Problèmes d'évolution

Frédéric Legoll et Pierre Lissy

Février-Juin 2025

Table des matières

Chapitre 1

Rappels

Ce chapitre a l'objectif de rappeler plusieurs notions élémentaires. Nous en profitons pour faire un certain nombre de remarques, illustrées par plusieurs exercices, et montrant la spécificité de la dimension infinie par rapport à la dimension finie.

On rappelle tout d'abord la notation suivante pour un espace vectoriel normé E.

Définition 1.1. *La boule unité fermée de E est*

$$B_E = \{x \in E;\ \|x\|_E \leq 1\}.$$

1.1 Espaces de Hilbert

Dans cette section, on se place dans un espace de Hilbert V. On rappelle que V est donc un espace vectoriel muni d'un produit scalaire, qu'on note $\langle x, y\rangle$, que la norme induite par ce produit scalaire est $\|x\| = \sqrt{\langle x, x\rangle}$, et que V est complet pour cette norme.

1.1.1 Théorèmes fondamentaux

On rappelle maintenant quelques théorèmes fondamentaux pour les espaces de Hilbert.

Théorème 1.2 (Théorème de projection orthogonale)**.** *Soit V un espace de Hilbert et K un sous-espace vectoriel fermé de V. Pour tout $u \in V$, il existe un unique $v = P_K u \in K$, appelé projection orthogonale de u sur K, tel que*

$$\|P_K u - u\| = \inf_{w \in K} \|w - u\|.$$

De plus, $P_K u$ est caractérisé par

$$P_K u \in K \qquad et \qquad \forall w \in K,\ \langle u - P_K u, w\rangle = 0. \tag{1.1}$$

Démonstration. Cf. le cours de première année [5]. □

On peut faire un peu mieux, et simplement supposer que K est un sous-ensemble convexe et fermé de V.

Définition 1.3. *Soit E un espace vectoriel et C un sous-ensemble de E. L'ensemble C est convexe si, pour tout x et y dans C et tout $\lambda \in [0,1]$, on a $\lambda x + (1-\lambda)y \in C$.*

Théorème 1.4 (Théorème de projection sur un convexe)**.** *Soit V un espace de Hilbert et K un sous-ensemble fermé et convexe de V. Pour tout $u \in V$, il existe un unique $v = P_K u \in K$, appelé projection de u sur K, tel que*

$$\|P_K u - u\| = \inf_{w \in K} \|w - u\|.$$

De plus, $P_K u$ est caractérisé par

$$P_K u \in K \qquad et \qquad \forall w \in K,\ \langle u - P_K u, w - P_K u\rangle \leq 0. \tag{1.2}$$

Démonstration. La preuve est très similaire à celle du théorème de projection orthogonale donnée dans [5]. □

Le théorème suivant permet d'identifier un espace de Hilbert V avec son dual $V' = \mathcal{L}(V, \mathbb{R})$:

Théorème 1.5 (Théorème de Riesz)**.** *Soit V un espace de Hilbert. Etant donné $\varphi \in V'$, il existe un unique $u \in V$ tel que*

$$\forall w \in V, \quad \varphi(w) = \langle u, w\rangle.$$

De plus, on a $\|u\|_V = \|\varphi\|_{V'}$. En d'autres termes, l'application de V' dans V qui à φ associe u permet d'identifier l'espace de Hilbert V avec son dual.

Démonstration. Cf. le cours de première année [5]. □

La notion d'application bilinéaire coercive joue un rôle fondamental pour l'étude des équations aux dérivées partielles.

Définition 1.6. *Soit V un espace de Hilbert et soit a une forme bilinéaire sur V. On dit que a est coercive sur V s'il existe un réel $\alpha > 0$ tel que*

$$\forall u \in V, \quad a(u,u) \geq \alpha \|u\|^2.$$

Théorème 1.7 (Théorème de Lax-Milgram)**.** *Soit V un espace de Hilbert et a une forme bilinéaire sur V, symétrique, continue et coercive. Soit b une forme linéaire continue sur V. Alors le problème*

$$\begin{cases} \text{Chercher } u \in V \text{ tel que} \\ \forall w \in V, \quad a(u,w) = b(w) \end{cases} \tag{1.3}$$

admet une unique solution. De plus, le problème (1.3) est équivalent au problème de minimisation

$$\begin{cases} \text{Chercher } u \in V \text{ tel que} \\ J(u) = \inf\limits_{w \in V} J(w) \end{cases} \tag{1.4}$$

où la fonctionnelle d'énergie $J(w)$ est définie par $J(w) = \frac{1}{2}a(w,w) - b(w)$.

Démonstration. Cf. le cours de première année [9]. □

Remarque 1.8. *On peut supprimer l'hypothèse de symétrie sur la forme bilinéaire a. Alors le problème (1.3) admet encore une unique solution, mais il n'y a plus d'équivalence de (1.3) avec un problème de minimisation du type (1.4).*

1.1.2 Bases hilbertiennes

La notion de base hilbertienne généralise en dimension infinie la notion de base orthonormée.

Définition 1.9. *Soit V un espace de Hilbert. On appelle base hilbertienne de V une suite $(e_n)_{n\geq 1}$ d'éléments de V tels que*
- *pour tout n, $\|e_n\| = 1$ et pour tous $m \neq n$, $\langle e_n, e_m\rangle = 0$.*
- *l'espace vectoriel engendré par la famille $(e_n)_{n\geq 1}$ est dense dans V.*

Proposition 1.10. *Soit V un espace de Hilbert admettant une base hilbertienne $(e_n)_{n\geq 1}$. Soit $u \in V$ et posons $u_n = \langle u, e_n\rangle$ pour tout $n \geq 1$. Alors, les séries $\sum_{n\geq 1} u_n e_n$ et $\sum_{n\geq 1} |u_n|^2$ sont convergentes dans V et $\mathbb{R}$ respectivement, et on a*

$$u = \sum_{n\geq 1} u_n e_n \qquad \text{et} \qquad \|u\|^2 = \sum_{n\geq 1} |u_n|^2.$$

Démonstration. Cf. le cours de première année [5]. □

1.1.3 Orthogonal d'un sous-espace

Définition 1.11. *Soit V un espace de Hilbert, et $W \subset V$ un sous-espace vectoriel. On note*

$$W^\perp = \{v \in V; \quad \forall w \in W, \quad \langle v, w\rangle = 0\}.$$

Lemme 1.12. *Soit V un espace de Hilbert, et $W \subset V$ un sous-espace vectoriel. Alors $W^\perp$ est un sous-espace vectoriel fermé de V.*

Démonstration. Soit $(v_n)_{n\geq 1}$ une suite d'éléments de $W^\perp$ qui converge vers $v \in V$. Pour tout $w \in W$, et tout $n \geq 1$, on a $\langle v_n, w\rangle = 0$. En passant à la limite, on obtient donc $\langle v, w\rangle = 0$ et par conséquent $v \in W^\perp$. □

Lemme 1.13. *Soit V un espace de Hilbert, et $W \subset V$ un sous-espace vectoriel. Alors*

$$\left(W^\perp\right)^\perp = \overline{W}.$$

Démonstration. Par définition,

$$\left(W^\perp\right)^\perp = \left\{v \in V; \quad \forall w \in W^\perp,\ \langle v, w\rangle = 0\right\}.$$

On a immédiatement que $W \subset \left(W^\perp\right)^\perp$. D'après le lemme 1.12, $\left(W^\perp\right)^\perp$ est fermé, donc $\overline{W} \subset \left(W^\perp\right)^\perp$. Soit maintenant $x \in \left(W^\perp\right)^\perp$. Comme $\overline{W}$ est fermé, on peut appliquer le théorème de projection orthogonale de V sur $\overline{W}$ et décomposer x selon

$$x = P_{\overline{W}}x + y, \tag{1.5}$$

avec $y \in (\overline{W})^\perp$, et donc $\langle y, P_{\overline{W}}x\rangle = 0$. On a aussi $y \in W^\perp$, et comme $x \in \left(W^\perp\right)^\perp$, ceci implique $\langle x, y\rangle = 0$. Donc

$$0 = \langle x, y\rangle - \langle P_{\overline{W}}x, y\rangle = \langle x - P_{\overline{W}}x, y\rangle = \langle y, y\rangle,$$

ce qui conduit à $y = 0$. La relation (1.5) implique alors que $x \in \overline{W}$. On a donc montré que $\left(W^\perp\right)^\perp \subset \overline{W}$, ce qui termine la preuve. □

Théorème 1.14. *Si W est fermé dans V, et que $W^\perp = \{0\}$, alors $W = V$ tout entier.*

Démonstration. Soit $x \in V$. Comme W est fermé, on peut appliquer le théorème de projection orthogonale et décomposer x selon

$$u = P_W x + y. \tag{1.6}$$

La caractérisation (1.1) donne $\langle y, w\rangle = 0$ pour tout $w \in W$. Donc $y \in W^\perp$, et par conséquent $y = 0$. On déduit de (1.6) que $x = P_W x$, soit $x \in W$. Par conséquent, $W = V$. □

1.2 Espaces de Sobolev

Les espaces de Sobolev jouent un rôle central dans l'étude des équations aux dérivées partielles.

1.2.1 Définitions principales

Soit Ω un ouvert de $\mathbb{R}^d$. On rappelle que, pour tout $p \geq 1$, l'ensemble $L^p(\Omega)$ est l'ensemble des fonctions dont la puissance p-ième est intégrable sur Ω.

On rappelle qu'un multi-indice $\alpha = (\alpha_1, \ldots, \alpha_d)$ est un élément de $\mathbb{N}^d$. Sa longueur est $|\alpha| = \sum_{i=1}^d \alpha_i$ et on adopte la notation suivante : pour toute distribution $u \in \mathcal{D}'(\Omega)$,

$$\partial^\alpha u = \frac{\partial^{|\alpha|}\, u}{\partial^{\alpha_1} x_1 \ldots \partial^{\alpha_d} x_d} = \frac{\partial^{\alpha_1+\ldots+\alpha_d}\, u}{\partial^{\alpha_1} x_1 \ldots \partial^{\alpha_d} x_d}.$$

Définition 1.15. *Pour $k \geq 1$, l'espace de Sobolev $H^k(\Omega)$ est l'ensemble des fonctions $f \in L^2(\Omega)$ telles que les dérivées de f au sens des distributions, jusqu'à l'ordre k, s'identifient à des fonctions de $L^2(\Omega)$. Autrement dit,*

$$H^k(\Omega) = \left\{ f \in L^2(\Omega) \quad \textit{telles que} \quad \forall \alpha \in \mathbb{N}^d,\ |\alpha| \leq k,\ \partial_\alpha f \in L^2(\Omega) \right\}.$$

Comme l'espace $L^2(\Omega)$, les espaces $H^k(\Omega)$ sont des espaces de Hilbert.

Théorème 1.16. *Muni du produit scalaire*

$$(f, g)_{H^k} = \int_\Omega f(x)\, g(x)\, dx + \sum_{1 \leq |\alpha| \leq k} \int_\Omega \partial_\alpha f(x)\, \partial_\alpha g(x)\, dx,$$

l'espace $H^k(\Omega)$ est un espace de Hilbert. Sa norme est notée $\| \cdot \|_{H^k(\Omega)}$.

On rappelle maintenant un théorème de densité de l'ensemble des fonctions test.

Théorème 1.17. *Pour tout ouvert Ω de $\mathbb{R}^d$, l'ensemble $\mathcal{D}(\Omega)$ est dense dans $L^2(\Omega)$ pour la norme $L^2(\Omega)$.*

De plus, pour tout $k \geq 1$, l'ensemble $\mathcal{D}(\mathbb{R}^d)$ est dense dans $H^k(\mathbb{R}^d)$ pour la norme $H^k(\mathbb{R}^d)$.

Pour tout $k \geq 1$, si $\Omega \subset \mathbb{R}^d$ avec $\Omega \neq \mathbb{R}^d$, alors $\mathcal{D}(\Omega)$ n'est pas dense dans $H^k(\Omega)$.

Définition 1.18. *Pour $k \geq 1$, on définit $H_0^k(\Omega)$ comme la fermeture de $\mathcal{D}(\Omega)$ dans $H^k(\Omega)$ (pour la norme de $H^k(\Omega)$).*

On donne maintenant un résultat propre à la dimension 1.

Théorème 1.19. *Soit I un intervalle de $\mathbb{R}$ et $u \in H^1(I)$. Alors u s'identifie à une fonction continue et, pour tout x et y dans I,*

$$u(x) - u(y) = \int_y^x u'(s) ds.$$

On souligne que ce théorème est faux en dimension plus grande.

Démonstration. On esquisse ici la preuve, dont les détails sont laissés au lecteur. Soit $x_0 \in I$ fixé. Pour $u \in H^1(I)$, on définit

$$w(x) = \int_{x_0}^{x} u'(s)ds.$$

Grâce à l'inégalité de Cauchy-Schwarz, cette définition a bien un sens, et on montre que w est une fonction continue sur I. On calcule ensuite la dérivée de w au sens des distributions, en utilisant le théorème de Fubini. On montre ainsi que $w' = u'$ dans $\mathcal{D}'(I)$. Par conséquent, $w - u$ est une constante, et u s'identifie donc bien à une fonction continue. □

1.2.2 Trace

Pour une fonction définie dans un ouvert Ω, on souhaite définir sa valeur au bord de Ω. Pour les fonctions $u \in L^2(\Omega)$, cette notion n'a pas de sens. Par contre, si u est plus régulière, alors on peut définir rigoureusement cette notion.

Proposition 1.20. *Soit Ω un ouvert borné et régulier. On peut définir une application linéaire et continue*

$$\begin{array}{rccc} \gamma \ : & H^1(\Omega) & \longrightarrow & L^2(\partial\Omega) \\ & u & \mapsto & \gamma(u), \end{array}$$

et qui prolonge l'application trace pour les fonctions continues sur $\overline{\Omega}$: pour tout $u \in H^1(\Omega) \cap C^0(\overline{\Omega})$, $\gamma(u) = u|_{\partial\Omega}$.

L'application trace est continue de $H^1(\Omega)$ dans $L^2(\partial\Omega)$, ce qui signifie qu'il existe une constante C_Ω telle que

$$\forall u \in H^1(\Omega),\ \|\gamma(u)\|_{L^2(\partial\Omega)} \leq C_\Omega\ \|u\|_{H^1(\Omega)}. \tag{1.7}$$

Remarque 1.21. *L'application trace n'est pas surjective sur $L^2(\partial\Omega)$, mais sur un espace plus petit, qui est $H^{1/2}(\partial\Omega)$. Elle est en fait continue de $H^1(\Omega)$ vers $H^{1/2}(\partial\Omega)$, si bien qu'il existe C_Ω tel que*

$$\forall u \in H^1(\Omega),\ \|\gamma(u)\|_{H^{1/2}(\partial\Omega)} \leq C_\Omega\ \|u\|_{H^1(\Omega)}.$$

Enfin, pour tout $u \in H^{1/2}(\partial\Omega)$, on a $\|u\|_{L^2(\partial\Omega)} \leq \|u\|_{H^{1/2}(\partial\Omega)}$.

L'espace $H_0^1(\Omega)$, défini comme la fermeture dans $H^1(\Omega)$ de $\mathcal{D}(\Omega)$, s'identifie à l'espace des fonctions à trace nulle :

Proposition 1.22. *Soit Ω un ouvert de $\mathbb{R}^d$. On a*

$$H_0^1(\Omega) = \left\{ u \in H^1(\Omega), \quad \gamma(u) = 0 \right\}.$$

1.2.3 Inégalité de Poincaré

On rappelle la notation suivante :

Définition 1.23. *Soit Ω un ouvert de $\mathbb{R}^d$. Pour une fonction u à valeur vectorielle $u = (u_1, \dots, u_d) \in L^2(\Omega)^d$, on note*

$$\|u\|_{L^2(\Omega)} = \sqrt{\sum_{i=1}^{d} \|u_i\|^2_{L^2(\Omega)}} \, .$$

Proposition 1.24 (Inégalité de Poincaré)**.** *Soit Ω un ouvert borné de $\mathbb{R}^d$. Alors il existe une constante C_Ω telle que*

$$\forall u \in H^1_0(\Omega), \quad \|u\|_{L^2(\Omega)} \leq C_\Omega \ \|\nabla u\|_{L^2(\Omega)} \, . \tag{1.8}$$

Démonstration. Cette inégalité est démontrée dans le cours [9]. L'exercice 18 en propose une autre démonstration. L'exercice 23 donne une caractérisation de la meilleure constante C_Ω en terme de valeur propre du laplacien. □

1.2.4 Injections de Sobolev

On considère une fonction $u \in H^1(\Omega)$. Bien sûr, $u \in L^2(\Omega)$. On peut se demander si u n'est pas plus régulière que ceci, du fait que ∇u soit dans $L^2(\Omega)$. Le théorème suivant répond à cette question.

Théorème 1.25. *Soit Ω un ouvert régulier de $\mathbb{R}^d$, et soit k un entier. On a les injections continues suivantes :*

- *si $d > 2k$, alors $H^k(\Omega) \subset L^{p^*}(\Omega)$ avec $1/p^* = 1/2 - k/d$.*
- *si $d = 2k$, alors $H^k(\Omega) \subset L^q(\Omega)$ pour tout $q \in [2, +\infty[$.*
- *si $d < 2k$, alors $H^k(\Omega) \subset C^0(\overline{\Omega})$.*

On rappelle maintenant l'inégalité de Hölder.

Lemme 1.26 (Inégalité de Hölder)**.** *Soient p et q deux réels compris (au sens large) entre 1 et $+\infty$, avec $1/p + 1/q = 1$. Soient $f \in L^p(\Omega)$ et $g \in L^q(\Omega)$. Alors le produit $f\,g$ est dans $L^1(\Omega)$ et*

$$\|f\,g\|_{L^1(\Omega)} \leq \|f\|_{L^p(\Omega)} \ \|g\|_{L^q(\Omega)}.$$

On déduit de cette inégalité (le faire en exercice !) le résultat suivant :

Lemme 1.27. *Soient p et q deux réels compris (au sens large) entre 1 et $+\infty$, avec $p < q$. Soit $f \in L^p(\Omega) \cap L^q(\Omega)$. Alors, pour tout $r \in [p, q]$, on a $f \in L^r(\Omega)$, avec*

$$\|f\|_{L^r(\Omega)} \leq \|f\|^{\alpha}_{L^p(\Omega)} \ \|f\|^{1-\alpha}_{L^q(\Omega)},$$

où α est tel que $1/r = \alpha/p + (1-\alpha)/q$.

Ainsi, soit Ω un ouvert régulier de $\mathbb{R}^d$, et soit k un entier, avec par exemple $d > 2k$. On a vu que $H^k(\Omega) \subset L^{p^*}(\Omega)$ avec $1/p^* = 1/2 - k/d$. De plus, $H^k(\Omega) \subset L^2(\Omega)$. Donc $H^k(\Omega) \subset L^r(\Omega)$ pour tout $r \in [2, p^*]$.

1.3 Convergence faible

On rappelle qu'une suite d'éléments $(u_n)_{n\geq 0}$ d'un espace de Hilbert V converge vers $u \in V$ si $\lim_n \|u_n - u\| = 0$. On introduit ici une notion de convergence plus faible, la *convergence faible*. Pour éviter les confusions, on parlera alors de *convergence forte* pour la convergence usuelle.

Avant d'introduire cette nouvelle notion, on rappelle ici quelques notions liées à la compacité de sous-ensembles d'un espace vectoriel.

1.3.1 Compacité

On se place dans un espace vectoriel normé E. On rappelle la définition suivante :

Définition 1.28. *Un sous-ensemble $K \subset E$ est compact si, de toute suite $(u_n)_{n\geq 0}$ d'éléments de K, on peut extraire une sous-suite convergente dans K.*

Nous aurons besoin dans la suite de ce cours d'une notion plus fine que celle d'ensemble compact, et que nous introduisons maintenant :

Définition 1.29. *Un sous-ensemble $K \subset E$ est relativement compact si, de toute suite $(u_n)_{n\geq 0}$ d'éléments de K, on peut extraire une sous-suite convergente dans E.*

La différence avec la notion d'ensemble compact est donc que la limite de la suite n'appartient pas nécessairement à K.

La preuve de la proposition suivante est laissée en exercice :

Proposition 1.30. *Un sous-ensemble $K \subset E$ est relativement compact si et seulement si $\overline{K}$ est compact.*

On rappelle que les sous-ensembles compacts de E sont nécessairement des ensembles fermés et bornés. La réciproque n'est vraie que dans le cas où E est un espace de dimension finie. On a en effet le résultat suivant, caractéristique de la dimension infinie :

Théorème 1.31. *Soit V un espace de Hilbert de dimension infinie. Alors la boule unité fermée de V n'est pas compacte.*

Démonstration. Comme l'espace est de dimension infinie, on peut construire une suite orthonormée infinie $(e_n)_{n\geq 1}$ (en utilisant le procédé de Gram-Schmidt). Cette suite appartient bien à la boule unité fermée. Par ailleurs, pour $n \neq p$, on a

$$\|e_n - e_p\|^2 = \|e_n\|^2 + \|e_p\|^2 - 2\langle e_n, e_p\rangle = 2. \tag{1.9}$$

Supposons que la boule unité fermée est compacte. Alors on peut extraire de la suite $(e_n)_{n\geq 1}$ une sous-suite convergente, donc de Cauchy. Or ceci est contradictoire avec (1.9). □

1.3.2 Définition de la convergence faible

Avant de donner la définition de la notion de convergence faible, nous avons besoin de rappeler la définition de la limite inférieure d'une suite de réels.

Définition 1.32. *Soit u_n une suite de réels. On définit sa limite inférieure par*

$$\liminf u_n = \lim_{n\to\infty} \left(\inf_{k\geq n} u_k \right).$$

La suite $I_n = \inf_{k\geq n} u_k$ est une suite croissante de réels, qui admet donc bien une limite (éventuellement infinie).

Le lemme suivant montre que la notion de limite inférieure généralise la notion de limite.

Lemme 1.33. *Soit u_n une suite de réels qui converge vers λ. Alors $\lambda = \liminf u_n$.*

Dans le cas d'une suite quelconque, on a le résultat suivant :

Lemme 1.34. *Soit u_n une suite de réels, et soit $\lambda = \liminf u_n$. On peut extraire de u_n une sous-suite qui converge vers λ.*

Démonstration. On suppose $\lambda \in \mathbb{R}$ (le cas $\lambda = +\infty$ se traite de la même façon). On pose $I_n = \inf_{k\geq n} u_k$: par définition, $\lambda = \lim_n I_n$. Soit $\varepsilon > 0$ et $N > 0$. Il existe $n_0 > N$ tel que $\lambda \geq I_{n_0} \geq \lambda - \varepsilon$. De plus, il existe $k_0 \geq n_0$ tel que $\varepsilon + \inf_{k\geq n_0} u_k \geq u_{k_0} \geq \inf_{k\geq n_0} u_k$. Donc on a $\varepsilon + \lambda \geq u_{k_0} \geq \lambda - \varepsilon$, ce qui conclut la preuve. □

On introduit maintenant la notion de convergence faible.

Définition 1.35. *Soit V un espace de Hilbert. On dit qu'une suite u_n de V converge faiblement vers u dans V si $u \in V$ et*

$$\forall w \in V, \ \lim_{n\to+\infty} \langle u_n, w\rangle = \langle u, w\rangle.$$

On note $u_n \rightharpoonup u$.

Si V est de dimension finie, alors la convergence au sens faible est équivalente à la convergence au sens fort. En dimension infinie, les deux notions sont différentes.

On a également la caractérisation équivalente suivante de la convergence faible.

Proposition 1.36. *Soit V un espace de Hilbert, $u \in V$ et $(u_n)_{n\in N}$ une suite d'élements de V. Les deux propositions suivantes sont équivalentes :*

(i) $(u_n)_{n\in\mathbb{N}}$ converge faiblement vers u dans V ;

(ii) pour toute forme linéaire continue $\varphi \in V'$,

$$\varphi(u_n) \underset{n\to+\infty}{\longrightarrow} \varphi(u).$$

Démonstration. On montre que (ii) implique (i). Ceci découle du fait que, pour tout $w \in V$, l'application $\varphi : v \in V \mapsto \langle v, w\rangle \in \mathbb{R}$ est une forme linéaire continue. Montrons maintenant que (i) implique (ii). Ceci est une conséquence du théorème de Riesz. En effet, pour tout $\varphi \in V'$, il existe $w \in V$ tel que pour tout $v \in V$, $\varphi(v) = \langle w, v\rangle$. D'où le résultat. □

1.3.3 Propriétés de la convergence faible

Nous commençons par énoncer les liens entre convergence faible et convergence forte (au sens usuel).

Théorème 1.37. *Soit u_n une suite de V.*
- *si u_n converge fortement vers u dans V, alors u_n converge faiblement vers u dans V ;*
- *si u_n converge faiblement vers u dans V, alors la suite u_n est bornée dans V et $\|u\| \leq \liminf_{n\to\infty} \|u_n\|$.*
- *Si u_n converge vers u faiblement et w_n converge vers w fortement, alors on a $\lim_{n\to\infty}\langle u_n, w_n\rangle = \langle u, w\rangle$.*

Démonstration. La preuve de la première et de la troisième affirmation sont laissées au lecteur (utiliser l'inégalité de Cauchy-Schwarz). Le fait qu'une suite qui converge faiblement soit bornée est une propriété plus difficile à démontrer, et qui sera ici admise. Elle repose sur le théorème de Banach-Steinhaus (cf. par exemple [2, Théorème II.1 et Proposition III.5]). On prouve maintenant l'inégalité dans la deuxième affirmation. Supposons que u_n converge faiblement vers u. L'inégalité de Cauchy-Schwarz donne que

$$\left\langle \frac{u}{\|u\|}, u_n \right\rangle \leq \|u_n\|.$$

On passe à la limite inférieure et on utilise que le membre de gauche converge :

$$\lim_{n\to\infty} \left\langle \frac{u}{\|u\|}, u_n \right\rangle = \liminf_{n\to\infty} \left\langle \frac{u}{\|u\|}, u_n \right\rangle \leq \liminf_{n\to\infty} \|u_n\|,$$

d'où le fait que $\|u\| \leq \liminf_{n\to\infty} \|u_n\|$. □

L'intérêt de la convergence faible réside dans la proposition suivante, que nous admettrons.

Proposition 1.38. *Soit V un espace de Hilbert. La boule unité de V est faiblement compacte : de toute suite bornée de V, on peut extraire une sous-suite qui converge faiblement dans V.*

Dans un espace de Hilbert, pour montrer qu'une suite converge faiblement (à extraction près), il suffit donc de montrer qu'elle est bornée.

La définition d'ensemble fermé pour la topologie faible est naturelle :

Définition 1.39. *Soit V un espace de Hilbert, et C un sous-ensemble de V. On dit que C est faiblement fermé si, pour toute suite d'éléments $(u_n)_{n\geq 0}$ de C qui converge faiblement vers u dans V, on a $u \in C$.*

Comme la convergence forte implique la convergence faible, un ensemble faiblement fermé (i.e. fermé pour la topologie faible) est fortement fermé (i.e. fermé pour la topologie forte). La réciproque est fausse, sauf si l'ensemble est convexe, comme le montre le résultat suivant :

Proposition 1.40. *Soit V un espace de Hilbert, et C un sous-ensemble de V qui soit convexe et fortement fermé. Alors C est faiblement fermé.*

Démonstration. Soit u_n est une suite de points de C qui converge faiblement vers $u \in V$. Comme C est convexe et fortement fermé dans V, on peut considérer la projection de V sur C, qu'on note P_C. D'après le théorème 1.4, on a

$$\forall w \in C,\ \langle u - P_C u, w - P_C u\rangle \leq 0.$$

On écrit cette inégalité avec $w = u_n$ et on passe à la limite $n \to +\infty$ en utilisant la convergence faible de u_n vers u. Donc $\langle u - P_C u, u - P_C u\rangle \leq 0$, ce qui implique que $u = P_C u$ et donc $u \in C$. □

Proposition 1.41. *Soit V un espace de Hilbert et $J : V \to \mathbb{R}$ une fonction continue (pour la topologie forte de V) et convexe sur V. Pour toute suite u_n qui converge faiblement dans V vers u, on a*

$$J(u) \leq \liminf J(u_n).$$

Démonstration. Pour tout $\lambda \in \mathbb{R}$, l'ensemble $C(\lambda) = \{u \in V;\ J(u) \leq \lambda\}$ est convexe, car J est convexe. Comme J est continue, cet ensemble est fortement fermé. On utilise la proposition 1.40 : $C(\lambda)$ est faiblement fermé.

Soit $\lambda_0 = \liminf J(u_n)$. Le lemme 1.34 donne l'existence d'une sous-suite extraite $u_{\varphi(n)}$ telle que $\lim_n J(u_{\varphi(n)}) = \lambda_0$. Par conséquent, pour tout $\varepsilon > 0$, et pour tout $n \geq n_0(\varepsilon)$, on a $J(u_{\varphi(n)}) \leq \varepsilon + \lambda_0$, et donc $u_{\varphi(n)} \in C(\varepsilon + \lambda_0)$. Par ailleurs, la suite $u_{\varphi(n)}$ converge faiblement vers u. Donc $u \in C(\varepsilon + \lambda_0)$, soit $J(u) \leq \varepsilon + \lambda_0$, et ce pour tout ε. Donc $J(u) \leq \lambda_0$, ce qui conclut la preuve. □

On a donc vu que les notions de topologie faible et de convexité sont reliées.

En guise d'application de ces notions aux espaces de Sobolev, nous donnons la proposition suivante :

Proposition 1.42. *De toute suite bornée de $H_0^1(\Omega)$, on peut extraire une-suite qui converge faiblement vers u dans $H^1(\Omega)$. De plus, $u \in H_0^1(\Omega)$.*

Démonstration. La proposition 1.38 donne l'existence d'une sous-suite qui converge faiblement vers u dans $H^1(\Omega)$. L'espace $H_0^1(\Omega)$ est fermé dans $H^1(\Omega)$ et convexe, donc il est faiblement fermé en vertu de la proposition 1.40, et donc $u \in H_0^1(\Omega)$. □

Chapitre 2

Introduction à la théorie spectrale

Nous présentons dans ce chapitre les fondements de la théorie spectrale des opérateurs (définis en Section 2.1). Cette théorie est particulièrement utile et importante pour l'étude des équations aux dérivées partielles. En effet, un des buts premiers de l'étude d'un opérateur est la détermination de son spectre (Section 2.2), qui est la généralisation en dimension infinie de l'ensemble des valeurs propres d'une matrice. Dans les cas les plus simples, notamment pour les opérateurs dits compacts (Section 2.3), on peut déterminer complètement de manière qualitative le spectre d'un opérateur, et ensuite l'approcher numériquement. Ceci permet de résoudre des problèmes aux valeurs propres définis par une équation aux dérivées partielles (voir le Chapitre 3), ainsi que des problèmes d'évolution en mécanique, physique, etc, comme l'équation de la chaleur, l'équation des ondes, ou l'équation de Schrödinger (cf. la deuxième partie du polycopié).

Nous verrons des applications concrètes de cette théorie dans le Chapitre 3.

2.1 Applications linéaires

2.1.1 Applications linéaires et continues

Proposition 2.1. *Soit A une application linéaire de E dans F, où E et F sont deux espaces vectoriels normés. Les 3 propositions suivantes sont équivalentes :*

- *A est continue.*
- *A est continue en 0.*
- *il existe une constante $c \geq 0$ telle que*

$$\forall u \in E, \quad \|Au\|_F \leq c\|u\|_E.$$

Démonstration. Cf. le cours de première année [5]. □

Attention, comme le montre l'exercice suivant, la norme choisie joue un role.

Exercice 1. *On considère les espaces de fonctions $C^0([0,1])$ et $C^1([0,1])$, qu'on munit de la norme*

$$\|f\| = \sup_{t\in[0,1]} |f(t)|.$$

L'application

$$\begin{array}{rcl} A : C^1([0,1]) & \longrightarrow & C^0([0,1]) \\ f & \longmapsto & f' \end{array}$$

est linéaire. Montrer qu'elle n'est pas continue.

Définition 2.2. *On note $\mathcal{L}(E,F)$ l'espace vectoriel des opérateurs linéaires et continus de E dans F. L'application $\|\cdot\|$ définie par*

$$\forall A \in \mathcal{L}(E,F), \quad \|A\| := \sup_{x\in E\setminus\{0\}} \frac{\|Ax\|_F}{\|x\|_E} = \sup_{x\in E,\ \|x\|_E=1} \|Ax\|_F, \tag{2.1}$$

est une norme sur cet espace.

Le seul point éventuellement délicat est de montrer l'inégalité triangulaire $\|A+B\| \leq \|A\| + \|B\|$. Pour ce faire, on fixe $f \in E \setminus \{0\}$ et on écrit

$$\|(A+B)f\|_F \leq \|Af\|_F + \|Bf\|_F \leq \Big(\|A\| + \|B\|\Big)\|f\|_E.$$

Ceci montre que

$$\frac{\|(A+B)f\|_F}{\|f\|_E} \leq \|A\| + \|B\|,$$

d'où le résultat en prenant le supremum sur $f \in E \setminus \{0\}$.

Exercice 2. *Soient E, F et G trois espaces de Banach, et $A \in \mathcal{L}(E,F)$ et $B \in \mathcal{L}(F,G)$. Montrer que $BA \in \mathcal{L}(E,G)$ et $\|BA\| \leq \|A\|\ \|B\|$.*

Un cas particulier important est lorsque l'espace d'arrivée est $\mathbb{R}$.

Définition 2.3. *L'ensemble $\mathcal{L}(E,\mathbb{R})$ des applications linéaires continues de E dans $\mathbb{R}$ est appelé espace dual de E et est noté E'. Un élément de E' est appelé forme linéaire continue et son action sur un élément $u \in E$ est notée à l'aide du crochet de dualité :*

$$\langle A, u\rangle_{E',E} = Au \in \mathbb{R}.$$

L'espace E' est équipé de la norme

$$\|A\|_{E'} = \sup_{u\in E, u\neq 0} \frac{|Au|}{\|u\|_E}.$$

Donnons quelques exemples d'applications linéaires et continus.

Exemple 2.4 (Opérateurs de shift)**.** *On considère $E = F = \ell^p(\mathbb{N}, \mathbb{C})$ (pour $1 \leq p \leq +\infty$ fixé), où*

$$\ell^p(\mathbb{N}, \mathbb{C}) = \left\{ (x_1, x_2, \dots, x_n, \dots) \in \mathbb{C}^{\mathbb{N}} \;\middle|\; \sum_{n=1}^{+\infty} |x_i|^p < +\infty \right\}, \qquad 1 \leq p < +\infty,$$

et

$$\ell^\infty(\mathbb{N}, \mathbb{C}) = \left\{ (x_1, x_2, \dots, x_n, \dots) \in \mathbb{C}^{\mathbb{N}} \;\middle|\; \sup_{i \in \mathbb{N}} |x_i| < +\infty \right\}.$$

On définit les opérateurs de shift à droite et de shift à gauche, sur $\ell^p(\mathbb{N}, \mathbb{C})$, par

$$\tau_{\mathrm{d}}(x_1, x_2, \dots, x_n, \dots) = (0, x_1, x_2, \dots, x_n, \dots) \tag{2.2}$$

et

$$\tau_{\mathrm{g}}(x_1, x_2, \dots, x_n, \dots) = (x_2, x_3, \dots, x_n, \dots). \tag{2.3}$$

Ces deux applications sont linéaires et continues. Il est immédiat que $\|\tau_{\mathrm{d}} x\| = \|x\|$ pour tout $x \in \ell^p(\mathbb{N}, \mathbb{C})$ et donc $\|\tau_{\mathrm{d}}\| = 1$. Pour τ_{g}, on note tout d'abord que $\|\tau_{\mathrm{g}} x\| \leq \|x\|$, avec égalité par exemple pour $x = (0, 1, 0, \dots)$, ce qui donne $\|\tau_{\mathrm{g}}\| = 1$.

Exercice 3 (Opérateur de convolution)**.** *Soit $E = F = \mathrm{L}^2(\mathbb{R}^d)$ et $k \in \mathrm{L}^1(\mathbb{R}^d)$. Montrer que l'opérateur $T : E \to E$ d'action $Tf = k \star f$ est bien défini, qu'il est linéaire et continu et vérifie $\|T\| \leq \|k\|_{\mathrm{L}^1}$.*

Exercice 4 (Opérateur intégral)**.** *On considère $E = \mathrm{L}^1([0,1], \mathbb{R})$, $F = C^0([0,1], \mathbb{R})$, et $k \in C^0([0,1]^2, \mathbb{R})$. On rappelle que la norme sur l'espace de Banach F est $\|g\|_F = \sup_{x \in [0,1]} |g(x)|$. On considère l'opérateur K défini par*

$$Kf(x) = \int_0^1 k(x, y) f(y)\, dy.$$

Vérifier que $Kf \in F$ lorsque $f \in E$ puis que $K \in \mathcal{L}(E, F)$.

Exemple 2.5 (Opérateur de multiplication)**.** *Soit $E = F = \mathrm{L}^2(\mathbb{R}^d)$. Pour une fonction $V \in \mathrm{L}^\infty(\mathbb{R}^d, \mathbb{C})$ donnée, on définit l'opérateur A sur E par*

$$A\varphi = V\varphi.$$

On constate que, pour tout $\varphi \in E$, on a $A\varphi \in F$, et que $\|A\varphi\|_F \leq \|V\|_{\mathrm{L}^\infty} \|\varphi\|_E$. Donc A est linéaire et continu.

Exercice 5. *Montrer que si, dans l'Exemple 2.5, la fonction V est continue et bornée, alors $\|A\| = \sup_{x \in \mathbb{R}^d} |V(x)|$.*

Concluons cette section par un résultat important.

Proposition 2.6. *Si F est un espace de Banach et E un espace normé, alors $\mathcal{L}(E,F)$ est un espace de Banach.*

Démonstration. Considérons une suite de Cauchy $(A_n)_{n\geq 0}$ de $\mathcal{L}(E,F)$ pour la norme donnée par (2.1). Alors, pour tout $\varepsilon > 0$, il existe $N_\varepsilon \in \mathbb{N}$ tel que

$$\|A_n - A_m\| \leq \varepsilon \tag{2.4}$$

si $n, m \geq N_\varepsilon$. En particulier, la suite $(\|A_n\|)_{n\geq 0}$ est bornée, et il existe $C > 0$ tel que $0 \leq \|A_n\| \leq C < +\infty$ pour tout $n \in \mathbb{N}$. Pour $x \in E$ donné, on a

$$\|A_n x - A_m x\|_F \leq \varepsilon \|x\|_E \tag{2.5}$$

si $n, m \geq N_\varepsilon$. La suite $(A_n x)_{n\geq 0}$ est ainsi une suite de Cauchy dans l'espace de Banach F, et admet donc une limite $a_x \in F$. On peut construire un opérateur limite A en posant $Ax = a_x$. On vérifie facilement que A est linéaire (par unicité de la limite). Par ailleurs, en passant à la limite $m \to +\infty$ dans (2.5), on obtient

$$\|A_n x - Ax\|_F \leq \varepsilon \|x\|_E,$$

et donc, pour $n \geq N_\varepsilon$,

$$\|Ax\|_F \leq \|Ax - A_n x\|_F + \|A_n x\|_F \leq (\varepsilon + C)\|x\|_E.$$

Ainsi, A est dans $\mathcal{L}(E,F)$ et on peut passer à la limite dans (2.4) (ou prendre le supremum sur les $x \in E$ avec $\|x\|_E \leq 1$) et obtenir que, pour tout $\varepsilon > 0$, il existe $N_\varepsilon \in \mathbb{N}$ tel que

$$\|A_n - A\| \leq \varepsilon$$

pour tout $n \geq N_\varepsilon$. Ceci montre bien que $A_n \to A$ dans $\mathcal{L}(E,F)$. □

Finissons cette section en prouvant le résultat suivant :

Proposition 2.7. *Soient V et W deux espaces de Hilbert et $A \in \mathcal{L}(V,W)$ une application linéaire et continue de V dans W. Soit $(u_n)_{n\in\mathbb{N}}$ une suite d'éléments de V qui converge faiblement vers un élément $u \in V$. Alors la suite $(Au_n)_{n\in\mathbb{N}}$ converge faiblement vers Au dans W.*

Démonstration. Soit $w \in W$. Soit $\varphi : v \in V \mapsto \langle Av, w\rangle_W$. Comme $A \in \mathcal{L}(V,W)$, on vérifie facilement que $\varphi \in V'$. D'après la caractérisation équivalente de la convergence faible donnée par la Proposition 1.36, on a alors $\varphi(u_n) \underset{n\to+\infty}{\longrightarrow} \varphi(u)$, ce qui se réécrit

$$\langle Au_n, w\rangle_W \underset{n\to+\infty}{\longrightarrow} \langle Au, w\rangle_W.$$

Cette convergence a lieu pour tout $w \in W$, ce qui implique bien que la suite $(Au_n)_{n\in\mathbb{N}}$ converge faiblement vers Au dans W. □

2.1.2 Injectivité et surjectivité

En dimension finie, on a le résultat classique suivant :

Proposition 2.8. *Soit E un espace vectoriel de dimension finie et A une application linéaire de E dans E. Alors A est continue, et de plus les 3 propositions suivantes sont équivalentes :*

- *A est injective sur E.*
- *A est surjective sur E.*
- *A est bijective de E dans E.*

Comme le montre l'exercice suivant, la situation en dimension infinie est plus complexe : une application linéaire continue peut être injective sans être surjective.

Exemple 2.9. *L'opérateur de shift à droite* (2.2) *est injectif, mais pas surjectif car $(1, 0, \dots) \notin \mathrm{Ran}(\tau_{\mathrm{d}})$. L'opérateur de shift à gauche* (2.3) *est surjectif, mais pas injectif.*

Enonçons une propriété qui nous sera utile par la suite (la preuve, omise, repose sur le lemme de Baire, voir par exemple [10]).

Proposition 2.10. *Si $A \in \mathcal{L}(E, F)$ et A est une bijection de E vers F, alors $A^{-1} \in \mathcal{L}(F, E)$.*

2.1.3 Adjoint

Définition 2.11. *Soit H un espace de Hilbert, muni d'un produit scalaire (complexe) noté $\langle \cdot, \cdot \rangle$, et $T \in \mathcal{L}(H)$. L'adjoint de T est l'opérateur T^* défini par*

$$\forall u \in H,\ \forall v \in H, \qquad \langle T^* u, v \rangle = \langle u, Tv \rangle.$$

On dit que T est auto-adjoint *si $T^* = T$.*

Exemple 2.12. *On vérifie facilement que l'adjoint sur $\ell^2(\mathbb{N}, \mathbb{C})$ de l'opérateur τ_{d} de shift à droite* (2.2) *est l'opérateur τ_{g} de shift à gauche* (2.3) *(et réciproquement).*

Exercice 6. *Soit $V \in \mathrm{L}^\infty([a, b], \mathbb{R})$. Vérifier que l'opérateur $T : \mathrm{L}^2([a, b]) \to \mathrm{L}^2([a, b])$ défini par $Tf(x) = V(x)\, f(x)$ est autoadjoint.*

Exercice 7 (Opérateurs de Hilbert-Schmidt)**.** *Soit $H = \mathrm{L}^2(\mathbb{R}^d, \mathbb{C})$ et $K \in \mathrm{L}^2(\mathbb{R}^{2d}, \mathbb{C})$. On considère l'opérateur intégral $\widehat{K} : H \to H$ défini par*

$$\widehat{K} f(x) = \int_{\mathbb{R}^d} K(x, y) f(y)\, dy.$$

On dit que K est le noyau *de $\widehat{K}$. Montrer que $\widehat{K} \in \mathcal{L}(H)$ et que*

$$\left\| \widehat{K} \right\| \leq \|K\|_{\mathrm{L}^2} = \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |K(x, y)|^2\, dx\, dy \right)^{1/2}.$$

Montrer également que $\widehat{K}^*$ est un opérateur intégral de noyau $\overline{K(y, x)}$.

On pourra vérifier en exercice la propriété suivante (voir [7, Section 4.2]).

Proposition 2.13. *Si $T \in \mathcal{L}(H)$ alors $T^* \in \mathcal{L}(H)$, $\|T^*\| = \|T\|$ et $T^{**} = T$. Si T_1 et T_2 sont dans $\mathcal{L}(H)$, alors $(T_1T_2)^* = T_2^*T_1^*$.*

Le résultat suivant sera utile dans la suite :

Proposition 2.14. *Soit $T \in \mathcal{L}(H)$ et $\lambda \in \mathbb{C}$. Alors*

$$\Big(\mathrm{Ran}(\lambda - T)\Big)^{\perp} = \mathrm{Ker}\left(\overline{\lambda} - T^*\right). \tag{2.6}$$

Démonstration. Par définition, on a, pour tout x et y dans H, que

$$\langle(\lambda - T)x, y\rangle = \langle x, (\overline{\lambda} - T^*)y\rangle.$$

Soit $\widetilde{x} \in \mathrm{Ran}(\lambda - T)$ et $y \in \mathrm{Ker}(\overline{\lambda} - T^*)$. Il existe x tel que $\widetilde{x} = (\lambda - T)x$ et ainsi $\langle\widetilde{x}, y\rangle = \langle x, (\overline{\lambda} - T^*)y\rangle = 0$. Ceci montre que $\mathrm{Ker}(\overline{\lambda} - T^*) \subset \Big(\mathrm{Ran}(\lambda - T)\Big)^{\perp}$. On montre l'inclusion inverse. Soit $y \in \Big(\mathrm{Ran}(\lambda - T)\Big)^{\perp}$. Pour tout $x \in H$, on a $\langle y, (\lambda - T)x\rangle = 0 = \langle(\overline{\lambda} - T^*)y, x\rangle$. Ceci étant vrai pour tout $x \in H$, on obtient $(\overline{\lambda} - T^*)y = 0$ et donc l'inclusion contraire $\Big(\mathrm{Ran}(\lambda - T)\Big)^{\perp} \subset \mathrm{Ker}(\overline{\lambda} - T^*)$. □

2.2 Théorie spectrale des opérateurs linéaires et continus

On va à présent étudier de plus près l'inversibilité d'opérateurs linéaires et continus d'un espace de Banach E dans lui-même. De telles considérations sont particulièrement intéressantes lorsqu'il s'agit de résoudre une équation du type

$$(\lambda\,\mathrm{Id} - A)u = f$$

avec $u, f \in E$ et $\lambda \in \mathbb{C}$. En effet, si l'inverse de l'opérateur $\lambda\,\mathrm{Id} - A$ est bien défini, alors $u = (\lambda\,\mathrm{Id} - A)^{-1}f$ est l'unique solution de cette équation.

2.2.1 Théorie générale

On peut définir aisément l'inverse d'un opérateur $\mathrm{Id} - A$ lorsque A est de norme suffisamment petite par le biais d'une série infinie. Plus précisément, la notion pertinente est le rayon spectral.

Lemme 2.15 (Rayon spectral)**.** *Soit $A \in \mathcal{L}(E)$. Alors la limite suivante existe :*

$$r(A) = \lim_{n\to+\infty} \|A^n\|^{1/n} = \inf_{n\geq 1} \|A^n\|^{1/n},$$

et est appelée rayon spectral. *On a en particulier $r(A) \leq \|A\|$.*

On peut avoir $r(A) < \|A\|$. Le cas le plus frappant est celui des opérateurs *nilpotents*, c'est-à-dire tels qu'il existe $N \in \mathbb{N}$ tel que $A^N = 0$. Dans ce cas, $r(A) = 0$. Par exemple, l'opérateur sur $E = \mathbb{R}^2$ dont la représentation matricielle dans la base canonique est

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

est tel que $\|A\| = 1$ mais $A^2 = 0$ et donc $r(A) = 0$.

Démonstration. On suit la preuve de [8, Section I.4.2]. Pour $n, m \in \mathbb{N}$, on a clairement

$$\|A^{n+m}\| \leq \|A^n\| \, \|A^m\|, \qquad \|A^n\| \leq \|A\|^n, \tag{2.7}$$

avec la convention $A^0 = \mathrm{Id}$. Ces inégalités proviennent de l'inégalité générale $\|AB\| \leq \|A\| \, \|B\|$ pour $A, B \in \mathcal{L}(E)$ (voir Exercice 2). Notons

$$a_n = \ln \|A^n\|.$$

Alors $a_n/n \leq \ln \|A\|$. Il s'agit de montrer que la suite $(a_n/n)_{n\geq 1}$ converge.

Les inégalités (2.7) montrent que $a_{n+m} \leq a_n + a_m$. Pour $m \in \mathbb{N}^*$ donné, considérons la division euclidienne de n par m : $n = qm + r$ avec $q, r \in \mathbb{N}$ et $r < m$. On montre alors que $a_n \leq q a_m + a_r$ et ainsi

$$\frac{a_n}{n} \leq \frac{q}{n} \, a_m + \frac{1}{n} \, a_r.$$

Lorsque $n \to +\infty$, $q/n \to 1/m$ alors que les valeurs de r sont limitées à $0, \ldots, m-1$. Ainsi,

$$\sup_{r=0,\ldots,m-1} \frac{1}{n} \, a_r \longrightarrow 0$$

lorsque $n \to +\infty$, et donc

$$\limsup_{n\to+\infty} \frac{a_n}{n} \leq \frac{a_m}{m}.$$

Comme m est arbitraire, on en déduit que

$$\limsup_{n\to+\infty} \frac{a_n}{n} \leq \inf_{m\geq 1} \frac{a_m}{m}.$$

Par ailleurs, on a trivialement

$$\liminf_{n\to+\infty} \frac{a_n}{n} \geq \inf_{m\geq 1} \frac{a_m}{m},$$

et on en déduit donc

$$\limsup_{n\to+\infty} \frac{a_n}{n} \leq \inf_{m\geq 1} \frac{a_m}{m} \leq \liminf_{n\to+\infty} \frac{a_n}{n}.$$

Les inégalités ci-dessus sont finalement des égalités, ce qui montre que la suite $(a_n/n)_{n\geq 1}$ est bien convergente, et qu'elle converge vers $\inf_{m\geq 1}(a_m/m)$. □

Exercice 8. *Soient τ_{d} et τ_{g} les opérateurs de shift définis par* (2.2) *et* (2.3)*. Montrer que $r(\tau_{\mathrm{d}}) = r(\tau_{\mathrm{g}}) = 1$.*

Le lemme suivant, simple, va nous être utile dans la suite :

Lemme 2.16. *Soit $A \in \mathcal{L}(E)$ et soit $z \in \mathbb{C}$. La série $\sum_n z^n A^n$ est convergente dans $\mathcal{L}(E)$ si et seulement si $|z| < 1/r(A)$.*

On remarque facilement que, si $|z| < 1/\|A\|_E \leq 1/r(A)$, alors la série $\sum_n z^n A^n$ est normalement convergente, c'est à dire que $\sum_n |z|^n \, \|A^n\|_E < \infty$. Comme E est un Banach, l'espace $\mathcal{L}(E)$ est un espace de Banach (cf. la Proposition 2.6), et d'après le cours de première année [5], on sait que, si la série est normalement convergente, alors elle est convergente dans $\mathcal{L}(E)$. La preuve ci-dessous montre que le résultat est aussi vrai pour un ensemble de z un peu plus général, i.e. que les z tels que $1/\|A\|_E < |z| < 1/r(A)$ fonctionnent aussi.

Démonstration. Comme E est un Banach, l'espace $\mathcal{L}(E)$ est un espace de Banach (cf. la Proposition 2.6). D'après le cours de première année [5], on sait que, si la série est normalement convergente, i.e. si $\sum_n |z|^n \, \|A^n\|_E < \infty$, alors la série $\sum_n z^n A^n$ est convergente dans $\mathcal{L}(E)$.

Supposons $|z| < 1/r(A)$. Soit $\varepsilon > 0$. Par définition du rayon spectral, il existe N_ε tel que, pour tout $n > N_\varepsilon$, on a $\|A^n\|^{1/n} \leq r(A) + \varepsilon$, donc $|z|^n \, \|A^n\|_E \leq |z|^n (r(A)+\varepsilon)^n$. Grace à l'hypothèse sur z, on peut trouver ε tel que $|z| \, (r(A)+\varepsilon) < 1$. La série $\sum_n |z|^n \, \|A^n\|_E$ est donc convergente, donc la série $\sum_n z^n A^n$ est convergente dans $\mathcal{L}(E)$.

Supposons maintenant que la série $\sum_n z^n A^n$ est convergente dans $\mathcal{L}(E)$. Ceci implique que $z^n A^n$ converge vers 0 dans $\mathcal{L}(E)$: $\lim_n |z|^n \|A^n\|_E = 0$. Or $r(A) = \inf_n \|A^n\|_E^{1/n}$. On a donc $(|z|r(A))^n \leq |z|^n \|A^n\|_E$, et donc $\lim_n (|z|r(A))^n = 0$. Ceci implique que $|z|r(A) < 1$, d'où $|z| < 1/r(A)$. □

On peut à présent définir l'inverse de l'opérateur $\mathrm{Id} - A$ lorsque A a un rayon spectral strictement plus petit que 1.

Lemme 2.17 (Série de Neumann)**.** *Soit $A \in \mathcal{L}(E)$ tel que $r(A) < 1$. Alors l'opérateur $\mathrm{Id} - A$ est bijectif de E sur E, vérifie $(\mathrm{Id} - A)^{-1} \in \mathcal{L}(E)$ et*

$$(\mathrm{Id} - A)^{-1} = \sum_{n=0}^{+\infty} A^n. \tag{2.8}$$

Démonstration. Le lemme 2.16 montre que, pour tout z tel que $|z| < 1/r(A)$, la série $\sum_{n=0}^{+\infty} z^n A^n$ converge dans $\mathcal{L}(E)$. C'est donc en particulier le cas pour $z = 1$, ce qui indique que la série du membre de droite de (2.8) est une série convergente dans $\mathcal{L}(E)$.

On écrit ensuite que, pour tout N, on a

$$(\mathrm{Id} - A) \sum_{n=0}^{N} A^n = \mathrm{Id} - A^{N+1}. \tag{2.9}$$

On passe à la limite $N \to \infty$. Le membre de gauche converge vers $(\mathrm{Id} - A) \sum_{n=0}^{+\infty} A^n$. Pour étudier le membre de droite, on utilise le fait que $\|A^N\|^{1/N} \to r(A) < 1$. Il existe donc $\varepsilon > 0$ et N_ε tel que, pour tout $n > N_\varepsilon$, on a $\|A^N\|^{1/N} \leq 1 - \varepsilon$, si bien que $\|A^N\| \leq (1 - \varepsilon)^N$, et donc $\lim_{N\to\infty} \|A^N\| = 0$. On peut maintenant passer à la limite $N \to \infty$ dans (2.9), ce qui donne $(\mathrm{Id} - A) \sum_{n=0}^{\infty} A^n = \mathrm{Id}$, et donc le résultat escompté. □

Théorème-Définition 2.18. *Soit E un espace de Banach et $T \in \mathcal{L}(E)$. D'après la proposition 2.10, si $\lambda - T$ est bijectif, alors son inverse $(\lambda - T)^{-1}$ est continu.*

1. *On appelle* ensemble résolvant *de T l'ensemble*

$$\rho(T) = \Big\{\lambda \in \mathbb{C}, \quad \lambda - T \text{ est bijectif}\Big\}.$$

   L'ensemble résolvant $\rho(T)$ est un ouvert de $\mathbb{C}$.

2. *Pour $\lambda \in \rho(T)$, on note $R(\lambda) = (\lambda - T)^{-1}$. La famille d'opérateurs linéaires continus $(R(\lambda))_{\lambda\in\rho(T)}$ est appelée la résolvante de T. La fonction $\lambda \mapsto R(\lambda)$ est analytique de $\rho(T)$ dans $\mathcal{L}(E)$ et on a, pour tout $(\lambda, \mu) \in \rho(T) \times \rho(T)$, l'identité de la résolvante*

$$R(\lambda) - R(\mu) = (\mu - \lambda)R(\lambda)R(\mu).$$

3. *On appelle* spectre *de T l'ensemble*

$$\sigma(T) = \mathbb{C} \setminus \rho(T) = \Big\{\lambda \in \mathbb{C}, \quad \lambda - T \text{ non bijectif}\Big\}.$$

   L'ensemble $\sigma(T)$ est un compact de $\mathbb{C}$.

4. *On a*

$$\sigma(T) \subset \overline{D(0, r(T))},$$

   où $\overline{D(0, r(T))}$ est le disque fermé centré en 0 et de rayon $r(T)$. On a aussi que

$$\sigma(T) \cap C(0, r(T)) \neq \emptyset$$

   où $C(0, r(T))$ est le cercle de centre 0 et de rayon $r(T)$. En particulier le spectre d'un opérateur lineaire et continue n'est jamais vide.

5. *L'ensemble* $\sigma(T)$ *se décompose en l'union disjointe*

$$\sigma(T) = \sigma_{\mathrm{p}}(T) \cup \sigma_{\mathrm{r}}(T) \cup \sigma_{\mathrm{c}}(T),$$

avec

$$\sigma_{\mathrm{p}}(T) = \Big\{\lambda \in \mathbb{C}, \quad \lambda - T \text{ non injectif}\Big\},$$

$$\sigma_{\mathrm{r}}(T) = \Big\{\lambda \in \mathbb{C}, \quad \lambda - T \text{ injectif et } \overline{(\lambda - T)E} \neq E\Big\},$$

et

$$\sigma_{\mathrm{c}}(T) = \Big\{\lambda \in \mathbb{C}, \quad \lambda - T \text{ injectif et } (\lambda - T)E \neq \overline{(\lambda - T)E} = E\Big\}.$$

L'ensemble $\sigma_{\mathrm{p}}(T)$ *est appelé le* spectre ponctuel *de* T, $\sigma_{\mathrm{c}}(T)$ *le* spectre continu *de* T, $\sigma_{\mathrm{r}}(T)$ *le* spectre résiduel *de* T.

Notons que les trois types de spectre définis ci-dessus ont été classés par ordre croissant de défaut d'inversibilité :

— pour le spectre ponctuel, on a un défaut d'injectivité ;
— pour le spectre résiduel, on a un défaut majeur de surjectivité : même en prenant l'adhérence de l'image de E, on ne retrouve pas E ;
— pour le spectre continu, l'inverse est bien défini sur un sous-ensemble dense de F, mais n'est pas continu. Montrons ceci par l'absurde.
L'opérateur linéaire $\lambda - T$ est bijectif de E sur $(\lambda - T)E$. On introduit son inverse $B : (\lambda - T)E \to E$, qui est défini sur un sous-ensemble dense de E. Supposons B continu de $(\lambda - T)E$ sur E. On peut alors l'étendre par continuité comme un opérateur de E sur E. Soit $y \in E$ et $u = By$ (qui existe car B est maintenant défini sur tout E). Montrons que $y = (\lambda - T)u$:
  — Si $y \in (\lambda - T)E$, c'est évident.
  — Sinon, on sait qu'il existe une suite $y_n \in (\lambda - T)E$ telle que $y_n \to y$. Puisque $y_n \in (\lambda - T)E$, il existe $u_n \in E$ tel que $y_n = (\lambda - T)u_n$, et donc $u_n = By_n$. La suite y_n est convergente, donc de Cauchy. Puisque B est continu, on voit que u_n est aussi de Cauchy, donc convergente. Par définition, on a $u = By = \lim_n u_n$. On peut donc passer à la limite dans l'égalité $y_n = (\lambda - T)u_n$ (puisque $\lambda - T$ est continu), ce qui donne $y = (\lambda - T)u$.

  On vient donc de démontrer que, pour tout $y \in E$, il existe $u \in E$ tel que $y = (\lambda - T)u$, ce qui donne $(\lambda - T)E = E$. On obtient donc une contradiction.

Démonstration. Soit $\lambda \in \mathbb{C}$ tel que $|\lambda| > r(T)$. On écrit

$$\lambda - T = \lambda\left(\mathrm{Id} - \frac{T}{\lambda}\right)$$

et $r(T/\lambda) = r(T)/|\lambda| < 1$. En utilisant le lemme 2.17, on voit que $\lambda - T$ est inversible, donc $\lambda \in \rho(T)$. Il en découle que

$$\sigma(T) \subset \overline{D(0, r(T))}.$$

Soit maintenant $\mu \in \rho(T)$. On écrit

$$\lambda - T = \mu - T + (\lambda - \mu)\mathrm{Id} = (\mu - T)\Big(\mathrm{Id} + (\lambda - \mu)(\mu - T)^{-1}\Big). \tag{2.10}$$

Donc, si $|\lambda - \mu|\, r\big((\mu - T)^{-1}\big) < 1$, alors $\lambda - T$ est inversible (en vertu du lemme 2.17). On en déduit que $\rho(T)$ est un ouvert de $\mathbb{C}$.

Comme $\sigma(T) = \mathbb{C} \setminus \rho(T)$, on obtient que $\sigma(T)$ est un fermé de $\mathbb{C}$. Comme $\sigma(T)$ est borné, c'est un compact de $\mathbb{C}$.

La relation (2.10) montre que $R(\lambda)$ est analytique dans $\rho(T)$.

En multipliant les deux membres de l'égalité

$$(\lambda - T) = (\mu - T) + (\lambda - \mu)\mathrm{Id}$$

à gauche par $R(\lambda)$ et à droite par $R(\mu)$, on obtient l'identité de la résolvante.

Supposons que $\sigma(T) \cap C(0, r(T)) = \emptyset$. Comme $\sigma(T)$ est compact, il existe $\varepsilon \in]0, r(T)[$ tel que

$$\mathbb{C} \setminus \overline{D(0, r(T) - \varepsilon)} \subset \rho(T).$$

Comme $R(\lambda)$ est analytique sur $\rho(T)$, il en résulte que $f(z) = R(1/z)$ est analytique sur $D(0, (r(T) - \varepsilon)^{-1})$. Or, un calcul explicite montre que le développement en série entière de $f(z)$ en 0 est donné par

$$f(z) = z \sum_{n \in \mathbb{N}} z^n T^n.$$

Sur l'ensemble $\mathcal{C} = \{z \in \mathbb{C};\ 1/r(T) < |z| < 1/(r(T) - \varepsilon)\}$, on obtient donc que $f(z)$ est analytique, alors que la série est divergente, d'après le lemme 2.16. On obtient donc une contradiction. □

Remarque 2.19. *Notons que $\sigma_{\mathrm{p}}(T)$ est l'ensemble des valeurs propres de T, i.e. l'ensemble des $\lambda \in \mathbb{C}$ tels qu'il existe $u \in E \setminus \{0\}$ tel que*

$$Tu = \lambda u.$$

En dimension finie, un opérateur linéaire injectif est bijectif. Ainsi,

$$\sigma(T) = \sigma_{\mathrm{p}}(T)$$

est simplement l'ensemble des valeurs propres de T dans ce cas.

Prouvons ici le lemme suivant qui sera utile par la suite.

Lemme 2.20. *Soit $T \in \mathcal{L}(E)$. Soit $(\lambda_k)_{k \geq 1}$ une suite de $\sigma_{\mathrm{p}}(T)$ de valeurs propres toutes distinctes, et soit $(u_k)_{k \geq 1}$ une suite de vecteurs propres associés. Alors les vecteurs $(u_k)_{k \geq 1}$ sont linéairement indépendants.*

Démonstration. On procède par récurrence. On suppose que les vecteurs $u_1, \dots, u_n$ sont indépendants. Si, au rang $n+1$, l'hypothèse de récurrence n'est pas vraie, alors il existe $(\alpha_k)_{1\leq k\leq n}$ tels que $u_{n+1} = \sum_{k=1}^n \alpha_k u_k$. Alors

$$Tu_{n+1} = \sum_{k=1}^n \alpha_k \lambda_k u_k = \lambda_{n+1} u_{n+1} = \lambda_{n+1} \sum_{k=1}^n \alpha_k u_k.$$

Par hypothèse de récurrence, la famille $(u_1, \dots, u_n)$ est libre, donc $\lambda_{n+1}\alpha_k = \alpha_k \lambda_k$ pour tout $1 \leq k \leq n$. Les valeurs propres étant distinctes deux à deux, on a ainsi $\alpha_k = 0$, ce qui donne $u_{n+1} = 0$, ce qui est contradictoire. On a donc démontré l'hypothèse de récurrence au rang $n+1$. □

Remarque 2.21 (Autre décomposition du spectre)**.** *Dans certains cas, il est plus commode de décomposer $\sigma(T)$ sous la forme $\sigma(T) = \sigma_{\rm d}(T) \cup \sigma_{\rm ess}(T)$, où $\sigma_{\rm d}(T) \subset \sigma_{\rm p}(T)$ est le* spectre discret*, qui est composé des valeurs propres isolées de multiplicité finie :*

$$\sigma_{\rm d}(T) = \Big\{ \lambda \in \mathbb{C} \;\Big|\; 0 < \dim(\mathrm{Ker}(\lambda - T)) < +\infty,\ \exists \varepsilon > 0,\ \ ]\lambda-\varepsilon, \lambda+\varepsilon[\cap \sigma(T) = \{\lambda\} \Big\}.$$

Donnons à présent quelques exemples de spectre résiduel et continu, afin de donner un début d'intuition sur ces notions.

Exercice 9 (Spectre résiduel)**.** *On considère l'opérateur de shift à droite $\tau_{\rm d}$ dans $\ell^2(\mathbb{N}, \mathbb{C})$ défini par* (2.2)*.*

1. *Vérifier que $\sigma_{\rm p}(\tau_{\rm d}) = \emptyset$ et que $\lambda - \tau_{\rm d}$ est injectif pour tout $\lambda \in \mathbb{C}$.*
2. *Montrer que $0 \in \sigma_{\rm r}(\tau_{\rm d})$.*
3. *Montrer que $\{\lambda \in \mathbb{C},\ |\lambda| < 1\} \subset \sigma_{\rm r}(\tau_{\rm d})$. Indication : considérer $x_\lambda = \left(1, \overline{\lambda}, \overline{\lambda}^2, \dots\right)$ et vérifier que $x_\lambda \in (\mathrm{Ran}(\lambda - \tau_{\rm d}))^\perp$.*

Exercice 10 (Spectre continu)**.** *Soit $a < b$ deux réels, $E = \mathrm{L}^2([a,b], \mathbb{C})$ et $T \in \mathcal{L}(E)$ défini par*

$$Tf(x) = x\, f(x).$$

Montrer que $\sigma(T) = \sigma_{\rm c}(T) = [a,b]$, en suivant les étapes ci-dessous :

1. *Montrer que $\sigma(T) \subset [a,b]$.*
2. *Montrer que $\sigma(T) = [a,b]$ (en supposant qu'il existe $\lambda \in [a,b]$ tel que $\lambda - T$ soit inversible, et en considérant $\varphi \in \mathbb{C}^\infty([a,b], \mathbb{C})$ valant 1 au voisinage de λ).*
3. *Montrer que $\sigma(T) = \sigma_{\rm c}(T)$. Pour cela, établir d'abord que $\sigma_{\rm p}(T) = \emptyset$, puis prouver que $\overline{\mathrm{Ran}(\lambda - T)} = E$ pour tout $\lambda \in [a,b]$. Pour ce dernier point, pour $f \in E$ donnée, considérer la suite $(\varphi_n)_{n\geq 1}$ de E définie par*

$$\varphi_n(x) = \begin{cases} \dfrac{f(x)}{\lambda - x} & \text{si } |x-\lambda| \geq \dfrac{1}{n} \text{ et } x \in [a,b], \\ 0 & \text{sinon.} \end{cases}$$

2.2.2 Cas des opérateurs lineaires, continus et autoadjoints

Les opérateurs lineaires, continus et auto-adjoints ont des propriétés intéressantes, qui se traduisent sur leur spectre.

Proposition 2.22. *Soit H un espace de Hilbert et $T \in \mathcal{L}(H)$. Si T est auto-adjoint, on a*

$$\sigma(T) \subset \mathbb{R}.$$

De plus, $r(T) = \|T\|$, $\sigma(T) \subset [-\|T\|, \|T\|]$ et l'une au moins des deux extrémités du segment est dans $\sigma(T)$. Enfin, $\sigma_{\mathrm{r}}(T) = \emptyset$ et les vecteurs propres associés à des éléments différents de $\sigma_{\mathrm{p}}(T)$ sont orthogonaux.

Démonstration. Pour prouver ce résultat, nous allons établir plusieurs résultats intermédiaires.

— Commençons par montrer que si $\lambda \in \mathbb{C}$ est tel que $\alpha = |\mathrm{Im}(\lambda)| \neq 0$, alors $\lambda - T$ est inversible.

Montrons tout d'abord que l'opérateur $\lambda - T$ est injectif. En effet, pour tout $x \in H$, on a

$$\langle (\lambda - T)x, x\rangle = -\langle Tx, x\rangle + \mathrm{Re}(\lambda)\,\langle x, x\rangle - \mathrm{i}\,\mathrm{Im}(\lambda)\,\langle x, x\rangle.$$

On voit que $\langle Tx, x\rangle = \overline{\langle x, Tx\rangle} = \overline{\langle T^*x, x\rangle} = \overline{\langle Tx, x\rangle}$. Donc $\langle Tx, x\rangle$ est réel. Il en résulte que

$$|\langle (\lambda - T)x, x\rangle| \geq \alpha \|x\|^2. \tag{2.11}$$

On en déduit par l'inégalité de Cauchy-Schwarz que

$$\|(\lambda - T)x\| \geq \alpha \|x\|. \tag{2.12}$$

Cette inégalité implique que l'opérateur $\lambda - T$ est injectif.

Montrons ensuite que l'opérateur $\lambda - T$ est surjectif. Soit $V = \mathrm{Ran}(\lambda - T)$. Nous allons montrer que $V = H$. Pour cela, montrons tout d'abord que V est fermé dans H. Soit $w_n = (\lambda - T)v_n$ une suite dans V qui converge vers $w \in H$. En utilisant (2.12), on obtient

$$\|w_p - w_q\| \geq \alpha \|v_p - v_q\|.$$

La suite $(w_n)_{n\geq 0}$ est de Cauchy, donc la suite $(v_n)_{n\geq 0}$ aussi. Elle converge donc vers un certain $v \in H$. Par continuité de l'application T,

$$w_n = (\lambda - T)v_n \longrightarrow (\lambda - T)v$$

dans H. Donc $w = (\lambda - T)v$, ce qui prouve que $w \in V$. Donc V est fermé dans H. Montrons enfin que V est dense. Une technique standard pour montrer

cela est de prouver que $V^\perp = \{0\}$ (ce qui donne, grace au lemme 1.13, que $\overline{V} = (V^\perp)^\perp = H$). Soit donc $w \in V^\perp$. Pour tout $v \in H$, on a alors

$$\langle (\lambda - T)v, w\rangle = 0.$$

En particulier, pour $v = w$,

$$\langle (\lambda - T)w, w\rangle = 0.$$

En utilisant (2.11), on obtient $w = 0$, ce qui montre que $V^\perp = \{0\}$ d'où la densité de V dans H. Comme V est dense dans H et fermé dans H, on en déduit que $V = H$, et donc la surjectivité de $\lambda - T$.
Comme l'opérateur $\lambda - T \in \mathcal{L}(H)$ est bijectif, il est inversible (cf. la proposition 2.10). Noter également que l'inégalité (2.12) donne la borne suivante sur la résolvante :

$$\|(\lambda - T)^{-1}\| \leq \frac{1}{|\mathrm{Im}(\lambda)|}.$$

On a donc démontré que $\sigma(T) \subset \mathbb{R}$.

— Le théorème 2.18 implique alors que

$$\sigma(T) \subset \overline{D(0, r(T))} \cap \mathbb{R} = [-r(T), r(T)]$$

et que

$$\sigma(T) \cap C(0, r(T)) = \sigma(T) \cap C(0, r(T)) \cap \mathbb{R} = \sigma(T) \cap \{-r(T), r(T)\} \neq \emptyset.$$

— Nous allons maintenant prouver que $r(T) = \|T\|$. Tout d'abord, notons que $\|T^*T\| \leq \|T\|\,\|T^*\| = \|T\|^2$. Par ailleurs, comme $|\langle x, T^*Tx\rangle| \leq \|T^*T\|\,\|x\|^2$, on a

$$\|T^*T\| \geq \sup_{\|x\|=1} |\langle x, T^*Tx\rangle| = \sup_{\|x\|=1} \|Tx\|^2 = \left(\sup_{\|x\|=1} \|Tx\|\right)^2 = \|T\|^2,$$

ce qui montre que $\|T^2\| = \|T^*T\| = \|T\|^2$. Par récurrence, on a ensuite $\|T^{2^p}\| = \|T\|^{2^p}$. Pour $n \in \mathbb{N}$ quelconque, on considère p tel que $n \leq 2^p$ et on écrit

$$\|T\|^{2^p} = \|T^{2^p}\| \leq \|T^n\|\,\|T^{2^p-n}\| \leq \|T^n\|\,\|T\|^{2^p-n}.$$

Ceci montre que $\|T\|^n \leq \|T^n\|$. L'inégalité contraire étant par ailleurs toujours satisfaite, on en déduit que $\|T\|^n = \|T^n\|$, et donc $\|T^n\|^{1/n} = \|T\|$ pour tout $n \geq 1$. On a donc finalement $r(T) = \lim_{n\to\infty} \|T^n\|^{1/n} = \|T\|$.

— Montrons maintenant que $\sigma_{\rm r}(T) = \emptyset$. Pour ce faire, on considère $\lambda \in \sigma(T) \subset \mathbb{R}$ tel que $\mathrm{Ker}(\lambda - T) = \{0\}$. On a vu (cf. la proposition 2.14) que

$$\Big(\mathrm{Ran}(\lambda - T)\Big)^{\perp} = \mathrm{Ker}(\overline{\lambda} - T^*).$$

Dans le cas présent, ceci implique que $\Big(\mathrm{Ran}(\lambda - T)\Big)^{\perp} = \mathrm{Ker}(\lambda - T) = \{0\}$, ce qui implique (cf. le lemme 1.13) que signifie que $\overline{\mathrm{Ran}(\lambda - T)} = H$ et donc $\lambda \notin \sigma_{\rm r}(T)$.

— Enfin, soient u et v deux vecteurs propres associés respectivement à deux éléments $\lambda \neq \mu$ de $\sigma_{\rm p}(T)$. Alors,

$$\lambda\langle u, v\rangle = \langle Tu, v\rangle = \langle u, Tv\rangle = \mu\langle u, v\rangle.$$

Ceci montre que $\langle u, v\rangle = 0$.

□

Remarque 2.23. *On fait ici le lien entre le spectre résiduel d'un opérateur et le spectre ponctuel de son adjoint.*

La relation (2.6) *montre de manière générale que, pour un opérateur linéaire et continu $T \in \mathcal{L}(E)$, si $\lambda \in \sigma_{\rm r}(T)$, alors $\overline{\lambda} \in \sigma_{\rm p}(T^*)$. Bien sûr, dans le cas des opérateurs autoadjoints, on a $T^* = T$ et donc $\lambda \in \sigma_{\rm r}(T) \cap \sigma_{\rm p}(T) = \emptyset$ par définition des différentes parties du spectre. Ceci montre bien que $\sigma_{\rm r}(T) = \emptyset$ pour des opérateurs autoadjoints.*

Par ailleurs, on peut montrer que, si $\lambda \in \sigma_{\rm p}(T)$, alors $\overline{\lambda} \in \sigma_{\rm p}(T^*) \cup \sigma_{\rm r}(T^*)$.

Exercice 11. *Donner un exemple d'opérateur lineaire et continu tel que $\overline{\lambda} \in \sigma_{\rm p}(T^*)$ lorsque $\lambda \in \sigma_{\rm p}(T)$, et un exemple d'opérateur lineaire et continu tel que $\overline{\lambda} \in \sigma_{\rm r}(T^*)$ lorsque $\lambda \in \sigma_{\rm p}(T)$.*

Exercice 12. *Soit V un espace de Hilbert et soit $T \in \mathcal{L}(V)$ un opérateur lineaire, continu et auto-adjoint. On suppose que $\langle Tu, u\rangle = 0$ pour tout $u \in V$. Montrer qu'alors $T = 0$.*

2.3 Opérateurs compacts

2.3.1 Définition et premières propriétés

Définition 2.24. *Soient E et F deux espaces de Banach et T un opérateur linéaire de E dans F. On dit que l'opérateur T est* compact *si, pour tout $B \subset E$,*

$$B \text{ borné dans } E \quad \Rightarrow \quad T(B) \text{ relativement compact dans } F.$$

On note $\mathcal{K}(E, F)$ l'ensemble des opérateurs compacts de E dans F.

Ainsi, un opérateur compact transforme une suite bornée en une suite convergente (à extraction près).

Proposition 2.25. *Tout opérateur linéaire compact est continu, i.e.* $\mathcal{K}(E,F) \subset \mathcal{L}(E,F)$.

Démonstration. Soit E et F deux espaces de Banach et T un opérateur linéaire compact de E dans F. Soit $\overline{B}_1 = \{x \in E,\ \|x\| \leq 1\}$ la boule unité fermée de E. L'ensemble $\overline{B}_1$ étant borné, son image par T est relativement compacte donc bornée : il existe une constante C telle que

$$\forall x \in \overline{B}_1, \qquad \|Tx\|_F \leq C.$$

On en déduit que

$$\forall x \in E \setminus \{0\}, \qquad \|Tx\|_F = \|x\|_E \left\| T\left(\frac{x}{\|x\|_E}\right)\right\|_F \leq C\|x\|_E.$$

L'opérateur linéaire T est donc continu. □

Nous avons la caractérisation équivalente suivante des opérateurs compacts dans le cas où les espaces E et F sont des opérateurs de Hilbert.

Proposition 2.26. *Soit E et F deux espaces de Hilbert. Soit $T \in \mathcal{L}(E,F)$. Alors les deux propositions suivantes sont équivalentes :*

(i) $T \in \mathcal{K}(E,F)$;

(ii) Pour toute suite $(u_n)_{n\in\mathbb{N}}$ qui converge faiblement vers u dans E, on peut extraire une sous-suite de la suite $(Tu_n)_{n\in\mathbb{N}}$ qui converge fortement vers Tu dans F.

Démonstration. On démontre l'implication (i) ⇒ (ii). Soit $T \in \mathcal{K}(E,F)$. Soit $(u_n)_{n\in\mathbb{N}}$ une suite d'éléments de E qui converge faiblement vers un élément $u \in E$. On utilise la Proposition 2.7 : comme T est un opérateur continu, la suite $(Tu_n)_{n\in\mathbb{N}}$ converge faiblement vers Tu dans F. Par ailleurs, la suite $(u_n)_{n\in\mathbb{N}}$ est bornée, et T est compact, donc on peut extraire une sous-suite de $(Tu_n)_{n\in\mathbb{N}}$ qui converge fortement vers un élément $w \in F$. Comme la convergence forte implique la convergence faible, par unicité de la limite, on a nécessairement $w = Tu$.

On prouve maintenant l'implication (ii) ⇒ (i). Soit $T \in \mathcal{L}(E,F)$ qui vérifie la propriété (ii). Montrons que T est compact. Soit B un sous-ensemble borné de E. Montrons que $T(B)$ est un ensemble relativement compact dans F. Soit $(u_n)_{n\in\mathbb{N}}$ une suite d'éléments de B. Comme B est borné, la suite $(u_n)_{n\in\mathbb{N}}$ l'est aussi, et on peut donc en extraire une sous-suite qui converge faiblement dans E vers un élément $u \in E$. D'après la caractérisation (ii), il existe une extraction φ telle que la suite $(Tu_{\varphi(n)})_{n\in\mathbb{N}}$ converge fortement dans F vers Tu. Ceci montre qu'il existe une sous-suite de $(Tu_n)_{n\in\mathbb{N}}$ qui converge fortement dans F. L'ensemble $T(B)$ est donc relativement compact. □

Exercice 13. *Montrer que les opérateurs suivants sont compacts :*

1. *l'identité de E est compacte si et seulement si E est de dimension finie ;*
2. *si l'un des espaces E ou F est de dimension finie, alors tout opérateur linéaire continu T de E dans F est compact (en particulier, si $T \in \mathcal{L}(E,F)$ avec $\mathrm{Ran}(T)$ de dimension finie, alors $T \in \mathcal{K}(E,F)$) ;*
3. *si T_1 et T_2 sont deux opérateurs linéaires compacts de E dans F, alors T_1+T_2 est un opérateur compact ;*
4. *la restriction d'un opérateur compact $T \in \mathcal{K}(E,F)$ à un sous-espace vectoriel $\widetilde{E}$ de E est compacte.*

Exercice 14. *On considère l'opérateur de l'Exercice 4. Montrer que $K \in \mathcal{K}(E,F)$ en admettant le résultat de compacité suivant, connu sous le nom de lemme d'Ascoli :*

Soit $\mathcal{F}$ un sous-ensemble borné de $F = C^0([0,1],\mathbb{R})$ tel que la propriété d'équicontinuité suivante soit satisfaite : pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que

$$|x - x'| \leq \delta \;\Rightarrow\; \forall u \in \mathcal{F},\ |u(x) - u(x')| \leq \varepsilon.$$

Alors $\mathcal{F}$ est relativement compact dans F.

Théorème 2.27. *Soit E et F deux espaces de Banach. L'ensemble $\mathcal{K}(E,F)$ est un sous-espace vectoriel fermé de l'espace vectoriel $\mathcal{L}(E,F)$.*

Démonstration. Il est facile de montrer que $\mathcal{K}(E,F)$ est un espace vectoriel. Grace à la Proposition 2.25, on sait qu'il est inclus dans $\mathcal{L}(E,F)$. Il reste à prouver que c'est un sous-espace fermé de $\mathcal{L}(E,F)$. Considérons pour cela une suite d'opérateurs compacts $(T_k)_{k\in\mathbb{N}^*}$ qui converge dans $\mathcal{L}(E,F)$ vers un opérateur $T \in \mathcal{L}(E,F)$ et montrons que T est compact. Soit B un borné de E, soit $R > 0$ un réel tel que $B \subset \{x \in E,\ \|x\| \leq R\}$ et soit $(u_n)_{n\in\mathbb{N}}$ une suite de $T(B)$. Il faut montrer que on peut extraire de $(u_n)_{n\in\mathbb{N}}$ une sous-suite convergente (ceci prouvera que $T(B)$ est relativement compact et donc que T est compact).

Soit $(w_n)_{n\in\mathbb{N}}$ une suite d'éléments de B tels que pour tout $n \in \mathbb{N}$, $T(w_n) = u_n$. On va extraire de $(u_n)_{n\in\mathbb{N}}$ une sous-suite convergente en utilisant un procédé diagonal. On pose $\{w_n^0\}_n = \{w_n\}_n$ et on construit, par récurrence sur k, la suite $\{w_n^k\}_n$, qui est une sous-suite de $(w_n^{k-1})_{n\in\mathbb{N}}$ telle que $(T_k(w_n^k))_{n\in\mathbb{N}}$ soit convergente. On utilise pour cela le fait que T_k est un opérateur compact, et que $\{w_n^{k-1}\}_n$, suite extraite de $(w_n)_n$, est bornée. On définit maintenant la suite $(v_n)_{n\in\mathbb{N}}$ par $v_n = w_n^n$. Pour tout $k \in \mathbb{N}^*$, $(v_n)_{n\geq k}$ est une sous-suite de $(w_n^k)_{n\in\mathbb{N}}$: la suite $(T_k(v_n))_{n\in\mathbb{N}}$ est donc convergente.

On pose $\widetilde{u}_n = T(v_n)$. La suite $(\widetilde{u}_n)_{n\in\mathbb{N}}$ est une sous-suite de $(u_n)_{n\in\mathbb{N}}$. On va montrer qu'elle est de Cauchy. Soit $\varepsilon > 0$ et $k \in \mathbb{N}^*$ tel que

$$\|T - T_k\|_{\mathcal{L}(E,F)} \leq \frac{\varepsilon}{3R}.$$

Soit ensuite $N \geq 0$ tel que $\forall q > p \geq N$,

$$\|T_k(v_p) - T_k(v_q)\|_F \leq \frac{\varepsilon}{3}.$$

Il vient que, pour tout $q > p \geq N$,

$$\begin{aligned} \|\widetilde{u}_p - \widetilde{u}_q\| &= \|T(v_p) - T(v_q)\|_F \\ &\leq \|T(v_p) - T_k(v_p)\|_F + \|T_k(v_p) - T_k(v_q)\|_F + \|T_k(v_q) - T(v_q)\|_F \\ &\leq \|T - T_k\|_{\mathcal{L}(E,F)} \left(\|v_p\|_E + \|v_q\|_E\right) + \|T_k(v_p) - T_k(v_q)\|_F \\ &\leq \varepsilon. \end{aligned}$$

La suite $(\widetilde{u}_n)_{n\in\mathbb{N}}$ est donc de Cauchy. Ceci conclut la preuve. □

Une des conséquences importantes de ce résultat est que, si T est la limite d'une suite d'opérateurs $(T_n)_{n\geq 0}$ de rang fini (*i.e.* tels que la dimension de $\mathrm{Ran}(T_n)$ est finie), au sens où

$$\|T_n - T\| \longrightarrow 0$$

où la norme est définie en (2.1), alors l'opérateur limite T est compact. En général, la réciproque est fausse : on ne peut pas approcher n'importe quel opérateur compact par une suite d'opérateurs de rang fini. Cette réciproque est cependant vraie si on considère $\mathcal{K}(E, F)$ avec F un espace de Hilbert (cf. [2, Section VI.1] ou la Remarque 2.38 pour le cas où $E = F$ est un espace de Hilbert).

Proposition 2.28. *Soient E, F et G trois espaces de Banach, et soient $T_1 \in \mathcal{L}(E, F)$ et $T_2 \in \mathcal{L}(F, G)$.*

Si T_1 est compact, ou bien si T_2 est compact, alors l'application $T_2 \circ T_1$ est compacte : $T_2 \circ T_1 \in \mathcal{K}(E, G)$.

Démonstration. On suppose que $T_1 \in \mathcal{L}(E, F)$ et $T_2 \in \mathcal{K}(F, G)$. Comme T_1 est continue, l'image par T_1 de la boule unité de E, qu'on note $T_1(B_E)$, est bornée. Comme T_2 est linéaire compacte, l'image par T_2 d'un ensemble borné est relativement compacte dans G. Donc $T_2 \circ T_1(B_E)$ est relativement compacte dans G, et $T_2 \circ T_1$ est une application compacte.

Supposons maintenant que $T_1 \in \mathcal{K}(E, F)$ et $T_2 \in \mathcal{L}(F, G)$. Soit $w_n = T_2 \circ T_1(u_n)$ une suite d'éléments de $T_2 \circ T_1(B_E)$, avec $u_n \in B_E$. On pose $v_n = T_1(u_n) \in F$. Comme T_1 est compacte, on peut extraire de v_n une sous-suite convergente dans F, qu'on note $v_{\varphi(n)}$, avec $\lim_{n\to\infty} v_{\varphi(n)} = v$. Par conséquent, comme T_2 est continue, on a

$$\lim_{n\to\infty} w_{\varphi(n)} = \lim_{n\to\infty} T_2(v_{\varphi(n)}) = T_2(v).$$

On peut donc extraire de toute suite de $T_2 \circ T_1(B_E)$ une sous-suite convergente : donc $T_2 \circ T_1$ est une application compacte. □

Concluons enfin avec quelques exercices d'application.

Exercice 15 (Opérateurs de Hilbert-Schmidt)**.** *Montrer que l'opérateur $\widehat{K}$ de l'Exercice 7 est compact.*

Exercice 16. *Soit V un espace de Hilbert de dimension infinie. Montrer que, si $A \in \mathcal{K}(V, V)$, alors A n'est pas bijectif.*

Exercice 17. *Soit $u = (u_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ une suite à valeur réelle. On considère l'ensemble $\ell_2 = \{u \in \mathbb{R}^{\mathbb{N}};\ \sum_{i\geq 0} u_i^2 < +\infty\}$ des suites de carré sommable, qu'on munit du produit scalaire $\langle u, v\rangle = \sum_{i\geq 0} u_i v_i$.*

Soit $(a_i)_{i\geq 0}$ une suite de réels bornés : $|a_i| \leq C < +\infty$ pour tout $i \geq 0$. On définit l'application linéaire A sur ℓ_2 par $Au = (a_i u_i)_{i\geq 0}$. Montrer que $Au \in \ell_2$ et que A est continue. Montrer que A est compacte si et seulement si $\lim_{i\to+\infty} a_i = 0$ (Indication : pour montrer que $\lim_{i\to+\infty} a_i = 0$ implique A est compacte, on pourra utiliser un principe d'extraction diagonale).

Proposition 2.29. *Soit V un espace de Hilbert et $A \in \mathcal{K}(V, V)$. Alors* $\mathrm{Ker}(\mathrm{Id} - A)$ *est de dimension finie.*

Démonstration. Soit $E_1 = \mathrm{Ker}(\mathrm{Id} - A)$. Montrons que la boule unité fermée de E_1 est compacte. Soit $v \in \mathrm{Ker}(\mathrm{Id} - A)$ avec $\|v\| \leq 1$: on a donc $v = Av$, donc $v \in A(B_V)$, et ainsi $B_{E_1} \subset A(B_V)$. Comme A est compacte, $A(B_V)$ est relativement compacte, et donc B_{E_1} est relativement compact. Comme B_{E_1} est fermée, on a donc que B_{E_1} est compacte. En application de la proposition 1.31, on a donc que E_1 est de dimension finie. □

2.3.2 Le théorème de Rellich

Définition 2.30. *Soient V et H deux espaces de Hilbert avec $V \subset H$. On note respectivement $\langle\cdot,\cdot\rangle_V$ et $\langle\cdot,\cdot\rangle_H$ leur produit scalaire. On dit que l'injection $V \subset H$ est compacte si l'application*

$$\begin{array}{rcl} \mathcal{I} : V & \longrightarrow & H \\ u & \longmapsto & u \end{array}$$

est continue et compacte, autrement dit :

- *il existe C tel que, pour tout $u \in V$, on a $\|u\|_H \leq C\ \|u\|_V$;*
- *de toute suite bornée de V (pour la norme $\|\cdot\|_V$), on peut extraire une sous-suite convergente dans H (pour la norme $\|\cdot\|_H$).*

On va à présent énoncer un résultat de compacité important (et très utile dans l'étude des équations aux dérivées partielles).

Théorème 2.31. *Soit Ω un ouvert* borné *de $\mathbb{R}^d$. L'injection canonique de* $\mathrm{H}_0^1(\Omega)$ *dans* $\mathrm{L}^2(\Omega)$ *est compacte.*

Un des intérêts de ce résultat est que, si on arrive à obtenir une borne (en norme $\mathrm{H}^1(\Omega)$) sur une suite de fonctions approchant la solution d'une équation (par exemple, en montrant qu'une énergie est uniformément bornée), alors on peut extraire de cette suite une sous-suite convergente (en norme $\mathrm{L}^2(\Omega)$). Cette limite est alors un candidat naturel pour être une solution de l'équation.

Dans ce chapitre, ce résultat va nous permettre de montrer que les inverses de certains opérateurs sont compacts, ce qui permettra de décrire complètement le spectre de l'opérateur en question.

Démonstration. La preuve comprend trois étapes.

— On commence par traiter le cas où $\Omega =]0, \pi[$. On note $e_k(x) = \sqrt{2/\pi}\,\sin(kx)$ le k-ième mode de Fourier valant 0 au bord de Ω. On note que $e_k \in \mathrm{H}^1_0(0,\pi)$, $\|e_k\|_{\mathrm{L}^2} = 1$ et $\|e_k\|^2_{\mathrm{H}^1} = 1 + k^2$.
En utilisant la transformée de Fourier, on peut montrer (et ce sera admis ici) qu'on peut caractériser l'espace $\mathrm{L}^2(0,\pi)$ par

$$\mathrm{L}^2(0,\pi) = \left\{ u(x) = \sum_{k=1}^{+\infty} c_k\, e_k(x), \quad \sum_{k=1}^{+\infty} |c_k|^2 < +\infty \right\} \tag{2.13}$$

avec $\|u\|_{\mathrm{L}^2} = \left(\sum_{k=1}^{+\infty} |c_k|^2 \right)^{1/2}$. Chaque $u \in \mathrm{L}^2(0,\pi)$ se décompose donc comme une série convergente (dans $\mathrm{L}^2(0,\pi)$) de terme général $c_k\, e_k$ où $(c_k)_{k\geq 1} \in \ell_2$. Réciproquement, pour chaque suite $\{c_k\}_{k\geq 1}$ de réels tels que $\sum_{k=1}^{+\infty} |c_k|^2 < +\infty$, on peut considérer la série $\sum_{k=1}^{+\infty} c_k\, e_k$, qui se trouve être convergente dans $\mathrm{L}^2(0,\pi)$.

On montre maintenant que

$$\mathrm{H}^1_0(0,\pi) = \left\{ u(x) = \sum_{k=1}^{+\infty} c_k\, e_k(x), \quad \sum_{k=1}^{+\infty} (1+k^2)\, |c_k|^2 < +\infty \right\}. \tag{2.14}$$

avec $\|u\|_{\mathrm{H}^1} = \left(\sum_{k=1}^{+\infty} (1+k^2)\, |c_k|^2 \right)^{1/2}$. Soit $\{c_k\}_{k\geq 1}$ une suite de réels tels que $\sum_{k=1}^{+\infty} (1+k^2)\, |c_k|^2 < +\infty$. Pour tout entier N, on considère la fonction $u_N(x) = \sum_{k=1}^{N} c_k\, e_k(x)$. Pour tout $p \geq 1$, on a $u_{N+p}(x) - u_N(x) = \sum_{k=N+1}^{N+p} c_k\, e_k(x)$, donc $\|u_{N+p} - u_N\|^2_{H^1(0,\pi)} = \sum_{k=N+1}^{N+p} (1+k^2)\, |c_k|^2$. Puisque la série $\sum_{k=1}^{+\infty} (1+k^2)\, |c_k|^2$ converge, on voit que u_N est de Cauchy dans $H^1(0,\pi)$, et elle converge donc (au sens de la norme $H^1(0,\pi)$) vers un certain $u \in H^1(0,\pi)$. On a donc

$$\left\{ u(x) = \sum_{k=1}^{+\infty} c_k\, e_k(x), \quad \sum_{k=1}^{+\infty} (1+k^2)\, |c_k|^2 < +\infty \right\} \subset H^1(0,\pi).$$

En notant γ l'application trace, on voit bien sur que $\gamma(u_N) = 0$, puique $e_k(0) = e_k(\pi) = 0$ pour tout $k \geq 1$. Puisque γ est continu sur $H^1(0,\pi)$ et que u_N converge vers u dans $H^1(0,\pi)$, on obtient que $\gamma(u) = 0$, si bien que $u \in H^1_0(0,\pi)$, et donc

$$\left\{ u(x) = \sum_{k=1}^{+\infty} c_k\, e_k(x), \quad \sum_{k=1}^{+\infty} (1+k^2)\, |c_k|^2 < +\infty \right\} \subset H^1_0(0,\pi).$$

Réciproquement, soit $u \in H_0^1(0,\pi) \subset \mathrm{L}^2(0,\pi)$. Grace à (2.13), on peut écrire u sous la forme $u = \sum_{k=1}^{+\infty} c_k\, e_k$, où la série converge dans $\mathrm{L}^2(0,\pi)$, avec $\sum_{k=1}^{+\infty} |c_k|^2 < +\infty$. Puisque $u' \in \mathrm{L}^2(0,\pi)$, les coefficients c_k vérifient, pour tout $k \geq 1$,

$$\begin{aligned} c_k = \int_0^\pi u(x)\, e_k(x)\, dx &= \sqrt{2/\pi} \int_0^\pi u(x)\, \sin(kx)\, dx \\ &= -\frac{\sqrt{2/\pi}}{k}\, [u(x)\, \cos(kx)]_0^\pi + \frac{\sqrt{2/\pi}}{k} \int_0^\pi u'(x)\, \cos(kx)\, dx. \end{aligned}$$

En utilisant maintenant que u est nul au bord, on obtient

$$c_k = \frac{\sqrt{2/\pi}}{k} \int_0^\pi u'(x)\, \cos(kx)\, dx.$$

Posons $z_k(x) = \sqrt{2/\pi}\, \cos(kx)$. Puisque les fonctions $\{z_k,\ k \in \mathbb{N}\}$ forment une base orthonormée de $\mathrm{L}^2(0,\pi)$ (au même titre que les fonctions $\{e_k,\ k \in \mathbb{N}^\star\}$), on a que

$$\|u'\|^2_{\mathrm{L}^2(0,\pi)} = \sum_{k\geq 0} |\langle u', z_k\rangle|^2 = |\langle u', z_0\rangle|^2 + \sum_{k\geq 1} k^2\, |c_k|^2 = \sum_{k\geq 1} k^2\, |c_k|^2,$$

la dernière égalité provenant du fait que $\langle u', z_0\rangle = \sqrt{2/\pi} \displaystyle\int_0^\pi u'(x)\, dx = 0$ puisque u est nul au bord. On vient donc d'obtenir que la suite de réels $\{c_k\}_{k\geq 1}$ satisfait $\sum_{k=1}^{+\infty} (1+k^2)\, |c_k|^2 < +\infty$. La suite des sommes partielles $u_N(x) = \sum_{k=1}^N c_k\, e_k(x)$ est donc de Cauchy dans $H^1(0,\pi)$, elle converge donc dans $H^1(0,\pi)$ vers un certain $\overline{u}$. Puisque la convergence dans $H^1(0,\pi)$ implique la convergence dans $\mathrm{L}^2(0,\pi)$, on a donc que $\overline{u} = u$, si bien qu'on peut effectivement écrire u sous la forme $\sum_{k=1}^{+\infty} c_k\, e_k$, avec convergence de la série dans $H^1(0,\pi)$ et $\sum_{k=1}^{+\infty} (1+k^2)\, |c_k|^2 < +\infty$. Ceci achève la preuve de (2.14).

On vient donc de caractériser les espaces $\mathrm{L}^2(0,\pi)$ et $\mathrm{H}_0^1(0,\pi)$ par (2.13) et (2.14). Soit

$$\begin{array}{rccc} I\ :\ \mathrm{H}_0^1(0,\pi) & \longrightarrow & \mathrm{L}^2(0,\pi) \\ u & \mapsto & u \end{array}$$

l'injection canonique de $\mathrm{H}_0^1(0,\pi)$ dans $\mathrm{L}^2(0,\pi)$. Pour tout $N \in \mathbb{N}^*$, soit I_N l'opérateur linéaire défini par

$$\begin{array}{rccc} I_N\ :\ \mathrm{H}_0^1(0,\pi) & \longrightarrow & \mathrm{L}^2(0,\pi) \\ u = \displaystyle\sum_{k=1}^{+\infty} c_k e_k & \mapsto & I_N(u) = \displaystyle\sum_{k=1}^{N} c_k e_k. \end{array}$$

Montrons que la suite $(I_N)_{N\in\mathbb{N}^*}$ converge vers I dans $\mathcal{L}(\mathrm{H}^1_0, \mathrm{L}^2)$. On calcule

$$\begin{aligned}
\|I - I_N\|^2_{\mathcal{L}(\mathrm{H}^1_0,\mathrm{L}^2)} &= \sup_{u\in\mathrm{H}^1_0(\Omega),\ u\neq 0} \frac{\|(I-I_N)(u)\|^2_{\mathrm{L}^2}}{\|u\|^2_{\mathrm{H}^1_0}} \\
&= \sup_{(c_k)_{k\in\mathbb{N}^*}\neq 0,\ \sum(1+k^2)|c_k|^2<+\infty} \frac{\displaystyle\sum_{k=N+1}^{+\infty} |c_k|^2}{\displaystyle\sum_{k=1}^{+\infty}(1+k^2)|c_k|^2} \\
&\leq \sup_{(c_k)_{k\in\mathbb{N}^*}\neq 0,\ \sum(1+k^2)|c_k|^2<+\infty} \frac{\displaystyle\sum_{k=N+1}^{+\infty} |c_k|^2}{\displaystyle\sum_{k=N+1}^{+\infty}(1+k^2)|c_k|^2} \\
&\leq \frac{1}{1+(N+1)^2} \underset{N\to+\infty}{\longrightarrow} 0.
\end{aligned}$$

Par ailleurs, pour tout $N \in \mathbb{N}^*$, l'opérateur I_N est de rang fini (égal à N). C'est donc un opérateur compact. Il en résulte que I est limite dans $\mathcal{L}(\mathrm{H}^1_0, \mathrm{L}^2)$ d'opérateurs compacts. C'est donc lui-même un opérateur compact d'après le Théorème 2.27.

— Pour $\Omega =]0, \pi[^d$, on montre de la même manière que l'injection canonique de $\mathrm{H}^1_0(\Omega)$ dans $\mathrm{L}^2(\Omega)$ est compacte. Il suffit de développer les fonctions $u \in \mathrm{H}^1_0(\Omega)$ dans la base tensorielle de Fourier :

$$u(x_1, x_2, \cdots, x_d) = \sum_{k_1,k_2,\cdots,k_d=1}^{+\infty} c_{k_1k_2\cdots k_d} \sin(k_1x_1)\, \sin(k_2x_2)\cdots\sin(k_dx_d).$$

— Enfin, si Ω est un ouvert borné quelconque de $\mathbb{R}^d$, on peut se ramener par homothétie et translation au cas où $\Omega \subset \omega =]0, \pi[^d$. Il suffit alors de remarquer que l'injection I_Ω de $\mathrm{H}^1_0(\Omega)$ dans $\mathrm{L}^2(\Omega)$ peut se décomposer en

$$I_\Omega\ :\ \mathrm{H}^1_0(\Omega) \ \xrightarrow{\ p\ }\ \mathrm{H}^1_0(\omega) \ \xrightarrow{\ I_\omega\ }\ \mathrm{L}^2(\omega) \ \xrightarrow{\ r\ }\ \mathrm{L}^2(\Omega)$$

où p désigne l'opérateur linéaire qui transforme une fonction de $\mathrm{H}^1_0(\Omega)$ en une fonction de $\mathrm{H}^1_0(\omega)$ en la prolongeant par 0 dans $\omega \setminus \Omega$, I_ω est l'injection canonique de $\mathrm{H}^1_0(\omega)$ dans $\mathrm{L}^2(\omega)$ et r est l'opérateur de restriction qui à $u \in \mathrm{L}^2(\omega)$ associe la fonction $u|_\Omega$ (qui est dans $\mathrm{L}^2(\Omega)$). Comme p et r sont des opérateurs continus et I_ω est un opérateur compact, il en résulte (cf. la proposition 2.28) que I_Ω est lui-même un opérateur compact.

Ceci conclut la preuve. □

Remarque 2.32 (Injection compacte de $\mathrm{H}^1(\Omega)$ dans $\mathrm{L}^2(\Omega)$)**.** *Une modification de la preuve ci-dessus permet de montrer facilement que l'injection de* $\mathrm{H}^1(\Omega)$ *dans* $\mathrm{L}^2(\Omega)$ *est compacte lorsque le domaine* Ω *est un parallélépipède* $\Omega = \prod_{i=1}^{d}]a_i, b_i[$*. Pour des domaines généraux, la question est plus difficile. Ce qui pose problème dans la preuve ci-dessus, c'est de montrer que l'opérateur d'extension (celui qui à une fonction* $f \in \mathrm{H}^1(\Omega)$ *associe une fonction* $\widetilde{f} \in \mathrm{H}^1(\omega)$ *où* ω *est un cube contenant* Ω *et* $\widetilde{f}\big|_{\Omega} = f$*) est bien défini et est continu. De tels résultats existent pour des domaines bornés* réguliers*, voir par exemple [4, Théorème 7.1.7] et [2, Théorème IX.7] et les résultats ci-dessous.*

On a le résultat suivant :

Théorème 2.33 (de Rellich-Kondrachov)**.** *Soit* Ω *ouvert* régulier *borné de* $\mathbb{R}^d$*. On a les injections compactes :*

- *si* $d > 2$*, alors* $H^1(\Omega) \subset L^q(\Omega)$ *pour tout* $q \in [1, p^*[$*, avec* $1/p^* = 1/2 - 1/d$*.*
- *si* $d = 2$*, alors* $H^1(\Omega) \subset L^q(\Omega)$ *pour tout* $q \in [1, +\infty[$*.*
- *si* $d = 1$*, alors* $H^1(\Omega) \subset C^0(\overline{\Omega})$*.*

On en déduit en particulier le résultat suivant.

Corollaire 2.34. *Soit* Ω *un ouvert régulier borné de* $\mathbb{R}^d$*. Alors l'injection* $H^1(\Omega) \subset L^2(\Omega)$ *est compacte.*

Donc, si Ω est un ouvert régulier borné, alors, de toute suite bornée de $H^1(\Omega)$, on peut extraire une sous-suite convergente dans $L^2(\Omega)$.

Démontration du Corollaire 2.34. Si $d \geq 2$, le résultat découle directement du théorème de Rellich-Kondrachov. Si $d = 1$, on remarque que l'injection $I : H^1(\Omega) \longrightarrow L^2(\Omega)$ est la composition de deux injections

$$I_1 : H^1(\Omega) \longrightarrow C^0(\overline{\Omega}) \quad \text{ et } I_2 : C^0(\overline{\Omega}) \longrightarrow L^2(\Omega).$$

L'injection I_1 est compacte d'après le théorème de Rellich-Kondrachov, et l'injection I_2 est continue. L'injection $I = I_1 \circ I_2$ est donc compacte. □

Le corollaire suivant est alors une conséquence immédiate des Propositions 2.7 et 2.26.

Corollaire 2.35. *Soit* Ω *un ouvert régulier borné de* $\mathbb{R}^d$*. Soit* u_n *une suite bornée de* $H^1(\Omega)$*. On peut extraire de la suite* u_n *une sous-suite qui converge faiblement vers* u *dans* $H^1(\Omega)$ *et qui converge fortement vers* u *dans* $L^2(\Omega)$*.*

Exercice 18. *En utilisant le corollaire ci-dessus, démontrer l'inégalité de Poincaré (1.8) par un raisonnement par l'absurde.*

2.3.3 Théorie spectrale des opérateurs autoadjoints compacts

Les opérateurs autoadjoints compacts ont une structure spectrale très particulière, qui ressemble beaucoup à celle des opérateurs linéaires en dimension finie.

Théorème 2.36 (Diagonalisation des opérateurs auto-adjoints compacts)**.** *Soit H un espace de Hilbert séparable de dimension infinie et $T \in \mathcal{L}(H)$ un opérateur autoadjoint compact. Alors il existe une suite (μ_n) de réels non nuls, finie ou tendant vers 0, et une base hilbertienne $(e_n) \cup (f_n)$ de H, telles que*

1. *$\sigma(T) = (\mu_n) \cup \{0\}$,*
2. *$Te_n = \mu_n e_n$ (et donc $\mu_n \in \sigma_{\mathrm{p}}(T)$),*
3. *(f_n) est une base de $\mathrm{Ker}(T)$.*

En outre, pour tout $\lambda \in \sigma(T) \setminus \{0\}$, l'espace propre $E_\lambda = \mathrm{Ker}(\lambda - T)$ est de dimension finie.

On note qu'on a toujours $0 \in \sigma(T)$. En effet :
- soit T n'est pas injectif, et alors $0 \in \sigma_{\mathrm{p}}(T)$;
- soit T n'est pas surjectif, et alors $0 \in \sigma_{\mathrm{r}}(T) \cup \sigma_{\mathrm{c}}(T)$ (en effet, si T est injectif et surjectif, alors il est bijectif, ce qui n'est pas possible en vertu de l'exercice 16) ; d'après la Proposition 2.22, on a que $\sigma_{\mathrm{r}}(T) = \emptyset$, donc $0 \in \sigma_{\mathrm{c}}(T)$.

Remarque 2.37. *La preuve ci-dessous montre que plusieurs cas (et uniquement ceux-là) peuvent se présenter :*

1. *on peut avoir $\sigma(T) = \sigma_{\mathrm{p}}(T)$, avec les cas suivants :*
   (a) *ou bien $\sigma(T) = \sigma_{\mathrm{p}}(T) = \{0\}$, auquel cas $T = 0$. Dans ce cas, la base (f_n) engendre tout l'espace, et la base (e_n) est vide ;*
   (b) *ou bien $\sigma(T) = \sigma_{\mathrm{p}}(T) = \{\mu_n\}_{n \in \{1,\dots,N\}} \cup \{0\}$, c'est-à-dire que T est de rang fini (et bien sur T n'est pas injectif). Dans ce cas, la base (e_n) est de cardinal fini N, et la base (f_n) est de cardinal infini ;*
   (c) *ou bien $\sigma(T) = \sigma_{\mathrm{p}}(T) = \{\mu_n\}_{n \geq 0} \cup \{0\}$, auquel cas T est non injectif. La base (e_n) est de cardinal infini, alors que la base (f_n) peut être de cardinal fini ou infini en fonction de la dégénerescence de la valeur propre 0 ;*
2. *si $\sigma_{\mathrm{p}}(T) \subsetneq \sigma(T)$, alors $\sigma(T)$ est l'union disjointe de $\sigma_{\mathrm{p}}(T)$ et de $\{0\}$. Dans ce cas, T est injectif (car $0 \notin \sigma_{\mathrm{p}}(T)$) et on a $\sigma_{\mathrm{p}}(T) = \{\mu_n\}_{n \geq 0}$ et $\sigma_{\mathrm{c}}(T) = \{0\}$ (en effet, $\{0\} = \sigma(T) \setminus \sigma_{\mathrm{p}}(T) = \sigma_{\mathrm{c}}(T) \cup \sigma_{\mathrm{r}}(T)$ et $\sigma_{\mathrm{r}}(T) = \emptyset$ d'après la Proposition 2.22).*

Démonstration. Nous décomposons cette (longue) preuve en plusieurs étapes.

1. Montrons pour commencer que

$$\sigma(T) \subset \sigma_{\mathrm{p}}(T) \cup \{0\}\,. \tag{2.15}$$

On rappelle que $\sigma(T) \subset \mathbb{R}$ par la Proposition 2.22. Pour montrer (2.15), considérons $\lambda \in \mathbb{R} \setminus \{0\}$ tel que $\lambda \notin \sigma_{\mathrm{p}}(T)$. Il s'agit de montrer que $\lambda \notin \sigma(T)$. Comme $\lambda \notin \sigma_{\mathrm{p}}(T)$, $(\lambda - T)$ est injectif. Etudions alors la surjectivité en nous intéressant à $V = \mathrm{Ran}(\lambda - T)$, et plus particulièrement, montrons que $V = H$, ce qui donnera le résultat escompté.

(a) On montre que V est fermé.
En effet, soit une suite $(w_n)_{n\in\mathbb{N}}$ d'éléments de V qui converge vers w dans H. Soit $(v_n)_{n\in\mathbb{N}}$ l'unique suite d'éléments de H définie par $w_n = (\lambda - T)v_n$ pour tout $n \in \mathbb{N}$. On a alors

$$v_n = \frac{1}{\lambda}[w_n + Tv_n].$$

Montrons d'abord que la suite (v_n) admet une sous-suite bornée. Par l'absurde, supposons que $\|v_n\| \longrightarrow +\infty$. En utilisant le fait que w_n converge, on aurait dans ce cas

$$\lambda \frac{v_n}{\|v_n\|} - T\frac{v_n}{\|v_n\|} = \frac{w_n}{\|v_n\|} \longrightarrow 0.$$

En utilisant la compacité de l'opérateur T, on extrait de (v_n) une sous-suite (v_{n_k}) telle que

$$T\frac{v_{n_k}}{\|v_{n_k}\|} \longrightarrow u \in H.$$

D'où

$$\frac{v_{n_k}}{\|v_{n_k}\|} \longrightarrow z = \frac{1}{\lambda}u$$

et z vérifie $(\lambda - T)z = 0$. Il en résulte que $z = 0$ puisque $\lambda - T$ est injectif. C'est impossible car z est la limite forte d'une suite de points de la sphère unité de H.
La suite $(v_n)_{n\in\mathbb{N}}$ admet donc une sous-suite bornée. L'opérateur T étant compact, (v_n) admet une sous-suite (v_{n_k}) bornée telle qu'on ait

$$Tv_{n_k} \longrightarrow w' \in H.$$

En utilisant à nouveau que w_n converge, il en résulte que

$$v_{n_k} \longrightarrow v = \frac{1}{\lambda}[w + w'] \in H,$$

ce qui indique que la suite $(v_n)_{n\in\mathbb{N}}$ admet une sous-suite convergente. Comme T est continu, on a finalement

$$w = \lim_{k\to+\infty} w_{n_k} = \lim_{k\to+\infty} (\lambda - T)v_{n_k} = (\lambda - T)v \in V,$$

ce qui montre bien que V est fermé.

(b) On montre que V est dense.
En effet, soit $w \in V^\perp$. Alors $\langle(\lambda - T)v, w\rangle = 0$ pour tout $v \in H$. Comme T est auto-adjoint et λ est réel, on en déduit que $\langle v, (\lambda - T)w\rangle = 0$ pour tout $v \in H$. Ceci implique que $(\lambda - T)w = 0$, et donc $w = 0$ puisque $(\lambda - T)$ est injectif. Donc $V^\perp = \{0\}$, et en utilisant le lemme 1.13, on en déduit que $\overline{V} = (V^\perp)^\perp = H$.

Ceci conclut la preuve de (2.15).

2. Montrons que $\sigma_{\mathrm{p}}(T)$ est ou bien une suite finie, ou bien une suite infinie qui converge vers 0.
Dans le cas contraire, on pourrait extraire de $\sigma_{\mathrm{p}}(T)$ une suite $(\lambda_n)_{n\in\mathbb{N}}$ de réels non nuls *tous distincts* qui converge vers un réel $\mu \neq 0$. Soit $e_n \in \mathrm{Ker}(\lambda_n - T)$ tel que $\|e_n\| = 1$. On a, pour tout $n \in \mathbb{N}$,

$$e_n = \frac{1}{\lambda_n} T e_n.$$

La suite $(e_n)_{n\in\mathbb{N}}$ étant bornée et T étant compact, on peut extraire une sous-suite (Te_{n_k}) qui converge dans H vers un certain u, d'où

$$e_{n_k} \longrightarrow \frac{1}{\mu} u.$$

Or la suite (e_n) est orthonormale par la Proposition 2.22, ce qui montre que la suite (e_{n_k}) n'est pas de Cauchy, donc ne peut pas converger. On a obtenu une contradiction, ce qui donne le résultat annoncé. En particulier, $\sigma_{\mathrm{p}}(T)$ est dénombrable.

3. A tout élément $\lambda_n \in \sigma_{\mathrm{p}}(T)$ tel que $\lambda_n \neq 0$, on associe $E_n = \mathrm{Ker}(\lambda_n - T)$. Montrons que les espaces E_n sont de dimension finie.
Soit en effet $T_n = T|_{E_n}$. Il est clair que $T_n = \lambda_n \,\mathrm{Id}_{E_n}$ (avec $\lambda_n \neq 0$) et que T_n est compact de E_n dans E_n (car c'est la restriction d'un opérateur compact à l'ensemble $E_n = \mathrm{Ker}(\lambda_n - T)$). L'opérateur T_n est donc compact et bijectif de E_n dans E_n. L'exercice 16 indique alors que E_n est de dimension finie.

4. Les espaces E_n sont deux à deux orthogonaux (par la Proposition 2.22) et sont orthogonaux à $F = \mathrm{Ker}(T)$.
Pour le second point, on procède comme dans la preuve de la Proposition 2.22. En effet, soit $\lambda_n \in \sigma_{\mathrm{p}}(T)$ avec $\lambda_n \neq 0$, $u \in E_n$ et $v \in F$. Alors

$$\lambda_n \langle u, v\rangle = \langle Tu, v\rangle = \langle u, Tv\rangle = 0,$$

d'où $\langle u, v\rangle = 0$ puisque $\lambda_n \neq 0$. Donc $F \subset E^\perp$.

5. Soit enfin $E = \bigoplus_n E_n$. Montrons que $H = E \oplus F$, où les sommes directes sont des sommes orthogonales dans les deux cas (selon le point précédent).

(a) Remarquons tout d'abord que E est stable par T. En effet, soit $x \in E$. On peut écrire

$$x = \sum_n x_n, \qquad x_n \in E_n, \qquad \sum_n \|x_n\|^2 < +\infty.$$

Comme par ailleurs (λ_n) est finie ou tend vers 0, la série $\sum_n \lambda_n x_n$ converge dans H. On a donc

$$Tx = \sum_n \lambda_n x_n, \qquad \lambda_n x_n \in E_n, \qquad \sum_n \|\lambda_n x_n\|^2 < +\infty,$$

ce qui montre que $Tx \in E$.

(b) Par ailleurs, $E^\perp$ est aussi stable par T. En effet, si $w \in E^\perp$, alors $\langle Tw, v\rangle = \langle w, Tv\rangle = 0$ pour tout $v \in E$ (on a utilisé que $Tv \in E$). Ceci montre que $Tw \in E^\perp$.

(c) Définissons maintenant $\widetilde{T}$, la restriction de T à l'ensemble fermé $E^\perp$:

$$\begin{array}{rccc} \widetilde{T} \ : & E^\perp & \to & E^\perp \\ & v & \mapsto & Tv. \end{array}$$

L'opérateur $\widetilde{T}$ est auto-adjoint et compact. En vertu de (2.15), on a $\sigma(\widetilde{T}) \subset \sigma_{\rm p}(\widetilde{T}) \cup \{0\}$. Supposons que $\sigma_{\rm p}(\widetilde{T}) \not\subset \{0\}$. Il existe alors $\lambda \in \sigma_{\rm p}(\widetilde{T})$ avec $\lambda \neq 0$, et il existe donc $v \in E^\perp \setminus \{0\}$ tel que

$$\widetilde{T}v = \lambda v,$$

d'où aussi $Tv = \lambda v$. Donc $\lambda \in \sigma_{\rm p}(T)$. Ceci signifie cependant que $\lambda = \lambda_n$ et que $v \in E_n$ pour un certain n. D'où

$$v \in E_n \cap E^\perp = \{0\},$$

ce qui contredit l'hypothèse $v \neq 0$. Donc $\sigma_{\rm p}(\widetilde{T}) \subset \{0\}$.

Il en résulte que $\sigma(\widetilde{T}) \subset \{0\}$, et comme le spectre n'est jamais vide, on obtient

$$\sigma(\widetilde{T}) = \{0\}.$$

D'après la proposition 2.22, la relation ci-dessus implique que $\|\widetilde{T}\| = 0$ et donc que $\widetilde{T} = 0$. Ainsi, $E^\perp \subset \mathrm{Ker}(T) = F$.

(d) On a $H = E \oplus E^\perp$ et on a vu ci-dessus que $F \subset E^\perp$. On vient de montrer que $E^\perp \subset \mathrm{Ker}(T) = F$. Donc $F = E^\perp$, ce qui donne bien que $H = E \oplus F$.

6. La base (e_n) et la suite (μ_n) sont construites de la manière suivante. Notons n_k la dimension de E_k. On prend $\mu_1 = \mu_2 = \cdots = \mu_{n_1} = \lambda_1$ et $(e_1, \cdots, e_{n_1})$ une base orthonormale de E_1. Puis on pose $\mu_{n_1+1} = \cdots = \mu_{n_1+n_2} = \lambda_2$ et $(e_{n_1+1}, \cdots, e_{n_1+n_2})$ une base orthonormale de E_2. On procède de même pour tous les espaces E_n.

Ceci conclut la preuve. □

Remarque 2.38. *Soit H est un espace de Hilbert. La preuve précédente montre qu'on peut écrire tout opérateur autoadjoint de $\mathcal{K}(H)$ (donc compact) comme une limite d'opérateurs de rang fini (voir [2]). En effet, comme $(e_n) \cup (f_n)$ forme une base hilbertienne de H, on peut écrire tout $u \in H$ sous la forme*

$$u = \sum_{n=1}^{+\infty} u_n,$$

et l'application T est diagonale dans cette base :

$$Tu = \sum_{n=1}^{+\infty} \lambda_n u_n, \tag{2.16}$$

avec $\lambda_n \to 0$ lorsque $n \to +\infty$ (éventuellement, il est possible que $\lambda_n = 0$ à partir d'un certain rang). Définissant les opérateurs de rang fini T_N par

$$T_N u = \sum_{n=1}^{N} \lambda_n u_n,$$

on voit facilement que $\|T - T_N\| \leq \sup_{m \geq N} |\lambda_m| \longrightarrow 0$ lorsque $N \to +\infty$.

Remarque 2.39 (Calcul fonctionnel)**.** *Notons également que la décomposition* (2.16) *permet de définir des opérateurs $f(T)$ par la formule*

$$f(T)u = \sum_{n=1}^{+\infty} f(\lambda_n) u_n.$$

Ceci généralise les opérations faites sur les matrices symétriques réelles.

2.3.4 Opérateurs autoadjoints compacts définis positifs

Dans la suite du cours, nous aurons besoin en particulier d'appliquer le théorème de décomposition spectrale à des opérateurs autoajoints compacts *définis positifs*. Donnons-en tout d'abord la définition.

Définition 2.40. *Soit V un espace de Hilbert, et soit A un opérateur linéaire et continu de V dans V. On dit que A est défini positif si*

$$\forall u \in V \setminus \{0\}, \quad \langle Au, u \rangle > 0.$$

Remarque 2.41. *Soit V un espace de Hilbert, et soit A un opérateur linéaire et continu de V dans V. On lui associe la forme bilinéaire a définie par*

$$a(u, w) = \langle Au, w \rangle.$$

En dimension finie, A est défini positif si et seulement si a est coercive. En dimension infinie, ce n'est plus le cas, comme le montre l'exercice 19 ci-dessous.

Exercice 19. *Soit Ω un ouvert borné de $\mathbb{R}^d$. On se place dans l'espace de Hilbert $L^2(\Omega)$. Pour tout $f \in L^2(\Omega)$, le problème*

$$\left\{ \begin{array}{l} \text{Chercher } u \in H_0^1(\Omega) \text{ tel que} \\ -\Delta u = f \quad \text{dans } \mathcal{D}'(\Omega) \end{array} \right. \tag{2.17}$$

admet une unique solution. On considère l'opérateur

$$\begin{array}{rcl} A : L^2(\Omega) & \longrightarrow & L^2(\Omega) \\ f & \longmapsto & u \quad \text{solution du problème (2.17).} \end{array}$$

Montrer que A est un opérateur linéaire et continu et que A est défini positif. Pour montrer que la forme bilinéaire associée a n'est pas coercive, on pourra supposer que Ω est la boule ouverte de centre 0 et de rayon 1, et considérer les fonctions $f_n(x) = n^{d/2}\chi(nx)$, où χ est une fonction fixée de $\mathcal{D}(\Omega)$.

Le théorème ci-dessous est alors un corollaire du Theorème 2.36 (on est dans le dernier cas évoqué dans la Remarque 2.37).

Théorème 2.42. *Soit V un espace de Hilbert de dimension infinie, et A un opérateur linéaire, continu, défini positif, auto-adjoint et compact de V dans V. Alors les valeurs propres de A forment une suite $(\lambda_k)_{k\geq 1}$ de réels strictement positifs qui tend vers 0, et il existe une base hilbertienne $(u_k)_{k\geq 1}$ de V formée de vecteurs propres de A, avec*

$$\forall k \geq 1, \quad Au_k = \lambda_k u_k.$$

De plus, le sous-espace propre associé à chaque valeur propre est de dimension finie.

On remarque que le théorème ci-dessus ne caractérise que le spectre ponctuel de l'opérateur, alors que le Theorème 2.36 caractérise tout le spectre.

Remarque 2.43. *Comme $(u_k)_{k\geq 1}$ forme une base hilbertienne de V, on peut appliquer la proposition 1.10 et on a donc les relations suivantes pour tout $w \in V$:*

$$w = \sum_{k\geq 1} \langle w, u_k\rangle u_k \quad \text{et} \quad \|w\|^2 = \sum_{k\geq 1} |\langle w, u_k\rangle|^2.$$

Exercice 20. *On reprend les notations et hypothèses du théorème 2.42. Montrer que, pour $w \in V$, l'équation $Au = w$ admet une unique solution $u \in V$ si et seulement si w vérifie*

$$\sum_{k\geq 1} \frac{|\langle w, u_k\rangle|^2}{\lambda_k^2} < +\infty.$$

Exercice 21. *Soit $V = L^2(0,1)$ et A l'application linéaire de V dans V définie par $(Af)(x) = (x^2+1)f(x)$. Vérifier que A est continue, définie positive, auto-adjointe, mais pas compacte. Montrer que A n'a pas de valeurs propres. Montrer que $A - \lambda\mathrm{Id}$ est inversible si et seulement si $\lambda \notin [1,2]$.*

Nous présentons ici une démonstration directe du théorème 2.42. Dans ce but, nous aurons besoin des deux lemmes suivants.

Lemme 2.44. *Soit V un espace de Hilbert (non réduit au seul vecteur nul) et A une application linéaire continue auto-adjointe compacte de V dans V. On définit*

$$m = \inf_{u \in V \setminus \{0\}} \frac{\langle Au, u\rangle}{\langle u, u\rangle} \quad et \quad M = \sup_{u \in V \setminus \{0\}} \frac{\langle Au, u\rangle}{\langle u, u\rangle}.$$

Alors $\|A\|_{\mathcal{L}(V)} = \max(|m|, |M|)$ et soit m, soit M, est valeur propre de A.

Lemme 2.45. *Soit V un espace de Hilbert et A une application linéaire continue compacte de V dans V. Pour tout réel $\delta > 0$, il n'existe au plus qu'un nombre fini de valeurs propres de A en dehors de l'intervalle $]-\delta, \delta[$.*

Démontration du lemme 2.44. On voit que $|\langle Au, u\rangle| \leq \|A\|_{\mathcal{L}(V)}\|u\|^2$, par conséquent $\max(|m|, |M|) \leq \|A\|_{\mathcal{L}(V)}$. Comme A est auto-adjoint, on a, pour tout u et w dans V, que

$$\begin{aligned} 4\langle Au, w\rangle &= \langle A(u+w), u+w\rangle - \langle A(u-w), u-w\rangle \\ &\leq M\|u+w\|^2 - m\|u-w\|^2 \\ &\leq \max(|m|, |M|)\left(\|u+w\|^2 + \|u-w\|^2\right) \\ &\leq 2\max(|m|, |M|)\left(\|u\|^2 + \|w\|^2\right). \end{aligned}$$

Si $Au \neq 0$, on peut choisir $w = Au/\|Au\|$ dans l'inégalité précédente, et on obtient

$$2\|Au\| \leq \max(|m|, |M|)\left(\|u\|^2 + 1\right).$$

Cette dernière inégalité reste vraie si $Au = 0$. On prend maintenant le supremum sur les $u \in V$, $\|u\| = 1$, ce qui donne $2\|A\|_{\mathcal{L}(V)} \leq 2\max(|m|, |M|)$. En combinant cette inégalité avec l'inégalité inverse obtenue ci-dessus, on obtient que $\max(|m|, |M|) = \|A\|_{\mathcal{L}(V)}$.

On montre maintenant la deuxième partie du lemme. Si $m = M = 0$, alors, pour tout $u \in V$, on a $\langle Au, u\rangle = 0$. En utilisant l'exercice 12, on obtient que $A = 0$, ce qui termine la preuve du lemme. On suppose maintenant que soit m, soit M, est non nul, et donc $\max(|m|, |M|) > 0$. Par définition, on a $M \geq m$. Si $M \leq |m|$, alors on est dans un des deux cas suivants :

- soit $0 \geq M \geq m$: on change alors A en $-A$ ce qui permet de revenir au cas $M \geq m > 0$.
- soit $M \geq 0 \geq m$ et $M \leq |m|$: on change alors A en $-A$ ce qui permet de revenir au cas $M \geq |m| \geq 0$.

Sans perte de généralité, on peut donc supposer que $M \geq |m|$ et $M > 0$. Montrons que M est valeur propre de A. En utilisant la première partie du lemme et la définition de M, on a

$$\|A\|_{\mathcal{L}(V)} = M = \sup_{u \in V \setminus \{0\}} \frac{\langle Au, u\rangle}{\langle u, u\rangle}.$$

Soit $u_n \in V$ une suite maximisante, avec $\|u_n\| = 1$. On a donc $\lim_{n\to+\infty}\langle Au_n, u_n\rangle = M$. Comme u_n est bornée et A est compacte, on peut extraire de Au_n une sous-suite convergente : $\lim_{n\to+\infty} Au_{\varphi(n)} = v$. On a aussi

$$\langle Au_n, u_n\rangle \leq \|Au_n\|\ \|u_n\| \leq \|A\|_{\mathcal{L}(V)}\ \|u_n\|^2 = \|A\|_{\mathcal{L}(V)} = M.$$

Or $\lim_{n\to+\infty}\langle Au_n, u_n\rangle = M$, ce qui donne que $\lim_{n\to+\infty}\|Au_n\|\ \|u_n\| = M$. Comme $\|u_n\| = 1$, on en déduit que $\lim_{n\to+\infty}\|Au_n\| = M$. Sachant que Au_n converge à extraction près vers v, on obtient que $\|v\| = M$.

On voit aussi que

$$\|Au_n - Mu_n\|^2 = \|Au_n\|^2 + M^2 - 2M\langle Au_n, u_n\rangle \to_{n\to+\infty} 0,$$

ce qui implique $\lim_{n\to\infty} Au_n - Mu_n = 0$. Or $\lim\limits_{n\to+\infty} Au_{\varphi(n)} = v$, donc $\lim\limits_{n\to+\infty} Mu_{\varphi(n)} = v$. Comme A est continue, on a $\lim\limits_{n\to+\infty} MAu_{\varphi(n)} = Av$, et par unicité de la limite, on déduit que $Mv = Av$, avec $v \neq 0$. Donc M est bien valeur propre de A. □

Démontration du lemme 2.45. On procède par contradiction, et on suppose donc qu'il existe une suite infinie de valeurs propres $(\lambda_k)_{k\geq 1}$ distinctes telles que $|\lambda_k| \geq \delta$. Soient $(u_k)_{k\geq 1}$ les vecteurs propres associés, et E_k le sous-espace vectoriel engendré par $u_1, \ldots, u_k$.

Grâce au lemme 2.20, les vecteurs propres $(u_k)_{k\geq 1}$ sont linéairement indépendants, et donc E_{k-1} est strictement inclus dans E_k. Donc il existe w_k de norme 1, avec $w_k \in E_k$ et w_k orthogonal à E_{k-1}. Comme λ_k est isolé de 0, on voit que la suite de vecteurs w_k/λ_k est bornée. L'application A étant compacte, on en déduit que, à extraction près, la suite Aw_k/λ_k converge. Par ailleurs, pour $j < k$, on voit que

$$\begin{aligned}\frac{1}{\lambda_k}Aw_k - \frac{1}{\lambda_j}Aw_j &= \frac{1}{\lambda_k}\left(Aw_k - \lambda_k w_k\right) + w_k - \frac{1}{\lambda_j}Aw_j \\ &= \left(A - \lambda_k \mathrm{Id}\right)\frac{w_k}{\lambda_k} + w_k - \frac{1}{\lambda_j}Aw_j.\end{aligned}$$

Or, pour tout $w \in E_k$, on a $(A - \lambda_k\mathrm{Id})\, w \in E_{k-1}$. Par conséquent, les vecteurs $(A - \lambda_k\mathrm{Id})\,\dfrac{w_k}{\lambda_k}$ et $\dfrac{1}{\lambda_j}Aw_j$ sont dans E_{k-1}, tandis que w_k est orthogonal à E_{k-1}. Donc

$$\begin{aligned}\left\|\frac{1}{\lambda_k}Aw_k - \frac{1}{\lambda_j}Aw_j\right\|^2 &= \left\|\left(A - \lambda_k\mathrm{Id}\right)\frac{w_k}{\lambda_k} - \frac{1}{\lambda_j}Aw_j\right\|^2 + \|w_k\|^2 \\ &\geq \|w_k\|^2 = 1.\end{aligned}$$

Ceci est contradictoire avec le fait que la suite Aw_k/λ_k converge à extraction près. □

Démontration du théorème 2.42. Le lemme 2.44 montre que l'ensemble des valeurs propres n'est pas vide, tandis que le lemme 2.45 montre que cet ensemble est soit fini,

soit infini dénombrable avec 0 comme seul point d'accumulation. On note $(\lambda_k)_{k\geq 1}$ les valeurs propres de A et $V_k = \mathrm{Ker}(A - \lambda_k \mathrm{Id})$ les sous-espaces vectoriels propres associés. Comme A est défini positif, on voit que les valeurs propres sont toutes strictement positives.

Comme $\lambda_k \neq 0$, l'application $\dfrac{1}{\lambda_k}A$ est compacte, et la proposition 2.29 montre que $V_k = \mathrm{Ker}(\dfrac{1}{\lambda_k}A - \mathrm{Id})$ est de dimension finie.

Les sous-espaces propres sont orthogonaux deux à deux. En effet, si $v_k \in V_k$ et $v_j \in V_j$ avec $k \neq j$, alors, comme A est auto-adjoint,

$$\langle Av_j, v_k\rangle = \lambda_j\langle v_j, v_k\rangle = \langle v_j, Av_k\rangle = \lambda_k\langle v_j, v_k\rangle.$$

On déduit de $\lambda_k \neq \lambda_j$ que $\langle v_j, v_k\rangle = 0$.

Soit

$$W = \left\{ v \in V;\ \exists K \geq 1 \text{ tel que } v = \sum_{k=1}^{K} v_k,\ v_k \in V_k \right\}$$

l'espace vectoriel engendré par les $(v_k)_{k\geq 1}$. Montrons que W est dense dans V. Il est clair que W est stable par A, c'est-à-dire $A(W) \subset W$. L'application A étant auto-adjointe, ceci implique que $W^\perp$ est lui-aussi stable par A. On considère alors la restriction A_0 de A à $W^\perp$, qui est encore une application linéaire continue auto-adjointe compacte. Si $W^\perp \neq \{0\}$, on peut appliquer le lemme 2.44, et donc A_0 a une valeur propre λ. Soit u le vecteur propre associé : $u \in W^\perp$ et $Au = \lambda u$. Donc λ est une valeur propre de A, et par conséquent $u \in W$. Donc $u \in W \cap W^\perp$, ce qui est contradictoire avec le fait que $u \neq 0$. Donc $W^\perp = \{0\}$. Par conséquent, $V = \{0\}^\perp = (W^\perp)^\perp = \overline{W}$ (on a utilisé le lemme 1.13 pour obtenir la dernière égalité), ce qui montre que W est dense dans V.

On construit maintenant une base hilbertienne de V. Pour cela, on considère dans chacun des V_k (qui sont de dimension finie) une base orthonormée. Les réunions de ces bases forme une base hilbertienne de V, car les V_k sont orthogonaux deux à deux et W est dense dans V.

Comme V est de dimension infinie et que les V_k sont de dimension finie, on obtient aussi que A possède un nombre infini dénombrable de valeurs propres. □

Chapitre 3

Equations aux dérivées partielles et problèmes aux valeurs propres

3.1 Motivation

Ce chapitre est une introduction à l'étude mathématique et numérique des phénomènes vibratoires. Ces phénomènes ont une grande importance pour de nombreuses sciences de l'ingénieur : génie civil, acoustique (des instruments de musique mais aussi des véhicules), détection de fissure dans des matériaux (par contrôle non destructif), ...

D'un point de vue mathématique, il s'agit d'étudier les valeurs propres et vecteurs propres d'équations aux dérivées partielles. Illustrons notre propos sur un exemple concret. On considère une membrane élastique homogène et isotrope, dont le bord est maintenu fixe, initialement au repos, et on cherche à étudier sa réponse à une excitation dépendant du temps.

Lorsqu'on néglige les forces de gravitation devant les forces de tension superficielle, et qu'on se place dans le cadre de l'élasticité linéaire, le système vérifié par le déplacement vertical $u(t,x)$ d'un point de la membrane situé au repos à la position $x \in \Omega$ s'écrit :

$$\begin{cases} \dfrac{1}{c^2}\dfrac{\partial^2 u}{\partial t^2}(t,x) - \Delta u(t,x) = f(t,x) & \text{dans } \mathbb{R}^{+*} \times \Omega, \\ u(t,x) = 0 & \text{sur } \mathbb{R}^{+*} \times \partial\Omega, \\ u(0,x) = 0 & \text{sur } \Omega, \\ \dfrac{\partial u}{\partial t}(0,x) = 0 & \text{sur } \Omega, \end{cases} \tag{3.1}$$

où $c = \sqrt{S/\rho}$, S désignant la tension superficielle et ρ la masse surfacique de la membrane. On reconnaît dans l'EDP du système (3.1) une équation d'onde de célérité c comportant un terme source f.

L'analogue discret (en espace) de ce problème est le système dynamique d'incon-

nue $U(t) \in \mathbb{R}^N$ suivant :

$$\begin{cases} M \dfrac{d^2U}{dt^2}(t) + AU(t) = B(t), \\ U(0) = \dfrac{dU}{dt}(0) = 0, \end{cases} \tag{3.2}$$

où M et A sont deux matrices de taille $N \times N$ et $B(t)$ est un vecteur de $\mathbb{R}^N$ dépendant du temps.

Nous verrons plus loin dans le cours (cf. la deuxième partie du polycopié) qu'on peut effectivement passer du système (3.1) au système (3.2) par une formulation variationnelle de (3.1), qui est ensuite approximée par une méthode de Galerkin (par exemple une méthode d'éléments finis).

Supposons ici pour simplifier que M est la matrice identité, et que A est une matrice symétrique. Une méthode classique pour résoudre (3.2) est de diagonaliser la matrice A, ce qui consiste à chercher les couples $(\lambda_k, U_k)_{1 \leq k \leq N}$ de valeurs propres et de vecteurs propres de A, qui vérifient donc

$$\forall k, \quad AU_k = \lambda_k U_k. \tag{3.3}$$

Puisque A est symétrique, ses vecteurs propres forment une base orthonormée de $\mathbb{R}^N$. On cherche alors une solution de (3.2) comme une combinaison linéaire sur ces vecteurs propres :

$$U(t) = \sum_{k=1}^{N} \alpha_k(t) U_k \quad \text{avec} \quad \alpha_k(t) \in \mathbb{R}.$$

En insérant cette décomposition dans (3.2), on trouve que les α_k vérifient

$$\frac{d^2\alpha_k}{dt^2} + \lambda_k \alpha_k(t) = b_k(t) \tag{3.4}$$

avec $b_k(t) = \langle B(t), U_k \rangle$. On est donc ramené à la résolution d'une équation différentielle ordinaire scalaire.

L'argument clé qui a permis de ramener le système (3.2), posé en dimension N éventuellement grande, à la résolution des N équations scalaires *indépendantes* (3.4), est la diagonalisation de la matrice A et la recherche d'une solution comme combinaison linéaire de vecteurs propres. Essayons maintenant d'utiliser la même stratégie pour résoudre le problème (3.1). L'analogue de la matrice A, qui associe au vecteur U le vecteur $A\,U$, est l'opérateur $-\Delta$, qui à la distribution u associe la distribution $-\Delta u$. Il est donc naturel d'essayer de chercher des fonctions u_k, définies sur Ω, et des réels λ_k, tels que

$$-\Delta u_k = \lambda_k u_k \quad \text{dans} \quad \Omega. \tag{3.5}$$

Ce problème aux valeurs propres est l'équivalent en dimension infinie du problème (3.3). En fait, cette équation aux valeurs propres apparaît aussi naturellement si on

s'intéresse à l'équation sans second membre associée à (3.1), et qu'on en cherche une solution sous la forme $u(t,x) = \varphi(t)v(x)$. Oublions les conditions initiales : les fonctions φ et v doivent alors vérifier

$$\begin{cases} \dfrac{1}{c^2}\varphi''(t)v(x) - \varphi(t)\Delta v(x) = 0 & \text{pour tout } t > 0,\ x \in \Omega, \\ v(x) = 0 & \text{sur } \partial\Omega. \end{cases} \tag{3.6}$$

Formellement, on a donc

$$\forall t > 0,\ \forall x \in \Omega, \quad \frac{\varphi''(t)}{\varphi(t)} = \frac{\Delta v}{v} = -\lambda,$$

où $\lambda \in \mathbb{R}$ est une constante, et donc la fonction $v(x)$ est un vecteur propre du laplacien avec conditions de Dirichlet nulles au bord (on retrouve la relation (3.5)), tandis que φ suit l'équation suivante, similaire à (3.4) :

$$\varphi''(t) + \lambda\varphi(t) = 0.$$

Supposons $\lambda > 0$ (nous montrerons au théorème 3.3 ci-dessous que c'est effectivement le cas). Alors $\varphi(t) = a\cos(\sqrt{\lambda}t) + b\sin(\sqrt{\lambda}t)$, et la fonction

$$u(t,x) = av(x)\cos(\sqrt{\lambda}t) + bv(x)\sin(\sqrt{\lambda}t) \tag{3.7}$$

est solution de l'EDP apparaissant dans (3.1) avec $f = 0$. La fonction u s'interprète comme un mode propre de vibration de la membrane. La signification mécanique de λ se comprend sur la relation (3.7) : il s'agit du carré des pulsations propres de vibration.

La discussion ci-dessus permet donc de comprendre l'importance des valeurs propres et des vecteurs propres du laplacien, et de la signification du point de vue vibratoire de ces quantités.

La suite de ce chapitre est organisée ainsi. Les théorèmes abstraits qui ont été présentés au Chapitre 2 sont utilisés dans la section 3.2 pour étudier les modes propres du laplacien et de l'élasticité linéarisée. En pratique, on ne peut calculer qu'une approximation numérique des valeurs et vecteurs propres, et l'analyse d'erreur est discutée dans la section 3.3. Enfin, la mise en oeuvre numérique d'une méthode de discrétisation aboutit au bout du compte à un problème d'algèbre linéaire, qui consiste à diagonaliser une matrice. Quelques algorithmes pour la résolution d'un tel problème seront discutés dans la section 3.4.

3.2 Valeurs propres d'un problème elliptique

Pour commencer cette section, on se place dans un cadre assez général, qu'on pourra ensuite appliquer à différents modèles. Nous suivons en fait la même démarche que dans le cours d'Analyse de première année [9], dans lequel on a tout d'abord démontré, dans un cadre assez général, le théorème de Lax-Milgram, qu'on a ensuite appliqué à différentes équations. Nous appliquerons le résultat abstrait démontré à la section 3.2.1 dans la section 3.2.2, pour l'étude des valeurs propres du laplacien.

3.2.1 Problème variationnel abstrait

On se donne un espace de Hilbert V et une forme bilinéaire $a(\cdot,\cdot)$ sur V, qui est symétrique, continue et coercive. On se donne aussi un autre espace de Hilbert H, tel que

$$\begin{cases} V \subset H \text{ avec injection compacte au sens de la définition 2.30,} \\ V \text{ dense dans } H. \end{cases}$$

Pour ne pas confondre les produits scalaires sur H et sur V, nous les noterons respectivement $\langle\cdot,\cdot\rangle_H$ et $\langle\cdot,\cdot\rangle_V$. Les normes associées sont notées $\|\cdot\|_H$ et $\|\cdot\|_V$. Les hypothèses sur la forme a donnent donc l'existence de $M > 0$ et $\alpha > 0$ tels que

$$\begin{gathered} \forall u \in V,\ \forall w \in V, \quad |a(u,w)| \leq M\|u\|_V\|w\|_V, \\ \forall u \in V, \quad a(u,u) \geq \alpha\|u\|_V^2. \end{gathered}$$

Le problème qui nous intéresse ici est : trouver $\lambda \in \mathbb{R}$ et $u \in V \setminus \{0\}$ tels que

$$\forall w \in V, \quad a(u,w) = \lambda\langle u, w\rangle_H. \tag{3.8}$$

On dira alors que λ est valeur propre de la forme bilinéaire a (ou du probème variationnel (3.8)), et que u est le vecteur propre associé.

On donne dès à présent un cas typique d'application du cadre abstrait développé ici. Soit Ω un ouvert borné de $\mathbb{R}^d$. On pose $V = H_0^1(\Omega)$, $H = L^2(\Omega)$, et

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v.$$

Nous montrerons à la section 3.2.2 que les hypothèses faites ci-dessus sont vérifiées, et que résoudre (3.8) est alors équivalent à chercher $\lambda \in \mathbb{R}$ et $u \in H_0^1(\Omega)$, $u \neq 0$, tels que

$$-\Delta u = \lambda u \text{ dans } \Omega.$$

Ainsi, λ et u seront valeur propre et vecteur propre du laplacien dans Ω avec conditions aux limites de Dirichlet.

Théorème 3.1. *Soient V et H deux espaces de Hilbert de dimension infinie. On suppose $V \subset H$ avec injection compacte et V dense dans H. Soit $a(\cdot,\cdot)$ une forme bilinéaire symétrique, continue et coercive sur V. Alors les valeurs propres de (3.8) forment une suite croissante $(\lambda_k)_{k\geq 1}$ de réels strictement positifs qui tend vers l'infini, et il existe une base hilbertienne de H de vecteurs propres associés, c'est-à-dire :*

$$u_k \in V \quad et \quad \forall w \in V,\ a(u_k,w) = \lambda_k\langle u_k, w\rangle_H. \tag{3.9}$$

De plus, $u_k/\sqrt{\lambda_k}$ est une base hilbertienne de V pour le produit scalaire $a(\cdot,\cdot)$.

Démonstration. L'injection $V \subset H$ étant continue, on sait qu'il existe $C > 0$ tel que

$$\forall w \in V, \quad \|w\|_H \leq C\|w\|_V. \tag{3.10}$$

Pour $f \in H$, on considère le problème variationnel

$$\begin{cases} \text{Chercher } u \in V \text{ tel que} \\ \forall w \in V, \quad a(u,w) = \langle f, w\rangle_H. \end{cases} \tag{3.11}$$

Grâce au théorème de Lax-Milgram, ce problème admet une unique solution $u \in V$. On définit les applications linéaires

$$\begin{array}{rcl} \mathcal{A} : H & \longrightarrow & V \\ f & \longmapsto & u \text{ unique solution de (3.11),} \end{array}$$

et

$$\begin{array}{rcl} A : H & \longrightarrow & H \\ f & \longmapsto & \mathcal{A}f. \end{array}$$

Comme a est coercive sur V, on a, pour u solution de (3.11),

$$\alpha\|u\|_V^2 \leq a(u,u) = \langle f, u\rangle_H \leq \|f\|_H \ \|u\|_H.$$

En utilisant (3.10), on obtient

$$\|\mathcal{A}f\|_V = \|u\|_V \leq \frac{C}{\alpha}\|f\|_H.$$

Donc $\mathcal{A}$ est linéaire continue de H dans V. En utilisant à nouveau (3.10), on obtient que A est linéaire continue de H dans H.

Montrons que A est définie positive, auto-adjointe et compacte sur H.

Comme A est la composition de $\mathcal{A} \in \mathcal{L}(H,V)$ et de l'injection de V dans H, qui est compacte, on a que A est compacte. Soient maintenant f et g dans H. On a

$$\langle f, Ag\rangle_H = \langle f, \mathcal{A}g\rangle_H = a(\mathcal{A}f, \mathcal{A}g) = a(\mathcal{A}g, \mathcal{A}f) = \langle g, \mathcal{A}f\rangle_H = \langle g, Af\rangle_H,$$

et donc A est auto-adjointe sur H. On montre enfin que A est définie positive sur H. En prenant $g = f$ dans l'égalité précédente, on voit que, pour tout $f \in H$,

$$\langle f, Af\rangle_H = a(\mathcal{A}f, \mathcal{A}f) \geq \alpha\|\mathcal{A}f\|_V^2 \geq 0.$$

Supposons que $\langle f, Af\rangle_H = 0$. Alors l'inégalité ci-dessus donne que $\mathcal{A}f = 0$. Par définition, on a

$$\forall w \in V, \quad a(\mathcal{A}f, w) = \langle f, w\rangle_H.$$

On déduit de $\mathcal{A}f = 0$ que $\langle f, w\rangle_H = 0$ pour tout $w \in V$. Or V est dense dans H, donc ceci implique que $\langle f, w\rangle_H = 0$ pour tout $w \in H$, et par conséquent $f = 0$.

Finalement, pour tout $f \in H$, $f \neq 0$, on a $\langle f, Af\rangle_H > 0$ et donc A est définie positive sur H.

On peut donc appliquer le théorème 2.42. Il existe donc une base hilbertienne de H formée des vecteurs propres u_k de A, associés aux valeurs propres $(\mu_k)_{k\geq 1}$, qui forme une suite décroissante vers 0 :

$$\forall k \geq 1, \quad Au_k = \mu_k u_k.$$

Comme $\mu_k > 0$ et $Au_k \in V$, on voit que $u_k \in V$. On montre maintenant que les u_k sont vecteurs propres de la forme bilinéaire a. Par définition de A, on a

$$\forall w \in V,\ a(Au_k, w) = \langle u_k, w\rangle_H = \mu_k a(u_k, w),$$

et donc, en posant

$$\lambda_k = \frac{1}{\mu_k},$$

on obtient (3.9). Montrons que les v_k définis par

$$v_k = \frac{u_k}{\sqrt{\lambda_k}}$$

forment une base hilbertienne de V pour le produit scalaire $a(\cdot,\cdot)$. On a $v_k \in V$ et l'espace vectoriel engendré par les v_k est dense dans H, donc dense dans V. Enfin, les vecteurs v_k sont orthogonaux deux à deux, car

$$\begin{aligned} a(v_k, v_p) &= a\left(\frac{u_k}{\sqrt{\lambda_k}}, \frac{u_p}{\sqrt{\lambda_p}}\right) \\ &= \frac{1}{\sqrt{\lambda_k\lambda_p}} a(u_k, u_p) \\ &= \frac{\sqrt{\lambda_k}}{\sqrt{\lambda_p}} \langle u_k, u_p\rangle_H = \delta_{kp}. \end{aligned}$$

Ceci conclut la preuve du théorème. □

On donne maintenant une caractérisation très utile des valeurs propres du problème (3.8), appelé principe du min-max ou de Courant-Fisher. Nous introduisons le quotient de Rayleigh défini, pour chaque $v \in V \setminus \{0\}$, par

$$R(v) = \frac{a(v,v)}{\|v\|_H^2}. \tag{3.12}$$

Proposition 3.2. *Soient V et H deux espaces de Hilbert de dimension infinie. On suppose $V \subset H$ avec injection compacte et V dense dans H. Soit $a(\cdot,\cdot)$ une forme bilinéaire symétrique, continue et coercive sur V. Pour $k \geq 0$, on note E_k l'ensemble des sous-espaces vectoriels de dimension k de V. On note $(\lambda_k)_{k\geq 1}$ la suite croissante*

des valeurs propres du problème variationnel (3.8). Alors, pour tout $k \geq 1$, la k-ième valeur propre est donnée par

$$\lambda_k = \min_{W \in E_k} \left(\max_{v \in W \setminus \{0\}} R(v) \right) = \max_{W \in E_{k-1}} \left(\min_{v \in W^\perp \setminus \{0\}} R(v) \right). \tag{3.13}$$

En particulier, la première valeur propre vérifie

$$\lambda_1 = \min_{v \in V \setminus \{0\}} R(v), \tag{3.14}$$

et tout point de minimum dans (3.14) est un vecteur propre associé à λ_1.

Démonstration. Soit u_k une base hilbertienne de H formée des vecteurs propres de (3.8). On commence par caractériser H et V. On a

$$H = \left\{ v = \sum_{k \geq 1} \alpha_k u_k \text{ tel que } \sum_{k \geq 1} \alpha_k^2 < +\infty \right\}.$$

En effet, soit $v \in H$: comme u_k est une base hilbertienne de H, en utilisant la proposition 1.10, on a bien $v = \sum_{k \geq 1} \alpha_k u_k$ avec $\alpha_k = \langle v, u_k \rangle_H$. La série $\sum_{k \geq 1} \alpha_k^2$ est bien convergente car égale à $\|v\|_H^2$. Réciproquement, soit une suite α_k telle que $\sum_{k \geq 1} \alpha_k^2 < +\infty$. La suite $\sum_{k=1}^K \alpha_k u_k$ est bien dans H, et elle est de Cauchy, donc elle converge vers un élément de H.

On montre maintenant que

$$V = \left\{ v = \sum_{k \geq 1} \alpha_k u_k \text{ tel que } \sum_{k \geq 1} \lambda_k \alpha_k^2 < +\infty \right\}.$$

Soit $v \in V$: les $v_k = u_k / \sqrt{\lambda_k}$ forment une base hilbertienne de V pour $a(\cdot, \cdot)$, donc on peut décomposer v suivant ces v_k selon

$$v = \sum_{k \geq 1} \alpha_k v_k \text{ avec } a(v, v) = \sum_{k \geq 1} \alpha_k^2.$$

Posant $\beta_k = \alpha_k / \sqrt{\lambda_k}$, on obtient $v = \sum_{k \geq 1} \beta_k u_k$ avec $a(v, v) = \sum_{k \geq 1} \lambda_k \beta_k^2 < +\infty$. Réciproquement, supposons $v = \sum_{k \geq 1} \alpha_k u_k$ avec $\sum_{k \geq 1} \lambda_k \alpha_k^2 < +\infty$. Alors la suite $\sum_{k=1}^K \alpha_k u_k$ est une suite d'éléments de V qui est de Cauchy pour la norme induite par $a(\cdot, \cdot)$. Donc cette suite converge vers un élément de V.

Soit maintenant $v \in V \setminus \{0\}$. Alors on écrit $v = \sum_{k \geq 1} \alpha_k u_k$ et le quotient de Rayleigh s'écrit

$$R(v) = \frac{\sum_{k \geq 1} \lambda_k \alpha_k^2}{\sum_{k \geq 1} \alpha_k^2}.$$

L'égalité (3.14) est donc claire. Soit u un point de minimum : $R(u) = \lambda_1$. Soit $v \in V$ quelconque. La fonction $f(t) = R(u+tv)$ est minimale en $t = 0$, donc $f'(0) = 0$. Or

$$f'(0) = 2\frac{a(u,v)\|u\|_H^2 - \langle u, v\rangle_H a(u,u)}{\|u\|_H^4}.$$

Comme $f'(0) = 0$ et $a(u,u) = \lambda_1\|u\|_H^2$, on obtient $a(u,v) = \lambda_1\langle u, v\rangle_H$ pour tout $v \in V$, et donc u est vecteur propre associé à la valeur propre λ_1.

On démontre maintenant (3.13). Soit W_k l'espace vectoriel engendré par $(u_1, \ldots, u_k)$, qui est de dimension k. Soit $v \in W_k$: on a $R(v) = \dfrac{\sum_{j=1}^k \lambda_j \alpha_j^2}{\sum_{j=1}^k \alpha_j^2}$ donc

$$\lambda_k = \max_{v\in W_k, v\neq 0} R(v) \geq \min_{W\in E_k} \left(\max_{v\in W\setminus\{0\}} R(v) \right). \tag{3.15}$$

De même, pour $v \in W_{k-1}^\perp$, on a $R(v) = \dfrac{\sum_{j\geq k} \lambda_j \alpha_j^2}{\sum_{j\geq k} \alpha_j^2}$ et donc

$$\lambda_k = \min_{v\in W_{k-1}^\perp, v\neq 0} R(v) \leq \max_{W\in E_{k-1}} \left(\min_{v\in W^\perp\setminus\{0\}} R(v) \right).$$

Soit maintenant W un sous-espace vectoriel de V de dimension k. On a $V = W_{k-1} \oplus W_{k-1}^\perp$, donc $W = (W\cap W_{k-1})\oplus(W\cap W_{k-1}^\perp)$. Si $W\cap W_{k-1}^\perp = \{0\}$, alors $W = W\cap W_{k-1}$, ce qui n'est pas possible car W est de dimension k et $W \cap W_{k-1}$ est de dimension inférieure ou égale à $k-1$. Donc $(W \cap W_{k-1}^\perp) \setminus \{0\} \neq \emptyset$. On a

$$\begin{aligned}
\max_{v\in W\setminus\{0\}} R(v) &\geq \max_{v\in (W\cap W_{k-1}^\perp)\setminus\{0\}} R(v) \\
&\geq \min_{v\in (W\cap W_{k-1}^\perp)\setminus\{0\}} R(v) \\
&\geq \min_{v\in W_{k-1}^\perp\setminus\{0\}} R(v) = \lambda_k.
\end{aligned}$$

Par conséquent,

$$\min_{W\in E_k} \left(\max_{v\in W\setminus\{0\}} R(v) \right) \geq \lambda_k.$$

En rassemblant cette inégalité avec (3.15), on obtient la première égalité de (3.13). La seconde égalité de (3.13) s'obtient de manière analogue, en considérant $W \in E_{k-1}$ et en s'appuyant sur le fait que $W^\perp \cap W_k$ n'est pas réduit à $\{0\}$. □

3.2.2 Application : valeurs propres du laplacien

Dans cette section, nous mettons en oeuvre le théorème 3.1, démontré dans un cadre abstrait, pour étudier les valeurs propres du laplacien.

Théorème 3.3. *Soit Ω un ouvert borné régulier de classe C^1 de $\mathbb{R}^d$. Il existe une suite croissante $(\lambda_k)_{k\geq 1}$ de réels strictement positifs qui tend vers l'infini, et il existe une base hilbertienne de $L^2(\Omega)$, notée $(u_k)_{k\geq 1}$, telle que chaque u_k appartient à $H_0^1(\Omega)$ et vérifie*

$$\left\{ \begin{array}{rcll} -\Delta u_k & = & \lambda_k u_k & \text{dans } \mathcal{D}'(\Omega), \\ u_k & = & 0 & \text{sur } \partial\Omega. \end{array} \right. \tag{3.16}$$

Les $(\lambda_k)_{k\geq 1}$ et les $(u_k)_{k\geq 1}$ sont appelés les valeurs propres et vecteurs propres du laplacien avec conditions aux limites de Dirichlet sur l'ouvert Ω.

Démonstration. On va appliquer le théorème 3.1, avec les choix $V = H_0^1(\Omega)$ (muni du produit scalaire $(\cdot,\cdot)_{H^1}$), $H = L^2(\Omega)$ (muni du produit scalaire $(\cdot,\cdot)_{L^2}$), et

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v.$$

Comme $C_0^\infty(\Omega)$ est dense dans $L^2(\Omega)$ et inclus dans $H_0^1(\Omega)$, on a bien que V est dense dans H. Comme Ω est borné, on peut appliquer le théorème de Rellich 2.33, et l'injection $V \subset H$ est bien compacte. La forme a est bien bilinéaire, symétrique, continue et coercive sur V (ce dernier point résulte directement de l'inégalité de Poincaré (1.8)). Par conséquent, il existe une suite croissante $(\lambda_k)_{k\geq 1}$ de réels positifs et une base hilbertienne $(u_k)_{k\geq 1}$ de $L^2(\Omega)$ tels que $u_k \in H_0^1(\Omega)$ et

$$\forall v \in H_0^1(\Omega),\ \int_\Omega \nabla u_k \cdot \nabla v = \lambda_k \int_\Omega u_k v.$$

On obtient alors (3.16) par une simple intégration par partie. □

Remarque 3.4. *Supposons que Ω soit de classe C^∞. Alors les u_k solutions de (3.16) sont bien plus réguliers que $H_0^1(\Omega)$. On voit en effet que $-\Delta u_k = \lambda_k u_k$ avec $\lambda_k u_k$ de régularité H^1. Donc $\Delta u_k \in H^1(\Omega)$. Comme Ω est très régulier, ceci impose que $u_k \in H^3(\Omega)$, et donc $\Delta u_k \in H^3(\Omega)$, ce qui donne $u_k \in H^5(\Omega)$, ... On obtient finalement que $u_k \in C^\infty(\Omega)$.*

Remarque 3.5. *L'hypothèse que Ω est borné est fondamentale. Sans cette hypothèse, le théorème de Rellich est faux, et le théorème 3.3 est lui aussi faux.*

Exercice 22. *On se place en dimension 1 et on considère $\Omega =]0,1[$. Calculer explicitement toutes les valeurs propres et les fonctions propres du laplacien avec conditions aux limites de Dirichlet (3.16). En déduire que la série $\sum_{k\geq 1} a_k \sin(k\pi x)$ converge dans $L^2(0,1)$ si et seulement si $\sum_{k\geq 1} a_k^2 < +\infty$, et que la même série converge dans $H^1(0,1)$ si et seulement si $\sum_{k\geq 1} k^2 a_k^2 < +\infty$.*

En utilisant le principe de Courant-Fisher, on pourra résoudre l'exercice suivant.

Exercice 23. *On reprend les notations et hypothèses du théorème 3.3. Trouver une relation entre la plus petite constante C_Ω possible dans l'inégalité de Poincaré (1.8) et la première valeur propre λ_1 de (3.16).*

On donne enfin un résultat qualitatif très important à propos de la première valeur propre.

Théorème 3.6 (de Krein-Rutman)**.** *On reprend les notations et hypothèses du théorème 3.3. On suppose que l'ouvert Ω est connexe. Alors la première valeur propre λ_1 est simple (le sous-espace vectoriel associé est de dimension 1), et le premier vecteur propre peut être choisi positif presque partout dans Ω.*

Remarque 3.7. *Ce théorème est spécifique aux équations scalaires, c'est-à-dire pour lesquelles l'inconnue u est à valeurs dans $\mathbb{R}$. Dans le cas vectoriel (comme par exemple dans le cas de l'élasticité linéaire), le résultat est faux.*

3.3 Méthodes numériques

Dans la section 3.2.1, nous nous sommes intéressés à la résolution du problème aux valeurs propres (3.8). Nous expliquons maintenant comment discrétiser ce problème pour aboutir à une méthode numérique permettant de calculer une approximation des valeurs propres (et éventuellement des vecteurs propres) de (3.8).

3.3.1 Discrétisation du problème

On réalise une approximation interne du problème (3.8). Soit donc $V_h \subset V$ un sous-espace de dimension finie de V. Typiquement, V_h est un espace d'éléments finis, tandis que H est l'espace $L^2(\Omega)$. Le problème discrétisé est : trouver $\lambda_h \in \mathbb{R}$ et $u_h \in V_h \setminus \{0\}$ tels que

$$\forall w_h \in V_h, \quad a(u_h, w_h) = \lambda_h \langle u_h, w_h \rangle_H. \tag{3.17}$$

Théorème 3.8. *On reprend les hypothèses du théorème 3.1 : soient V et H deux espaces de Hilbert de dimension infinie. On suppose $V \subset H$ avec injection compacte et V dense dans H. Soit $a(\cdot,\cdot)$ une forme bilinéaire symétrique, continue et coercive sur V, et soit $V_h \subset V$ un sous-espace de dimension finie J.*

Alors les valeurs propres de (3.17) forment une suite croissante finie

$$0 < \lambda_{1,h} \leq \ldots \leq \lambda_{J,h},$$

et il existe une base de V_h, orthonormale dans H, de vecteurs propres associés, c'est-à-dire : pour tout m, $1 \leq m \leq J$,

$$u_{m,h} \in V_h \quad et \quad \forall w_h \in V_h,\ a(u_{m,h}, w_h) = \lambda_{m,h} \langle u_{m,h}, w_h \rangle_H. \tag{3.18}$$

Pour démontrer ce théorème, nous aurons besoin du résultat d'algèbre linéaire suivant :

Proposition 3.9 (Factorisation de Cholesky)**.** *Soit A une matrice réelle symétrique définie positive. Il existe une unique matrice réelle B, triangulaire inférieure, telle que tous ses éléments diagonaux soient positifs, et qui vérifie*

$$A = BB^t.$$

Démonstration. Plutôt que de démontrer ce théorème en suivant le schéma de preuve du théorème 3.1, on suit ici une preuve plus algébrique. Soit $(\varphi_j)_{1\leq j\leq J}$ une base de V_h (ce sont par exemple les fonctions de base d'une méthode d'éléments finis). On cherche u_h solution de (3.17) sous la forme

$$u_h(x) = \sum_{j=1}^{J} U_j \varphi_j(x).$$

On introduit les matrices de masse $\mathcal{M}_h$ et de rigidité $\mathcal{K}_h$ définies par, pour tout i et j, $1 \leq i, j \leq J$,

$$(\mathcal{M}_h)_{ij} = \langle \varphi_i, \varphi_j \rangle_H, \quad (\mathcal{K}_h)_{ij} = a(\varphi_i, \varphi_j).$$

Alors le problème (3.17) se récrit : trouver $\lambda_h \in \mathbb{R}$ et $U \in \mathbb{R}^J$, $U \neq 0$, tels que

$$\mathcal{K}_h U = \lambda_h \mathcal{M}_h U. \tag{3.19}$$

La terminologie matrice de masse et de rigidité est liée à la mécanique des solides. La matrice de rigidité $\mathcal{K}_h$ est la même que celle apparaissant dans la résolution par approximation interne du problème variationnel $a(u, w) = \langle f, w \rangle_H$. Les matrices $\mathcal{M}_h$ et $\mathcal{K}_h$ sont symétriques définies positives.

Pour résoudre le problème (3.19), on commence par calculer la factorisation de Cholesky de $\mathcal{M}_h$, c'est-à-dire calculer la matrice Q_h telle que $\mathcal{M}_h = Q_h Q_h^t$.

Une fois ceci fait, le problème (3.19) revient au problème classique

$$\tilde{\mathcal{K}}_h \tilde{U} = \lambda_h \tilde{U}, \tag{3.20}$$

avec $\tilde{U} = Q_h^t U$ et $\tilde{\mathcal{K}}_h = Q_h^{-1} \mathcal{K}_h (Q_h^t)^{-1}$. On note que la matrice $\tilde{\mathcal{K}}_h$ est symétrique et positive. Si ξ est tel que $\xi^t \tilde{\mathcal{K}}_h \xi = 0$, alors, puisque $\mathcal{K}_h$ est symétrique définie positive, on a $(Q_h^t)^{-1}\xi = 0$, donc $\xi = 0$. La matrice $\tilde{\mathcal{K}}_h$ est donc symétrique définie positive.

Pour le problème (3.20), on dispose d'algorithmes de calculs de valeurs propres et de vecteurs propres, dont certains seront décrits à la section 3.4.

On note $(\lambda_m, \tilde{U}_m)$ les éléments propres de $\tilde{\mathcal{K}}_h$: $\tilde{\mathcal{K}}_h \tilde{U}_m = \lambda_m \tilde{U}_m$. On définit $U_m = (Q_h^t)^{-1} \tilde{U}_m$ et on a donc $\mathcal{K}_h U_m = \lambda_m \mathcal{M}_h U_m$.

Soit U_m et U_n associés à des valeurs propres distinctes : $\lambda_m \neq \lambda_n$. Alors, en utilisant la symétrie de $\mathcal{K}_h$, on a

$$\lambda_m U_n^t \mathcal{M}_h U_m = U_n^t \mathcal{K}_h U_m = (U_n^t \mathcal{K}_h U_m)^t = U_m^t \mathcal{K}_h U_n = \lambda_n U_m^t \mathcal{M}_h U_n.$$

Puisque $\lambda_m \neq \lambda_n$, ceci implique que $U_m^t \mathcal{M}_h U_n = 0$. Les vecteurs propres solution de (3.19) sont donc orthogonaux pour $\mathcal{M}_h$ (et donc pour $\mathcal{K}_h$). □

Pour éviter d'avoir à calculer la factorisation de Cholesky de $\mathcal{M}_h$, on peut utiliser une formule de quadrature pour évaluer $\langle \varphi_i, \varphi_j \rangle_H$ qui rende la matrice de masse diagonale. Un tel procédé est appelé condensation de masse (ou mass lumping) et est souvent utilisé en pratique, par exemple dans l'esprit de l'exercice suivant.

Exercice 24. *On suppose que Ω est un ouvert borné de $\mathbb{R}^d$,*

$$V = H_0^1(\Omega), \quad H = L^2(\Omega), \quad a(u,v) = \int_\Omega \nabla u \cdot \nabla v.$$

On étudie donc $-\Delta u = \lambda u$ dans $H_0^1(\Omega)$. On suppose qu'on utilise une méthode d'éléments finis P_1 sur un maillage formé de triangles (en 2D) ou de tetraèdres (en 3D) de sommets $(a_i)_{1\le i\le d+1}$. On utilise la formule de quadrature

$$\int_K \psi(x)dx \approx \frac{\mathit{Volume(K)}}{d+1} \sum_{i=1}^{d+1} \psi(a_i), \tag{3.21}$$

où K est un triangle (ou un tétraèdre) du maillage. Ceci revient donc à choisir pour noeud d'intégration les sommets de K, qu'on affecte tous du même poids.

Vérifier que la formule de quadrature (3.21) conduit effectivement à une matrice de masse $\mathcal{M}_h$ diagonale.

3.3.2 Convergence et estimation d'erreur

Nous estimons ici la différence entre les valeurs propres du problème continu (3.8) et les valeurs propres du problème (3.18) (identique à (3.17)), qui est son approximation discrète. Cette estimation est fondée sur la caractérisation suivante des valeurs propres $(\lambda_{m,h})_{1\le m\le J}$ du problème discrétisé (3.18), analogue en dimension finie du principe de Courant-Fisher (cf. la proposition 3.2) :

$$\lambda_{m,h} = \min_{W\in E_{m,h}} \left(\max_{v\in W\setminus\{0\}} R(v) \right), \tag{3.22}$$

où $E_{m,h}$ est l'ensemble des sous-espaces vectoriels de dimension m de V_h, et $R(v)$ est le quotient de Rayleigh défini par (cf. (3.12))

$$R(v) = \frac{a(v,v)}{\|v\|_H^2}.$$

La comparaison de (3.13) et de (3.22) donne déjà que, pour $1 \le m \le J$,

$$\lambda_m \le \lambda_{m,h}.$$

Pour obtenir une majoration de $\lambda_{m,h}$, on introduit l'opérateur de projection $\Pi_h \in \mathcal{L}(V, V_h)$ défini, pour tout $u \in V$, par

$$\forall w_h \in V_h, \quad a(\Pi_h u, w_h) = a(u, w_h). \tag{3.23}$$

Soient $(u_m)_{m\ge1}$ les vecteurs propres de (3.8), et soit W_m le sous-espace vectoriel de V engendré par $(u_1, \dots, u_m)$, qui est de dimension m.

Lemme 3.10. *Pour tout $1 \leq m \leq J$, on pose*

$$\sigma_{m,h} = \inf_{v \in W_m, \|v\|_H = 1} \|\Pi_h v\|_H.$$

Si $\sigma_{m,h} > 0$, on a

$$\lambda_{m,h} \leq \frac{\lambda_m}{\sigma_{m,h}^2}.$$

Démonstration. On utilise le principe de Courant-Fisher (caractérisation (3.22)) avec le choix $W_{m,h} = \text{Vect } \{\Pi_h u_1, \ldots, \Pi_h u_m\}$. On a bien $W_{m,h} \subset V_h$ et $\dim W_{m,h} \leq m$. Montrons que $W_{m,h}$ est de dimension m. Si ce n'est pas le cas, alors il existe $(\alpha_i)_{1 \leq i \leq m}$ non tous nuls tels que

$$0 = \sum_{i=1}^{m} \alpha_i \Pi_h u_i = \Pi_h \left(\sum_{i=1}^{m} \alpha_i u_i \right),$$

ce qui contredit l'hypothèse $\sigma_{m,h} > 0$. Donc $\dim W_{m,h} = m$ et (3.22) implique que

$$\lambda_{m,h} \leq \max_{v \in W_{m,h} \setminus \{0\}} R(v) = \max_{v \in W_m, \|v\|_H = 1} \frac{a(\Pi_h v, \Pi_h v)}{\|\Pi_h v\|_H^2}.$$

Pour tout $v \in V$, on a

$$a(v,v) = a\left(\Pi_h v, \Pi_h v\right) + a\left(v - \Pi_h v, v - \Pi_h v\right) + 2a\left(v - \Pi_h v, \Pi_h v\right).$$

Par définition de $\Pi_h v$, le dernier terme est nul. Par coercivité de a, le second terme est positif. Donc $a(v,v) \geq a\left(\Pi_h v, \Pi_h v\right)$ et donc

$$\lambda_{m,h} \leq \max_{v \in W_m, \|v\|_H = 1} \frac{a(v,v)}{\|\Pi_h v\|_H^2}.$$

Pour $v \in W_m$ tel que $\|v\|_H = 1$, on a $v = \sum_{i=1}^m \alpha_i u_i$ avec $\sum_{i=1}^m \alpha_i^2 = 1$, donc $a(v,v) \leq \lambda_m$, d'où

$$\lambda_{m,h} \leq \lambda_m \max_{v \in W_m, \|v\|_H = 1} \frac{1}{\|\Pi_h v\|_H^2} = \frac{\lambda_m}{\sigma_{m,h}^2}.$$

Ceci conclut la preuve. □

On a donc l'estimation

$$\lambda_m \leq \lambda_{m,h} \leq \lambda_m / (\sigma_{m,h}^2). \tag{3.24}$$

On voit donc que la différence entre $\lambda_{m,h}$ et λ_m est reliée aux propriétés d'approximation de V par V_h. Plus V_h est "proche" de V, plus on s'attend à ce que la solution $\Pi_h u \in V_h$ du problème (3.23) soit proche de u, donc en particulier que $\|\Pi_h u\|_H$ soit

proche de $\|u\|_H$. Ceci implique alors que $\sigma_{m,h}$ est proche de 1 (puisqu'on minimise $\|\Pi_h v\|_H$ sur des vecteurs v de norme 1). On remarque donc que, pour aller plus loin dans l'estimation de $\lambda_{m,h}$, il n'est plus nécessaire de faire appel à la spécificité du problème (c'est un problème aux valeurs propres). Disposer de propriétés d'approximation de V par V_h suffit.

Mentionnons enfin que ces propriétés d'approximation sont souvent reliées à l'existence d'une application r_h de V dans V_h telle que, pour tout $v \in V$, on a $\lim_{h\to 0} \|v - r_h(v)\|_V = 0$. Dans le cas d'une approximation par éléments finis P_1, l'application r_h est par exemple l'interpolation de v sur les noeuds du maillage.

Précisons tout ceci dans un cas particulier. On revient à la définition (3.23) de l'opérateur Π_h. En utilisant le fait que la forme bilinéaire a est coercive et continue, on a, pour tout $u \in V$,

$$\begin{aligned} \alpha\|u - \Pi_h u\|_V^2 &\leq a(u - \Pi_h u, u - \Pi_h u) \\ &\leq a(u - \Pi_h u, u - \Pi_h u + w_h) \\ &\leq M\|u - \Pi_h u\|_V \, \|u - \Pi_h u + w_h\|_V \end{aligned}$$

pour tout $w_h \in V_h$. Donc

$$\|u - \Pi_h u\|_V \leq \frac{M}{\alpha} \inf_{w_h \in V_h} \|u - w_h\|_V. \tag{3.25}$$

On suppose maintenant que $V = H_0^1(\Omega)$ pour un ouvert Ω borné de $\mathbb{R}^n$, et que V_h est le sous-espace de V correspondant à la méthode des éléments finis P_k, avec $k + 1 > n/2$. On considère alors l'interpolée $r_h v$ d'une fonction v. C'est un résultat classique [1] que cette application r_h est bien définie sur $H^{k+1}(\Omega)$ et qu'il existe une constante C vérifiant

$$\forall v \in H^{k+1}(\Omega), \quad \|v - r_h v\|_{H^1(\Omega)} \leq C h^k \|v\|_{H^{k+1}(\Omega)}. \tag{3.26}$$

Supposons maintenant que W_m, l'espace vectoriel engendré par les m premiers vecteurs propres de la forme bilinéaire a, soit inclus dans $H^{k+1}(\Omega)$. Alors, il existe C_m tel que, pour tout $v \in W_m$ de norme 1, on a

$$\|v - r_h v\|_{H^1(\Omega)} \leq C_m h^k. \tag{3.27}$$

Détaillons ceci. On peut toujours supposer que les m premiers vecteurs propres de a, notés u_j, $1 \leq j \leq m$, sont orthogonaux deux à deux pour le produit scalaire de H^1, et sont de norme 1 : $\|u_j\|_{H^1} = 1$. On a supposé que $W_m \subset H^{k+1}(\Omega)$, donc $u_j \in H^{k+1}(\Omega)$ vérifie la majoration (3.26). En posant $\overline{C}_m = C \sup_{1\leq j\leq m} \|u_j\|_{H^{k+1}(\Omega)}$, on a donc

$$\forall j, \ 1 \leq j \leq m, \quad \|u_j - r_h u_j\|_{H^1(\Omega)} \leq \overline{C}_m h^k. \tag{3.28}$$

Soit maintenant $v \in W_m$, avec $\|v\|_{H^1} = 1$. On décompose v sur la base des u_j :

$$v = \sum_{j=1}^{m} \alpha_j u_j \quad \text{avec} \quad \|v\|_{H^1}^2 = \sum_j \alpha_j^2 = 1.$$

La dernière relation implique que $|\alpha_j| \leq 1$ pour tout j. On calcule maintenant

$$\|v - r_h v\|_{H^1(\Omega)} = \left\| \sum_j \alpha_j (u_j - r_h u_j) \right\|_{H^1(\Omega)} \leq \sum_j |\alpha_j| \, \|u_j - r_h u_j\|_{H^1(\Omega)} .$$

En utilisant $|\alpha_j| \leq 1$ et la majoration (3.28), on arrive à

$$\|v - r_h v\|_{H^1(\Omega)} \leq \sum_{j=1}^{m} \overline{C}_m h^k = C_m h^k,$$

ce qui est exactement (3.27).

En rassemblant (3.25) et (3.27), on a donc, pour tout $v \in W_m$ de norme 1, que

$$\|v - \Pi_h v\|_{H^1(\Omega)} \leq \frac{M}{\alpha} C_m h^k,$$

soit $\|\Pi_h v\|_{H^1(\Omega)} \geq 1 - \tilde{C}_m h^k$. Ceci implique $\sigma_{m,h} \geq 1 - \tilde{C}_m h^k$. L'estimation (3.24) donne donc, pour une constante C_m, l'encadrement $\lambda_m \leq \lambda_{m,h} \leq \lambda_m(1 + C_m h^k)$, soit

$$0 \leq \lambda_{m,h} - \lambda_m \leq C_m h^k. \tag{3.29}$$

Nous finissons cette section en énonçant un résultat précis de convergence pour les valeurs propres et les vecteurs propres du laplacien, définis par (3.16), approximés par une méthode d'éléments finis triangulaires P_k. Un tel résultat se généralise à d'autres problèmes et d'autres types d'éléments finis.

Théorème 3.11. *Soit Ω un ouvert borné et régulier de $\mathbb{R}^d$. Soit $(\mathcal{T}_h)_{h\geq 0}$ une suite de maillages triangulaires réguliers de Ω. Soit V_{0h} le sous-espace de $H^1_0(\Omega)$ défini par la méthode des éléments finis P_k, de dimension J.*

Soient $(\lambda_m, u_m)_{m\geq 1}$ les valeurs propres et vecteurs propres du problème (3.16), et soit $(\lambda_{m,h})_{1\leq m\leq J}$ les valeurs propres de l'approximation variationnelle (3.17) correspondante sur l'espace de dimension finie V_{0h}. Pour tout $m \geq 1$ fixé, on a

$$\lim_{h\to 0} |\lambda_m - \lambda_{m,h}| = 0.$$

Il existe une famille de vecteurs propres $(u_{m,h})_{1\leq m\leq J}$ de (3.17) dans V_{0h} telle que, si λ_m est valeur propre simple, alors

$$\lim_{h\to 0} \|u_m - u_{m,h}\|_{H^1(\Omega)} = 0.$$

Si le sous-espace engendré par $(u_1, \ldots, u_m)$ est inclus dans $H^{k+1}(\Omega)$ avec $k+1 > d/2$, alors il existe C_m indépendant de h tel que

$$|\lambda_m - \lambda_{m,h}| \leq C_m h^{2k}. \tag{3.30}$$

Si λ_m est valeur propre simple, alors

$$\|u_m - u_{m,h}\|_{H^1(\Omega)} \leq C_m h^k. \tag{3.31}$$

Il est important à ce stade de faire plusieurs remarques :

— la constante C_m dans (3.30) et (3.31) tend vers $+\infty$ lorsque m tend vers $+\infty$. Donc, à h fixé, les plus grandes valeurs propres discrètes (par exemple, $\lambda_{J,h}$) ne sont pas nécessairement une bonne approximation des valeurs propres exactes. Pour avoir une bonne approximation de λ_J, il peut donc être nécessaire de travailler avec un espace d'approximation V_{0h} de dimension bien plus grande que J.

— la convergence des vecteurs propres ne peut s'obtenir que si la valeur propre est simple. Si λ_m est multiple, alors il se peut que la suite $u_{m,h}$ ne converge pas, mais admette plusieurs points d'accumulation, qui sont des combinaisons linéaires de vecteurs propres associés à λ_m.

— l'ordre de convergence des valeurs propres est le double de celui pour les vecteurs propres [1]. On retrouvera ce phénomène (lié au caractère auto-adjoint de l'opérateur) dans les algorithmes de calcul des valeurs propres et vecteurs propres d'une matrice (cf. par exemple la proposition 3.13).

3.4 Algorithmes pour le calcul de valeurs et de vecteurs propres

Les valeurs propres d'une matrice sont les racines de son polynôme caractéristique $P(\lambda) = \det(A - \lambda \mathrm{Id})$. Cependant, il n'existe pas de méthodes directes (c'est-à-dire qui donnent le résultat en un nombre fini d'opérations) pour calculer les racines d'un polynôme quelconque, dès que son ordre est supérieur ou égal à 5. De plus, tout polynôme est le polynôme caractéristique d'une matrice, donc le calcul des valeurs propres d'une matrice est un problème aussi difficile que celui du calcul des racines d'un polynôme quelconque.

Calculer les valeurs propres d'une matrice est en fait un problème beaucoup plus difficile que la résolution d'un système linéaire. Il n'existe que des méthodes itératives. Nous nous concentrons dans cette section sur le cas des matrices réelles symétriques, pour lesquelles le problème est plus simple.

Nous mentionnons ici trois méthodes typiques pour une matrice symétrique :

— la méthode de la puissance, analysée dans la section 3.4.1. C'est la méthode la plus simple, mais elle ne permet (au mieux) que de calculer les valeurs propres de plus grande et de plus petite valeur absolue.

— la méthode de Given-Householder, qui permet de calculer une ou plusieurs valeurs propres de rang quelconque sans avoir à calculer toutes les valeurs propres. Cette méthode est en fait la concaténation de deux algorithmes, l'algorithme de Householder qui permet de transformer une matrice symétrique

1. On voit aussi que l'estimation (3.29) sur les valeurs propres n'est pas optimale, si la forme bilinéaire a correspond au laplacien.

en une matrice tridiagonale de mêmes valeurs propres, et l'algorithme de Givens qui permet le calcul des valeurs propres d'une matrice tridiagonale. Nous n'en dirons pas plus et renvoyons à la bibliographie pour plus de détails.

— la méthode de Lanczos, analysée dans la section 3.4.2. Comme l'algorithme de gradient conjugué, cette méthode fait appel aux espaces de Krylov. Nous en décrirons ci-dessous l'esprit. Cette méthode est à la base de nombreux développements récents qui conduisent aux méthodes les plus efficaces pour de grandes matrices creuses.

3.4.1 Méthode de la puissance

Il s'agit de la méthode la plus simple pour calculer la valeur propre de plus grande (ou de plus petite) valeur absolue. Une limitation de la méthode est que cette valeur propre doit être simple.

Algorithme 3.12 (Méthode de la puissance)**.** *Soit A une matrice symétrique réelle d'ordre n, et ε une précision souhaitée.*

1. *Initialisation : soit $x_0 \in \mathbb{R}^n$ avec $\|x_0\| = 1$.*
2. *Itération : pour $k \geq 1$,*
   (a) *on calcule $y_k = Ax_{k-1}$.*
   (b) *on pose $x_k = y_k/\|y_k\|$.*
   (c) *test de convergence : si $\|x_k - x_{k-1}\| \leq \varepsilon$, on s'arrête.*

La proposition suivante indique sous quelles conditions et à quelle vitesse cet algorithme converge.

Proposition 3.13. *On suppose que A est une matrice réelle symétrique de taille n, de valeurs propres $(\lambda_1, \ldots, \lambda_n)$ rangées par ordre de valeur absolue croissante, et que λ_n est positive et simple : $|\lambda_1| \leq \ldots \leq |\lambda_{n-1}| < \lambda_n$. Soit $(e_1, \ldots, e_n)$ une base de vecteurs propres orthonormés. On suppose que x_0 n'est pas orthogonal à e_n. Alors la méthode de la puissance converge, au sens où*

$$\lim_{k\to+\infty} \|y_k\| = \lambda_n, \quad \lim_{k\to+\infty} x_k = x_\infty \text{ avec } x_\infty = \pm e_n.$$

La convergence est géométrique, avec une vitesse proportionnelle à $|\lambda_{n-1}|/|\lambda_n|$:

$$|\|y_k\| - \lambda_n| \leq C \left|\frac{\lambda_{n-1}}{\lambda_n}\right|^{2k}, \quad \|x_k - x_\infty\| \leq C \left|\frac{\lambda_{n-1}}{\lambda_n}\right|^{k}.$$

Remarque 3.14. *Comme on l'a remarqué dans le théorème 3.11, la convergence de la valeur propre se fait à un ordre deux fois plus grand que la convergence du vecteur propre.*

Démonstration. On décompose le vecteur initial sur les vecteurs propres de A : $x_0 = \sum_{i=1}^n \beta_i e_i$, avec $\beta_n \neq 0$ par hypothèse. Le vecteur x_k est proportionnel à $A^k x_0 = \sum_{i=1}^n \beta_i \lambda_i^k e_i$ et de norme 1, donc

$$x_k = \frac{\beta_n e_n + \sum_{i=1}^{n-1} \beta_i (\lambda_i/\lambda_n)^k e_i}{\left(\beta_n^2 + \sum_{i=1}^{n-1} \beta_i^2 (\lambda_i/\lambda_n)^{2k}\right)^{1/2}}. \tag{3.32}$$

Comme $|\lambda_i| < \lambda_n$, on voit que x_k converge vers $x_\infty = \text{signe}\,(\beta_n) e_n$. On déduit de (3.32) que

$$y_{k+1} = \frac{\beta_n \lambda_n e_n + \sum_{i=1}^{n-1} \beta_i (\lambda_i/\lambda_n)^k \lambda_i e_i}{\left(\beta_n^2 + \sum_{i=1}^{n-1} \beta_i^2 (\lambda_i/\lambda_n)^{2k}\right)^{1/2}},$$

ce qui donne la convergence de $\|y_{k+1}\|$ vers λ_n au rythme $|\lambda_{n-1}/\lambda_n|^{2k}$. □

On est souvent intéressé par le calcul des valeurs propres petites. L'algorithme suivant, très inspiré de la méthode de la puissance, permet de calculer la valeur propre de valeur absolue la plus petite.

Algorithme 3.15 (Méthode de la puissance inverse)**.** *Soit A une matrice symétrique réelle inversible d'ordre n, et ε une précision souhaitée.*

1. *Initialisation : soit $x_0 \in \mathbb{R}^n$ avec $\|x_0\| = 1$.*
2. *Itération : pour $k \geq 1$,*
   (a) *résoudre $Ay_k = x_{k-1}$.*
   (b) *on pose $x_k = y_k/\|y_k\|$.*
   (c) *test de convergence : si $\|x_k - x_{k-1}\| \leq \varepsilon$, on s'arrête.*

La proposition suivante indique sous quelles conditions et à quelle vitesse cet algorithme converge.

Proposition 3.16. *On suppose que A est une matrice réelle symétrique inversible de taille n, de valeurs propres $(\lambda_1, \ldots, \lambda_n)$ rangées par ordre de valeur absolue croissante, et que λ_1 est positive et simple : $0 < \lambda_1 < |\lambda_2| \leq \ldots \leq |\lambda_n|$. Soit $(e_1, \ldots, e_n)$ une base de vecteurs propres orthonormés. On suppose que x_0 n'est pas orthogonal à e_1. Alors la méthode de la puissance inverse converge, au sens où*

$$\lim_{k\to+\infty} \frac{1}{\|y_k\|} = \lambda_1, \quad \lim_{k\to+\infty} x_k = x_\infty \text{ avec } x_\infty = \pm e_1.$$

La convergence est géométrique, avec une vitesse proportionnelle à $|\lambda_1|/|\lambda_2|$:

$$\left|\|y_k\|^{-1} - \lambda_1\right| \leq C\left|\frac{\lambda_1}{\lambda_2}\right|^{2k}, \quad \|x_k - x_\infty\| \leq C\left|\frac{\lambda_1}{\lambda_2}\right|^{k}.$$

Démonstration. La preuve de cette proposition est similaire à celle de la proposition 3.13. □

3.4.2 Méthode de Lanczos

Cette méthode utilise la notion d'espace de Krylov, qui apparait aussi dans l'algorithme de gradient conjugué, et qu'on rappelle ci-dessous. Comme nous l'avons précisé ci-dessus, cette méthode (et ses généralisations) est très efficace pour les matrices de grande taille. On donne ici l'esprit de la méthode plutôt qu'une description précise d'une implémentation numérique efficace.

Dans toute la suite, A est une matrice symétrique réelle d'ordre n, $r_0 \neq 0$ est un vecteur de $\mathbb{R}^n$ donné, et K_k est l'espace de Krylov associé :

Théorème-Définition 3.17. *Soit $r_0 \neq 0$ un vecteur de $\mathbb{R}^n$ donné. Pour tout $k \geq 1$, l'espace de Krylov K_k associé est*

$$K_k = \mathit{Vect}\ \left\{r_0, Ar_0, \ldots, A^k r_0\right\}.$$

Il existe un entier $k_0 \leq n-1$, appelé dimension critique de Krylov, tel que :
- *si $k \leq k_0$, alors la famille $(r_0, \ldots, A^k r_0)$ est libre et $\dim K_k = k+1$;*
- *si $k > k_0$, alors $K_k = K_{k_0}$.*

L'algorithme de Lanczos consiste à construire une suite de vecteurs v_j par la formule de récurrence

$$\forall j \geq 2, \quad \hat{v}_j = Av_{j-1} - \langle Av_{j-1}, v_{j-1}\rangle v_{j-1} - \|\hat{v}_{j-1}\| v_{j-2} \ \text{ et } \ v_j = \frac{\hat{v}_j}{\|\hat{v}_j\|}, \qquad (3.33)$$

avec les initialisations $v_0 = 0$ et $v_1 = r_0/\|r_0\|$. On montrera ci-dessous que, tant que $j \leq k_0 + 1$, on a $\hat{v}_j \neq 0$ et donc v_j est bien défini, tandis que $\hat{v}_{k_0+2} = 0$. La relation entre les v_j et les espaces de Krylov sera explicitée dans le lemme ci-dessous.

Pour tout entier $k \leq k_0 + 1$, on définit la matrice V_k de taille $n \times k$ dont les colonnes sont les vecteurs $v_1, \ldots, v_k$, ainsi que la matrice symétrique tridiagonale de taille $k \times k$ définie par

$$(T_k)_{i,i} = \langle Av_i, v_i\rangle, \quad (T_k)_{i,i+1} = (T_k)_{i+1,i} = \|\hat{v}_{i+1}\|, \quad (T_k)_{i,j} = 0 \text{ sinon.}$$

Lemme 3.18. *Pour tout $j \leq k_0 + 1$, on a $\hat{v}_j \neq 0$ et donc v_j est bien défini, tandis que $\hat{v}_{k_0+2} = 0$.*

Pour $1 \leq k \leq 1 + k_0$, la famille $(v_1, \ldots, v_k)$ coïncide avec la base orthonormée de l'espace de Krylov K_{k-1} construite par le procédé de Gram-Schmidt appliqué à la famille $(r_0, \ldots, A^{k-1} r_0)$.

Soit e_k le k-ième vecteur de la base canonique de $\mathbb{R}^k$, et Id_k la matrice identité de taille $k \times k$. Alors, pour $1 \leq k \leq 1 + k_0$, on a

$$AV_k = V_k T_k + \hat{v}_{k+1} e_k^t \tag{3.34}$$

et

$$V_k^t A V_k = T_k \quad et \quad V_k^t V_k = \mathrm{Id}_k. \tag{3.35}$$

Démonstration. On introduit la suite de vecteurs w_j définie par $w_0 = 0$, $w_1 = r_0/\|r_0\|$ et, pour $j \geq 2$,

$$\hat{w}_j = Aw_{j-1} - \sum_{i=1}^{j-1} \langle Aw_{j-1}, w_i \rangle w_i \quad \text{et} \quad w_j = \frac{\hat{w}_j}{\|\hat{w}_j\|}. \tag{3.36}$$

On montrera ci-dessous que $w_j = v_j$. On montre par récurrence que les vecteurs w_j (tant qu'ils existent) sont orthonormés. Supposons que ce soit vrai jusqu'au rang $j-1$: pour tout $p, q \leq j-1$, on suppose que $\langle w_q, w_p \rangle = \delta_{qp}$. On prouve maintenant l'hypothèse de récurrence au rang j. Soit $p \leq j-1$: alors

$$\begin{aligned} \langle \hat{w}_j, w_p \rangle &= \langle Aw_{j-1}, w_p \rangle - \sum_{i=1}^{j-1} \langle Aw_{j-1}, w_i \rangle \langle w_i, w_p \rangle \\ &= \langle Aw_{j-1}, w_p \rangle - \langle Aw_{j-1}, w_p \rangle = 0, \end{aligned}$$

donc $\langle w_j, w_p \rangle = \delta_{pj}$ pour tout $p \leq j$, ce qui donne l'hypothèse de récurrence au rang j.

Par récurrence, on montre aussi que $w_j \in K_{j-1}$, tant que les vecteurs w_j existent.

Supposons maintenant que l'algorithme stoppe à l'indice j (c'est-à-dire que j est le premier indice tel que $\hat{w}_j = 0$), avec $j \leq k_0 + 1$. Alors

$$Aw_{j-1} = \sum_{i=1}^{j-1} \langle Aw_{j-1}, w_i \rangle w_i. \tag{3.37}$$

Or $w_i \in K_{i-1}$ pour tout $i \leq j-1$, donc on a $w_i = \sum_{p=0}^{i-1} \beta_i^p A^p r_0$. On insère cette décomposition dans (3.37), ce qui donne

$$\sum_{p=0}^{j-2} \beta_{j-1}^p A^{p+1} r_0 = \sum_{i=1}^{j-1} \langle Aw_{j-1}, w_i \rangle \sum_{p=0}^{i-1} \beta_i^p A^p r_0,$$

soit, en isolant le terme de plus haut degré à gauche,

$$\beta_{j-1}^{j-2} A^{j-1} r_0 = \sum_{i=1}^{j-1} \langle Aw_{j-1}, w_i \rangle \sum_{p=0}^{i-1} \beta_i^p A^p r_0 - \sum_{p=0}^{j-3} \beta_{j-1}^p A^{p+1} r_0.$$

Le vecteur du membre de droite est dans K_{j-2}. Comme $j-1 \leq k_0$, la famille $(r_0, \ldots, A^{j-1}r_0)$ est libre, donc $A^{j-1}r_0 \notin K_{j-2}$. Donc $\beta_{j-1}^{j-2} = 0$. Par conséquent, la décomposition de w_{j-1} s'écrit

$$w_{j-1} = \sum_{p=0}^{j-3} \beta_i^p A^p r_0 \in K_{j-3}.$$

Donc la famille $(w_1, \ldots, w_{j-1})$ est une famille de $j-1$ vecteurs orthogonaux deux à deux et qui appartiennent tous à K_{j-3}, qui est de dimension $j-2$. Ceci est contradictoire : donc l'algorithme stoppe à un indice $j > k_0 + 1$.

Supposons maintenant que $\hat{w}_{k_0+2} \neq 0$. Alors la famille $(w_1, \ldots, w_{k_0+2})$ est une famille de k_0+2 vecteurs orthogonaux deux à deux et qui appartiennent tous à $K_{k_0+1} = K_{k_0}$, qui est de dimension k_0+1. Ceci est à nouveau contradictoire. Donc l'algorithme stoppe exactement à l'indice k_0+2.

Pour tout $j \leq k_0+1$, la famille $(w_1, \ldots, w_j)$ est une famille de j vecteurs orthonormés et qui appartiennent tous à K_{j-1}, qui est de dimension j : donc cette famille constitue une base orthonormée de K_{j-1}, qui coïncide avec la base orthonormée construite par le procédé de Gram-Schmidt appliqué à la famille $(r_0, \ldots, A^{j-1}r_0)$.

On montre maintenant que $w_j = v_j$ pour tout $j \leq k_0+1$. Comme A est symétrique, on a

$$\begin{aligned} \langle Aw_p, w_{j-1} \rangle &= \langle w_p, Aw_{j-1} \rangle \\ &= \langle w_p, \hat{w}_j \rangle + \sum_{i=1}^{j-1} \langle Aw_{j-1}, w_i \rangle \langle w_p, w_i \rangle. \end{aligned}$$

Supposons $j \leq p-1$: alors, pour les i tels que $1 \leq i \leq j-1$, on a $i \leq p-2 < p$ et $\langle w_p, w_i \rangle = 0$. Donc, pour $j \leq p-1$, on a $\langle Aw_p, w_{j-1} \rangle = 0$. On voit aussi que

$$\langle Aw_p, w_{p-1} \rangle = \langle w_p, \hat{w}_p \rangle = \|\hat{w}_p\|.$$

Donc la récurrence (3.36) définissant $\hat{w}_j$ se récrit

$$\begin{aligned} \hat{w}_j &= Aw_{j-1} - \langle Aw_{j-1}, w_{j-1} \rangle w_{j-1} - \langle Aw_{j-1}, w_{j-2} \rangle w_{j-2} \\ &= Aw_{j-1} - \langle Aw_{j-1}, w_{j-1} \rangle w_{j-1} - \|\hat{w}_{j-1}\| w_{j-2}, \end{aligned}$$

ce qui est exactement la récurrence (3.33). Par conséquent, on a bien $w_j = v_j$ pour tout $j \leq k_0+1$.

On montre maintenant (3.34). La colonne p de la matrice AV_k est exactement, pour $1 \leq p \leq k$, égale à

$$\mathrm{Col}_p(AV_k) = Av_p = \hat{v}_{p+1} + \langle Av_p, v_p \rangle v_p + \|\hat{v}_p\| v_{p-1}.$$

Un simple calcul montre que les colonnes de $V_k T_k$ sont

$$\begin{aligned} \forall p, \quad 2 \leq p \leq k-1, \quad \mathrm{Col}_p(V_k T_k) &= \hat{v}_{p+1} + \langle A v_p, v_p \rangle v_p + \|\hat{v}_p\| v_{p-1}, \\ \mathrm{Col}_1(V_k T_k) &= \hat{v}_2 + \langle A v_1, v_1 \rangle v_1, \\ \mathrm{Col}_k(V_k T_k) &= \langle A v_k, v_k \rangle v_k + \|\hat{v}_k\| v_{k-1}. \end{aligned}$$

Enfin, la colonne p de $\hat{v}_{k+1} e_k^t$ est nulle si $p < k$, tandis que la colonne k vaut exactement $\hat{v}_{k+1}$. On a donc bien la relation (3.34).

Les vecteurs v_k étant orthogonaux deux à deux et de norme 1, on a $V_k^t V_k = \mathrm{Id}_k$. On multiplie enfin à gauche la relation (3.34) par V_k^t : du fait que $\hat{v}_{k+1}$ est orthogonal aux v_j pour $j \leq k$, on a $V_k^t \hat{v}_{k+1} = 0$ et on obtient finalement la relation (3.35). □

Nous comparons maintenant les valeurs propres de A et celle de la matrice T_{k_0+1}. Notons que ces deux matrices ne sont pas en général de même taille. On note $\lambda_1 < \lambda_2 < \ldots < \lambda_m$ les valeurs propres distinctes de la matrice A qui est de taille $n \times n$ (donc $1 \leq m \leq n$), et soient P_i les matrices de projection orthogonale sur les sous-espaces propres correspondants de A. Par construction,

$$A = \sum_{i=1}^{m} \lambda_i P_i, \quad \mathrm{Id}_n = \sum_{i=1}^{m} P_i, \quad P_i P_j = 0 \text{ si } i \neq j, \quad P_i^2 = P_i \text{ pour tout } i.$$

Lemme 3.19. *Les valeurs propres de T_{k_0+1} sont aussi valeurs propres de A.*

Réciproquement, si on suppose que $P_i r_0 \neq 0$ pour tout i, alors toutes les valeurs propres de A sont aussi valeurs propres de T_{k_0+1} et $k_0 + 1 = m$. Les valeurs propres de T_{k_0+1} sont simples.

Dans le cas où $P_i r_0 \neq 0$ pour tout i, la récurrence de Lanczos permet donc de construire une matrice T_{k_0+1} qui est tridiagonale et dont les valeurs propres sont exactement les valeurs de A. On pourrait alors penser calculer les valeurs propres de A de la façon suivante :

- on applique la récurrence de Lanczos jusqu'à l'ordre $k_0 + 1$, ce qui permet de construire la matrice T_{k_0+1}.
- on calcule les valeurs propres de la matrice T_{k_0+1}. Le problème sur T_{k_0+1} est plus simple que le problème initial sur A, car T_{k_0+1} est tridiagonale et il existe des algorithmes pour le calcul des valeurs propres qui sont spécifiques aux matrices tridiagonales, comme l'algorithme de Givens.
- comme (dans les bons cas) T_{k_0+1} a exactement les mêmes valeurs propres que A, on a ainsi calculé les valeurs propres de A.

Une telle approche n'est cependant pas la meilleure façon d'exploiter la récurrence de Lanczos, à cause d'instabilités numériques liées à des erreurs d'arrondi. Une bonne façon d'exploiter la récurrence de Lanczos sera donnée par le lemme 3.20 ci-dessous. On démontre maintenant le lemme 3.19.

Démontration du lemme 3.19. Soit λ valeur propre de T_{k_0+1}, et soit $y \neq 0$ un vecteur propre associé : $T_{k_0+1} y = \lambda y$. Comme $\hat{v}_{k_0+2} = 0$, on déduit de (3.34) que

$AV_{k_0+1} = V_{k_0+1}T_{k_0+1}$, et donc que

$$AV_{k_0+1}y = \lambda V_{k_0+1}y.$$

Si $V_{k_0+1}y = 0$, alors les colonnes de V_{k_0+1} sont liées (puisque $y \neq 0$), ce qui est contradictoire avec le fait que la famille $(v_1, \ldots, v_{k_0+1})$ forme une base orthonormée de K_{k_0}. Donc $V_{k_0+1}y \neq 0$, et λ est valeur propre de A.

Réciproquement, on suppose que $P_i r_0 \neq 0$ pour tout $1 \leq i \leq m$. Supposons la famille $(P_1 r_0, \ldots, P_m r_0)$ liée : alors, par exemple, il existe $\alpha_1, \ldots, \alpha_{m-1}$ tels que

$$P_m r_0 = \sum_{i=1}^{m-1} \alpha_i P_i r_0.$$

Comme $P_m P_i = 0$ pour tout $i < m$, on obtient $0 = P_m^2 r_0 = P_m r_0$, ce qui est contradictoire avec les hypothèses. Donc la famille $(P_1 r_0, \ldots, P_m r_0)$ est libre et $E_m =$ Vect $\{P_1 r_0, \ldots, P_m r_0\}$ est de dimension m.

Montrons que $m = k_0 + 1$. On voit que $A^k r_0 = \sum_{i=1}^m \lambda_i^k P_i r_0$ donc $A^k r_0 \in E_m$, et par conséquent $K_k \subset E_m$ pour tout k. Donc $k_0 + 1 = \dim K_{k_0} \leq \dim E_m = m$.

On montre l'inégalité inverse. La famille $(P_1 r_0, \ldots, P_m r_0)$ est libre. On se place dans cette base. La famille $(r_0, \ldots, A^{m-1} r_0)$ est représentée dans cette base par la matrice

$$M = \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{m-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_m & \lambda_m^2 & \ldots & \lambda_m^{m-1} \end{pmatrix},$$

qui est une matrice de Van Der Monde inversible car les λ_j sont distincts deux à deux. Donc la famille $(r_0, \ldots, A^{m-1} r_0)$ est libre, ce qui implique $m - 1 \leq k_0$. On a donc bien $m = k_0 + 1$ et $E_m = K_{k_0}$.

Soit λ_i une valeur propre de A : le vecteur $P_i r_0$ est vecteur propre associé. Or $P_i r_0 \in E_m = K_{k_0}$, et les colonnes de V_{k_0+1} forment une base orthonormée de K_{k_0}. Donc il existe $y \neq 0$ tel que $V_{k_0+1}y = P_i r_0$. La relation (3.35) donne

$$\begin{aligned} T_{k_0+1}y &= V_{k_0+1}^t A V_{k_0+1} y \\ &= V_{k_0+1}^t A P_i r_0 \\ &= \lambda_i V_{k_0+1}^t P_i r_0 \\ &= \lambda_i V_{k_0+1}^t V_{k_0+1} y = \lambda_i y, \end{aligned}$$

donc λ_i est aussi valeur propre de T_{k_0+1}. La matrice A possède $m = k_0 + 1$ valeurs propres distinctes, et toutes ces valeurs propres sont aussi valeurs propres de T_{k_0+1}, qui est de dimension $k_0 + 1$. Donc les valeurs propres de T_{k_0+1} sont simples. □

Comme nous l'avons précisé plus haut, la bonne façon d'exploiter la récurrence de Lanczos n'est pas de calculer la matrice T_{k_0+1} pour ensuite la diagonaliser. Il est plus intéressant d'exploiter le lemme que nous donnons maintenant :

Lemme 3.20. *Soit un entier k, $1 \leq k \leq k_0 + 1$. Soit λ valeur propre de T_k et soit $y \in \mathbb{R}^k$ un vecteur propre associé. Alors il existe une valeur propre λ_i de la matrice A telle que*

$$|\lambda - \lambda_i| \leq \sqrt{m}\ \|\hat{v}_{k+1}\|\ \frac{|\langle e_k, y\rangle|}{\|y\|},$$

où e_k est le k-ième vecteur de la base canonique de $\mathbb{R}^k$.

Ce lemme vient compléter la discussion qui fait suite au lemme 3.19. Une façon efficace d'utiliser la récurrence de Lanczos est en effet la suivante : si la dernière composante d'un vecteur propre de T_k est petite, i.e. $|\langle e_k, y\rangle| \ll \|y\|$, alors la valeur propre correspondante est une bonne approximation d'une valeur propre de A. Ainsi, le calcul (d'une approximation) des valeurs propres de A passe toujours par la diagonalisation de la matrice T_k. Cependant, le lemme ci-dessus donne une estimation d'erreur qu'il est possible d'évaluer en pratique.

Démonstration. Soit λ valeur propre de T_k et y vecteur propre associé : $T_k y = \lambda y$. La relation (3.34) donne

$$AV_k y = \lambda V_k y + \langle y, e_k\rangle \hat{v}_{k+1}.$$

En utilisant les projections P_i, on a donc

$$\sum_{i=1}^{m} (\lambda_i - \lambda) P_i V_k y = \langle y, e_k\rangle \hat{v}_{k+1}.$$

Soit $\varepsilon_j = \text{signe}\,(\lambda_j - \lambda)$, on prend le produit scalaire de l'égalité ci-dessus avec $\varepsilon_j P_j V_k y$:

$$\varepsilon_j (\lambda_j - \lambda) \|P_j V_k y\|^2 = \langle y, e_k\rangle \varepsilon_j \langle \hat{v}_{k+1}, P_j V_k y\rangle.$$

On somme sur les j, avec $\varepsilon_j(\lambda_j - \lambda) = |\lambda_j - \lambda| \geq \min_i |\lambda_i - \lambda|$:

$$\min_i |\lambda_i - \lambda| \sum_{j=1}^{m} \|P_j V_k y\|^2 \leq \langle y, e_k\rangle \sum_{j=1}^{m} \varepsilon_j \langle \hat{v}_{k+1}, P_j V_k y\rangle.$$

Or $\sum_{j=1}^{m} \|P_j V_k y\|^2 = \|V_k y\|^2 = \|y\|^2$, donc

$$\begin{aligned}
\min_i |\lambda_i - \lambda| &\leq \frac{|\langle e_k, y\rangle|}{\|y\|^2} \left| \sum_{j=1}^{m} \varepsilon_j \langle \hat{v}_{k+1}, P_j V_k y\rangle \right| \\
&\leq \frac{|\langle e_k, y\rangle|}{\|y\|^2} \sum_{j=1}^{m} \|\hat{v}_{k+1}\|\ \|P_j V_k y\| \\
&\leq \frac{|\langle e_k, y\rangle|}{\|y\|^2} \|\hat{v}_{k+1}\| \sqrt{m} \sqrt{\sum_{j=1}^{m} \|P_j V_k y\|^2}.
\end{aligned}$$

En utilisant à nouveau $\sum_{j=1}^{m} \|P_j V_k y\|^2 = \|y\|^2$, on obtient le résultat annoncé. □

Chapitre 4

Introduction aux lois de conservation

Ce chapitre est une brève introduction à l'étude mathématique d'un type précis d'équations d'évolution, les *lois de conservation*, qui interviennent naturellement dans différentes modèles physiques (quelques exemples seront donnés par la suite).

Donnons pour commencer une manière heuristique de dériver physiquement ce type d'équations, au travers d'un exemple jouet de trafic routier. Supposons que l'on considère une route rectiligne et infinie, sur laquelle il y a une unique voie de circulation (il est donc impossible de dépasser). On considère alors deux quantités d'intérêt :

— La *densité* de véhicules à l'instant t et au point x (autrement dit, on adopte un point de vue dit macroscopique où on ne "compte" pas les véhicules un à un, mais on en compte la quantité moyenne par unité de volume), notée $\rho(t,x)$. Le nombre de véhicules entre les positions a et b à l'instant t est donc donné par la quantité $\int_a^b \rho(t,x)dx$.

— La vitesse du véhicule qui passe au point x à l'instant t, notée $v(t,x)$.

On s'intéresse alors aux équations que satisfont ρ et v. Nous allons dériver plusieurs points de vues qui seront utiles par la suite.

La première idée est de comparer le nombre de véhicules dans une certain portion $[a,b]$ entre les instants proches t et $t+\delta t$, puis de faire tendre δt vers 0 pour obtenir la "variation instantanée" des quantités d'intérêt (c'est un raisonnement très souvent utilisé en physique). La variation du nombre de véhicules entre a et b est alors donnée par

$$\int_a^b \rho(t+\delta t,x)dx - \int_a^b \rho(t,x)dx.$$

Comme on suppose qu'aucun véhicule ne peut disparaître ou apparaître, cette quantité correspond aussi au nombre de véhicules entrants moins le nombre de

véhicules sortants. En supposant que les véhicules vont de la gauche vers la droite, le nombre de véhicules entrants est donné la quantité de véhicules qui entrent au point a entre les temps t et $t+\delta t$, qui est donnée par $\rho(t,a)(\delta t \cdot v(t,a))$. De même, le nombre de véhicules sortants est donné par $\rho(t,b)(\delta t \cdot v(t,b))$. On en déduit donc la formule

$$\int_a^b \rho(t+\delta t,x)dx - \int_a^b \rho(t,x)dx = \rho(t,a)(\delta t v(t,a)) - \rho(t,b)(\delta t v(t,b)).$$

On divise par δt. On en déduit

$$\begin{aligned}\int_a^b \frac{\rho(t+\delta t,x)-\rho(t,x)}{\delta t}dx &= \rho(t,a)(v(t,a)) - \rho(t,b)(v(t,b)) \\ &= -\int_a^b \frac{\partial}{\partial x}(\rho(t,x)v(t,x))dx.\end{aligned} \tag{4.1}$$

En faisant $\delta t \to 0$, on obtient

$$\int_a^b \left(\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x}\right) = 0.$$

Ceci étant vrai pour tout a, b, on peut par exemple utiliser le théorème fondamental de l'analyse, en fixant a et en dérivant par rapport à b, pour en déduire que

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0. \tag{4.2}$$

On a alors une seule équation pour les deux inconnues ρ et v, ce qui ne permettra pas d'assurer l'unicité des solutions à ce problème. Autrement dit, il faut "fermer" ce système en rajoutant une loi de comportement qui permette de lier ρ à v. Autrement dit, comment les véhicules adaptent-ils leur vitesse à la densité de véhicules et donc au trafic ?

Dans le cas où par exemple il y a très peu de trafic, les véhicules n'interagissent pas entre eux, on peut donc supposer que les véhicules vont à la vitesse maximum autorisée (qui dépend éventuellement de l'endroit où on est sur la route). On a donc dans ce cas $\rho(x) = V_{max}(x)$ et une équation de transport dite conservative

$$\frac{\partial}{\partial t}\rho(t,x) + \frac{\partial}{\partial x}(V_{max}(x)\rho(t,x)) = 0.$$

On peut aussi supposer que la vitesse dépend directement de la densité de véhicules et considérer $v = g(\rho)$ avec $g : \mathbb{R} \to \mathbb{R}$ une fonction donnée. On obtient alors une équation non linéaire sur la densité de la forme

$$\frac{\partial \rho}{\partial t} + \frac{\partial(f(\rho))}{\partial x} = 0, \tag{4.3}$$

où $f(\rho) = \rho g(\rho)$. Par exemple, un flux très souvent utilisé dans les modèles de trafic routier est $g = V_{max}(1 - \rho)$.

Un autre point de vue possible est de s'intéresser plutôt à un véhicule particulier roulant sur la route dans une approche plus "trajectorielle". On suppose qu'en temps 0, le véhicule est en position x_0. On appelle $X(t, x_0)$ la trajectoire de ce véhicule (donc $X(t, x_0)$ est la position du véhicule au temps t). Par définition de la vitesse, clairement, on a

$$\begin{aligned} \frac{d}{dt}X(t, x_0) &= v(t, X(t, x_0)), \\ X(t_0, x_0) &= x_0. \end{aligned} \tag{4.4}$$

X est ce qu'on appelle le *flot* associé au champ de vitesse v. On peut alors retrouver l'équation (4.2) de la manière suivante : on considère deux véhicules x_1 et x_2, alors le nombre de véhicules entre x_1 et x_2 est nécessairement constant, donc

$$\frac{d}{dt}\int_{X(t,x_1)}^{X(t,x_2)} \rho(t,)dx = 0.$$

On a donc, en utilisant un théorème de dérivation composée et en dérivant sous le signe intégral,

$$\frac{d}{dt}X(t,)\rho(t, X(t, x_2)) - \frac{d}{dt}X(t, X(t, x_1))\rho(t, X(t, x_1)) + \int_{X(t,x_1)}^{X(t,x_2)} \frac{\partial \rho}{\partial t}(t, x)dx = 0.$$

En revenant à (4.4),

$$v(t, X(t, x_2))\rho(t, X(t, x_2)) - v(t, X(t, x_1))\rho(t, X(t, x_1)) + \int_{X(t,x_1)}^{X(t,x_2)} \frac{\partial \rho}{\partial t}(t, x)dx = 0,$$

i.e.

$$\int_{X(t,x_1)}^{X(t,x_2)} \partial_x \left(v(t, X(t, x))\rho(t, x) + \frac{\partial \rho}{\partial t}(t, x) \right) dx = 0,$$

et un raisonnement similaire à celui pour trouver (4.4) permet de retrouver (4.2).

Pour conclure, justifions la terminologie "Loi de conservation". On revient à l'équation (4.1). En intégrant à droite en espace, en supposant que

$$\lim_{x \to \pm\infty} f(\rho u(t, x)) = 0, \ \forall t \geqslant 0,$$

et enfin en supposant $\rho_0(x) := \rho(0, x)$ d'intégrale finie (autrement dit il y a un nombre fini de véhicules), on en déduit en faisant $a, b \to +\infty$ que l'intégrale de ρ est conservée au cours du temps :

$$\forall t \in \mathbb{R}^+, \int_{\mathbb{R}} \rho(t, x)dx = \int_{\mathbb{R}} \rho(0, x)dx.$$

4.1 L'équation de transport linéaire

4.1.1 Quelques rappels et compléments sur les équations différentielles ordinaires (EDO)

On verra que l'étude des équations de transport et des lois de conservations scalaires repose beaucoup sur une méthode qui suppose naturellement de savoir résoudre des EDO éventuellement non linéaires. Donnons donc quelques résultats classiques que nous utiliserons par la suite. On considère $\mathcal{O}$ un ouvert de $\mathbb{R}\times\mathbb{R}^d$ ($d \geqslant 1$) (la première variable joue le rôle du temps, la deuxième jouera essentiellement ici celle de l'espace).

Définition 4.1. *Soit $f : \mathcal{O} \to \mathbb{R}^d$. On dit que f est localement Lipschitzienne par rapport à sa seconde variable si pour tout $(t_0, x_0) \in \mathcal{O}$, il existe un voisinage $\mathcal{V}$ de (t_0, x_0), il existe $K > 0$ tel que pour tout $(t, x) \in \mathcal{V}$ et $(t, y) \in \mathcal{V}$, on ait*

$$||f(t, x) - f(t, y)|| \leqslant K_{\mathcal{V}}||x - y||.$$

Supposons maintenant que $\mathcal{O} = I \times \mathbb{R}^d$, avec I un intervalle ouvert de $\mathbb{R}$. On dit que f est globalement Lipschitzienne par rapport à sa seconde variable si pour tout $(t_0, x_0) \in \mathcal{O}$, il existe un voisinage $\mathcal{W}$ de t_0, il existe $K_{\mathcal{W}} > 0$ tel que pour tout $t \in \mathcal{W}$ et tout $(x, y) \in \mathbb{R}^d$, on ait

$$||f(t, x) - f(t, y)|| \leqslant K_{\mathcal{W}}||x - y||.$$

Théorème 4.2 (Cauchy-Lipschitz)**.** *On considère $f : \mathcal{O} \to \mathbb{R}^d$ continue et localement Lipschtizienne par rapport à sa seconde variable. Alors, pour tout $(t_0, x_0) \in \mathcal{O}$, il existe un intervalle ouvert J contenant t_0 tel que le système différentiel*

$$\begin{cases} x'(t) = f(t, x(t)), \\ x(t_0) = x_0 \end{cases} \tag{4.5}$$

admette une unique solution y de classe C^1 sur J. De plus, si $\mathcal{O} = I \times \mathbb{R}^d$ avec I un intervalle ouvert de $\mathbb{R}$, et si f est continue, et globalement Lipschitzienne par rapport à la seconde variable, alors, il existe une unique solution à (4.5) définie globalement sur tout I.

Pour ce qui suit, nous allons avoir besoin de propriétés un peu plus fines des solutions. Cela suppose d'introduire la notion de flot associé à une équation différentielle, qui permet de prendre en compte la dépendance de la solution par rapport à la donnée initiale et au temps initial.

Définition 4.3. *On suppose f définie $\mathcal{O} = I \times \mathbb{R}^d$ avec I un intervalle ouvert de $\mathbb{R}$, et f continue et globalement Lipschitzienne par rapport à la seconde variable. On*

appelle flot global de l'équation (4.5) *l'application* $X : I \times I \times \mathbb{R}^d$ *qui soit telle que* $X(t, t_0, x_0) = y(t)$, *où* y *est la solution de* (4.5).

Autrement dit, $X(t, t_0, x_0)$ *est la valeur au temps* t *de l'unique solution de* $x'(t) = f(t, x(t))$ *qui passe par la valeur* x_0 *au temps* t_0.

L'introduction de cette notion va nous permettre de donner un sens à la régularité (continuité, dérivabilité) par rapport aux donnes initiales t_0 et x_0.

Théorème 4.4. *On suppose (en plus des hypothèses assurant l'existence du flot global) que* f *est de classe* C^p *sur* $I \times \mathbb{R}$ *pour* $p \in \mathbb{N}^*$. *Alors* X *est de classe* C^p.

Nous allons admettre ce théorème. Une fois ce théorème admis, il est facile d'en déduire les expressions des dérivées partielles de X, ce qu'on fera par la suite.

4.1.2 Équations de transport conservatives et non conservatives, cas d'une vitesse constante

En reprenant l'exemple précédent de trafic à vitesse maximale, on obtient le modèle suivant, en supposant que la vitesse maximale puisse dépendre du temps et de l'espace.

Définition 4.5. *On appelle équation de transport scalaire unidimensionnelle conservative une équation de la forme*

$$\begin{cases} \partial_t y(t,x) + \partial_x(a(t,x)y(t,x)) = 0,\ (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ y(0,x) = y^0(x), \end{cases} \tag{4.6}$$

où y^0 *est une condition initiale (on précisera plus tard dans quelle espace on la prendra) et* $a : \mathbb{R} \to \mathbb{R}$ *est un coefficient suffisamment régulier, appelé* vitesse.

On peut aussi introduire un modèle légèrement différent.

Définition 4.6. *On appelle équation de transport scalaire unidimensionnelle non conservative une équation de la forme*

$$\begin{cases} \partial_t y(t,x) + a(t,x)\partial_x y(t,x) = 0,\ (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ y(0,x) = y^0(x), \end{cases} \tag{4.7}$$

où y^0 *est une condition initiale et* $a : \mathbb{R} \to \mathbb{R}$ *est un coefficient suffisamment régulier, appelé* vitesse.

Remarque 4.7. — *L'équation de transport conservative* (4.6) *correspond à une vraie situation "physique"" de transport de particules, concentration, etc. présentée précédemment. Pour comprendre comment résoudre* (4.6) *une fois que l'on sait résoudre l'équation* (4.7)*, on pourra faire l'Exercice 26.*

— *L'équation de transport non conservative* (4.7) *a moins de sens physique, mais elle est un peu plus simple à étudier et constituent un préliminaire indispensable à l'étude des lois de conservation non linéaires qui seront étudiées par la suite.*
— *Ces deux équations sont des équations aux dérivées partielle dites* linéaires *(au sens où l'ensemble des solutions forme un espace vectoriel, une combinaison linéaire de solutions reste une solution).*

Avant de se lancer dans une étude générale de ces équations, commençons par regarder le cas d'une vitesse constante :

$$\begin{cases} \partial_t y(t,x) + a\partial_x y(t,x) = 0,\ (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ y(0,x) = y^0(x),\ x \in \mathbb{R}, \end{cases} \tag{4.8}$$

où a est une constante (non nulle, sinon, la solution reste constante à sa valeur initiale). Supposons y^0 de classe C^1. Alors, on aurait envie de dire que comme on transporte la condition initiale à vitesse a constante, une solution du problème au temps $t > 0$ est donnée par une certaine translation bien choisie de la condition initiale, sachant que la "'distance" parcourue est at. Il est donc raisonnable de penser qu'une solution (en fait, ce serait la seule, mais on le démontrera proprement plus tard) est donnée par

$$y(t,x) = y^0(x - at).$$

Il est clair que $y(0,x) = y^0(x)$, que y^0 est de classe C^1 et que par dérivation de fonctions composées, on a

$$\partial_x y(t,x) = (y^0)'(x-at) \ \text{ et } \partial_t y(t,x) = -a(y^0)'(x-at),$$

donc (4.8) est bien vérifié.

Une autre manière (qui nous sera utile par la suite) pour aboutir à ces solutions est de rechercher les *courbes caractéristiques* de l'équation, autrement dit les courbes le long de laquelle la solution reste constante au cours du temps. Comme on considère des équations qui "transportent" des quantités, il est raisonnable de penser que de telles courbes existent. On appelle $(s, X(s))$ une telle courbe caractéristique, et pour éviter de tout confondre, nous allons réserver les notations (t,x) aux solutions de l'équation (4.8). On souhaite que y soit constante le long de la trajectoire, autrement dit que

$$\frac{d}{ds} y(s, X(s)) = 0.$$

En utilisant le théorème de dérivation des fonctions composées, on doit donc avoir que

$$\partial_t y(s, X(s)) + X'(s)\partial_x y(s, X(s)) = 0.$$

Compte-tenu de l'équation (4.8) vérifiée par y, les courbes caractéristiques "doivent" donc vérifier $X'(s) = a$. Donc $X(s) = as + c$ pour un certain c. Reste à déterminer

c. Si l'on souhaite avoir la valeur de la solution y à (4.8) au temps (t, x), on doit être sûr que X passe par x au temps t. c doit donc être choisi de telle sorte que $X(t) = x$, *i.e.* $c = x - at$. On a donc $X(s) = a(s - t) + x$. Enfin, y étant constante le long des caractéristiques, on a pour tout s que

$$y(s, X(s)) = y(t, X(t)) = y(t, x).$$

Il nous reste à exploiter le fait que l'on connaisse la condition initiale. On choisit donc $s = 0$ et on en déduit que

$$y(t, x) = y(0, X(0)) = y^0(x - at).$$

On voit donc que la méthode revient à résoudre une certaine EDO en forçant une position t en temps x, puis "retourner" en arrière pour aller chercher la valeur de la solution en temps initial et pouvoir exploiter le fait que l'on sait ce que vaut y au temps $t = 0$.

Cette méthode de résolution est appelée *méthode des caractéristiques*, et sera utilisée fortement dans la suite.

4.1.3 Solutions classiques dans le cas non conservatif

La méthode des caractéristiques que nous venons d'expliquer va nous permettre de démontrer le théorème suivant.

Théorème 4.8. *On suppose que y^0 est de classe C^1 sur $\mathbb{R}$. On suppose que a est C^1 sur $\mathbb{R}^2$ et globalement Lipschitzienne par rapport à la seconde variable. Alors il existe une unique solution de classe C^1 sur $\mathbb{R}^2$ à (4.7). De plus, cette solution est donnée par la formule*

$$y(t, x) = y^0(X(0, t, x)), \tag{4.9}$$

où $X(0, t, x)$ est la valeur en 0 de la caractéristique associée à a passant par x au temps t (qui existe et est globalement définie sur $\mathbb{R}^3$, par le théorème de Cauchy-Lipschitz global).

Remarque 4.9. *On voit qu'ici, la solution est aussi définie en temps négatif. L'équation est dite réversible en temps. On aurait pu aussi tout à fait supposer a seulement définie sur $\mathbb{R}^+ \times \mathbb{R}$, cela ne changerait pas grand chose.*

Remarque 4.10. *Dans le cadre fonctionnel du Théorème 4.8, les solutions sont appelées* solutions classiques *ou* solutions fortes*.*

Preuve : Nous allons faire un raisonnement par analyse-synthèse : nous allons d'abord montrer que si une solution existe, elle ne peut être donnée que l'expression (4.9), puis nous allons montrer que cette expression fournit effectivement une solution au problème.

— Analyse : soit y une solution dérivable de (4.7). Soient $(t,x) \in \mathbb{R}^2$. On considère l'équation différentielle

$$\begin{cases} Y'(s) = a(s, Y(s)), \\ Y(t) = x. \end{cases} \tag{4.10}$$

Puisque a est globalement Lipzchitzienne par rapport à la deuxième variable, le théorème de Cauchy-Lipschtiz global nous donne l'existence et l'unicité d'une solution globale à (4.10). On appelle $X(s,t,x)$ le flot. Alors par définition $X(t,t,x) = x$. De plus, y est constante le long de X. En effet, pour tout $s \in \mathbb{R}$, on a, par dérivation de fonctions composées et en utilisant (4.10) et (4.7).

$$\begin{aligned} \frac{d}{ds} y(s, X(s)) &= \partial_t y(s, X(s)) + \partial_1 X(s,t,x) \partial_x y(s, X(s)) \\ &= \partial_t y(s, X(s,t,x)) + a(s, X(s,t,x)) \partial_x y(s, X(s,t,x)) \\ &= 0. \end{aligned}$$

Autrement dit,

$$y^0(X(0,t,x)) = y(0, X(0,t,x)) = y(t, X(t,t,x)) = y(t,x).$$

Ainsi, si une solution à (4.7) existe, elle ne peut être donnée que par l'expression (4.9).

— Synthèse. On pose

$$y(t,x) = y^0(X(0,t,x)).$$

La fonction $(s,t,x) \mapsto X(s,t,x)$ est de classe C^1 sur $\mathbb{R}^3$ par le théorème 4.4, ce qui va nous permettre de calculer des dérivées partielles. On a donc déjà que $y(t,x)$ est C^1 par théorème de composition. On commence par remarquer qu'on a la propriété suivante : pour tout $(t,x) \in \mathbb{R}^2$,

$$x = X(t, 0, X(0,t,x)). \tag{4.11}$$

On dérive ceci par rapport à t. On en déduit par dérivation de fonctions composées que pour tout $(t,x) \in \mathbb{R}^2$,

$$\partial_1 X(t, 0, X(0,t,x)) + \partial_2 X(0,t,x) \partial_3 X(t, 0, X(0,t,x)) = 0.$$

Or par (4.10),

$$\partial_1 X(t, 0, X(0,t,x)) = a(t, X(t, 0, X(0,t,x))) = a(t,x).$$

Donc

$$\partial_2 X(0,t,x) \partial_3 X(0, t, X(0,t,x)) = -a(t,x). \tag{4.12}$$

De même, en dérivant maintenant la relation (4.11) par rapport à x, on en déduit que pour tout $(t,x) \in \mathbb{R}^2$,

$$\partial_2 3X(0,t,x)\partial_3 X(t,0,X(0,t,x)) = 1. \tag{4.13}$$

On multiplie ceci par $a(t,x)$ et on somme avec (4.12) pour obtenir

$$\partial_3 X(t,0,X(0,t,x))\left(\partial_2 X(0,t,x) + a(t,x)\partial_3 X(0,t,x)\right) = 0.$$

Par (4.13), on a $\partial_3 X(t,0,X(0,t,x)) \neq 0$ et donc

$$\partial_2 X(0,t,x) + a(t,x)\partial_3 X(0,t,x) = 0. \tag{4.14}$$

On a donc bien que $y(t,x)$ vérifie (4.7) puisque

$$\partial_t y(t,x) = (y^0)'(X(0,t,x))\partial_2 X(0,t,x)$$

et

$$\partial_x y(t,x) = (y^0)'(X(0,t,x))\partial_3 X(0,t,x).$$

♢

4.1.4 Solutions faibles

Le but de ce paragraphe est d'expliquer comment donner un sens à des solutions *non regulières* de (4.7). Les motivations sont diverses. D'abord, il se peut très bien que l'on rencontre dans la nature des phénomènes discontinus (les ondes de choc comme le mur du son). En outre, dans le cas de l'équation de transport à vitesse constante (4.8), il est très facile de "transporter" une fonction qui serait non régulière : si l'on part d'une condition initiale discontinue (par exemple, une fonction en escalier avec un saut en $x = 0$) appelée y_0, alors $y_0(x - at)$ correspond bien à une translation de la solution à vitesse a, et est donc un bon candidat pour être une solution de l'équation (4.7). Toutefois, comme y_0 n'est pas dérivable, elle n a aucune chance de pouvoir vérifier (4.7) *stricto sensu*. Le but de ce paragraphe va donc de donner une nouvelle formulation de l'équation (4.7), appelée *formulation faible*. Bien sûr, par souci de cohérence, il faut que cette notion de solution soit compatible avec les solutions classiques (*i.e.* quand y_0 est de classe C^1), au sens où toute solution classique doit aussi être une solution faible, et que cette notion de solution faible permette aussi d'avoir une propriété d'unicité. La bonne manière est de faire un peu comme pour la théorie des distributions, et donc de tester l'équation sur des fonctions tests C^∞ bien choisies. On va donc démontrer la proposition cruciale suivante. Par simplicité, nous allons nous restreindre à l'équation (4.8).

Proposition 4.11. *Soit $y^0 \in C^1(\mathbb{R})$ et $y \in C^1(\mathbb{R}^2)$. Alors y est la solution de* (4.8) *si et seulement si*

$$\int_0^{+\infty}\int_{\mathbb{R}} y(t,x)(\partial_t\varphi(t,x)+a\partial_x\varphi(t,x))dxdt = -\int_{\mathbb{R}} y^0(x)\varphi(0,x)dx, \forall\varphi \in C_0^\infty(\mathbb{R}^+\times\mathbb{R}). \tag{4.15}$$

Remarque 4.12. *Attention, les fonctions de $C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$ ne sont pas forcément nulles en $t = 0$! Elles sont nulles en dehors d'un compact $K \subset \mathbb{R}^+ \times \mathbb{R}$. Par contre, les fonctions de $C_0^\infty(\mathbb{R}^{+*} \times \mathbb{R})$ sont nulles au voisinage de $t = 0$.*

Preuve : Le sens direct est assez simple et repose sur des intégrations par parties (IPP dans la suite). Soit $y^0 \in C^1(\mathbb{R})$ et y la solution de classe $C^1(\mathbb{R}^+ \times \mathbb{R})$ de (4.8). Soit $\varphi \in C_0^\infty([0, +\infty[\times\mathbb{R})$. On considère la première ligne de (4.8), on la multiplie par φ, puis on intègre. Comme tout le monde est de classe C^1 et que l'on travaille en réalité sur un compact de $\mathbb{R}^2$, cette opération a un sens. On obtient donc

$$\int_0^{+\infty} \int_{\mathbb{R}} \left(\partial_t y + a\partial_x y\right)(t,x)\varphi(t,x)dxdt = 0,$$

i.e.

$$\int_0^{+\infty} \int_{\mathbb{R}} \partial_t y(t,x)\varphi(t,x)dxdt + a\int_0^{+\infty} \int_{\mathbb{R}} \partial_x y(t,x)\varphi(t,x)dxdt = 0.$$

Dans la première intégrale, on applique le théorème de Fubini (valide car φ est à support compact donc on regarde bien une fonction intégrable) puis on fait une intégration par parties (IPP) en temps pour écrire

$$\begin{aligned}
\int_0^{+\infty} \int_{\mathbb{R}} \partial_t y(t,x)\varphi(t,x)dxdt &= \int_{\mathbb{R}} \left(\int_0^{+\infty} \partial_t y(t,x)dt\right) dx \\
&= \int_{\mathbb{R}} \left([\varphi(\cdot,x)y(\cdot,x)]_0^{+\infty} - \int_0^{+\infty} \partial_t\varphi(t,x)y(t,x)dt\right) dx \\
&= -\int_{\mathbb{R}} y^0(x)\varphi(0,x)dx - \int_{\mathbb{R}}\int_0^{+\infty} \partial_t\varphi(t,x)y(t,x)dt \\
&= -\int_{\mathbb{R}} y^0(x)\varphi(0,x)dx - \int_0^{+\infty}\int_{\mathbb{R}} \partial_t\varphi(t,x)y(t,x)dt..
\end{aligned}$$

Maintenant, dans la deuxième intégrale, on fait une IPP en espace à l'intérieur, de la même manière.

$$\begin{aligned}
a\int_0^{+\infty} \left(\int_{\mathbb{R}} \partial_x y(t,x)\varphi(t,x)dx\right) dt &= -a\int_0^{+\infty} \left(\int_{\mathbb{R}} y(t,x)\partial_x\varphi(t,x)dx\right) dt \\
&= -a\int_0^{+\infty} \int_{\mathbb{R}} y(t,x)\partial_x\varphi(t,x)dxdt.
\end{aligned}$$

On obtient donc exactement (4.15) en sommant les deux expressions obtenues.

Le sens réciproque est la partie non triviale de la démonstration. On aura besoin de la propriété suivante, appelée *lemme fondamental du calcul des variations.*

Lemme 4.13. *Soit Ω un ouvert de $\mathbb{R}^d$ ($d \in \mathbb{N}^*$) et $f \in L^1_{loc}(\Omega)$ (l'espace des fonctions intégrables sur tout compact inclus dans Ω). Alors*

$$f = 0 \quad p.p. \quad \Leftrightarrow \forall \psi \in C_0^\infty(\Omega),\ \int_\Omega f\psi = 0.$$

Preuve : Le sens direct est trivial, c'est le sens réciproque qui est intéressant. Une idée pour démontrer le sens réciproque serait de remarquer que si on pouvait remplacer ψ par f, alors on aurait $\int_\Omega |f|^2 = 0$ et donc $f = 0$ p.p. par un résultat classique d'intégration de Lebesgue. Toutefois, f n'est ni régulière, ni à support compact, et $|f|^2$ n 'est pas forcément intégrable. On va donc procéder d'une autre manière, pas très intuitive mais simple et élégante. On commence par se ramener au cas de $\mathbb{R}^d$ tout entier, en étendant f par 0 en dehors de Ω. alors $f \in L^1_{loc}(\mathbb{R}^d)$.

D'abord, on "rappelle" le résultat suivant (théorème de Lebesgue) : pour $f \in L^1_{loc}(\mathbb{R}^d)$ et pour presque tout $x \in \mathbb{R}^d$, on a

$$t^{-d} \int_{x+[-t,t]^d} |f(x) - f(y)| dy \to 0 \text{ quand } t \to 0. \tag{4.16}$$

Soit maintenant $\varphi \in C_0^\infty([-1,1]^d)$ étendu par 0 en dehors de $[-1,1]^d$ telle que $\int_{\mathbb{R}^d} \varphi = 1$. Soit $x \in \Omega$. Pour tout $t > 0$, on a alors par le changement de variable $y' = (x-y)/t$ (dont le jacobien vaut t^{-d}) que

$$\int_{\mathbb{R}^d} \varphi\left(\frac{x-y}{t}\right) t^{-d} dy = \int_{\mathbb{R}^d} \varphi(y') dy' = 1.$$

Ainsi, on a

$$\begin{aligned}
|f(x)| &= \left| \int_{\mathbb{R}^d} f(x) \varphi\left(\frac{x-y}{t}\right) t^{-d} dy \right| \\
&= \left| \int_{\mathbb{R}^d} (f(x) - f(y) + f(y)) \varphi\left(\frac{x-y}{t}\right) t^{-d} dy \right| \\
&\leqslant \left| \int_{\mathbb{R}^d} |f(x) - f(y)| \, |\varphi\left(\frac{x-y}{t}\right) t^{-d} \right| dy + \left| \int_{\mathbb{R}^d} f(y) \varphi\left(\frac{x-y}{t}\right) t^{-d} dy \right|.
\end{aligned}$$

$\varphi((x-\cdot)/t)$ étant encore $C_0^\infty(\mathbb{R}^d)$, en prenant t suffisamment petit pour que son support soit inclus dans Ω (c'est possible, son support étant $x + [-t,t]^d$ et Ω étant ouvert), on a par hypothèse que

$$\int_{\mathbb{R}^d} f(y) \varphi\left(\frac{x-y}{t}\right) t^{-d} dy = \int_\Omega f(y) \varphi\left(\frac{x-y}{t}\right) t^{-d} dy = 0.$$

Quant à la première intégrale, on remarque que l'on intègre en fait sur $x+[-t,t]^d \subset \Omega$ et dont on peut la majorer par

$$t^{-d} \int_{x+[-t,t]^d} |f(x) - f(y)| dy ||\varphi||_\infty$$

pour en déduire par le théorème de Lebesgue que cette quantité tend vers 0 quand $t \to 0$ pour presque tout x. On a donc bien $f(x) = 0$ p.p. comme voulu. $\diamondsuit$

Revenons à la preuve de la Proposition. On suppose que pour tout $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$, on a (4.15). Malheureusement, $\mathbb{R}^+ \times \mathbb{R}$ n'est pas un ouvert, on ne peut donc pas lui appliquer le lemme précédent. Ce n'est pas un énorme problème : on commence par prendre $\varphi \in C_0^\infty(\mathbb{R}^{+*} \times \mathbb{R})$ (qui est aussi clairement dans $C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$ en prolongeant par 0 en $t = 0$). On a alors $\varphi(0, x) = 0$ pour tout $x \in \mathbb{R}$, et donc, pour tout $\varphi \in C_0^\infty(\mathbb{R}^{+*} \times \mathbb{R})$, on a

$$\int_0^{+\infty} \int_{\mathbb{R}} y(t,x)(\partial_t \varphi(t,x) + a\partial_x \varphi(t,x))dxdt = 0.$$

En faisant les IPP inverses de celles du sens direct (qui sont valides car y et φ sont régulières et qu'on travaille en fait sur des segments), on obtient

$$\int_0^{+\infty} \int_{\mathbb{R}} \varphi(t,x)(\partial_t y(t,x) + a\partial_x y(t,x))dxdt = 0.$$

En appliquant le lemme fondamental du calcul des variations, on en déduit donc que $\partial_t y(t,x) + a\partial_x y(t,x) = 0$ p.p. sur $\mathbb{R}^{+*} \times \mathbb{R}$, et donc partout puisque l'expression de gauche est une fonction continue en (t, x). Il reste donc à montrer que $y(0, x) = y^0(x)$ pour conclure la preuve. Il est clair que dans cette optique, "supprimer" le bord en $t = 0$ à l'aide d'une fonction $\varphi \in C_0^\infty(\mathbb{R}^{+*} \times \mathbb{R})$ ne pourra jamais nous aider. On repart donc de la formulation de départ (4.15) avec $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$. En refaisant les mêmes IPP sur la première intégrale que précédemment, on obtient (il reste maintenant un terme de bord en $t = 0$)

$$\int_0^{+\infty} \int_{\mathbb{R}} \varphi \left(\partial_t y + a\partial_x y\right) - \int_{\mathbb{R}} \varphi(0,x)y(0,x)dx = -\int_{\mathbb{R}} y^0(x)\varphi(0,x)dx.$$

Mais on vient de prouver que la première intégrale est nulle. On obtient donc que pour tout $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$,

$$\int_{\mathbb{R}} \varphi(0,x)(y(0,x) - y^0(x))dx = 0.$$

Pour conclure, il est très tentant de vouloir utiliser une fois de plus le lemme fondamental du calcul des variations, mais ce n'est pas possible de manière immédiate, car il faudrait pouvoir tester sur toutes fonctions $\psi \in C_0^\infty(\mathbb{R})$, mais ici on teste sur les traces en temps $t = 0$ de fonctions $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$. Ces fonctions sont bien dans $C_0^\infty(\mathbb{R})$. Une question naturelle est donc de savoir si on peut "récupérer" toutes les fonctions $C_0^\infty(\mathbb{R})$ à l'aide de traces de la forme $\varphi(0, x)$. Autrement dit, de manière ensembliste, on voudrait que l'application

$$\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}) \mapsto \varphi(0,x) \in C_0^\infty(\mathbb{R})$$

soit surjective. C'est assez simple. On considère n'importe quelle fonction $\psi \in C_0^\infty(\mathbb{R})$. Maintenant, on fixe n'importe quel $\eta \in C_0^\infty([0, +\infty[)$ telle que $\eta(0) = 1$. Alors, il est clair que $\varphi(t, x) = \eta(t)\psi(x)$ est $C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$ et que $\varphi(0, x) = \psi(x)$.

♢

Il y a deux avantages majeurs à la formulation (4.15).

— La condition initiale est directement incorporée dans la formulation de l'équation.
— L'expression (4.15) à un sens même si y^0 n'est pas régulier ! On peut par exemple prendre prendre y^0 dans n'importe quel espace L^p, et même $y^0 \in L^1_{loc}(\mathbb{R})$. C'est l'intérêt majeur de cette formulation faible.

Cela suggère donc la définition suivante. Pour simplifier les choses, on va se placer dans le cas de données initiales L^∞, et on va uniquement traiter le cas des vitesses constantes. Le cas des vitesses variables se traite de la même manière (sous des hypothèses appropriées de régularité sur $a(t,x)$), mais les calculs seraient plus lourds.

Définition 4.14. *Soit $y^0 \in L^\infty(\mathbb{R})$. Une fonction $y \in L^\infty(\mathbb{R}^+ \times \mathbb{R})$ est une solution faible de* (4.8) *si* (4.15) *est vérifiée.*

Cette définition est en cohérence avec les solutions usuelles, au sens où les solutions usuelles vérifient (4.15) et que (4.15) caractérise les solutions usuelles dans le cas où $y^0 \in C^1(\mathbb{R})$.

On a alors le théorème d'existence et d'unicité suivant.

Théorème 4.15. *Soit $y^0 \in L^\infty(\mathbb{R})$. Il existe une unique solution à* (4.15) *donnée par*

$$y(t,x) = y^0(x - at),$$

et cette solution est dans $L^\infty(\mathbb{R}^+ \times \mathbb{R})$.

Remarque 4.16. *Bien sûr, cette expression fournit aussi une unique solution en tout temps, même négatif.*

Preuve : Puisque le théorème suggère quelle est la bonne forme de la solution, posons $y(t,x) = y^0(x-at)$ et regardons si (4.15) est vérifiée. Soit $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$. Par le changement de variable en espace $x' = x - at$ (qui ne change par les bornes d'intégrations), on a

$$\begin{aligned}&\int_0^{+\infty}\int_{\mathbb{R}} y^0(x-at)(\partial_t\varphi(t,x) + a\partial_x\varphi(t,x))dxdt \\ &= \int_0^{+\infty}\int_{\mathbb{R}} y^0(x')(\partial_t\varphi(t,x'+at) + a\partial_x\varphi(t,x'+at))dxdt.\end{aligned}$$

Posons maintenant $\psi(t,x') = \varphi(x'+at)$. Par dérivation de fonctions composées, on obtient que

$$\partial_t\psi(t,x') = (\partial_t\varphi(t,x'+at) + a\partial_x\varphi(t,x'+at)).$$

L'expression précédente se réécrit donc

$$\int_0^{+\infty}\int_{\mathbb{R}} y^0(x-at)(\partial_t\varphi(t,x) + a\partial_x\varphi(t,x))dxdt = \int_0^{+\infty}\int_{\mathbb{R}} y^0(x')\partial_t\psi(t,x')dx'dt.$$

En appliquant le théorème de Fubini (valide car ψ est à support compact), on obtient (seul le terme de bord en $t = 0$ reste)

$$\int_0^{+\infty} \int_{\mathbb{R}} y^0(x - at)(\partial_t \varphi(t, x) + a\partial_x \varphi(t, x))dxdt = \int_{\mathbb{R}} y^0(x') \left(\int_0^{+\infty} \partial_t \psi(t, x)dt \right) dx$$
$$= - \int_{\mathbb{R}} y(x')\psi(0, x')dx'.$$

On a donc bien que (4.15) est vérifié.

L'unicité est un peu plus difficile à obtenir. On ne peut pas faire comme dans le cas des solutions régulières, car on ne peut pas faire d'intégrations par parties pour des fonctions seulement dans L^∞.

On prend $y_1, y_2 \in L^\infty(\mathbb{R}^+ \times \mathbb{R})$ vérifiant (4.15), avec bien sûr la même condition initiale y^0. Alors, par linéarité de l'intégrale, la différence $z = y_1 - y_2$ vérifie que pour tout $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$, on a

$$\int_0^{+\infty} \int_{\mathbb{R}} z(t, x)(\partial_t \varphi(t, x) + a\partial_x \varphi(t, x))dxdt = 0.$$

Encore une fois, on souhaiterait pouvoir utiliser le lemme fondamental du calcul des variations, mais pour ce faire, il faudrait être capable de "faire en sorte" que $\partial_t \varphi(t, x) + a\partial_x \varphi(t, x)$ soit égal à n'importe quelle fonction quelconque de $C_0^\infty(\mathbb{R}^{+*} \times \mathbb{R})$. Autrement, dit, si l'on se donne $\psi \in C_0^\infty(\mathbb{R}^{+*} \times \mathbb{R})$, peut-on trouver $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$ telle que

$$\partial_t \varphi(t, x) + a\partial_x \varphi(t, x) = \psi(t, x)? \tag{4.17}$$

Il est assez simple de résoudre cette équation en

$$\varphi(t, x) = - \int_t^{+\infty} \psi(s, x + a(s - t))dt.$$

Alors :

1. φ est bien C^∞. C'est assez intuitif mais pas complètement élémentaire à démontrer, puisque l'on a du temps en même temps dans la borne et à l'intérieur de l'intégrale. On ne peut donc pas appliquer directement les résultats habituels de dérivation sous le signe intégral. On va en fait "découpler" les deux problèmes en séparant ce qui se passe au bord de l'intervalle et à l'intérieur, en utilisant un argument de type fonctions composées. On écrit

   $$\varphi(t, x) = f(t, t, x), \tag{4.18}$$

   avec

   $$f(\alpha, \beta, x) = - \int_\alpha^{+\infty} \psi(s, x + a(s - \beta))dt.$$

   Comme on intègre en fait sur un segment des fonctions C^∞, les théorèmes usuels de dérivation sous le signe somme s'appliquent pour dire que toutes les

dérivées croisées par rapport à β et x existent et sont continues par rapport aux trois variables (α, β, x).

De plus, par le théorème fondamental de l'analyse, f est dérivable par rapport à α, et

$$\partial_\alpha f(\alpha, \beta, x) = \psi(\alpha, x + a(\alpha - \beta)). \tag{4.19}$$

Une telle fonction est alors clairement de classe C^∞ par rapport à chacune des variables. Ainsi, au total, que l'on ait dérivé une fois par rapport à α ou non, toutes les dérivées partielles existent et sont continues, donc f est bien de classe C^∞. Par composition, c'est aussi le cas pour φ.

2. De plus, on a

$$\partial_\beta f(\alpha, \beta, x) = \int_\alpha^{+\infty} a\partial_x \psi(s, x + a(s - \beta))dt \tag{4.20}$$

et

$$\partial_x f(\alpha, \beta, x) = -\int_\alpha^{+\infty} a\partial_x \psi(s, x + a(s - t))dt. \tag{4.21}$$

Par dérivation de fonctions composées, en utilisant (4.19) et (4.20), on en déduit par (4.18) que

$$\partial_t \varphi(t, x) = \partial_\alpha f(t, t, x) + \partial_\beta f(t, t, x) = \psi(t, x) + \int_\alpha^{+\infty} a\partial_x \psi(s, x + a(s - t))dt.$$

De même, en utilisant (4.20), on a

$$\partial_x \varphi(t, x) = \partial_x f(t, t, x) = -\int_t^{+\infty} a\partial_x \psi(s, x + a(s - t))dt.$$

Donc (4.17) est bien vérifié.

3. Enfin, φ est à support compact. Ceci se vérifie aisément sur l'expression explicite de φ. En effet, on suppose que le support de ψ est inclus dans $[0, A] \times [-B, B]$, avec $A, B > 0$.

Il est clair que pour $t \geqslant A$, on a pour $s \geqslant t$ que $\psi(s, x + a(t - s)) = 0$ et donc $\varphi = 0$. Maintenant, pour $x \geqslant B$, on a à t fixé et pour $s \geqslant t$ que $x + a(s - t) \geqslant B$ donc on a aussi que $\psi(s, x + a(s - t)) = 0$ et donc $\varphi = 0$. Pour conclure, remarquons qu'on a pour $t \leqslant A$ que

$$\varphi(t, x) = \int_t^A \psi(s, x + a(s - t))dt.$$

Donc on a toujours $x + a(s - t) \leqslant x + aA$, et donc pour $x \leqslant -B - aA$, on a bien$\psi(s, x + a(s - t))dt = 0$ sur $[t, A]$ et donc $\varphi = 0$.

$\diamondsuit$

4.2 Introduction aux lois de conservations scalaire non linéaires : l'exemple de l'équation de Burgers

4.2.1 Solutions classiques de l'équation de Burgers

On considère un flux v donné par $v(\rho) = \frac{\rho}{2}$. Dans ce cas, l'équation

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0$$

devient (en changeant le ρ en y)

$$\begin{cases} \partial_t y(t,x) + \dfrac{1}{2}\partial_x(y^2)(t,x) = 0,\ (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ y(0,x) = y^0(x), \end{cases} \tag{4.22}$$

Nous n'allons pas expliquer trop en détail d'où vient cette équation. Elle correspondrait à une approximation unidimensionnelle d'une équation célèbre de la mécanique des fluides (l'équation d'Euler) dans un régime supersonique, permettant de faire disparaître le terme de pression. Il s'agit maintenant d'une équation non linéaire en l'état y. Cela va poser un certain nombre de difficultés spécifiques. On remarque qu'en développant la dérivée en espace, c'est une sorte d'équation de transport où le transport est donné par la densité elle-même. On peut donc essayer de trouver des solutions en utilisant la méthode des caractéristiques. Toutefois, comme on va le voir tout de suite, de telles solutions peuvent ne pas être définies globalement en temps : on a un phénomène dit d'explosion en temps fini.

Théorème 4.17. *On suppose que $y^0 \in C^1(\mathbb{R})$ et que $y^0, (y^0)' \in L^\infty(\mathbb{R})$.*

- *Si $(y^0)' \geqslant 0$ (autrement dit si y^0 est croissante), alors il existe une unique solution globale à (4.22) dans $C^1([0,+\infty[\times\mathbb{R})$.*
- *Sinon, il existe une unique solution $C^1([0,T^*[\times\mathbb{R})$, avec*

$$T^* = -\frac{1}{\inf_{x\in\mathbb{R}}(y^0)'(x)} \in]0,+\infty[, \tag{4.23}$$

et la solution ne peut pas être prolongée en un temps $t > T^*$.

Remarque 4.18. *Dans le cadre fonctionnel du Théorème 4.17, les solutions sont appelées* solutions classiques *ou* solutions fortes*.*

Preuve : Comme on cherche une solution C^1, on peut développer le carré et se ramener à étudier

$$\begin{cases} \partial_t y(t,x) + y(t,x)\partial_x y(t,x) = 0,\ (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ y(0,x) = y^0(x). \end{cases} \tag{4.24}$$

On a donc une sorte d'équation de transport, mais dont la vitesse dépend de la solution elle-même. On fait un raisonnement par analyse-synthèse, et on commence par supposer donc l'existence d'une solution $y \in C^1([0,T[\times\mathbb{R})$ à (4.24), sur un certain intervalle de temps $[0,T[$ à déterminer.

On va procéder comme dans le cas linéaire et appliquer la méthode des caractéristiques. En prenant exemple sur ce qui a été fait pour l'équation (4.7), on introduit le flot $X(t,0,x) = X_t(x)$ associé à l'équation différentielle $x'(t) = y(t,x(t))$ (on n'aura pas besoin de changer t_0 ici, donc on peut se permettre de changer les notations), qui part donc de x en $t = 0$. Appelons I l'intervalle maximal sur lequel y est bien définie et de classe C^1 (en se restreignant aux $t \geqslant 0$). Alors le théorème de Cauchy-Lipschitz global s'applique et on a existence et unicité d'une solution sur cet intervalle, et on a bien un flot global bien défini et de classe C^1.Remarquons maintenant que y est constante le long de X_t. On pose $z(t) = y(t, X_t(x))$. Par dérivation de fonctions composées, On a bien

$$\begin{aligned}\frac{d}{ds}z(t) &= \partial_t y(t,X_t(x)) + \partial_t X_t(x)\partial_x y(t,X(t)) \\ &= \partial_t y(t,X_t(x)) + y(t,X_t(x))\partial_x y(x,X_t(x)) = 0.\end{aligned}$$

Puisque y est constante le long de X_t, on a donc la simplification importante suivante

$$\partial_t X_t(x) = y(t,X_t(x)) = y^0(x).$$

Donc

$$X_t(x) = x + ty^0(x). \tag{4.25}$$

Donc, encore une fois, comme X_t est constante le long des caractéristiques, on a

$$y^0(x) = y(t,X_t(x)) = y(t,x+ty_0(x)).$$

Ceci fournit donc une solution de manière implicite : si on était capable de résoudre en x l'équation $x+ty_0(x) = x'$ pour $x' \in \mathbb{R}$ et tout $t \in I$, on aurait alors, en appelant $\xi(t,x')$ une telle solution,

$$y(t,x') = y^0(\xi(t,x')).$$

Commençons par remarquer que par dérivabilité du flot et par (4.25), on a

$$\partial_x X_t(x) = 1 + ty_0'(x). \tag{4.26}$$

Notamment, $\partial_x X_0(x) = 1 > 0$. Posons alors

$$m = \inf_{\mathbb{R}} y_0'(x). \tag{4.27}$$

On sait que $m \in \mathbb{R}$, car y_0' est continue bornée.

Si $m \geqslant 0$, alors pour tout $x \in \mathbb{R}$, $\partial_x X_t(x)) > 0$, et donc pour tout $t \geqslant 0$, $x \mapsto X_t(x)$ est une bijection strictement croissante de $\mathbb{R}$ vers $X_t(\mathbb{R})$. l'expression

explicite de X_t donnée en (4.25) assure la surjectivité de X_t, en effet, y^0 est bornée, donc $X_t \to \pm\infty$ quand $x \to \pm\infty$. Dans ce cas, on pose $T^* = +\infty$.

Si $m < 0$, on pose

$$T^* = -\frac{1}{m} > 0.$$

La même analyse que précédemment permet de conclure que $t \in [0, \tau[$, $x \mapsto X_t(x))$ est une bijection strictement croissante de $\mathbb{R}$ vers $\mathbb{R}$.

Ainsi, si y est solution de (4.24) et tant que X_t est une bijection, on a que

$$y(t,x) = y^0(X_t^{-1}(x)). \tag{4.28}$$

Reste donc à vérifier que cette expression fournit bien une solution de classe C^1 sur $[0, T^*[$. On utilise la remarque suivante. $X_t^{-1}(x)$ est l'unique $z \in \mathbb{R}$ tel que $X_t(z) = x$, *i.e.* $X(t, 0, z) = x$. En revenant à (4.11), on en déduit que nécessairement $z = X(0, t, x)$. Donc (4.28) se transforme en

$$y(t,x) = y^0(X(0,t,x)).$$

En reprenant alors exactement les mêmes calculs que dans la partie "synthèse" de la preuve du Théorème 4.8, on en déduit bien que y est solution de (4.24).

Dans le cas où $T^* = +\infty$ (*i.e.* le premier cas du théorème), on a une solution globale en temps comme annoncée. Inspectons maintenant plus en détail le cas $T^* < +\infty$ et montrons que l'on ne peut pas étendre la solution en temps $t > T^*$. Le phénomène qui explique ceci est qu'en temps T^*, il y a des caractéristiques qui vont obligatoirement se croiser, ce qui va entrer en contradiction avec le fait que y soit constante le long des caractéristiques, puisqu'on va montrer qu'on peut faire en sorte que la valeur de y soit différente sur les deux caractéristiques. Prenons n'importe quel $\overline{x} \in \mathbb{R}$ tel que $y_0'(x) < 0$, et posons

$$\bar{t} = -\frac{1}{y_0'(x)} (\geqslant T^*).$$

Prenons un temps $\tilde{t} > \bar{t}$ et montrons qu'il n'existe pas de solutions y à (4.24) définie sur $[0, \tilde{t}]$. On a par définition de $\bar{t}$ que $\tilde{t} y_0'(\overline{x}) < -1$. y_0' étant continue, cette propriété reste donc vraie localement autour de $\overline{x}$: il existe $\varepsilon > 0$ tel que pour tout $x \in [\overline{x} - \varepsilon, \overline{x} + \varepsilon]$, on ait $\tilde{t} y_0'(x) < -1$. Notamment, $y^0(\overline{x} - \varepsilon) \neq y^0(\overline{x})$. Comme la caractéristique $X_t(x)$ est de classe C^1, par le théorème des accroissements finis, on a existence de $y \in [\overline{x} - \varepsilon, \overline{x}]$ tel que (on utilise ici (4.26))

$$X_{\tilde{t}}(\overline{x} - \varepsilon) = X_{\tilde{t}}(\overline{x}) - \varepsilon \partial_x X_{\tilde{t}}(y) = X_{\tilde{t}}(\overline{x}) - \varepsilon(1 + \tilde{t} y_0'(x)) > X_{\tilde{t}}(\overline{x}).$$

Or en $t = 0$, on a

$$X_0(\overline{x} - \varepsilon) = \overline{x} - \varepsilon < x = X_0(\overline{x}).$$

Par continuité et le théorème des valeurs intermédiaires, on a donc existence de $t_0 \in [0, \tilde{t}]$ tel que $X_{t_0}(\overline{x}-\varepsilon) = X_{t_0}(\overline{x})$. Or comme on a déjà remarqué, on a $y^0(\overline{x}-\varepsilon) \neq y^0(\overline{x})$ et donc la contradiction voulue : par constance le long des caractéristiques,

$$y^0(\overline{x}-\varepsilon) = y(t_0, X_{t_0}(\overline{x}-\varepsilon)) = y(t_0, X(t_0, \overline{x})) = y^0(\overline{x}) \neq y^0(\overline{x}-\varepsilon).$$

Ainsi, pour tout $\overline{x} \in \mathbb{R}$ tel que $y_0'(x) < 0$, on ne peut trouver de solution définie sur $[0, \tilde{t}]$, avec

$$\tilde{t} > \bar{t} = -\frac{1}{y_0'(x)} (\geqslant T^*).$$

Par définition de m donnée en (4.27) et en prenant une suite minimisante, on en déduit donc que pour tout $\varepsilon > 0$ et tout $\tilde{t} > T^* - \varepsilon$, on ne peut trouver de solution définie sur $[0, \tilde{t}]$. Autrement dit, en faisant $\varepsilon \to 0$, pour tout $\tilde{t} > T^*$, il n'existe pas de solutions définies sur $[0, \tilde{t}]$. Donc une solution est au mieux définie sur $[0, T^*]$, ce qui est ce qu'on voulait démontrer. ♢

Le fait que les solutions classiques sont dans de nombreux cas pas définies globalement justifient ici l'introduction de la notion de solution faible, dont la définition est calquée sur l'équation de transport (autrement dit, on fait des IPP formelles pour "passer" les dérivées sur une fonction test, puis on prend cette nouvelle formulation comme définition de solutions).

Définition 4.19. *Soit $y^0 \in L^\infty(\mathbb{R})$. On appelle solution faible de l'équation de Burgers* (4.22) *toute fonction $y \in L^\infty \mathbb{R}^+ \times \mathbb{R})$ telle que pour tout $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$, on ait*

$$\int_0^{+\infty} \int_{\mathbb{R}} \left(y(t,x)\partial_t \varphi(t,x) + \frac{y^2(t,x)}{2} \partial_x \varphi(t,x) \right) dxdt = -\int_{-\infty}^{+\infty} y^0(x)\varphi(0,x)dx. \tag{4.29}$$

Bien sûr, il est important de vérifier que

Proposition 4.20. *Toute solution classique de*(4.22) *est une solution faible, et toute solution faible de* (4.22) *de classe C^1 est une solution classique.*

Nous n'allons pas faire la preuve, qui repose sur le même principe que pour les équations de transport (on multiplie la première ligne de (4.22) par φ puis on fait des IPP).

Les deux questions principales sont celles de l'existence et de l'unicité de solutions faibles. Ces deux questions sont relativement difficiles. Dans ce chapitre, nous nous intéresserons uniquement à la question de l'unicité des solutions fortes. La question de l'existence sera traitée en DM.

Commençons par démontrer qu'en général, même si des solutions existent, elle peuvent ne pas être uniques. Donnons-en un exemple très simple.

Exemple 4.21. *On considère comme condition initiale $y^0(x) = 0$, Clairement, 0 est une solution faible L^∞ de (4.22), car dans l'identité (4.24), les deux termes sont égaux à 0. On peut aussi exhiber un nombre infini non dénombrable de solutions faibles. On considère $\alpha > 0$ et une fonction définie par morceaux par*

$$y(t,x) = 2\alpha signe(x)\mathbb{1}_{[-\alpha t,\alpha t]}(x).$$

y est clairement dans L^∞. De plus, elle vérifie (4.29), comme le montre un calcul direct : soit $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$, on a

$$\begin{aligned}
&\int_0^{+\infty}\int_{\mathbb{R}}\left(y(t,x)\partial_t\varphi(t,x) + \frac{y^2(t,x)}{2}\partial_x\varphi(t,x)\right)dxdt \\
&\quad = \int_0^{+\infty}\int_{-\alpha t}^{0}\left(-2\alpha\partial_t\varphi(t,x) + 2\alpha^2\partial_x\varphi(t,x)\right)dxdt \\
&\qquad + \int_0^{+\infty}\int_0^{\alpha t}\left(2\alpha\partial_t\varphi(t,x) + 2\alpha^2\partial_x\varphi(t,x)\right)dxdt.
\end{aligned}$$

Regardons la deuxième intégrale. Pour la partie "dérivée en espace", on a

$$\int_0^{+\infty}\int_0^{\alpha t}\partial_x\varphi(t,x)dxdt = \int_0^{+\infty}\left(\varphi(t,\alpha t) - \varphi(0,t)\right)dt.$$

Pour la partie "dérivée en temps", on commence par remarquer que par dérivation de fonctions composées (on peut intervertir sans problèmes la dérivation et l'intégrale car toutes les fonctions sont de classe C^1, et on travaille en fait sur des segments),

$$\frac{d}{dt}\left(\int_0^{\alpha t}\varphi(t,x)dx\right) = \int_0^{\alpha t}\partial_t\varphi(t,x)dx + \alpha\varphi(t,\alpha t).$$

On en tire donc une expression de $\int_0^{\alpha t}\partial_t\varphi(t,x)dx$, ce qui nous donne donc (puisque φ est à support compact)

$$\begin{aligned}
\int_0^{+\infty}\int_0^{\alpha t}\partial_t\varphi(t,x)dxdt &= \int_0^{+\infty}\frac{d}{dt}\int_0^{\alpha t}\varphi(t,x)dxdt - \int_0^{+\infty}\alpha\varphi(t,\alpha t)dt \\
&= [\int_0^{\alpha t}\varphi(t,x)dx]_{t=0}^{t=+\infty} - \alpha\int_0^{+\infty}\varphi(t,\alpha t)dt. \\
&= -\alpha\int_0^{+\infty}\varphi(t,\alpha t)dt.
\end{aligned}$$

Donc on en déduit que

$$\begin{aligned}
&\int_0^{+\infty}\int_0^{\alpha t}\left(2\alpha\partial_t\varphi(t,x) + 2\alpha^2\partial_x\varphi(t,x)\right)dxdt \\
&= -2\alpha^2\int_0^{+\infty}\varphi(t,\alpha t)dt - 2\alpha^2\int_0^{+\infty}\left(\varphi(t,\alpha t) + \varphi(0,t)\right)dt. \\
&= -2\alpha^2\int_0^{+\infty}\varphi(0,t)dt.
\end{aligned}$$

On se convainc aisément que les calculs sont très similaires pour la première intégrale : on a

$$\int_0^{+\infty}\int_{-\alpha t}^{0}\partial_x\varphi(t,x)dxdt = \int_0^{+\infty}\left(\varphi(0,t)-\varphi(t,-\alpha t)\right)dt,$$

et un raisonnement similaire donne

$$\begin{aligned}\int_0^{+\infty}\int_{-\alpha t}^{0}\partial_t\varphi(t,x)dxdt &= \int_0^{+\infty}\frac{d}{dt}\int_{-\alpha t}^{0}\varphi(t,x)dxdt + \int_0^{+\infty}\alpha\varphi(t,-\alpha t)dt\\ &= \alpha\int_0^{+\infty}\varphi(t,-\alpha t)dt.\end{aligned}$$

Donc

$$\int_0^{+\infty}\int_{-\alpha t}^{0}\left(2\alpha\partial_t\varphi(t,x)+2\alpha^2\partial_x\varphi(t,x)\right)dxdt = 2\alpha^2\int_0^{+\infty}\varphi(0,t)dt.$$

En sommant ces deux identités, on obtient bien 0*, comme voulu puisqu'ici* $y^0(x) = 0$.

Pour restaurer une forme d'unicité, il va donc falloir se donner des critères supplémentaires qui permettent de "sélectionner" une bonne solution. Différents critères sont possibles, et proviennent en général de considérations physiques. Autrement dit, il faut comprendre quelles sont les solutions qui ont un sens physique et celles qui n'en ont pas. Nous donnerons un cas particulier de critère d'unicité dans la suite.

4.2.2 Problème de Riemann, relations de Rankine-Hugoniot

Une bonne manière de comprendre un peu mieux le comportement des solutions faibles est de regarder le cas de ce qu'on appelle le problème de Riemann, qui consiste à regarder un cas très particulier de données initiales constantes par morceaux et discontinues en espace, à savoir des données initiales de la forme

$$y^0(x) = y_g,\ x\leqslant 0, \quad y^0(x) = y_d,\ x>0, \tag{4.30}$$

avec $y_g \neq y_d \in \mathbb{R}$, puis nous allons chercher des solutions particulières sous la forme

$$y(t,x) = y^0(x-\sigma t), \tag{4.31}$$

avec σ à déterminer, sur le modèle du transport linéaire, en espérant que de telles solutions existent. Remarquons que dans ce cas, par les relations (4.25) les courbes caractéristiques (le long desquelles la solution est constante) sont données, à x fixé, par des droites de la forme $(t, x+y_g t)$ si $x<0$ et de la forme $(t, x+y_d t)$ si $x>0$.

Proposition 4.22. *la fonction y donnée en* (4.31) *est solution faible de* (4.22) *avec donnée initiale* (4.30) *si et seulement si*

$$\sigma = \frac{y_g + y_d}{2}. \tag{4.32}$$

Remarque 4.23. (4.32) *est appelée relation de Rankine-Hugoniot.*

Preuve : On cherche une solution faible de l'équation sous la forme (4.31). Notamment, on voit que si $x > \sigma t$, alors $y(t,x) = y_d$ et si $x < \sigma t$, on a $y(t,x) = y_g$.

Soit $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$. On veut avoir (4.29). Nous allons considérer des fonctions à support compact bien particulières. Ici, les boules seront des boules ouvertes euclidiennes dans $\mathbb{R}^2$. Soit $\varepsilon > 0$. On pose

$$B_\varepsilon = B((t,\sigma t),\varepsilon),$$

$$D_{\varepsilon,g} = B((t,\sigma t),\varepsilon) \cap \{\sigma t > x\},$$

$$D_{\varepsilon,d} = B((t,\sigma t),\varepsilon) \cap \{\sigma t < x\}.$$

(on "coupe" B_ε en deux selon la droite d'équation $\sigma t = x$). On prend maintenant ε suffisamment petit pour que B_ε ne touche pas la droite $t = 0$, et on considère $\varphi \in C_0^\infty(B_\varepsilon)$ (étendue à zéro sur $\mathbb{R}^+ \times \mathbb{R}$). Notamment, $\varphi(0,x) = 0$ et donc (4.29) devient (puisqu'un morceau de droite est de mesure de Lebesgue nulle)

$$\int_{D_{\varepsilon,g}} \left(y(t,x)\partial_t\varphi(t,x) + \frac{y^2(t,x)}{2}\partial_x\varphi(t,x) \right) dxdt$$
$$+ \int_{D_{\varepsilon,d}} \left(y(t,x)\partial_t\varphi(t,x) + \frac{y^2(t,x))}{2}\partial_x\varphi(t,x) \right) dxdt = 0.$$

Au vue de la forme particulière de la solution étudiée ici, on doit donc avoir que

$$\int_{D_{\varepsilon,g}} \left(y_g\partial_t\varphi(t,x) + \frac{y_g^2}{2}\partial_x\varphi(t,x) \right) dxdt + \int_{D_{\varepsilon,d}} \left(y_d\partial_t\varphi(t,x) + \frac{y_d^2}{2}\partial_x\varphi(t,x) \right) dxdt = 0.$$

On "rappelle" le résultat suivant, appelé *formule de Green*, qui est une sorte de formule d'intégration par parties en dimension supérieure.

Lemme 4.24. *Soit Ω un ouvert borné suffisamment régulier (au sens où par exemple, on peut localement en tout point "redresser" la frontière en une portion d'hyperplan qui sépare l'extérieur et i'intérieur de l'ouvert, le redressement étant de classe C^1), et soit $u, v \in C^1(\overline{\Omega})$. Soit $n = (n_1, \dots n_d)$ la normale extérieure sur $\partial\Omega$. Alors on a, pour $i \in [|1,n|]$,*

$$\int_\Omega u\partial_i v dx = -\int_\Omega v\partial_i u dx + \int_{\partial\Omega} uvn_i dS.$$

Remarque 4.25. *L'intégrale de surface dS est assez compliquée à définir. On peut par exemple imaginer que Ω est paramétrisée par une certaine fonction $\gamma : \mathcal{O} \to \partial\Omega$ ($\mathcal{O}$ ouvert de $\mathbb{R}^{d-1}$) régulière peut être définie en se donnant un paramétrage de la frontière et en appliquant une formule type changement de variables : pour f régulière définie sur $\partial\Omega$,*

$$\int_{\partial\Omega} f dS := \int_{\mathcal{O}} f(\gamma(s)) J_\gamma(s) ds,$$

où J_γ est le Jacobien. Si on a seulement des paramétrisations locales, on fait tout ceci localement et on recolle les intégrales (il y en a un nombre fini comme on peut le démontrer par un argument de compacité).

Revenons à notre preuve. Ici, en appliquant la formule de Green séparément sur les deux parties $D_{\varepsilon,g}$ et $D_{\varepsilon,d}$, et en remarquant que la dérivée de y_g et y_d est nulle, on obtient (en notant n_t et n_x les composantes de la normale par rapport à x)

$$\int_{\partial D_{\varepsilon,g}} \left(y_g n_t + \frac{y_g^2}{2} n_x \right) \varphi(t,x) dS + \int_{\partial D_{\varepsilon,d}} \left(y_d n_t + \frac{y_d^2}{2} n_x \right) \varphi(t,x) dS = 0.$$

Comme $\varphi \in C_0^\infty(B_\varepsilon)$, le seul endroit où cette intégrale est non nulle est sur le segment $\{x = at\}$. De plus, sur ce segment, on a que n_g est positivement colinéaire à $\begin{pmatrix} -\sigma \\ 1 \end{pmatrix}$ et n_d est positivement colinéaire à $\begin{pmatrix} \sigma \\ -1 \end{pmatrix}$. On en déduit donc que

$$\left(-\sigma y_g + \frac{y_g^2}{2} + \sigma y_d - \frac{y_d^2}{2} \right) \int_{\partial D_{\varepsilon,g} \cap \{x=\sigma t\}} \varphi(t,x) dS = 0,$$

et ceci pour tout $\varphi \in C_0^\infty(B_\varepsilon)$. Comme on peut construire φ de telle sorte que $\int_{\partial D_{\varepsilon,g} \cap \{x=\sigma t\}} \varphi(t,x) dS \neq 0$, on a donc

$$\left(-\sigma y_g + \frac{y_g^2}{2} + \sigma y_d - \frac{y_d^2}{2} \right) = 0,$$

ce qui donne (4.32) en factorisant par $y_g - y_d$. ♢

On peut largement généraliser le raisonnement précédent à une classe bien particulière de solutions, que nous définissons ici.

Définition 4.26. *Une fonction u définie sur $\mathbb{R}^+ \times \mathbb{R}$ est de classe C^1 par morceaux si pour tout ouvert borné $\mathcal{O}$ de $\mathbb{R}^+ \times \mathbb{R}$, il existe un nombre fini de courbes $\Sigma_1, \dots \Sigma_p$ qui peuvent se paramétriser localement sous la forme $(t, \xi_i(t))$ ($i \in \{1, \dots p\}$) avec ξ de classe C^1, et telles que u est de classe C^1 dans chaque composante connexe de $\mathcal{O} \setminus \bigcup_{i=1}^d \Sigma_i$ et admette une limite à droite et à gauche de chacun des Σ_i, ainsi que ses dérivées partielles.*

On admettra alors le théorème suivant, dont la preuve est très similaire au cas du problème de Riemann. On appellera aussi le résultat de ce théorème *relations de Rankine-Hugoniot*.

Théorème 4.27. *On suppose que y_0 est de classe C^1 par morceaux sur $\mathbb{R}$ (au sens usuel) et dans $L^\infty(\mathbb{R})$. Une fonction y de classe C^1 par morceaux est une solution faible de* (4.22) *si et seulement si elle vérifie la première équation de* (4.22) *sur tout domaine où y est de classe C^1, et elle vérifie le long des courbes où elle n'est pas régulière la relation*

$$\xi'(t) = \frac{y^+(t) + y^-(t)}{2}, \tag{4.33}$$

où $y^+(t)$ et $y^-(t)$ sont respectivement les limites à droite et à gauche au point $(t, \xi(t))$.

Pour comprendre comment utiliser ce théorème, on pourra regarder les exercices 27,34,35.

4.2.3 Critère d'Oleinik

Une manière de restaurer l'unicité est de rajouter la condition suivante.

Définition 4.28. *Soit $u \in L^\infty([0, +\infty[\times\mathbb{R})$. On dit que u satisfait à une inégalité d'Oleinik si et seulement s'il existe $C :]0, +\infty[\to \mathbb{R}^+$ décroissante telle que*

$$u(t, x') - u(t, x) \leqslant C(t)(y - x), \ \forall x' \geqslant x, \ \forall t > 0. \tag{4.34}$$

Une solution faible y de (4.22) *est dite admissible si elle vérifie* (4.34) *(où u est remplacé par y).*

On va voir que ce critère suffit à assurer l'unicité des solutions faibles à (4.22) (on verra des exemples d'utilisation de ce critère en exercice). Commençons par le meme suivant, sur les EDO, qui nous sera utile par la suite.

Lemme 4.29. *Sous les hypothèses du théorème de Cauchy-Lipschitz global et en appelant $X(t, t_0, x_0)$ le flot associé à* (4.5)*, si de plus il existe $C > 0$ tel que pour tout*

$$a(t, x') - a(t, x) \leqslant C(x' - x), \ \forall y \geqslant x, \ \forall t > 0,$$

alors pour tout (t, t_0, x, y), le flot vérifie

$$|X(t, t_0, x') - X(t, t_0, x)| \leqslant e^{C(t-t_0)}|y - x|, \ \forall t \geqslant t_0. \tag{4.35}$$

et pour tout (t, t_0, x),

$$|\partial_3 X(t, t_0, x)| \leqslant e^{C(t-t_0)}, \ \forall t \geqslant t_0. \tag{4.36}$$

Remarque 4.30. *Il s'agit d'une estimée de* stabilité *: si y n'est "pas trop loin" de x, alors, tant que t ne grossit pas trop, $X(t, t_0, y)$ n'est "pas trop loin" de $X(t, t_0, x)$.*

Preuve : On pose, à x, y, x_0 fixé,

$$D(t) = (X(t, t_0, y) - X(t, t_0, x))^2.$$

Notamment,

$$D(0) = (x - y)^2.$$

$$\begin{aligned} D'(t) &= 2D(t)D'(t) \\ &= 2(X(t, t_0, y) - X(t, t_0, x))(\partial_1 X(t, t_0, y) - \partial_1 X(t, t_0, x)) \\ &= 2(X(t, t_0, y) - X(t, t_0, x))(a(t, X(t, t_0, y)) - a(t, X(t, t_0, x))) \\ &\leqslant 2C(X(t, t_0, y) - X(t, t_0, x))^2 \\ &\leqslant 2CD(t). \end{aligned}$$

On a donc

$$D'(t) - 2CD(t) \leqslant 0.$$

On multiplie par e^{-2Ct}et on remarque qu'on a identifié la dérivée de $D(t)e^{-2Ct}$. On obtient

$$\frac{d}{dt}(D(t)e^{-2Ct}) = (D'(t) - 2CD(t))e^{-2Ct_0} \leqslant 0.$$

Donc pour $t \geqslant t_0$, on obtient

$$D(t)e^{-2Ct} \leqslant D(0)e^{-2Ct_0}|y - x|^2 = |x - y|^2 e^{-2ct_0},$$

Ce qui donne (4.35) en passant à la racine. (4.36) s'en déduit en prenant $y \neq x$, en divisant (4.35) par $|x - y|$, puis en faisant $y \to x$. ♢

De manière assez étonnante, dans le cas de l'équation de Burgers, un argument simple nous permettra de nous ramener à une équation de transport linéaire conservative. Commençons donc par énoncer la formulation faible associée à (4.6). En suivant point par point la démonstration effectuée dans le cas des vitesses constantes, on obtient la proposition suivante.

Proposition 4.31. *Si $y^0 \in C^1(\mathbb{R})$, $y \in C^1(\mathbb{R}^+ \times \mathbb{R})$ et si $a \in C^1(\mathbb{R}^+ \times \mathbb{R})$, alors y est solution de* (4.6) *si et seulement si*

$$\int_0^{+\infty} \int_{\mathbb{R}} y\,(\partial_t \varphi + a\partial_x \varphi) = -\int_{\mathbb{R}} y^0(x)\varphi(0, x)dx, \forall \varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}). \tag{4.37}$$

Ceci suggère donc d'introduire la définition suivante.

Définition 4.32. *Soit $y^0 \in L^\infty(\mathbb{R})$ et $a \in L^\infty(\mathbb{R}^+ \times \mathbb{R})$. Une fonction $y \in L^\infty(\mathbb{R}^+ \times \mathbb{R})$ est une solution faible de* (4.6) *si* (4.37) *est vérifiée.*

Le point crucial est alors le suivant.

Proposition 4.33. *Si $a \in L^\infty(\mathbb{R}^+, \mathbb{R})$ vérifie l'inégalité* (4.34) *(où y est remplacé par a), alors il existe au plus une solution faible dans $L^\infty(\mathbb{R}^+ \times \mathbb{R})$ à* (4.6).

Remarque 4.34. *La question de l'existence de solutions faibles à* (4.6) *dans le cas où a ne vérifie par les conditions de Cauchy-Lipschitz est très difficile et nous n'en parlerons pas ici.*

Preuve : Cela ressemble un peu à la partie unicité de la preuve du Théorème 4.15, mais il va falloir être beaucoup plus fin. Considérons deux solutions faibles y et z de (4.6), associées à la même condition initiale y^0. Soit $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$. Alors

$$\int_0^{+\infty} \int_{\mathbb{R}} y(t,x)(\partial_t \varphi(t,x) + a(t,x)\partial_x \varphi(t,x))dxdt = -\int_{\mathbb{R}} y^0(x)\varphi(0,x)dx$$

et

$$\int_0^{+\infty} \int_{\mathbb{R}} z(t,x)(\partial_t \varphi(t,x) + a(t,x)\partial_x \varphi(t,x))dxdt = -\int_{\mathbb{R}} y^0(x)\varphi(0,x)dx.$$

En retranchant ces deux expressions, on trouve

$$\int_0^{+\infty} \int_{\mathbb{R}} u(t,x)(\partial_t \varphi(t,x) + a(t,x)\partial_x \varphi(t,x))dxdt = 0, \tag{4.38}$$

où on a posé $u = y - z$. Si l'on souhaitait conclure comme dans la preuve du Théorème 4.15, il suffirait de montrer que pour tout $\psi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$, on peut trouver $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$ tel que

$$\partial_t \varphi + a\partial_x \varphi = \psi,$$

Mais ceci n'a aucune chance d'être vrai, puisque a n'est pas suffisamment régulière (donc le φ est forcément pas de classe C^∞ là où a ne l'est pas). De plus, si l'on regarde bien le détail de la preuve qui suit, on va avoir un petit problème au voisinage de $t = 0$, il va donc falloir s'en éloigner un peu. On va donc prendre $\varepsilon > 0$ aussi petit que l'on veut, et on va plutôt partir de la propriété suivante : pour tout $\psi \in C_0^\infty(]\varepsilon, +\infty[\times\mathbb{R})$, on a

$$\int_\varepsilon^{+\infty} \int_{\mathbb{R}} u(t,x)(\partial_t \varphi(t,x) + a(t,x)\partial_x \varphi(t,x))dxdt = 0, \tag{4.39}$$

Si on arrive à en déduire que $u = 0$ p.p. sur $[\varepsilon, +\infty[\times\mathbb{R}$, ceci étant vrai pour tout $\varepsilon > 0$, cela nous permettrait de conclure.

On va donc devoir régulariser la vitesse a. Pour ce faire, on considère $\rho \in C_0^\infty(]-1, 0[)$ telle que $\rho \geqslant 0$ et $\int_{\mathbb{R}^2} \rho = 1$, puis on pose

$$\rho_n(t,x) = n^2 \rho(nt, nx). \tag{4.40}$$

Alors,

— $\forall n \in \mathbb{N}^*$, on a $\rho_n \geqslant 0$ et $\int_{\mathbb{R}^2} \rho_n = 1$ par le changement de variable $(nt, nx) = (s, z)$ de jacobien $1/n^2$.

— $\forall n \in \mathbb{N}^*$, ρ_n est à support compact dans $]-1/n, 0[^2$.

—
$$\forall n \in \mathbb{N}^*, ||\rho_n||_\infty \leqslant n^2 ||\rho||_\infty. \tag{4.41}$$

On étend alors a à 0 sur en dehors de $[\varepsilon, +\infty[$, puis on pose

$$a_n = a * \rho_n.$$

Alors :

— Pour tout $n \in \mathbb{N}^*$, on a que $a_n \in C^\infty(\mathbb{R}^2)$ (c'est un résultat classique sur la convolution).

— Pour tout $n \in \mathbb{N}^*$ et tout $(t,x) \in \mathbb{R}^2$, on a

$$|a_n(t,x)| \leqslant \int_{\mathbb{R}^2} |a(t-s, x-z)| \rho_n(s,z) ds dz \leqslant ||a||_\infty \int_{\mathbb{R}^2} \rho_n(s,z) ds dz = ||a||_\infty. \tag{4.42}$$

— Pour tout $n \in \mathbb{N}^*$, tout $t \geqslant \varepsilon$, et tout $x, y \in \mathbb{R}$ avec $y \geqslant x$, et en utilisant (4.34) (où u est remplacé par a) et la positivité des ρ_n,

$$\begin{aligned} a_n(t,y) - a_n(t,x) &= \int_{\mathbb{R}^2} (a(t-s, y-z) - a(t-s, x-z)) \rho_n(s,z) ds dz \\ &\leqslant \int_{\mathbb{R}^2} C(t-s)(y-z-(x-z)) \rho_n(s,z) ds dz \\ &\int_{\mathbb{R}^2} C(t-s)(y-x) \rho_n(s,z) ds dz. \end{aligned}$$

Maintenant, il suffit de remarquer que l'on intègre en fait sur $]-1/n, 0[^2$, et pour $s \in]-1/n, 0[$, par croissance de C, on a $C(t-s) \leqslant C(t)$ pour $t \geqslant \varepsilon$. On a donc que a_n vérifie aussi la condition d'Oleinik (en utilisant que $\int_{\mathbb{R}^2} \rho_n = 1$)

$$a_n(t,y) - a_n(t,x) \leqslant C(t)(y-x) \int_{\mathbb{R}^2} \rho_n = C(t)(y-x), \ t \geqslant \varepsilon. \tag{4.43}$$

— $a_n \to a$ presque partout. Ce point est un peu délicat. On utilise encore une fois que $\int_{\mathbb{R}^2} \rho_n = 1$ ainsi que (4.41). On écrit de manière classique (en appelant $(t,x) = X$ pour compactifier un peu les notations)

$$\begin{aligned} |a_n(X) - a(X)| &= \left| \int_{\mathbb{R}^2} a(X-Y) \rho_n(Y) dY - a(X) \int_{\mathbb{R}^2} \rho_n(Y) dY \right| \\ &\leqslant \int_{[-1/n,0]^2} |a(X-Y) - a(X)| \rho_n(Y) dY \\ &\leqslant n^2 ||\rho||_\infty \int_{[-1/n,0]^2} |a(X-Y) - a(X)| dY. \end{aligned}$$

Le théorème de Lebesgue assure que cette quantité converge bien vers 0 p.p. par (4.16) où t est remplacé par $1/n$.

On pose alors

$$\varphi_n(t,x) = -\int_t^\infty \psi(s, X_n(s,t,x))ds,$$

où X_n est le flot associé à la vitesse a_n. Remarquons tout de suite que grâce à (4.42), on a que pour tout $(t,t_0,x) \in [\varepsilon,+\infty[\times[\varepsilon,+\infty[\times\mathbb{R}$,

$$|\partial_1 X_n(t,t_0,x)| \leqslant ||a||_\infty. \tag{4.44}$$

En reprenant exactement les calculs de la preuve du Théorème 4.15, on montre que $\varphi_n \in C_0^\infty([\varepsilon,+\infty[\times\mathbb{R})$ et qu'elle vérifie

$$\partial_t\varphi_n(t,x) + a_n(t,x)\partial_x\varphi(t,x) = \psi(t,x).$$

Donc notamment,

$$\partial_t\varphi_n(t,x) + a(t,x)\partial_x\varphi(t,x) = \psi(t,x) + (a(t,x) - a_n(t,x))\varphi_n(t,x). \tag{4.45}$$

On repart alors de (4.39) et on utilise comme fonction test les φ_n crées précédemment. On en déduit

$$0 = \int_\varepsilon^{+\infty}\int_{\mathbb{R}} u(t,x)(\partial_t\varphi(t,x) + a(t,x)\partial_x\varphi(t,x))dxdt,$$

et donc par l'identité (4.45) que

$$0 = \int_\varepsilon^{+\infty}\int_{\mathbb{R}} u(t,x)\psi(t,x)dxdt = \int_0^{+\infty}\int_{\mathbb{R}} w(t,x)(a(t,x) - a_n(t,x))\varphi_n(t,x)dxdt.$$

Pour conclure à l'aide du lemme fondamental du calcul des variations, il suffit donc de démontrer que

$$\int_\varepsilon^{+\infty}\int_{\mathbb{R}} u(t,x)(a(t,x) - a_n(t,x))\varphi_n(t,x)dxdt \to 0 \text{ quand } n \to +\infty.$$

Ceci va être une conséquence du théorème de convergence dominée et du Lemme 4.29.

On considère K un compact de $[\varepsilon,+\infty[\times\mathbb{R}$ qui englobe le support de φ_n pour tout $n \in \mathbb{N}^*$. Un tel compact K existe. Pour le montrer, on peut remarquer que le support en temps de ψ est toujours inclus dans un certain $[-\alpha,\alpha]$ pour un certain $\alpha > 0$. En intégrant l'inégalité (4.42), on en déduit donc que pour tout $(s,t,x) \in [\varepsilon,+\infty[\times\mathbb{R}^+\times\mathbb{R}$, on a

$$X_n(s,t,x) \in [x - \alpha||a||_\infty, x + \alpha||a||_\infty].$$

A partir de là, en revenant à la définition de φ_n, on en déduit aisément que φ_n est à support compact inclus dans un certain K ne dépendant pas de n, le point crucial étant que le compact $[x - \alpha||a||_\infty, x + \alpha||a||_\infty]$ n'en dépende pas. On en déduit que

$$\int_{[\varepsilon,+\infty[} int_{\mathbb{R}} u(a_n - a)\partial_x\varphi_n = \int_K u(a_n - a)\partial_x\varphi_n.$$

On va essayer d'appliquer le théorème de convergence dominée.

Le point crucial ici est que $|\partial_x\varphi_n|$ est bornée sur K, par un certain $B > 0$. En effet, commençons par remarquer que pout tout $(t, x) \in [\varepsilon, +\infty[\times\mathbb{R}$, on a ($\psi$ et toutes ses dérivées sont bornées)

$$\begin{aligned} |\partial_x\varphi_n(t,x)| &= \left| \int_t^{+\infty} \partial_x(\psi(s, X_n(s,t,x)))\partial_3 X_n(s,t,x) \right| \\ &\leqslant ||\partial_x\psi||_\infty \int_t^a |\partial_x(\psi(s, X_n(s,t,x)))| \end{aligned}.$$

Maintenant, on applique le lemme 4.29 avec t_0 remplacé par t, t remplacé par s, et $C = C(t)$. En effet, C est décroissante donc pour tout $s \geqslant t$, on a $C(s) \leqslant C(t) \leqslant C(\varepsilon)$. On en déduit par (4.36) que

$$\partial_x\varphi_n(t,x)| \leqslant ||\partial_x\psi||_\infty \int_t^a e^{C(\varepsilon)(s-t)}dt \leqslant ||\partial_x\psi||_\infty \frac{e^{aC(\varepsilon)-1}}{C(\varepsilon)} =: B. \tag{4.46}$$

On en déduit donc que $u(a_n - a)\partial_x\varphi_n$ converge p.p. vers 0. De plus, grâce à (4.42) et (4.46), on en déduit que

$$|u(a_n - a)|\partial_x\varphi_n| \leqslant B||u||_\infty (||a_n||_\infty + ||a||_\infty) \leqslant 2B||u||_\infty||a||_\infty,$$

et la fonction de droite est un chapeau intégrable indépendant de n, puisque c'est une constante intégrée sur un compact. Le théorème de convergence dominée s'applique, on en déduit donc que

$$\int_0^{+\infty} \int_{\mathbb{R}} u(t,x)(a(t,x) - a_n(t,x))\varphi_n(t,x)dxdt \to 0 \text{ quand } n \to +\infty$$

et la preuve est terminée. ♢

Remarque 4.35. *Il se peut très bien que $C(t) \to +\infty$ quand $t \to 0$ (c.f. Exercice 27). Si l'on regarde bien la preuve précédente, on voit donc qu'on est obligé de s'éloigner de 0, sinon on ne peut pas appliquer le théorème de convergence dominée.*

Théorème 4.36. *Il existe au plus une solution faible admissible à (4.22) dans l'espace $L^\infty(]0, \infty[\times\mathbb{R})$.*

Preuve : Considérons $y, z \in L^\infty(]0, \infty[\times\mathbb{R})$ qui partent de la même condition initiale y_0. Posons $w = u - v$. Soit $\varphi \in C_0^\infty([0, \infty[\times\mathbb{R})$. On va essayer d'identifier quelle formulation faible w vérifie. Il faut faire un peu attention, car maintenant le

problème est non linéaire, donc on ne peut pas superposer les solutions comme dans le cas linéaire. On a

$$\int_0^{+\infty}\int_{\mathbb{R}}\left(y(t,x)\partial_t\varphi(t,x)+\frac{y^2(t,x)}{2}\partial_x\varphi(t,x)\right)dxdt=-\int_{-\infty}^{+\infty}y^0(x)\varphi(0,x)dx$$

et

$$\int_0^{+\infty}\int_{\mathbb{R}}\left(z(t,x)\partial_t\varphi(t,x)+\frac{z^2(t,x)}{2}\partial_x\varphi(t,x)\right)dxdt=-\int_{-\infty}^{+\infty}y^0(x)\varphi(0,x)dx.$$

On fait la différence de ces deux expressions. En utilisant que $a^2-b^2=(a-b)(a+b)$, on en déduit que

$$\int_0^{+\infty}\int_{\mathbb{R}}\left(w(t,x)\partial_t\varphi(t,x)+w(t,x)\frac{y(t,x)+z(t,x)}{2}\partial_x\varphi(t,x)\right)dxdt=0.$$

Autrement dit, w est une solution de (4.37) avec comme vitesse $a(t,x)=\frac{y(t,x)+z(t,x)}{2}\in L^\infty(\mathbb{R}^+\times\mathbb{R})$ et avec condition initiale nulle. Par la Proposition 4.33, $w=0$ p.p. et donc $u=v$ p.p., comme on le souhaitait. ♢

Exemple 4.37. *On revient à l'exemple 4.21. Alors, clairement la fonction identiquement nulle est clairement la solution entropique car elle vérifie bien la condition d'Oleinik* (4.34) *avec $C(t)=0$ (qui est décroissante). Au moins dans ce cas, notre critère de sélection de solutions est raisonnable.*

On a vu des candidats naturels pour résoudre le problème de Riemann. Reste à comprendre lesquels vérifient le critère d'Oleinik (4.34).

Proposition 4.38. *On considère la fonction y donnée en* (4.31)*, solution faible de* (4.22) *avec donnée initiale* (4.30)*, vérifiant donc la relation de Rankine-Hugoniot* (4.32)*. Alors y est admissible si et seulement si $y_g > y_d$. Une telle solution est appelée* onde de choc *qui se propage à vitesse σ ou qui se propage selon la courbe $x=\sigma t$.*

Preuve : On suppose dans un premier temps que $y_g > y_d$. Soient $(t,x,x')\in\mathbb{R}^{+*}\times\mathbb{R}^2$ avec $x'\geqslant x$. On a alors $t-\sigma x\geqslant t-\sigma x'$.

De trois choses l'une.

— Soit $t-\sigma x'\geqslant 0$. Alors automatiquement $t-\sigma x\geqslant 0$ et on a alors

$$y(t,x')-y(t,x)=y^0(t-\sigma x')-y^0(t-\sigma x)=y_d-y_d=0.$$

La condition (4.34) est donc vérifiée pour la fonction décroissante $C(t)=0$.

— Soit $t-\sigma x'<0$ et $t-\sigma x<0$. On a alors

$$y(t,x')-y(t,x)=y^0(t-\sigma x')-y^0(t-\sigma x)=y_g-y_g=0.$$

La condition (4.34) est donc vérifiée pour la fonction décroissante $C(t)=0$.

— Soit $t - \sigma x' < 0$ et $t - \sigma x \geqslant 0$. On a alors

$$y(t, x') - y(t, x) = y^0(t - \sigma x') - y^0(t - \sigma x) = y_d - y_g \leqslant 0.$$

La condition (4.34) est donc vérifiée pour la fonction décroissante $C(t) = 0$.

Globalement, la condition (4.34) est donc vérifiée pour la fonction décroissante $C(t) = 0$.

Maintenant supposons que $y_g < y_d$ et raisonnons par l'absurde en supposant l'existence de $C(t)$ décroissante telle que pour tout $x, x' \in \mathbb{R}$ avec $x' > x$, on ait

$$y(t, x') - y(t, x) \leqslant C(t)(x' - x).$$

Notamment, en $t = 1$,

$$y(1, x') - y(1, x) \leqslant C(1)(x' - x).$$

On choisit alors n'importe quel x' tel que $1 - \sigma x' < 0$, *i.e.* $x' > \frac{1}{\sigma}$, et n'importe quel x tel que $1 - \sigma x > 0$, *i.e.* $x < \frac{1}{\sigma}$. On a alors par hypothèse

$$y(t, x') - y(t, x) = y^0(t - \sigma x') - y^0(t - \sigma x) = y_d - y_g \leqslant C(1)(x' - x).$$

On peut faire $x' \to \frac{1}{\sigma^+}$ et $x \to \frac{1}{\sigma^-}$, ce qui conduit à

$$y_d - y_g \leqslant 0.$$

Ceci est absurde car $y_d - y_g > 0$. ♢

Remarque 4.39. *Dans le cas $y_g < y_d$, on peut aussi exhiber une solution admissible appelée* onde de raréfaction*, c.f. Exercice 27.*

4.3 Exercices

Les exercices 1 à 5 sont à effectuer en autonomie. Un corrigé est proposé à la fin du chapitre. Certains des exercices 6 à 14 seront traités en cours, le reste étant laissé en entraînement.

Exercices corrigés

Exercice 25. *1. Résoudre l'équation*

$$\partial_t y(t, x) + x\partial_x y(t, x) = 0,\ (t, x) \in \mathbb{R}^+ \times \mathbb{R}, y(0, x) = y^0(x),$$

avec $y_0 \in C^1(\mathbb{R})$.

2. On suppose de plus que $y_0 \in L^p(\mathbb{R})$ pour un certain $p \in [1, +\infty]$. A-t-on encore $y(t, \cdot) \in L^p(\mathbb{R})$ pour tout $t > 0$?

3. *Si $y^0 \in L^1(\mathbb{R})$, la masse est-elle conservée au cours du temps ?*

Exercice 26 (Équation de transport conservative)**.** *Le but de cet exercice est de comprendre comment résoudre* (4.6) *à l'aide de la méthode des caractéristiques. Pour ce faire, on supposera que y^0 est de classe $C^1(\mathbb{R})$ et a est de classe $C^2(\mathbb{R}^2)$ et vérifie les hypothèses du théorème de Cauchy-Lipschitz global.*

1. *Montrer que pour tout s, t, x, on a*

$$\partial_2 X(s,t,x) + a(t,x)\partial_3 X(s,t,x) = 0,$$

où les courbes caractéristiques sont les mêmes que pour l'équation (4.7)*.*

2. *On pose*

$$\rho(t,x) = \exp\left(-\int_0^t \partial_x a(s, X(s,t,x))ds\right)$$

En posant

$$\xi(t,x) = -\int_0^t \partial_x a(s, X(s,t,x))ds,$$

montrer que

$$\partial_t \rho + a\partial_x \rho + \rho \partial_x a = 0.$$

3. *En utilisant la méthode des caractéristiques, montrer que si alors* (4.6) *admet une unique solution donnée par*

$$y(t,x) = y^0(X(0,t,x)) \exp\left(-\int_0^t \partial_x a(s, X(s,t,x))ds\right), \tag{4.47}$$

4. *Montrer que pour tout t, z, on a*

$$\partial_3 X(t,0,z) = \exp\left(\int_0^t \partial_x a(s, X(s,0,z)ds\right) dz. \tag{4.48}$$

5. *Montrer que la masse est conservée : si on suppose de plus que $y^0 \in L^1(\mathbb{R})$, alors, pour tout $t > 0$, $y(t,\cdot) \in L^1(\mathbb{R})$ et*

$$\int_{\mathbb{R}} y(t,x)dx = \int_{\mathbb{R}} y^0(x)dx.$$

Indication : on admettra que le flot est de classe C^2.

Exercice 27 (Construction d'une onde de raréfaction pour l'équation de Burgers)**.** *On s'intéresse au problème de Riemann pour* (4.22) *dans le cas où $y_g < y_d$. On a alors vu que la solution constante par morceaux donnée par les relations de Rankine-Huguniot dans le cas du problème de Riemann n'est pas admissible.*

1. *Montrer que $f(t,x) = \frac{x}{t}$ vérifie, pour tout $(t,x) \in \mathbb{R}^{+*} \times \mathbb{R}$,*

$$\partial_t f + \frac{1}{2}\partial_x(f^2) = 0.$$

2. *En utilisant la méthode des caractéristiques, les conditions de Rankine-Hugoniot généralisées et la première question, trouver une solution à y à* (4.22) *qui soit continue sur* $\mathbb{R}^{+*} \times \mathbb{R}$, *dans le cas particulier où* $y_g = 0$ *et* $y_d = 1$.
3. *Montrer qu'elle est admissible.*
4. *Donner la solution admissible dans le cas général* $y_g < y_d$.

Exercice 28 (Paires d'entropie-Flux d'entropie)**.** *On appelle paire d'entropie-flux d'entropie pour l'équation* (4.22) *tout couple* (S, F) *de fonctions* $L^1_{loc}(\mathbb{R})$ *tel que* S *est continue convexe, et* $F'(x) = xS'(x)$ *au sens des distributions.*

1. *Soit* $k \in \mathbb{R}$. *Montrer que le couple* $S(s) = |s-k|$ *et* $F(s) = signe(s-k)\frac{s^2-k^2}{2}$ *est une paire d'entropie-flux d'entropie.*
2. *On suppose que* F *et* S *sont suffisamment régulières (on précisera quelle régularité on peut prendre à la fin du raisonnement). Sout* y *une solution régulière de* (4.22). *Montrer que*

$$\partial_t S(y) + \partial_x F(y) = 0.$$

3. *Considérons maintenant* $y \in L^\infty(\mathbb{R}^+ \times \mathbb{R})$ *satisfaisant, pour tout couple d'entropie-flux d'entropie,*

$$\partial_t S(y) + \partial_x F(y) \leqslant 0$$

au sens faible, autrement dit que pour tout $\varphi \in C^\infty_0(\mathbb{R}^+ \times \mathbb{R})$ *à valeurs positives, on ait*

$$\int_0^{+\infty} \int_{\mathbb{R}} (S(y)(t,x)\partial_t\varphi(t,x) + F(y)\partial_x\varphi(t,x))\, dxdt + \int_{-\infty}^{+\infty} S(y^0)(x)\varphi(0,x)dx \geqslant 0.$$

Montrer que y *est une solution faible de* (4.22). *Indication : on admettra un résultat de densité des fonctions* $C^\infty_0(\mathbb{R}^+ \times \mathbb{T})$ *positives dans* $W^{1,1}(\mathbb{R}^+ \times \mathbb{R})$ *positives.*

Exercice 29 (Conditions de Rankine-Hugoniot pour des flux plus généraux)**.** *On revient au cas général d'une loi de conservation donnée par*

$$\begin{cases} \partial_t y(t,x) + \partial_x f(y(t,x)) = 0,\ (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ y(0,x) = y^0(x), \end{cases} \tag{4.49}$$

avec f *de classe* C^1.

1. *Donner une formulation faible équivalente à* (4.49).
2. *En reprenant les calculs effectués dans le cas de l'équation de Burgers, montrer une relation de Rankine-Hugoniot de la forme*

$$\sigma = \frac{f(y_d) - f(y_g)}{y_d - y_g}.$$

3. *Montrer que les discontinuités d'amplitude très petites se déplacent à une vitesse proche de la vitesse caractéristique $f'(y_g)$.*
4. *Donner une CNS sur y_g et y_d pour que la solution du problème de Riemann vérifie l'inégalité d'Oleinik* (4.34).

Exercices non corrigés

Exercice 30 (Équation des ondes)**.** *On s'intéresse de l'équation des ondes posée sur tout $\mathbb{R}$:*

$$\begin{cases} \dfrac{\partial^2}{\partial t^2}u(t,x) - c^2\dfrac{\partial^2}{\partial x^2}u(t,x) = 0, \quad (t,x) \in]0;\infty[\times\mathbb{R} \\ u(0,x) = u_0(x), \\ \frac{\partial}{\partial t}u(0,x) = u_1(x). \end{cases} \tag{4.50}$$

1. *En remarquant que*

$$\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial x^2} = \left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right),$$

   montrer qu'il existe au plus une solution à (4.50).
2. *On suppose que $u_0 \in C^2(\mathbb{R})$ et $u_1 \in C^1(\mathbb{R})$. Montrer qu'il existe une unique solution donnée par la formule de D'Alembert*

$$u(t,x) = \frac{1}{2}(u_0(x-ct) + u_0(x+ct)) + \frac{1}{2c}\int_{x-ct}^{x+ct} u_1(s)\,ds.$$

3. *Donner la formulation faible associée à* (4.50)*. Pour quelles régularités minimales sur u_0 et u_1 cette formulation faible a-t-elle un sens ?*
4. *Pour $u_0, u_1 \in L^1_{loc}(\mathbb{R})$, montrer qu'il existe une unique solution faible dans $L^1_{loc}(\mathbb{R})$ à* (4.50) *donnée par la formule de d'Alembert précédente. Si $u_0, u_1 \in L^\infty(\mathbb{R})$, a-t-on nécessairement $u \in L^\infty(\mathbb{R}^+ \times \mathbb{R})$?*

Exercice 31 (Équation de transport avec condition au bord)**.** *On s'intéresse à la résolution du problème suivant, posé sur une demi-droite en espace,*

$$\frac{\partial u}{\partial t}(t,x) + a\,\frac{\partial u}{\partial x}(t,x) = 0,\ t > 0,\ x \in \mathbb{R}_+,$$

$$u(0,x) = u_0(x),\ x \in \mathbb{R}_+,$$

où a est une constante et $u_0 \in C^1(\mathbb{R}_+)$.

1. *On suppose que $a < 0$. Montrer que le problème admet une unique solution. On suppose maintenant que $a > 0$.*

2. *Montrer que le problème est mal posé (autrement dit, soit il peut ne pas exister de solutions, soit il peut en exister une infinité).*
3. *On ajoute la condition au bord*

$$u(t,0) = g(t),\ t > 0,$$

où $g \in C^1(\mathbb{R})$. Montrer que le problème admet une solution de classe C^1, que l'on déterminera, si et seulement si l'on a

$$g(0) = u_0(0) \text{ et } g'(0) + a\, u_0'(0) = 0.$$

4. *On s'intéresse maintenant au cas des solutions faibles. Donner la formulation faible associée au problème dans le cadre $L^\infty(0,+\infty)$, et démontrer l'existence et l'unicité des solutions dans ce cadre-là.*

Exercice 32 (Système d'équations de transport)**.** *On considère $n \in \mathbb{N}^*$, ainsi que $A \in \mathcal{M}_n(\mathbb{R})$, supposée **diagonalisable** à valeurs propres réelles non nulles. Soit $Y^0 \in C^1(\mathbb{R}, \mathbb{R}^n)$.*

1. *Montrer qu'il existe une unique solution $Y \in C^1(\mathbb{R}, \mathbb{R}^n)$ au système d'équations de transport*

$$\begin{cases} \partial_t Y(t,x) + \partial_x AY(t,x) = 0,\ (t,x) \in \mathbb{R} \times \mathbb{R}, \\ Y(0,\cdot) = Y_0. \end{cases}$$

2. *En faisant des IPP formelles, donner la formulation faible naturellement associée à ce système, et démontrer (en s'appuyant sur les résultats du cours) l'existence et l'unicité des solutions faibles.*

Exercice 33 (Équation de transport en dimension quelconque)**.** *On s'intéresse à l'équation de transport dans tout l'espace $\mathbb{R}^n$*

$$\begin{cases} \dfrac{\partial}{\partial t} u(t,x) + b \cdot \nabla_x u(t,x) = 0, & (t,x) \in (0,\infty) \times \mathbb{R}^n, \\ u(0,x) = g(x), & x \in \mathbb{R}^n. \end{cases} \tag{4.51}$$

où b est un vecteur fixe de $\mathbb{R}^n$ (indépendant de x et t).

1. *On suppose d'abord que $g \in C^1(\mathbb{R}^n)$. Montrer qu'il existe une unique solution $u \in C^1(\mathbb{R}^2)$ à (4.51), dont on donnera une solution explicite.*
2. *On suppose maintenant que $g \in L^\infty(\mathbb{R})$. Donner une formulation faible associée à ce problème pour tout temps même négatif.*
3. *Démontrer l'existence et l'unicité d'une solution faible, et donner son expression explicite.*

Exercice 34. *1. En utilisant la méthode des caractéristiques, montrer l'existence et l'unicité d'une solution faible de l'équation de Burgers associée à la condition initiale*

$$y_0(x) = \begin{cases} 1+x, \ -1 \leqslant x < 0, \\ 1-2x, 0 < x < 1/2, \\ 0 \, sinon, \end{cases}$$

sur $[0, \frac{1}{2}[$ *et donner son expression explicite. S'agit-il d'une solution classique ?*

2. En utilisant les relations de Rankine-Hugoniot, déterminer une solution faible globale et montrer qu'il s'agit de l'unique solution entropique.

Exercice 35 (Interaction entre une onde de raréfaction et une onde de choc)**.** *On cherche à calculer la solution entropique pour* (4.22) *associée à la condition initiale*

$$y_0(x) = \mathbb{1}_{[0,1]}(x).$$

1. Montrer que p.p.,

$$\mathbb{1}_{[0,1]} = \mathbb{1}_{]0,+\infty]} - \mathbb{1}_{[1,+\infty[}.$$

2. Donner la solution entropique associée à la condition initiale

$$y_0(x) = -\mathbb{1}_{[1,+\infty[}.$$

Faire un dessin dans le plan (t, x) *représentant les valeurs de* y^0*.*

3. A l'aide de l'exercice 27 et de la question précédente, donner l'expression de la solution entropique pour $t < 2$*, en partant du principe que l'onde de choc et l'onde de raréfaction "n'interagissent pas". On pourra faire un dessin dans le plan* (t, x)*.*

4. On s'intéresse maintenant au cas $t \geqslant 2$*. Montrer qu'une solution est donnée par une onde de choc se propageant selon la courbe caractéristique* $x = \sqrt{2t}$ *(la fonction étant* C^1 *en dehors de cette courbe), et que cette solution est admissible.*

Exercice 36 (Approximation d'une onde de raréfaction)**.** *1. Considérons une condition initiale* $y^0 \in C^1(\mathbb{R})$*, croissante. Montrer que pour tout* $n \in \mathbb{N}^*$*,* y *est une solution classique de* (4.22) *associée à la condition initiale si et seulement* $z_n(t, x) = y(nt, nx)$ *est une solution classique de* (4.22) *associée à une condition initiale que l'on précisera.*

2. On introduit

$$y_0(x) = \begin{cases} 0, x < -1/2, \\ 2(x+\frac{1}{2})^2, x \in [-\frac{1}{2}, 0], \\ 1 - 2\left(x - \frac{1}{2}\right)^2, x \in [0, \frac{1}{2}], \\ 1, x \geqslant \frac{1}{2}. \end{cases}$$

Montrer que y_0 est de classe C^1 sur $\mathbb{R}$ et qu'elle admet une unique solution globale dont on précisera l'expression.

3. *Donner la solution y_n associée à la condition initiale $y^0(n\cdot)$.*
4. *Montrer que quand $n \to \infty$, y_n converse simplement vers la solution donnée à l'exercice 27.*

Exercice 37 (Équation de Burgers avec potentiel)**.** *On considère l'équation suivante :*

$$\begin{cases} \partial_t y(t,x) + \dfrac{1}{2}\partial_x(y^2)(t,x) + y(t,x) = 0,\ (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ y(0,x) = y^0(x), \end{cases} \tag{4.52}$$

1. *On suppose dans un premier temps y^0 croissante. En reprenant la preuve du Théorème 4.17, montrer qu'il existe une unique solution de la forme*

$$y(t,x) = y^0(x_0)e^{-t},$$

où x_0 est à déterminer.

2. *Proposer une définition adéquate de solution faible pour* (4.52)*, sur le modèle de l'équation de Burgers.*
3. *Montrer que les relations de Rankine-Hugoniot sont inchangées pour le problème de Riemann.*
4. *En déduire une solution associée au problème de Riemann pour $y_g = 1$ et $y_d = 0$.*

Exercice 38. *On considère l'équation* (4.22) *avec donnée initiale*

$$y_0(x) = \begin{cases} 1,\ x < 0, \\ 1 - x,\ x \in [0,1], \\ 0,\ x > 1. \end{cases}$$

et on cherche à en donner une solution sur $\mathbb{R}^+ \times \mathbb{R}$.

1. *Identifier les courbes caractéristiques et en et dessiner quelques-unes dans le plan (on mettra le temps $t \geqslant 0$ en abcsisse). En quel(s) points mes courbes caractéristiques se croisent-elles ?*
2. *Montrer que pour $t \leqslant 1$, on peut trouver une solution faible continue y que l'on construira à l'aide de la méthode des caractéristiques.*
3. *Pour "sélectionner" des courbes caractéristiques, on effectue la modification suivante dans le dessin précédent : dès qu'une courbe caractéristique parcourue dans le sens $t > 0$ rencontre la droite oblique $(t, t/2 + 1/2)$, elle s'arrête. Effectuer un nouveau dessin. À partir de ce dessin, trouver un candidat raisonnable y pour être solution du problème, en utilisant la méthode des caractéristiques.*
4. *Montrer que c'est effectivement une solution faible. Est-elle continue ?*
5. *La solution ainsi trouvée est-elle la solution entropique ?*

4.4 Corrigés des exercices 1 à 5

Exercice 1. *On remarque que $a(t,x) = x$ est de classe C^1 et globalement Lipschitzienne par rapport à la seconde variable. On a donc existence et unicité d'une solution donnée par la méthode des caractéristiques. On commence donc par résoudre l'EDO*

$$X'(s) = X(s), X(t) = x.$$

On trouve comme solution $X(s) = xe^{s-t}$. Donc $X(0) = xe^{-t}$. Donc la solution de l'équation est donnée par $y(t,x) = y^0(xe^{-t})$. On pourra dériver cette expression par rapport à t et x pour se convaincre que c'est bien une solution.

Si maintenant $y^0 \in L^p(\mathbb{R})$, il est évident que y aussi (en séparant le cas $+\infty$ et en faisant le changement de variable $xe^{-t} = z$ dans le cas $p < \infty$). Enfin, pour $p = 1$, le changement de variable $xe^{-t} = z$ donne que pour tout $t > 0$,

$$\int_{\mathbb{R}} y(t,x)dx = \int_{\mathbb{R}} y^0(xe^{-t})dx = e^t \int_{\mathbb{R}} y^0(x)dx.$$

Donc l'intégrale est conservée seulement si $\int_{\mathbb{R}} y^0(x)dx = 0$. Sinon, l'intégrale croit vers $+\infty$ ou décroit vers $-\infty$ en $+\infty$, selon le signe de $\int_{\mathbb{R}} y^0(x)dx$.

Exercice 2. 1. *On repart alors de*(4.11) *où* 0 *est remplacé par s :*

$$x = X(t,s,X(s,t,x)).$$

On dérive ceci par rapport à t. (4.12) *est changé en*

$$\partial_2 X(s,t,x)\partial_3 X(s,t,X(s,t,x)) = -a(t,x).$$

De même, en dérivant par rapport à x, on obtient que

$$\partial_3 X(t,s,X(s,t,x))(\partial_2 X(s,t,x) + a(t,x)\partial_3 X(s,t,x)) = 0. \tag{4.53}$$

Comme pour obtenir (4.14)*, on en déduit que*

$$\partial_2 X(s,t,x) + a(t,x)\partial_3 X(s,t,x) = 0.$$

2. *Pour simplifier les notations, on pose*

$$\rho(t,x) = \exp\left(-\int_0^t \partial_x a(s,X(s,t,x))ds\right)$$

On veut montrer que

$$\partial_t \rho + a\partial_x \rho + \rho\partial_x a = 0.$$

En posant

$$\xi(t,x) = -\int_0^t \partial_x a(s,X(s,t,x))ds,$$

On a que $\rho = e^{\xi}$ *et donc*

$$\partial_t \rho + a\partial_x \rho + \rho \partial_x a = \rho \left(\partial_t \xi + a\partial_x \xi + \partial_x a \right).$$

On se ramène donc à montrer que

$$\partial_t \xi + a\partial_x \xi + \partial_x a = 0. \tag{4.54}$$

Pour conclure, on a

$$\begin{aligned} \partial_t \xi(t,x) &= -\partial_x a(t, X(t,t,x)) - \int_0^t \partial_2 X(s,t,x) \partial_{xx} a(s, X(s,t,x)) ds \\ &= -\partial_x a(t,x) - \int_0^t \partial_2 X(s,t,x) \partial_{xx} a(s, X(s,t,x)) ds, \end{aligned} \tag{4.55}$$

et par dérivation sous l'intégrale,

$$\partial_x \xi(t,x) = -\int_0^t \partial_3 X(s,t,x) \partial_{xx} a(s, X(s,t,x)) ds. \tag{4.56}$$

D'où le résultat voulu par la question précédente.

3. *On raisonne par analyse-synthèse. On considère une solution* y *de l'équation* (4.6)*, et on regarde son comportement le long d'une courbe caractéristique* $(s, X(s))$*. On a alors, en utilisant l'équation* (4.6) *où on a développé la dérivée en* x*,*

$$\begin{aligned} \frac{d}{ds} y(s, X(s)) &= \partial_t y(s, X(s)) + X'(s) \partial_x y(s, X(s)) \\ &= \partial_t y(s, X(s)) + a(s, X(s)) \partial_x y(s, X(s)) \\ &= -\partial_x a(s, X(s)) y(s, X(s)). \end{aligned}$$

La solution n'est plus constante le long des caractéristiques, mais suit une dynamique que nous pouvons suivre de manière explicite :

$$y(s, X(s)) = y(0, X(0)) \exp\left(-\int_0^t \partial_x a(s, X(s)) ds \right).$$

En appelant $X(t, t_0, x_0)$ *le flot associé à* a*, on obtient que*

$$y(t,x) = y(t, X(t,t,x)) = y(0, X(0,t,x)) \exp\left(-\int_0^t \partial_x a(s, X(s,t,x)) ds \right),$$

et donc l'expression (4.47)*. Posons*

$$z(t,x) = y(0, X(0,t,x)).$$

On sait déjà que z *vérifie* (4.7)*. On a donc par la question précédente*

Inversement, y donnée par (4.47) vérifie bien $y(0,\cdot) = y^0$, est de classe C^1 sur $\mathbb{R}^{+*} \times \mathbb{R}$, et vérifie bien la première équation de (4.6) par la question prexédente :

$$\begin{aligned}\partial_t y + \partial_x(ay) &= \partial_t y + a\partial_x y + y\partial_x a \\ &= (\partial_t z + a\partial_x z)\rho + (\partial_t \rho + a\partial_x \rho)z + \rho z \partial_x a \\ &= (\partial_t \rho + a\partial_x \rho + \rho \partial_x a)z \\ &= 0.\end{aligned}$$

4. *De l'équation*
$$\partial_1 X(t,0,z) = a(t, X(t,0,z)),$$
On tire en dérivant par rapport à z que
$$\partial_3 \partial_1 X(t,0,z) = \partial_3(X(t,0,z))\partial_x a(t, X(t,0,z)).$$
En admettant que le flot est C^2, on voit que l'on peut échanger les dérivées partielles, donc $Y = \partial_3 X$ est solution de l'EDO
$$\partial_1 Y(t,0,z) = Y(t,0,z)\partial_x a(t, X(t,0,z)).$$
De plus, $X(0,0,z) = z$, donc $\partial_3 X(0,0,z) = 1$. On en déduit donc (4.48) en intégrant cette EDO, avec comme condition initiale 1.

5. *On fixe $t > 0$. On sait que a est globalement Lipschitizienne et C^1.*
On effectue le changement de variable $X(0,t,x) = z$, qui s'inverse (comme vu pour résoudre (4.28)) en $x = X(t,0,z)$. On a alors
$$dx = \partial_3 X(t,0,z)dz.$$
On en déduit que
$$\begin{aligned}&\int_{\mathbb{R}} |y^0(X(0,t,x))| \exp\left(-\int_0^t \partial_x a(s, X(s,t,x))ds\right) \\ &= \int_{\mathbb{R}} |y^0(z)| \partial_3 X(t,0,z) \exp\left(-\int_0^t \partial_x a(s, X(s,t,X(t,0,z)))ds\right) dz.\end{aligned}$$
On remarque que
$$X(s,t,X(t,s,z)) = X(s,0,z).$$
Donc, par (4.48),
$$\begin{aligned}&\int_{\mathbb{R}} |y^0(X(0,t,x))| \exp\left(-\int_0^t \partial_x a(s, X(s,t,x))ds\right) \\ &= \int_{\mathbb{R}} |y^0(z)| \partial_3 X(t,0,z) \exp\left(-\int_0^t \partial_x a(s, X(s,0,z)ds\right) dz \\ &= \int_{\mathbb{R}} |y^0(z)| dz.\end{aligned}$$

Donc pour tout $t > 0$, $y(t, \cdot) \in L^1(\mathbb{R})$.
On peut faire le même calcul en enlevant les valeurs absolues, ce qui montre que la norme L^1 est conservée.

Exercice 3. *1. f est clairement de classe C^1 sur $\mathbb{R}^{+*} \times \mathbb{R}$. De plus,*

$$\partial_t f(t, x) = -\frac{x}{t^2}$$

et

$$\frac{1}{2}\partial_x f^2(t, x) = \frac{1}{2}\partial_x \frac{x^2}{t^2} = \frac{x}{t^2},$$

d'où le résultat.

2. La condition initiale est L^∞ et C^1 par morceaux. De plus, elle est croissante. On rappelle que les caractéristiques sont données par $X_t(x) = x + ty_0(x)$. Ainsi, ici, pour $x < 0$, les caractéristiques sont données par

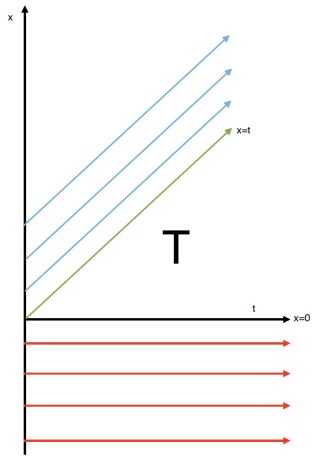

FIGURE 4.1 – Courbes caractéristiques

$$(t, X_t(x)) = (t, x).$$

Ce sont les lignes en rouge sur la figure 2. Pour $x \geqslant 0$, les caractéristiques sont données par

$$(t, X_t(x)) = (t, x + t).$$

Ce sont les lignes en bleu et la ligne en vert sur la figure 2. De plus, ces caractéristiques ne se croisent jamais, puisque dans le premier cas, on a toujours $x < 0$ et dans le deuxième cas, on a toujours $x + t > 0$. Prenons maintenant un point quelconque $(t, x) \in \mathbb{R}^{+*} \times \mathbb{R}$. De trois choses l'une.
— *Soit $x < 0$. Auquel cas, par constance le long des caractéristiques, on a $y(x) = y^0(x) = 0$.*

— *Soit $x > t$. Auquel cas, on peut résoudre l'équation des caractéristiques $y + t = x$ sous la forme $y = x - t$. Donc le pied de la caractéristique est donné par $x - t$ est on a alors, puisque $x - t > 0$, $y(x) = y^0(x - t) = 1$.*
— *Soit $0 < x < t$. On a alors un problème, car aucune caractéristique ne passe (zone en triangle T sur la figure 2).*

On voit donc qu'il faut "comprendre" comment remplir la zone T. On remarque alors que si l'on pose sur T, $y(t, x) = f(t, x)$, on obtient une fonction C^1 par morceaux. De plus, de part et d'autre de la droite d'équation $x = 0$ paramétrée par $(t, 0) = (t, \xi(t))$, on a que $\xi'(t) = 0$ et

$$y^+(t) = \lim_{x\to 0, s\to t^+} \frac{x}{s} = 0 = y^-(t).$$

Enfin, de part et d'autre de la droite d'équation $x = t$ paramétrée par $(t, t) = (t, \xi(t))$, on a que $\xi'(t) = 1$ et

$$y^-(t) = \lim_{x\to 1, s\to t^-} \frac{x}{s} = 1 = y^+(t).$$

Le Théorème 4.33 assure donc que y est bien une solution de (4.22) pour cette condition initiale.

3. *On fixe $t > 0$. La fonction $x \mapsto y(t, x)$ est alors croissante et continue sur $\mathbb{R}$, et dérivable sauf aux points $x = 0$ et $x = t$, de dérivée soit 0, soit $1/t$. Il est donc tentant de poser $C(t) = \frac{1}{t}$ et de voir si cela convient. Faisons une disjonction des cas.*
— *Si on a soit $x' > x > t$ ou $x < x'' < 0$, on a*

$$y(t, x') - y(t, x) = 0 \leqslant \frac{x' - x}{t}.$$

— *Si $x' > t$ et $x < 0$, on a*

$$y(t, x') - y(t, x) = 1 = \frac{x' - x}{x' - x} \leqslant \frac{x' - x}{x'} \leqslant \frac{x' - x}{t}.$$

— *Si $0 < x' < x < t$, on a*

$$y(t, x') - y(t, x) = \frac{x' - x}{t}.$$

— *Si $0 < x' < t$ et $x < 0$, on a*

$$y(t, x') - y(t, x) = \frac{x}{t} \leqslant \frac{x' - x}{t}.$$

— *Si $0 < x < t$ et $x' > t$, on a*

$$y(t, x') - y(t, x) = 1 - \frac{x}{t} = \frac{t - x}{t} \leqslant \frac{x' - x}{t}.$$

D'où le résultat.

4. *Ainsi, ici, pour $x < 0$, les caractéristiques sont données par*

$$(t, X_t(x)) = (t, x + y_g t).$$

Pour $x \geqslant 0$, les caractéristiques sont données par

$$(t, X_t(x)) = (t, x + y_d t).$$

On voit que pour $x < y_g t$, il faut prendre $y(t,x) = y_g$. Pour $x > y_d t$, il faut prendre $y(t,x) = y_d$. Il reste à comprendre comment prolonger par continuité sur l'ensemble $\{x \in [y_g t, y_d t]\}$. Il s'avère que la même fonction convient, et donc on prend

$$y(t,x) = \frac{x}{t}.$$

Les mêmes calculs que les questions précédentes montrent que la version étendue du théorème de Rankine-Hugoniot s'applique et que la solution est admissible.

Exercice 4. 1. *On rappelle que la dérivée au sens des distributions de la fonction $|x|$ est la fonction signe(x). On en déduit donc qu'au sens des distributions, $S'(x) = signe(x-k)$.*

De même, F est continue et de classe C^1 en dehors du point $s = k$. On en déduit donc que sa dérivée au sens des distributions est donnée pour $x < k$ par

$F'(x) = -x = signe(x-k)x$ et pour $x > k$ par $F'(x) = x = signe(x-k)x$. D'où le résultat voulu.

2. *On suppose S et F de classe C^1. On a alors par dérivation de fonctions composées et en utilisant la relation $xS'(x) = F'(x)$ que*

$$\partial_t S(y(t,x)) = \partial_t y(t,x) S'(y(t,x))$$

et

$$\partial_x F(y(t,x)) = \partial_x y(t,x) F'(y(t,x)) = \partial_x y(t,x) y(t,x) S'(y(t,x)).$$

On en déduit le résultat voulu en se rappelant que (4.24) est vérifié pour toute solution régulière.

3. *On prend $S(x) = signe(x)$ et $F(x) = signe(x)\frac{x^2}{2}$, qui sont tous les deux des paires d'entropie-flux d'entropie comme on le vérifie par un calcul simple. On en déduit donc déjà que pour tout $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$ à valeurs positives, on a*

$$\int_0^{+\infty} \int_{\mathbb{R}} \left(y(t,x)\partial_t \varphi(t,x) + \frac{y^2(t,x)}{2} \partial_x \varphi(t,x) \right) dxdt = -\int_{-\infty}^{+\infty} y^0(x)\varphi(0,x)dx.$$

En remplaçant φ en $-\varphi$, on a donc aussi que tout $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$ à valeurs négatives, on a

$$\int_0^{+\infty} \int_{\mathbb{R}} \left(y(t,x)\partial_t\varphi(t,x) + \frac{y^2(t,x)}{2}\partial_x\varphi(t,x) \right) dxdt = -\int_{-\infty}^{+\infty} y^0(x)\varphi(0,x)dx.$$

Remarquons que par densité des fonctions $C_0^\infty(\mathbb{R}^+ \times \mathbb{T})$ positives dans $W^{1,1}(\mathbb{R}^+ \times \mathbb{R})$ positives (il suffit de remarquer que le raisonnement par convolution qui permet d'obtenir ce résultat préserve la positivité), ce résultat reste vrai pour tout $\varphi \in W^{1,1}(\mathbb{R}^+ \times \mathbb{R})$ de signe constant. Il reste à comprendre comment passer à des $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$ de signe quelconque. On décompose φ en sa partie positive φ^+ et sa partie négative φ^-. Ces deux fonctions ne sont plus forcément C^∞, mais par contre, elles restent dans $W^{1,1}(\mathbb{R}^+ \times \mathbb{R})$, puisque la dérivée au sens des distributions est la dérivée usuelle ici, sauf aux points de discontinuité de la dérivée. Ainsi, on peut appliquer les résultats précédents

$$\int_0^{+\infty} \int_{\mathbb{R}} \left(y\partial_t\varphi^\pm + \frac{y^2}{2}\partial_x\varphi^\pm \right) dxdt = -\int_{-\infty}^{+\infty} y^0(x)\varphi(0,x)dx.$$

En sommant la partie positive et la partie négative, on obtient le résultat voulu.

Exercice 5. *1. On fait des IPP formelles et on se rend compte qu'on est dans le même cas que Burgers : Soit $y^0 \in L^\infty(\mathbb{R})$. On appelle solution faible toute fonction $y \in L^\infty\mathbb{R}^+ \times \mathbb{R})$ telle que pour tout $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$, on ait*

$$\int_0^{+\infty} \int_{\mathbb{R}} \left(y\partial_t\varphi + f(y)\partial_x\varphi \right) dxdt = -\int_{-\infty}^{+\infty} y^0(x)\varphi(0,x)dx. \tag{4.57}$$

2. On procède come pour Burgers, en abrégeant les arguments redondants. Soit $\varphi \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R})$. On veut avoir (4.57). Soit $\varepsilon > 0$. On pose

$$B_\varepsilon = B((t,\sigma t),\varepsilon),$$

$$D_{\varepsilon,g} = B((t,\sigma t),\varepsilon) \cap \{\sigma t > x\},$$
$$D_{\varepsilon,d} = B((t,\sigma t),\varepsilon) \cap \{\sigma t < x\}.$$

On prend maintenant ε suffisamment petit pour que B_ε n'intersete pas la droite $t = 0$, et on considère $\varphi \in C_0^\infty(B_\varepsilon)$. Donc (4.57) devient (puisqu'un morceau de droite est de mesure de Lebesgue nulle)

$$\int_{D_{\varepsilon,g}} \left(y(t,x)\partial_t\varphi(t,x) + f(y(t,x))\partial_x\varphi(t,x) \right) dxdt$$
$$+ \int_{D_{\varepsilon,d}} \left(y(t,x)\partial_t\varphi(t,x) + f(y(t,x))\partial_x\varphi(t,x) \right) dxdt = 0.$$

Au vue de la forme particulière de la solution étudiée ici, on doit donc avoir que

$$\int_{D_{\varepsilon,g}} (y_g \partial_t \varphi(t,x) + f(y_g)\partial_x \varphi(t,x))\, dxdt + \int_{D_{\varepsilon,d}} (y_d \partial_t \varphi(t,x) + f(y_d)\partial_x \varphi(t,x))\, dxdt = 0.$$

En appliquant la formule de Green ,on obtient

$$\int_{\partial D_{\varepsilon,g}} (y_g n_t + f(y_g) n_x)\, \varphi(t,x) dS + \int_{\partial D_{\varepsilon,d}} (y_d n_t + f(y_d) n_x)\, \varphi(t,x) dS = 0.$$

Comme $\varphi \in C_0^\infty(B_\varepsilon)$, le seul endroit où cette intégrale est non nulle est sur le segment $\{x = at\}$. De plus, sur ce segment, on a que n_g est positivement colinéaire à $\begin{pmatrix} -\sigma \\ 1 \end{pmatrix}$ et n_d est positivement colinéaire à $\begin{pmatrix} \sigma \\ -1 \end{pmatrix}$. On en déduit donc que

$$(-\sigma y_g + f(y_g) + \sigma y_d - f(y_d)) \int_{\partial D_{\varepsilon,g} \cap \{x = \sigma t\}} \varphi(t,x) dS = 0,$$

et ceci pour tout $\varphi \in C_0^\infty(B_\varepsilon)$. Donc

$$-\sigma y_g + f(y_g) + \sigma y_d - f(y_d) = 0,$$

ce qui donne ce qu'on voulait.

3. *C'est juste un développement limité : quand $y_d \to y_g$, on a*

$$f(y_d) = f(y_g) + (y_d - y_g) f'(y_g) + o(y_d - y_g).$$

On obtient ce qu'on veut en passant $f(y_g)$ à gauche et en divisant par $(y_d - y_g)$.

4. *On reprend les calculs faits dans le cas du problème de Riemann pour Burgers, et on se rend compte que la condition est exactement la même : $y_g > y_d$.*

Chapitre 5

Problèmes d'évolution paraboliques : l'exemple de l'équation de la chaleur

Ce chapitre est une brève introduction à l'étude mathématique d'équations aux dérivées partielles dépendant du temps. Ici, nous étudierons surtout l'équation de la chaleur.

5.1 Préliminaires

On se pose généralement différentes questions lors de l'étude d'une équation d'évolution.

La première est bien sûr *l'existence et l'unicité de solutions* dans un espace fonctionnel bien choisi. Pour cela, il pourra être très utile de commencer par choisir des espaces fonctionnels assez "gros", c'est-à-dire contenant beaucoup plus de fonctions que celles qui sont régulières par rapport à toutes leurs variables. On parle alors de *solutions faibles*. Intuitivement, plus l'espace fonctionnel est grand et plus il sera facile de démontrer l'existence de la solution.

Pour effectuer l'analyse de ces solutions faibles, nous aurons besoin d'introduire un outil essentiel : l'*intégrale de Bochner*. Cette nouvelle notion d'intégrale généralise la notion d'intégrale de Lebesgue (que vous connaissez bien dans le cas de fonctions à valeurs *scalaires*) pour des fonctions à valeurs dans des *espaces de Banach*. L'introduction de cette nouvelle notion fera l'objet de la Section 5.1.3.

Une autre question importante est celle de la dépendance de la solution en fonction des conditions initiales et des divers paramètres apparaissant dans l'équation. D'une part on peut se demander comment la solution varie (dans l'espace fonctionnel choisi) si on change un peu ces paramètres. Mais on peut aussi se demander quelle est la régularité de la solution lorsque la condition initiale est elle-même régulière ainsi que les autres paramètres de l'équation. On peut ainsi obtenir l'existence et

l'unicité de solutions régulières *a posteriori*.

L'utilisation de solutions faibles peut donc être soit un intermédiaire utile pour démontrer l'existence de solutions plus régulières, soit une nécessité lors de l'étude d'équations pour lesquelles on ne s'attend pas à ce que la solution soit ou reste régulière au cours du temps.

Enfin, on cherche généralement ensuite à décrire un peu plus précisément le comportement de la solution au cours du temps, en particulier en relation avec des motivations physiques (signe de la solution, vitesse de propagation, comportement en temps grand, etc). Le comportement qualitatif peut alors être très différent suivant le type d'équation considérée.

Dans cette première partie, nous présentons certains outils qui seront nécessaires à l'étude des équations d'évolution.

5.1.1 Théorème de représentation de Riesz (complément) et triplets de Gelfand

Dans le cas des espaces de Hilbert, on peut caractériser de manière très simple l'ensemble des éléments du dual.

Théorème 5.1. *[Théorème de représentation de Riesz] Soit $(H, \langle\cdot,\cdot\rangle)$ un espace de Hilbert. Pour tout $\varphi \in H'$, il existe un unique $f \in H$ tel que pour tout $y \in H$, on ait $\varphi(y) = \langle y, f\rangle$. f est appelé le représentant de φ. De plus, l'application*

$$\Gamma : \varphi \in H' \mapsto f \in H$$

est linéaire si $\mathbb{K} = \mathbb{R}$ et antilinéaire si $\mathbb{K} = \mathbb{C}$, isométrique (i.e. pour tout $\varphi \in H'$, on a $|||\varphi||| = ||f||$), et donc continue.

Remarque 5.2.
- *H et H' sont donc isomorphes, au sens où il existe une bijection isométrique entre ces deux espaces.*
- *On peut notamment munir H' d'une structure d'espace de Hilbert, en identifiant un élément de H' avec son représentant : si $\varphi \in H'$ de représentant f et $\varphi' \in H'$ de représentant f', on pose $\langle\varphi, \varphi'\rangle_{H'} = \langle f, f'\rangle_H$. De plus, cette norme associée est équivalente (égale même) à la norme d'opérateur habituelle.*
- *Pour un opérateur antilinéaire continu, on peut aussi définir une norme d'opérateur de la même manière que pour les applications linéaires continues, dont l'existence caractérise la continuité.*
- *Toute isométrie surjective (i.e. bijective) entre deux espaces vectoriels normés est telle que son inverse est aussi une isométrie, et est donc notamment continue.*
- *On peut montrer que toute application isométrique entre deux espaces de Hilbert conserve le produit scalaire, et est de plus forcément une application affine (et donc linéaire si elle envoie 0 sur 0).*

Preuve : Le théorème de représentation de Riesz a déjà été démontré en première année, on l'admet donc, et on va se concentrer sur les raffinements. Montrons maintenant la linéarité dans le cas $\mathbb{K} = \mathbb{R}$. Si $(\varphi_1, \varphi_2) \in (H')^2$, de représentants respectifs $f_1 \in H$ et $f_2 \in H$, et $\lambda \in \mathbb{R}$, alors on a, pour tout $y \in H$ et en utilisant la bilinéarité du produit scalaire que

$$(\varphi_1 + \lambda\varphi_2)(y) = \varphi_1(y) + \lambda\varphi_2(y) = \langle y, f_1\rangle + \langle \lambda y, f_2\rangle = \langle y, f_1 + \lambda f_2\rangle.$$

Par unicité du représentant, en identifiant, le représentant de $\varphi_1 + \lambda\varphi_2$ est bien nécessairement $f_1 + \overline{\lambda} f_2$. En revenant à la définition de Γ, ceci se réécrit

$$\Gamma(\varphi_1 + \lambda\varphi_2) = \Gamma(\varphi_1) + \lambda\Gamma(\varphi_2).$$

Montrons le caractère antilinéaire dans le cas $\mathbb{K} = \mathbb{C}$. La preuve est essentiellement la même. Si $(\varphi_1, \varphi_2) \in (H')^2$, de représentants respectifs $f_1 \in H$ et $f_2 \in H$, et $\lambda \in \mathbb{R}$, alors on a, pour tout $y \in H$ et en utilisant la sesquilinéarité du produit scalaire que

$$(\varphi_1 + \lambda\varphi_2)(y) = \varphi_1(y) + \lambda\varphi_2(y) = \langle y, f_1\rangle + \langle \lambda y, f_2\rangle = \langle y, f_1 + \overline{\lambda} f_2\rangle.$$

Par unicité du représentant, en identifiant, le représentant de $\varphi_1 + \overline{\lambda}\varphi_2$ est bien nécessairement $f_1 + \lambda f_2$. En revenant à la définition de Γ, ceci se réécrit

$$\Gamma(\varphi_1 + \lambda\varphi_2) = \Gamma(\varphi_1) + \overline{\lambda}\Gamma(\varphi_2).$$

En ce qui concerne la norme de Γ, on remarque que par l'inégalité de Cauchy-Schwartz, on a

$$|||\varphi||| = \sup_{||x|| \leqslant 1} |\varphi(x)| = \sup_{||x|| \leqslant 1} |\langle x, f\rangle| \leqslant ||f||\,||x|| \leqslant ||f||.$$

De plus, en prenant $x = \frac{f}{||f||}$ (sauf si $f = 0$, auquel cas tout est beaucoup plus simple), on a que la borne supérieure est atteinte. Ainsi, $|||\varphi||| = ||f||$, *i.e.* $|||\varphi||| = ||\Gamma(\varphi)||$. Γ est donc notamment continue (de norme d'opérateur 1), injective (c'est le cas de toute isométrie, puisque le noyau est réduit à 0). Elle est surjective puisque c'est exactement ce que nous dit le théorème de Riesz : pour tout $\varphi \in H'$, on peut trouver un représentant $f \in H$, *i.e.* $\varphi = \Gamma(f)$. ♢

Remarque 5.3. *Si $\varphi \in H'$, alors $Ker(\varphi)$ est fermé. Inversement, si φ est une forme linéaire sur H (pas forcément supposée continue) et que $Ker(\varphi)$ est fermé, alors φ est continue.*

Dorénavant, on identifiera implicitement H et H' au travers de l'isométrie Γ. Autrement dit, on identifiera $\varphi \in H'$ avec son représentant $f = \Gamma(\varphi)$, et on fera "comme si $\varphi = f$". Cette identification peut parfois poser des problèmes techniques (notamment pour les espaces imbriqués les uns dans les autres). Une situation typique est la suivante.

On considère $(H, \langle \cdot, \cdot \rangle)$ un espace de Hilbert réel pour simplifier, que l'on identifie à son dual. On considère un autre espace de Hilbert $(V, (\cdot|\cdot))$, supposé inclus strictement dans H, dense dans H, et avec inclusion continue au sens suivant : il existe $C > 0$ tel que pour tout $u \in V$, on ait $||u||_H \leqslant C||u||_V$ (on retrouve au passage que si $u \in V$, alors $u \in H$.) Dans toute la suite, pour une injection continue entre deux espaces par exemple V et H, on utilisera la notation $V \hookrightarrow H$.

On suppose que l'on identifie H à H', de telle sorte que tout élément de H' est représenté par un certain $h \in H$. On peut alors regarder l'application

$$i : h \in H \mapsto (v \in V \mapsto \langle v, h \rangle).$$

D'abord, $v \mapsto \langle v, h \rangle$ est trivialement linéaire. Elle est aussi continue pour la topologie de V (qu'elle le soit pour H est évident par l'inégalité de Cauchy-Schwartz), en effet, on a pour tout $v \in V$ que

$$|\langle v, h \rangle| \leqslant ||v||_H ||h||_H \leqslant C||v||_V ||h||_H.$$

Donc $i(h) \in V'$ et $||i(h)||_{V'} \leqslant C||h||_H$. Donc $i : H \to V'$ et est continue.

i est de plus injective. En effet, si $i(h) = 0$, alors pour tout $v \in V$, on a $\langle v, h \rangle = 0$. V étant dense dans H, il existe une suite $(v_n)_{n \in \mathbb{N}^*}$ une suite d'éléments de V qui converge vers h. Pour tout $n \in \mathbb{N}^*$, on a $\langle v_n, h \rangle = 0$. Par continuité du produit scalaire, on en déduit pour $n \to \infty$ que $||h||^2 = 0$ et donc $h = 0$. Ainsi $Ker(i) = \{0\}$ et i est bien injective.

Ainsi, on peut identifier $i(H)$ avec H, et faire "comme si" $H \subset V'$. On a alors la suite d'injections continues

$$V \hookrightarrow H \hookrightarrow V'.$$

On a donc $V \subset H = H' \subset V'$ avec $V \neq H$: il est alors impossible d'identifier simultanément H et H' ainsi que V et V' ! Autrement dit, dans ce cadre, il est raisonnable d'identifier un seul espace de Hilbert avec lui-même, mais pas les autres. La relation $V \hookrightarrow H \hookrightarrow V'$ est appelé triplet de Gelfand. L'espace H est appelé espace pivot dans ce cadre.

Enfin, signalons que l'on peut en fait identifier de manière explicite le dual V', à l'aide de la Proposition suivante, que nous admettrons.

Proposition 5.4. *On suppose que $V \hookrightarrow H$ et V dense dans H. On définit $|| \cdot ||_*$ sur H par*

$$||h||_* = \sup_{v \in V, ||v||_V = 1} |\langle v, h \rangle_H|.$$

C'est une norme sur H et le complété de H pour cette norme est isomorphe à V'.

Exemple 5.5. *Un exemple très éclairant est le suivant. On considère $H = L^2(\mathbb{R})$ muni du produit scalaire canonique et de la norme associée notée $|| \cdot ||_H$, et $\rho : \mathbb{R} \to \mathbb{R}^+$ une fonction continue telle que $\rho(x) \to +\infty$ quand $|x| \to +\infty$, bornée par en*

dessous : il existe un certain $C > 0$ tel que pour tout $x \in \mathbb{R}$, on ait $\rho(x) \geqslant C$. On pose alors

$$V = \{f \in L^2(\mathbb{R}) | \int_{\mathbb{R}} \rho f^2 < +\infty\}.$$

On peut très facilement munir V d'une structure d'espace de Hilbert en introduisant le produit scalaire (c'est ici que sert l'hypothèse de positivité stricte sur ρ)

$$\langle f, g \rangle_V := \int_{\mathbb{R}} \rho f g,$$

et on note $|| \cdot ||_V$ la norme canoniquement associée.

Alors $V \subset H$ avec injection continue, tout simplement car si $f \in V$, on a

$$||f||_V^2 = \int_{\mathbb{R}} \rho f^2 \geqslant C||f||_H^2.$$

De plus, V est strictement inclus dans H. En effet, si on avait $V = H$, on aurait sur l'espace H deux normes $||f||_V$ et $||f||_H$ qui rendent l'espace complet et qui seraient comparables, elles sont donc équivalentes par un théorème célèbre d'analyse fonctionnelle (corollaire du théorème de l'application ouverte). On aurait donc existence de $C' > 0$ tel que pour tout $f \in H$, on ait

$$\int_{\mathbb{R}} \rho f^2 \leqslant C' \int_{\mathbb{R}} f^2.$$

Ceci entre en contradiction avec le fait que $\rho(x) \to +\infty$ quand $|x| \to +\infty$: on prend $A > 0$ suffisamment grand pour que

$$x \geqslant A \Rightarrow \rho(x) \geqslant 2C'.$$

On pose alors $f = \mathbb{1}_{[A,A+1]}$, on devrait alors avoir

$$2C' \leqslant \int_{\mathbb{R}} \rho f^2 \leqslant C' \int_{\mathbb{R}} f^2 = C',$$

ce qui est bien sûr faux.

Enfin, il est facile de voir que V est dense dans H, car V contient les fonctions $\varphi \in C_0^\infty(\mathbb{R}^d)$ (puisque $\rho\varphi^2 \in L^2(\mathbb{R})$ en tant que fonction continue à support inclus dans un compact).

Maintenant, nous cherchons à identifier V'. Le calcul principal est le suivant : si $f \in H$ et $g \in V$, alors

$$\langle f, g \rangle_H = \int_{\mathbb{R}} fg = \int_{\mathbb{R}} \rho \left(\frac{f}{\rho}\right) g = \langle \frac{f}{\rho}, g \rangle_V.$$

Donc par l'inégalité de Cauchy-Schwarz (ainsi que le cas d'égalité), pour $f \in H$, on a clairement que

$$||f||_* = \left|\left|\frac{f}{\rho}\right|\right|_V.$$

Il est alors très facile de voir que

$$V^* = \{f \text{ mesurables} | \int_{\mathbb{R}} \frac{f^2}{\rho} < +\infty\}.$$

On peut donc identifier V' à V^*, ce qui revient à prendre l'inverse du poids ρ.

5.1.2 Fonctions absolument continues et lemme de Gronwall

Avant de rappeler le lemme de Gronwall, nous donnons ici la définition d'une fonction absolument continue à valeurs dans un espace de Banach.

Définition 5.6. *Soit $I \subset \mathbb{R}$ un intervalle de $\mathbb{R}$ et soit X un espace de Banach. On dit qu'une fonction continue $u : I \to X$ est une fonction absolument continue si et seulement si pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour toute suite $(\alpha_n)_{n\in\mathbb{N}}, (\beta_n)_{n\in\mathbb{N}} \subset I$ tels que $(\alpha_n, \beta_n) \cap (\alpha_m, \beta_m) = \emptyset$ pour tout $n \neq m$ et $\sum_{n\in\mathbb{N}} |\beta_n - \alpha_n| \leq \delta$, alors $\sum_{n\in\mathbb{N}} \|u(\beta_n) - u(\alpha_n)\|_X \leq \epsilon$.*

Remarque 5.7. *On admettra les points suivants :*

- *On peut montrer qu'une fonction $f \in L^1_{\text{loc}}(I)$ à valeurs réelles est* absolument continue *sur $I \subset \mathbb{R}$ si sa dérivée au sens des distributions appartient aussi à $L^1_{\text{loc}}(I)$. Les fonctions absolument continues sont uniformément continues et différentiables presque partout.*
- *Si $I = [a, b]$, on admettra que f est absolument continue sur $[a, b]$ si et seulement s'il existe $g \in L^1([a, b])$ tel que pour tout $x \in [a, b]$, on ait*

$$f(x) - f(a) = \int_a^x g(t)dt.$$

Auquel cas, g est la dérivée au sens de distributions de f. Donc l'ensemble des fonctions absolument continues sur $[a, b]$ n'est rien d'autre que $W^{1,1}([a, b])$.

Le lemme suivant est très classique et très utile :

Lemme 5.8 (Gronwall)**.** *Soit η une fonction positive absolument continue sur $[0; T]$ vérifiant :*

$$\eta'(t) \leq \varphi(t)\eta(t) + \psi(t)$$

pour tout $t \in [0; T]$, où φ et ψ sont des fonctions positives de $L^1([0; T])$. Alors

$$\forall t \in [0; T], \qquad \eta(t) \leq e^{\int_0^t \varphi(s)ds} \left(\eta(0) + \int_0^t \psi(s)ds\right).$$

Preuve : On pose

$$g(t) = e^{-\int_0^t \varphi(s)ds}\eta(t).$$

On a que $g \in W^{1,1}(\mathbb{R})$ comme produit de tels fonctions par la Remarque 5.7, et on a donc

$$g'(t) = \eta'(t)e^{-\int_0^t \varphi(s)ds} - \varphi(t)\eta(t)e^{-\int_0^t \varphi(s)ds}.$$

L'hypothèse sur η assure donc que

$$g'(t) \leqslant \psi(t)e^{-\int_0^t \varphi(s)ds} \leqslant \psi(t),$$

puisque $\varphi \geqslant 0$. Donc en intégrant et en utilisant la Remarque 5.7,

$$g(t) \leqslant g(0) + \int_0^t \psi(s)ds,$$

ce qui donne le résultat voulu. ♢

5.1.3 Intégrale de Bochner

Dans cette section, nous introduisons la notion d'intégrale de Bochner, qui permet de généraliser la notion d'intégrale de Lebesgue, à des fonctions à valeurs dans un espace de Banach.

Dans toute cette section, on considère $a, b \in \mathbb{R} \cup \{\pm\infty\}$ et X un espace de Banach.

Définition 5.9. *Une fonction $f : [a, b] \to X$ est dite mesurable si et seulement si pour tout ensemble ouvert $B \subset X$, l'ensemble $f^{-1}(B)$ est un ensemble borélien de $[a, b]$.*

On voit aisément que cette définition est une extension directe de la notion de mesurabilité pour des fonctions à valeurs scalaires (ou à valeurs dans $\mathbb{R}^n$ avec $n \in \mathbb{N}^*$). Le théorème suivant est également une extension directe d'un résultat que vous connaissez bien pour des fonctions à valeurs scalaires.

Théorème 5.10. *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables définies sur $[a, b]$ à valeurs dans X. Si $(f_n)_{n\in\mathbb{N}}$ converge simplement (dans X) vers une fonction $f : [a, b] \to X$, alors f est une fonction mesurable.*

Preuve : La preuve classique dans le cas réel (elle se généraliserait d'ailleurs très simplement au cas où X est un espace vectoriel de dimension finie) est de d'abord démontrer que le sup et l'inf d'une suite de fonctions est mesurable, en déduire en utilisant deux fois les propriétés précédentes que la limsup et la liminf d'une fonction sont aussi mesurables, ce qui permet d'en déduire le résultat pour la limite simple. Ici, il n'est pas possible de procéder ainsi, car la limsup et la liminf n'ont pas de sens si on est dans un espace de Banach X quelconque. Il va falloir donc procéder autrement, de manière directe.

Soit $B \subset X$ ouvert. On veut montrer que $f^{-1}(B)$ est mesurable. Pour tout $m \in \mathbb{N}^*$, on pose

$$F_m = \{y \in B | B\left(y, \frac{1}{m}\right) \subset B\}.$$

F_m est un ensemble fermé de X. En effet, si $y_j \to y$, avec $y_j \in F_m$ et $y \in X$. Soit $v \in B(y, 1/m)$. On pose $r = ||v - y|| < 1/m$. On sait alors que pour j suffisamment grand, on a $||y_j - y|| < 1/m - r$. Donc par inégalité triangulaire,

$$||y_j - v|| \leqslant ||y_j - y|| + ||v - y|| < 1/m - r + r = 1/m.$$

Donc $v \in B(y_j, 1/m)$ et donc $v \in B$. Ainsi, $B(y, v) \subset B$, ce qui implique notamment que son centre $y \in B$ et donc $y \in F_m$ par définition.

Si $f(x) \in B$, alors $f(x) \in F_m$ pour un certain m puisque U est ouvert. Or, il existe un certain $N \in \mathbb{N}$ tel que

$$n \geqslant N \to ||f_n(x) - f(x)|| \leqslant \frac{1}{2m}.$$

On en déduit donc que si $v \in B(f_n(x), 1/2m)$, on a

$$||f(x) - v|| \leqslant ||f_n(x) - f(x)|| + ||f(x) - v|| < \frac{1}{m},$$

donc $v \in U$ (car $f(x) \in F_m$) et donc $f_n(x) \in F_m$, *i.e.* $x \in f_n^{-1}(F_m)$. Ainsi, on a

$$f^{-1}(B) \subset \bigcup_{m\in\mathbb{N}^*} \bigcup_{k\in\mathbb{N}} \bigcap_{n\geqslant k} f_n^{-1}(F_m).$$

Inversement, si $f_n(x) \in F_m$ pour $n \geqslant N$, un raisonnement totalement analogue au précédent assure que $x \in f^{-1}(B)$, et ainsi

$$\bigcup_{m\in\mathbb{N}^*} \bigcup_{k\in\mathbb{N}} \bigcap_{n\geqslant k} f_n^{-1}(F_m) \subset f^{-1}(B).$$

On en déduit donc que

$$f^{-1}(B) = \bigcup_{m\in\mathbb{N}^*} \bigcup_{k\in\mathbb{N}} \bigcap_{n\geqslant k} f_n^{-1}(F_m),$$

et ce dernier ensemble est mesurable par intersection et union d'ensembles mesurables, puisque les F_m sont fermés (donc complémentaires d'ouverts, la propriété de mesurabilité se transfère donc aux images réciproques de fermés). ♢

Comme pour l'intégrale de Lebesgue, nous allons définir l'intégrale de Bochner comme la limite d'intégrales d'une suite de fonctions étagées. Dans notre cas, on appellera une fonction étagée toute fonction $s : [a, b] \to X$ telle qu'il existe $M \in \mathbb{N}^*$, $u_1, \cdots, u_M \in X$ et $B_1, \cdots, B_M$ des sous-ensembles boréliens de $[a, b]$ de mesure de Lebesgue finie tels que

$$\forall t \in [a, b], \quad s(t) = \sum_{m=1}^{M} u_m \chi_{B_m}(t),$$

où $\chi_B : [a,b] \to \{0,1\}$ désigne la fonction caractéristique du sous-ensemble $B \subset [a,b]$.

Pour une telle fonction, on peut définir son intégrale de Bochner comme l'élément de X suivant :

$$\int_{[a,b]} s(t)\, dt = \sum_{m=1}^{M} u_m \lambda(B_m) \in X,$$

où λ désigne la mesure de Lebesgue sur l'intégrale $[a,b]$. Comme dans le cas de l'intégrale de Lebesgue, on peut démontrer que cette définition est indépendante de la manière dont on représente s (il y a une infinité de représentations possibles), et que l'on peut supposer sans perte de généralité que les B_m sont disjoints. On a alors la propriété très simple suivante.

Proposition 5.11. *Si $s : [a,b] \to X$ est une fonction étagée, alors*

$$\left\| \int_{[a,b]} s(t)\, dt \right\|_X \leq \int_{[a,b]} \|s(t)\|_X\, dt.$$

Preuve : Cela repose tout simplement sur l'inégalité triangulaire dans X.

En reprenant les notations précédentes et en supposant les B_m disjoints, on a

$$\begin{aligned} \left\| \int_{[a,b]} s(t)\, dt \right\|_X &= \left\| \sum_{m=1}^{M} u_m \lambda(B_m) \right\|_X \\ &\leqslant \sum_{m=1}^{M} ||u_m||_X \lambda(B_m) \\ &= \int_{[a,b]} \|s(t)\|_X\, dt, \end{aligned}$$

la dernière égalité venant de la définition d'une fonction étagée sur $\mathbb{R}$ et du fait que les B_m sont supposés disjoints, ce qui implique que

$$\|s(t)\|_X = \sum_{m=1}^{M} ||u_m|| \chi_{B_m}(t).$$

♢

Pour définir l'intégrale de Lebesgue, nous utilisions dans le cas scalaire le résultat crucial suivant : toute fonction mesurable (à valeurs scalaires) peut être vue comme la limite simple d'une suite de fonctions étagées. Il se trouve que ce résultat n'est pas toujours valide dans le cas d'espaces de Banach généraux, ce qui justifie la définition suivante :

Définition 5.12. *On dit qu'une fonction $f : [a,b] \to X$ est Lebesgue-mesurable s'il existe une suite de fonctions étagées $(s_n)_{n\in\mathbb{N}}$ qui converge simplement vers f presque partout sur $[a,b]$ (au sens de la mesure de Lebesgue).*

Les notions de mesurabilité et de Lebesgue-mesurabilité ne sont pas équivalentes en général. La Lebesgue-mesurabilité implique la mesurabilité comme l'indique la proposition suivante.

Proposition 5.13. *Soit une fonction $f : [a, b] \to X$ Lebesgue-mesurable. Alors f est mesurable.*

Preuve : Toute fonction Lebesgue-mesurable f est limite simple p.p. de fonctions simples $\{s_n\}_{n\in\mathbb{N}}$. Donc à un ensemble de mesure nulle E près, cette suite de fonctions étagées converge simplement sur $[a, b] \setminus E$. De telles fonctions sont toutes mesurables comme combinaisons linéaires finies de fonctions mesurables (c'est facile à démontrer). On remarque alors que le Théorème 5.10 reste valable si on change l'espace de départ $[a, b]$ en $[a, b] \setminus E$. Donc f est mesurable sur $[a, b] \setminus E$. E étant de mesure nulle, f est bien mesurable sur $[a, b]$. ♢

La réciproque est fausse en général. Elle est cependant vraie dans le cas où l'espace X est un espace *séparable* au sens de la définition suivante.

Définition 5.14. *Un espace de Banach X est dit séparable s'il existe un sous-ensemble dense de X au plus dénombrable.*

Exemple 5.15. *Un espace de Hilbert réel H séparable (i.e. muni d'une base hilbertienne) est un espace de Banach séparable au sens de la Définition 5.14. En effet, si $\{e_k\}_{k\in\mathbb{N}^*}$ est une base hilbertienne de H, on considère E l'ensemble des combinaisons linéaires finies à coefficients rationnels dans la base hilbertienne :*

$$E = \{\sum_{k=1}^{K} a_k e_k | K \in \mathbb{N}^*, a_k \in \mathbb{Q}\}.$$

C'est clairement un ensemble dénombrable comme union dénombrable d'ensembles dénombrables. Il est dense. En effet, soit $\varepsilon > 0$. Soit $x \in H$. Alors x se représente de manière unique sous la forme $x = \sum_{k=1}^{+\infty} x_k e_k$ où $(x_k)_{k\in\mathbb{N}^*} \in l^2(\mathbb{N}^*)$. Notamment, si $K \in \mathbb{N}^*$, on a par l'identité de Parseval que

$$||x - \sum_{k=1}^{K} x_k e_k||^2 = || \sum_{k=K+1}^{+\infty} x_k e_k||^2 = \sum_{k=K+1}^{+\infty} x_k^2 \to 0 \text{ quand } K \to +\infty.$$

Il existe donc un certain $K > 0$ tel que

$$|| \sum_{k=K+1}^{+\infty} x_k||^2 \leqslant \frac{\epsilon^2}{2}.$$

Maintenant, $\mathbb{Q}$ étant dense dans $\mathbb{R}$, pour tout $k \in \{1, \dots K\}$, il existe $a_k \in \mathbb{Q}$ tel que $|a_k - x_k| \leqslant \varepsilon/\sqrt{K}$. On pose alors $y = \sum_{k=1}^{K} a_k e_k$. Par déinition, $y \in E$ et de plus, en utilisant encore Parseval,

$$||x - y||^2 = || \sum_{k=1}^{K} (x_k - a_k e_k) + \sum_{k=K+1}^{+\infty} x_k e_k||^2 = \sum_{k=1}^{K} |x_k - a_k|^2 + \varepsilon^2 \leqslant \sum_{k=1}^{K} \frac{\varepsilon^2}{K} + \varepsilon \leqslant 2\varepsilon^2,$$

ce qui donne le résultat voulu, ε étant arbitraire.

En pratique, la plupart des espaces de Banach que vous connaissez (espaces de Lebesgue L^p, de Sobolev H^k...) sont des espaces séparables (hormis certains espaces L^∞ ou $W^{p,\infty}$ et certains espaces de fonctions). On admettra que sauf si le contraire est spécifié, les espaces que nous rencontrerons dans ce cours sont séparables. Si X est un espace de Banach séparable, on a alors le résultat suivant, que l'on admettra :

Théorème 5.16. *Soit X un espace de Banach séparable. Alors, pour toute fonction $f : [a,b] \to X$ mesurable, il existe une suite $(s_n)_{n\in\mathbb{N}}$ de fonctions étagées définies sur $[a,b]$ à valeurs dans X telle que $(s_n)_{n\in\mathbb{N}}$ converge simplement vers f presque partout (au sens de la mesure de Lebesgue) sur $[a,b]$.*

Autrement dit, si X est un espace de Banach séparable, toute fonction $f : [a,b] \to X$ est mesurable si et seulement si elle est Lebesgue-mesurable.

Pour pouvoir définir l'intégrale de Bochner, nous avons besoin de définir la notion de fonction intégrable dans notre contexte. C'est le but de la définition suivante.

Définition 5.17. *Une fonction $f : [a,b] \to X$ est dite intégrable si et seulement si il existe une suite $(s_n)_{n\in\mathbb{N}}$ de fonctions étagées telles que*

(i) $(s_n)_{n\in\mathbb{N}}$ converge simplement vers f presque partout sur $[a,b]$;

(ii) $\displaystyle\int_{[a,b]} \|f - s_n\|_X \underset{n\to+\infty}{\longrightarrow} 0.$

Le théorème suivant énonce une formulation équivalente de la notion d'intégrabilité, qui vous sera probablement plus familière, et que l'on admettra

Théorème 5.18 (Critère d'intégrabilité de Bochner)**.** *On suppose X séparable. Une fonction $f : [a,b] \to X$ est intégrable si et seulement si*

$$\int_{[a,b]} \|f(t)\|_X \, dt < +\infty.$$

Nous sommes armés à présent pour pouvoir définir l'intégrale de Bochner d'une fonction intégrable.

Théorème-Définition 5.19. *Soit $f : [a,b] \to X$ une fonction intégrable et $(s_n)_{n\in\mathbb{N}}$ une suite de fonctions étagées vérifiant les propriétés (i) et (ii) de la Définition 5.17. Alors, l'intégrale de Bochner de f sur $[a,b]$ est définie par*

$$\int_{[a,b]} f = \lim_{n\to+\infty} \int_{[a,b]} s_n,$$

la limite existant et étant indépendante de la suite de fonctions étagées choisie. On a de plus la propriété suivante :

$$\left\| \int_{[a,b]} f \right\|_X \leq \int_{[a,b]} \|f\|_X . \tag{5.1}$$

Une propriété très importante de l'intégrale de Bochner est donnée dans la proposition suivante, que l'on admettra.

Proposition 5.20. *Soit Y un espace de Banach et $T : X \to Y$ une application linéaire continue. Si $f : [a,b] \to X$ est intégrable, alors $T(f) : [a,b] \to Y$ est intégrable et*

$$T\left(\int_{[a,b]} f\right) = \int_{[a,b]} T(f).$$

On peut montrer que le théorème de convergence dominée est encore valable dans le cas de fonctions à valeurs dans un espace de Banach. Plus précisément, on admettra le résultat suivant :

Théorème 5.21. *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions définies sur $[a,b]$ et à valeurs dans X, et $f : [a,b] \to X$ telles que*

- *$f_n(t) \longrightarrow_{n\to+\infty} f(t)$ dans X pour presque tout $t \in [a,b]$;*
- *Il existe une fonction $g \in L^1([a,b],\mathbb{R})$ telle que pour tout $n \in \mathbb{N}$, $\|f_n(t)\|_X \leq g(t)$ pour presque tout $t \in [a;b]$.*

Alors, f est intégrable et

$$\int_{[a,b]} \|f_n(t) - f(t)\|_X \, dt \longrightarrow 0,$$

ce qui implique que

$$\int_{[a,b]} f_n(t)\, dt \underset{n\to+\infty}{\longrightarrow} \int_{[a,b]} f(t)\, dt \quad \text{dans } X.$$

Remarque 5.22. *Une extension immédiate permet de définir les fonctions $f : [a,b] \times [c,d] \to X$ intégrables sur $[a,b] \times [c,d]$ ainsi que leur intégrale de Bochner. On peut alors aussi montrer facilement que le théorème de Fubini est toujours valide dans ce contexte.*

5.1.4 Espaces dépendant du temps

Dans cette section, nous introduisons plusieurs espaces fonctionnels adaptés à l'étude des équations d'évolution, et qui seront à la base de la définition des *solutions faibles*. Le contenu de ce chapitre peut être trouvé en détail dans [6, Section 5.9.2 et Appendice E.5].

L'idée générale est de séparer la variable temporelle en voyant $u(t,x)$ non pas comme une fonction des deux variables t et x, mais plutôt comme une fonction de t à valeurs dans un espace de fonctions de la variable x :

$$u : t \mapsto \{x \mapsto u(t,x)\}.$$

Soit X un espace de Banach et I un intervalle de $\mathbb{R}$. Nous noterons $C^k(I,X)$, $k \geq 0$, l'espace des fonctions k fois continuement dérivables sur I à valeurs dans

X. De même on peut définir l'espace $L^p(I,X)$ contenant les fonctions $u : I \to X$ (définies presque partout et mesurables en un sens approprié, voir l'appendice E.5 de [6]) telles que la fonction $t \mapsto \|u(t)\|_X$ appartient à l'espace usuel $L^p(I,\mathbb{R})$:

$$\int_I \|u(t)\|_X^p \, dt < \infty.$$

Tous ces espaces sont eux-mêmes des espaces de Banach lorsqu'ils sont munis des normes associées :

$$\|u\|_{\mathcal{C}^k(I,X)} = \sum_{m=0}^{k} \sup_{t\in I} \left\|u^{(m)}(t)\right\|_X ,$$

$$\|u\|_{L^p(I,X)} = \left(\int_I \|u(t)\|_X^p \, dt\right)^{1/p}, \quad 1 \le p < \infty,$$

$$\|u\|_{L^\infty(I,X)} = \sup_{t\in I} \|u(t)\|_X .$$

De façon similaire, on dit que $u \in L^1_{\text{loc}}(I,X)$ si $u \in L^1([a;b],X)$ pour tout $[a;b] \subset I$, $a,b \in \mathbb{R}$. Notons que si X est un espace de Hilbert, alors $L^2(I,X)$ est aussi un espace de Hilbert muni du produit scalaire

$$\langle u,v\rangle_{L^2(I,X)} = \int_I \langle u(t),v(t)\rangle_X dt.$$

Nous utiliserons souvent des espaces du type $L^2(I,H_0^p(\Omega))$ ou $L^2(I,H^{-r}(\Omega))$, où $H^{-r}(\Omega)$ est le dual de $H_0^r(\Omega)$ avec espace pivot $L^2(\Omega)$. Ce sont tous des espaces de Hilbert.

Nous aurons besoin dans la suite de définir la notion de *dérivée faible* temporelle pour des fonctions appartenant à de tels espaces. Cette notion est donnée dans le Théorème-Définition 5.23.

Théorème-Définition 5.23 (Dérivée faible temporelle)**.** *Soit I un intervalle ouvert de $\mathbb{R}$, X un espace de Banach. On dit que $v \in L^1_{\text{loc}}(I,X)$ est la* dérivée faible *de $u \in L^1_{\text{loc}}(I,X)$ (et on note $v = u'$) si et seulement si*

$$\forall \varphi \in \mathcal{C}_c^\infty(I,\mathbb{R}), \quad \int_I \varphi(t)v(t)dt = -\int_I \varphi'(t)u(t)dt \text{ dans } X. \tag{5.2}$$

Comme dans le cas réel, on peut montrer que si $w_1, w_2 \in L^1_{\text{loc}}(I,X)$ vérifient $\int_I w_1(t)\varphi(t)\,dt = \int_I w_2(t)\,\varphi(t)\,dt$ pour toute fonction $\varphi \in \mathcal{C}_c^\infty(I)$, alors nécessairement $w_1(t) = w_2(t)$ pour presque tout $t \in I$, et donc que la dérivée faible de u définie par (5.2) est définie de manière unique.

Remarque 5.24. *La dérivée faible coïncide la dérivée au sens des distributions, mais pour la distribution u à valeurs vectorielles, c'est-à-dire dans l'espace X. Attention, la dérivée au sens des distributions existe toujours, mais pas la dérivée faible car on a une hypothèse d'intégrabilité en plus.*

Nous avons le lemme suivant, qui se montre de manière analogue que dans le cas de fonctions à valeurs scalaires.

Lemme 5.25. *Soit $u \in L^1_{\text{loc}}(I, X)$ telle que $u'(t) = 0$ pour tout $t \in I$. Alors, il existe $u_0 \in X$ tel que $u(t) = u_0$ pour presque tout $t \in I$.*

Preuve : Soit $\eta \in \mathcal{C}^\infty_c(I)$ telle que $\int_I \eta = 1$. Soit $\varphi \in \mathcal{C}^\infty_c(I)$, on suppose que le support de φ est inclus dans un certain segment $[a, b]$. Alors

$$\varphi(t) = A\eta(t) + \psi'(t),$$

où $A = \int_a^b \varphi(t)\,dt$ et $\psi(t) = \int_a^t [\varphi(s) - A\eta(s)]\,ds$ (remarquer que ψ est bien dans $C^\infty_0(I)$) et que son support est en fait même inclus dans $[a, b]$). On a alors

$$\begin{aligned}\int_I u(t)\varphi(t)\,dt &= A\int_I u(t)\eta(t)\,dt + \int_I u(t)\psi'(t)\,dt,\\ &= \left(\int_I \varphi(t)\,dt\right) u_0 - \int_I u'(t)\psi(t)\,dt = u_0\left(\int_I \varphi(t)\,dt\right),\end{aligned}$$

où $u_0 := \int_I \eta(t)u(t)\,dt$. En utilisant des arguments similaires à ceux permettant de démontrer l'unicité d'une solution faible, ceci implique bien que $u(t) = u_0$ pour presque tout $t \in I$. ♢

La dérivée faible possède des propriétés intéressantes qui nous seront utiles par la suite, que nous donnons dans la Proposition 5.26.

Proposition 5.26. *Soit X un espace de Banach. Soit $u \in L^1_{\text{loc}}(]a, b[, X)$ tel que $u' \in L^1_{\text{loc}}(]a, b[, X)$. Alors,*

$$u(t) - u(s) = \int_s^t u'(\tau)\,d\tau, \qquad \textit{pour presque tout } t, s \in I. \tag{5.3}$$

Preuve : Montrons d'abord qu'il existe $u_0 \in X$ tel que $u(t) - \int_s^t u'(\tau)\,d\tau = u_0$ pour presque tout $t \in I$. Notons $v(t) := \int_s^t u'(\tau)\,d\tau$ pour tout $t \in I$. Montrons tout d'abord que

$$v'(t) = u'(t).$$

Soit $\varphi \in \mathcal{C}^\infty_c(I, \mathbb{R})$ et soit $c, d \in]a, b[$ tel que $[c, d] \subset]a, b[$ et $\{s\} \cup \text{Supp}\varphi \subset [c, d]$.

$$\int_{]c,d[} v'(t)\varphi(t)\,dt = -\int_{]a,b[} \varphi'(t)\left(\int_s^t u'(\tau)\,d\tau\right) dt.$$

La fonction $(t,\tau) \in [c,d] \times [c,d] \mapsto \varphi'(t)u'(\tau)$ est une fonction intégrable sur $[c,d] \times [c,d]$. On peut donc appliquer le théorème de Fubini pour obtenir

$$\begin{aligned} -\int_{]c,d[} \varphi'(t) \left(\int_s^t u'(\tau)\, d\tau \right) dt &= -\int_s^d \left(\int_\tau^d \varphi'(t)u'(\tau)\, dt \right) d\tau + \int_c^s \left(\int_c^\tau \varphi'(t)u'(\tau)\, dt \right) d\tau \\ &= \int_s^d \varphi(\tau)u'(\tau)\, d\tau + \int_c^s \varphi(\tau)u'(\tau)\, d\tau \\ &= \int_{[c,d]} \varphi'(\tau)u'(\tau)\, d\tau. \end{aligned}$$

Cette dernière égalité prouve bien que $u'(t) = v'(t)$ pour tout $t \in]a,b[$. Donc il existe $u_0 \in X$ tel que

$$u(t) = u_0 + \int_s^t u'(\tau)\, d\tau. \tag{5.4}$$

Il nous reste à prouver que $u_0 = u(s)$. En utilisant le théorème de convergence dominée, on peut montrer aisément que la fonction v est continue. Ceci implique, en utilisant la formule (5.4) que la fonction u est continue. De plus, toujours en utilisant le théorème de convergence dominée, on montre que $v(t) \underset{t\to s}{\longrightarrow} 0$, ce qui montre que nécessairement $u_0 = u(s)$. ♢

Donnons enfin une dernière propriété, simple à démontrer grâce au théorème de représentation précédent.

Proposition 5.27. *Soit X un espace de Banach. Soit $u \in L^1_{\mathrm{loc}}(]a,b[,X)$ tel que $u' \in L^1_{\mathrm{loc}}(]a,b[,X)$. On a alors*

$$\lim_{h\to 0} \frac{u(t+h)-u(t)}{h} = u'(t) \text{ dans } X, \quad \text{pour presque tout } t \in I, \tag{5.5}$$

Preuve : Par (5.3), on a pour presque tout t, h tels sur $t, t+h \in I$,

$$\frac{u(t+h)-u(t)}{h} = \frac{1}{h}\int_t^{t+h} u'(\tau)\, d\tau.$$

Le théorème de Lebesgue assure que pour presque tout $t \in I$, on a

$$\frac{1}{h}\int_t^{t+h} u'(\tau)\, d\tau \to u'(t),$$

d'où le résultat. ♢

S'il n'est pas très difficile de définir ce qu'est une solution faible pour une équation aux dérivées partielles (linéaire), il n'est en revanche *a priori* pas du tout évident de donner un sens précis aux conditions initiales : que peut bien vouloir

dire $u(t=0)=u_0$ si u n'est définie que presque partout en t ? Voici maintenant un résultat fournissant une meilleure régularité pour u lorsque l'on sait dans quel espace fontionnel vit u', et qui va être crucial dans la suite pour définir correctement les conditions initiales. Ceci est à comparer avec les injections de Sobolev usuelles en dimension un.

On considère un espace de Hilbert H séparable que l'on identifie avec son dual, et un autre espace de Hilbert V tel que $V \hookrightarrow H$ (injection continue), avec V dense dans H. On a donc

$$V \hookrightarrow H \hookrightarrow V'.$$

Théorème 5.28. *Soient $a, b \in \mathbb{R}$. Si $u \in L^2(]a;b[,V)$ est tel que $u' \in L^2(]a;b[,V')$, alors on a :*

1. *$u \in C^0([a;b],H)$;*
2. *$\sup_{t\in[a;b]} \|u(t)\|_H \leq C\left(\|u\|_{L^2(]a;b[,V)} + \|u'\|_{L^2(]a;b[,V')}\right)$ pour une constante C ne dépendant pas de u ;*
3. *Soient $u, v \in L^2(]a;b[,V)$ tels que $u', v' \in L^2(]a;b[,V')$. Alors la fonction $t \mapsto \langle u(t), v(t)\rangle_H$ est absolument continue et on a*

$$\frac{d}{dt}\langle u(t), v(t)\rangle_H = {}_{V'}\langle u'(t), v(t)\rangle_V + {}_{V'}\langle v'(t), u(t)\rangle_V.$$

Nous donnerons la preuve uniquement dans le cas où $V = H = V'$. La preuve dans le cas général est plus longue : nous renvoyons par exemple à [3, Chap. XVIII § 1] pour le cas général ou à [6] pour le cas où $V = H_0^1(\Omega)$ et $H = L^2(\Omega)$.

Preuve : L'idée lorsque $V = H$ est d'utiliser la formule (5.3). D'après la Proposition 5.26, on a

$$u(t) - u(s) = \int_s^t u'(\tau)\, d\tau,$$

pour presque tout $s, t \in]a, b[$.

Notons que l'intégrale du terme à droite a un sens puisque $u' \in L^2(]a;b[,H)$ par hypothèse et que $1_{[s,t]} \in L^2(]a;b[)$. Les points *1.* et *2.* du théorème sont des conséquences faciles de la formule (5.3). En effet, on a par l'inégalité de Cauchy-Schwarz

$$\|u(t) - u(s)\|_H \leq |t-s|^{1/2} \|u'\|_{L^2(]a;b[,H)}$$

qui démontre la continuité (en fait $t \mapsto u(t)$ est même Hölder). En utilisant l'inégalité triangulaire et en intégrant par rapport à s, on trouve aussi

$$(b-a)\|u(t)\|_H \leq (b-a)^{1/2}\|u\|_{L^2(]a;b[,H)} + \frac{4}{3}(b-a)^{3/2}\|u'\|_{L^2(]a;b[,H)},$$

donc

$$\sup_{t\in[a;b]} \|u(t)\|_H \leq (b-a)^{-1/2}\|u\|_{L^2(]a;b[,H)} + \frac{4}{3}(b-a)^{1/2}\|u'\|_{L^2(]a;b[,H)}.$$

On admettra le dernier point. ♢

Remarque 5.29. *Si $u \in L^2(]a;b[,V)$, alors on a également $u \in L^2(]a;b[,V')$ car $V \hookrightarrow V'$. L'hypothèse du théorème signifie simplement que u est différentiable dans V' (au sens de la définition 5.23 avec $X = V'$), et que sa dérivée u' appartient en plus à $L^2(]a;b[,V')$.*

Remarque 5.30. *Introduisons l'espace fonctionnel (de Banach)*

$$W(]a;b[,V,V') := \{u \in L^2(]a;b[,V) \mid u' \in L^2(]a;b[,V')\}.$$

Alors les points 1. *et* 2. *du Théorème 5.28 signifient que l'on a une injection continue*

$$W(]a;b[,V,V') \hookrightarrow C^0([a;b],H)$$

où $C^0([a;b],H)$ est muni de la norme uniforme $\|\cdot\|_{L^\infty([a;b],H)}$.

Ceci est très important car cela permet en particulier de donner un sens à $u(a)$ et $u(b)$ dans H, donc aux conditions aux limites.

Remarque 5.31. *En prenant $u = v$ dans* 3., *on trouve que*

$$\frac{1}{2}\frac{d}{dt}\|u(t)\|_H^2 = {}_{V'}\langle u'(t), u(t)\rangle_V.$$

De même on trouve que si $v \in V$, alors $t \mapsto \langle u(t), v\rangle_V$ est absolument continue et on a

$${}_{V'}\langle u'(t), v\rangle_V = \frac{d}{dt}\langle u(t), v\rangle_H.$$

Remarque 5.32. *Dans la pratique, nous utiliserons souvent le théorème précédent avec*

$$V = H_0^1(\Omega) \subset H = L^2(\Omega) \subset V' := H^{-1}(\Omega),$$

où Ω est un ouvert borné de $\mathbb{R}^n$ (ou $\Omega = \mathbb{R}^n$, auquel cas $H_0^1(\mathbb{R}^n)$ s'identifie avec $H^1(\mathbb{R}^n)$). Le choix de $V = H_0^1(\Omega)$ correspond aux conditions au bord de Dirichlet.

On obtient donc que si $u \in L^2(]0;T[,H_0^1(\Omega))$ est tel que $u' \in L^2(]0;T[,H^{-1}(\Omega))$, alors $u \in C^0([0;T],L^2(\Omega))$, donc $u(0)$ et $u(T)$ ont un sens dans $L^2(\Omega)$.

Si $u \in L^p(I,H^k(\Omega))$ pour un ouvert régulier $\Omega \subset \mathbb{R}^3$ avec $p \geq 1$ et $k \geq 1$, alors on peut évidemment définir ∇u par

$$\nabla u : t \mapsto \{x \mapsto \nabla_x u(t,x)\}.$$

Bien sûr dans ce cas $\nabla u \in L^p(I,H^{k-1}(\Omega))$.

Nous étudions dans le reste de ce chapitre avec plus de détails l'équation de la chaleur, qui est l'exemple prototype par excellence d'une équation parabolique.

Nous commencerons par le cas simple de tout l'espace avant de traiter celui d'un domaine borné. Nous n'aborderons pas le cas d'un domaine non borné différent de $\mathbb{R}^n$ dont l'approche classique est basée sur des considérations plus compliquées (théorie des semi-groupes).

5.2 L'équation de la chaleur dans tout l'espace

Commençons par chercher une solution particulière $G(t,x)$ régulière (pour $t>0$) de l'équation de la chaleur

$$\frac{\partial}{\partial t}G - \Delta G = 0.$$

Pour cela, on utilise la tranformée de Fourier définie par

$$\mathcal{F}(f)(k) = \widehat{f}(k) = (2\pi)^{-n/2}\int_{\mathbb{R}^n} f(x)e^{-ik\cdot x}dx.$$

On trouve donc que $\widehat{G}$ doit résoudre l'équation suivante (on notera $||\cdot||$ la norme euclidienne

$$\frac{\partial}{\partial t}\widehat{G}(t,k) + ||k||^2\widehat{G}(t,k) = 0. \tag{5.6}$$

On peut alors par exemple prendre

$$\widehat{G}(t,k) = Ce^{-t||k||^2}.$$

Remarquons que G n'est bien définie que lorsque $t>0$. Si $t=0$, $\widehat{G}=C$ donc G est égal à une constante multipliée par la distribution de dirac δ_0 (dont la transformée de Fourier est $\frac{1}{(2\pi)^{n/2}}$). Si $t<0$, $\widehat{G}$ n'est pas dans $\mathcal{S}'$ et on ne peut pas définir G. On voit donc apparaître dès maintenant une propriété importante de l'équation de la chaleur : la *non-réversibilité*. La solution ne sera définie que pour les temps futurs, c'est-à-dire $t\geq 0$ si la condition initiale est donnée en $t=0$.

On rappelle que pour une Gaussienne centrée réduite définie sur $\mathbb{R}$, sa transformée de Fourier est

$$\mathcal{F}\left(e^{-|\cdot|^2}\right)(\xi) = \frac{1}{\sqrt{2}}e^{-\frac{|\xi|^2}{4}}.$$

En tensorisant ceci pour avoir une intégrale sur $\mathbb{R}^n$ et en utilisant les propriétés habituelles de dilatation pour la transformée de Fourier, on obtient

$$\mathcal{F}\left(e^{-t||\cdot||^2}\right)(k) = \frac{1}{(2t)^{n/2}}e^{-\frac{||k||^2}{4t}}.$$

En revenant donc dans les variables d'espace et choisissant $C=\frac{1}{(2\pi)^{n/2}}$ (la transformée de Fourier et la transformée de Fourier inverse coïncident pour de telles gaussiennes, qui sont des fonctions réelles paires), on obtient donc une solution de l'équation de la chaleur :

$$G(t,x) = (4\pi t)^{-n/2}e^{\frac{-|x|^2}{4t}} > 0.$$

De plus, on a

$$\forall t>0, \qquad \int G(t,x)dx = 1,$$

et donc $G(t,\cdot) \in L^1(\mathbb{R}^n)$. En effet, pour $t>0$, on a par le théorème de Fubini pour les fonctions positives que

$$\begin{aligned}\int_{\mathbb{R}^n} G(t,x)dx &= (4\pi t)^{-n/2}\int_{\mathbb{R}^n} e^{\frac{-|x|^2}{4t}}dx \\ &= (4\pi t)^{-n/2}\int_{\mathbb{R}^n} e^{\frac{-|x_1|^2+|x_2|^2\ldots+|x_n|^2}{4t}}dx_1\ldots dx_n \\ &= (4\pi t)^{-n/2}\int_{\mathbb{R}^n}\prod_{j=1}^n e^{\frac{-|x_j|^2}{4t}}dx_1\ldots dx_n \\ &= \prod_{j=1}^n \frac{1}{\sqrt{4\pi t}}\int_{\mathbb{R}^n} e^{\frac{-|x_j|^2}{4t}}dx_j.\end{aligned}$$

Le changement de variable $x_j=\sqrt{4t}y$ donne le résultat voulu, étant donné la valeur de l'intégrale de Gauss $\int_{\mathbb{R}} e^{-y^2}dy=\sqrt{\pi}$.

Comme on a $G(t,\cdot)\to\delta_0$ (par un résultat classique sur les approximations de l'unité) au sens des distributions, on dit que G est la solution fondamentale (ou le noyau de Green) de l'équation de la chaleur, c'est-à-dire formellement celle de

$$\left\{\begin{array}{ll}\dfrac{\partial}{\partial t}G-\Delta G &= 0, \quad t>0\\ G(0) &= \delta_0.\end{array}\right. \tag{5.7}$$

On peut maintenant utiliser la fonction G pour construire une solution de l'équation de la chaleur avec une condition initiale différente. En effet, ici, on a résolu l'équation (5.7). Supposons maintenant que l'on souhaite résoudre l'équation

$$\left\{\begin{array}{l}\dfrac{\partial}{\partial t}u-\Delta u=0, \quad t>0\\ u(0)=g,\end{array}\right. \tag{5.8}$$

De manière formelle, comme δ_0 vérifie l'égalité fonctionnelle $f*\delta_0=f$, en passant en Fourier, ceci revient à résoudre la famille d'EDOs

$$\begin{aligned}\frac{\partial}{\partial t}\widehat{G}(t,k)+|k|^2\widehat{G}(t,k)&=0,\\ \widehat{G}(k)&=\widehat{g}(k).\end{aligned} \tag{5.9}$$

On remarque que (5.9) est vérifiée si on multiplie $\widehat{G}$ par une fonction ne dépendant que de k (à savoir $\widehat{g}(k)$), ce qui est une manifestation puisque l'identité $g*\delta_0=g$ dans l'espace de départ devient en Fourier à $\widehat{g}(k)\widehat{\delta_0}(k)=\widehat{g}(k)\frac{1}{(2\pi)^{n/2}}$ et que l'on a résolu l'équation en Fourier avec comme condition initiale $\widehat{\delta_0}(k)=\frac{1}{(2\pi)^{n/2}}$. On introduit donc pour $x\in\mathbb{R}^n$ et $t>0$

$$u(t,x)=(G(t,\cdot)*g)(x)=\int_{\mathbb{R}^n}G(t,x-y)g(y)dy=(4\pi t)^{-n/2}\int_{\mathbb{R}^n}e^{\frac{-|x-y|^2}{4t}}g(y)dy, \tag{5.10}$$

que l'on prolonge par convention en $t = 0$ par $u(0, x) = u^0(x)$, ce qui est raisonnable au vu de la discussion précédente. Comme $G(t, \cdot) \in L^1(\mathbb{R}^n)$ pour tout $t > 0$, on déduit que si $g \in L^p(\mathbb{R}^n)$ pour un certain $p \in [1, +\infty]$, alors $u(t) \in L^p(\mathbb{R}^n)$ pour tout $t > 0$ (il s'agit d'un résultat classique sur le produit de convolution). Ici, nous allons nous concentrer sur le cas $L^2(\mathbb{R}^n)$ (ou éventuellement des sous-espaces de cet espace), pour des raisons de simplicité. C'est ce qui justifie l'introduction du noyau de Green : celui-ci permet par convolution de résoudre notre problème de Cauchy (5.8). Commençons par regarder quelques propriétés d'une solution donnée par la formule (5.10).

Proposition 5.33. *Si $g \in L^2(\mathbb{R}^n)$, la fonction u fournie par la formule* (5.10) *est dans $C^\infty((0; \infty) \times \mathbb{R}^n)$.*

Ainsi, bien que nous ayons seulement supposé $g \in L^2(\mathbb{R}^n)$, on obtient que la fonction $u(t, x)$ fournie par (5.10) est de classe $\mathcal{C}^\infty$ par rapport à x pour tout $t > 0$. On dit que l'équation de la chaleur a un *effet régularisant*.

Preuve : Il s'agit juste de remarquer que G est de classe C^∞ sur $(0; \infty) \times \mathbb{R}^n$ pour tout $\delta > 0$, puis que toutes les dérivées partielles sont continues et intégrables en espace sur $\mathbb{R}^n$ par croissances comparées (au pire des cas, on a sorti des polynômes en $||x||^2$ quand on dérive, qui sont absorbés par la Gaussienne). On peut donc appliquer les résultats classiques de régularité d'intégrales dépendant d'un paramètre, et en déduire notamment que toutes les dérivées partielles de u existent. De plus, si $i, j \in \mathbb{N}$, on a que l'on peut dériver à l'intérieur du signe somme :

$$\partial_t^i \partial_x^j u(t, x) = \int_{\mathbb{R}^n} \partial_t^i \partial_x^j G(t, y) g(x - y) dy. \tag{5.11}$$

En appliquant encore une fois les théorèmes de continuité sous le signe somme (qui s'appliquent pour les mêmes raisons que précédemment), on a donc que conjointement en la variable (t, x) pour tout $\delta > 0$, $(t, x) \in [\delta, +\infty) \times \mathbb{R}^n \mapsto \partial_t^i \partial_x^j u(t, x)$ est continue. Ceci étant vrai pour tout i, j et tout $\delta > 0$, on a bien que u est de classe C^∞ sur $\mathbb{R}^{+*} \times \mathbb{R}^n$. ♢

Corollaire 5.34. *u donnée par* (5.10) *vérifie*

$$\frac{\partial}{\partial t} u - \Delta u = 0$$

sur $\mathbb{R}^{+*} \times \mathbb{R}^n$.

Preuve : C'est une simple conséquence de (5.11) : on sait que G vérifie (5.7), donc

$$\frac{\partial}{\partial t} u - \Delta u = \int_{\mathbb{R}^n} (\frac{\partial}{\partial t} G(t, x) - \Delta G(t, x)) g(x - y) dy = 0.$$

♢

Il reste donc à comprendre en quel sens la condition initiale dans (5.8) est vérifiée, et pourquoi la solution est unique. C'est entre autres ce que dit le théorème suivant.

Théorème 5.35 (Solution de l'équation de la chaleur dans $\mathbb{R}^n$)**.** *Soit $g \in L^2(\mathbb{R}^n)$. Le problème* (5.8) *a une solution unique $u \in C^0([0;\infty), L^2(\mathbb{R}^n)) \cap C^1((0;\infty), H^2(\mathbb{R}^n))$, donnée par la formule* (5.10)*.*

Preuve : Montrons d'abord que la définition (5.10) fournit une solution $u \in C^0([0;\infty), L^2(\mathbb{R}^n))$.

D'abord, pour tout $t > 0$, il est clair que $u(t,\cdot) \in L^2(\mathbb{R}^n)$, puisque $G(t,\cdot) \in L^1(\mathbb{R}^n)$ et $g \in L^2(\mathbb{R}^n)$. On remarque que, par le théorème de Plancherel et la linéarité de la transformée de Fourier, pour $t, t' > 0$, on a par (5.9) que

$$||u(t,\cdot) - u(t',\cdot)||^2_{L^2(\mathbb{R}^n)} = ||\widehat{u}(t,\cdot) - \widehat{u}(t',\cdot)||^2_{L^2(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |e^{-t||k|^2} - e^{-t'||k||^2}|^2|\, |g(k)|^2 dk.$$

Il est alors très simple de voir qu'une application du théorème de convergence dominée assure que pour si $t > 0$ et $t' \to t$, alors

$$||u(t,\cdot) - u(t',\cdot)||^2_{L^2(\mathbb{R}^n)} \to 0,$$

ce qui donne bien la continuité pour $t > 0$ et donc $u \in C^0((0;\infty), L^2(\mathbb{R}^n))$. Reste à prolonger par continuité en $t = 0$. Ceci est une conséquence du fait que si $t_k \to 0$, alors $G(t_k,\cdot)$ est une approximation de l'unité pour la convolution (*c.f.* cours de première année), auquel cas on sait effectivement que

$$u(t_k,\cdot) = G(t_k,\cdot) * g \to g \text{ dans } L^2(\mathbb{R}^n) \text{ quand } k \to +\infty.$$

On a donc bien que $u \in C^0([0;\infty), L^2(\mathbb{R}^n))$ et notamment $u(t,\cdot) \to g$ dans $L^2(\mathbb{R}^n)$ quand $t \to 0^+$.

Reste à montrer que $u \in C^1((0;\infty), H^2(\mathbb{R}^n))$. On rappelle que l'espace $H^2(\mathbb{R}^n)$ peut être décrit par

$$\{u \in L^2(\mathbb{R}^n) | \int_{\mathbb{R}^n} (1 + ||k||^2)^2 \widehat{u}^2(k) dk < +\infty$$

et qu'on peut alors poser comme norme

$$||u||^2_{H^2(\mathbb{R}^n)} = \int_{\mathbb{R}^n} (1 + ||k||^2)^2 \widehat{u}^2(k) dk.$$

Montrons pour commencer que $u \in C^0((0;\infty), H^2(\mathbb{R}^n))$. Ceci repose sur le même calcul que précédemment. Par la formule de Parseval :

$$||u(t,\cdot))||^2_{H^2(\mathbb{R}^n)} = \int_{\mathbb{R}^n} (1 + ||k||^2)^2 \, |e^{-t||k|^2}|^2 \, |g(k)|^2 dk < +\infty,$$

par croissances comparées, et le même calcul que pour le cas $L^2(\mathbb{R}^n)$ avec $t > 0$ donne bien par convergence dominée que $u \in C^0((0;\infty), H^2(\mathbb{R}^n))$. De plus, $\partial_t u = \Delta u$, donc quitte à remplacer u par Δu dans le calcul précédent, on a aussi que

$$||\Delta u(t,\cdot))||^2_{H^2(\mathbb{R}^n)} < +\infty,$$

ainsi que la continuité par rapport à t pour $t > 0$ grâce encore une fois au théorème de convergence dominée et des arguments de croissance comparées.

Si maintenant $v \in C^0([0;\infty), L^2(\mathbb{R}^n)) \cap C^1((0;\infty), H^2(\mathbb{R}^n))$ résout (5.8) avec $g \equiv 0$, on peut prendre le produit scalaire avec la fonction $x \mapsto v(t,x) \in L^2(\mathbb{R}^n)$ et on intègre sur $[0;t_0]$ avec $t_0 > 0$ quelconque. On obtient

$$\|v(t_0,\cdot)\|^2_{L^2(\mathbb{R}^n)} + \int_0^{t_0} dt \int_{\mathbb{R}^n} dx |\nabla v(t,x)|^2 = \frac{1}{2}\|v(0,\cdot)\|^2_{L^2(\mathbb{R}^n)} = 0,$$

donc $v \equiv 0$. Ceci démontre l'unicité, par linéarité de l'équation (si u et v vérifient (5.8) avec la même condition initiale g, alors $u - v$ vérifie (5.8) avec la condition initiale 0 et on est ramené au cas juste étudié). ♢

Remarque 5.36. *Il est absolument crucial ici de démontrer l'unicité dans la classe $L^2(\mathbb{R}^n)$. En effet, si on se place dans des classes plus larges de solutions (qui explosent à une certaine vitesse exponentielle à l'infini), on peut construire des solutions à support compact en temps de l'équation de la chaleur non triviales et valant 0 en $t = 0$.*

De même, notons que si $g \geq 0$ alors $u(t,x) > 0$ pour tout $x \in \mathbb{R}^n$ et $t > 0$, puisque $G > 0$. Même si g est non nulle et positive, à support compact, la solution sera strictement positive sur tout l'espace quand $t > 0$. On parle de *propagation à vitesse infinie*.

Ces propriétés de l'équation de la chaleur sont très spécifiques aux équations de type parabolique et ne seront plus vraies pour l'équation des ondes, par exemple. Démontrer ces propriétés dans le cas d'un ouvert borné nous prendra un peu plus de temps mais tout restera vrai.

Donnons maintenant trois propriétés très simples à démontrer dans ce cas, mais qui seraient vérifiées dans des situations beaucoup plus générales.

Proposition 5.37 (Conservation de la positivité)**.** *Si $g \in L^2(\mathbb{R}^n)$ et $g \geqslant 0$ p.p. et non identiquement nulle, alors $g(t,x) > 0$ pour tout $t > 0$ et tout $x > 0$.*

Preuve : C'est évident sur l'expression

$$u(t,x) = \int_{\mathbb{R}^n} G(t,x-y)g(y)dy = (4\pi t)^{-n/2} \int_{\mathbb{R}^n} e^{\frac{-|x-y|^2}{4t}} g(y)dy.$$

Comme $G \geqslant 0$ et $g \geqslant 0$, on a $u(t,x) \geqslant 0$ et $u(t,x) = 0$ signifie que pour tout x, y, t, on ait

$$e^{\frac{-|x-y|^2}{4t}} g(y) = 0,$$

ce qui est impossible car $e^{\frac{-|x-y|^2}{4t}} > 0$ et g est non identiquement nulle. ♢

Proposition 5.38 (Principe du maximum)**.** *On suppose que $g \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$. Alors u est dans $L^\infty((0;\infty), L^\infty(\mathbb{R}^n))$ et*

$$\sup_{t>0} \|u(t)\|_{L^\infty(\mathbb{R}^n)} \leq \|g\|_{L^\infty(\mathbb{R}^n)}.$$

Preuve : C'est évident, compte tenu de la formule (5.10) et du fait que $\int_{\mathbb{R}^n} G(t,\cdot) = 1$, pour tout $t > 0$. ♢

Proposition 5.39 (Comportement asymptotique)**.** *Si $g \in L^2(\mathbb{R}^n)$, on a*

$$\forall t > 0, \qquad \lim_{|x|\to\infty} u(t,x) = 0,$$

$$\forall x \in \mathbb{R}^n, \qquad \lim_{t\to\infty} u(t,x) = 0,$$

et

$$\lim_{t\to\infty} ||u(t,\cdot)||_{L^2(\mathbb{R}^n)} = 0.$$

Remarque 5.40. *La dernière propriété se traduit par le fait que l'équation de la chaleur est* dissipative *(l'énergie du système est dissipée au cours du temps).*

Preuve : Ce sont des conséquences immédiates du théorème de convergence dominée (pour la norme L^2, penser à appliquer l'identité de Parseval). ♢

5.3 L'équation de la chaleur sur un ouvert borné Ω

Pour étudier l'équation de la chaleur sur un ouvert borné, on ne peut utiliser la transformée de Fourier comme nous l'avons fait dans tout l'espace.

Considérons un ouvert borné régulier $\Omega \subset \mathbb{R}^n$ et un réel $T > 0$. On désire résoudre

$$\begin{cases} \dfrac{\partial}{\partial t}u - \Delta u = f, & \text{dans } (0;T)\times\Omega \\ u_{|(0;T)\times\partial\Omega} = 0 \quad \text{(conditions au bord de Dirichlet)} \\ u(0,x) = g(x). \end{cases} \tag{5.12}$$

5.3.1 Théorème d'existence de solutions faibles

On considère $g \in L^2(\Omega)$ et $f \in L^2(]0;T[, L^2(\Omega))$.

Définition 5.41 (Solutions faibles)**.** *Soit $u \in L^2(]0;T[, H_0^1(\Omega))$ telle que $u' \in L^2(]0;T[, H^{-1}(\Omega))$. On dit que u est une* solution faible *de (5.12) si on a*

(C1) ${}_{H^{-1}(\Omega)}\langle u', v\rangle_{H_0^1(\Omega)} + \displaystyle\int_\Omega \nabla u(t)\cdot\nabla v = \int_\Omega f(t)v$ *pour tout $v \in H_0^1(\Omega)$ et presque partout en $t \in]0;T[$.*

(C2) $u(0) = g$.

Rappelons que d'après le Théorème 5.28, on a $u \in C^0([0;T], L^2(\Omega))$ qui permet de donner un sens à $(C2)$. Rappelons aussi que pour tout $v \in H_0^1(\Omega)$,

$$\frac{d}{dt}\langle u, v\rangle_{L^2(\Omega)} = {}_{H^{-1}(\Omega)}\langle u', v\rangle_{H_0^1(\Omega)}.$$

Dans $(C1)$, la fonction v ne dépend pas du temps.

Le but de cette section est principalement de démontrer le résultat suivant :

Théorème 5.42 (Existence et unicité de solutions faibles)**.** *On suppose que $g \in L^2(\Omega)$ et $f \in L^2(]0;T[, L^2(\Omega))$. Alors le problème* (5.12) *admet une unique solution faible u, vérifiant de plus*

$$\max_{0\le t\le T} \|u(t)\|_{L^2(\Omega)} + \|u\|_{L^2(]0;T[,H_0^1(\Omega))} + \|u'\|\, L^2(]0;T[, H^{-1}(\Omega)) \\ \le C\left(\|f\|_{L^2(]0;T[,L^2(\Omega))} + \|g\|_{L^2(\Omega)}\right). \quad (5.13)$$

Remarque 5.43. *Il est clair que les solutions dépendent linéairement de g et f : si u_1 et u_2 sont des solutions faibles associées aux problèmes avec respectivement (f_1, g_1) et (f_2, g_2), alors $u_1 + u_2$ est solution du problème associé au couple $(f_1 + f_2, g_1 + g_2)$. On parle de* principe de superposition*. Une fois que nous aurons démontré l'unicité de la solution, on obtient donc une application linéaire qui à tout (f, g) associe la solution u. Alors la formule* (5.14) *signifie que cette application linéaire est* continue *dans les bons espaces fonctionnels.*

Il existe deux méthodes de preuve de ce théorème, que nous allons voir dans ce cours. Une première méthode, dite par approximation de Galerkin, est présentée par la suite. Cette méthode est utile car elle est à l'origine de la méthode d'approximation numérique la plus couramment utilisée pour discrétiser ce type d'équations paraboliques, et elle pourrait de plus se généraliser à des classes d'équations paraboliques plus générales. La deuxième méthode utilise la décomposition de la solution sur les fonctions propres de l'opérateur Laplacien. Cette dernière permet de prouver très facilement des propriétés qualitatives fines sur le comportement de la solution, qui ne pourraient pas être facilement accessibles via une méthode d'approximation de Galerkin. Pour cette raison, nous vous présentons ces deux approches dans le détail dans le cadre de ce cours.

Preuve : On utilise la méthode des approximations successives par des espaces de dimension finie (méthode de Galerkin).

Étape 1 : Approximations de Galerkin.

Considérons une famille de fonctions $(w_k)_{k\ge1} \subset H_0^1(\Omega)$, telle que
— $(w_k)_{k\ge1}$ est une base *orthogonale* de $H_0^1(\Omega)$;
— $(w_k)_{k\ge1}$ est une base *orthonormée* de $L^2(\Omega)$.

Par exemple, on peut prendre les fonctions propres du Laplacien avec conditions de Dirichlet au bord de Ω, qui vérifient $-\Delta w_k = \lambda_k w_k$ où $\mathrm{Sp}_{H_0^1(\Omega)}(-\Delta) = \{\lambda_k\}$ (voir le théorème 3.3), mais ce n'est pas indispensable à ce stade.

On pose alors

$$V_m := \mathrm{Vect}(w_1, ..., w_m)$$

et on cherche une solution $u_m \in C^1([0;T], V_m)$ faible dans V_m, c'est-à-dire vérifiant (en décomposant sur la base donnée des V_m)

$(i)_m$ $\langle u_m', w_k\rangle_{L^2} + \displaystyle\int_\Omega \nabla u_m(t)\cdot\nabla w_k = \langle f(t), w_k\rangle_{L^2}$ pour tout $k = 1...m$ et presque partout en $t \in [0;T]$;

$(ii)_m$ $\langle u_m(0), w_k\rangle = \langle g, w_k\rangle$ pour tout $k = 1...m$.

Si on écrit

$$u_m(t,x) = \sum_{k=1}^m d_k^m(t) w_k(x),$$

alors $(i)_m$ et $(ii)_m$ équivalent à

$(i)'_m$ $\dfrac{d}{dt} d_k^m(t) + d_k^m(t)\,\|\nabla w_k\|_{L^2}^2 = \langle f(t), w_k\rangle_{L^2}$, $\forall k = 1...m$, p.p. $t \in [0;T]$;

$(ii)'_m$ $d_k^m(0) = \langle g, w_k\rangle$, $\forall k = 1...m$.

Il s'agit d'un système (diagonal) d'équations différentielles ordinaires qui admet une unique solution absolument continue $(d_k^m(t))_{k=1}^m$, définie sur tout $[0;T]$ et donnée par une certain formule de Duhamel.

Étape 2 : estimées d'énergie.

On désire maintenant passer à la limite quand $m \to \infty$. Pour cela, nous commençons par démontrer le lemme suivant.

Lemme 5.44 (Estimées d'énergie)**.** *Il existe une constante C qui ne dépend que de Ω et $T > 0$ telle que pour tout $m \geq 1$,*

$$\max_{0\leq t\leq T} \|u_m(t)\|_{L^2(\Omega)} + \|u_m\|_{L^2(]0;T[,H_0^1(\Omega))} + \|u_m'\|_{L^2(]0;T[,H^{-1}(\Omega))} \leq C\left(\|f\|_{L^2(]0;T[,L^2(\Omega))} + \|g\|_{L^2(\Omega)}\right). \quad (5.14)$$

Remarque 5.45. *Il est clair que les solutions dépendent linéairement de g et f : si u_1 et u_2 sont des solutions faibles associées aux problèmes avec respectivement (f_1, g_1) et (f_2, g_2), alors $u_1 + u_2$ est solution du problème associé au couple $(f_1 + f_2, g_1 + g_2)$. On parle de* principe de superposition*. Une fois que nous aurons démontré l'unicité de la solution, on obtient donc une application linéaire qui à tout (f, g) associe la solution u. Alors la formule* (5.14) *ave $m \to +\infty$ signifie que cette application linéaire est* continue *dans les bons espaces fonctionnels.*

Preuve : (du Lemme 5.44). En effectuant des combinaisons linaires dépendant éventuellement du temps, on peut prendre $w = u_m$ dans $(i)_m$. On obtient :

$$\langle u'_m, u_m \rangle_{L^2} + \int_\Omega |\nabla u_m(t)|^2 = \langle f(t), u_m(t) \rangle_{L^2}. \tag{5.15}$$

On a

$$\langle f(t), u_m(t) \rangle_{L^2} \leq \frac{1}{2} \left(\|f(t)\|^2_{L^2(\Omega)} + \|u_m(t)\|^2_{L^2(\Omega)} \right)$$

donc,

$$\langle u'_m, u_m \rangle_{L^2} + \int_\Omega |\nabla u_m(t)|^2 \leq \frac{1}{2} \left(\|f(t)\|^2_{L^2(\Omega)} + \|u_m(t)\|^2_{L^2(\Omega)} \right). \tag{5.16}$$

En posant $\eta(t) = \|u_m(t)\|^2_{L^2(\Omega)}$, on obtient

$$\frac{1}{2}\eta'(t) \leq \frac{1}{2}\eta(t) + \frac{1}{2} \|f(t)\|^2_{L^2(\Omega)}.$$

D'après le lemme de Gronwall, on déduit

$$\begin{aligned} \|u_m(t)\|^2_{L^2(\Omega)} &\leq e^t \left(\|g\|^2_{L^2(\Omega)} + \int_0^t \|f(s)\|^2_{L^2(\Omega)} \, ds \right) \\ &\leq e^T \left(\|g\|^2_{L^2(\Omega)} + \|f\|^2_{L^2(]0;T[,L^2(\Omega))} \right). \end{aligned}$$

Ceci fournit bien l'estimée sur $\max_{0\leq t\leq T} \|u_m(t)\|_{L^2(\Omega)}$. Intégrons maintenant l'inégalité (5.16) sur $[0;T]$. Nous obtenons

$$\frac{1}{2} \|u_m(T)\|^2_{L^2(\Omega)} + \|u_m\|^2_{L^2(]0;T[,H^1_0(\Omega))} \leq \frac{1}{2} \|g\|^2_{L^2(\Omega)} + \frac{1}{2} \|f\|^2_{L^2(]0;T[,L^2(\Omega))} + \frac{T}{2} \max_{0\leq t\leq T} \|u_m(t)\|^2_{L^2(\Omega)}.$$

On déduit alors l'estimée sur $\|u_m\|_{L^2(]0;T[,H^1_0(\Omega))}$ en utilisant ce que nous avons déjà démontré pour majorer le dernier terme ci-dessus.

On estime maintenant $\|u'_m\|_{L^2(]0;T[,H^{-1}(\Omega))}$ par dualité. On considère une fonction fixée $v \in H^1_0(\Omega)$, telle que $\|v\|_{H^1_0(\Omega)} \leq 1$. On peut alors écrire $v = v^1 + v^2$ avec $v^1 \in V_m$ et $v^2 \in V_m^\perp = \text{Vect}(w_k,\ k \geq m+1)$. Bien sûr $\|v^1\|_{H^1_0(\Omega)} \leq 1$. On a alors

$$\begin{aligned} {}_{H^{-1}(\Omega)}\langle u'_m, v \rangle_{H^1_0(\Omega)} &= {}_{L^2(\Omega)}\langle u'_m, v^1 \rangle_{L^2(\Omega)} \\ &= \langle f(t), v^1 \rangle_{L^2} - \int_\Omega \nabla u_m(t) \cdot \nabla v^1. \end{aligned}$$

En appliquant l'inégalité de Cauchy-Schwarz et l'inégalité de Poincaré, on a donc que pour une certaine constante $C > 0$,

$$\|u'_m\|_{H^{-1}(\Omega)} \leq C \left(\|f(t)\|_{L^2(\Omega)} + \|\nabla u_m\|_{L^2(\Omega)} \right).$$

On obtient alors l'estimée voulue en passant au carré, en intégrant sur $[0;T]$ et en utilisant les résultats précédents. ♢

Étape 3 : existence.

Nous pouvons maintenant démontrer l'existence d'au moins une solution en passant à la limite faible.

Comme u_m et u'_m sont des suites bornées, respectivement dans les espaces de Hilbert $L^2(]0;T[,H^1_0(\Omega))$ et $L^2(]0;T[,H^{-1}(\Omega))$, on peut extraire des sous-suites $u_{\varphi(m)}$ et $u'_{\varphi(m)}$ telles que $u_{\varphi(m)} \rightharpoonup u$ dans $L^2(]0;T[,H^1_0(\Omega))$ et $u'_{\varphi(m)} \rightharpoonup v$ dans $L^2(]0;T[,H^{-1}(\Omega))$. Il est facile de voir que $v = u'$, donc que $u \in C^0([0;T],L^2(\Omega))$.

Soit maintenant une fonction test de la forme

$$v(t) = \sum_{k=1}^{M} d_k(t) w_k, \tag{5.17}$$

avec $d_k(t)$ des fonctions régulières de t. Comme $v(t) \in V_m$ pour tout $m \geq M$ et tout $t \in [0;T]$, on a d'après (5.15)

$$\int_0^T \langle u'_m(t), v(t)\rangle_{L^2} dt + \int_0^T \int_\Omega \nabla u_m(t) \cdot \nabla v(t) dt = \int_0^T \langle f(t), v(t)\rangle_{L^2} dt.$$

Par convergence faible pour la sous-suite $u_{\varphi(m)}$, on a donc

$$\int_0^T \langle u'(t), v(t)\rangle_{L^2} dt + \int_0^T \int_\Omega \nabla u(t) \cdot \nabla v(t) dt = \int_0^T \langle f(t), v(t)\rangle_{L^2} dt$$

pour tout $v(t)$ de la forme (5.17) ci-dessus, donc pour tout $v \in L^2(]0;T[,H^1_0(\Omega))$ par densité. Pour conclure que $(O1)$ est vérifié, il va falloir se débarrasser de l'intégrale en temps. Pour ce faire, on réécrit ceci comme : pour tout pour tout $v \in L^2(]0;T[,H^1_0(\Omega))$, on a

$$\int_0^T \langle u'(t), v(t)\rangle_{L^2} dt + \int_0^T \int_\Omega \nabla u(t) \cdot \nabla v(t) dt = \int_0^T \langle f(t), v(t)\rangle_{L^2} dt.$$

En prenant v l'indicatrice de $[0,s]$ multiplié par une fonction w ne dépendant que de x, on obtient donc que pour tout $s \in [0,T]$ et tout $w \in H^1_0(\Omega)$, on a

$$\int_0^s \langle u'(t), w\rangle_{L^2} dt + \int_0^s \int_\Omega \nabla u(t) \cdot \nabla w dt = \int_0^s \langle f(t), w\rangle_{L^2} dt.$$

On peut dériver par rapport à s et obtenir (C1). Il reste à vérifier que $u(0) = g$. Soit pour cela une fonction v de la forme (5.17) qui est régulière et satisfait de plus $v(T) \equiv 0$. En intégrant l'égalité ci-dessus par parties, on obtient

$$-\int_0^T \langle u(t), v'(t)\rangle_{L^2} dt + \int_0^T \int_\Omega \nabla u(t) \cdot \nabla v(t) dt = \int_0^T \langle f(t), v(t)\rangle_{L^2} dt + \langle u(0), v(0)\rangle.$$

En intégrant par parties l'équation pour u_m, on trouve de même :

$$-\int_0^T \langle u_m(t), v'(t)\rangle_{L^2} dt + \int_0^T \int_\Omega \nabla u_m(t) \cdot \nabla v(t) dt = \int_0^T \langle f(t), v(t)\rangle_{L^2} dt + \langle g, v(0)\rangle.$$

Par passage à la limite faible, on trouve donc

$$\langle u(0), v(0)\rangle = \langle g, v(0)\rangle,$$

c'est-à-dire $u(0) = g$ puisque $v(0)$ était quelconque.

Étape 4 : démonstration de (5.13)**.**

Il suffit de faire $m \to +\infty$ dans (5.14) appliqué à la sous-suite qui converge vers u, et utiliser la Proposition 1.41.

Étape 5 : Unicité.

On utilise la linéarité. Si $u$et v sont deux solutions faibles de (5.12) associées au même f et g, alors la différence $u - v$ est toujours une solution à (5.12) associée à $f = 0$ et $g = 0$. L'inégalité (5.13) donne alors immédiatement que $u - v = 0$, ce qui était le résultat voulu.

♢

Preuve : [Preuve 2 : Décomposition sur les fonctions propres du Laplacien] On utilise la méthode de décomposition sur les fonctions propres du Laplacien.

Étape 1 : forme de la solution.

Soit $u \in L^2(]0;T[, H_0^1(\Omega))$ telle que $u' \in L^2(]0;T[, H^{-1}(\Omega))$, une solution faible de (5.12). D'après le Théorème 5.28, on a $u \in C^0([0;T], L^2(\Omega))$.

Considérons maintenant la famille $(w_k)_{k\geq 1} \subset H_0^1(\Omega)$ des fonctions propres du Laplacien avec conditions de Dirichlet au bord de Ω :

$$-\Delta w_k = \lambda_k w_k$$

où les λ_k sont les valeurs propres du Laplacien de Dirichlet, voir le Théorème 3.3. Rappelons que l'on a $\langle w_k, w_\ell\rangle = \delta_{k\ell}$ et $\langle \nabla w_k, \nabla w_\ell\rangle = \lambda_k \delta_{k\ell}$. Comme $u \in C^0([0;T], L^2(\Omega))$, on peut écrire pour tout t

$$u(t) = \sum_{k\geq 1} \alpha_k(t) w_k$$

où chaque $\alpha_k(t) = \langle u(t), w_k\rangle_{L^2(\Omega)}$ est une fonction absolument continue sur $[0;T]$ d'après le Théorème 5.28. En choisissant $v = w_k$ dans $(C1)$, on obtient que chaque α_k est une solution du problème

$$\begin{cases} \alpha_k'(t) + \lambda_k \alpha_k(t) = \beta_k(t) \quad \text{dans }]0;T[\\ \alpha_k(0) = \alpha_k^0 \end{cases}$$

où

$$\beta_k(t) = \langle f(t), w_k\rangle, \qquad \alpha_k^0 = \langle g, w_k\rangle.$$

Il s'agit pour chaque k d'une équation différentielle ordinaire dont l'unique solution est

$$\alpha_k(t) = \alpha_k^0 e^{-\lambda_k t} + \int_0^t \beta_k(s) e^{-\lambda_k(t-s)} ds, \qquad t > 0.$$

Ainsi, on trouve que u doit vérifier

$$u(t) = \sum_{k\geq 1} e^{-\lambda_k t} \langle g, w_k\rangle w_k + \int_0^t \sum_{k\geq 1} e^{-\lambda_k(t-s)} \langle f(s), w_k\rangle w_k \tag{5.18}$$

si cette formule a un sens. L'unicité est donc automatique si nous pouvons montrer que cette formule a un sens dans les espaces fonctionnels adaptés. Introduisons l'opérateur

$$U(t) = \sum_{k\geq 1} e^{-\lambda_k t} |w_k\rangle\langle w_k| \tag{5.19}$$

où la notation $|w_k\rangle\langle w_k|$ désigne le projecteur orthogonal dans $L^2(\Omega)$ sur $\mathrm{Vect}(w_k)$. Comme $-\Delta \geq 0$, on obtient que $\lambda_k \geq 0$ pour tout k, donc que pour chaque $t > 0$ $U(t)$ définit un opérateur linéaire continu auto-adjoint tel que

$$0 \leq U(t) \leq 1. \tag{5.20}$$

La formule (5.18) s'écrit alors

$$u(t) = U(t)g + \int_0^t U(t-s) f(s)\, ds. \tag{5.21}$$

Si nous pouvons montrer que cette formule fournit bien une fonction $u \in L^2(]0;T[, H_0^1(\Omega))$ telle que $u' \in L^2(]0;T[, H^{-1}(\Omega))$, nous aurons démontré à la fois l'existence et l'unicité.

Étape 2 : propriétés du propagateur $U(t)$.

Nous démontrons ici certaines propriétés utiles de l'opérateur $U(t)$. Nous noterons $B(X,Y)$ l'espace de Banach des opérateurs linéaires continus auto-adjoints entre les espaces de Hilbert X et Y, muni de la norme usuelle

$$\|U\|_{X\to Y} = \sup_{x\in X,\ \|x\|_X=1} \|Ux\|_Y .$$

Remarquons que

$$U \in L^\infty([0;T], B(L^2(\Omega), L^2(\Omega)))$$

d'après (5.20).

Prouvons maintenant le

Lemme 5.46. *On a*

$$U \in B(L^2(\Omega), L^2([0,T[; H_0^1(\Omega)))$$

et

$$U' \in B(L^2(\Omega), L^2(]0,T[, H^{-1}(\Omega))).$$

Preuve : Commençons par estimer $\|U(t)\|_{L^2(\Omega)\to H_0^1(\Omega)}$. Pour cela, nous prenons une fonction $\psi \in L^2(\Omega)$ et calculons

$$\begin{aligned}\|U(t)\psi\|^2_{H_0^1(\Omega)} &= \|\nabla U(t)\psi\|^2_{L^2(\Omega)} \\ &= \sum_{k\geq 1} e^{-2t\lambda_k}\lambda_k \langle w_k,\psi\rangle^2\end{aligned}$$

d'après la formule de Parseval. Ainsi en utilisant le théorème de Fubini pour les fonctions positives, on a

$$\begin{aligned}\|U(t)\psi\|^2_{L^2(]0,T[,H_0^1(\Omega)} &= \int_{]0,T[}\sum_{k\geq 1} e^{-2t\lambda_k}\lambda_k\langle w_k,\psi\rangle^2 \\ &= \sum_{k\geq 1}\int_{]0,T[} e^{-2t\lambda_k}\lambda_k\, dt\langle w_k,\psi\rangle^2 \\ &= \sum_{k\geq 1}\frac{1}{2}\left(1-e^{-2T\lambda_k}\right)\langle w_k,\psi\rangle^2 \\ &\leq \frac{1}{2}\left(1-e^{-2T\lambda_1}\right)\sum_{k\geq 1}\langle w_k,\psi\rangle^2 \\ &= \frac{1}{2}\left(1-e^{-2T\lambda_1}\right)\|\psi\|^2_{L^2(\Omega)}.\end{aligned}$$

On a donc bien que $U \in B(L^2(\Omega); L^2(]0;T[, H_0^1(\Omega)))$.

Ensuite, on remarque que

$$-U'(t) = \sum_{k\geq 1}\lambda_k e^{-\lambda_k t}|w_k\rangle\langle w_k| = (-\Delta)U(t) = U(t)(-\Delta).$$

On a donc pour tout $(\varphi,\psi)\in L^2(\Omega)\times H_0^1(\Omega)$, et pour presque tout $t\in]0,T[$,

$$\begin{aligned}{}_{H^{-1}(\Omega)}\langle -U'(t)\varphi,\psi\rangle_{H_0^1(\Omega)} &= {}_{H^{-1}(\Omega)}\langle(-\Delta)U(t)\varphi,\psi\rangle_{H_0^1(\Omega)} \\ &= {}_{L^2(\Omega)}\langle\nabla U(t)\varphi,\nabla\psi\rangle_{L^2(\Omega)} \\ &\leq \|\psi\|_{H_0^1(\Omega)}\|U(t)\varphi\|_{H_0^1(\Omega)}.\end{aligned}$$

Ainsi,

$$\|U'(t)\varphi\|_{H^{-1}(\Omega)} \leq \|U(t)\varphi\|_{H_0^1(\Omega)}$$

et

$$\|U'(t)\varphi\|_{L^2(]0,T[,H^{-1}(\Omega))} \leq \|U(t)\varphi\|_{L^2(]0,T[,H_0^1(\Omega))} \leq \|U(t)\|_{L^2(\Omega)\to L^2(]0,T[,H_0^1(\Omega))}\|\varphi\|_{L^2(\Omega)}.$$

En conséquence, on obtient que

$$\|U'(t)\|_{L^2(\Omega)\to L^2(]0,T[,H^{-1}(\Omega))} \leq \|U(t)\|_{L^2(\Omega)\to L^2(]0,T[,H_0^1(\Omega))}.$$

$\diamondsuit$

Étape 3 : conclusion.

Montrons maintenant que (5.21) fournit bien une fonction de $L^2(]0;T[,H_0^1(\Omega))$ telle que $u' \in L^2(]0;T[,H^{-1}(\Omega))$ en utilisant le Lemme 5.46. Posons

$$u_1(t) = U(t)g \quad \text{et} \quad u_2(t) = \int_0^t U(t-s)f(s)\,ds.$$

D'après le Lemme 5.46, on a d'abord que $u_1(t) \in L^2(]0,T[,H_0^1(\Omega))$. De plus, comme $u_1'(t) = U'(t)g$, on a, toujours d'après le Lemme 5.46, que $u_1' \in L^2(]0,T[,H^{-1}(\Omega))$.

On a aussi

$$\begin{aligned}
\|u_2(t)\|^2_{L^2(]0,T[,H_0^1(\Omega))} &= \int_{]0,T[} \|u_2(t)\|^2_{H_0^1(\Omega)}\,dt,\\
&= \int_{]0,T[} \|\int_0^t U(t-s)f(s)\,ds\|^2_{H_0^1(\Omega)}\,dt,\\
&\le \int_{]0,T[}\int_0^t \|U(t-s)f(s)\|^2_{H_0^1(\Omega)}\,ds\,dt,\\
&= \int_{]0,T[}\int_0^t \|U(t')f(t'-s)\|^2_{H_0^1(\Omega)}\,dt'\,ds,\\
&\le T\int_{]0,T[} \|U\|^2_{L^2(\Omega)\to L^2(]0,T[,H_0^1(\Omega))}\|f(s)\|^2_{L^2(\Omega)}\,ds,\\
&= T\|U\|^2_{L^2(\Omega)\to L^2(]0,T[,H_0^1(\Omega))}\|f\|^2_{L^2(]0,T[,L^2(\Omega))}.
\end{aligned}$$

Ceci montre bien que $u_2 \in L^2(]0,T[,H_0^1(\Omega))$.

Finalement, on a

$$u_2'(t) = f(t) + \int_0^t U'(t-s)f(s)\,ds.$$

Or $f \in L^2(]0;T[,L^2(\Omega)) \subset L^2(]0;T[,H^{-1}(\Omega))$ et le second terme est traité comme ci-dessus et on obtient bien que $u_2' \in L^2(]0,T[,H^{-1}(\Omega))$.

Ainsi on obtient en particulier que u appartient à $C^0([0;T],L^2(\Omega))$. Notons que u satisfait par construction la formulation faible (i) pour $v = w_k$ pour tout $k \ge 1$, donc pour tout $v \in H_0^1(\Omega)$. Il faut encore vérifier que l'on a bien

$$\lim_{t\to 0} \|u(t) - g\|_{L^2(\Omega)} = 0.$$

Notons d'abord que

$$\begin{aligned}
\left\|\int_0^t U(t-s)f(s)\,ds\right\|_{L^2(\Omega)} &\le \int_0^t \|U(t-s)\|_{L^2(\Omega)\to L^2(\Omega)}\,\|f(s)\|_{L^2(\Omega)}\\
&\le \sqrt{t}\,\|f\|_{L^2(]0;T[,L^2(\Omega))}
\end{aligned}$$

où nous avons utilisé que $\|U(t)\|_{L^2(\Omega)\to L^2(\Omega)} \leq 1$ et l'inégalité de Cauchy-Schwarz. Ceci démontre que le dernier terme de (5.21) tend vers 0 dans $L^2(\Omega)$ quand $t \to 0$. Ainsi, nous devons juste prouver le

Lemme 5.47. *Soit $g \in L^2(\Omega)$. Alors on a*

$$\lim_{t\to 0} \|U(t)g - g\|_{L^2(\Omega)} = 0.$$

Preuve : On a, comme $(w_k)_{k\geq 1}$ est une base orthonormée de $L^2(\Omega)$,

$$U(t)g - g = \sum_{k\geq 1} (e^{-\lambda_k t} - 1)\langle g, w_k\rangle w_k$$

donc

$$\|U(t)g - g\|^2_{L^2(\Omega)} = \sum_{k\geq 1} (e^{-\lambda_k t} - 1)^2 \langle g, w_k\rangle^2$$

qui tend vers 0 par convergence dominée (ou en coupant la série en deux). ♢

Ceci termine la preuve du Théorème 5.42. ♢

Remarque 5.48. *Si $f \in L^2(]0;T[, L^2(\Omega))$ pour tout $T > 0$, alors on obtient une unique solution définie pour tout $t > 0$, mais les estimées sur cette solution dépendent du temps final considéré.*

Voici maintenant un résultat fournissant la régularité par rapport aux conditions initiales :

Théorème 5.49 (Régularité par rapport aux conditions initiales)**.** *Il existe une constante C (dépendant de T et Ω) telle que pour tous $g \in L^2(\Omega)$ et $f \in L^2(]0;T[, L^2(\Omega))$, l'unique solution u de* (5.12) *vérifie :*

$$\max_{0\leq t\leq T} \|u(t)\|_{L^2(\Omega)} + \|u\|_{L^2(]0;T[,H^1_0(\Omega))} + \|u'\|_{L^2(]0;T[,H^{-1}(\Omega))} \leq C\left(\|f\|_{L^2(]0;T[,L^2(\Omega))} + \|g\|_{L^2(\Omega)}\right). \quad (5.22)$$

Preuve : On peut prendre $v = u$ dans la formulation faible $(C1)$. On obtient, pour presque tout $t > 0$,

$$_{H^{-1}(\Omega)}\langle u', u\rangle_{H^1_0(\Omega)} + \int_\Omega |\nabla u(t)|^2 = \langle f(t), u(t)\rangle_{L^2(\Omega)}. \quad (5.23)$$

On a

$$\langle f(t), u(t)\rangle_{L^2(\Omega)} \leq \frac{1}{2}\left(\|f(t)\|^2_{L^2(\Omega)} + \|u(t)\|^2_{L^2(\Omega)}\right)$$

donc,

$$_{H^{-1}(\Omega)}\langle u', u\rangle_{H^1_0(\Omega)} + \int_\Omega |\nabla u(t)|^2 \leq \frac{1}{2}\left(\|f(t)\|^2_{L^2(\Omega)} + \|u(t)\|^2_{L^2(\Omega)}\right). \quad (5.24)$$

En posant $\eta(t) = \|u(t)\|^2_{L^2(\Omega)}$ et en utilisant le Théorème 5.28, on obtient

$$\eta'(t) \leq \eta(t) + \|f(t)\|^2_{L^2(\Omega)}.$$

D'après le lemme de Gronwall, on déduit

$$\|u(t)\|^2_{L^2(\Omega)} \leq e^t \left(\|g\|^2_{L^2(\Omega)} + \int_0^t \|f(s)\|^2_{L^2(\Omega)}\, ds \right)$$

donc

$$\max_{0\leq t\leq T} \|u(t)\|^2_{L^2(\Omega)} \leq e^T \left(\|g\|^2_{L^2(\Omega)} + \|f\|^2_{L^2(]0;T[,L^2(\Omega))} \right).$$

D'après l'inégalité (5.24), on a aussi, en intégrant sur $[0;T]$,

$$\|u\|^2_{L^2(]0;T[,H^1_0(\Omega))} \leq \frac{1}{2}\|g\|^2_{L^2(\Omega)} + \frac{1}{2}\|f\|^2_{L^2(]0;T[,L^2(\Omega))} + \frac{T}{2}\max_{0\leq t\leq T}\|u(t)\|^2_{L^2(\Omega)}.$$

On déduit alors l'estimée sur $\|u\|_{L^2(]0;T[,H^1_0(\Omega))}$ en utilisant ce que nous avons déjà démontré pour majorer le dernier terme ci-dessus.

On estime maintenant $\|u'\|_{L^2(]0;T[,H^{-1}(\Omega))}$ par dualité. Soit une fonction fixée $v \in H^1_0(\Omega)$, telle que $\|v\|_{H^1_0(\Omega)} \leq 1$. On a alors par définition des solutions faibles

$$_{H^{-1}(\Omega)}\langle u', v\rangle_{H^1_0(\Omega)} = \langle f(t), v\rangle_{L^2} - \int_\Omega \nabla u(t)\cdot\nabla v. \tag{5.25}$$

Ceci démontre par un raisonnement déjà vu que

$$\|u'(t)\|_{H^{-1}(\Omega)} \leq \|f(t)\|_{L^2(\Omega)} + \|\nabla u(t)\|_{L^2(\Omega)}.$$

On obtient alors l'estimée voulue en passant au carré, en intégrant sur $[0;T]$ et en utilisant les résultats précédents. ♢

5.3.2 Propriétés qualitatives des solutions faibles

Théorème 5.50 (Comportement asymptotique)**.** *Soit Ω un ouvert borné régulier, $g \in L^2(\Omega)$ et $u \in C^0([0;T], L^2(\Omega))$ l'unique solution faible obtenue avec le Théorème 5.42, avec $f \equiv 0$. Alors on a :*

$$\lim_{t\to+\infty} \|u(t)\|_{L^2(\Omega)} = 0.$$

Preuve : D'après l'inégalité de Poincaré (cf. la Proposition 1.24 et l'Exercice 23), on sait que la première valeur propre du Laplacien sur Ω avec conditions de Dirichlet est strictement positive. On a donc $-\Delta \geq \epsilon > 0$ au sens des formes quadratiques, ce qui signifie aussi que $\lambda_k \geq \epsilon > 0$ où les λ_k sont les valeurs propres introduites dans

la preuve du Théorème 5.42. Ceci démontre en particulier que $0 \leq U(t) \leq e^{-\epsilon t}$ et donc que

$$\|U(t)\|_{L^2(\Omega)\to L^2(\Omega)} \leq e^{-\epsilon t}.$$

Or si $f \equiv 0$, on a $u(t) = U(t)g$ d'après (5.21), donc

$$\|u(t)\|_{L^2(\Omega)} \leq e^{-\epsilon t} \|g\|_{L^2(\Omega)} \to_{t\to+\infty} 0.$$

♢

Examinons maintenant la régularité de la solution lorsque les données initiales sont plus ou moins régulières.

Théorème 5.51 (Effet régularisant avec $f \equiv 0$)**.** *On suppose que Ω est un ouvert borné de $\mathbb{R}^n$, de classe C^∞. Soit $g \in L^2(\Omega)$ une condition initiale et u l'unique solution faible obtenue par le Théorème 5.42. Alors pour tout $0 < \epsilon < T$, on a*

$$u \in C^\infty([\epsilon; T] \times \overline{\Omega}).$$

Preuve : La preuve est plus difficile que dans le cas de l'espace tout entier et nous ne donnons que les idées générales. Fixons $0 < \epsilon < T$. Nous voulons montrer que $(t, x) \mapsto (U(t)g)(x)$ est une fonction régulière par rapport au couple (t, x) lorsque $g \in L^2(\Omega)$. L'idée est de prouver que pour tous $\ell \geq 0$ et $m \geq 0$, il existe une constante $C_{\ell,m}$ telle que

$$\left\|\partial_t^\ell (-\Delta)^m U(t)g\right\|_{L^2(]\epsilon;T[,L^2(\Omega))} \leq C_{\ell,m} \|g\|_{L^2(\Omega)}. \tag{5.26}$$

Ceci signifie par régularité elliptique que

$$u = U(t)g \in H^r(]\epsilon; T[\times\Omega)$$

pour tout $r \geq 0$. D'après les injections de Sobolev, on obtient bien que $u \in C^\infty([\epsilon; T] \times \overline{\Omega})$.

Pour démontrer (5.26), on peut par densité prendre $g \in \mathrm{Vect}(w_1, ..., w_m)$ et se rendre compte suivant un argument précédent que

$$\left\|\partial_t^\ell (-\Delta)^m U(t)g\right\|_{L^2(\Omega)} \leq \sup_{k\geq 1} \left(\lambda_k^{\ell+m} e^{-\lambda_k t}\right) \|g\|_{L^2(\Omega)}.$$

On obtient bien (5.26) avec

$$C_{\ell,m} := (T - \epsilon)^{1/2} \sup_{x>0} x^{\ell+m} e^{-x\epsilon}.$$

♢

Obtenir la régularité jusqu'à $t = 0$ ou avec un terme source $f \neq 0$ est plus difficile. Schématiquement, on peut démontrer

g	f	u
$H^{2m+1}(\Omega)$	$\frac{d^k}{dt^k} f \in L^2(]0;T[, H^{2m-2k}(\Omega))$ $k=0,...,m$	$\frac{d^k}{dt^k} u \in L^2(]0;T[, H^{2m+2-2k}(\Omega))$ $k=0,...,m+1$

mais il faut ajouter des *conditions de compatibilité*. Par exemple la dérivée u' vérifie aussi l'équation de la chaleur mais avec condition initiale $u'(0) = \Delta g + f(0,\cdot)$. Pour pouvoir utiliser les résultats précédents, il faut donc que cette fonction soit au moins dans $L^2(\Omega)$, ce qui impose des conditions sur f et g. Voir par exemple [6] pour plus de détails.

Nous nous contenterons du résultat partiel suivant :

Théorème 5.52 (Régularité)**.** *On suppose que Ω est un ouvert de bord C^∞ et que $g \in C_0^\infty(\Omega)$, $f \in C^\infty([0;T], C_0^\infty(\Omega))$. Alors*

$$u \in C^\infty([0;T] \times \overline{\Omega}).$$

Preuve : Comme précédemment, on démontre que $\partial_t^\ell (-\Delta)^m u \in L^2(]0;T[\times\Omega)$ pour tous $\ell, m \geq 0$. On utilise la propriété fondamentale vue plus haut

$$U'(t) = (-\Delta)U(t) = U(t)(-\Delta).$$

Ainsi

$$(-\Delta)^m u = U(t)(-\Delta)^m g + \int_0^t U(t-s)(-\Delta)^m f(s) ds.$$

Or $(-\Delta)^m g \in L^2(\Omega)$ et $(-\Delta)^m f \in L^2(]0;T[, L^2(\Omega))$ par hypothèse. Donc toute l'étude précédente implique que

$$(-\Delta)^m u \in L^2(]0;T[, H_0^1(\Omega)) \cap C^0([0;T], L^2(\Omega)), \quad \partial_t (-\Delta)^m u \in L^2(]0;T[, H^{-1}(\Omega))$$

pour tout $m \geq 1$. Rappelons que $\partial_t u = \Delta u + f$ donc

$$\partial_t (-\Delta)^{m-1} u = -(-\Delta)^m u + (-\Delta)^{m-1} f$$

au moins au sens des distributions. Or $(-\Delta)^m u \in C^0([0;T]; L^2(\Omega))$ et bien sûr $(-\Delta)^{m-1} f \in C^0([0;T]; L^2(\Omega))$ donc finalement

$$\partial_t (-\Delta)^{m-1} u \in C^0([0;T]; L^2(\Omega))$$

pour tout $m \geq 1$.

Ensuite on a

$$\partial_t^2 (-\Delta)^{m-1} u = -\partial_t (-\Delta)^m u + \partial_t (-\Delta)^{m-1} f$$

au moins au sens des distributions, donc

$$\partial_t^2 (-\Delta)^{m-1} u \in C^0([0;T], L^2(\Omega)).$$

La démonstration suit en itérant l'argument précédent. ♢

Théorème 5.53 (Principe du maximum faible)**.** *Soient Ω un ouvert borné de $\mathbb{R}^n$, $T > 0$, $g \in L^2(\Omega)$, $f \in L^2(]0;T[, L^2(\Omega))$, et u l'unique solution faible obtenue grâce au Théorème 5.42. Si $f \geq 0$ presque partout dans $[0;T]\times\Omega$ et $g \geq 0$ presque partout dans Ω, alors $u \geq 0$ presque partout dans $[0;T]\times\Omega$.*

Preuve : Nous commençons par démontrer ce résultat en supposant que f et g sont respectivement dans $C^\infty([0;T], C_0^\infty(\Omega))$ et $C_0^\infty(\Omega)$ et que $f(t) > 0$ sur Ω. D'après le théorème précédent, u est très régulière (en fait on a seulement besoin que u soit de classe C^2 par rapport à (t,x)).

Soit $(t_0, x_0) \in [0;T]\times\overline{\Omega}$ un point où u atteint son minimum. Si $t_0 = 0$ ou $x_0 \in \partial\Omega$, on a clairement $u(t_0,x_0) = 0$ par positivité de g et la condition de Dirichlet au bord. On peut donc supposer que $x_0 \in \Omega$ et $t_0 \in]0;T]$. Supposons pour commencer que $t_0 < T$. Alors comme le minimum de u est atteint dans l'ouvert $]0;T[\times\Omega$, on a

$$\partial_t u(t_0,x_0) = 0 \quad \text{ et } \nabla u(t_0,x_0) = 0.$$

Comme il s'agit d'un minimum, la Hessienne de u est nécessairement positive en (t_0,x_0), donc on obtient

$$-\Delta u(t_0,x_0) = -\text{tr } (\text{Hess}(u)(t_0,x_0)) \leq 0.$$

Or d'après l'équation,

$$-\Delta u(t_0,x_0) = f(t_0,x_0) > 0$$

donc c'est absurde.

Si maintenant le minimum est atteint en (T,x_0), on a seulement

$$\frac{\partial}{\partial t}u(T,x_0) \leq 0$$

mais on a toujours

$$-\Delta u(t_0,x_0) = -\text{tr } (\text{Hess}(u)(t_0,x_0)) \leq 0$$

car $x \mapsto u(T,x)$ admet un minimum local en x_0 dans l'ouvert Ω. L'équation donne alors

$$0 < f(T,x_0) = \frac{\partial}{\partial t}u(T,x_0) - \Delta u(T,x_0) \leq 0$$

qui est aussi absurde. Nous avons prouvé que soit $t_0 = 0$, soit $x_0 \in \partial\Omega$. On a donc bien $\min_{[0;T]\times\overline{\Omega}} u(t,x) = 0$.

Nous venons donc de démontrer que si f et g sont des fonctions régulières strictement positives sur Ω, alors $u \geq 0$. Le cas général s'obtient par densité des fonctions régulières positives dans $L^2(\Omega)$ et $L^2(]0;T[, L^2(\Omega))$, et en utilisant la continuité de u par rapport aux données f et g, prouvée au Théorème 5.49. $\diamondsuit$

Remarque 5.54. *Dans le cas où $g \in H^{1/2}(\Omega)$, il existe une autre preuve de ce résultat qu'il est utile de connaître, et dont on donne ici les grandes idées sans rentrer dans les détails. On prend $v = u^-$ dans la formulation variationnelle $(C1)$ et on obtient donc (on admet que l'on peut prendre $v = u^-$ comme fonction test, même si u^-, qui est bien une fonction de $H_0^1(\Omega)$ pour presque tout temps, dépend du temps)*

$$\int_\Omega \frac{\partial u}{\partial t} u^- + \int_\Omega \nabla u \cdot \nabla u^- = \int_\Omega f u^-.$$

On en déduit :

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |u^-|^2 + \int_\Omega |\nabla u^-|^2 \leq 0,$$

et donc, comme $\int_\Omega |u_0^-|^2 = 0$, pour tout $t \geq 0$,

$$\frac{1}{2}\int_\Omega |u^-|^2 + \int_0^t \int_\Omega |\nabla u^-|^2 \leq 0.$$

Ceci permet de conclure que $u^- = 0$ et donc $u \geq 0$ presque partout.

Pour être tout à fait rigoureux et justifier l'égalité $\int_\Omega \frac{\partial u}{\partial t} u^- = \frac{1}{2}\frac{d}{dt}\int_\Omega |u^-|^2$, on peut utiliser la méthode des troncatures de Stampacchia. On renvoie à [2, Théorème X.3]. Ainsi, si u et v désignent deux solutions de l'équation de la chaleur, avec même second membre f et mêmes conditions aux limites g, et si les conditions initiales satisfont $u_0 \leq v_0$ alors $u \leq v$.

Remarque 5.55. *Le fait que u reste ≥ 0 lorsque les données sont ≥ 0 est important physiquement, par exemple si u représente une température.*

Voici maintenant un résultat plus précis quand $f \equiv 0$ et qui traduit l'existence d'une *propagation à vitesse infinie* : même si la condition initiale s'annule à l'intérieur de Ω, la solution u vérifie $u(t,x) > 0$ pour tout $t > 0$ et $x \in \Omega$.

Théorème 5.56 (Propagation à vitesse infinie)**.** *Soit Ω un ouvert borné régulier de $\mathbb{R}^n$, un temps final $T > 0$ et une fonction $g \in L^2(\Omega)$ telle que $g \neq 0$ et $g \geq 0$ presque partout. Alors la solution u obtenue par le Théorème 5.42 avec $f \equiv 0$ vérifie*

$$u(t,x) > 0 \qquad \forall x \in \Omega$$

pour tout temps $t > 0$.

La démonstration, complexe, repose sur une inégalité de type Harnack parabolique, ou une formule de la moyenne parabolique. Voir [6] pour plus de détails.

5.4 Exercices

Exercice 39. *Soit H un espace de Hilbert, $v \in H$ et $f : [a,b] \to H$ une fonction intégrable. Montrer que*

$$\left\langle \int_{[a,b]} f(t)\,dt, v \right\rangle_H = \int_{[a,b]} \langle f(t), v\rangle_H \,dt.$$

Exercice 40.

1. *Montrer que si si $y^0 \in H^1(\mathbb{R})$, il existe une unique solution faible y à (4.8) qui satisfasse de plus*

$$y \in C^0([0,\infty), H^1(\mathbb{R})) \cap C^1([0,\infty), L^2(\mathbb{R})), \quad \partial_x y \in C^0([0,\infty), L^2(\mathbb{R})), \tag{5.27}$$

2. *Montrer que l'égalité de la première ligne de (4.8) a lieu dans l'espace*

$$C^0([0,\infty), L^2(\mathbb{R})).$$

3. *Montrer que*

$$\lim_{t\to 0} \left\| u(t,\cdot) - y^0 \right\|_{H^1(\mathbb{R})} = 0. \tag{5.28}$$

4. *Inversement, montrer que la solution faible à (4.8) est l'unique fonction satisfaisant (5.27), la première ligne de (4.8) dans $C^0([0,\infty), L^2(\mathbb{R}))$ et (5.28).*

Exercice 41 (Équation de la chaleur dans tout l'espace avec second membre)**.** *Soient $g \in L^2(\mathbb{R}^n)$ et $f \in C^1([0;\infty), L^2(\mathbb{R}^n))$. Montrer que le problème*

$$\begin{cases} \dfrac{\partial}{\partial t} u - \Delta u = f, & t > 0 \\ u(0) = g, \end{cases} \tag{5.29}$$

admet une solution unique $u \in C^0([0;\infty), L^2(\mathbb{R}^n)) \cap C^1((0;\infty), L^2(\mathbb{R}^n))$, donnée par la formule de Duhamel

$$u(t) = U(t)g + \int_0^t U(t-s) f(s)\,ds. \tag{5.30}$$

Exercice 42. (Un théorème général)

1. *En s'inspirant de l'une des deux démonstrations du Théorème 5.42, démontrer le résultat général suivant :*

   Théorème 5.57. *Soient H et V deux espaces de Hilbert tels que $V \hookrightarrow H$ avec injection compacte et V est dense dans H. Soit $a(\cdot,\cdot)$ une forme bilinéaire symétrique continue et coercive dans V. Soit un temps final $T > 0$, une condition initiale $g \in H$ et un terme source $f \in L^2(]0;T[, H)$. Il existe une unique solution faible $u \in L^2(]0;T[, V)$ telle que $u' \in L^2(]0;T[, V')$ au problème*

$$\begin{cases} \frac{d}{dt}\langle u(t), v\rangle_H + a(u(t), v) = \langle f(t), v\rangle_H & \forall v \in V,\ t \in]0;T[\\ u(0) = g. \end{cases}$$

De plus il existe une constante C telle que

$$\|u\|_{L^2(]0;T[,V)} + \|u\|_{C^0([0;T],H)} \leq C(\|f\|_{L^2(]0;T[,H)} + \|g\|_H).$$

2. (Équation de la chaleur en milieu inhomogène). *Soit $\Omega \subset \mathbb{R}^n$ un ouvert borné régulier et A une fonction définie sur Ω à valeurs dans les matrices symétriques réelles définies positives de taille n, telle que*

$$\alpha I_n \leq A(x) \leq \beta I_n$$

p.p. $x \in \Omega$, où $\alpha, \beta > 0$ et I_n est l'identité de $\mathbb{R}^n$. En déduire l'existence d'une unique solution faible au problème

$$\begin{cases} \frac{\partial}{\partial t}u(t,x) - \operatorname{div}(A(x)\nabla u(t,x)) = f, & (t,x) \in]0;T[\times\Omega, \\ u(t,x) = 0, & (t,x) \in]0;T[\times\partial\Omega \\ u(0,x) = g(x), \end{cases}$$

où $g \in L^2(\Omega)$ et $f \in L^2(]0;T[, L^2(\Omega))$.

Exercice 43. *Soit u l'unique solution faible de (5.12). On suppose $f \in L^2((0,T)\times\Omega)$ et $g \in H_0^1(\Omega)$. Montrer alors que la solution $u \in L^\infty((0,T), H_0^1(\Omega)) \cap H^1((0,T), L^2(\Omega))$ et satisfait l'estimée d'énergie : $\forall t \in [0,T]$,*

$$\int_\Omega |\nabla u|^2(t) + \int_0^t \int_\Omega |\partial_t u|^2 \leq \int_\Omega |\nabla g|^2 + \int_0^T \|f\|^2_{L^2(\Omega)}.$$

En déduire que $u \in L^2((0,T), H^2(\Omega))$.

Exercice 44. *Montrer que si $f \in L^2(\Omega)$ ne dépend pas du temps, l'unique solution faible obtenue par le Théorème 5.42 vérifie*

$$\lim_{t\to+\infty} \|u(t) - v\|_{L^2(\Omega)} = 0$$

où v est l'unique solution de l'équation de Laplace

$$-\Delta v = f$$

dans $H_0^1(\Omega)$.

Exercice 45. *Tout comme pour l'équation de la chaleur, nous étudions maintenant l'équation des ondes dans un ouvert borné $\Omega \subset \mathbb{R}^n$, avec des conditions de Dirichlet au bord et terme source f :*

$$\begin{cases} \frac{\partial^2}{\partial t^2}u(t,x) - \Delta u(t,x) = f(t,x), & (t,x) \in]0;T[\times\Omega, \\ u(t,x) = 0 \text{ si } x \in \partial\Omega, \\ u(0,x) = g(x), \\ \frac{\partial}{\partial t}u(0,x) = h(x), \end{cases} \tag{5.31}$$

Comme pour l'équation de la chaleur, nous commençons par introduire une notion de solution faible. On considère $f \in L^2(]0;T[, L^2(\Omega))$, $g \in H_0^1(\Omega)$ et $h \in L^2(\Omega)$.

Définition 5.58 (Solutions faibles)**.** *Soit $u \in L^2(]0;T[, H_0^1(\Omega))$ telle que $u' \in L^2(]0;T[, L^2(\Omega))$ et $u'' \in L^2(]0;T[, H^{-1}(\Omega))$. On dit que u est une* solution faible *de* (5.31) *si on a*

$(O1)$ $_{H^{-1}(\Omega)}\langle u''(t), v\rangle_{H_0^1(\Omega)} + \displaystyle\int_\Omega \nabla u(t)\cdot\nabla v = \int_\Omega f(t)v$ *pour tout $v \in H_0^1(\Omega)$ et presque tout en $t \in]0;T[$.*

$(O2)$ $u(0) = g$.

$(O3)$ $u'(0) = h$.

1. *Expliquer ce qui permet de donner un sens à $(O2)$ et $(O3)$.*

Le but de cet exercice est principalement de démontrer le résultat suivant :

Théorème 5.59 (Existence et unicité de solutions faibles)**.** *Soit $\Omega \subset \mathbb{R}^n$ un ouvert borné régulier. On suppose que $f \in L^2(]0;T[, L^2(\Omega))$, $g \in H_0^1(\Omega)$ et $h \in L^2(\Omega)$. Alors le problème* (5.31) *admet une unique solution faible u.*

De plus, on a

$$u \in L^\infty(]0;T[, H_0^1(\Omega)), \qquad u' \in L^\infty(]0;T[, L^2(\Omega))$$

et

$$\sup_{0\le t\le T}\left(\|u(t)\|_{H_0^1(\Omega)} + \|u'(t)\|_{L^2(\Omega)}\right) + \|u''\|_{L^2(]0;T[,H^{-1}(\Omega))} \le C\left(\|f\|_{L^2(]0;T[,L^2(\Omega))} + \|g\|_{H_0^1(\Omega)} + \|h\|_{L^2(\Omega)}\right) \quad (5.32)$$

pour une constante C indépendante de u.

Considérons la famille des fonctions propres du Laplacien $(w_k)_{k\ge1} \subset H_0^1(\Omega)$, introduite dans la preuve pour l'équation de la chaleur. Rappelons que

— *$(w_k)_{k\ge1}$ est une base* orthogonale *de $H_0^1(\Omega)$;*

— *$(w_k)_{k\ge1}$ est une base* orthonormée *de $L^2(\Omega)$.*

On pose alors

$$V_m := \mathrm{vect}(w_1, ..., w_m)$$

et on cherche une solution $u_m \in C^2([0;T], V_m)$ faible dans V_m.

2. *On écrit*

$$u_m(t,x) = \sum_{k=1}^m d_k^m(t) w_k(x).$$

En exhibant les EDO que doivent vérifier les d_k^m, montrer l'existence d'une unique solution vérifiant $u_m \in C^1([0;T], H^2(\Omega)\cap H_0^1(\Omega))$, $u_m' \in C^0([0;T], H^2(\Omega)\cap H_0^1(\Omega))$ et $u_m'' \in L^2(]0;T[, H^2(\Omega)\cap H_0^1(\Omega))$

3. *Montrer qu'il existe une constante C qui ne dépend que de Ω et $T > 0$ telle que pour tout $m \geq 1$,*

$$\max_{0\leq t\leq T}\left(\|u_m(t)\|_{H_0^1(\Omega)} + \|u_m'(t)\|_{L^2(\Omega)}\right) \leq C\left(\|f\|_{L^2(]0;T[,L^2(\Omega))} + \|g\|_{H_0^1(\Omega)} + \|h\|_{L^2(\Omega)}\right). \quad (5.33)$$

4. *Montrer qu'il existe une constante C qui ne dépend que de Ω et $T > 0$ telle que pour tout $m \geq 1$,*

$$\|u_m''(t)\|_{L^2(]0;T[,H^{-1}(\Omega))} \leq C\left(\|f\|_{L^2(]0;T[,L^2(\Omega))} + \|g\|_{H_0^1(\Omega)} + \|h\|_{L^2(\Omega)}\right). \quad (5.34)$$

Indication : considérer, $v \in H_0^1(\Omega)$, écrire $v = v_1 + v_2$ avec $v_1 \in V_m$ et $v_2 \in (V_m)^\perp$, puis regarder ${}_{H^{-1}(\Omega)}\langle u_m'', v\rangle_{H_0^1(\Omega)}$.

5. *En déduire que $(O1)$ est vérifié.*
6. *En considérant une fonction régulière quelconque $v \in C^\infty([0;T], V_{m'})$, telle que $v(T) = v'(T) \equiv 0$ et en intégrant (**??**) par parties, montrer que u vérifie $(O2)$ et $(O3)$.*
7. *Pour montrer l'unicité, se ramener au cas où $f = g = h = 0$, fixer $t_0 \in]0, T[$, introduire*

$$v(t) = \int_t^{t_0} u(s)\, ds,$$

puis montrer que

$$\int_0^{t_0} {}_{H^{-1}(\Omega)}\langle u'(s), u(s)\rangle_{H_0^1(\Omega)} ds - \int_0^{t_0} \langle v'(s), v(s)\rangle_{H_0^1(\Omega)}\, ds = 0$$

et conclure.

8. *Montrer L'inégalité* (5.32).

Exercice 46. (Formules explicites pour la décomposition sur les modes propres du Laplacien). *Considérons une solution faible u de l'équation des ondes sur Ω. On décompose u sous la forme*

$$u(t) = \sum_{k\geq 1} \alpha_k(t) w_k$$

et on pose $\beta_k(t) = \langle f(t), w_k\rangle$.

1. *Trouver l'équation différentielle ordinaire vérifiée par α_k et écrire la forme de la solution pour tout k.*

2. *En déduire que u doit s'écrire sous la forme*

$$u(t) = V'(t)g + V(t)h + \int_0^t V(t-s)f(s)\,ds \tag{5.35}$$

où

$$V(t) = \sum_{k\geq 1} \frac{\sin(\sqrt{\lambda_k}t)}{\sqrt{\lambda_k}} |w_k\rangle\langle w_k|.$$

3. *Montrer que*

$$\|V(t)\|_{L^2(\Omega)\to L^2(\Omega)} \leq Ct \qquad \|V'(t)\|_{L^2(\Omega)\to L^2(\Omega)} \leq C,$$

$$\|V(t)\|_{L^2(\Omega)\to H_0^1(\Omega)} \leq C, \qquad \|V'(t)\|_{H_0^1(\Omega)\to H_0^1(\Omega)} \leq C.$$

Vérifier que $V''(t) = \Delta V(t) = V(t)\Delta$ et en déduire que

$$\|V''(t)\|_{H_0^1(\Omega)\to L^2(\Omega)} \leq C.$$

4. *Montrer que la formule* (5.35) *fournit une fonction telle que $u \in L^\infty(]0;T[, H_0^1(\Omega))$, $u' \in L^\infty(]0;T[, L^2(\Omega))$ et $u'' \in L^2(]0;T[, L^2(\Omega))$, qui est l'unique solution faible de l'équation des ondes sur Ω.*

Exercice 47. (Un théorème général)

1. *En s'inspirant de la démonstration du Théorème 5.59, démontrer le résultat général suivant :*

   Théorème 5.60. *Soit H et V deux espaces de Hilbert tel que $V \hookrightarrow H$ avec injection compacte et V est dense dans H. Soit $a(\cdot,\cdot)$ une forme bilinéaire symétrique continue et coercive dans V. Soit un temps final $T > 0$, une condition initiale $(g,h) \in V \times H$ et un terme source $f \in L^2(]0;T[, H)$. Il existe une unique solution faible $u \in L^2(]0;T[, V)$ telle que $u' \in L^2(]0;T[, H)$ et $u'' \in L^2(]0;T[, V')$ au problème*

$$\begin{cases} \frac{d^2}{dt^2}\langle u(t), v\rangle_H + a(u(t), v) = \langle f(t), v\rangle_H \quad \forall v \in V,\ t \in]0;T[\\ \qquad\qquad u(0) = g,\ u'(0) = h. \end{cases}$$

   De plus il existe une constante C telle que

$$\|u\|_{L^\infty([0;T],V)} + \|u'\|_{L^\infty([0;T],H)} + \|u''\|_{L^2(]0;T[,V')} \leq C(\|f\|_{L^2(]0;T[,H)} + \|g\|_V + \|h\|_H).$$

2. (Équation des ondes en milieu inhomogène). *Soit $\Omega \subset \mathbb{R}^n$ un ouvert borné régulier et A une fonction définie sur Ω à valeurs dans les matrices symétriques réelles définies positives de taille n, telle que*

$$\alpha I_n \leq A(x) \leq \beta I_n$$

p.p. $x \in \Omega$, *où* $\alpha, \beta > 0$ *et* I_n *est l'identité de* $\mathbb{R}^n$. *En déduire l'existence d'une unique solution faible au problème*

$$\begin{cases} \frac{\partial^2}{\partial t^2} u(t,x) - \operatorname{div}(A(x)\nabla u(t,x)) = f, & (t,x) \in]0;T[\times\Omega, \\ u(t,x) = 0, & (t,x) \in]0;T[\times\partial\Omega \\ u(0,x) = g(x), \ \frac{\partial}{\partial t} u(0,x) = h(x), & \end{cases}$$

où $g \in H^1(\Omega)$, $h \in L^2(\Omega)$ *et* $f \in L^2(]0;T[, L^2(\Omega))$.

Exercice 48 (Propriétés qualitatives des solutions faibles pour l'équation des ondes)**.** *1. Soit* $\Omega \subset \mathbb{R}^n$ *un ouvert borné régulier. On suppose que* $f \in L^2(]0;T[, L^2(\Omega))$, $g \in H_0^1(\Omega)$ *et* $h \in L^2(\Omega)$. *Montrer que l'équation des ondes rétrograde en temps*

$$\begin{cases} \dfrac{\partial^2}{\partial t^2} u(t,x) - \Delta u(t,x) = f(t,x), \quad (t,x) \in]0;T[\times\Omega, \\ u(t,x) = 0 \ \textit{si} \ x \in \partial\Omega, \\ u(T,x) = g(x), \\ \frac{\partial}{\partial t} u(T,x) = h(x), \end{cases} \tag{5.36}$$

admet une unique solution faible $\tilde{u} \in L^\infty(]0;T[, H_0^1(\Omega))$ *avec* $\tilde{u}' \in L^\infty(]0;T[, L^2(\Omega))$. *De plus si* u *est la solution de l'équation des ondes usuelle telle que* $u(T) = g$ *et* $u'(T) = h$, *alors on a* $u = \tilde{u}$.

2. On se place sous les hypothèses du Théorème 5.59. Montrer que u *vérifie*

$$u \in C^0([0;T], H_0^1(\Omega)), \qquad u' \in C^0([0;T], L^2(\Omega)),$$

et satisfait l'égalité

$$\int_\Omega \left(\left| \frac{\partial}{\partial t} u(t,x) \right|^2 + |\nabla u(t,x)|^2 \right) dx = \int_\Omega \left(h(x)^2 + |\nabla g(x)|^2 \right) dx + 2 \int_0^t \int_\Omega f(s,x) u'(s,x) dx\, ds \tag{5.37}$$

pour tout $t \in]0;T[$. *En particulier, si* $f \equiv 0$, *on a la conservation de l'énergie :*

$$\int_\Omega \left(\left| \frac{\partial}{\partial t} u(t,x) \right|^2 + |\nabla u(t,x)|^2 \right) dx = \int_\Omega \left(h(x)^2 + |\nabla g(x)|^2 \right) dx \tag{5.38}$$

pour tout $t \in]0;T[$.

Bibliographie

[1] G. Allaire, Analyse numérique et optimisation (Editions de l'Ecole Polytechnique, 2005).

[2] H. Brézis, Analyse fonctionnelle (Dunod, Paris, 1999).

[3] R. Dautray et J.-L. Lions, Analyse mathématique et calcul numérique pour les sciences et les techniques, vol. 8 : Évolution : semi-groupe, variationnel (Masson, Paris, 1988).

[4] E.B. Davies, Spectral Theory and Differential Operators (Cambridge University Press, 1995).

[5] V. Ehrlacher, Analyse et Equations aux dérivées partielles (Cours de première année de l'ENPC, 2023).

[6] L.C. Evans, Partial differential equations (Graduate Studies in Mathematics, 19, American Mathematical Society, Providence, RI, 1998).

[7] P.D. Hislop et I.M. Sigal, Introduction to Spectral Theory with Application to Schrödinger Operators (Springer-Verlag, Applied Mathematical Science, 113, 1996).

[8] T. Kato, Perturbation Theory for Linear Operators (Springer-Verlag, 1976).

[9] F. Legoll, Equations aux dérivées partielles : approches variationnelles (Cours de première année de l'ENPC, 2021).

[10] M. Reed et B. Simon, Methods of Modern Mathematical Physics. I. Functional Analysis (Academic Press, 1980).