Problèmes d'évolution

Frédéric Legoll et Pierre Lissy

Partie 1, Février 2024

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

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

Démonstration. Cf. le cours de première année [4]. □

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 [4]. □

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 [4]. □

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*

$$\left\{ \begin{array}{l} \textit{Chercher } u \in V \textit{ tel que} \\ \forall w \in V, \quad a(u,w) = b(w) \end{array} \right. \tag{1.3}$$

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

$$\left\{\begin{array}{l}\text{Chercher } u \in V \text{ tel que} \\ J(u) = \inf\limits_{w \in V} J(w)\end{array}\right. \tag{1.4}$$

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

Démonstration. Cf. le cours de première année [7]. □

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*

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

Démonstration. Cf. le cours de première année [4]. □

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 $H^1_0(\Omega)$, 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*

$$H^1_0(\Omega) = \left\{ u \in H^1(\Omega), \quad \gamma(u) = 0 \right\}.$$

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 $u = (u_1, \ldots, u_d) \in L^2(\Omega)^d$, 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 [7]. L'exercice 2.53 en propose une autre démonstration. L'exercice 3.7 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. $\square$

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*

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

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. $\square$

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 $u \in C$. $\square$

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. $\square$

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 $u \in H_0^1(\Omega)$. $\square$

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 [4]. □

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

Exercice 2.2. *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 2.3. *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 2.4. *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 2.5. *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 2.6 (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 2.7 (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 2.8 (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 2.9 (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 2.10. *Montrer que si, dans l'Exemple 2.9, la fonction V est continue et bornée, alors $\|A\| = \sup\limits_{x\in\mathbb{R}^d} |V(x)|$.*

Concluons cette section par un résultat important.

Proposition 2.11. *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 2.12. *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 2.13. *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 2.14. *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 [8]).

Proposition 2.15. *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 2.16. *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 2.17. *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 2.18. *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 2.19 (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 [5, Section 4.2]).

Proposition 2.20. *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 2.21. *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 2.22 (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 [6, 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 2.4). 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 2.23. *Soient $\tau_{\rm d}$ et $\tau_{\rm g}$ les opérateurs de shift définis par* (2.2) *et* (2.3). *Montrer que $r(\tau_{\rm d}) = r(\tau_{\rm g}) = 1$.*

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

Lemme 2.24. *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 2.11), et d'après le cours de première année [4], 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 2.11). D'après le cours de première année [4], 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 2.25 (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 2.24 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 2.26. *Soit E un espace de Banach et $T \in \mathcal{L}(E)$. D'après la proposition 2.15, si $\lambda - T$ est bijectif, alors son inverse $(\lambda - T)^{-1}$ est continu.*

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

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

   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 2.25, 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 2.25). 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 2.24. On obtient donc une contradiction. □

Remarque 2.27. *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 2.28. *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 $(\alpha_k)_{1 \le k \le n}$ 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 $1 \le k \le n$. 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 2.29 (Autre décomposition du spectre). *Dans certains cas, il est plus commode de décomposer $\sigma(T)$ sous la forme $\sigma(T) = \sigma_{\mathrm{d}}(T) \cup \sigma_{\mathrm{ess}}(T)$, où $\sigma_{\mathrm{d}}(T) \subset \sigma_{\mathrm{p}}(T)$ est le* spectre discret*, qui est composé des valeurs propres isolées de multiplicité finie :*

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

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

Exercice 2.30 (Spectre résiduel). *On considère l'opérateur de shift à droite τ_{d} dans $\ell^2(\mathbb{N}, \mathbb{C})$ défini par* (2.2)*.*

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

Exercice 2.31 (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 $\sigma(T) = \sigma_{\mathrm{c}}(T) = [a,b]$, 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 $\varphi \in \mathrm{C}^\infty([a,b], \mathbb{C})$ valant 1 au voisinage de λ).*
3. *Montrer que $\sigma(T) = \sigma_{\mathrm{c}}(T)$. Pour cela, établir d'abord que $\sigma_{\mathrm{p}}(T) = \emptyset$, 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 $(\varphi_n)_{n \ge 1}$ de E définie par*

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

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 2.32. *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 2.15). 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 2.26 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 $\sigma_{\mathrm{r}}(T) = \emptyset$. Pour ce faire, on considère $\lambda \in \sigma(T) \subset \mathbb{R}$ tel que $\operatorname{Ker}(\lambda - T) = \{0\}$. On a vu (cf. la proposition 2.21) que

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

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

— Enfin, soient u et v deux vecteurs propres associés respectivement à deux éléments $\lambda \neq \mu$ de $\sigma_{\mathrm{p}}(T)$. 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 2.33. *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 $\lambda \in \sigma_{\mathrm{r}}(T)$, alors $\overline{\lambda} \in \sigma_{\mathrm{p}}(T^*)$. Bien sûr, dans le cas des opérateurs autoadjoints, on a $T^* = T$ et donc $\lambda \in \sigma_{\mathrm{r}}(T) \cap \sigma_{\mathrm{p}}(T) = \emptyset$ par définition des différentes parties du spectre. Ceci montre bien que $\sigma_{\mathrm{r}}(T) = \emptyset$ pour des opérateurs autoadjoints.

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

Exercice 2.34. *Donner un exemple d'opérateur lineaire et continu tel que $\overline{\lambda} \in \sigma_{\mathrm{p}}(T^*)$ lorsque $\lambda \in \sigma_{\mathrm{p}}(T)$, et un exemple d'opérateur lineaire et continu tel que $\overline{\lambda} \in \sigma_{\mathrm{r}}(T^*)$ lorsque $\lambda \in \sigma_{\mathrm{p}}(T)$.*

Exercice 2.35. *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 2.36. *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 2.37. *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 2.38. *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 2.12 : 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 2.39. *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 2.40. *On considère l'opérateur de l'Exercice 2.8. 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 2.41. *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 2.37, 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 $q > p \geq N$,

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

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 2.56 pour le cas où $E = F$ est un espace de Hilbert).

Proposition 2.42. *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 2.43 (Opérateurs de Hilbert-Schmidt)**.** *Montrer que l'opérateur $\widehat{K}$ de l'Exercice 2.19 est compact.*

Exercice 2.44. *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 2.45. *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 2.46. *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 2.47. *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*

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

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 2.48. *Soit* Ω *un ouvert* borné *de* $\mathbb{R}^d$. *L'injection canonique de* $\mathrm{H}^1_0(\Omega)$ *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 les espaces $\mathrm{L}^2(0,\pi)$ et $\mathrm{H}^1_0(0,\pi)$ par

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

et

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

De plus,

$$\|u\|_{\mathrm{L}^2} = \left(\sum_{k=1}^{+\infty} |c_k|^2 \right)^{1/2}, \qquad \|u\|_{\mathrm{H}^1} = \left(\sum_{k=1}^{+\infty} (1+k^2)|c_k|^2 \right)^{1/2}.$$

On note que, pour montrer la complétude de la base des $\{e_k\}_{k\geq 1}$ dans $\mathrm{L}^2(0,\pi)$, il suffit de prendre une fonction de $\mathrm{L}^2(0,\pi)$, de l'antisymétriser pour en faire une fonction sur $]-\pi,\pi[$, d'étendre la fonction à tout $\mathbb{R}$ en la périodisant, et enfin de développer cette fonction sur la base des sinus et cosinus (en utilisant la théorie des séries de Fourier).

Soit

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

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

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

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 2.41.

— 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

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

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 2.42) que I_Ω est lui-même un opérateur compact.

Ceci conclut la preuve. □

Remarque 2.49 (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 borné. De tels résultats existent pour des domaines bornés* réguliers*, voir par exemple [3, Théorème 7.1.7] et [2, Théorème IX.7] et les résultats ci-dessous.*

On a le résultat suivant :

Théorème 2.50 (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 2.51. *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 2.51. Si $d \geq 2$, le résultat découle directement du théorème de Rellich-Kondrachov. Si $d = 1$, on remarque que l'injection $I : H^1(\Omega) \hookrightarrow L^2(\Omega)$ est la composition de deux injections

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

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 de la Proposition 2.38.

Corollaire 2.52. *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 2.53. *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 2.54 (Diagonalisation des opérateurs auto-adjoints compacts)**.** *Soit H un espace de Hilbert séparable de dimension infinie et $T \in \mathcal{L}(H)$ un opérateur auto-adjoint 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 $\lambda \in \sigma(T)\backslash\{0\}$, 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 2.44) ; d'après la Proposition 2.32, on a que $\sigma_{\mathrm{r}}(T) = \emptyset$, donc $0 \in \sigma_{\mathrm{c}}(T)$.

Remarque 2.55. *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 2.32).*

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

1. Montrons pour commencer que
$$\sigma(T) \subset \sigma_{\mathrm{p}}(T) \cup \{0\}\,. \tag{2.13}$$
On rappelle que $\sigma(T) \subset \mathbb{R}$ par la Proposition 2.32. Pour montrer (2.13), 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 (2.13).

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 2.32, 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 2.44 indique alors que E_n est de dimension finie.

4. Les espaces E_n sont deux à deux orthogonaux (par la Proposition 2.32) et sont orthogonaux à $F = \mathrm{Ker}(T)$.
Pour le second point, on procède comme dans la preuve de la Proposition 2.32. 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 (2.13), on a $\sigma(\widetilde{T}) \subset \sigma_{\mathrm{p}}(\widetilde{T}) \cup \{0\}$. Supposons que $\sigma_{\mathrm{p}}(\widetilde{T}) \not\subset \{0\}$. Il existe alors $\lambda \in \sigma_{\mathrm{p}}(\widetilde{T})$ 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 $\lambda \in \sigma_{\mathrm{p}}(T)$. 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 $\sigma_{\mathrm{p}}(\widetilde{T}) \subset \{0\}$.

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 2.32, 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 2.56. *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 :

$$Tu = \sum_{n=1}^{+\infty} \lambda_n u_n, \tag{2.14}$$

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 2.57 (Calcul fonctionnel)**.** *Notons également que la décomposition* (2.14) *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 2.58. *Soit V un espace de Hilbert, et soit A un opérateur borné 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 2.59. *Soit V un espace de Hilbert, et soit A un opérateur borné 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 2.60 ci-dessous.

Exercice 2.60. *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*

$$\left\{ \begin{array}{l} \textit{Chercher } u \in H_0^1(\Omega) \textit{ tel que} \\ -\Delta u = f \quad \textit{dans } \mathcal{D}'(\Omega) \end{array} \right. \tag{2.15}$$

admet une unique solution. On considère l'opérateur

$$\begin{array}{rcl} A : L^2(\Omega) & \longrightarrow & L^2(\Omega) \\ f & \longmapsto & u \quad \textit{solution du problème (2.15).} \end{array}$$

Montrer que A est un opérateur borné 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 2.54 (on est dans le dernier cas évoqué dans la Remarque 2.55).

Théorème 2.61. *Soit V un espace de Hilbert de dimension infinie, et A un opérateur borné, 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 2.54 caractérise tout le spectre.

Remarque 2.62. *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 $w \in V$:*

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

Exercice 2.63. *On reprend les notations et hypothèses du théorème 2.61. 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 2.64. *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 2.61. Dans ce but, nous aurons besoin des deux lemmes suivants.

Lemme 2.65. *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 2.66. *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 2.65. 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 2.35, 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 2.66. 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 2.28, 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

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

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

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

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

Démontration du théorème 2.61. Le lemme 2.65 montre que l'ensemble des valeurs propres n'est pas vide, tandis que le lemme 2.66 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 2.46 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 2.65, 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. $\square$

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'inconnue $U(t) \in \mathbb{R}^N$ suivant :

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

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 $(\lambda_k, U_k)_{1 \le k \le N}$ 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 AU, 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 [7], 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

$$\begin{cases} V \subset H \text{ avec injection compacte au sens de la définition 2.47,} \\ V \text{ dense dans } H. \end{cases}$$

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

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

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 :*

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

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 2.61. 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 $\{0\}$. $\square$

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*

$$\begin{cases} -\Delta u_k &= \lambda_k u_k \quad \text{dans } \mathcal{D}'(\Omega), \\ \quad u_k &= 0 \qquad \text{sur } \partial\Omega. \end{cases} \tag{3.16}$$

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 2.50, 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

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

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 3.6. *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 3.7. *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 3.8 (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 3.9. *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 3.10. *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, $1 \leq m \leq J$,

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

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

Proposition 3.11 (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 $\mathcal{K}_h$). $\square$

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 3.12. *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 $(a_i)_{1\leq i\leq d+1}$. On utilise la formule de quadrature

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

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 $(\lambda_{m,h})_{1\leq m\leq J}$ 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 $1 \leq m \leq J$,

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

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 $(u_m)_{m\geq 1}$ les vecteurs propres de (3.8), et soit W_m le sous-espace vectoriel de V engendré par $(u_1, \ldots, u_m)$, qui est de dimension m.

Lemme 3.13. *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 $W_{m,h} = \text{Vect }\{\Pi_h u_1, \dots, \Pi_h u_m\}$. 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. $\square$

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 3.14. *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 $H_0^1(\Omega)$ 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 $(u_1, \dots, u_m)$ 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 3.16).

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 3.15 (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 3.16. *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 $|\lambda_{n-1}|/|\lambda_n|$:

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

Remarque 3.17. *Comme on l'a remarqué dans le théorème 3.14, 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 3.18 (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 3.19. *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 3.16. □

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 3.20. *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*

$$K_k = Vect\ \left\{r_0, Ar_0, \dots, A^k r_0\right\}.$$

Il existe un entier $k_0 \leq n-1$, appelé dimension critique de Krylov, tel que :

— *si $k \leq k_0$, alors la famille $(r_0, \dots, A^k r_0)$ 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

$$\forall j \geq 2, \quad \hat{v}_j = Av_{j-1} - \langle Av_{j-1}, v_{j-1}\rangle v_{j-1} - \|\hat{v}_{j-1}\| v_{j-2} \ \text{ et } \ v_j = \frac{\hat{v}_j}{\|\hat{v}_j\|}, \tag{3.33}$$

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 $v_1, \dots, v_k$, 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 3.21. *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 $(v_1, \dots, v_k)$ coïncide avec la base orthonormée de l'espace de Krylov K_{k-1} construite par le procédé de Gram-Schmidt appliqué à la famille $(r_0, \dots, A^{k-1} r_0)$.

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 $j-1 \le k_0$, la famille $(r_0, \dots, A^{j-1}r_0)$ 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 $(w_1, \dots, w_{j-1})$ 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 $(w_1, \dots, w_{k_0+2})$ 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 $j \le k_0 + 1$, la famille $(w_1, \dots, w_j)$ 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 $(r_0, \dots, A^{j-1}r_0)$.

On montre maintenant que $w_j = v_j$ pour tout $j \le k_0 + 1$. Comme A est symétrique, on a

$$\begin{aligned} \langle A w_p, w_{j-1} \rangle &= \langle w_p, A w_{j-1} \rangle \\ &= \langle w_p, \hat{w}_j \rangle + \sum_{i=1}^{j-1} \langle A w_{j-1}, w_i \rangle \langle w_p, w_i \rangle. \end{aligned}$$

Supposons $j \le p-1$: alors, pour les i tels que $1 \le i \le j-1$, on a $i \le p-2 < p$ et $\langle w_p, w_i \rangle = 0$. Donc, pour $j \le p-1$, on a $\langle A w_p, w_{j-1} \rangle = 0$. On voit aussi que

$$\langle A w_p, w_{p-1} \rangle = \langle w_p, \hat{w}_p \rangle = \|\hat{w}_p\|.$$

Donc la récurrence (3.36) définissant $\hat{w}_j$ se récrit

$$\begin{aligned} \hat{w}_j &= A w_{j-1} - \langle A w_{j-1}, w_{j-1} \rangle w_{j-1} - \langle A w_{j-1}, w_{j-2} \rangle w_{j-2} \\ &= A w_{j-1} - \langle A w_{j-1}, w_{j-1} \rangle w_{j-1} - \|\hat{w}_{j-1}\| w_{j-2}, \end{aligned}$$

ce qui est exactement la récurrence (3.33). Par conséquent, on a bien $w_j = v_j$ pour tout $j \le k_0 + 1$.

On montre maintenant (3.34). La colonne p de la matrice AV_k est exactement, pour $1 \le p \le k$, égale à

$$\mathrm{Col}_p(AV_k) = A v_p = \hat{v}_{p+1} + \langle A v_p, v_p \rangle v_p + \|\hat{v}_p\| v_{p-1}.$$

Un simple calcul montre que les colonnes de V_kT_k sont

$$\begin{aligned}
\forall p, \quad 2 \le p \le k-1, \quad \mathrm{Col}_p(V_kT_k) &= \hat{v}_{p+1} + \langle Av_p, v_p\rangle v_p + \|\hat{v}_p\| v_{p-1},\\
\mathrm{Col}_1(V_kT_k) &= \hat{v}_2 + \langle Av_1, v_1\rangle v_1,\\
\mathrm{Col}_k(V_kT_k) &= \langle Av_k, v_k\rangle v_k + \|\hat{v}_k\| v_{k-1}.
\end{aligned}$$

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 $V_k^tV_k = \mathrm{Id}_k$. 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 $j \le k$, 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 $1 \le m \le n$), et soient P_i les matrices de projection orthogonale sur les sous-espaces propres correspondants de A. Par construction,

$$A = \sum_{i=1}^m \lambda_i P_i, \quad \mathrm{Id}_n = \sum_{i=1}^m P_i, \quad P_iP_j = 0 \text{ si } i \ne j, \quad P_i^2 = P_i \text{ pour tout } i.$$

Lemme 3.22. *Les valeurs propres de T_{k_0+1} sont aussi valeurs propres de A.*

Réciproquement, si on suppose que $P_ir_0 \ne 0$ 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 $P_ir_0 \ne 0$ 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 3.23 ci-dessous. On démontre maintenant le lemme 3.22.

Démonstration du lemme 3.22. Soit λ valeur propre de T_{k_0+1}, et soit $y \ne 0$ 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 $(v_1, \dots, v_{k_0+1})$ 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 $(P_1 r_0, \dots, P_m r_0)$ liée : alors, par exemple, il existe $\alpha_1, \dots, \alpha_{m-1}$ 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 $(P_1 r_0, \dots, P_m r_0)$ est libre et $E_m = \text{Vect } \{P_1 r_0, \dots, P_m r_0\}$ 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 $(P_1 r_0, \dots, P_m r_0)$ est libre. On se place dans cette base. La famille $(r_0, \dots, A^{m-1} r_0)$ est représentée dans cette base par la matrice

$$M = \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \dots & \lambda_1^{m-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_m & \lambda_m^2 & \dots & \lambda_m^{m-1} \end{pmatrix},$$

qui est une matrice de Van Der Monde inversible car les λ_j sont distincts deux à deux. Donc la famille $(r_0, \dots, A^{m-1} r_0)$ 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. $\square$

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 3.23. *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 3.22. 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 ε_j = 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. $\square$

Bibliographie

[1] G. Allaire, Analyse numérique et optimisation (Editions de l'Ecole Polytechnique, 2005).

[2] H. Brézis, Analyse fonctionnelle (Dunod, Paris, 1999).

[3] E.B. Davies, Spectral Theory and Differential Operators (Cambridge University Press, 1995).

[4] V. Ehrlacher, Analyse et Equations aux dérivées partielles (Cours de première année de l'ENPC, 2023).

[5] P.D. Hislop et I.M. Sigal, Introduction to Spectral Theory with Application to Schrödinger Operators (Springer-Verlag, Applied Mathematical Science, 113, 1996).

[6] T. Kato, Perturbation Theory for Linear Operators (Springer-Verlag, 1976).

[7] F. Legoll, Equations aux dérivées partielles : approches variationnelles (Cours de première année de l'ENPC, 2021).

[8] M. Reed et B. Simon, Methods of Modern Mathematical Physics. I. Functional Analysis (Academic Press, 1980).