partial t}(x_i, t_j) \\ - \kappa \left(\frac{T(x_{i+1}, t_j) - 2T(x_i, t_j) + T(x_{i-1}, t_j)}{dx^2} - \frac{\partial^2 T}{\partial x^2}(x_i, t_j) \right) = \varepsilon_i^j(T), \end{aligned} \quad (13)$$

on a

$$\varepsilon_i^j(T) = \frac{dt}{2} \frac{\partial^2 T}{\partial t^2}(x_i, \theta) - \kappa \frac{dx^2}{24} \left(\frac{\partial^4 T}{\partial x^4}(x_i, \eta_1) + \frac{\partial^4 T}{\partial x^4}(x_i, \eta_2) \right)$$

On obtient le résultat final en prenant

$$B = \max \left(\sup_{(x,t) \in [0,1] \times [0,t_{max}]} \left(\frac{1}{2} \left| \frac{\partial^2 T}{\partial t^2}(x, t) \right| \right), \sup_{(x,t) \in [0,1] \times [0,t_{max}]} \left(\kappa \frac{2}{24} \left| \frac{\partial^4 T}{\partial x^4}(x, t) \right| \right) \right),$$

où les suprema sont finis parce que T est de classe $\mathcal{C}^2$ en temps et $\mathcal{C}^4$ en espace sur des intervalles fermés et bornés. ■

Nous avons montré que la discrétisation approche bien l'équation, ce que l'on appelle être consistant²⁵ pour un schéma numérique. Le schéma d'Euler explicite est consistant d'ordre un en temps et d'ordre deux en espace.

Cependant, nous n'avons pas encore montré que la solution approchée converge vers la solution exacte. Nous n'avons fait qu'étudier ce qui se passe lorsque l'on injecte la solution exacte dans l'équation approchée, ce qui revient un peu à voir ce qui se passe en une étape. Ce qui nous intéresse vraiment, c'est la quantité

$$e_i^j(T) = T_i^j - T(x_i, t_j),$$

où T_i^j est le terme de la solution approchée produit par l'algorithme et $T(x_i, t_j)$ la solution exacte appliquée en (x_i, t_j) .

Notons comme précédemment :

$$e^j(T) = \left(e_i^j(T) \right)_{i \in \{0, \dots, n+1\}}.$$

25. Le terme consistence vient d'une erreur de traduction de l'anglais *consistency* qui est passée dans le langage des mathématiciens français. Un terme plus approprié serait cohérence.

Théorème 179

Supposons que $\frac{\kappa dt}{dx^2} \leq \frac{1}{2}$ et que la solution exacte T du problème est de classe $\mathcal{C}^2$ en temps et $\mathcal{C}^4$ en espace sur $[0, 1] \times [0, t_{max}]$. Alors il existe $A \in \mathbb{R}_+^*$ tel que

$$\forall j \in \{0, \dots, m+1\}, \quad \|e^j(T)\|_\infty \leq A(dt + dx^2).$$

Démonstration : Remarquons d'abord que les erreurs sont nulles sur les bords gauche, droits et bas :

$$e_0^j(T) = e_{n+1}^j(T) = e_i^0(T) = 0$$

Ensuite, pour $i \in \{1, n\}$ et $j \in \{0, m\}$, on a

$$\begin{aligned} e_i^{j+1}(T) &= T_i^{j+1} - T(x_i, t_{j+1}) \\ &= T_i^j + \frac{\kappa dt}{dx^2} (T_{i+1}^j - 2T_i^j + T_{i-1}^j) - T(x_i, t_{j+1}) \\ &= T_i^j + \frac{\kappa dt}{dx^2} (T_{i+1}^j - 2T_i^j + T_{i-1}^j) - T(x_i, t_j) + T(x_i, t_j) - T(x_i, t_{j+1}) \\ &= T_i^j + \frac{\kappa dt}{dx^2} (T_{i+1}^j - 2T_i^j + T_{i-1}^j) - T(x_i, t_j) - \left(T(x_i, t_{j+1}) - T(x_i, t_j) \right) \\ &= e_i^j(T) + \frac{\kappa dt}{dx^2} (T_{i+1}^j - 2T_i^j + T_{i-1}^j) - \left(T(x_i, t_{j+1}) - T(x_i, t_j) \right). \end{aligned}$$

Or, par définition de l'erreur de consistance $\varepsilon_i^j(T)$, on a

$$T(x_i, t_{j+1}) - T(x_i, t_j) = \kappa \frac{dt}{dx^2} \left(T(x_{i+1}, t_j) - 2T(x_i, t_j) + T(x_{i-1}, t_j) \right) + dt \varepsilon_i^j(T),$$

ce qui donne en reprenant le calcul précédent et en regroupant les termes pour faire apparaître les erreurs (petit effort mental!),

$$e_i^{j+1}(T) = e_i^j(T) + \frac{\kappa dt}{dx^2} \left(e_{i+1}^j(T) - 2e_i^j(T) + e_{i-1}^j(T) \right) - dt \varepsilon_i^j(T),$$

ou encore, écrit différemment,

$$e_i^{j+1}(T) = \left(1 - 2\frac{\kappa dt}{dx^2} \right) e_i^j(T) + \frac{\kappa dt}{dx^2} \left(e_{i+1}^j(T) + e_{i-1}^j(T) \right) - dt \varepsilon_i^j(T).$$

Par inégalité triangulaire et homogénéité de la valeur absolue,

$$|e_i^{j+1}(T)| \leq \left| \left(1 - 2\frac{\kappa dt}{dx^2} \right) e_i^j(T) \right| + \frac{\kappa dt}{dx^2} \left(|e_{i+1}^j(T)| + |e_{i-1}^j(T)| \right) + dt |\varepsilon_i^j(T)|.$$

Or, nous avons supposé $\frac{\kappa dt}{dx^2} \leq \frac{1}{2}$, c'est-à-dire $1 - 2\frac{\kappa dt}{dx^2} \geq 0$, d'où :

$$|e_i^{j+1}(T)| \leq \left(1 - 2\frac{\kappa dt}{dx^2} \right) |e_i^j(T)| + \frac{\kappa dt}{dx^2} \left(|e_{i+1}^j(T)| + |e_{i-1}^j(T)| \right) + dt |\varepsilon_i^j(T)|.$$

En prenant le maximum sur i de chaque côté de l'inégalité, et par linéarité du maximum sur le membre de droite, on obtient

$$\|e^{j+1}(T)\|_\infty \leq \left(1 - 2\frac{\kappa dt}{dx^2} \right) \|e^j(T)\|_\infty + \frac{\kappa dt}{dx^2} \left(\|e^j(T)\|_\infty + \|e^j(T)\|_\infty \right) + dt \|\varepsilon^j(T)\|_\infty,$$

ce qui donne en simplifiant,

$$\|e^{j+1}(T)\|_{\infty} \leq \|e^j(T)\|_{\infty} + dt \|e^j(T)\|_{\infty}.$$

Enfin, en appliquant le théorème précédent, on peut dire qu'il existe $B \in \mathbb{R}_+^*$ tel que

$$\|e^{j+1}(T)\|_{\infty} \leq \|e^j(T)\|_{\infty} + dtB(dt + dx^2),$$

ce qui donne par récurrence :

$$\|e^{j+1}(T)\|_{\infty} \leq \|e^0(T)\|_{\infty} + dt(j+1)B(dt + dx^2),$$

Remarquons enfin que nous avons déjà $\|e^0(T)\|_{\infty} = 0$ et qu'en outre

$$0 \leq (j+1)dt \leq (m+1)dt = t_{max},$$

ce qui donne le résultat sous la forme voulue :

$$\|e^{j+1}(T)\|_{\infty} \leq t_{max}B(dt + dx^2),$$

■

On remarquera que borner le temps entre 0 et t_{max} est d'une importance capitale pour obtenir la dernière majoration, mais que la majoration est valable pour tout t_{max} . Cela signifie que l'on a convergence de la solution approchée vers la solution exacte sur tout intervalle fermé et borné de temps, mais que plus l'intervalle de temps est grand moins le contrôle de la solution approchée est bon. On constate aussi que lorsque t_{max} est très grand, il faut prendre dt et dx très petits pour avoir une erreur faible de 0 à t_{max} .

L'étude théorique de la convergence de cet algorithme a donc révélé un important défaut : l'algorithme ne converge vers la solution exacte qu'à la condition que

$$\frac{\kappa dt}{dx^2} \leq \frac{1}{2},$$

que l'on appelle condition CFL (Courant-Friedrichs-Lewy).

Si l'on ne respecte pas cette condition, les simulations numériques font rapidement n'importe quoi. La condition semble tomber du ciel, mais elle paraît beaucoup plus naturelle si l'on regarde l'algorithme écrit sous la forme :

$$T_i^{j+1} = \left(1 - 2\frac{\kappa dt}{dx^2}\right)T_i^j + \frac{\kappa dt}{dx^2}(T_{i+1}^j + T_{i-1}^j).$$

Du point de vue de la programmation, il peut être pratique de poser $T^j = (T_i^j)_{i \in \{0, \dots, n+1\}}$, afin d'avoir le schéma d'Euler explicite sous la forme :

$$T^{j+1} = T^j + \frac{\kappa dt}{dx^2}AT^j,$$

avec A la matrice,

$$A = \begin{pmatrix} 2 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \\ 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & \dots & \dots & 0 & -1 & 2 \end{pmatrix}.$$

La matrice A est appelée matrice de raideur.

Pour éviter la condition contraignante de convergence, il est possible d'utiliser un autre schéma numérique, voisin, le schéma d'Euler implicite :

$$T_i^{j+1} = T_i^j + \frac{\kappa dt}{dx^2} (T_{i+1}^{j+1} - 2T_i^{j+1} + T_{i-1}^{j+1})$$

qui s'exprime sous forme matricielle ainsi,

$$T^{j+1} = T^j + \frac{\kappa dt}{dx^2} A T^{j+1},$$

ou encore,

$$(I_d - \frac{\kappa dt}{dx^2} A) T^{j+1} = T^j,$$

ce qui revient à dire que T^{j+1} est solution du système linéaire

$$(I_d - \frac{\kappa dt}{dx^2} A) X = Z,$$

où Z est connu.

On peut montrer théoriquement, avec les même outils que précédemment, que ce schéma implicite est toujours consistant et qu'il converge inconditionnellement vers la solution exacte quand les pas de temps et d'espace tendent vers 0. Le prix à payer est la formulation implicite qui nous oblige à résoudre un système linéaire $n \times n$ à chaque étape, ce qui peut coûter cher quand la dimension est grande et générer des erreurs d'arrondis importantes si le système est mal conditionné.

3. L'équation de Laplace

Soit Ω un ouvert de $\mathbb{R}^n$ de classe $\mathcal{C}^1$ dont nous notons le bord $\partial\Omega$. Soit f une fonction de classe $\mathcal{C}^1$. On s'intéresse aux fonctions régulières (ce terme est à préciser, bien entendu) u qui vérifient

$$\begin{cases} \Delta u &= 0 & \text{sur } \Omega \\ u &= f & \text{sur } \partial\Omega. \end{cases} \quad (14)$$

3.a. Solutions classiques

Pour donner facilement un sens à l'équation, nous allons chercher des solutions de classe $\mathcal{C}^2$ sur l'ouvert Ω et continues sur $\bar{\Omega} = \Omega \cup \partial\Omega$. Dans le cas de la dimension 1 et pour $\Omega =]0, 1[$, cela revient simplement à trouver une fonction u qui vérifie $u''(x) = 0$ pour $x \in]0, 1[$, et $u(0) = \alpha$, $u(1) = \beta$ où α et β sont deux nombres réels fixés à l'avance. Dans ce cas simple, il est facile de trouver une solution explicite :

$$u(x) = (\beta - \alpha)x + \alpha.$$

On remarque au passage qu'il n'existe qu'une seule fonction u de classe $\mathcal{C}^2$ sur l'ouvert $]0, 1[$ et continue sur $\bar{\Omega} = [0, 1]$ qui vérifie l'équation en dimension 1. C'est vrai en dimension supérieure si l'ouvert est borné et c'est l'occasion pour nous d'expérimenter une méthode classique de démonstration de l'unicité des solutions d'une EDP. On remarquera qu'il est important de bien préciser dans quel espace la solution est unique.

Théorème 180

Soit Ω un ouvert borné de $\mathbb{R}^n$. Il existe au plus une fonction u dans l'espace $\mathcal{C}^2(\Omega) \cap \mathcal{C}^0(\bar{\Omega})$ qui vérifie l'équation (14).

Démonstration : Supposons qu'il existe deux solutions $u_1, u_2 \in \mathcal{C}^2(\Omega) \cap \mathcal{C}^0(\bar{\Omega})$.

Soit $w = u_1 - u_2$. La fonction w est solution de (14) dans le cas particulier $f = 0$.

On a, d'après la formule de Green pour le Laplacien,

$$\int_{\Omega} w(x) \Delta w(x) \, dx = - \int_{\Omega} \|\nabla w(x)\|^2 \, dx + \int_{\partial\Omega} \frac{\partial w}{\partial \nu}(x) w(x) \, d\sigma(x),$$

Ce qui donne, puisque $\Delta w = 0$ sur Ω et $w = 0$ sur $\partial\Omega$,

$$\int_{\Omega} \|\nabla w(x)\|^2 \, dx = 0.$$

Le gradient de w est donc identiquement nul sur Ω . w est donc constante sur Ω et par continuité de w sur $\bar{\Omega}$ elle vaut partout sa valeur au bord, soit 0.

Donc $u_1 = u_2$. ■

Notons l'importance de la condition de bord. Elle détermine entièrement la solution. Notons aussi le fort lien qui existe entre l'équation de Laplace et l'analyse complexe. En effet, si on note $i : (x, y) \mapsto x + iy$ l'identification naturelle entre $\mathbb{R}^2$ et $\mathbb{C}$, si l'on note $Re : \mathbb{C} \rightarrow \mathbb{R}^2$ et $Im : \mathbb{C} \rightarrow \mathbb{R}^2$ les applications partie réelle et partie imaginaire du plan complexe, alors pour toute fonction holomorphe $f : \mathbb{C} \rightarrow \mathbb{C}$, les fonctions $g_{Re} = Re \circ f \circ i : \mathbb{R}^2 \rightarrow \mathbb{R}$ et $g_{Im} = Im \circ f \circ i : \mathbb{R}^2 \rightarrow \mathbb{R}$ sont de classe $\mathcal{C}^\infty$ en tous points et vérifient

$$\Delta g_{Re} = 0, \quad \Delta g_{Im} = 0, \quad \text{sur } \mathbb{R}^2.$$

Dans le cas particulier d'une boule de dimension 2 (un disque), on peut exprimer explicitement les solutions, et par conséquent avoir sans effort l'existence en plus de l'unicité.

Théorème 181

En dimension deux, soit $\Omega = B(0, R)$, $R \in \mathbb{R}_+^*$. On note $\partial\Omega = C(0, R)$. Soit $f : C(0, R) \rightarrow \mathbb{R}$ continue. Il existe une unique solution u au problème (14) dans l'espace $\mathcal{C}^2(\Omega) \cap \mathcal{C}^0(\bar{\Omega})$ et elle s'écrit

$$u(x) = \frac{R^2 - \|x\|^2}{2\pi R} \int_{C(0, R)} \frac{f(y)}{\|x - y\|^2} d\sigma(y).$$

Démonstration : Nous savons déjà que si la solution existe, elle est alors unique. Plaçons-nous en coordonnées polaires (r, θ) ; le laplacien s'écrit

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2}.$$

La donnée de bord continue f , vue comme une fonction de l'angle, peut se décomposer en série de Fourier :

$$\begin{aligned} \forall \theta \in [0, 2\pi[, \quad f(\theta) &= \frac{1}{2\pi} \int_0^{2\pi} f(\phi) d\phi \\ &+ \frac{1}{\pi} \sum_{n=0}^{+\infty} \left(\int_0^{2\pi} f(\phi) \cos(n\phi) d\phi \cos(n\theta) + \int_0^{2\pi} f(\phi) \sin(n\phi) d\phi \sin(n\theta) \right). \end{aligned} \quad (15)$$

En outre, pour tout, $y = (R, \phi) \in C(0, R)$ et $x = (r, \theta) \in \overline{B(0, R)}$, écrits en coordonnées polaires,

$$\begin{aligned} \frac{R^2 - \|x\|^2}{\|x - y\|^2} &= \frac{R^2 - r^2}{R^2 - 2rR \cos(\phi - \theta) + r^2} \\ &= 1 + 2 \frac{rR \cos(\phi - \theta) - r^2}{R^2 - 2rR \cos(\phi - \theta) + r^2} \\ &= 1 + 2Re \left(\sum_{n=0}^{+\infty} \left(\frac{r}{R} \right)^n e^{in(\phi - \theta)} \right) \\ &= 1 + 2 \sum_{n=0}^{+\infty} \left(\frac{r}{R} \right)^n \cos(n(\phi - \theta)) \end{aligned}$$

Notre expression de la solution u devient alors

$$\begin{aligned} u(r, \theta) &= \frac{1}{2\pi} \int_0^{2\pi} \left(1 + 2 \sum_{n=0}^{+\infty} \left(\frac{r}{R} \right)^n \cos(n(\phi - \theta)) \right) f(\phi) d\sigma(\phi) \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(\phi) d\phi + \frac{1}{\pi} \sum_{n=0}^{+\infty} \left(\frac{r}{R} \right)^n \int_0^{2\pi} f(\phi) \cos(n(\phi - \theta)) d\phi, \end{aligned}$$

et

$$\begin{aligned} \int_0^{2\pi} f(\phi) \cos(n(\phi - \theta)) d\phi &= \int_0^{2\pi} f(\phi) \cos(n\phi) d\phi \cos(n\theta) \\ &\quad + \int_0^{2\pi} f(\phi) \sin(n\phi) d\phi \sin(n\theta). \end{aligned}$$

La condition de bord est donc vérifiée d'après (15).

On peut ensuite vérifier que l'élément est solution dans le disque ouvert : dans la formulation de u que nous venons d'écrire, les intégrales sont sur des compacts et les séries convergent uniformément, nous pouvons donc intervertir dérivées et intégrales et dérivées et séries. Il suffit donc de vérifier que chaque terme de la forme

$$\left(\frac{r}{R}\right)^n \cos(n(\phi - \theta))$$

est solution, ce qui est le cas avec la formulation que nous avons donnée du Laplacien en coordonnées polaires (on dit que les solutions se superposent, ou encore qu'on utilise le principe de superposition). ■

La preuve que nous venons de faire triche dans le sens où nous posons la fonction qui marche et montrons qu'elle convient : nous partons de la fin et le lecteur n'a aucune idée de comment la solution a été réellement trouvée. En fait, il suffit de lire à l'envers et de remarquer que l'idée-clef est de séparer les variables en coordonnées polaires. Les solutions élémentaires, avant superposition, sont de la forme $\phi(r)\psi(\theta)$. Une construction plus naturelle et plus détaillée pourra être lue dans [3]. On trouvera dans le polycopié de Frédéric Paulin *Introduction à la théorie spectrale et à l'analyse harmonique* (https://www.math.u-psud.fr/paulin/notescours/cours_magistere2.pdf) une étude plus approfondie, plus théorique et beaucoup plus générale de ce résultat. En effet, le noyau de Poisson

$$(x, y) \mapsto P_x(y) = \frac{1 - \|x\|^2}{\|y - x\|^2},$$

sur le disque unité $\mathbb{D}$ et le cercle unité $\mathbb{S}_1$ est à mettre en lien avec l'ensemble $Aut(\mathbb{D})$ des automorphismes holomorphes du disque ouvert $\mathbb{D}$ qui sont tous de la forme

$$z \mapsto e^{i\theta} \frac{z - a}{1 - \bar{a}z},$$

où $a \in \mathbb{D}$ et $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. On peut appliquer P de manière (relativement) naturelle à toute mesure borélienne complexe μ sur le cercle unité $\mathbb{S}_1$ et l'application $P\mu$ qui en résulte est harmonique et est réelle si la mesure est réelle, positive si la mesure est positive. Cette considération qui peut paraître éloignée de l'étude des braves EDP qui servent à quelque chose sert ici à montrer que l'étude des solutions de l'équation de Laplace cache en fait des choses très profondes sur les fonctions harmoniques et holomorphes et que les deux sont intimement liées.

Notons aussi que toute solution u de (14) dans $\mathbb{R}^2$ sur un ouvert plus général qu'une boule est aussi une solution dans une petite boule contenue dans l'ouvert et donc admet la forme ci-dessus et vérifie toutes les propriétés que l'on peut en déduire dans cette petite boule. En particulier, on peut montrer avec cette formulation un résultat appelé principe du maximum et lui aussi très général et très importante en analyse harmonique et complexe.

Théorème 182

[†] En dimension deux, soit $\Omega = B(0, R)$. On note $\partial\Omega = C(0, R)$.

Soit $f : C(0, r) \rightarrow \mathbb{R}$ continue et u la solution du problème (14).
 Si pour $m, M \in \mathbb{R}$, $m \leq f \leq M$, alors pour tout $x \in B(0, R)$,
 $m \leq u(x) \leq M$.

Démonstration : La fonction

$$(x, y) \mapsto P_x(y) = \frac{1 - \|x\|^2}{\|y - x\|^2},$$

est positive. On a donc par intégration sur $C(0, R)$ de l'inégalité en hypothèse

$$\begin{aligned} \frac{R^2 - \|x\|^2}{2\pi R} \int_{C(0, r)} \frac{m}{\|x - y\|^2} d\sigma(y) \\ \leq \frac{R^2 - \|x\|^2}{2\pi R} \int_{C(0, r)} \frac{f(y)}{\|x - y\|^2} d\sigma(y) \\ \leq \frac{R^2 - \|x\|^2}{2\pi R} \int_{C(0, r)} \frac{M}{\|x - y\|^2} d\sigma(y). \end{aligned} \quad (16)$$

Or,

$$\begin{aligned} \frac{R^2 - \|x\|^2}{2\pi R} \int_{C(0, r)} \frac{m}{\|x - y\|^2} d\sigma(y) \\ = 1 + \frac{1}{\pi} \sum_{n=0}^{+\infty} \left(\frac{r}{R}\right)^n \int_0^{2\pi} \cos(n(\phi - \theta)) d\phi \\ = 1, \end{aligned} \quad (17)$$

Car pour tout $n \in \mathbb{N}$,

$$\int_0^{2\pi} \cos(n(\phi - \theta)) d\phi = 0.$$

On a le résultat en combinant (16), (17) et l'expression de u dans le théorème précédent. ■

3.b. Simulation numérique par les éléments finis

On va s'intéresser à un problème voisin du problème de Laplace. Soient Ω un ouvert borné de classe $\mathcal{C}^\infty$ de $\mathbb{R}^n$, f une fonction sur Ω et considérons le problème de Dirichlet

$$\begin{cases} -\Delta u + u = f & \text{sur } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \quad (18)$$

qui modélise, par exemple en dimension 2, une membrane tendue au dessus de l'ouvert Ω sur laquelle s'exerce une force décrite par f . C'est un problème stationnaire, on s'intéresse à la forme de la membrane à l'équilibre.

Nous allons décrire une méthode pour la recherche de solutions approchées calculables par un ordinateur à l'aide d'une méthode plus abstraite que les différences finies que nous avons déjà vues.

Remarquons pour commencer que si u est une solution de classe $\mathcal{C}^2$ sur Ω et continue sur $\overline{\Omega}$, pour toute fonction test $\phi \in \mathcal{C}_0^\infty(\Omega)$, on a, en intégrant l'équation sur Ω et en utilisant la formule de Green et la condition de bord,

$$\int_{\Omega} \left(\langle \nabla u(x) | \nabla \phi(x) \rangle + u(x)\phi(x) \right) dx = \int_{\Omega} f(x)\phi(x) dx.$$

Comme toute fonction v de $H_0^1(\Omega)$ est limite, par définition de cet espace, d'une suite $(\phi_n)_{n \in \mathbb{N}}$ de fonctions de $\mathcal{C}_0^\infty(\Omega)$, et comme on a pour tout $n \in \mathbb{N}$,

$$\int_{\Omega} \left(\langle \nabla u(x) | \nabla \phi_n(x) \rangle + u(x)\phi_n(x) \right) dx = \int_{\Omega} f(x)\phi_n(x) dx,$$

il est possible en faisant attention aux détails techniques (vérifier que tous les objets considérés sont dans $\mathbb{L}^2(\Omega)$, penser que convergence forte implique convergence faible, que $u \in H_0^1(\Omega), \dots$) de passer à la limite dans l'égalité précédente.

On appelle donc solution faible une fonction $u \in H_0^1(\Omega)$ telle que

$$\forall v \in H_0^1(\Omega), \quad \int_{\Omega} \left(\langle \nabla u(x) | \nabla v(x) \rangle + u(x)v(x) \right) dx = \int_{\Omega} f(x)v(x) dx.$$

Théorème 183

Si $f \in \mathbb{L}^2(\Omega)$, il existe une unique solution faible $u \in H_0^1(\Omega)$ au problème 18 et elle vérifie

$$\frac{1}{2} \int_{\Omega} \left(\|\nabla u\|^2 + u^2 \right) - \int_{\Omega} f u = \min_{v \in H_0^1(\Omega)} \left(\frac{1}{2} \int_{\Omega} \left(\|\nabla v\|^2 + v^2 \right) - \int_{\Omega} f v \right).$$

Démonstration : L'espace $H_0^1(\Omega)$ est un espace de Hilbert pour le produit scalaire issu de $H^1(\Omega)$ (il hérite de la structure hilbertienne ambiante car il est fermé). Soit la forme bilinéaire

$$a(u, v) = \int_{\Omega} \left(\langle \nabla u(x) | \nabla \phi_n(x) \rangle + u(x)\phi_n(x) \right) dx,$$

sur $H_0^1(\Omega)$ et l'application linéaire

$$l(v) = \int_{\Omega} f(x)\phi_n(x) dx,$$

sur $H_0^1(\Omega)$.

On a

$$\forall v \in H_0^1(\Omega), \quad |l(v)| \leq \|f\|_{\mathbb{L}^2} \|v\|_{\mathbb{L}^2} \leq \|f\|_{\mathbb{L}^2} \|v\|_{H_0^1},$$

donc l est continue.

La forme bilinéaire a est continue et coercive car c'est exactement le produit scalaire de $H_0^1(\Omega)$. Le théorème de Lax-Milgram donne donc le résultat : il existe un unique $u \in H_0^1(\Omega)$ tel que

$$\forall v \in H_0^1(\Omega), \quad a(u, v) = l(v),$$

et u minimise la fonctionnelle

$$\Phi : v \mapsto \frac{1}{2}a(v, v) - l(v). \quad \blacksquare$$

On peut montrer de nombreuses propriétés *a priori* pour les solutions faibles (dépendance de u en f , bornitude de u , régularité de $u, \dots$), comme le fait l'excellent polycopié de Frédéric Lagoutière *Équations aux dérivées partielles et leurs approximations*, disponible sur le lien <http://math.univ-lyon1.fr/lagoutiere/poly.pdf>, auquel nous avons déjà volé les démonstrations pour la méthode des différences finies (nous allons recommencer le plagiat pour les éléments finis).

Parlons-en, des éléments finis. Nous souhaitons calculer une solution faible approchée (en espérant que des développements théoriques puissent prouver qu'elle soit forte et régulière). Les ordinateurs ne savent pas travailler en dimension infinie et échouerons à trouver de manière analytique une fonction $u \in H_0^1(\Omega)$ qui vérifie

$$\forall v \in H_0^1(\Omega), \quad a(u, v) = l(v),$$

par contre ils réussiront sous de bonnes conditions, pour un sous-espace vectoriel $V_h \subset H_0^1(\Omega)$ de dimension finie, à calculer u_h tel que

$$\forall v_h \in V_h, \quad a(u_h, v_h) = l(v_h), \quad (19)$$

dans les limites du calcul intégral approché et de la représentation des fonctions par une (grande) liste de valeurs. La méthode que nous avons utilisée dans la preuve assure l'existence d'une solution u_h au problème 19.

L'idée intuitive des éléments finis est que si le sous-espace de dimension finie V_h indexé par $h \in \mathbb{R}_+^*$ "tend" vers $H_0^1(\Omega)$ lorsque h tend vers 0, la solution $u_h \in V_h$ devrait "tendre" vers l'unique solution faible u du problème posé sur $H_0^1(\Omega)$ tout entier.

On définit une notion de convergence de sous-espaces vectoriels pour pouvoir travailler mathématiquement.

Définition 184 (Approximation conforme). *Soit V un espace vectoriel normé de norme $\|\cdot\|_V$, on dit que $(V_h)_{h \in \mathbb{R}_+^*}$ est une approximation conforme de l'espace V si pour tout $h \in \mathbb{R}_+^*$, $V_h \subset V$ est de dimension finie et si pour tout $v \in V$, pour tout $\varepsilon \in \mathbb{R}_+^*$, il existe h_ε tel que pour tout $h \in]0, h_\varepsilon]$, il existe $v_h \in V_h$ vérifiant*

$$\|v_h - v\|_V \leq \varepsilon.$$

Nous pouvons montrer que cette définition est adaptée au problème que nous nous sommes posé.

Théorème 185 (*Lemme de C  a*)

Soit V_h un sous-espace vectoriel de dimension finie d'un espace de Hilbert V . Soit a une forme bilin  aire sur V , continue de module de continuit   C , coercive de module de coercivit   α , soit l une forme lin  aire continue sur V de module de continuit   D . Soit $u \in V$ l'unique solution de

$$\forall v \in H_0^1(\Omega), \quad a(u, v) = l(v),$$

et $u_h \in V_h$ l'unique solution de

$$\forall v_h \in V_h, \quad a(u_h, v_h) = l(v_h).$$

On a

$$\|u - u_h\|_V \leq \frac{C}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V$$

D  monstration : Soit $x_h, v_h \in V_h$.

Par d  finition de u et u_h , on a

$$a(u - u_h, v_h) = a(u, v_h) - a(u_h, v_h) = l(v_h) - l(v_h) = 0.$$

C'est en particulier vrai pour $x_h = u_h - v_h \in V_h$ (Stabilit   par somme d'un espace vectoriel), ce qui donne la proposition

$$\forall v_h \in V_h, \quad a(u - u_h, u_h - v_h) = 0.$$

Par cons  quent,

$$a(u - u_h, u - v_h) = a(u - u_h, u - u_h) - a(u - u_h, u_h - v_h) = a(u - u_h, u - u_h).$$

On a par continuit  

$$a(u - u_h, u - v_h) \leq C \|u - u_h\|_V \|u - v_h\|_V,$$

et par coercivit  

$$a(u - u_h, u - u_h) \geq \alpha \|u - u_h\|_V^2.$$

On d  duit de tout cela

$$\forall v_h \in V_h \quad \|u - u_h\|_V \leq \frac{C}{\alpha} \|u - v_h\|_V,$$

d'o   le r  sultat en passant    l'infimum. ■

Ce lemme est parfois appel   second lemme de Strang. Il dit que u est aussi proche de v_h que l'espace V_h le permet.

Intéressons nous à présent à la méthode de calcul. Soit n_h la dimension de V_h et $(e_h^1, \dots, e_h^{n_h})$ une base de V_h . La solution u_h du problème

$$\forall v_h \in V_h, \quad a(u_h, v_h) = l(v_h)$$

admet une décomposition dans cette base :

$$u_h = \sum_{i=1}^{n_h} \lambda_{i,h} e_h^i.$$

On peut donc écrire le problème sous la forme

$$\sum_{i=1}^{n_h} \sum_{j=1}^{n_h} \lambda_{i,h} \lambda_{j,h} a(e_h^i, e_h^j) = \sum_{i=1}^{n_h} \lambda_{i,h} l(e_h^i).$$

Cela revient donc à résoudre un système linéaire

$$A_h U_h = L_h,$$

où A_h est une matrice associée à a et V_h , L_h est un vecteur associé à l et V_h et U_h est le vecteur contenant les $\lambda_{i,h}$. Résoudre numériquement le problème revient à calculer le vecteur U_h pour h assez petit pour que u_h soit une bonne approximation de u , et étant donné la matrice A_h et le vecteur V_h , c'est du ressort de l'algèbre linéaire et du calcul numérique qui lui est associé.

Qu'on ne crie pas victoire trop vite. La borne du lemme de Céa ne garantie pas à elle seule la convergence de l'algorithme car la borne

$$\inf_{v_h \in V_h} \|u - v_h\|_V$$

dépend de l'espace V_h choisi.

De plus, il faut également choisir une base de V_h et calculer la matrice A_h et le vecteur L_h en un temps raisonnable. Pire encore, résoudre le système linéaire peut être très coûteux lorsque n_h est grand et stocker la matrice à un coût de complexité en espace. Il faut donc choisir une base de V_h qui rende la matrice A_h aussi creuse que possible (lui faire avoir des zéros sur le plus d'emplacements possibles).

On réalise en général un maillage de l'espace et on prend l'espace V_h engendré par une famille de polynômes par morceaux dont les supports ne sont compris que dans une seule maille.

Par exemple, en dimension 1 d'espace, le problème sur $\Omega =]0, 1[$ devient $-u''(x) + u(x) = f(x)$, $u(0) = u(1) = 0$. On peut alors discrétiser l'intervalle $]0, 1[$ en petits intervalles de même longueur au maximum h , et définir sur chaque

segment une fonction linéaire par morceau : pour $[a_i, b_i] \subset]0, 1[$ le $i^{\text{ème}}$ intervalle de la sorte, on défini e_h^i par la formule

$$\begin{aligned} e_h^i(x) &= 0 \quad \text{si } x \notin [a_i, b_i] \\ e_h^i(x) &= \frac{2(x - a_i)}{a_i + b_i} \quad \text{si } x \in [a_i, \frac{a_i + b_i}{2}] \\ e_h^i(x) &= \frac{2(b_i - x)}{a_i + b_i} \quad \text{si } x \in]\frac{a_i + b_i}{2}, b_i] \end{aligned} \quad (20)$$

C'est en gros un triangle de hauteur 1 au dessus de l'intervalle $[a_i, b_i]$.

Ces éléments finis sont une bonne première approche. On peut obtenir dans ce cas une borne du type

$$\|u - u_h\|_{H^1(]0,1])} \leq Kh,$$

pour une constante $K \in \mathbb{R}_+^*$.

Dès la dimension 2, le maillage devient plus compliqué qu'une simple subdivision d'intervalle (voir figure 7) et doit prendre en compte la géométrie du domaine.

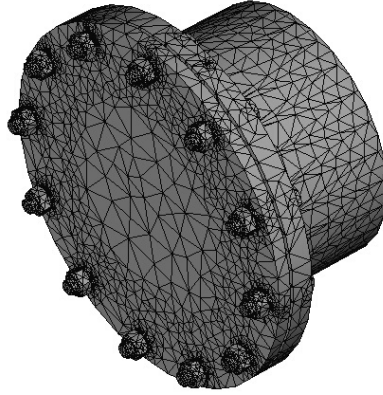


FIGURE 7 – Maillage pour l'étude mécanique des contraintes , ici sur une surface (extension naturelle mais difficile d'un maillage sur un ouvert de $\mathbb{R}^2$).

4. Modélisation plus avancée : la chimiotaxie

Nous allons ici présenter, sans faire de mathématiques dessus en raison de la difficulté des résultats les plus simples, un modèle aux dérivées partielles pour un phénomène biologique plus subtil afin de montrer ce que l'on peut trouver

comme modèle en dehors des sempiternelles équations linéaires.

La chimiotaxie²⁶, le phénomène par lequel des organismes dirigent leurs mouvements en fonction de certaines espèces chimiques présentes dans leur environnement, joue un rôle clef dans l'histoire de la vie et est un pilier du fonctionnement des écosystèmes et des organismes pluricellulaires de grande taille comme les mammifères. On trouve des comportements chimiotactiques dans toutes les branches du vivant, des bactéries les plus primitives aux cellules immunitaires humaines. Il peut s'agir d'attraction vers une substance nutritive ou d'évitement des substances nocives.

Quoique les premiers microscopes permissent déjà d'observer le mouvement des cellules, la description spécifique de la chimiotaxie a dû attendre la fin du XIXe siècle avec les travaux de Theodor Engelmann en 1881 et Wilhelm Pfeffer en 1884. Ce domaine de la biologie n'a été sérieusement exploré qu'au XXe siècle, avec notamment la contribution du prix Nobel Ilya Metchnikov pour ce qui est de la chimiotaxie eucaryote et ses liens avec le système immunitaire. Les connaissances en chimiotaxie ont ensuite bénéficié de l'explosion des biotechnologies et de la biochimie, pendant la seconde moitié du XXe siècle, avec au bout les approches de modélisation mathématique dont nous allons parler.

En raison de la grande diversité de l'usage que font les organismes de la chimiotaxie, la compréhension qualitative et quantitative des mouvements d'ensemble induits par la présence de substances chimiques dans un milieu a des retombées importantes aussi bien en biologie fondamentale qu'en écologie ou en médecine. En effet, il est très important de comprendre comment le système immunitaire humain localise et poursuit les pathogènes, ou comment les bactéries hospitalières fuient le phénol ou toute autre substance bactéricide.

La majeure partie des procaryotes présentant des comportements chimiotactiques utilisent à cette "fin" des flagelles enchâssés dans leur paroi cellulaire. On trouve une grande diversité de flagelles et de nombres de flagelles par organisme à travers le monde vivant. En particulier, et contrairement à ce qui a longtemps été cru, les bactéries et les archées ont des flagelles non-homologues. Les flagelles des procaryotes sont structurellement plus "simples" que ceux des eucaryotes et sont en général activés par une force proton motrice.

Le mécanisme de chimiotaxie le plus courant chez les procaryotes est dû à l'alternance entre des phases de natation rectiligne à l'aide du ou des flagelles et des phases de mouvements aléatoires dit de *tumble*. En particulier, chez *Escherichia Coli*, que nous abrègerons souvent en E. Coli, la paroi cellulaire comporte quatre à six flagelles qui en phase de tumble se désolidarisent les uns des autres, causant un mouvement aléatoire. Lorsque la bactérie va dans la bonne direction (gradient d'une substance favorable ou défavorable) elle change le sens de ro-

26. Du grec *taxis* qui signifie arrangement. La taxie est un mouvement d'un organisme provoqué par une stimulation extérieure.

tation de ses flagelles pour les arranger dans un faisceau synchrone et avancer dans la direction choisie.

D'autres méthodes existent pour des procaryotes ne possédant qu'un flagelle ou ayant une forme globale hélicoïdale.

Les procaryotes n'ayant pas à disposition un système nerveux et étant trop petits pour sentir une différence de concentration entre deux points de leur paroi cellulaire, ils ont besoin de capteurs pour évaluer la concentration des espèces chimiques favorables ou défavorables et d'une capacité à intégrer dans le temps les informations qu'ils trouvent.

L'évaluation des concentrations se fait grâce à des récepteurs, pour la plupart transmembranaires. Chaque molécule a ses propres récepteurs, et pour la même molécule, le dioxygène par exemple, on trouve des récepteurs transmembranaires qui fixent la molécule, d'autres sensibles à la force proton-motrice, d'autres inter-membranaires sensibles à l'état interne de la cellule,... Une fois les récepteurs activés, une cascade de réactions conduit à la modification du comportement du ou des flagelles. Le système s'adapte continuellement à la concentration extérieure : des groupes méthyles sont par exemple ajoutés au récepteur au fur et à mesure de son activation, ce qui le rend de plus en plus sensible. Le signal du récepteur entraîne sa propre déméthylation, ce qui constitue une boucle de rétroaction, de sorte que le système réagira toujours à un changement de la concentration du chimioattractant, parce que la sensibilité du récepteur a été ajustée au milieu que vient de quitter la cellule par ses mouvements aléatoires. Ce phénomène est du au jeu d'enzymes antagonistes qui méthylent et déméthylent la partie inter-membranaire du récepteur.

Les organismes eucaryotes étant en règle générale plus gros et à l'organisation plus complexe que leurs homologues procaryotes, ils n'ont pas à recourir à la méthode de marche aléatoire pour percevoir les différences de concentrations des substances favorables ou nocives ; leur taille leur permet de sentir un gradient entre deux points de leur paroi cellulaire. L'intégration des signaux issus de gradients chimiques en des voies métaboliques permettant la locomotion est encore mal comprise.

Les eucaryotes unicellulaires possèdent parfois des cils semblables aux flagelles des procaryotes mais présents en bien plus grand nombre et s'agitant de manière synchrone. On trouve aussi des cellules eucaryotes présentant un unique flagelle complexe. D'autres se déforment en créant des sortes de bras internes formés de filaments d'actine : la cellule n'a *a priori* pas d'avant ni d'arrière et la présence d'une plus grande quantité de chimioattractant d'un côté de sa paroi la fait se déformer de façon à avancer dans la direction favorable. L'excroissance que la cellule projette en avant, qui est une manifestation de la réorganisation du cytosquelette, est appelé pseudopode.

L'amibe²⁷ *Dictyostelium discoideum* du groupe Amoebozoa, vit sur les tapis de feuilles mortes et se nourrit d'organismes plus petits par phagocytose (bactéries, levures). Cette amibe, souvent appelée amibe sociale, est un modèle pour les mathématiciens qui s'intéressent à la chimiotaxie de par sa culture facile et les données de qualité qui ont été obtenue sur ses comportements chimiotactiques.

La chimiotaxie se fait chez *Dictyostelium discoideum* par la création de pseudopodes. Lorsque la concentration de chimioattractant est plus forte d'un certain côté de la cellule, il se forme dans le cytosol²⁸ des filaments d'actine qui poussent et déforment la membrane dans la direction du chimioattractant. La rétraction de ce qui est momentanément devenu l'arrière de la cellule est commandée par la même protéine motrice – la myosine – qui joue un rôle dans les contractions musculaires humaines.

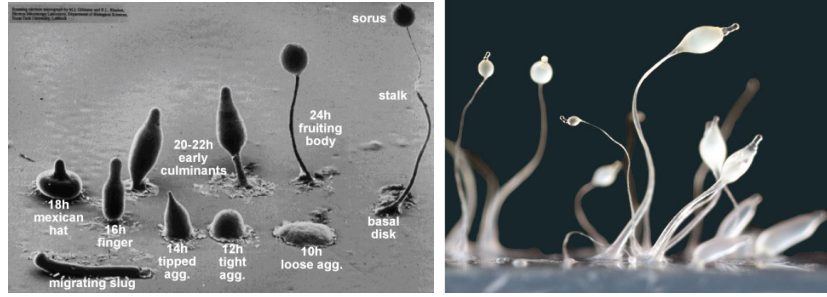
Dispersée dans un milieu riche en nourriture, les *Dictyostelium discoideum* se dirigent vers les zones à forte concentration en proies. Une fois les réserves de nourritures taries, les amibes se dispersent afin de couvrir toute l'étendue du territoire, puis elle commencent à être attirées par un chimioattractant qu'elles sécrètent elles mêmes – l'adénosine monophosphate cyclique, abrégé AMPc. Elles se regroupent alors en amas de plus en plus compacts. Ces amas étant une source d'AMPc, ils attirent toujours plus d'amibe jusqu'à ce que se produise un bourgeonnement : elles forment une tige, appelée sporocarpe, qui se termine en une boule d'amibes qui se différencient en spores et sont disséminées par le vent vers des contrées plus favorables. La formation du sporocarpe peut être précédée d'une phase où la colonie compacte d'amibes se déplace comme une limace vers des zones où la nourriture est plus abondante. On peut voir les différentes étapes de la formation du sporocarpe sur la figure 8a et une (jolie) photo de ce que cela peut donner une fois achevé sur la figure 8b.

Le système de détection des gradients de concentration chez *Dictyostelium discoideum* se base, comme chez les bactéries, sur des récepteurs qui se lient de manière réversible aux éléments chimiques recherchés ou évités, mais les récepteurs sont très nombreux (plusieurs dizaines de milliers) et c'est la répartition de l'état de ces récepteurs (liés à un composant chimique ou non) qui entre en jeu dans la détermination du sens de déplacement des cellules. Le mécanisme de la chimiotaxie de cette amibe est caractéristique de beaucoup d'autres espèces d'amibes et s'apparente, dans le comportement de groupe, à l'agrégation de certaines bactéries comme les souches chimiotactiques d'*Escherichia Coli*. En particulier, les mêmes modèles mathématiques décrivent le comportement des colonies.

Nous n'abordons pas les aspects probabilistes liés à l'étude du tumbling

27. Amibe est un nom vernaculaire qui englobe des groupes de protozoaires très différents les uns des autres et l'ensemble des espèces réunies sous ce nom ne forme pas un groupe monophylétique. On retrouve cependant à travers les amibes une similitude de formes et de comportements, et de nombreuses amibes sont concernées par ce que nous allons décrire.

28. La phase liquide dans laquelle baignent les organites de la cellule.



(a) Différents stades de la formation d'un sporocarpe.

(b) Ineffable beauté des comportements sociaux protozoaires.

FIGURE 8 – Agrégations de *Dictyostelium discoideum*

comme marche aléatoire chez les bactéries ou aux processus stochastiques associés à la création de pseudopodes chez les eucaryotes. Nous utilisons plutôt une vision macroscopique qui ne demande plus de distinction entre les organismes et leurs diverses techniques de déplacement et de perception du gradient chimique d'intérêt. Les équations aux dérivées partielles sont donc tout à fait adaptées. Plaçons-nous dans un environnement bidimensionnel simple, c'est-à-dire un ouvert Ω de $\mathbb{R}^2$, connexe, et régulier (de classe $\mathcal{C}^\infty$).

Le modèle que nous présentons est (à peu de choses près) celui développé par Evelyn Keller²⁹ et Lee Segel³⁰ dans l'article [6] qui tâchait de modéliser notre amibe favorite : *Dictyostelium discoideum*.

Notons $u(x, t)$ la concentration au temps t et à la position x d'organismes actifs et $v(x, t)$ la concentration de chimioattractant. Soit un volume fermé V quelconque délimité par une surface S dont on note n la normale extérieure sortante. En appliquant la conservation de la masse aux organismes dans le système délimité par V , on a

$$\partial_t \int_V u(x, t) \, dx = - \int_S \Phi(s, t) \cdot n(s, t) \, ds + \int_V f(u(x, t)) \, dx,$$

où Φ est le flux d'organismes et f le terme de création/mort des organismes. Par la formule de Stokes, et en passant la dérivée sous l'intégrale, on obtient,

$$\int_V \left(\partial_t u(x, t) + \nabla \cdot \Phi(x, t) - f(x, t) \right) \, dx = 0,$$

29. Evelyn Fox Keller (1936 -), également réputée pour être une militante et écrivaine féminisme, a contribué à la biophysique et à l'histoire des sciences et de la philosophie.

30. Lee Aaron Segel (1932 - 2005) était un mathématicien connu pour ses travaux sur la convection non-linéaire, la chimiotaxie et la modélisation d'agrégations spontanées d'organismes amiboïdes.

ce qui donne, puisque V est quelconque,

$$\partial_t u(x, t) = -\nabla \cdot \Phi(x, t) + f(u(x, t)).$$

Or, il y a dans le flux Φ un terme de diffusion et un terme dû au chimioattractant. La loi de Fick permet d'exprimer la diffusion comme

$$\Phi_{\text{diff}} = -d_u \nabla u,$$

avec $d_u > 0$ un coefficient de diffusion, et on peut exprimer le flux de chimioattractant comme le produit du gradient de v (un fort gradient de chimioattractant implique un fort déplacement), de la concentration d'organisme (plus d'organismes implique plus d'organismes qui bougent) et d'une constante χ qui traduit l'efficacité du chimioattractant, c'est-à-dire

$$\Phi_{\text{chemo}} = \chi u \nabla v,$$

ce qui conduit à l'équation

$$\partial_t u = \nabla \cdot (d_u \nabla u - \chi u \nabla v) + f(u),$$

En appliquant la même méthode pour le chimioattractant, avec $g(v(x, t))$ le terme de disparition du chimioattractant et une création de ce dernier par les organismes au taux α , on obtient

$$\partial_t v = d_v \nabla \cdot \nabla v - g(v) + \alpha u.$$

Supposons également, avant de réunir les deux équations en un système interdépendant, que la réaction qui concerne le chimioattractant se fait beaucoup plus vite que le déplacement des organismes, ce qui correspond au système

$$\begin{cases} \partial_t u &= \nabla \cdot (d_u \nabla u - \chi u \nabla v) + f(u) \\ \varepsilon \partial_t v &= d_v \Delta v - g(v) + \alpha u \end{cases}$$

Nous allons simplifier le système avant de l'étudier mathématiquement en posant $d_u = d_v = 1$ et $g(v) = v$. On considère aussi, et c'est une hypothèse réaliste dans une certaine mesure, que l'échelle de temps du déplacement des organismes est faible devant l'échelle de temps de leur naissances et morts, ce qui revient à prendre $f = 0$. Enfin, nous supposons également que les réactions qui concernent le chimioattractant sont quasi-instantanées, ce qui revient à dire que $\varepsilon = 0$.

Pour Ω un ouvert borné de $\mathbb{R}^2$, de bord lisse, on considère donc le système d'équations aux dérivées partielles

$$\begin{cases} \partial_t u &= \nabla \cdot (\nabla u - \chi u \nabla v) & \text{dans } \Omega, t > 0 \\ 0 &= \Delta v - v + \alpha u & \text{dans } \Omega, t > 0 \\ \frac{\partial u}{\partial n} &= \frac{\partial v}{\partial n} = 0 & \text{sur } \partial\Omega, t > 0 \\ u(x, 0) &= u_0(x) & \text{dans } \Omega \end{cases} \quad (\clubsuit)$$

où $\alpha, \chi \in \mathbb{R}_+^*$, et $u_0 \in \mathbb{L}^1(\Omega) \cap \mathbb{L}^\infty(\Omega)$.

La condition de Neumann homogène au bord signifie que rien n'entre ou ne sort du domaine Ω et la condition initiale u_0 est naturellement supposée positive et non uniformément nulle, c'est-à-dire non nulle sur un ensemble de mesure strictement positive puisqu'elle est définie à un ensemble de mesure nulle près.

Ce système parabolique-elliptique d'apparence simple a été intensivement étudié ces dernières décennies et son homologue parabolique-parabolique, plus complexe, résiste encore sur bien des points.

Le système (♣) a été beaucoup étudié et certaines de ses propriétés sont connues. En particulier, on peut montrer que si la condition initiale u_0 est positive, non uniformément nulle et régulière sur le compact $\bar{\Omega}$, alors il existe $T \in \mathbb{R}_+^*$ et une unique solution (u, v) sur $\Omega \times]0, T]$ strictement positive. Mais pour une trop grande quantité d'organismes, il peut y avoir explosion en temps fini de la solution au sens où l'unique solution locale ne peut pas être prolongée sur $\mathbb{R}$ tout entier. Plus précisément, on a le résultat suivant, dû à Toshitaka Nagai.

Théorème 186

Supposons que

$$\int_{\Omega} u_0(x) dx > \frac{8\pi}{\alpha\chi}.$$

Pour tout $q \in \Omega$, il existe $\varepsilon(q) \in \mathbb{R}_+^*$ tel que si

$$\int_{\Omega} u_0(x) \|x - q\|^2 dx < \varepsilon(q),$$

alors la solution locale ne peut pas être globale.

Contrairement aux équations différentielles ordinaires pour lesquelles l'explosion en temps fini est simple à caractériser, les équations aux dérivées partielles présentent des comportements très divers. Dans le cas de l'équation de Keller-Segel, on peut observer une convergence de la solution en temps fini et au sens des mesures vers une masse de Dirac δ . Le résultat suivant, obtenu dans le cas des solutions à symétrie radiale par Miguel Herrero et Juan Velázquez pour le système presque identique

$$\left\{ \begin{array}{ll} \partial_t u &= \nabla \cdot (\nabla u - \chi u \nabla v) & \text{dans } \Omega, t > 0 \\ 0 &= \Delta v + (u - 1) & \text{dans } \Omega, t > 0 \\ \frac{\partial u}{\partial n} &= \frac{\partial v}{\partial n} = 0 & \text{sur } \partial\Omega, t > 0 \\ u(x, 0) &= u_0(x) & \text{dans } \Omega \end{array} \right. \quad (HV)$$

donne un aperçu du cauchemar que peut être l'étude de l'explosion en temps fini d'une EDP (à titre indicatif, la démonstration fait une trentaine de pages dans l'article qui l'introduit).

Théorème 187

Il existe une solution radiale (u, v) de (HV) telle que

$$u(r, t) \rightarrow \frac{8\pi}{\chi} \delta + f(r),$$

au sens des mesures quand t tend vers T , avec

$$f(r) = \frac{C}{r^2} e^{-2|\log(T-t)|^{1/2}} (2|\log(r)|)^{\frac{1}{2\sqrt{2}|\log(r)|^{1/2}-\frac{1}{2}}} (1 + o(1)),$$

où C est une constante strictement positive.

5. Synthèse et pratique**5.a. Synthèse de la section**

Nous avons vu plusieurs types d'équations aux dérivées partielles aux comportements très différents et exploré (un peu) les méthodes appropriées à leur résolution. Il faut être précautionneux lorsque nous parlons de condition initiale, de solution, d'unicité. Les espaces de solutions des équations aux dérivées partielles sont beaucoup plus vastes que ceux des équations différentielles ordinaires et les conditions initiales ne sont pas des points mais des fonctions.

Nous avons d'abord travaillé sur les équations de transport en remontant le long des courbes caractéristiques sur lesquelles les solutions sont constantes. Les solutions sont de même régularité que la condition initiale en temps.

Puis nous avons travaillé sur l'équation de la chaleur et l'avons résolue par convolution avec le noyau de la chaleur. Ses solutions sont très régulières même pour des données initiales très irrégulières. La méthode des différences finies permet de résoudre numériquement l'équation de la chaleur et est simple à mettre en œuvre. Le schéma explicite est par contre coûteux en temps de calcul en raison d'une condition de stabilité numérique. Le schéma implicite contourne cette condition mais demande la résolution d'un système linéaire, ce qui est également coûteux.

L'équation de Laplace n'est pas un problème d'évolution, elle nécessite seulement une condition de bord. La solution est unique pour une donnée de bord continue fixée. Si le domaine est un disque du plan, on a une formulation explicite qui permet de déduire des choses sur la solution. Les solutions de l'équation de Laplace dans le plan sont dites harmoniques. La méthode des éléments finis est toute désignée pour résoudre l'équation de Laplace : il s'agit d'écrire le problème sous forme variationnelle (forme faible) puis de trouver une solution d'un problème approché en dimension finie *via* le théorème de Lax-Milgram.

Nous avons enfin présenté un modèle pour la chimiotaxie.

5.b. Exercices théoriques et de modélisation

Exercice 188 (Ondes progressives pour le modèle KPP). *Nous allons étudier, dans le cas de la dimension 1 d'espace et pour des paramètres égaux à 1, le modèle de Fisher-KPP dont nous avons parlé plus haut, soit*

$$\frac{\partial u}{\partial t}(x, t) = \frac{\partial^2 u}{\partial x^2}(x, t) + u(x, t)(1 - u(x, t)), \quad u(x, 0) = u_0, \quad x \in \mathbb{R}, \quad t \in \mathbb{R},$$

et notre objectif est de mettre en lumière le phénomène d'invasion du milieu par les animaux en explicitant une onde progressive.

1. Soit $c \in \mathbb{R}_+^*$ et soit une solution u de classe $\mathcal{C}^2$ sur $\mathbb{R}^2$ vérifiant $u(x, t) = u_0(x - ct)$ (on appelle ondes progressives de telles solutions). Montrer que si l'on pose $y = x - ct$ et $v(y) = u(x, t)$ alors v vérifie une équation différentielle de la forme

$$av''(y) + bv'(y) + f(v(y)) = 0,$$

où $a, b \in \mathbb{R}$ et $f : \mathbb{R} \rightarrow \mathbb{R}$ est continue.

2. On choisit $c = \frac{5}{\sqrt{6}}$. Montrer que la fonction

$$v : y \mapsto \frac{1}{1 + C e^{\frac{y}{\sqrt{6}}}}$$

est solution de l'équation différentielle obtenue.

3. Quelles sont les limites de v en $+\infty$ et $-\infty$. Qu'est-ce que cela signifie du point de vue de la modélisation ? Que représentent ces ondes progressives en écologie ?

A. COMPLÉMENTS EN ANALYSE FONCTIONNELLE ET EN GÉOMÉTRIE

« Cette fois c'est pas une bagarre entre deux piliers de comptoir ou un connard qui veut crâmer la gueule de quelqu'un avec sa bombe de gaz. C'est du lourd ma p'tite caille. »

Sin City, réalisé par Frank Miller et Robert Rodriguez, avec la participation de Quentin Tarantino (2005)

Nous présentons dans cette annexe un peu bric-à-brac des résultats (classiques et relativement simples) de topologie, d'analyse fonctionnelle et de géométrie utiles à la compréhension des autres annexes. La présentation est rapide car il était impossible de faire un cours complet là-dessus sans tripler la taille du photocopié et sacrifier la thèse de l'auteur. Le contenu de ces rappels est ~~sauvagement plagié~~ inspiré des livres [8], [12], [7], du cours de topologie algébrique donné à l'université Paris-Sud par Antoine Chambert-Loir et des cours spécifiques de magistère à l'université Paris-sud d'Orsay donnés par Frédéric Paulin et Frédéric Bourgeois.

1. Notions de topologie générale

La topologie générale est un cadre plus vaste que celui des espaces métriques. Son approche axiomatique est quasiment minimale pour l'étude des objets de l'analyse et de la géométrie et elle permet d'exprimer la plupart des notions importantes (continuité, limite, voisinage,...) avec très peu de contraintes sur la structure mise sur l'espace. Nous rappelons ici uniquement ce dont nous aurons besoin pour parler sans ambiguïté des variétés différentielles et des espaces de fonctions tests, c'est-à-dire assez peu de chose en somme. Il est intéressant de remarquer que du fait de son caractère peu intuitif, cette théorie n'est apparue que très tard dans l'histoire des mathématiques, au début du vingtième siècle avec notamment l'importante contribution de Felix Hausdorff.

Espaces topologiques :

Définition 189 (Topologie). *Soit E un ensemble. On appelle topologie sur E un sous-ensemble $\mathcal{T}$ de l'ensemble $\mathcal{P}(E)$ des parties de E vérifiant*

- $\forall U, V \in \mathcal{T}, U \cup V \in \mathcal{T}$;

- Pour toute famille $(U_i)_{i \in I}$ indexée par un ensemble quelconque,

$$\bigcup_{i \in I} U_i \in \mathcal{T} ;$$

- $E \in \mathcal{T}$ et $\emptyset \in \mathcal{T}$.

On donne à présent le vocabulaire qui permet d'apposer une intuition sur la notion de topologie.

Définition 190 (Espace topologique, ouvert, fermé, topologie plus finie). *Un ensemble muni d'une topologie est appelé un espace topologique. Les éléments d'une topologie sont appelés les ouverts. Le complémentaire dans l'espace topologique d'un ouvert est dit fermé pour la topologie. Une topologie $\mathcal{T}$ est dite plus fine qu'une topologie $\mathcal{S}$ si tous les ouverts de $\mathcal{S}$ sont dans $\mathcal{T}$.*

Remarque 191. *Tout espace métrique (E, d) est muni d'une topologie naturelle avec le cas particulier des ouverts définis dans notre section sur les espaces métriques. On vérifie facilement avec les propriétés que nous avons montrées que l'ensemble des ouverts au sens d'ouvert pour une distance dans un espace métrique correspond bien à une topologie au sens de la topologie générale. Les espaces vectoriels normés, de Banach et de Hilbert sont donc également associés de manière naturelle à des topologies.*

Les parties d'un espace topologique sont naturellement munies d'une topologie dite topologie induite.

Définition 192 (Topologie induite). *Soit $(E, \mathcal{T})$ un espace topologique et F une partie de E . On appelle topologie induite sur F la topologie*

$$\mathcal{T}|_F = \{U \cap F \mid U \in \mathcal{T}\}.$$

La plupart des topologies intéressantes sont issues de normes ou de distances, mais il faut connaître les deux exemples extrêmes :

Exemple 193. *Soit E un ensemble. $\{E, \emptyset\}$ est une topologie sur E dite grossière. L'ensemble $\mathcal{P}(E)$ des parties de E est une topologie également, appelée topologie discrète. Pour ces deux topologies, tous les ouverts sont fermés.*

Équivalent des notions d'analyse : On peut redéfinir de nombreuses notions d'analyse à partir de cette notion de topologie.

Définition 194 (Voisinage, suite convergente, application continue, adhérence, intérieur). *Soit $(E, \mathcal{T}_E)$ et $(F, \mathcal{T}_F)$ deux espaces topologiques.*

1. *Soit $x \in E$, on appelle voisinage de x tout ensemble $V \subset E$ tel qu'il existe $U \in \mathcal{T}_E$ tel que $U \subset V$ et $x \in U$.*
2. *Soit $x \in E$, on dit que la suite $(x_n)_{n \in \mathbb{N}}$ converge vers x si pour tout voisinage V de x , on a*

$$\exists n_0 \in \mathbb{N}, \forall n \in \mathbb{N}, (n > n_0 \implies x_n \in V).$$

3. Soit $f : E \rightarrow F$ une application. Elle est dite continue si

$$\forall U \in \mathcal{T}_F, \quad f^{-1}(U) \in \mathcal{T}_E,$$

où $f^{-1}(U) = \{x \in E \mid f(x) \in U\}$ est l'image réciproque de l'ensemble U .

4. Soit $A \subset E$. On appelle adhérence de A et on note $\overline{A}$ le plus petit fermé qui contient A , soit

$$\overline{A} = \bigcap_{A \subset F, F \text{ fermé}} F.$$

5. Soit $A \subset E$. On appelle intérieur de A et on note A° le plus grand ouvert contenu dans A , soit

$$A^\circ = \bigcup_{U \subset A, U \text{ ouvert}} U.$$

On prendra garde lors du maniement de la notion de limite dans un espace topologique de ne pas supposer l'unicité de la limite; c'est faux en général. Cela nous amène à la propriété de séparation sans laquelle il est difficile de dire quelque chose.

Définition 195 (Espace séparé). *Un espace topologique $(E, \mathcal{T})$ est dit séparé si pour tous $x, y \in E$, il existe des ouverts $U_x, U_y \in \mathcal{T}$ tels que $x \in U_x$, $y \in U_y$ et $U_x \cap U_y = \emptyset$.*

Si l'espace n'est pas séparé il y a des points qu'on ne peut pas distinguer et les limites ne sont pas uniques. Dans un espace séparé il y a trivialement unicité de la limite (réfléchir quelques secondes!).

Remarque 196. *Tous les espaces métriques sont des espaces topologiques séparés.*

Définition 197 (Espace connexe). *Un espace topologique est dit connexe si ses seules parties à la fois ouvertes et fermées sont lui-même et l'ensemble vide.*

On dira en abrégé qu'une partie d'un espace topologique est connexe si elle l'est comme espace topologique pour la topologie induite.

Propriété 198. *Soit $(E, \mathcal{T})$ un espace topologique. Il y a équivalence entre les propriétés suivantes.*

- *L'ensemble E est connexe (ses seules parties à la fois ouvertes et fermées sont lui-même et l'ensemble vide);*
- *L'ensemble E n'est pas la réunion de deux ouverts non-vides et disjoints;*
- *L'ensemble E n'est pas la réunion de deux fermés non-vides et disjoints;*
- *Toute application continue de E dans un espace topologique à deux éléments muni de la topologie discrète est constante.*

On définit de même la notion de compacité.

Définition 199 (Espace compact). *Un espace topologique $(E, \mathcal{T})$ est dit compact s'il est séparé et si l'on peut extraire de tout recouvrement d'ouverts de cet espace un recouvrement fini : Si $(U_i)_{i \in I}$ est une famille d'éléments de $\mathcal{T}$ et si $E = \cup_{i \in I} U_i$, il existe des ouverts $U_{i_1}, \dots, U_{i_n}$ dans cette famille tels que $E = \cup_{k=1}^n U_{i_k}$.*

On se méfiera du fait que tous les pays n'ont pas la même définition de compact. La propriété d'être séparé n'est pas toujours demandée.³¹

Topologies produit et quotient : On munit naturellement le produit cartésien de deux (ou plus) espaces topologiques d'une topologie en disant que U est un ouvert de $E \times F$ si pour tout point $x \in U$, il existe $V_E \subset E$ et $V_F \subset F$ ouverts tels que $x \in V_E \times V_F$ et $V_E \times V_F \subset U$. On parle de topologie produit, c'est la topologie la moins fine rendant continues les projections de $E \times F$ sur E et F . La topologie produit est une topologie initiale au sens où c'est la moins fine qui rende continues des applications qui partent vers d'autres espaces.

Soit E un espace topologique et soit R une relation d'équivalence sur E . Soit $p : E \rightarrow E/R$ la projection naturelle. On munit E/R de la topologie la plus fine qui rende p continue et on appelle cette topologie topologie quotient. L'idée de la topologie quotient est de coller des points entre eux en recollant les ouverts de manière intelligente autour. La topologie quotient est une topologie finale au sens où elle est la plus fine qui rende continue une application qui aboutisse sur elle.

Homéomorphismes : Nous arrivons à présent à l'une des notions les plus centrales de la topologie ; nous l'avons déjà rencontrée précédemment.

Définition 200 (Homéomorphisme). *Soient $(E, \mathcal{T}_E)$ et $(F, \mathcal{T}_F)$ deux espaces topologiques. On appelle homéomorphisme d'espaces topologiques, ou en abrégé homéomorphisme, une application $f : E \rightarrow F$ continue, bijective et dont la bijection réciproque est continue. Les espaces E et F sont dits homéomorphes s'il existe un homéomorphisme entre eux.*

Deux espaces homéomorphes sont en quelque sorte indiscernables du point de vue de la topologie (mais ils peuvent être très différents sous d'autres aspects). La plupart des propriétés sont conservées par homéomorphismes.

Propriété 201. *Soient $(E, \mathcal{T}_E)$ et $(F, \mathcal{T}_F)$ deux espaces topologiques homéomorphes. Si E est connexe, F est connexe. Si E est compact, F est compact. L'image de tout ouvert de E par un homéomorphisme est un ouvert de F , même chose pour les fermés.*

31. Remarquons que la séparation n'étant pas une propriété locale, nous pouvons définir une notion de locale compacité qui n'est pas une propriété locale. Eh oui...

2. Quelques résultats importants

2.a. Prolongement d'applications

Théorème 202

Soient (E, d_E) un espace métrique et (F, d_F) un espace métrique complet. Soient A une partie dense de E et $f : A \rightarrow F$ une application uniformément continue sur A . Alors il existe une unique application continue $\tilde{f} : E \rightarrow F$ telle que $\tilde{f}|_A = f$ et elle est uniformément continue.

Démonstration : Soit $x \in E$. Comme A est dense dans E , il existe une suite $(x_n)_{n \in \mathbb{N}}$ à valeurs dans A qui converge vers x . Par uniforme continuité, pour tout $\varepsilon \in \mathbb{R}_+^*$, il existe $\eta \in \mathbb{R}_+^*$ tel que si $d_E(x_p, x_q) \leq \eta$, alors

$$d_F(f(x_p), f(x_q)) \leq \varepsilon.$$

On en déduit, puisque la suite convergente $(x_n)_{n \in \mathbb{N}}$ est de Cauchy, que la suite $(f(x_n))_{n \in \mathbb{N}}$ est également de Cauchy, et donc convergente puisque F est complet. On note $\tilde{f}(x)$ sa limite.

Si $(y_n)_{n \in \mathbb{N}}$ est une autre suite de A qui tend vers x , on a

$$d_F(\tilde{f}(x), f(y_n)) \leq d_F(\tilde{f}(x), f(x_n)) + d_F(f(x_n), f(y_n)).$$

La quantité de droite tend vers 0 par continuité de f et d'après ce qui précède, donc celle de gauche tend vers 0 et la définition de $\tilde{f}$ ne dépend pas de la suite choisie.

L'application $\tilde{f}$ est continue car pour tout x , pour tout y , pour toutes suites $(x_n)_{n \in \mathbb{N}}$ et $(y_n)_{n \in \mathbb{N}}$ qui convergent vers x et y respectivement, on a pour tout entier n ,

$$d_F(\tilde{f}(x), \tilde{f}(y)) \leq d_F(\tilde{f}(x), f(x_n)) + d_F(f(x_n), f(y_n)) + d_F(\tilde{f}(y), f(y_n)).$$

Pour tout $\varepsilon \in \mathbb{R}_+^*$, on peut choisir n assez grand tel que

$$d_F(\tilde{f}(x), f(x_n)) \leq \frac{\varepsilon}{3} \quad \text{et} \quad d_F(\tilde{f}(y), f(y_n)) \leq \frac{\varepsilon}{3},$$

et par uniforme continuité de f , il existe $\eta \in \mathbb{R}_+^*$ tel que si $d_E(x_n, y_n) \leq \eta$, alors $d_F(f(x_n), f(y_n)) \leq \frac{\varepsilon}{3}$. Donc, si $d_E(x, y) < \eta$, on peut prendre n suffisamment grand pour obtenir

$$d_F(\tilde{f}(x), \tilde{f}(y)) \leq \varepsilon.$$

Donc la fonction $\tilde{f}$ est uniformément continue.

L'unicité est immédiate par caractérisation séquentielle de la limite et unicité de la limite. ■

2.b. Conséquences de la propriété de Baire

Les espaces de Banach, comme tous les espaces métriques complets, satisfont à la propriété de Baire, ce qui a de nombreuses conséquences très importantes en analyse fonctionnelle. Dans toute sa généralité, le lemme de Baire s'énonce comme suit.

Théorème 203 (Lemme de Baire, 1899)

Soit E un espace métrique complet et soit $(F_n)_{n \in \mathbb{N}}$ une famille dénombrable de fermés de E d'intérieurs vides. Alors l'ensemble

$$F = \bigcap_{n=0}^{\infty} F_n$$

est une partie de E d'intérieur vide.

Remarque 204. Ce théorème est équivalent à dire que dans un espace métrique complet l'intersection d'une famille dénombrable d'ouverts denses est une partie dense de E . Ces deux manières d'énoncer le résultat ont chacune leur intérêt pratique mais transmettent la même idée.

Théorème 205 (Théorème de Banach-Steinhaus, 1927)

Soit E et F deux espaces de Banach, I un ensemble quelconque et pour tout $i \in I$, $T_i : E \rightarrow F$ une application linéaire continue. Si

$$\forall x \in E, \sup_{i \in I} \|T_i x\|_F < +\infty,$$

alors

$$\sup_{i \in I} \|T_i\|_{\mathcal{L}(E, F)} < +\infty.$$

Une conséquence directe est que les opérateurs convergeant ponctuellement ont une limite simple continue, comme l'exprime le résultat suivant.

Théorème 206 (Continuité de l'opérateur limite)

Soient E et F des espaces de Banach. Soit une suite $(T_n)_{n \in \mathbb{N}}$ d'applications linéaires continues telles que pour tout $x \in E$, la suite $(T_n(x))_{n \in \mathbb{N}}$ admet une limite Tx . Alors l'application T ainsi définie est une application linéaire et continue et on a

$$\sup_{n \in \mathbb{N}} \|T_n\| < +\infty \quad \text{et} \quad \|T\| \leq \liminf_{n \rightarrow +\infty} \|T_n\|.$$

Démonstration : Par convergence de la suite $(T_n(x))_{n \in \mathbb{N}}$, on a

$$\sup_{n \in \mathbb{N}} \|T_n x\|_F < +\infty.$$

Donc d'après le théorème de Banach-Steinhaus, le supremum des normes des opérateurs T_n est fini.

La linéarité de T provient directement de la linéarité des éléments de la suite. Notons

$$C = \sup_{n \in \mathbb{N}} \|T_n\|.$$

On a par hypothèses

$$\forall x \in E, \forall n \in \mathbb{N}, \|T_n x\| \leq C \|x\|,$$

donc par passage à la limite, T est continue. Enfin,

$$\|Tx\| = \lim_{n \rightarrow +\infty} \|T_n(x)\| \leq \|x\| \liminf_{n \rightarrow +\infty} \|T_n\|,$$

d'où la dernière inégalité. ■

Théorème 207 (Théorème de l'application ouverte, 1930)

Soient E et F des espaces de Banach et $T \in \mathcal{L}(E, F)$ un opérateur surjectif. Alors T est une application ouverte (l'image d'un ouvert de E par T est un ouvert de F).

Le théorème de l'application ouverte permet de démontrer deux résultats très utiles dans l'étude des opérateurs et des équations différentielles.

Théorème 208 (Théorème de Banach)

Soient E et F des espaces de Banach et $T \in \mathcal{L}(E, F)$. Si T est bijective, sa bijection réciproque T^{-1} est linéaire et continue.

En particulier, T est alors un isomorphisme d'espaces de Banach.

Théorème 209 (Théorème du graphe fermé)

Soit E et F des espaces de Banach et $T : E \rightarrow F$ une application linéaire. Alors T est continue si et seulement si l'ensemble

$$\text{Gr}(T) = \{(x, Tx) \mid x \in E\} \subset E \times F,$$

appelé graphe de T , est fermé dans $E \times F$ pour la topologie produit.

3. Dualité

Nous allons nous pencher un peu sur un objet que nous avons beaucoup vu dans le corps principal du document sans jamais entrer dans des détails spécifiques, le dual topologique d'un espace vectoriel normé. Soit E un espace de Banach sur $\mathbb{K}$ (l'hypothèse de complétude n'est pas toujours utile mais demeure confortable), l'espace dual noté E^* est l'ensemble des formes linéaires $f : E \rightarrow \mathbb{K}$ continues.

Le théorème suivant, laborieux à démontrer, est une conséquence du théorème de Radon-Nykodým.

Théorème 210

Soit une mesure σ -finie m sur $\mathbb{R}^n$, soient $p, q \in]1, +\infty[$ tels que

$$\left\{ \begin{array}{l} \frac{1}{p} + \frac{1}{q} = 1. \text{ On a alors} \\ L^p(\mathbb{R}^n, m)^* = L^q(\mathbb{R}^n, m) \quad \text{et} \quad L^1(\mathbb{R}^n, m)^* = L^\infty(\mathbb{R}^n, m). \end{array} \right.$$

3.a. Topologie faible

Définition 211. La topologie faible sur un espace vectoriel normé E est la topologie la moins fine pour laquelle toutes les formes linéaires continues (pour la topologie associée à la norme) sur E demeurent continues pour ladite topologie faible. On la note $\sigma(E, E^*)$ ou w comme weak³².

Pour bien distinguer les propriétés de la topologie faible et celles de la topologie issue de la norme on mettra en opposition les propriété *faibles* et *fortes*. On parlera de convergence faible pour la topologie faible et de convergence forte pour la norme ; un ensemble sera faiblement ou fortement fermé, respectivement ouvert, une application sera faiblement ou fortement continue.

Proposition 212. La topologie faible est moins fine que la topologie issue de la norme ; i.e. l'application identité $i : (E, \|\cdot\|) \rightarrow (E, w)$ est continue. En dimension finie, ces deux topologies sont identiques. En dimension infinie elles ne le sont jamais.

Proposition 213. Un espace de Banach muni de sa topologie faible est un espace vectoriel topologique localement convexe et séparé.

Proposition 214. Une suite $(x_n)_{n \in \mathbb{N}}$ converge vers x pour la topologie faible si et seulement si pour toute forme linéaire continue ϕ , la suite $(\phi(x_n))_{n \in \mathbb{N}}$ converge vers $\phi(x)$ dans $\mathbb{K}$.

Exemple 215. Par le théorème de représentation de Riesz, dans un espace de Hilbert H une suite $(x_n)_{n \in \mathbb{N}}$ converge faiblement vers x si et seulement si

$$\forall y \in H, \lim_{n \rightarrow +\infty} \langle x_n | y \rangle = \langle x | y \rangle.$$

De même, dans $L^p(\mathbb{R}^n)$, une suite $(f_n)_{n \in \mathbb{N}}$ converge faiblement vers f si et seulement si

$$\forall g \in L^q(\mathbb{R}^n), \lim_{n \rightarrow +\infty} \int_{\mathbb{R}^n} f_n(x)g(x) dx = \int_{\mathbb{R}^n} f(x)g(x) dx,$$

avec $\frac{1}{p} + \frac{1}{q} = 1$.

Théorème 216

Toute partie convexe fermée d'un espace de Banach qui est fortement fermée est faiblement fermée.

32. Or like *wicked*, cuz this mothafuckin' topology is a damn wicked one.

Démonstration : Les sous-espaces vectoriels fermés pour la norme sont tous images réciproques de formes linéaires continues. Comme par définition de la topologie faible ces formes linéaires sont aussi faiblement continues, leurs images réciproques sont également faiblement fermées. Les convexes fermés sont des intersections d'hyperplans fermés, donc les convexes fermés sont également les mêmes pour les topologies faibles et fortes. ■

Remarque 217 (Important). *C'est en particulier vrai pour les sous-espaces vectoriels fermés qui sont donc les mêmes pour les deux topologies. Attention cependant, ce n'est pas vrai pour toutes les parties fermées.*

Proposition 218. *Toute suite faiblement convergente dans un espace de Banach est bornée.*

Proposition 219. *Soit H un espace de Hilbert. Si une suite $(x_n)_{n \in \mathbb{N}}$ converge faiblement vers x et si la suite réelle $(\|x_n\|)_{n \in \mathbb{N}}$ converge vers $\|x\|$, alors la suite $(x_n)_{n \in \mathbb{N}}$ converge fortement (en norme) vers x .*

Théorème 220

Soit E et F deux espaces de Banach. Une application $T : E \rightarrow F$ linéaire est continue pour les topologies issues des normes sur E et F si et seulement si elle est continue pour les topologies faibles sur E et F .

3.b. Topologie faible—*

Pour certains besoins, la topologie faible n'est, en quelque sorte, pas assez faible, au sens où ses propriétés ne sont pas encore assez fortes (oulalaaaa :'(...). Plus prosaïquement, si un espace vectoriel F s'avère être le dual topologique d'un espace vectoriel E , c'est-à-dire $F = E^*$, on peut définir sur F une topologie qui présente des propriétés intéressantes : la topologie faible étoile.

Définition 221 (Topologie faible—*). *Soit E un espace vectoriel normé. On appelle topologie faible—* sur E^* , aussi appelée topologie préfaible, la topologie la moins fine rendant continues toutes les applications de la forme*

$$\begin{aligned} f_x : E^* &\rightarrow \mathbb{K} \\ \phi &\mapsto \phi(x), \end{aligned}$$

où $x \in E$. on la note $\sigma(E^*, E)$ ou $w - *$ comme weak-star.

Remarque 222. *On peut également définir une topologie faible $\sigma(E^*, E^{**})$ sur E^* . Et une topologie forte subordonnée à la norme de E . La topologie faible—* est alors moins fine que la topologie faible qui elle-même est moins fine que la topologie forte.*

Proposition 223. *Un espace de Banach muni de sa topologie faible—* est un espace vectoriel topologique localement convexe et séparé.*

Contrairement au cas de la topologie faible, on perd des formes linéaires continues avec la topologie faible- $*$.

Théorème 224

Les formes linéaires continues pour la topologie faible- $*$ sont et sont uniquement les applications de la forme

$$\begin{aligned} f_x : E^* &\rightarrow \mathbb{K} \\ \phi &\mapsto \phi(x), \end{aligned}$$

où $x \in E$.

Un théorème fondamental à propos de la topologie faible- $*$ est la compacité de la boule unité, même en dimension infinie. Banach l'avait montré dans le cas particulier des espaces séparables en 1929.

Théorème 225 (Théorème de Banach-Alaoglu, 1940)

La boule unité fermée de l'espace dual d'un espace vectoriel normé pour la topologie forte (issue de la norme subordonnée) est compacte pour la topologie faible- $*$.

Remarque 226. *La boule unité fermée forte est donc faiblement fermée.*

Comme on l'a vu, la topologie faible sur un dual E^* implique l'utilisation du bidual E^{**} . Rappelons que tout espace de Banach possède une injection dite canonique dans son bidual :

$$\begin{aligned} i : E &\rightarrow E^{**} \\ x &\mapsto \left\{ \begin{array}{ll} f_x : E^* &\rightarrow \mathbb{K} \\ \phi &\mapsto \phi(x) \end{array} \right. . \end{aligned}$$

Cela motive la définition suivante.

Définition 227. *Un espace de Banach est dit réflexif si l'injection canonique dans son bidual est surjective.*

Proposition 228. *Sur le dual d'un espace de Banach réflexif, les topologies faible et faible- $*$ coïncident.*

Exemple 229. *Tout espace de Hilbert est réflexif et pour tout $p \in]1, +\infty[$ l'espace L^p est réflexif.*

Ces espaces ont donc une boule unité faiblement compacte (pas dur mais pas absolument direct). On a même mieux.

Théorème 230 (Théorème de Kakutani, 1940)

Un espace de Banach est réflexif si et seulement si sa boule unité fermée est faiblement compacte.

4. Topologie des espaces de fonctions tests

4.a. Les espaces de fonctions différentiables

Soit Ω un ouvert de $\mathbb{R}^n$ et notons $\mathbb{K}$ pour désigner $\mathbb{R}$ ou $\mathbb{C}$. Comme précédemment, nous notons ici $\mathcal{C}^k(\Omega)$ l'ensemble des fonctions de Ω dans $\mathbb{K}$ qui sont k fois continûment différentiables, $\mathcal{C}^\infty(\Omega)$ l'ensemble des fonctions indéfiniment différentiables, $\mathcal{C}_0^k(\Omega)$ l'ensemble des fonctions de $\mathcal{C}^k(\Omega)$ qui sont à support compact et $\mathcal{C}_0^\infty(\Omega)$ l'ensemble des fonctions de $\mathcal{C}^\infty(\Omega)$ qui sont indéfiniment différentiables.

Nous allons étudier (un peu) la topologie de ces espaces. Ce sont pour commencer tous des espaces vectoriels sur $\mathbb{K}$.

La topologie de ces espaces sur un compact au lieu d'un ouvert étant beaucoup plus simple, nous allons écrire Ω comme une réunion dénombrable de compacts. Nous disons que c'est une suite d'exhaustion de Ω . Remarquons que si, pour $i \in \mathbb{N}$, $i \neq 0$,

$$K_i = \{x \in \Omega \mid \|x\| \leq i, d(x, \Omega^c) \geq \frac{1}{i}\},$$

où Ω^c est le complémentaire de Ω dans $\mathbb{R}^n$ et la norme utilisée la norme euclidienne de $\mathbb{R}^n$, alors K_i est compact, $K_i \subset K_{i+1}^\circ$, pour tout compact K de Ω il existe un K_i tel que $K \subset K_i$ et

$$\Omega = \bigcup_{i=1}^{+\infty} K_i = \bigcup_{i=1}^{+\infty} K_i^\circ.$$

Tous ces résultats se vérifient facilement.

On peut alors considérer les semi-normes suivantes, pour $i \in \mathbb{N}$, $i \neq 0$, et $f \in \mathcal{C}^k(\Omega)$

$$p_i(f) = \sum_{|\alpha| \leq k} \sup_{x \in K_i} |\partial^\alpha f(x)|,$$

et pour $f \in \mathcal{C}^\infty(\Omega)$,

$$p_i(f) = \sum_{|\alpha| \leq i} \sup_{x \in K_i} |\partial^\alpha f(x)|.$$

Ces applications sont des semi-normes, elles vérifient tous les axiomes des normes sauf la propriété de séparation. Nous allons les utiliser pour définir une topologie.

Pour $f \in \mathcal{C}^k(\Omega)$, $k \in \bar{\mathbb{N}}$, $\varepsilon \in \mathbb{R}_+^*$ et $i \in \mathbb{N}^*$, posons

$$V_{i,\varepsilon} = \{g \in \mathcal{C}^k(\Omega) \mid p_i(g - f) < \varepsilon\}.$$

On dira qu'un ensemble est voisinage de $f \in \mathcal{C}^k(\Omega)$ s'il contient un ensemble de type $V_{i,\varepsilon}$. On définit les ouverts de $\mathcal{C}^k(\Omega)$ comme les ensembles qui sont

voisinages de tous leurs points.

On peut vérifier qu'en réunissant tous les ouverts ainsi définis, on obtient une topologie $\mathcal{T}$ sur $\mathcal{C}^k(\Omega)$, $k \in \bar{\mathbb{N}}$.

On vérifie également qu'une suite $(f_n)_{n \in \mathbb{N}}$ de $\mathcal{C}^k(\Omega)$, $k \in \bar{\mathbb{N}}$, converge vers f pour la topologie $\mathcal{T}$ définie ci-dessus si et seulement si pour tout $|\alpha| \leq k$, (donc pour tout α si $k = +\infty$) et pour tout compact K de Ω , la suite $(\partial^\alpha f_n)_{n \in \mathbb{N}}$ converge uniformément vers $\partial^\alpha f$ sur K .

Théorème 231

Soit sur $\mathcal{C}^k(\Omega)^2$, $k \in \bar{\mathbb{N}}$, l'application

$$d : (f, g) \mapsto \sum_{i=1}^{+\infty} \frac{1}{2^i} \frac{p_i(f - g)}{1 + p_i(f - g)}.$$

Alors d est une distance sur $\mathcal{C}^k(\Omega)$ qui définit la même topologie que celle que nous avons déjà construite sur cet espace et l'espace $(\mathcal{C}^k(\Omega), d)$ est un espace métrique complet.

Théorème 232

La topologie que nous avons construite est métrisable mais pas normable.

4.b. Les espaces de fonctions différentiables à support compact

La topologie des espaces $\mathcal{C}_0^k(\Omega)$ est un poil plus délicate et nous n'entrons pas dans tous les détails. Reprenons les compacts K_i définis ci-dessus et remarquons que

$$\mathcal{C}_0^k(\Omega) = \bigcup_{i=1}^{+\infty} \mathcal{C}_0^k(K_i).$$

On peut définir sur chaque ensemble $\mathcal{C}_0^k(K_i)$ la topologie associée à la norme

$$\|f\| = \sum_{|\alpha| \leq k} \sup_{x \in K_i} |\partial^\alpha f(x)|,$$

si $k \in \mathbb{N}$ ou à la norme

$$\|f\| = \sum_{|\alpha| \leq i} \sup_{x \in K_i} |\partial^\alpha f(x)|,$$

si $k = +\infty$.

Ces topologies définissent par le principe dit de limite inductive une topologie sur $\mathcal{C}_0^k(\Omega)$, comme l'exprime le théorème suivant.

Théorème 233

Il existe une unique topologie $\mathcal{T}$ sur $\mathcal{C}_0^k(\Omega)$, $k \in \bar{\mathbb{N}}$, appelée limite inductive stricte des topologies $\mathcal{T}_i$ précédentes, telle que $\mathcal{T}$ induit la topologie $\mathcal{T}_i$ sur $\mathcal{C}_0^k(K_i)$ pour tout $i \in \mathbb{N}^*$. De plus, une suite $(f_n)_{n \in \mathbb{N}}$ de $\mathcal{C}_0^k(\Omega)$, $k \in \bar{\mathbb{N}}$, converge vers f pour la topologie $\mathcal{T}$ si et seulement si il existe $i_0 \in \mathbb{N}^*$ tel que pour tout $j \in \mathbb{N}$, $\text{supp}(f_j) \subset K_{i_0}$ et la suite $(f_n)_{n \in \mathbb{N}}$ converge vers f dans $\mathcal{C}_0^k(K_{i_0})$.

Remarque 234 (Importante). *Par conséquent, une forme linéaire est continue sur $\mathcal{C}_0^k(\Omega)$, $k \in \bar{\mathbb{N}}$, si et seulement si sa restriction à tous les $\mathcal{C}_0^k(K_i)$ est continue.*

Cette topologie n'est pas normable.

5. Éléments de géométrie différentielle

De nombreux objets simples comme les cercles, les sphères, les tores sont adaptés pour de nombreux problèmes. Il est donc important de développer un attirail technique qui permette de travailler avec de tels objets de façon intrinsèque sans pour autant perdre les bénéfices du cadre vectoriel dont nous avons l'habitude. Nous allons présenter ici quelques définitions basiques de géométrie différentielle.

5.a. Les variétés différentielles

Une solution efficace pour travailler sur un cercle est de le couper en deux et de remarquer que chacun des deux morceaux est un segment ouvert de $\mathbb{R}$. Cela revient à dire que le cercle est localement un segment de $\mathbb{R}$. Plus généralement, on peut s'intéresser à toutes les parties de $\mathbb{R}^n$ qui sont constituées de morceaux de $\mathbb{R}^p$, $p \leq n$ déformés puis recollés. Nous aboutissons à la définition suivante, technique au premier abord mais charriant une idée relativement simple.

Définition 235 (Sous-variété différentielle). *Une partie $M \subset \mathbb{R}^n$ est une sous-variété différentielle de dimension $p \in \mathbb{N}$, de codimension $n - p$ et de classe $\mathcal{C}^k$ si pour tout $x \in M$ il existe un voisinage ouvert U de x dans $\mathbb{R}^n$, un voisinage ouvert V de 0 dans $\mathbb{R}^n$ et un $\mathcal{C}^k$ -difféomorphisme $f : U \rightarrow V$ tel que*

$$f(U \cap M) = V \cap (\mathbb{R}^p \times \{0\}^{n-p}).$$

Remarque 236. *La dimension d'une sous-variété est unique.*

Il existe plusieurs façons de vérifier sans trop de difficulté qu'un objet est bien une sous-variété. Nous en résumons quelques unes dans le théorème suivant (attention aux hypothèses, aux espaces, aux dimensions).

Théorème 237

| Soit M une partie de $\mathbb{R}^n$. Les propriétés suivantes sont équivalentes.

- La partie M est une sous-variété de dimension p de $\mathbb{R}^n$.
- Pour tout $x \in M$, il existe un ouvert U de $\mathbb{R}^n$ contenant x et une submersion $g : U \rightarrow \mathbb{R}^{n-p}$ telle que $U \cap M = g^{-1}(0)$.
- Pour tout $x \in M$, il existe un ouvert U de $\mathbb{R}^n$ contenant x , un ouvert V de $\mathbb{R}^p$ contenant 0 et une immersion $h : \Omega \rightarrow \mathbb{R}^n$ qui est un homéomorphisme de V sur $U \cap M$.
- Pour tout $x \in M$, il existe un ouvert U de $\mathbb{R}^n$ contenant $x = (x_1, \dots, x_n)$, un ouvert V de $\mathbb{R}^p$ contenant $(x_1, \dots, x_p)$ et une application f de classe $\mathcal{C}^k$ de V dans $\mathbb{R}^{n-p}$ tels que après permutation éventuelle des coordonnées, $U \cap M$ soit égal au graphe de f .

Exemple 238. Pour tout entier naturel n non nul, la sphère $\mathbb{S}_n = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \dots + x_{n+1}^2 = 1\}$ ³³ est une sous-variété de $\mathbb{R}^{n+1}$ de dimension n et de classe $\mathcal{C}^\infty$. En effet, l'application définie par $f(x) = x_1^2 + \dots + x_{n+1}^2 - 1$ a pour différentielle en tout point x de $\mathbb{S}_n$ $df_x = (2x_1, \dots, 2x_{n+1})$ et est donc une submersion.

Il est possible de généraliser la définition de sous-variété pour l'abstraire du cadre extérieur que constitue $\mathbb{R}^n$. On ne considère plus une partie de $\mathbb{R}^n$ mais un ensemble pris en lui-même. Cela facilite la construction de certaines notions. Comme les sous-variétés sont aussi des variétés, les notions ainsi construites s'y appliquent aussi bien.

S'il n'y a qu'une structure continue et qu'il n'est pas question de différentiabilité, il suffit de dire qu'autour de chaque point on ressemble à $\mathbb{R}^n$.

Définition 239 (Variété topologique³⁴). Une variété topologique de dimension n est un espace topologique séparé dont tout point est contenu dans un ouvert homéomorphe à un ouvert de $\mathbb{R}^n$.

Pour définir une variété différentielle, nous avons besoin de l'outil suivant.

Définition 240 (Cartes, Atlas). Soit X un espace topologique. On appelle carte de classe $\mathcal{C}^k$ de X un couple (U, ϕ) où U est un ouvert de X et $\phi : U \rightarrow \mathbb{R}^n$ est un $\mathcal{C}^k$ -difféomorphisme. L'ouvert U est le domaine de la carte. Un atlas de X est une famille $((U_i, \phi_i))_{i \in I}$ de cartes de X telle que

$$\bigcup_{i \in I} U_i = X.$$

33. De nombreux auteurs notent $\mathbb{S}^n$ la sphère de dimension n . À la suite de Frédéric Paulin et d'autres avant lui, nous rejetons cette notation pour la confusion qu'elle engendre avec le tore qui est le produit cartésien de n cercle, soit en notation courte $\mathbb{S}^n$.

34. Je n'ai pu me retenir de plagier mot pour mot, lettre pour lettre la prouesse de concision de Jacques Lafontaine, [7].

À présent, il faut introduire une notion de compatibilité des cartes de l'atlas afin d'être certain qu'il n'y a pas de problème à l'endroit où deux (ou plus) domaines se recouvrent dans l'espace topologique.

Définition 241 (Compatibilité, Atlas régulier). *On dit que deux cartes (U_1, ϕ_1) , (U_2, ϕ_2) de classe $\mathcal{C}^k$ d'un espace topologique sont compatibles d'ordre k si $U_1 \cap U_2 = \emptyset$ ou si l'application*

$$\phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \rightarrow \phi_2(U_1 \cap U_2)$$

est un $\mathcal{C}^k$ -difféomorphisme.

Un atlas de classe $\mathcal{C}^k$ est un atlas de cartes de classes $\mathcal{C}^k$ deux à deux compatibles d'ordre k .

Et voici enfin la bête.

Définition 242 (Variété différentielle). *Une variété différentielle de dimension n et de classe $\mathcal{C}^k$ est une variété topologique de dimension n munie d'un atlas de classe $\mathcal{C}^k$.*

Remarque 243. *On munit naturellement les variétés différentielles de l'atlas maximal associé à son atlas, c'est-à-dire celui qui contient toutes les cartes compatibles avec l'atlas de départ. On parle alors de structure différentielle. On peut définir directement la variété différentielle comme une variété topologique munie d'une structure différentielle, soit d'un atlas $\mathcal{C}^k$ maximal. L'atlas maximal est aussi dit complet ou saturé.*

En pratique, on définira toujours la structure différentielle en donnant le moins de cartes possibles, par exemple deux seulement pour un cercle, un sphère, etc.

Propriété 244. *Si M et N sont deux variétés différentielles de dimensions respectives n et m et de classe $\mathcal{C}^k$, le produit cartésien $M \times N$ peut être doté d'une structure de variété différentielle de classe $\mathcal{C}^k$ et de dimension $n + m$.*

Exemple 245 (Important, Tore). *On peut définir ainsi facilement le tore $\mathbb{T}^n$ de dimension n comme le produit cartésien de n cercles $\mathbb{S}$:*

$$\mathbb{T}^n = \mathbb{S}^n.$$

Il est également possible de le construire en considérant l'espace topologique constitué de l'hypercube unité de $\mathbb{R}^n$, $[0, 1]^n$, et en identifiant les côtés opposés deux à deux à l'aide de quotients d'espaces topologiques.

Le tore de dimension n peut également être construit comme une partie (et même une sous-variété) de $\mathbb{R}^{n+1}$ via des coordonnées locales. Le tore de dimension 1 est un cercle dans le plan (ou le segment $[0, 1]$ quotienté par le sous-ensemble $\{0, 1\}$). Le tore de dimension 2 est une "bouée" dans l'espace $\mathbb{R}^3$ (ou le carré $[0, 1]^2$ quotienté deux fois pour identifier les côtés opposés).

La propriété suivante permet de justifier facilement que les sous-variétés différentielles sont des variétés différentielles.

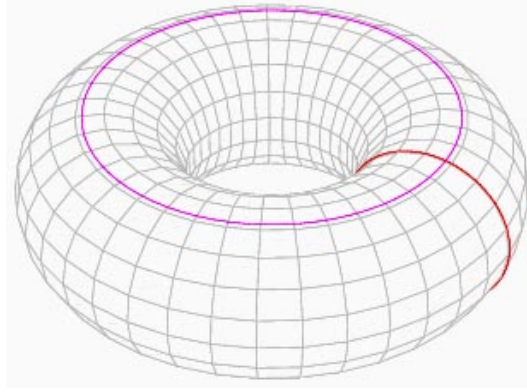


FIGURE 9 – Un tore de dimension 2, vivant sa vie de tore dans son environnement naturel : un polycopié de mathématiques.

Propriété 246. Soit M une sous-variété différentielle de dimension p et de classe $\mathcal{C}^k$. Soient (U_1, ϕ_1) et (U_2, ϕ_2) deux paramétrisations locales. Alors, si son domaine de définition est non-vidé, l'application

$$\phi_2^{-1} \circ \phi_1 : U_1 \cap \phi_1(\phi_2(U_2)) \rightarrow U_2 \cap \phi_2^{-1}(\phi_1(U_1))$$

est un $\mathcal{C}^k$ -difféomorphisme.

La réciproque est vraie en un certain sens. Le théorème de plongement de Whitney (forme faible en 1936 puis forte en 1944) assure que toute variété différentielle de dimension n à base dénombrable (dont la topologie admet une base dénombrable) peut être plongée dans $\mathbb{R}^{2n}$. La valeur $2n$ est optimale.

5.b. Espaces tangents

Considérons deux variétés différentielles M et N de dimensions respectives n_M et n_N et de classe $\mathcal{C}^k$.

Définition 247 (Application de classe $\mathcal{C}^k$). Une application continue $f : M \rightarrow N$ est dite de classe $\mathcal{C}^k$ si pour tout $m \in M$ il existe une carte (U, ϕ) de M avec $m \in U$, une carte (V, ψ) de N avec $f(m) \in V$ telles que l'application

$$\psi \circ f \circ \phi^{-1} : \phi(f^{-1}(V) \cap U) \rightarrow \psi(V)$$

soit de classe $\mathcal{C}^k$, en tant que fonction de $\mathbb{R}^{n_M}$ dans $\mathbb{R}^{n_N}$.

On vérifie facilement que la composée d'applications de classe $\mathcal{C}^k$ est de classe $\mathcal{C}^k$.

Notons $\mathcal{C}_m^M$ l'ensemble des courbes lisses $\gamma : I \rightarrow M$ définies sur un intervalle ouvert I contenant 0 et tel que $\gamma(0) = m$.

Définition 248 (Courbes tangentes). Deux courbes $\gamma_1: I_1 \rightarrow M$ et $\gamma_2: I_2 \rightarrow M$ de $\mathbb{C}_m^M$ sont dites tangentes en m si $\gamma_1(0) = \gamma_2(0) = m$ et s'il existe une carte (U, ϕ) telle que $m \in U$ et

$$(\phi \circ \gamma_1)'(0) = (\phi \circ \gamma_2)'(0) \quad (\in \mathbb{R}^{n_M})$$

On vérifie que cette définition ne dépend pas de la carte et induit une relation d'équivalence, ce qui mène à la définition qui nous intéresse.

Définition 249 (Vecteur tangent). Un vecteur tangent à M en un point m de M est une classe d'équivalence de la relation ci-dessus. L'ensemble des vecteurs tangents à M en m forme l'espace tangent à M en m , on le note $T_m M$.

Dans le cas où $M = \mathbb{R}^{n_M}$, l'espace tangent en tout point est de manière évidente $\mathbb{R}^{n_M}$ (pas de questions à se poser sur les cartes). Cette idée se généralise.

Proposition 250. $T_m M$ admet une structure d'espace vectoriel réel de dimension n_M .

Démonstration : i Commençons par montrer que $T_m M$ est en bijection avec $\mathbb{R}^{n_M}$.

Soit (U, ϕ) une carte telle que $m \in U$. Puisque la quantité $(\phi \circ \gamma)'(0)$ ne dépend que de la classe d'équivalence de γ , on peut définir l'application

$$\begin{aligned} \theta_\phi : T_m M &\rightarrow \mathbb{R}^{n_M} \\ \xi &\mapsto (\phi \circ \gamma)'(0) \end{aligned}$$

Par définition, θ_ϕ est injective. Voyons qu'elle est surjective : tout vecteur $v \in \mathbb{R}^{n_M}$ est l'image par θ_ϕ de la classe d'équivalence de la courbe γ_v définie par $t \mapsto \phi^{-1}(tv)$. En effet :

$$(\phi \circ \gamma_v)'(0) = \frac{d}{dt} (t \mapsto tv)_{|t=0} = v$$

et donc θ_ϕ est bijective.

ii Ensuite, on munit $T_m M$ d'une structure d'espace vectoriel.

Si (V, ψ) est une autre carte telle que $m \in V$ et $v \in \mathbb{R}^{n_M}$.

$$(\theta_\phi \circ \theta_\psi^{-1})(v) = D_{\psi(m)} (\phi \circ \psi^{-1}) \cdot v$$

Ainsi, par propriété de la différentielle, $\theta_\phi \circ \theta_\psi^{-1}$ est une application linéaire et on gagne donc "gratuitement" la compatibilité avec la structure d'espace vectoriel des opérations $+$ et $\cdot$ définies par : $\forall \xi, \zeta \in T_m M, \forall \lambda \in \mathbb{R}$,

$$\xi + \zeta := \theta_\phi^{-1}(\theta_\phi(\xi) + \theta_\phi(\zeta)) \quad \text{et} \quad \lambda \cdot \xi := \theta_\phi^{-1}(\lambda \theta_\phi(\xi))$$

et par définition de ϕ , la construction ci-dessus ne dépend pas de ϕ . ■

Ces espaces tangents permettent de définir un analogue de la dérivée sur une variété.

Définition 251 (Application linéaire tangente). Soit $f : M \rightarrow N$ une application de classe $\mathcal{C}^k$, $k > 1$, et $m \in M$. On définit une application linéaire tangente de f en m , notée $T_m f$, par passage au quotient de l'application $\gamma \mapsto f \circ \gamma$ de $\mathcal{C}_m^M$ dans $\mathcal{C}_{f(m)}^N$.

Exemple 252. Si M et N sont des espaces vectoriels, l'application $T_m f$ est la différentielle $d_m f$ de f en m .
En effet, $(f \circ \gamma)'(0) = d_m f \cdot \gamma'(0)$ d'après la formule de dérivation composée d'où le résultat en quotientant.

En réalité, nous considérerons par la suite des applications de I intervalle de $\mathbb{R}$ dans une variété M , ce qui élimine une grande partie des difficultés liées à la manipulation de cartes.

5.c. Fibrations

Soit $k \in [0, +\infty]$. On définit ici la notion d'espaces fibrés, liée à celle de fibré tangent (proposition 319) et de champ de vecteur sur une variété (définition 320).

Définition 253 (Fibration, Espace fibré). Soient E et B deux variétés $\mathcal{C}^k$. Une application p de classe $\mathcal{C}^k$ de E dans B est une fibration – de base B et d'espace total E – si pour tout $b \in B$ il existe un ouvert U contenant b , une variété F et un difféomorphisme $\phi : U \times F \rightarrow p^{-1}(U)$ tel que $p(\phi(y, z)) = y$ pour tout y de U et z de F . On dit alors que E est un espace fibré.

Pour tout $b \in B$, $E_b := p^{-1}(b)$ est une sous-variété fermée de E , de dimension $\dim(E) - \dim(B)$ et que l'on appelle *fibres* de b . Par construction E_b est difféomorphe à F .

Intuitivement et au moins localement, cette définition signifie que l'on peut voir E comme un "empilement" indexé par F , éventuellement continu donc, d'ouverts difféomorphes à U . Dans le cas où la base B est connexe, il existe une "unique" façon d'indexer ces ouverts; d'où le fait qu'on parle souvent de *fibres type*...

Exemple 254. Un exemple important d'espaces fibrés est celui des revêtements. Dans ce cas, "la" fibre F est discrète. Le revêtement du cercle $\mathbb{S}^1$ par une hélice est un exemple parmi d'autres...

Définition 255 (Fibration triviale, Fibration trivialisable, Trivialisation). La fibration triviale est celle pour laquelle $E = B \times F$ et $p = pr_1$. On dit qu'une fibration est trivialisable s'il existe un difféomorphisme $\phi : E \rightarrow B \times F$ tel que $p = pr_1 \circ \phi$. On dit alors que le fibré E est trivial (resp. trivialisable) et que ϕ est une trivialisation.

Remarque 256. Par définition, les fibrations sont localement trivialisables. Autrement dit, la fibration devient trivialisable au-dessus d'ouverts suffisamment petits de la base.

Définition 257 (Fibré vectoriel). *Un fibré vectoriel réel (resp. complexe) de rang k sur une variété B est un espace fibré (E, p, B) tel que :*

- i. la fibre type F et les fibres $p^{-1}(b)$, $b \in B$, sont des espaces vectoriels réels (resp. complexes),*
- ii. pour toute trivialisation locale ϕ , la restriction de ϕ à $p^{-1}(b)$ induit un isomorphisme d'espaces vectoriels sur F .*

De même que précédemment, on note E_b la fibre $p^{-1}(b)$ pour $b \in B$.

Définition 258 (Section). *Une section d'un fibré vectoriel E de base B est une application lisse s de B dans E telle que $p \circ s = Id_B$, autrement dit telle que $s(b) \in E_b$ pour tout b de B .*

6. Dimension fractale

Il est possible de caractériser les objets à géométrie fractale à l'aide d'un outil relativement simple : la dimension de Hausdorff.

Soient P une partie de $\mathbb{R}^n$, s un réel positif et δ un réel strictement positif. Posons

$$\mathcal{H}_\delta^s(P) = \inf \left\{ \sum_{i \in I} r_i^s \mid P \subset \bigcup_{i \in I} B(x_i, r_i), r_i < \delta, I \subset \mathbb{N} \right\}.$$

On remarque que, de manière évidente, si $\delta \leq \delta'$, alors

$$\mathcal{H}_{\delta'}^s(P) \leq \mathcal{H}_\delta^s(P).$$

On peut donc définir sans risque l'objet d'intérêt, ce que nous nous empressons de faire.

Définition 259 (s -mesure de Hausdorff). *On appelle s -mesure de Hausdorff l'application $\mathcal{H}^s : \mathcal{P}(\mathbb{R}^n) \rightarrow [0, +\infty]$ définie par*

$$\mathcal{H}^s(P) = \lim_{\delta \rightarrow 0} \mathcal{H}_\delta^s(P) = \sup_{\delta > 0} \mathcal{H}_\delta^s(P).$$

Définition 260 (Mesure extérieure). *On appelle mesure extérieure sur $\mathbb{R}^n$ une application $m : \mathcal{P}(\mathbb{R}^n) \rightarrow [0, +\infty]$ telle que*

- $m(\emptyset) = 0$;
- $\forall A, B \in \mathcal{P}(\mathbb{R}^n), A \subset B \implies m(A) \leq m(B)$;
- Pour tout ensemble I dénombrable, $m(\bigcup_{i \in I} A_i) \leq \sum_{i \in I} m(A_i)$.

Théorème 261

Pour tout $s \in \mathbb{R}_+$, la s -mesure de Hausdorff est une mesure extérieure sur $\mathbb{R}^n$.

Remarque 262. Si $s \in \mathbb{N}$, $\mathcal{H}^s$ est proportionnelle à la mesure extérieure de Lebesgue de dimension s .

Théorème 263

Pour tout $P \subset \mathbb{R}^n$, on a

- si $\mathcal{H}^{s_0}(P) < +\infty$, alors pour tout $s > s_0$, $\mathcal{H}^s(P) < +\infty$;
- si $\mathcal{H}^{s_0}(P) > 0$, alors pour tout $s < s_0$, $\mathcal{H}^s(P) = +\infty$;
- si $s > n$, alors $\mathcal{H}^s(P) = 0$.

Corollaire 264. Pour tout $P \subset \mathbb{R}^n$, il existe un unique $s_0 \in \mathbb{R}_+$ tel que

- $\forall s < s_0$, $\mathcal{H}^s(P) = +\infty$;
- $\forall s > s_0$, $\mathcal{H}^s(P) = 0$.

Définition 265 (Dimension de Hausdorff). Soit P une partie de $\mathbb{R}^n$, on appelle dimension de Hausdorff, ou dimension fractale, de P et on note $\dim_{\mathcal{H}}(P)$ l'unique réel s_0 tel que $\forall s < s_0$, $\mathcal{H}^s(P) = +\infty$ et $\forall s > s_0$, $\mathcal{H}^s(P) = 0$.

On appellera fractales les parties de $\mathbb{R}^n$ qui possèdent une dimension de Hausdorff non-entière.

Exemple 266 (Ensemble de Cantor – un grand classique). Soit l'ensemble

$$\mathcal{C} = \bigcap_{n=1}^{+\infty} \bigcup_{k=0}^{3^n-1} \left(\left[\frac{3^k}{3^n}, \frac{3^{k+1}}{3^n} \right] \cup \left[\frac{3^{k+2}}{3^n}, \frac{3^{k+3}}{3^n} \right] \right),$$

qui est une partie de $\mathbb{R}$. On peut aussi dire que l'on obtient l'ensemble de Cantor en prenant le segment $[0, 1]$, en le découpant en trois parts égales, en retirant le morceau du milieu, puis en réitérant l'opération sur les deux segments restants (ce qui conduit à retirer une infinité de segments); voir figure 10a. L'ensemble de Cantor est très étrange, il est indénombrable mais de mesure de Lebesgue nulle. De plus, on peut calculer facilement sa dimension de Hausdorff : on remarque en faisant un dessin (et on montre sans difficulté au besoin) que

$$\mathcal{H}^s(3\mathcal{C}) = 2\mathcal{H}^s(\mathcal{C}).$$

Or, par définition de la mesure de Hausdorff, $\mathcal{H}^s(3\mathcal{C}) = 3^s \mathcal{H}^s(\mathcal{C})$. Donc, si $\mathcal{H}^s(\mathcal{C}) \in \mathbb{R}_+^*$, alors $3^s = 2$, d'où

$$\dim_{\mathcal{H}}(\mathcal{C}) = \frac{\ln(2)}{\ln(3)} \simeq 0,6309.$$

Ce cas particulier se généralise à toute une classe d'objets à géométrie fractale sous la forme d'un théorème.

Théorème 267

Soit P une partie compacte de $\mathbb{R}^n$ telle que son image par une ho-

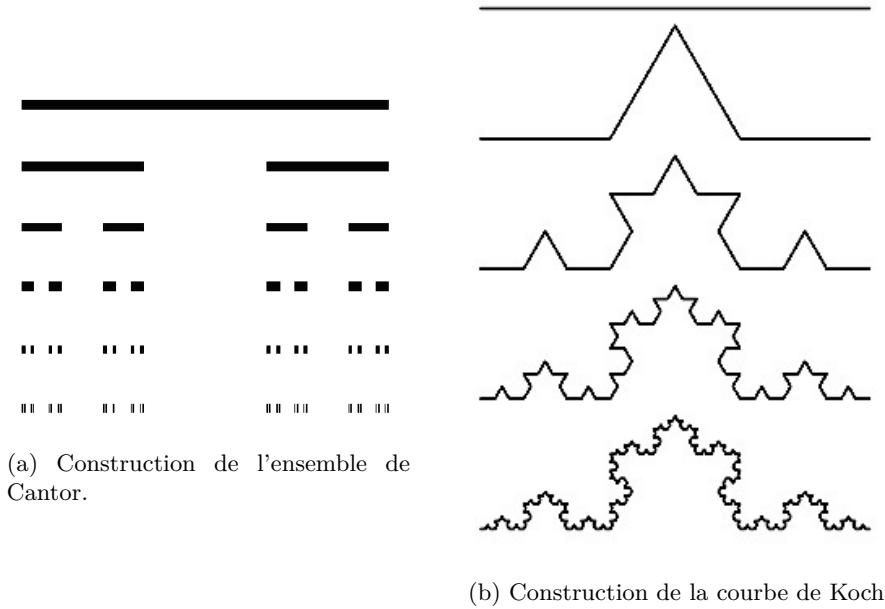


FIGURE 10 – Exemples de fractales

mothétique de rapport $r > 1$ est l'union d'un nombre N d'ensembles isométriques à P contenus dans des parallélépipèdes ayant des intersections de mesure de Lebesgue nulle. Alors,

$$\dim_{\mathcal{H}}(P) = \frac{\ln(N)}{\ln(r)}.$$

Ce résultat est très puissant et donne immédiatement la dimension de l'ensemble de Cantor. On trouve de même que la dimension fractale de la célèbre courbe de Koch (voir figure 10b) vaut

$$\dim_{\mathcal{H}}(\mathcal{K}) = \frac{\ln(4)}{\ln(3)}.$$

La dimension fractale de l'attracteur étrange de Lozentz est estimée à 2,06.

7. Synthèse et pratique

7.a. Synthèse de la section

7.b. Exercices

Exercice 268. Soit E un espace de Banach. Soit deux normes $\|\cdot\|_1$ et $\|\cdot\|_2$ telles que l'une est plus fine que l'autre (voir définition 31). Montrer que ces

deux normes sont alors équivalentes.

Exercice 269. *Quelle est la dimension fractale d'une variante de l'ensemble de Cantor obtenue en retirant successivement le neuvième central de chaque intervalle à partir du segment $[0, 1]$.*

Exercice 270. *Soit M, N, P trois variétés différentielles, $f \in \mathcal{C}^k(M, N)$ et $g \in \mathcal{C}^k(N, P)$. Montrer avec tous les détails que $f \circ g$ est de classe $\mathcal{C}^k$.*

B. INTRODUCTION À LA THÉORIE DES DISTRIBUTIONS

Nous nous basons principalement sur les livres [12] et [8].

1. Distributions de Schwartz

Définition 271 (Distribution). *On appelle distribution sur Ω toute forme linéaire T continue sur $\mathcal{C}_0^\infty(\Omega)$, ou de manière équivalente toute forme linéaire T sur $\mathcal{C}_0^\infty(\Omega)$ telle que pour tout compact K de Ω , il existe un entier m et une constante C tels que :*

$$\forall \phi \in \mathcal{C}_0^\infty(\Omega) \text{ nulle en dehors de } K, \quad |\langle T | \phi \rangle| \leq C \sup_{x \in K} \sup_{|\alpha| \leq m} |\partial^\alpha \phi(x)|.$$

On dit alors que la distribution est d'ordre fini si il existe $m \in \mathbb{N}$ qui vérifie la définition précédente pour tout compact K et on dit qu'elle est d'ordre infini sinon. L'espace des distributions sur Ω est noté $\mathcal{D}'(\Omega)$ et contient $L_{loc}^1(\Omega)$ l'espace des fonctions localement intégrables sur Ω , en identifiant fonction et distribution, *i.e*

$$\forall f \in L_{loc}^1(\Omega), \quad \forall \phi \in \mathcal{C}_0^\infty(\Omega), \quad \langle f | \phi \rangle = \int_{\Omega} f(x)\phi(x) dx.$$

Ce sont des distributions d'ordre 0.

Les distributions forment en fait le dual topologique de $\mathcal{D}'(\Omega) = \mathcal{C}_0^\infty(\Omega)$ pour la topologie que nous avons brièvement évoquée.

On généralise aux distributions la notion de dérivée en posant, pour tout $j \in \llbracket 1, n \rrbracket$ et toute distribution $T \in \mathcal{D}'(\Omega)$

$$\forall \phi \in \mathcal{C}_0^\infty(\Omega), \quad \langle \partial_{x_j} T | \phi \rangle := - \langle T | \partial_{x_j} \phi \rangle.$$

Le signe négatif provient de la formule d'intégration par partie pour les fonctions $\mathcal{C}^1(\Omega)$ qui sont en particulier des distributions avec l'identification précédente.

Propriété 272. *Pour tout indice j et toute distribution T , $\partial_{x_j} T$ ainsi définit est une distribution. De plus, elle est d'ordre au plus $m+1$ si T est d'ordre m .*

Corollaire 273. *Toute distribution est indéfiniment dérivable.*

Une autre opération courante est la multiplication par une fonction $\mathcal{C}^\infty$:

Propriété 274. *Pour toute distribution $T \in \mathcal{D}'(\Omega)$ et toute fonction f de classe $\mathcal{C}^\infty$, on définit une distribution que l'on note fT en posant :*

$$\forall \phi \in \mathcal{C}_0^\infty(\Omega), \quad \langle fT | \phi \rangle := \langle T | f\phi \rangle.$$

Remarque 275. *Si T est d'ordre m alors il suffit en fait que f soit de classe $\mathcal{C}^m$ pour pouvoir définir fT .*

La dérivation étant une opération linéaire, on retrouve des résultats similaires à ceux des fonctions :

$$\forall f \in \mathcal{C}^\infty(\Omega), \quad \forall T \in \mathcal{D}'(\Omega), \quad \partial(fT) = (\partial f)T + f(\partial T).$$

De plus, tout opérateur différentiel $P := \sum_\alpha a_\alpha \partial^\alpha$ se prolonge en une application linéaire de $\mathcal{D}'(\Omega)$ dans lui-même par la formule :

$$\forall T \in \mathcal{D}'(\Omega), \quad \forall \phi \in \mathcal{C}_0^\infty(\Omega), \quad \langle PT \mid \phi \rangle = \langle T \mid P^T \phi \rangle$$

où

$$\forall \phi \in \mathcal{C}_0^\infty(\Omega), \quad P^T \phi := \sum_\alpha (-1)^{|\alpha|} \partial^\alpha (a_\alpha \phi).$$

Enfin, on introduit une dernière opération sur les distributions, qui sert néanmoins moins fréquemment. Par soucis de lisibilité, dans toute la fin de ce paragraphe on notera T_f la distribution induite par une fonction mesurable f plutôt que f . Ainsi, on sera à même de distinguer T'_f qui désigne la dérivée de T_f dans l'espace des distributions et $T_{f'}$ qui est la distribution induite par f' .

Propriété 276. *Soit I un intervalle de $\mathbb{R}$ et $f \in L^1_{loc}(I)$. Alors les solutions de l'équation $T' = f$ dans $\mathcal{D}'(I)$ sont les distribution T_{F+C} où C est une constante et F la fonction continue définie par :*

$$\text{pour } a \in I \text{ fixé, } \quad \forall x \in I, \quad F(x) = \int_a^x f(t) dt.$$

En particulier, si u est dans l'espace de Sobolev $\mathcal{H}^1(I)$ et en notant u' sa dérivée faible alors $(T_u)' = T_{u'}$.

Corollaire 277. *Si I est un intervalle ouvert de $\mathbb{R}$ et $T \in \mathcal{C}_0^\infty(\Omega)$, alors $T' = \mathbf{0}$ si et seulement si T est constante, au sens où elle est associée à une fonction constante.*

Il existe des distributions qui ne sont pas des fonctions, comme par exemple la distribution δ dite de Dirac qui vérifie

$$\forall \phi \in \mathcal{C}_0^\infty(\Omega), \quad \langle \delta \mid \phi \rangle = \phi(0).$$

La dérivée au sens des distributions de la fonction de Heaviside $H := \mathbb{1}_{\mathbb{R}^+}$ vaut δ .

Enfin, le résultat suivant est souvent utile pour dériver au sens des distributions des fonctions discontinues :

Propriété 278 (Formule des sauts). *Soit $f: \mathbb{R} \rightarrow \mathbb{C}$ une fonction de classe $\mathcal{C}^1$ par morceaux, discontinues en $a_1, \dots, a_n$. Alors, si de plus $f \in L^1_{loc}(\mathbb{R})$,*

$$T'_f = T_{f'} + \sum_{k=1}^n [f(a_k^+) - f(a_k^-)] \delta_{a_k}$$

Comme restreindre les distributions revient naturellement à restreindre l'espace des fonctions tests, on introduit la définition suivante :

Définition 279 (Zéro d'une distribution). *Soit $T \in \mathcal{D}'(\Omega)$ et soit $U \subset \Omega$ un ouvert. On dit que T s'annule sur U si*

$$\forall \phi \in \mathcal{C}_0^\infty(U), \langle T | \phi \rangle = 0.$$

On voudrait alors définir, comme pour les fonctions, un support et un support compact, ce qui est rendu possible par la proposition suivante,

Propriété 280. *Soit $T \in \mathcal{D}'(\Omega)$. Si T s'annule sur chacun des ouverts de la collection $(U_j)_{j \in J}$, alors T s'annule sur $U = \cup_{j \in J} U_j$.*

Il existe donc pour toute distribution un plus grand ouvert de Ω sur lequel elle s'annule.

Définition 281 (Support d'une distribution). *On appelle support d'une distribution le complémentaire du plus grand ouvert sur lequel elle s'annule. Une distribution est dite à support compact si son support est compact. L'ensemble des distributions à support compact sur Ω sera noté $\mathcal{E}'(\Omega)$.*

2. Extension de la notion de convolution

2.a. Convolution d'une fonction avec une distribution

La convolution peut être efficacement étendue aux fonctions L^1 , mais l'extension qui nous intéresse est la suivante :

Définition 282 (Produit de convolution – Fonction & distribution). *Soit $T \in \mathcal{D}'(\Omega)$ et $\phi \in \mathcal{C}_0^\infty(\Omega)$. On appelle produit de convolution de T par ϕ , et on note $T * \phi$, la fonction*

$$T * \phi : y \mapsto \langle T | \phi(y - \cdot) \rangle$$

Propriété 283. *Soit $T \in \mathcal{D}'(\Omega)$ et $\phi \in \mathcal{C}_0^\infty(\Omega)$. La fonction $T * \phi$ est de classe $\mathcal{C}^\infty$ sur Ω et pour tout multi-indice α , on a*

$$\partial^\alpha (T * \phi) = (\partial^\alpha T) * \phi = T * (\partial^\alpha \phi).$$

Le fait que la convolution avec une fonction de $\mathcal{C}_0^\infty(\Omega)$ transforme un objet potentiellement très irrégulier comme une distribution en une fonction de classe $\mathcal{C}^\infty$ est une invitation à utiliser la convolution pour *régulariser*.

Théorème 284

Soit $\phi \in \mathcal{C}_0^\infty(\Omega)$ positive et d'intégrale 1. On pose pour tout $\varepsilon \in \mathbb{R}_+^*$,

$$\phi_\varepsilon(x) = \varepsilon^{-n} \phi\left(\frac{x}{\varepsilon}\right).$$

Alors, pour tout $T \in \mathcal{D}'(\Omega)$, si l'on pose $T_\varepsilon = T * \phi_\varepsilon$, on a

$$\forall \psi \in \mathcal{C}_0^\infty(\Omega), \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}^n} T_\varepsilon(x) \psi(x) dx = \langle T \mid \psi \rangle.$$

2.b. Convolution de deux distributions

Nous allons à présent étendre le produit de convolution au cas de deux distributions, mais, comme nous le constaterons, il y a alors une restriction. Pour plus de lisibilité nous noterons, pour toute fonction f , $\tilde{f}$ la fonction $x \mapsto f(-x)$. De même, pour une distribution $T \in \mathcal{D}'(\Omega)$, nous noterons $\tilde{T}$ la distribution telle que

$$\forall \phi \in \mathcal{C}_0^\infty(\Omega), \langle \tilde{T} \mid \phi \rangle = \langle T \mid \tilde{\phi} \rangle.$$

Définition 285 (Produit de convolution – Distribution & Distribution). *Soit $T \in \mathcal{D}'(\Omega)$ et $S \in \mathcal{E}'(\Omega)$, on appelle produit de convolution de T par S et on note $T * S$ la distribution qui vérifie*

$$\forall \phi \in \mathcal{C}_0^\infty(\Omega), \langle T * S \mid \phi \rangle = \langle T \mid \tilde{S} * \phi \rangle.$$

Remarque 286. *Du fait que l'on puisse étendre l'ensemble de définition d'une distribution à support compact à $\mathcal{C}^\infty(\Omega)$, écrire $\langle S \mid \tilde{T} * \phi \rangle$ a un sens pour $S \in \mathcal{E}'(\Omega)$, ce qui permet de définir le produit de convolution $S * T$ de S et de T .*

Théorème 287

Soit $T \in \mathcal{D}'(\Omega)$, $S \in \mathcal{E}'(\Omega)$ et α un multi-indice. On a

$$T * S = S * T,$$

ce qui revient à dire que la convolution des distribution est commutative, et

$$\partial^\alpha(T * S) = (\partial^\alpha T) * S = T * (\partial^\alpha S).$$

Théorème 288

Soit $T \in \mathcal{D}'(\Omega)$ et $R, S \in \mathcal{E}'(\Omega)$. La convolution de R , S et T est associative au sens où

$$R * (S * T) = (R * S) * T.$$

Remarque 289. *On peut donc définir la convolution $R * S * T$, et plus généralement pour $S_1, \dots, S_k \in \mathcal{E}'(\Omega)$ et $T \in \mathcal{D}'(\Omega)$, on peut écrire sans ambiguïté*

$$S_1 * \dots * S_k * T.$$

Il faut cependant se méfier de cette associativité, parce que l'hypothèse d'avoir au plus une distribution qui n'est pas à support compact est fondamentale, comme le montre l'exemple suivant :

Exemple 290. Soit H la fonction de Heaviside, δ la distribution de Dirac au point 0 et 1 la fonction constante égale à 1. On a $H, 1 \in \mathcal{D}'(\mathbb{R})$ et $\delta \in \mathcal{E}'(\mathbb{R})$. De plus δ' , la dérivée de δ , est également dans $\mathcal{E}'(\mathbb{R})$ et $H' = \delta$. Cependant,

$$(1 * \delta') * H = (1' * \delta) * H = 0 * H = 0,$$

et

$$1 * (\delta' * H) = 1 * (\delta * H') = 1 * (\delta * \delta) = 1 * \delta = 1.$$

La convolution a également un comportement agréable vis-à-vis du support :

Propriété 291. Soit $T \in \mathcal{D}'(\Omega)$ et $S \in \mathcal{E}'(\Omega)$. On a

$$\text{supp}(T * S) \subset \text{supp}(T) + \text{supp}(S),$$

où $\text{supp}(T) + \text{supp}(S) = \{x + y \mid x \in \text{supp}(T), y \in \text{supp}(S)\}$.

La convolution respecte également un autre type de support, le *support singulier*.

Définition 292 (Support singulier). Soit $T \in \mathcal{D}'(\Omega)$, on appelle support singulier de T et on note $\text{suppsing}(T)$ l'ensemble

$$\text{suppsing}(T) = \Omega \setminus \{x \in \Omega \mid \exists U \subset \Omega, \text{ouvert}, x \in U, T|_U \in \mathcal{C}^\infty(U)\}.$$

Le support singulier est le complémentaire du plus grand ouvert sur lequel la restriction de la distribution est une fonction indéfiniment différentiable.

Propriété 293. Soit $T \in \mathcal{D}'(\Omega)$ et $S \in \mathcal{E}'(\Omega)$. On a

$$\text{suppsing}(T * S) \subset \text{suppsing}(T) + \text{suppsing}(S).$$

3. Transformée de Fourier

En s'inspirant de ce qu'on a fait pour la dérivation et la convolution, on aimerait pouvoir définir la transformation de Fourier d'une distribution par une formule du type

$$\forall \phi \in \mathcal{C}_0^\infty(\Omega), \quad \langle \mathcal{F}T \mid \phi \rangle = \langle T \mid \mathcal{F}\phi \rangle$$

éventuellement avec un changement de signe. Mais une telle définition ne peut convenir car, en règle générale, $\mathcal{F}\phi \notin \mathcal{C}_0^\infty(\Omega)$. On va donc construire un espace que l'on note $\mathcal{S}(\Omega)$ et tel que $\phi \in \mathcal{S}(\Omega) \implies \mathcal{F}\phi \in \mathcal{S}(\Omega)$.

Remarque 294. L'espace $L^2(\mathbb{R})$ vérifie en particulier cette propriété, la transformée de Fourier étant une isométrie sur $L^2(\mathbb{R})$. Cependant, $L^2(\mathbb{R})$ n'est pas stable par passage à la dérivation, d'où la nécessité de se munir d'un espace plus restrictif.

3.a. Distributions tempérées

Définition 295 (Distribution tempérée). *Une forme linéaire $T: \mathcal{C}_0^\infty(\mathbb{R}^n) \rightarrow \mathbb{C}$ est une distribution tempérée s'il existe une constante $C > 0$ et un entier $m \in \mathbb{N}$ tels que :*

$$\forall \phi \in \mathcal{C}_0^\infty(\mathbb{R}^n), \quad |\langle T | \phi \rangle| \leq \sup \{ |N_{\alpha, \beta}(\phi)| \mid \alpha, \beta \in \mathbb{N}^n \text{ et } |\alpha|, |\beta| \leq m \}$$

Une telle appellation est notamment justifiée par la proposition qui suit.

Propriété 296. *Toute distribution tempérée est une distribution sur $\mathbb{R}^n$ et est, de plus, d'ordre fini.*

Communément, on définit plutôt les distributions tempérées *via* la proposition 297 et par analogie avec la notation $\mathcal{D}'$ et $\mathcal{E}'$, on note $\mathcal{S}'(\mathbb{R}^n)$ l'espace des distributions tempérées. De plus, $\mathcal{E}'(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n)$ et $\mathcal{S}(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n)$ en faisant agir $\mathcal{S}(\mathbb{R}^n)$ sur lui-même par la formule

$$\forall \phi, \psi \in \mathcal{S}(\mathbb{R}^n), \quad \langle \phi | \psi \rangle = \int_{\mathbb{R}^n} \phi(x) \psi(x) dx$$

En fait, on peut même montrer que $\mathcal{S}(\mathbb{R}^n)$ est dense dans $\mathcal{S}'(\mathbb{R}^n)$.

Propriété 297. *Une forme linéaire $T: \mathcal{C}_0^\infty(\mathbb{R}^n) \rightarrow \mathbb{C}$ est une distribution tempérée si et seulement si elle se prolonge en une forme linéaire continue $\tilde{T}: \mathcal{S}(\mathbb{R}^n) \rightarrow \mathbb{C}$.*

Corollaire 298. *La dérivée de toute distribution tempérée est encore une distribution tempérée.*

Une classe d'exemples de distributions tempérées est donnée par les fonctions à croissance lente :

Définition 299 (Fonction à croissance lente). *Une fonction $\phi \in \mathcal{C}^\infty(\mathbb{R}^n)$ est dite à croissance lente si*

$$\forall \alpha \in \mathbb{N}^n, \exists m \in \mathbb{N}, \exists C > 0, \quad \forall x \in \mathbb{R}^n, \quad |\partial^\alpha \phi(x)| \leq C (1 + |x|)^m$$

Exemple 300. *Les fonctions polynomiales sont $\mathcal{C}^\infty$ à croissance lente, de même pour les fonctions $L^p(\mathbb{R}^n)$ avec $1 \leq p \leq +\infty$.*

Propriété 301. *Toute fonction à croissance lente définit une distribution tempérée.*

3.b. Transformée de Fourier d'une distribution

D'après le théorème de Fubini, on a

$$\forall \phi, \psi \in \mathcal{S}(\mathbb{R}^n), \quad \int_{\mathbb{R}^n} \hat{\phi}(\xi) \psi(\xi) d\xi = \int_{\mathbb{R}^n} \phi(\xi) \hat{\psi}(\xi) d\xi$$

et on définit donc la transformée de Fourier sur $\mathcal{S}'(\mathbb{R}^n)$ en prolongeant, de manière continue, l'application $\mathcal{F}: \mathcal{S}(\mathbb{R}^n) \rightarrow \mathcal{S}'(\mathbb{R}^n)$ à $\mathcal{F}: \mathcal{S}'(\mathbb{R}^n) \rightarrow \mathcal{S}'(\mathbb{R}^n)$.

Définition 302 (Transformée de Fourier). *Pour toute distribution tempérée T , on définit sa transformée de Fourier $\mathcal{F}T$ comme la distribution tempérée définie par*

$$\forall \phi \in \mathcal{S}(\mathbb{R}^n), \quad \langle \mathcal{F}T \mid \phi \rangle := \langle T \mid \mathcal{F}\phi \rangle$$

On peut vérifier que $\mathcal{F}T$ ainsi définie est bien une distribution tempérée.

Commençons par remarquer que $\hat{\delta} = \mathbf{1}$ et donc, d'après 108.iv., $\hat{\mathbf{1}} = c\delta$ pour un certain $c \in \mathbb{C}$. D'où, pour toute fonction $\phi \in \mathcal{S}(\mathbb{R}^n)$ de l'espace de Schwartz :

$$\hat{\phi}(0) = \langle \delta \mid \hat{\phi} \rangle = \langle \mathbf{1} \mid \hat{\phi} \rangle = \langle \hat{\mathbf{1}} \mid \phi \rangle = \langle c\delta \mid \phi \rangle = c\phi(0)$$

en effectuant le calcul dans un cas particulier – typiquement la gaussienne $\phi(x) = e^{-|x|^2/2}$ – on montre que $c = (2\pi)^n$ et avec 108.iv. il vient que :

$$\forall x \in \mathbb{R}^n, \quad \hat{\phi}(x) = (2\pi)^n \phi(-x)$$

Autrement dit :

$$\forall x \in \mathbb{R}^n, \quad \phi(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ix \cdot \xi} \hat{\phi}(\xi) d\xi := \check{\mathcal{F}}\phi(x)$$

et on a en fait montré la formule d'inversion de Fourier sur $\mathcal{S}(\mathbb{R}^n)$.

Théorème 303 (Formule d'inversion de Fourier)

La transformation de Fourier $\mathcal{F}$ est un isomorphisme de $\mathcal{S}(\mathbb{R}^n)$ sur $\mathcal{S}(\mathbb{R}^n)$ et également de $\mathcal{S}'(\mathbb{R}^n)$ sur $\mathcal{S}'(\mathbb{R}^n)$. Son inverse est donnée par $(2\pi)^{-n} \check{\mathcal{F}}$ où $\check{\mathcal{F}}$ est la composée de $\mathcal{F}$ et de la relation symétrie $T \mapsto \check{T}$ définie sur $\mathcal{S}'(\mathbb{R}^n)$ par :

$$\forall \phi \in \mathcal{S}(\mathbb{R}^n), \quad \langle \check{T} \mid \phi \rangle = \langle T \mid \check{\phi} \rangle$$

Autrement dit, pour toute distribution tempérée $T \in \mathcal{S}'(\mathbb{R}^n)$:

$$\forall \phi \in \mathcal{S}(\mathbb{R}^n), \quad \langle \check{\mathcal{F}}T \mid \phi \rangle = \langle T \mid \check{\mathcal{F}}\phi \rangle$$

Enfin, on conclut en énonçant les résultats suivants à mettre en parallèle avec le point iii. de la proposition 108 :

- i. Pour toutes distributions $T \in \mathcal{S}'(\mathbb{R}^n)$ et $R \in \mathcal{E}'(\mathbb{R}^n)$: $\mathcal{F}(T \star R) = (\mathcal{F}T) \times (\mathcal{F}R)$;
- ii. Pour toutes distributions T_1 et $T_2 \in \mathcal{S}'(\mathbb{R}^n)$: $\mathcal{F}(T_1 \times T_2) = (2\pi)^{-n} \mathcal{F}T_1 \star \mathcal{F}T_2$.

et le théorème suivant (dit de Plancherel). Pour cela, on munit $L^2(\mathbb{R}^n)$ de son produit scalaire usuel :

$$\forall f, g \in L^2(\mathbb{R}^n), \quad (f, g) := \int_{\mathbb{R}^n} f(x) \bar{g}(x) dx$$

Théorème 304 (Théorème de Plancherel)

Pour toute fonction $f, g \in L^2(\mathbb{R}^n)$,

$$(\hat{f}, \hat{g}) = (2\pi)^n (f, g)$$

Autrement dit, $\mathcal{F}$ se prolonge en un automorphisme de $L^2(\mathbb{R}^n)$ tel que $(2\pi)^{-n/2} \mathcal{F}$ soit unitaire.

4. Masse de Dirac et solutions fondamentales

Nous avons déjà défini et utilisé dans les exemple la distribution de Dirac δ , qui est d'ordre 0 et à support compact, de support et de support singulier réduits à $\{0\}$, et définie par la formule

$$\forall \phi \in \mathcal{C}_0^\infty(\mathbb{R}^n), \langle \delta \mid \phi \rangle = \phi(0)$$

Nous allons voir qu'elle est utile pour résoudre des équations aux dérivées partielles. Pour cela, nous avons besoin des propriétés suivantes :

Propriété 305. Soit $T \in \mathcal{D}'(\mathbb{R}^n)$. On a

$$T * \delta = \delta * T = T,$$

et, comme nous l'avons déjà remarqué,

Propriété 306. La distribution de Dirac est tempérée et elle vérifie,

$$\hat{\delta} = \mathbf{1} \quad \text{et} \quad \hat{\mathbf{1}} = (2\pi)^n \delta.$$

Démonstration : La distribution de Dirac est à support compact, donc elle est tempérée.

Pour toute fonction $f \in \mathcal{S}(\mathbb{R}^n)$,

$$\begin{aligned} \langle \hat{\delta} \mid f \rangle &= \langle \delta \mid \hat{f} \rangle \\ &= \hat{f}(0) \\ &= \int_{\mathbb{R}^n} f(x) dx \\ &= \langle \mathbf{1} \mid f \rangle. \end{aligned}$$

Donc $\hat{\delta} = \mathbf{1}$.

Pour toute fonction $f \in \mathcal{S}(\mathbb{R}^n)$,

$$\begin{aligned} \langle \hat{\mathbf{1}} \mid f \rangle &= \langle \mathbf{1} \mid \hat{f} \rangle \\ &= \int_{\mathbb{R}^n} \hat{f}(x) dx \\ &= (2\pi)^n f(0) \quad \text{par la formule de transformée inverse} \\ &= (2\pi)^n \langle \delta \mid f \rangle. \end{aligned}$$

Donc $\hat{\mathbf{1}} = (2\pi)^n \delta$. ■

On constate que la distribution la "plus concentrée" qui soit est envoyée par la transformée de Fourier sur la fonction constante égale à 1, c'est-à-dire la "plus étalée" que l'on puisse imaginer. Cet exemple traduit un fait plus général : il est impossible de concentrer à la fois la masse d'une distribution et celle de sa transformée de Fourier. C'est cette idée qui conduit en physique quantique au principe d'incertitude de Heisenberg-Gabor.

Définition 307 (Solution fondamentale, Fonction de Green). *Soit P un opérateur différentiel à coefficients constants sur $\mathbb{R}^n$. Une solution fondamentale (aussi appelée solution élémentaire ou fonction de Green) de P est une distribution $E \in \mathcal{D}'(\mathbb{R}^n)$ telle que*

$$PE = \delta$$

Exemple 308. *L'opérateur différentiel*

$$P = \frac{\partial}{\partial t} - \Delta,$$

associé à l'équation de la chaleur admet pour solution fondamentale sur $\mathbb{R}_+ \times \mathbb{R}^n$ la fonction

$$E : (t, x) \mapsto H(t) \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{\|x\|^2}{4t}},$$

où H est toujours la fonction de Heaviside.

Théorème 309

Soit P un opérateur différentiel à coefficients constants qui admet une solution fondamentale E . Soit $S \in \mathcal{E}'(\mathbb{R}^n)$. La distribution $E * S$ est solution de l'équation

$$PT = S$$

Démonstration : Posons $T = E * S$. E est solution fondamentale, donc $PE = \delta$.

On a alors,

$$PT = P(E * S) = (PE) * S = \delta * S = S,$$

d'après les propriétés de la convolution et de la distribution de Dirac. ■

Définition 310 (Opérateur différentiel hypoelliptique). *Un opérateur différentiel P est dit hypoelliptique si,*

$$\forall T \in \mathcal{D}'(\Omega), \text{suppsing}(PT) = \text{suppsing}(T)$$

Théorème 311

Soit P un opérateur différentiel à coefficients constants. Si P admet une solution fondamentale E telle que

$$\text{suppsing}(E) = \{0\},$$

c'est-à-dire si E est de classe $\mathcal{C}^\infty$ en dehors de l'origine, alors P est hypoelliptique.

Propriété 312. Soit P un opérateur hypoelliptique sur Ω . Pour tout ouvert $U \subset \Omega$, pour toute distribution $T \in \mathcal{D}'(U)$, on a l'implication

$$PT \in \mathcal{C}_0^\infty(U) \implies T \in \mathcal{C}_0^\infty(U)$$

Corollaire 313. Les distributions harmoniques (vérifiant $\Delta T = 0$) sont des fonctions de classe $\mathcal{C}^\infty$. De même, les distributions holomorphes (vérifiant $\bar{\partial}T = 0$) sur un ouvert sont les fonctions holomorphes sur ce même ouvert.

C. LES SEMI-GROUPES D'OPÉRATEURS

1. Semi-groupes et générateur infinitésimal

Définition 314 (Semi-groupe d'opérateurs). *Soit E un espace de Banach. On appelle semi-groupe d'opérateurs sur E (ou encore semi-groupe fortement continu d'opérateurs linéaires sur E) une application S de $\mathbb{R}_+$ dans $\mathcal{L}(E)$ vérifiant*

- *l'image $S(0)$ de 0 est l'application identité sur E :*

$$S(0) = I_E ; \quad (21)$$

- *on a*

$$\forall t_1, t_2 \in \mathbb{R}_+, \quad S(t_1 + t_2) = S(t_1)S(t_2) ; \quad (22)$$

- *pour tout $u \in E$, l'application*

$$t \mapsto S(t)u \quad (23)$$

définie sur $\mathbb{R}_+$ et à valeurs dans E est continue.

Remarque 315. (Importante) *La définition de semi-groupe n'implique pas que l'application S est continue de $\mathbb{R}_+$ dans $\mathcal{L}(E)$. Elle ne l'est d'ailleurs pas en général et les semi-groupes sont, en quelque sorte, surtout utiles dans le cas où elle ne l'est pas.*

Le théorème suivant caractérise de tels opérateurs, qui demeurent un exemple important car intuitif.

Théorème 316

Soit $S : \mathbb{R}_+ \rightarrow \mathcal{L}(E)$ un semi-groupe d'opérateurs sur un espace de Banach E . Si S est continu alors il existe $A \in \mathcal{L}(E)$ tel que

$$\forall t \in \mathbb{R}_+, \quad S(t) = e^{tA} := \sum_{n=0}^{+\infty} \frac{(tA)^n}{n!}.$$

Définition 317 (Générateur infinitésimal). *Soit $S : \mathbb{R}_+ \rightarrow \mathcal{L}(E)$ un semi-groupe d'opérateurs sur un espace de Banach E . Notons $D(A)$ l'ensemble des éléments $u \in E$ tels que l'application $t \mapsto S(t)u$ est dérivable à droite en 0. On appelle générateur infinitésimal de S l'application*

$$\begin{aligned} A : D(A) &\rightarrow E \\ u &\mapsto -\frac{d}{dt}(S(t)u)_{|t=0^+}. \end{aligned}$$

On notera que $D(A)$, le domaine de A , est de manière évidente un sous-espace vectoriel de E .

2. Le théorème de Hille-Yosida

3. Applications

D. APPROCHE GÉNÉRALE DES SYSTÈMES DYNAMIQUES

Il est possible d'avoir une approche plus vaste des systèmes dynamiques que la seule considération d'équations différentielles sur un ouvert de $\mathbb{R}^n$. Ce que nous appelons maintenant système dynamique est la donnée d'un espace (un ensemble muni d'une ou plusieurs structures) qui représente les états du système, d'un paramètre temporel, continu ou discret, réversible ou irréversible, et d'une loi d'évolution dans l'espace des états.

Les espaces d'états seront des espaces topologiques, parfois des ouverts d'espaces vectoriel ou des variétés différentielles. On les munira au besoin de métriques, de tribus et de mesures.

Nous considérons ici uniquement des systèmes sans mémoire et déterministes et les lois d'évolutions seront des équations différentielles ou des fonctions qui agissent de façon discrète.

1. Systèmes continus sur une variété différentielle

Il peut être intéressant de considérer des systèmes dynamiques sur d'autres objets que $\mathbb{R}^n$, par exemple des espaces topologiques possédant une structure de variété différentielle. D'une part, de nombreux objets réels, les formes géométriques en particulier, se modélisent bien par des variétés différentielles et définir directement les systèmes dynamiques sur ces objets permet de prendre en compte les contraintes géométriques de manière plus naturelle. D'autre part, les variétés compactes et sans bord, comme les tores, sont le lieu idéal pour le développement d'une théorie du chaos basée sur le calcul de l'entropie. Faute de place et de temps, nous ne démontrerons pas les généralisations aux variétés des résultats que nous avons démontrés dans $\mathbb{R}^n$ – la plupart des adaptations de preuves sont purement formelles, demandant de passer par les cartes de l'atlas différentiable associé à la variété.

1.a. Champs de vecteurs

Considérons deux variétés différentielles M et N de dimensions respectives n_M et n_N et de classe $\mathcal{C}^k$.

Maintenant que nous avons des bases de calcul différentiel, nous avons besoin de champs de vecteurs. Le problème, c'est que les vecteurs vivent sur les espaces tangents. Nous avons besoin d'un nouvel objet sur lequel faire vivre les champs de vecteurs sur les variétés.

Définition 318 (Fibré tangent). *On appelle fibré tangent et on note TM l'union disjointe des espaces tangents à M en chacun de ses points.*

On peut montrer que le fibré tangent admet une structure de variété différentielle de dimension $2n_M$ et de classe $\mathcal{C}^{k-1}$ (il y a perte de régularité). Il

admet également une structure de fibré vectoriel pour la projection canonique qui envoie chaque espace tangent sur le point dont il est issu :

Proposition 319. *La projection canonique p de TM sur M est une fibration.*

Démonstration : Soit $(U_i, \phi_i)_{i \in I}$ un atlas de M . On définit, pour tout i de I ,

$$\begin{aligned} \psi_i : \quad p^{-1}(U_i) = TU_i &\rightarrow U_i \times \mathbb{R}^{n_M} \\ \xi &\mapsto (m, T_m \phi_i \cdot \xi) \quad \text{où } m \text{ est tel que } \xi \in T_m M \end{aligned}$$

Pour tout i de I , ψ_i est clairement un difféomorphisme et : $\forall m \in U_i, \forall v \in \mathbb{R}^{n_M}$,

$$p(\psi_i^{-1}(m, v)) = p(\xi) = m \quad \text{où } \xi \in T_m M \text{ tel que } v = T_m \phi_i \cdot \xi$$

et donc (cf annexe A., partie sur les fibrés) TM est espace fibré, de base M et de fibre $\mathbb{R}^{n_M}$ ■

En fait, le fibré tangent est même un fibré vectoriel réel de rang n_M .

On peut définir des sections pour cette projection, d'où la définition qui nous intéresse pour parler de systèmes dynamiques :

Définition 320 (Champ de vecteurs). *On appelle champ de vecteurs sur une variété une section de classe $\mathcal{C}^\infty$ de son fibré tangent pour la projection canonique. On note $\mathcal{C}^\infty(TM)$ l'ensemble des champs de vecteurs de la variété M .*

L'ensemble $\mathcal{C}^\infty(TM)$ peut également être muni d'une structure d'espace vectoriel et permet la plupart des opérations possibles sur les champs de vecteurs de $\mathbb{R}^n$.

1.b. Flot d'un champ de vecteurs

On notera $X : y \mapsto X_y$ les champs de vecteurs dans cette sous-section.

Définition 321 (Courbe intégrale). *On appelle courbe intégrale d'un champ de vecteurs X sur une variété M toute courbe $c : I \rightarrow M$, I intervalle ouvert de $\mathbb{R}$, de classe $\mathcal{C}^1$, telle que*

$$\forall t \in I, \quad c'(t) = X_{c(t)}.$$

On a pour ces courbes intégrales un équivalent du théorème de Cauchy-Lipschitz.

Théorème 322

Soit X un champ de vecteurs sur une variété M et soit $x \in M$. Il existe un intervalle ouvert I contenant 0 et une courbe intégrale $c : I \rightarrow M$ de X tels que $c(0) = x$. Si $\tilde{c} : J \rightarrow M$ est une courbe intégrale de X telle que $\tilde{c}(0) = x$, c et $\tilde{c}$ coïncident sur $I \cap J$.

Si l'on note alors $c_x(t)$ la valeur en t de la courbe intégrale qui passe par x en 0, on peut poser

$$\phi_t^X(x) = c_x(t).$$

Définition 323 (Flot). *On appelle flot du champ de vecteurs X sur la variété M l'application $\phi^X : \Omega \rightarrow M$, où Ω est l'ensemble des couples (t, x) tels que $c_x(t)$ existe.*

Sous réserve d'existence des quantités, on obtient aisément que

$$\phi_{t_1+t_2}^X(x) = \phi_{t_1}^X(\phi_{t_2}^X(x)) = \phi_{t_2}^X(\phi_{t_1}^X(x)).$$

On montre que le flot est continu sur son ensemble de définition et que les ϕ_t^X sont des difféomorphismes sur leurs images. On omettra le X en exposant s'il n'y a pas ambiguïté sur le champ de vecteurs. Le flot d'un champ de vecteurs est un groupe local à un paramètre de difféomorphismes et le champ de vecteurs est son générateur infinitésimal.

Le principal intérêt de considérer un système dynamique sur une variété est de choisir une variété compacte, car alors il n'y a plus de questions à se poser sur les temps d'existence. En effet :

Théorème 324

Soit X un champ de vecteurs sur une variété M compacte. Alors le flot de X est défini sur $\mathbb{R} \times M$ et pour tout $t \in \mathbb{R}$, ϕ_t^X est un difféomorphisme de M .

Démonstration : Soit $\Omega \subset \mathbb{R} \times M$ le domaine de définition du flot de X . On a $\{0\} \times M \subset \Omega$. Pour tout x on note $I_x = \Omega \cap (\mathbb{R} \times \{x\})$ le temps de vie de la courbe intégrale maximale c_x .

Soit $x \in M$; I_x est un intervalle. Supposons $I_x =]a, b[$ avec $b < +\infty$.

Soit $(t_n)_{n \in \mathbb{N}}$ une suite qui tend vers b . Alors la suite $(c_x(t_n))_{n \in \mathbb{N}}$ admet une valeur d'adhérence y .

Or, il existe $\varepsilon \in \mathbb{R}_+^*$ et U un ouvert de M contenant y tels que $]\varepsilon, \varepsilon[\times U \subset \Omega$ et il existe n_0 tel que $|t_n - b| < \varepsilon$ et $(t_n, c_x(t_n)) \in]-\varepsilon, \varepsilon[\times U \subset \omega$.

Or, la courbe intégrale $c_{c_x(t_n)}$ est définie sur $]-\varepsilon, \varepsilon[$ par définition de Ω , mais on a

$$c_{c_x(t_n)}(t) = c_x(t + t_n),$$

pour tout t tel que $t_n + t \in I_x$ et $t \in I_{c_x(t_n)}$. On peut donc prolonger c_x en posant

$$\tilde{c}_x(t) = \begin{cases} c_x(t) & \text{si } t \in I_x \\ c_{c_x(t_n)}(t - t_n) & \text{si } t_n \leq t < t_n + \varepsilon \end{cases},$$

et $\tilde{c}$ est également une courbe intégrale passant par x en 0, ce qui contredit la maximalité de c .

Le même raisonnement fonctionne si $-\infty < a$.

On a donc pour tout $x \in M$, $I_x = \mathbb{R}$ et,

$$\forall t_1, t_2 \in \mathbb{R}, \phi_{t_1+t_2}^X = \phi_{t_1}^X \circ \phi_{t_2}^X,$$

ce dont on déduit que ϕ_t^X est pour tout $t \in \mathbb{R}$ un difféomorphisme d'inverse ϕ_{-t}^X . ■

Le résultat reste vrai pour des champs de vecteurs à support compact sur une variété non-compacte.

On pourrait être chagriné de constater que nous n'avons jusque là parlé que d'équations différentielles autonomes, mais la situation se règle avec un simple jeu d'écriture.

Définition 325 (Champ de vecteur). *Si M est une variété de fibré tangent TM associé à la projection canonique p et I un intervalle ouvert de $\mathbb{R}$, on appelle champ de vecteurs dépendant du temps sur M une application X de classe $\mathcal{C}^k$ de $I \times M$ dans TM telle que*

$$\forall t \in I, \forall x \in M, p(X(t, x)) = x.$$

On peut associer à tout champ de vecteur X dépendant du temps sur une variété M un champ de vecteur autonome $\tilde{X}$ sur la variété $I \times M$ défini par

$$\tilde{X}_{(t,x)} = (1, X(t, x)),$$

ce qui permet d'adapter les théorèmes précédents aux champs de vecteurs non-autonomes. Les courbes intégrales de X sont donc, par projection, les courbes qui vérifient pour un sous-intervalle J de I ,

$$\forall t \in J, c'(t) = X(t, c(t)) := X_{c(t)}(t).$$

En particulier, les champs de vecteurs dépendant du temps sur une variété compacte ou les champs de vecteurs dépendant du temps à support compact sur une variété quelconque ont toujours des courbes intégrales définies sur I tout entier, quoique l'on perde cependant la structure de groupe local à un paramètre du flot.

2. Systèmes dynamiques discrets

On peut définir sur toute variété M de dimension n , et donc en particulier sur $\mathbb{R}^n$ muni de sa topologie naturelle, des systèmes dynamiques discrets en fournissant un point de départ et une fonction d'itération qui le déplace à chaque étape. On remarquera pour commencer que tout système dynamique continu dont les trajectoires existent pour tout temps définit de manière naturelle un système dynamique discret en faisant agir le flot sur le point de départ pendant une durée constante à chaque étape. Quoique ce lien soit fondamental et permette l'irrigation réciproque de ces deux aspects de la dynamique, des comportements nouveaux et parfois contre-intuitifs émergent dans les dynamiques discrètes, invalidant par exemple le théorème de Poincaré-Bendixson en dimension deux.

Donnons sans attendre une définition satisfaisante de la bête.

Définition 326 (Système dynamique discret). *On appelle système dynamique discret sur la variété différentielle M , étant donné un point $x_0 \in M$ et une application bijective et continue $f : M \rightarrow M$, la suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de M définie par la formule de récurrence*

$$x_{n+1} = f(x_n),$$

ou écrit autrement

$$x_n = f^n(x_0).$$

3. Systèmes topologiquement transitifs

Définition 327. *Soit M une variété différentielle. Soit un système dynamique défini par un champ de vecteur X ou une application d'itération f .*

Le système est dit topologiquement transitif si il existe une orbite $(\{\phi_t^X(x_0) \mid t \in I_{max}\})$ ou $(f(x_0))_{n \in \mathbb{N}}$ selon le cas) dense dans M .

Le système est dit minimal si toutes les orbites sont denses dans M .

Théorème 328

Soit (M, f) un système dynamique discret sur une variété différentielle M . Il est topologiquement transitif si et seulement si pour tous ouverts U, V de M , il existe $n_0 \in \mathbb{N}$ tel que $f^{n_0}(U) \cap V \neq \emptyset$.

Corollaire 329. *Soit (M, f) un système dynamique discret sur une variété différentielle M . Il est topologiquement transitif si et seulement si il n'existe pas deux ouverts U et V de M disjoints ($U \cap V = \emptyset$) et invariants par f ($f(U) = U$ et $f(V) = V$).*

Corollaire 330. *Soit (M, f) un système dynamique discret sur une variété différentielle M . S'il est topologiquement transitif alors il n'existe aucune fonction continue $\phi : M \rightarrow \mathbb{R}$ non constante et invariante par f ($\phi \circ f = \phi$)*

Théorème 331

Soit $\mathbb{T}^n$ le tore plat muni de sa topologie naturelle. Soit $f : \mathbb{T}^n \rightarrow \mathbb{T}^n$ définie par $f(x_1, \dots, x_n) = (x_1 + \omega_1, \dots, x_n + \omega_n)$. Le système dynamique discret $(\mathbb{T}^n, f)$ est minimal si et seulement si les nombres réels $\omega_1, \dots, \omega_n, 1$ sont rationnellement indépendants.

Théorème 332

Soit $\mathbb{T}^n$ le tore plat muni de sa topologie naturelle et du champ de vecteurs

$$X = \omega_1 \frac{\partial}{\partial x_1} + \dots + \omega_n \frac{\partial}{\partial x_n}.$$

Le système dynamique continu $(\mathbb{T}^n, \phi_t^X)$ est minimal si et seulement si les nombres réels $\omega_1, \dots, \omega_n$ sont rationnellement indépendants.

4. Moyenne temporelle d'un système dynamique

L'idée de moyenne périodique d'une équation différentielle a pris son essor au XVIII^e siècle, quand les scientifiques européens se sont penchés sur la théorie de la gravitation de Newton. Le modèle à deux corps donnait une approximation raisonnable du mouvement réel des corps célestes, mais il était clair, même avec les instruments de l'époque, que considérer chaque corps comme isolé avec le soleil dans un système indépendant était une approximation trop grossière. Les équations devenant inextricables dès que l'on ajoutait toutes les variables pertinentes (actions des grosses planètes comme Jupiter, de la lune, résistance de l'usine à gaz de l'époque qu'était l'éther), il fallut trouver des approximations, qui à l'époque étaient dispensées généreusement et sans aucune théorie mathématique solide. On trouvait cependant des méthodes très élaborées, et validées aujourd'hui par la théorie, en particulier dans les travaux de Lagrange et Laplace (Voir par exemple le *Traité de Mécanique Céleste* de Laplace).

L'idée principale est qu'une petite perturbation périodique donne presque le même résultat qu'un système similaire qui subirait à la place la moyenne sur une période de cette petite perturbation. On peut imaginer par exemple un pendule oscillant dans le vide subissant un champ magnétique constant, périodique en temps et de faible amplitude. Sa trajectoire serait semblable à celle du même pendule soumis à la moyenne temporelle de ce champ magnétique.

Il a fallu attendre le XX^e siècle avant que soient formalisées et rigoureusement étudiées les techniques empiriques de moyennes périodiques. La première contribution dans le domaine vient de Fatou, en 1928 – soit un an avant sa mort. Bogoliubov et Mitropolsky se sont ensuite penché sur la question et ont étendu certains résultats à des champs de vecteurs non périodiques. Depuis, ce domaine de recherche a connu un essor considérable et s'est ramifié en de nombreuses branches mathématiques. Nous présentons ici une infime partie de ce qui existe sur le sujet. Nous nous limiterons en particulier à des approximations du premier ordre.

4.a. Cas d'un système périodique

Posons le système dynamique

$$\dot{x} = \varepsilon f(t, x) + \varepsilon^2 g(t, x, \varepsilon), \quad x(0) = x_0 \quad (\mathfrak{A})$$

où $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ et $g : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}_+ \rightarrow \mathbb{R}^n$ sont continues et T-périodiques par rapport à la variable temporelle.

Posons

$$\begin{aligned} \bar{f} : \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ z &\mapsto \frac{1}{T} \int_0^T f(t, z) dt, \end{aligned}$$

et définissons un deuxième système dynamique

$$\dot{z} = \varepsilon \bar{f}(z), \quad z(0) = x_0 \quad (\Omega)$$

Nous souhaitons donner un argument mathématiques pour justifier que les solutions de $(\mathcal{R})$ et (Ω) sont très semblables dès que ε est assez petit.

Nous avons besoin au préalable d'un lemme technique qui est une sorte de descendant du lemme de Gronwall (théorème 150) au sens où leurs preuves comportent de très fortes similitudes et font appel aux mêmes types d'arguments. Cependant, il nous est plus facile ici de démontrer le lemme en partant de rien que de le déduire du lemme de Gronwall.

Lemme 333 (Variante de Gronwall). *Soit ϕ une fonction continue sur $[t_0, t_0 + T]$, et soit $a \in \mathbb{R}_+^*$ et $b, c \in \mathbb{R}_+$. Supposons que pour tout $t \in [t_0, t_0 + T]$,*

$$\phi(t) \leq a \int_{t_0}^t \phi(s) \, ds + b(t - t_0) + c.$$

Alors,

$$\forall t \in [t_0, t_0 + T], \quad \phi(t) \leq \left(\frac{b}{a} + c\right) e^{a(t-t_0)} - \frac{b}{a}$$

Démonstration :

- Soit ψ une fonction continue sur $[t_0, t_0 + T]$. Soit $\alpha \in \mathbb{R}_+^*$, $\beta \in \mathbb{R}_+$. Supposons que

$$\forall t \in [t_0, t_0 + T], \quad \psi(t) \leq \alpha \int_{t_0}^t \psi(s) \, ds + \beta.$$

Alors si l'on pose

$$\Gamma(t) = \alpha \int_{t_0}^t \psi(s) \, ds + \beta,$$

on a $\psi(t) - \Gamma(t) \leq 0$ et $\Gamma'(t) = \alpha\psi(t)$, donc de même que dans la preuve du lemme de Gronwall, on peut multiplier par $e^{-a(t-t_0)}$ et reconnaître une dérivée,

$$\frac{d}{dt} \left(\Gamma(t) e^{-a(t-t_0)} \right) \leq 0,$$

soit en intégrant,

$$\Gamma(t) e^{-a(t-t_0)} - \beta \leq 0.$$

On a donc

$$\psi(t) \leq \Gamma(t) \leq \beta e^{-a(t-t_0)}.$$

- Nous allons démontrer notre lemme comme cas particulier de l'inégalité que nous venons d'obtenir. On peut réécrire l'hypothèse sous la forme

$$\phi(t) \leq a \int_{t_0}^t \phi(s) \, ds + b \int_{t_0}^t 1 \, ds + c,$$

soit en réarrangeant les termes,

$$\frac{b}{a} + \phi(t) \leq a \int_{t_0}^t \left(\frac{b}{a} + \phi(s) \right) ds + \frac{b}{a} + c.$$

On obtient le résultat désiré en prenant dans l'inégalité précédemment obtenue $\psi = \frac{b}{a} + \phi$, $\alpha = a$ et $\beta = \frac{b}{a} + c$. ■

Théorème 334 (*Moyenne périodique*)

Supposons f continue et lipschitzienne par rapport à la deuxième variable et g continue. Alors il existe $\varepsilon_0, L, C \in \mathbb{R}_+^*$ tel que

$$\forall \varepsilon \in [0, \varepsilon_0], \forall t \in [0, \frac{L}{\varepsilon}], \quad \|x(t, \varepsilon) - z(t, \varepsilon)\| < C\varepsilon,$$

Démonstration :

- Contrôle des solutions :

Soit $\varepsilon_0, T_0 \in \mathbb{R}_+^*$ et D un ouvert borné (donc d'adhérence compacte puisque nous sommes en dimension finie) contenant x_0 . Pour tout $\varepsilon \in [0, \varepsilon_0]$, les fonctions f et g sont continues et périodiques en temps, donc bornées sur $\mathbb{R} \times \bar{D}$ et $\mathbb{R} \times \bar{D} \times [0, \varepsilon_0]$ par M_1 et M_2 respectivement. $\bar{f}$ est continue donc bornée par M_3 sur $\bar{D}$. On en déduit,

$$\forall t \in \mathbb{R}, \forall x \in \bar{D}, \left(\varepsilon \left(f(t, x) + \varepsilon g(t, x) \right) \leq \varepsilon(M_1 + \varepsilon M_2) \wedge \varepsilon \bar{f}(x) \leq \varepsilon M_3 \right).$$

D'après une adaptation évidente de la démonstration du théorème 139, il existe un rayon $r_0 \in \mathbb{R}_+^*$ strictement inférieur à la moitié du diamètre du compact $\bar{D}$ et une constante

$$L(\varepsilon) = \frac{r_0}{\max(M_1 + \varepsilon M_2, M_3)}$$

dépendant de ε tels que

$$\left[-\frac{L(\varepsilon)}{\varepsilon}, \frac{L(\varepsilon)}{\varepsilon} \right] \times B(x_0, r_0)$$

est un cylindre de sécurité pour les deux équations (P) et (Q). Cela signifie que les solutions des deux systèmes vivent dans D en espace pour tout temps inférieur en valeur absolue à $\frac{L(\varepsilon)}{\varepsilon}$. Comme $\varepsilon \in [0, \varepsilon_0]$, on a le même résultat pour tout temps inférieur en valeur absolue à $\frac{L}{\varepsilon}$ pour

$$L = \frac{r_0}{\max(M_1 + \varepsilon_0 M_2, M_3)}.$$

- Evaluation de la différence entre x et z :

Posons, pour tout $\varepsilon \in [0, \varepsilon_0]$ et tout $t \in [0, \frac{L}{\varepsilon}]$,

$$E(t, \varepsilon) = x(t, \varepsilon) - z(t, \varepsilon)$$

ce qui donne, en appliquant le fait que x et z soient issues d'équations différentielles,

$$\begin{aligned}
 E(t, \varepsilon) &= \varepsilon \int_0^t \left(f(s, x(s, \varepsilon)) + \varepsilon g(s, x(s, \varepsilon), \varepsilon) - \bar{f}(z(s, \varepsilon)) \right) ds \\
 &= \varepsilon \int_0^t \left(f(s, x(s, \varepsilon)) - f(s, z(s, \varepsilon)) \right. \\
 &\quad \left. + \varepsilon g(s, x(s, \varepsilon), \varepsilon) + f(s, z(s, \varepsilon)) - \bar{f}(z(s, \varepsilon)) \right) ds \\
 &= \varepsilon \int_0^t \left(f(s, x(s, \varepsilon)) - f(s, z(s, \varepsilon)) \right) ds \\
 &\quad + \varepsilon^2 \int_0^t g(s, x(s, \varepsilon), \varepsilon) ds + \varepsilon \int_0^t \left(f(s, z(s, \varepsilon)) - \bar{f}(z(s, \varepsilon)) \right) ds.
 \end{aligned}$$

Comme g est bornée sur $\mathbb{R} \times \bar{D} \times [0, \varepsilon_0]$ par une constante M_2 et que f est λ -lipschitzienne par rapport à la deuxième variable, nous avons, pour tout $\varepsilon \in [0, \varepsilon_0]$ et tout $t \in [0, \frac{L}{\varepsilon}]$,

$$\|E(t, \varepsilon)\| \leq \varepsilon \lambda \int_0^t \|E(s, \varepsilon)\| ds + \varepsilon^2 M_2 t + \varepsilon \int_0^t \|f(s, z(s, \varepsilon)) - \bar{f}(z(s, \varepsilon))\| ds.$$

• Etude du terme $\int (f(s, z) - \bar{f}(z)) :$

Posons la fonction

$$\phi(t, x) = f(s, x) - \bar{f}(x).$$

La fonction ϕ est T -périodique, continue et son intégrale sur une période vaut, pour tout $x \in \mathbb{R}^n$,

$$\int_0^T f(t, x) dt - \int_0^T dt \frac{1}{T} \int_0^T f(s, x) ds = 0.$$

ϕ est également bornée sur $\mathbb{R} \times \bar{D}$ comme fonction continue et périodique en temps. Un rapide calcul permet de vérifier qu'elle est également lipschitzienne en x pour une constante de Lipschitz μ .

Découpons l'intégrale de 0 à t de ϕ en autant de périodes que possible, puis appliquons l'inégalité triangulaire, ce qui donne

$$\left\| \int_0^t \phi(s, z(s, \varepsilon)) ds \right\| \leq \sum_{i=0}^{m-1} \left\| \int_{iT}^{(i+1)T} \phi(s, z(s, \varepsilon)) ds \right\| + \left\| \int_{mT}^t \phi(s, z(s, \varepsilon)) ds \right\|.$$

En outre, en utilisant les propriétés de ϕ que nous avons mises à jour, nous pouvons écrire

$$\begin{aligned}
 \int_{iT}^{(i+1)T} \phi(s, z(s, \varepsilon)) ds &= \int_{iT}^{(i+1)T} \left(\phi(s, z(s, \varepsilon)) - \phi(s, z(iT, \varepsilon)) \right) ds \\
 &\leq \mu \int_{iT}^{(i+1)T} \|z(s, \varepsilon) - z(iT, \varepsilon)\| ds.
 \end{aligned}$$

Or,

$$\begin{aligned}
 \|z(s, \varepsilon) - z(iT, \varepsilon)\| &= \left\| z(iT, \varepsilon) + \int_{iT}^s \varepsilon \bar{f}(z(u, \varepsilon)) \, du - z(iT, \varepsilon) \right\| \\
 &\leq \int_{iT}^s \|\varepsilon \bar{f}(z(u, \varepsilon))\| \, du \\
 &\leq \varepsilon(s - iT)M_3 \\
 &\leq \varepsilon TM_3.
 \end{aligned}$$

De même, puisque ϕ est T-périodique en temps et donc bornée sur le cylindre infini $\mathbb{R} \times \bar{D}$, dans lequel est compris $(t, z(t, \varepsilon))$ d'après ce qui précède, il existe une constante M_4 telle que

$$\left\| \int_{mT}^t \phi(s, z(s, \varepsilon)) \, ds \right\| \leq TM_4$$

On a donc, en réunissant les résultats obtenus,

$$\left\| \int_0^t \phi(s, z(s, \varepsilon)) \, ds \right\| \leq m\mu T^2 M_3 \varepsilon + TM_4.$$

Mais comme

$$mT \leq t \leq \frac{L}{\varepsilon},$$

on obtient finalement

$$\int_0^t \|f(s, z(s, \varepsilon)) - \bar{f}(z(s, \varepsilon))\| \, ds \leq m\mu LTM_3 + TM_4.$$

• Conclusion :

Posons $M_5 = m\mu LTM_3 + TM_4$. On a,

$$\|E(t, \varepsilon)\| \leq \varepsilon \lambda \int_0^t \|E(s, \varepsilon)\| \, ds + \varepsilon^2 M_2 t + \varepsilon M_5.$$

D'après le lemme 333, cela implique

$$\begin{aligned}
 \|E(t, \varepsilon)\| &\leq \left(\frac{\varepsilon^2 M_2}{\varepsilon \lambda} + \varepsilon M_5 \right) e^{\varepsilon \lambda t} - \frac{\varepsilon^2 M_2}{\varepsilon \lambda} \\
 &\leq \varepsilon \left(\left(\frac{M_2}{\lambda} + M_5 \right) e^{\varepsilon \lambda t} - \frac{M_2}{\lambda} \right)
 \end{aligned}$$

qui devient, en appliquant encore $t \leq \frac{L}{\varepsilon}$,

$$\|E(t, \varepsilon)\| \leq \varepsilon \left(\left(\frac{M_2}{\lambda} + M_5 \right) e^{\lambda L} - \frac{M_2}{\lambda} \right).$$

Si l'on pose

$$C = \left(\frac{M_2}{\lambda} + M_5 \right) e^{\lambda L} - \frac{M_2}{\lambda},$$

le théorème est démontré. ■

Remarque 335. Dans la preuve, nous avons un voisinage D dans lequel vivent les solutions, une valeur ε_0 qui borne l'applicabilité du théorème et une constante L qui limite son efficacité sur les temps longs. Nous avons choisi D et ε_0 et le L dépendait de ces deux données. Il était parfaitement possible de faire l'inverse : choisir le temps L sur lequel on veut avoir une estimation de l'efficacité de la moyenne périodique et en déduire la gamme ε_0 des ε qui sont acceptable pour notre approximation et l'ouvert D qui contrôlera assez les solutions pour finir la preuve. Le fait qu'il y ait un choix à faire entre l'espace et le temps est implicite dans l'énoncé du théorème qui ne parle que d'existence de plusieurs constantes.

Remarque 336. Si l'on pose $\tau = t\varepsilon$, alors

$$\frac{d}{d\tau} \left(z\left(\frac{\tau}{\varepsilon}, \varepsilon\right) \right) = \frac{1}{\varepsilon} z' \left(\frac{\tau}{\varepsilon}, \varepsilon \right) = \bar{f} \left(z\left(\frac{\tau}{\varepsilon}, \varepsilon\right) \right),$$

donc si w , fonction de la variable τ , est solution de

$$w' = \bar{f}(w),$$

alors on a en tout temps où cela est défini,

$$z\left(\frac{\tau}{\varepsilon}, \varepsilon\right) = w(\tau),$$

ou encore

$$z(t, \varepsilon) = w(t\varepsilon).$$

Il y a donc bien un lien intime entre ε_0 , L et D dans le théorème au sens également où ε , t et x sont fortement liés.

4.b. Cas général

Nous esquissons sans démonstration une ébauche de la théorie dans le cas où le champ de vecteurs n'est plus périodique en temps, pour donner une idée des ouvertures possible de la moyenne périodique.

Définition 337. Soit $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ un champ de vecteurs continu. Pour tout $T \in \mathbb{R}_+^*$, on appelle moyenne locale de f le champ de vecteurs

$$f_T : (t, z) \mapsto \frac{1}{T} \int_0^T f(t + s, z) \, ds.$$

Remarque 338. Si f est T -périodique en la variable temporelle, alors $f_T(t, z) = \bar{f}(z)$.

Définition 339. Soit $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ un champ de vecteurs continu et lipschitzien par rapport à la deuxième variable. On dit que f est un champ de

vecteur de Krylov, Bogoliubov, Mitropolsky, abrégé KBM-champ de vecteurs si la limite

$$\bar{f}(z) = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T f(t, z) dt$$

existe et est uniforme sur tout compact pour la variable d'espace z .

Théorème 340 (Moyenne générale)

Si f est un KBM-champ de vecteur, alors

$$x(t) - z(t) = \mathcal{O}(\sqrt{\delta_1(\varepsilon)}), \quad \varepsilon \rightarrow 0,$$

dès lors que $t \in [0, \frac{1}{\varepsilon}]$ et avec $\lim_{\varepsilon \rightarrow 0} \delta_1(\varepsilon) = 0$.

Bibliographie

- [1] Serge ALINHAC et Patrick GÉRARD : *Opérateurs pseudo-différentiels et théorème de Nash-Moser*. EDP Sciences.
- [2] Jean-Pierre DEMAILLY : *Analyse numérique et équations différentielles*. EDP science.
- [3] Larent DI MENZA : *Analyse numérique des équations aux dérivées partielles*. Cassini.
- [4] Francis FILBET : *Analyse numérique*. Dunod.
- [5] Xavier GOURDON : *Analyse*. Ellipses.
- [6] Evelyn F. KELLER et Lee A. SEGEL : Initiation of slime mold aggregation viewed as an instability. *Journal of Theoretical Biology*, 26:399 – 415, 1970.
- [7] Jacques LAFONTAINE : *Introduction aux variétés différentielles*. EDP science.
- [8] Daniel LI : *Cours d'analyse fonctionnelle*. Ellipses.
- [9] François ROUVIÈRE : *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*. Cassini.
- [10] Walter RUDIN : *Analyse réelle et complexe*. Dunod.
- [11] Jan A. SANDERS, Ferdinand VERHULST et James MURDOCK : *Averaging methods in nonlinear dynamical systems, Second edition*. Springer.
- [12] Claude ZUILY : *Éléments de distributions et d'équations aux dérivées partielles*. Dunod.