

Polycopié du cours MAP 431
ANALYSE VARIATIONNELLE DES ÉQUATIONS
AUX DÉRIVÉES PARTIELLES

Grégoire ALLAIRE - François ALOUGES

École Polytechnique
16 janvier 2015

Table des matières

1	FORMULATION VARIATIONNELLE DES PROBLÈMES ELLIPTIQUES	1
1.1	Généralités	1
1.2	Approche variationnelle	2
1.2.1	Formules de Green	2
1.2.2	Formulation variationnelle	5
1.3	Théorie de Lax-Milgram	7
1.3.1	Cadre abstrait	7
1.3.2	Application au Laplacien	10
2	ESPACES DE SOBOLEV	13
2.1	Introduction et avertissement	13
2.2	Fonctions de carré sommable et dérivation faible	13
2.2.1	Quelques rappels d'intégration	13
2.2.2	Dérivation faible	14
2.3	Définition et principales propriétés	16
2.3.1	Espace $H^1(\Omega)$	16
2.3.2	Espace $H_0^1(\Omega)$	18
2.3.3	Traces et formules de Green	20
2.3.4	Un résultat de compacité	23
2.3.5	Espaces $H^m(\Omega)$	23
3	ÉTUDE MATHÉMATIQUE DES PROBLÈMES ELLIPTIQUES	27
3.1	Introduction	27
3.2	Étude du Laplacien	27
3.2.1	Conditions aux limites de Dirichlet	27
3.2.2	Conditions aux limites de Neumann	33
3.2.3	Coefficients variables	38
3.2.4	Propriétés qualitatives	40
3.3	Résolution d'autres modèles	43
3.3.1	Système de l'élasticité linéarisée	43
3.3.2	Équations de Stokes	49
4	MÉTHODE DES ÉLÉMENTS FINIS	53
4.1	Approximation variationnelle	53

4.1.1	Introduction	53
4.1.2	Approximation interne générale	53
4.1.4	Méthode des éléments finis (principes généraux)	56
4.2	Éléments finis en dimension $N = 1$	56
4.2.1	Éléments finis \mathbb{P}_1	57
4.2.2	Convergence et estimation d'erreur	61
4.2.3	Éléments finis \mathbb{P}_2	64
4.2.4	Propriétés qualitatives	66
4.3	Éléments finis en dimension $N \geq 2$	67
4.3.1	Éléments finis triangulaires	68
4.3.2	Convergence et estimation d'erreur	76
4.3.3	Éléments finis rectangulaires	78
4.4	Résolution des systèmes linéaires	81
4.4.1	Rappels sur les normes matricielles	81
4.4.2	Conditionnement et stabilité	83
4.4.3	Méthodes directes	85
4.4.4	Méthodes itératives	89
4.4.5	Méthode du gradient conjugué	90
5	PROBLÈMES AUX VALEURS PROPRES	93
5.1	Motivation et exemples	93
5.1.1	Introduction	93
5.1.2	Résolution des problèmes instationnaires	94
5.3	Valeurs propres d'un problème elliptique	96
5.3.1	Problème variationnel	96
5.3.2	Valeurs propres du Laplacien	99
5.3.3	Autres modèles	101
5.4	Méthodes numériques	103
5.4.1	Discrétisation par éléments finis	103
13.2	Calcul de valeurs et vecteurs propres	106
13.2.1	Méthode de la puissance	106
13.2.2	Méthode de Givens-Householder	108

Introduction

Ce polycopié est un abrégé du cours “Analyse numérique et optimisation” dont la version intégrale [1] est publiée aux éditions de l’École Polytechnique. L’objectif de ce cours est d’introduire le lecteur au monde de la **modélisation mathématique** et de la **simulation numérique** qui ont pris une importance considérable ces dernières décennies dans tous les domaines de la science et des applications industrielles (ou sciences de l’ingénieur). La modélisation mathématique est l’art (ou la science, selon le point de vue) de représenter (ou de transformer) une réalité physique en des modèles abstraits accessibles à l’analyse et au calcul. La simulation numérique est, bien sûr, le processus qui permet de calculer sur ordinateur les solutions de ces modèles, et donc de simuler la réalité physique.

Plus que pour tout autre discipline l’ordinateur a été une révolution pour les mathématiques : il en a fait une science expérimentale ! On fait des “expériences numériques” comme d’autres font des expériences physiques, et la conception ainsi que l’analyse des méthodes de calcul sur ordinateur sont devenues une nouvelle branche des mathématiques : c’est la simulation numérique. Ces progrès ont aussi permis aux mathématiques de s’attaquer à des problèmes beaucoup plus complexes et concrets, issus de motivations immédiates industrielles ou scientifiques, auxquels on peut apporter des réponses à la fois qualitatives mais aussi quantitatives : c’est la modélisation mathématique. L’analyse numérique est donc la discipline qui conçoit et analyse les méthodes ou algorithmes de calcul numérique. C’est donc un outil essentiel pour la modélisation.

Les **objectifs de ce cours** sont multiples. Il s’agit tout d’abord de comprendre comment le point de vue variationnel permet d’aborder certains problèmes d’équations aux dérivées partielles sous un abord inhabituel. Ce point de vue s’avère riche et puissant. Il permet notamment d’introduire les éléments théoriques qui conduiront à la résolution du problème (démontrer l’existence et l’unicité de la solution dans un cadre adéquat) puis de construire la méthode des éléments finis, qui s’appuie sur les considérations théoriques de façon à fournir naturellement un moyen d’approcher la solution (qui, bien souvent, n’est pas calculable explicitement autrement).

L’ambition de ce cours est de donner les bases qui permettront aux futurs ingénieurs de bureau d’études ou de recherche et développement de créer de **nouveaux modèles** et de **nouveaux algorithmes numériques** pour des problèmes plus compliqués non discutés ici. Cependant, même ceux qui ne se destinent pas à une telle carrière ont intérêt à bien comprendre les enjeux de la simulation numérique. En effet, de nombreuses décisions industrielles ou politiques se prennent désormais sur la foi de calculs ou de simulations numériques. Il importe donc que les décideurs aient la capacité de juger de la **qualité** et de la **fiabilité** des calculs qui leur sont présentés. Ce cours leur permettra de connaître les premiers critères qui garantissent la validité et la pertinence des simulations numériques.

Ce cours est d’un niveau introductif et n’exige aucun autre prérequis que le niveau de connaissances acquis en classes préparatoires ou en premier cycle universitaire.

Chapitre 1

FORMULATION VARIATIONNELLE DES PROBLÈMES ELLIPTIQUES

1.1 Généralités

Dans ce chapitre nous nous intéressons à l'analyse mathématique des **équations aux dérivées partielles de type elliptique** qui correspondent à des modèles physiques stationnaires, c'est-à-dire indépendants du temps. Nous allons montrer que les problèmes aux limites sont bien posés pour ces e.d.p. elliptiques, c'est-à-dire qu'elles admettent une solution, unique, et dépendant continûment des données. L'approche que nous allons suivre est appelée **approche variationnelle**. Disons tout de suite que l'intérêt de cette approche dépasse, et de loin, le cadre des e.d.p. elliptiques et même le cadre d'analyse mathématique "pure" auquel nous nous restreignons pour l'instant. En effet, nous reprendrons cette approche variationnelle pour les problèmes d'évolution en temps (e.d.p. de type parabolique ou hyperbolique), et elle sera cruciale pour comprendre la méthode numérique des éléments finis que nous développerons au Chapitre 4. Par ailleurs, cette approche admet une interprétation physique ou mécanique très naturelle. Autant dire que le lecteur ne peut pas faire l'économie de la présentation qui suit de cette approche variationnelle!

Au cours de ce chapitre et des suivants, l'exemple prototype d'équation aux dérivées partielles de type elliptique sera le Laplacien pour lequel nous étudierons le problème aux limites suivant

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \quad (1.1)$$

où nous imposons des conditions aux limites de Dirichlet. Dans (1.1), Ω est un ouvert de l'espace \mathbb{R}^N , $\partial\Omega$ est son bord (ou frontière), f est un second membre (une donnée du problème), et u est l'inconnue. Bien sûr, nous donnerons au Chapitre 3 de nombreux autres exemples d'équations aux dérivées partielles de type elliptique qui peuvent s'étudier grâce à l'approche variationnelle.

Définition 1.1.1 Soit Ω un ouvert de \mathbb{R}^N , $\overline{\Omega}$ sa fermeture. On note $C(\Omega)$ (respectivement, $C(\overline{\Omega})$) l'espace des fonctions continues dans Ω (respectivement, dans $\overline{\Omega}$). Soit un entier $k \geq 0$. On note $C^k(\Omega)$ (respectivement, $C^k(\overline{\Omega})$) l'espace des fonctions k fois continûment dérivables dans Ω (respectivement, dans $\overline{\Omega}$).

Une **solution classique** (on parle aussi de **solution forte**) de (1.1) est une solution $u \in C^2(\Omega) \cap C(\overline{\Omega})$, ce qui implique que le second membre f doit appartenir à $C(\Omega)$. Cette formulation classique pose malheureusement un certain nombre de problèmes pour démontrer l'existence d'une solution. C'est pourquoi nous remplacerons la formulation classique de (1.1) par une formulation, dite variationnelle, beaucoup plus avantageuse.

Le plan de ce chapitre est le suivant. Dans la Section 1.2 nous rappelons quelques formules d'intégration par parties, dites **formules de Green**, puis nous définissons ce qu'est une **formulation variationnelle**. La Section 1.3 est consacrée au **théorème de Lax-Milgram** qui sera l'outil essentiel permettant de démontrer des résultats d'existence et d'unicité de solutions de formulation variationnelle. Nous verrons que pour pouvoir appliquer ce théorème il est inéluctable de devoir abandonner l'espace $C^1(\overline{\Omega})$ des fonctions continûment différentiables au profit de sa "généralisation", l'espace de Sobolev $H^1(\Omega)$.

1.2 Approche variationnelle

Le principe de l'approche variationnelle pour la résolution des équations aux dérivées partielles est de remplacer l'équation par une formulation équivalente, dite variationnelle, obtenue en intégrant l'équation multipliée par une fonction quelconque, dite test. Comme il est nécessaire de procéder à des intégrations par parties dans l'établissement de la formulation variationnelle, nous commençons par donner quelques résultats essentiels à ce sujet.

1.2.1 Formules de Green

Dans toute cette sous-section Ω est un ouvert de l'espace \mathbb{R}^N (borné ou non), dont le bord (ou la frontière) est noté $\partial\Omega$. Nous supposons aussi que Ω est un ouvert **régulier** de classe C^1 . La définition précise d'un ouvert régulier est donné plus bas dans la Définition 1.2.5, mais sa connaissance n'est absolument pas nécessaire pour la bonne compréhension de la suite de ce cours. Il suffit juste de savoir qu'un ouvert régulier est *grosso modo* un ouvert dont le bord est une hypersurface (une variété de dimension $N - 1$) régulière, et que cet ouvert est localement situé d'un seul côté de sa frontière. On définit alors la **normale extérieure** au bord $\partial\Omega$ comme étant le vecteur unité $n = (n_i)_{1 \leq i \leq N}$ normal en tout point au plan tangent de Ω et pointant vers l'extérieur de Ω . Dans $\Omega \subset \mathbb{R}^N$ on note dx la mesure volumique, ou mesure de Lebesgue de dimension N . Sur $\partial\Omega$, on note ds la mesure surfacique, ou mesure de Lebesgue de dimension $N - 1$ sur la variété $\partial\Omega$. Le résultat principal de cette sous-section est le théorème suivant que nous admettrons (voir le théorème 5.4.9 dans [6]).

Théorème 1.2.1 (Formule de Green) Soit Ω un ouvert régulier de classe C^1 . Soit w une fonction de $C^1(\overline{\Omega})$ à support borné dans le fermé $\overline{\Omega}$. Alors elle vérifie la formule de Green

$$\int_{\Omega} \frac{\partial w}{\partial x_i}(x) dx = \int_{\partial\Omega} w(x)n_i(x) ds, \quad (1.5)$$

où n_i est la i -ème composante de la normale extérieure unité de Ω .

Remarque 1.2.2 Dire qu'une fonction régulière w a son support borné dans le fermé $\overline{\Omega}$ veut dire qu'elle s'annule à l'infini si le fermé n'est pas borné. On dit aussi que la fonction w a un support compact dans $\overline{\Omega}$ (attention : cela n'implique pas que w s'annule sur le bord $\partial\Omega$). En particulier, l'hypothèse du Théorème 1.2.1 à propos du support borné de la fonction w dans $\overline{\Omega}$ est inutile si l'ouvert Ω est borné. Si Ω n'est pas borné, cette hypothèse assure que les intégrales dans (1.5) sont finies. •

Le Théorème 1.2.1 a de nombreux corollaires qui sont tous des conséquences immédiates de la formule de Green (1.5). Le lecteur qui voudra économiser sa mémoire ne retiendra donc que la formule de Green (1.5)!

Corollaire 1.2.3 (Formule d'intégration par parties) Soit Ω un ouvert régulier de classe C^1 . Soit u et v deux fonctions de $C^1(\overline{\Omega})$ à support borné dans le fermé $\overline{\Omega}$. Alors elles vérifient la formule d'intégration par parties

$$\int_{\Omega} u(x) \frac{\partial v}{\partial x_i}(x) dx = - \int_{\Omega} v(x) \frac{\partial u}{\partial x_i}(x) dx + \int_{\partial\Omega} u(x)v(x)n_i(x) ds. \quad (1.6)$$

Démonstration. Il suffit de prendre $w = uv$ dans le Théorème 1.2.1. \square

Corollaire 1.2.4 Soit Ω un ouvert régulier de classe C^1 . Soit u une fonction de $C^2(\overline{\Omega})$ et v une fonction de $C^1(\overline{\Omega})$, toutes deux à support borné dans le fermé $\overline{\Omega}$. Alors elles vérifient la formule d'intégration par parties

$$\int_{\Omega} \Delta u(x)v(x) dx = - \int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx + \int_{\partial\Omega} \frac{\partial u}{\partial n}(x)v(x) ds, \quad (1.7)$$

où $\nabla u = \left(\frac{\partial u}{\partial x_i} \right)_{1 \leq i \leq N}$ est le vecteur gradient de u , et $\frac{\partial u}{\partial n} = \nabla u \cdot n$.

Démonstration. On applique le Corollaire 1.2.3 à v et $\frac{\partial u}{\partial x_i}$ et on somme en i . \square

Définition 1.2.5 On dit qu'un ouvert Ω de \mathbb{R}^N est régulier de classe C^k (avec un entier $k \geq 1$) s'il existe un nombre fini d'ouverts $(\omega_i)_{0 \leq i \leq I}$ tels que

$$\overline{\omega_0} \subset \Omega, \quad \overline{\Omega} \subset \cup_{i=0}^I \omega_i, \quad \partial\Omega \subset \cup_{i=1}^I \omega_i,$$

et que, pour chaque $i \in \{1, \dots, I\}$ (voir la Figure 1.1), il existe une application bijective ϕ_i de classe C^k , de ω_i dans l'ensemble

$$Q = \{y = (y', y_N) \in \mathbb{R}^{N-1} \times \mathbb{R}, |y'| < 1, |y_N| < 1\},$$

dont l'inverse est aussi de classe C^k , et telle que

$$\begin{aligned} \phi_i(\omega_i \cap \Omega) &= Q \cap \{y = (y', y_N) \in \mathbb{R}^{N-1} \times \mathbb{R}, y_N > 0\} = Q^+, \\ \phi_i(\omega_i \cap \partial\Omega) &= Q \cap \{y = (y', y_N) \in \mathbb{R}^{N-1} \times \mathbb{R}, y_N = 0\}. \end{aligned}$$

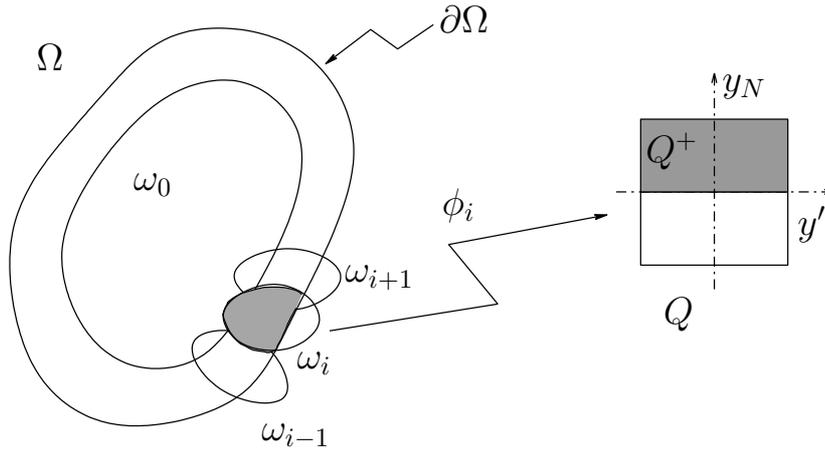


FIGURE 1.1 – Définition de la régularité d'un ouvert.

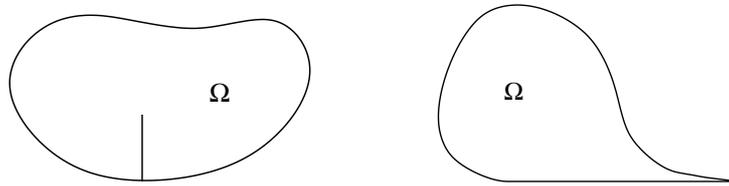


FIGURE 1.2 – Deux exemples d'ouvert non régulier : ouvert fissuré à gauche, ouvert avec un point de rebroussement à droite.

Remarque 1.2.6 Bien que la Figure 1.1 représente un ouvert régulier qui est borné, la Définition 1.2.5 s'applique aussi à des ouverts non bornés. La Définition 1.2.5 n'exclut pas seulement les ouverts dont le bord n'est pas une surface régulière, mais elle exclut aussi les ouverts qui ne sont pas localement situés d'un seul côté de leur frontière. La Figure 1.2 contient deux exemples typiques d'ouverts non réguliers qui présentent une singularité irrémédiable, soit le long de la fissure, soit en un point de rebroussement. Ces exemples ne sont pas des "inventions mathématiques" : l'ouvert fissuré est typiquement utilisé pour étudier les problèmes de fissures en mécanique des structures. On peut néanmoins généraliser un peu la classe des ouverts réguliers aux ouverts "réguliers par morceaux", à condition que ces morceaux de frontières se "recollent" en formant des angles différents de 0 (cas d'un point de rebroussement) ou de 2π (cas d'une fissure). •

Exercice 1.2.1 Dédurre de la formule de Green (1.5) la formule de Stokes

$$\int_{\Omega} \operatorname{div} \sigma(x) \phi(x) dx = - \int_{\Omega} \sigma(x) \cdot \nabla \phi(x) dx + \int_{\partial \Omega} \sigma(x) \cdot n(x) \phi(x) ds,$$

où ϕ est une fonction scalaire de $C^1(\overline{\Omega})$ et σ une fonction à valeurs vectorielles de $C^1(\overline{\Omega})$, à supports bornés dans le fermé $\overline{\Omega}$.

Exercice 1.2.2 En dimension $N = 3$ on définit le rotationnel d'une fonction de Ω dans

\mathbb{R}^3 , $\phi = (\phi_1, \phi_2, \phi_3)$, comme la fonction de Ω dans \mathbb{R}^3 définie par

$$\operatorname{rot}\phi = \left(\frac{\partial\phi_3}{\partial x_2} - \frac{\partial\phi_2}{\partial x_3}, \frac{\partial\phi_1}{\partial x_3} - \frac{\partial\phi_3}{\partial x_1}, \frac{\partial\phi_2}{\partial x_1} - \frac{\partial\phi_1}{\partial x_2} \right).$$

Pour ϕ et ψ , fonctions à valeurs vectorielles de $C^1(\overline{\Omega})$, à supports bornés dans le fermé $\overline{\Omega}$, déduire de la formule de Green (1.5)

$$\int_{\Omega} \operatorname{rot}\phi \cdot \psi \, dx - \int_{\Omega} \phi \cdot \operatorname{rot}\psi \, dx = - \int_{\partial\Omega} (\phi \times n) \cdot \psi \, ds.$$

1.2.2 Formulation variationnelle

Pour simplifier la présentation, nous supposons que l'ouvert Ω est borné et régulier, et que le second membre f de (1.1) est continu sur $\overline{\Omega}$. Le résultat principal de cette sous-section est la proposition suivante.

Proposition 1.2.7 *Soit u une fonction de $C^2(\overline{\Omega})$. Soit X l'espace défini par*

$$X = \{ \phi \in C^1(\overline{\Omega}) \text{ tel que } \phi = 0 \text{ sur } \partial\Omega \}.$$

Alors u est une solution du problème aux limites (1.1) si et seulement si u appartient à X et vérifie l'égalité

$$\int_{\Omega} \nabla u(x) \cdot \nabla v(x) \, dx = \int_{\Omega} f(x)v(x) \, dx \text{ pour toute fonction } v \in X. \quad (1.8)$$

*L'égalité (1.8) est appelée la **formulation variationnelle** du problème aux limites (1.1).*

Remarque 1.2.8 Un intérêt immédiat de la formulation variationnelle (1.8) est qu'elle a un sens si la solution u est seulement une fonction de $C^1(\overline{\Omega})$, contrairement à la formulation "classique" (1.1) qui requiert que u appartienne à $C^2(\overline{\Omega})$. On pressent donc déjà qu'il est plus simple de résoudre (1.8) que (1.1) puisqu'on est moins exigeant sur la régularité de la solution.

Dans la formulation variationnelle (1.8), la fonction v est appelée **fonction test**. La formulation variationnelle est aussi parfois appelée formulation faible du problème aux limites (1.1). En mécanique, la formulation variationnelle est connue sous le nom de "principe des travaux virtuels". En physique, on parle aussi d'équation de bilan ou de formule de réciprocité.

Lorsqu'on prend $v = u$ dans (1.8), on obtient ce qu'il est convenu d'appeler une **égalité d'énergie**, qui exprime généralement l'égalité entre une énergie stockée dans le domaine Ω (le terme de gauche de (1.8)) et une énergie potentielle associée à f (le terme de droite de (1.8)). •

Démonstration. Si u est solution du problème aux limites (1.1), on multiplie l'équation par $v \in X$ et on utilise la formule d'intégration par parties du Corollaire 1.2.4

$$\int_{\Omega} \Delta u(x)v(x) \, dx = - \int_{\Omega} \nabla u(x) \cdot \nabla v(x) \, dx + \int_{\partial\Omega} \frac{\partial u}{\partial n}(x)v(x) \, ds.$$

Or $v = 0$ sur $\partial\Omega$ puisque $v \in X$, donc

$$\int_{\Omega} f(x)v(x) dx = \int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx,$$

qui n'est rien d'autre que la formule (1.8). Réciproquement, si $u \in X$ vérifie (1.8), en utilisant "à l'envers" la formule d'intégration par parties précédente on obtient

$$\int_{\Omega} (\Delta u(x) + f(x))v(x) dx = 0 \text{ pour toute fonction } v \in X.$$

Comme $(\Delta u + f)$ est une fonction continue, grâce au Lemme 1.2.9 on conclut que $-\Delta u(x) = f(x)$ pour tout $x \in \Omega$. Par ailleurs, comme $u \in X$, on retrouve la condition aux limites $u = 0$ sur $\partial\Omega$, c'est-à-dire que u est solution du problème aux limites (1.1). \square

Lemme 1.2.9 *Soit Ω un ouvert de \mathbb{R}^N . Soit $g(x)$ une fonction continue dans Ω . Si pour toute fonction ϕ de $C^\infty(\Omega)$ à support compact dans Ω , on a*

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

alors la fonction g est nulle dans Ω .

Démonstration. Supposons qu'il existe un point $x_0 \in \Omega$ tel que $g(x_0) \neq 0$. Sans perte de généralité, on peut supposer que $g(x_0) > 0$ (sinon on prend $-g$). Par continuité, il existe un petit voisinage ouvert $\omega \subset \Omega$ de x_0 tel que $g(x) > 0$ pour tout $x \in \omega$. Soit alors une fonction test positive, non nulle, ϕ à support inclus dans ω . On a

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

qui est une contradiction avec l'hypothèse sur g . Donc $g(x) = 0$ pour tout $x \in \Omega$. \square

Remarque 1.2.10 En notation compacte on peut réécrire la formulation variationnelle (1.8) sous la forme : trouver $u \in X$ tel que

$$a(u, v) = L(v) \text{ pour toute fonction } v \in X,$$

avec

$$a(u, v) = \int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx$$

et

$$L(v) = \int_{\Omega} f(x)v(x) dx,$$

où $a(\cdot, \cdot)$ est une forme bilinéaire sur X et $L(\cdot)$ est une forme linéaire sur X . C'est sous cette forme abstraite que nous résoudrons (avec quelques hypothèses) la formulation variationnelle dans la prochaine section. \bullet

L'idée principale de **l'approche variationnelle** est de montrer l'existence et l'unicité de la solution de la formulation variationnelle (1.8), ce qui entraînera le même résultat pour l'équation (1.1) à cause de la Proposition 1.2.7. En effet, nous allons voir qu'il existe une théorie à la fois simple et puissante pour analyser les formulations variationnelles. Néanmoins cette théorie ne fonctionne que si l'espace dans lequel on cherche la solution et dans lequel on prend les fonctions tests (dans les notations précédentes, l'espace X) est un espace de Hilbert, ce qui n'est pas le cas pour $X = \{v \in C^1(\overline{\Omega}), v = 0 \text{ sur } \partial\Omega\}$ muni du produit scalaire "naturel" pour ce problème. La principale difficulté dans l'application de l'approche variationnelle sera donc qu'il faudra utiliser un autre espace que X , à savoir l'espace de Sobolev $H_0^1(\Omega)$ qui est bien un espace de Hilbert (voir le Chapitre 2).

Exercice 1.2.3 On considère le Laplacien avec condition aux limites de Neumann

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} = 0 & \text{sur } \partial\Omega. \end{cases} \quad (1.9)$$

Soit u une fonction de $C^2(\overline{\Omega})$. Montrer que u est une solution du problème aux limites (1.9) si et seulement si u appartient à $C^1(\overline{\Omega})$ et vérifie l'égalité

$$\int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx = \int_{\Omega} f(x)v(x) dx \text{ pour toute fonction } v \in C^1(\overline{\Omega}). \quad (1.10)$$

En déduire qu'une condition nécessaire d'existence d'une solution dans $C^2(\overline{\Omega})$ de (1.9) est que $\int_{\Omega} f(x)dx = 0$.

Exercice 1.2.4 On considère l'équation des plaques

$$\begin{cases} \Delta(\Delta u) = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \\ \frac{\partial u}{\partial n} = 0 & \text{sur } \partial\Omega \end{cases} \quad (1.11)$$

On note X l'espace des fonctions v de $C^2(\overline{\Omega})$ telles que v et $\frac{\partial v}{\partial n}$ s'annulent sur $\partial\Omega$. Soit u une fonction de $C^4(\overline{\Omega})$. Montrer que u est une solution du problème aux limites (1.11) si et seulement si u appartient à X et vérifie l'égalité

$$\int_{\Omega} \Delta u(x)\Delta v(x) dx = \int_{\Omega} f(x)v(x) dx \text{ pour toute fonction } v \in X. \quad (1.12)$$

1.3 Théorie de Lax-Milgram

1.3.1 Cadre abstrait

Nous décrivons une théorie abstraite pour obtenir l'existence et l'unicité de la solution d'une formulation variationnelle dans un espace de Hilbert réel V . Rappelons qu'un espace de Hilbert réel est un espace vectoriel sur \mathbb{R} , muni d'un produit scalaire, noté $\langle x, y \rangle$, qui est complet pour la norme associée à ce produit scalaire,

notée $\|x\| = \sqrt{\langle x, x \rangle}$. (Un espace vectoriel normé est complet si toute suite de Cauchy est une suite convergente dont la limite appartient à cet espace.) Suivant la Remarque 1.2.10 nous considérons une formulation variationnelle du type :

$$\text{trouver } u \in V \text{ tel que } a(u, v) = L(v) \text{ pour toute fonction } v \in V. \quad (1.13)$$

Les hypothèses sur a et L sont

1. $L(\cdot)$ est une forme linéaire continue sur V , c'est-à-dire que $v \rightarrow L(v)$ est linéaire de V dans \mathbb{R} et il existe $C > 0$ tel que

$$|L(v)| \leq C\|v\| \text{ pour tout } v \in V;$$

2. $a(\cdot, \cdot)$ est une forme bilinéaire sur V , c'est-à-dire que $w \rightarrow a(w, v)$ est une forme linéaire de V dans \mathbb{R} pour tout $v \in V$, et $v \rightarrow a(w, v)$ est une forme linéaire de V dans \mathbb{R} pour tout $w \in V$;

3. $a(\cdot, \cdot)$ est continue, c'est-à-dire qu'il existe $M > 0$ tel que

$$|a(w, v)| \leq M\|w\| \|v\| \text{ pour tout } w, v \in V; \quad (1.14)$$

4. $a(\cdot, \cdot)$ est **coercive** (ou elliptique), c'est-à-dire qu'il existe $\nu > 0$ tel que

$$a(v, v) \geq \nu\|v\|^2 \text{ pour tout } v \in V. \quad (1.15)$$

Comme nous le verrons au cours de cette sous-section, toutes les hypothèses ci-dessus sont nécessaires pour pouvoir résoudre (1.13). En particulier, la coercivité de $a(\cdot, \cdot)$ est essentielle.

Théorème 1.3.1 (Lax-Milgram) *Soit V un espace de Hilbert réel, $L(\cdot)$ une forme linéaire continue sur V , $a(\cdot, \cdot)$ une forme bilinéaire continue coercive sur V . Alors la formulation variationnelle (1.13) admet une unique solution. De plus cette solution dépend continûment de la forme linéaire L .*

Démonstration. Pour tout $w \in V$, l'application $v \rightarrow a(w, v)$ est une forme linéaire continue sur V : par conséquent, le théorème de représentation de Riesz entraîne qu'il existe un élément de V , noté $A(w)$, tel que

$$a(w, v) = \langle A(w), v \rangle \text{ pour tout } v \in V.$$

Par ailleurs, la bilinéarité de $a(w, v)$ implique évidemment la linéarité de l'application $w \rightarrow A(w)$. De plus, en prenant $v = A(w)$, la continuité (1.14) de $a(w, v)$ montre que

$$\|A(w)\|^2 = a(w, A(w)) \leq M\|w\| \|A(w)\|,$$

c'est-à-dire que $\|A(w)\| \leq M\|w\|$ et donc $w \rightarrow A(w)$ est continue. Une autre application du Théorème de représentation de Riesz implique qu'il existe un élément de V , noté f , tel que $\|f\|_V = \|L\|_{V'}$ et

$$L(v) = \langle f, v \rangle \text{ pour tout } v \in V.$$

Finalement, le problème variationnel (1.13) est équivalent à :

$$\text{trouver } u \in V \text{ tel que } A(u) = f. \quad (1.16)$$

Pour démontrer le théorème il nous faut donc montrer que l'opérateur A est bijectif de V dans V (ce qui implique l'existence et l'unicité de u) et que son inverse est continu (ce qui prouve la dépendance continue de u par rapport à L).

La coercivité (1.15) de $a(w, v)$ montre que

$$\nu \|w\|^2 \leq a(w, w) = \langle A(w), w \rangle \leq \|A(w)\| \|w\|,$$

ce qui donne

$$\nu \|w\| \leq \|A(w)\| \text{ pour tout } w \in V, \quad (1.17)$$

c'est-à-dire que A est injectif. Pour montrer que A est surjectif, c'est-à-dire que $\text{Im}(A) = V$ (ce qui n'est pas évident si V est de dimension infinie), il suffit de montrer que $\text{Im}(A)$ est fermé dans V et que $\overline{\text{Im}(A)}^\perp = \{0\}$. En effet, dans ce cas on voit que $V = \{0\}^\perp = (\text{Im}(A)^\perp)^\perp = \overline{\text{Im}(A)} = \text{Im}(A)$, ce qui prouve bien que A est surjectif. Soit $A(w_n)$ une suite dans $\text{Im}(A)$ qui converge vers b dans V . En vertu de (1.17) on a

$$\nu \|w_n - w_p\| \leq \|A(w_n) - A(w_p)\|$$

qui tend vers zéro quand n et p tendent vers l'infini. Donc w_n est une suite de Cauchy dans l'espace de Hilbert V , c'est-à-dire qu'elle converge vers une limite $w \in V$. Alors, par continuité de A on en déduit que $A(w_n)$ converge vers $A(w) = b$, c'est-à-dire que $b \in \text{Im}(A)$ et $\text{Im}(A)$ est donc fermé. D'autre part, soit $v \in \text{Im}(A)^\perp$; la coercivité (1.15) de $a(w, v)$ implique que

$$\nu \|v\|^2 \leq a(v, v) = \langle A(v), v \rangle = 0,$$

c'est-à-dire que $v = 0$ et $\text{Im}(A)^\perp = \{0\}$, ce qui prouve que A est bijectif. Soit A^{-1} son inverse : l'inégalité (1.17) avec $w = A^{-1}(v)$ prouve que A^{-1} est continu, donc la solution u dépend continûment de f . \square

Remarque 1.3.2 Si l'espace de Hilbert V est de dimension finie (ce qui n'est cependant jamais le cas pour les applications que nous visons), la démonstration du Théorème 1.3.1 de Lax-Milgram se simplifie considérablement. En effet, en dimension finie toutes les applications linéaires sont continues et l'injectivité (1.17) de A est équivalent à son inversibilité. On voit bien dans ce cas (comme dans le cas général) que l'hypothèse de coercivité de la forme bilinéaire $a(w, v)$ est essentielle puisque c'est elle qui donne l'injectivité de A . Remarquons pour finir que, si $V = \mathbb{R}^N$, une formulation variationnelle n'est que l'écriture, $\langle Au, v \rangle = \langle f, v \rangle$ pour tout $v \in \mathbb{R}^N$, d'un simple système linéaire $Au = f$. \bullet

Une formulation variationnelle possède souvent une interprétation physique, en particulier si la forme bilinéaire est symétrique. En effet dans ce cas, la solution de la formulation variationnelle (1.13) réalise le **minimum d'une énergie** (très naturelle en physique ou en mécanique).

Proposition 1.3.4 *On se place sous les hypothèses du Théorème 1.3.1 de Lax-Milgram. On suppose en plus que la forme bilinéaire est symétrique $a(w, v) = a(v, w)$ pour tout $v, w \in V$. Soit $J(v)$ l'énergie définie pour $v \in V$ par*

$$J(v) = \frac{1}{2}a(v, v) - L(v). \quad (1.18)$$

Soit $u \in V$ la solution unique de la formulation variationnelle (1.13). Alors u est aussi l'unique point de minimum de l'énergie, c'est-à-dire que

$$J(u) = \min_{v \in V} J(v).$$

Réciproquement, si $u \in V$ est un point de minimum de l'énergie $J(v)$, alors u est la solution unique de la formulation variationnelle (1.13).

Démonstration. Si u est solution de la formulation variationnelle (1.13), on développe (grâce à la symétrie de a)

$$J(u + v) = J(u) + \frac{1}{2}a(v, v) + a(u, v) - L(v) = J(u) + \frac{1}{2}a(v, v) \geq J(u).$$

Comme $u + v$ est quelconque dans V , u minimise bien l'énergie J dans V . Réciproquement, soit $u \in V$ tel que

$$J(u) = \min_{v \in V} J(v).$$

Pour $v \in V$ on définit une fonction $j(t) = J(u + tv)$ de \mathbb{R} dans \mathbb{R} (il s'agit d'un polynôme du deuxième degré en t). Comme $t = 0$ est un minimum de j , on en déduit que $j'(0) = 0$ qui, par un calcul simple, est exactement la formulation variationnelle (1.13). \square

1.3.2 Application au Laplacien

Essayons d'appliquer le Théorème 1.3.1 de Lax-Milgram à la formulation variationnelle (1.8) du Laplacien avec conditions aux limites de Dirichlet. Celle-ci s'écrit bien sous la forme (1.13) avec

$$a(u, v) = \int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx$$

et

$$L(v) = \int_{\Omega} f(x)v(x) dx,$$

où clairement $a(\cdot, \cdot)$ est une forme bilinéaire, et $L(\cdot)$ une forme linéaire. L'espace V (noté précédemment X) est

$$V = \{v \in C^1(\overline{\Omega}), v = 0 \text{ sur } \partial\Omega\}. \quad (1.19)$$

Comme produit scalaire sur V nous choisissons

$$\langle w, v \rangle = \int_{\Omega} \nabla w(x) \cdot \nabla v(x) dx, \quad (1.20)$$

qui a pour norme associée

$$\|v\| = \left(\int_{\Omega} |\nabla v(x)|^2 dx \right)^{1/2}.$$

On vérifie aisément que (1.20) définit un produit scalaire sur V : le seul point qui mérite de s'y attarder est la propriété $\|v\| = 0 \Rightarrow v = 0$. En effet, de l'égalité

$$\int_{\Omega} |\nabla v(x)|^2 dx = 0$$

on déduit que v est une constante dans Ω , et comme $v = 0$ sur $\partial\Omega$ on a bien $v = 0$. La motivation du choix de (1.20) comme produit scalaire est bien sûr le fait que la forme bilinéaire $a(\cdot, \cdot)$ est **automatiquement coercive** pour (1.20). On vérifie par ailleurs aisément que a est continue. Pour montrer que L est continue, il faut faire appel à l'inégalité de Poincaré du Lemme 1.3.5 : on a alors

$$\left| \int_{\Omega} f(x)v(x) dx \right| \leq \left(\int_{\Omega} |f(x)|^2 dx \right)^{1/2} \left(\int_{\Omega} |v(x)|^2 dx \right)^{1/2} \leq C\|v\|,$$

où C est une constante qui dépend de f mais pas de v . Donc L est continue sur V . Toutes les hypothèses du Théorème 1.3.1 de Lax-Milgram semblent vérifiées, et pourtant il en manque une qui empêche son application : l'espace V n'est pas un espace de Hilbert car il n'est pas complet pour la norme induite par (1.20) ! L'obstruction ne vient pas tant du choix du produit scalaire que de l'exigence de régularité C^1 des fonctions de l'espace V . Une façon immédiate, quoique peu explicite, de résoudre la difficulté est de remplacer V par \overline{V} , sa fermeture pour le produit scalaire (1.20). Évidemment, on n'a fait que déplacer la difficulté : à quoi peut bien ressembler l'espace \overline{V} ? La réponse sera apporté au Chapitre 2 : \overline{V} est l'espace de Sobolev $H_0^1(\Omega)$ dont les éléments ne sont plus des fonctions régulières mais seulement mesurables. Une autre difficulté sera de voir en quel sens la Proposition 1.2.7 (qui exprime l'équivalence entre le problème aux limites (1.1) et sa formulation variationnelle (1.8)) reste vrai lorsque on remplace l'espace V par \overline{V} .

Nous espérons avoir ainsi convaincu le lecteur du **caractère naturel et inéluctable des espaces de Sobolev dans la résolution des formulations variationnelles** d'équations aux dérivées partielles elliptiques. Terminons ce chapitre par un lemme technique, appelé inégalité de Poincaré, que nous avons utilisé un peu plus haut.

Lemme 1.3.5 *Soit Ω un ouvert de \mathbb{R}^N borné dans au moins une direction de l'espace. Il existe une constante $C > 0$ telle que, pour toute fonction $v \in C^1(\overline{\Omega})$ qui s'annule sur le bord $\partial\Omega$,*

$$\int_{\Omega} |v(x)|^2 dx \leq C \int_{\Omega} |\nabla v(x)|^2 dx.$$

Démonstration. L'hypothèse sur le caractère borné de Ω dit (après une éventuelle rotation) que pour tout $x \in \Omega$ la première composante x_1 est bornée, $-\infty < a \leq$

$x_1 \leq b < +\infty$. Soit v une fonction de $C^1(\overline{\Omega})$ qui est nulle sur $\partial\Omega$. On peut l'étendre par continuité par zéro en dehors de Ω (v est alors une fonction continue de classe C^1 par morceaux dans \mathbb{R}^N) et écrire, pour $x \in \Omega$,

$$v(x) = \int_a^{x_1} \frac{\partial v}{\partial x_1}(t, x_2, \dots, x_N) dt,$$

d'où l'on déduit par l'inégalité de Cauchy-Schwarz

$$|v(x)|^2 \leq (x_1 - a) \int_a^{x_1} \left| \frac{\partial v}{\partial x_1}(t, x_2, \dots, x_N) \right|^2 dt \leq (b - a) \int_a^b \left| \frac{\partial v}{\partial x_1}(t, x_2, \dots, x_N) \right|^2 dt.$$

Intégrant sur Ω on obtient

$$\int_{\Omega} |v(x)|^2 dx \leq (b - a) \int_{\Omega} \int_a^b \left| \frac{\partial v}{\partial x_1}(t, x_2, \dots, x_N) \right|^2 dt dx,$$

et permutant les deux intégrations par rapport à t et x_1 , on conclut

$$\int_{\Omega} |v(x)|^2 dx \leq (b - a)^2 \int_{\Omega} \left| \frac{\partial v}{\partial x_1}(x) \right|^2 dx \leq (b - a)^2 \int_{\Omega} |\nabla v(x)|^2 dx.$$

□

Exercice 1.3.1 Le but de cet exercice est de montrer que l'espace V , défini par (1.19) et muni du produit scalaire (1.20), n'est pas complet. Soit Ω la boule unité ouverte de \mathbb{R}^N . Si $N = 1$, on définit la suite

$$u_n(x) = \begin{cases} -x - 1 & \text{si } -1 < x < -n^{-1}, \\ (n/2)x^2 - 1 + 1/(2n) & \text{si } -n^{-1} \leq x \leq n^{-1}, \\ x - 1 & \text{si } n^{-1} < x < 1. \end{cases}$$

Si $N = 2$, pour $0 < \alpha < 1/2$, on définit la suite

$$u_n(x) = |\log(|x|^2 + n^{-1})|^{\alpha/2} - |\log(1 + n^{-1})|^{\alpha/2}.$$

Si $N \geq 3$, pour $0 < \beta < (N - 2)/2$, on définit la suite

$$u_n(x) = \frac{1}{(|x|^2 + n^{-1})^{\beta/2}} - \frac{1}{(1 + n^{-1})^{\beta/2}}.$$

Montrer que la suite u_n est de Cauchy dans V mais qu'elle ne converge pas dans V lorsque n tend vers l'infini.

Chapitre 2

ESPACES DE SOBOLEV

2.1 Introduction et avertissement

Dans ce chapitre nous définissons les espaces de Sobolev qui sont **les espaces “naturels” de fonctions permettant de résoudre les formulations variationnelles d’équations aux dérivées partielles**. Physiquement, les espaces de Sobolev s’interprètent comme des espaces de **fonctions d’énergie finie**. Ce chapitre est le plus “technique” de cet ouvrage et relève en partie d’un cours de mathématiques “pures”. Il n’est pas nécessaire de connaître les démonstrations de tous les résultats de ce chapitre (sauf pour les plus simples et les plus utiles) : ce qui importe ici, c’est **l’esprit des résultats plus que la lettre de leurs démonstrations**.

Le plan de ce chapitre est le suivant. Comme les espaces de Sobolev se construisent à partir de la notion de fonction mesurable et de l’espace L^2 des fonctions de carrés sommables, la Section 2.2 donne quelques rappels à ce sujet. On y introduit aussi la notion de **dérivation faible**. La Section 2.3 contient toutes les définitions et les résultats qu’il faut absolument connaître sur les espaces de Sobolev pour suivre le reste du cours. A la fin du chapitre le **Tableau 2.1 récapitule tous les résultats nécessaires pour la suite**.

2.2 Fonctions de carré sommable et dérivation faible

2.2.1 Quelques rappels d’intégration

Tous les résultats de cette sous-section sont détaillés dans le cours de mathématiques [6]. Soit Ω un ouvert de \mathbb{R}^N muni de la mesure de Lebesgue. On définit l’espace $L^2(\Omega)$ des fonctions mesurables de carré sommable dans Ω . Muni du produit scalaire

$$\langle f, g \rangle = \int_{\Omega} f(x)g(x) dx,$$

$L^2(\Omega)$ est un espace de Hilbert (voir le théorème 3.3.2 de [6]). On note

$$\|f\|_{L^2(\Omega)} = \left(\int_{\Omega} |f(x)|^2 dx \right)^{1/2}$$

la norme correspondante. Rappelons que les fonctions mesurables dans Ω sont définies **presque partout** dans Ω : si on change les valeurs d'une fonction mesurable f sur un sous-ensemble de Ω de mesure nulle, on ne change pas la fonction mesurable f . Autrement dit, deux fonctions mesurables f et g seront dites égales si $f(x) = g(x)$ presque partout dans Ω , c'est-à-dire s'il existe $E \subset \Omega$ tel que la mesure de Lebesgue de E est nulle et $f(x) = g(x)$ pour tout $x \in (\Omega \setminus E)$.

On note $C_c^\infty(\Omega)$ (ou $\mathcal{D}(\Omega)$) l'espace des fonctions de classe C^∞ à support compact dans Ω . Remarquons que l'espace $C_c^\infty(\Omega)$ n'est pas réduit à la seule fonction nulle partout (ce qui n'est pas évident ! Voir le corollaire 3.2.6 dans [6]). Notons aussi que les fonctions de $C_c^\infty(\Omega)$ s'annulent, ainsi que toutes leurs dérivées, sur le bord de Ω . Nous rappelons le résultat de densité suivant (voir le théorème 3.4.3 de [6])

Théorème 2.2.1 *L'espace $C_c^\infty(\Omega)$ est dense dans $L^2(\Omega)$, c'est-à-dire que pour tout $f \in L^2(\Omega)$ il existe une suite $f_n \in C_c^\infty(\Omega)$ telle que*

$$\lim_{n \rightarrow +\infty} \|f - f_n\|_{L^2(\Omega)} = 0.$$

Comme corollaire nous obtenons la généralisation immédiate du Lemme 1.2.9.

Corollaire 2.2.2 *Soit $f \in L^2(\Omega)$. Si pour toute fonction $\phi \in C_c^\infty(\Omega)$ on a*

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

alors $f(x) = 0$ presque partout dans Ω .

2.2.2 Dérivation faible

On définit tout d'abord le concept de dérivée faible dans $L^2(\Omega)$. Cette notion généralise la dérivation usuelle (parfois appelée, par opposition, dérivation forte) et est un cas particulier de la dérivation au sens des distributions (voir [6]).

Définition 2.2.3 *Soit v une fonction de $L^2(\Omega)$. On dit que v est dérivable au sens faible dans $L^2(\Omega)$ s'il existe des fonctions $w_i \in L^2(\Omega)$, pour $i \in \{1, \dots, N\}$, telles que, pour toute fonction $\phi \in C_c^\infty(\Omega)$, on a*

$$\int_{\Omega} v(x) \frac{\partial \phi}{\partial x_i}(x) dx = - \int_{\Omega} w_i(x) \phi(x) dx.$$

Chaque w_i est appelée la i -ème dérivée partielle faible de v et notée désormais $\frac{\partial v}{\partial x_i}$.

La Définition 2.2.3 a bien un sens : en particulier, la notation $w_i = \frac{\partial v}{\partial x_i}$ est univoque car, en vertu du Corollaire 2.2.2, les fonctions w_i sont uniques (si elles existent). Bien sûr, si v est dérivable au sens usuel et que ses dérivées partielles appartiennent à $L^2(\Omega)$, alors les dérivées usuelle et faible de v coïncident (utiliser le Corollaire 1.2.3). Donnons tout de suite un critère simple et pratique pour déterminer si une fonction est dérivable au sens faible.

Lemme 2.2.4 Soit v une fonction de $L^2(\Omega)$. S'il existe une constante $C > 0$ telle que, pour toute fonction $\phi \in C_c^\infty(\Omega)$ et pour tout indice $i \in \{1, \dots, N\}$, on a

$$\left| \int_{\Omega} v(x) \frac{\partial \phi}{\partial x_i}(x) dx \right| \leq C \|\phi\|_{L^2(\Omega)}, \quad (2.3)$$

alors v est dérivable au sens faible.

Démonstration. Soit L la forme linéaire définie par

$$L(\phi) = \int_{\Omega} v(x) \frac{\partial \phi}{\partial x_i}(x) dx.$$

A priori $L(\phi)$ n'est définie que pour $\phi \in C_c^\infty(\Omega)$, mais grâce à l'inégalité (2.3), on peut étendre L par continuité à toutes les fonctions de $L^2(\Omega)$ car $C_c^\infty(\Omega)$ est dense dans $L^2(\Omega)$ d'après le Théorème 2.2.1. En fait, l'inégalité (2.3) prouve que la forme linéaire L est continue sur $L^2(\Omega)$. En vertu du théorème de représentation de Riesz, il existe une fonction $(-w_i) \in L^2(\Omega)$ telle que

$$L(\phi) = - \int_{\Omega} w_i(x) \phi(x) dx,$$

ce qui prouve que v est dérivable au sens faible dans $L^2(\Omega)$. □

Exercice 2.2.1 Soit $\Omega = (0, 1)$. Montrer que la fonction x^α est dérivable au sens faible dans $L^2(\Omega)$ si et seulement si $\alpha > 1/2$.

Exercice 2.2.2 Soit Ω un ouvert borné. Montrer qu'une fonction continue sur $\overline{\Omega}$, et C^1 par morceaux est dérivable au sens faible dans $L^2(\Omega)$.

Exercice 2.2.3 Soit Ω un ouvert borné. Montrer qu'une fonction C^1 par morceaux mais pas continue n'est pas dérivable au sens faible dans $L^2(\Omega)$.

On retrouve un résultat bien connu pour la dérivée usuelle.

Proposition 2.2.5 Soit v une fonction de $L^2(\Omega)$ dérivable au sens faible et telle que toutes ses dérivées partielles faibles $\frac{\partial v}{\partial x_i}$, pour $1 \leq i \leq N$, sont nulles. Alors, pour chaque composante connexe de Ω , il existe une constante C telle que $v(x) = C$ presque partout dans cette composante connexe.

On peut facilement généraliser la Définition 2.2.3 de la dérivée faible à certains opérateurs différentiels qui ne font intervenir que certaines combinaisons de dérivées partielles (et non pas toutes). C'est par exemple le cas de la divergence d'une fonction à valeurs vectorielles qui nous sera utile par la suite.

Définition 2.2.6 Soit σ une fonction de Ω dans \mathbb{R}^N dont toutes les composantes appartiennent à $L^2(\Omega)$ (on note $\sigma \in L^2(\Omega)^N$). On dit que σ admet une divergence au sens faible dans $L^2(\Omega)$ s'il existe une fonction $w \in L^2(\Omega)$ telle que, pour toute fonction $\phi \in C_c^\infty(\Omega)$, on a

$$\int_{\Omega} \sigma(x) \cdot \nabla \phi(x) dx = - \int_{\Omega} w(x) \phi(x) dx.$$

La fonction w est appelée la divergence faible de σ et notée désormais $\operatorname{div} \sigma$.

La justification de la Définition 2.2.6 est que, si σ est une fonction régulière, alors une simple intégration par parties (voir le Corollaire 1.2.3) montre que l'on a bien $w = \operatorname{div} \sigma$. Une généralisation facile du critère de dérivation faible du Lemme 2.2.4 est donnée par le résultat suivant (dont nous laissons la démonstration au lecteur en guise d'exercice).

Lemme 2.2.7 *Soit σ une fonction de $L^2(\Omega)^N$. S'il existe une constante $C > 0$ telle que, pour toute fonction $\phi \in C_c^\infty(\Omega)$, on a*

$$\left| \int_{\Omega} \sigma(x) \cdot \nabla \phi(x) \, dx \right| \leq C \|\phi\|_{L^2(\Omega)},$$

alors σ admet une divergence au sens faible.

Exercice 2.2.4 Soit Ω un ouvert borné constitué de deux ouverts Ω_1 et Ω_2 séparés par une surface $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$. Montrer qu'une fonction vectorielle de classe C^1 sur chaque morceau Ω_1 et Ω_2 admet une divergence faible dans $L^2(\Omega)$ si et seulement si sa composante normale est continue à travers la surface Γ .

2.3 Définition et principales propriétés

2.3.1 Espace $H^1(\Omega)$

Définition 2.3.1 *Soit Ω un ouvert de \mathbb{R}^N . L'espace de Sobolev $H^1(\Omega)$ est défini par*

$$H^1(\Omega) = \left\{ v \in L^2(\Omega) \text{ tel que } \forall i \in \{1, \dots, N\} \frac{\partial v}{\partial x_i} \in L^2(\Omega) \right\}, \quad (2.5)$$

où $\frac{\partial v}{\partial x_i}$ est la dérivée partielle faible de v au sens de la Définition 2.2.3.

En physique ou en mécanique l'espace de Sobolev est souvent appelé **espace d'énergie** au sens où il est constitué des fonctions d'énergie finie (c'est-à-dire de norme $\|u\|_{H^1(\Omega)}$ finie). Les fonctions d'énergie finie peuvent éventuellement être "singulières" ce qui a un sens physique possible (concentration ou explosion locale de certaines grandeurs). On consultera avec intérêt les exemples explicites de l'Exercice 2.3.2 et du Lemme 3.2.33.

Proposition 2.3.2 *Muni du produit scalaire*

$$\langle u, v \rangle = \int_{\Omega} \left(u(x)v(x) + \nabla u(x) \cdot \nabla v(x) \right) dx \quad (2.6)$$

et de la norme

$$\|u\|_{H^1(\Omega)} = \left(\int_{\Omega} (|u(x)|^2 + |\nabla u(x)|^2) \, dx \right)^{1/2}$$

l'espace de Sobolev $H^1(\Omega)$ est un espace de Hilbert.

Démonstration. Il est évident que (2.6) est bien un produit scalaire dans $H^1(\Omega)$. Il reste donc à montrer que $H^1(\Omega)$ est complet pour la norme associée. Soit $(u_n)_{n \geq 1}$ une suite de Cauchy dans $H^1(\Omega)$. Par définition de la norme de $H^1(\Omega)$, $(u_n)_{n \geq 1}$ ainsi que $(\frac{\partial u_n}{\partial x_i})_{n \geq 1}$ pour $i \in \{1, \dots, N\}$ sont des suites de Cauchy dans $L^2(\Omega)$. Comme $L^2(\Omega)$ est complet, il existe des limites u et w_i telles que u_n converge vers u et $\frac{\partial u_n}{\partial x_i}$ converge vers w_i dans $L^2(\Omega)$. Or, par définition de la dérivée faible de u_n , pour toute fonction $\phi \in C_c^\infty(\Omega)$, on a

$$\int_{\Omega} u_n(x) \frac{\partial \phi}{\partial x_i}(x) dx = - \int_{\Omega} \frac{\partial u_n}{\partial x_i}(x) \phi(x) dx. \quad (2.7)$$

Passant à la limite $n \rightarrow +\infty$ dans (2.7), on obtient

$$\int_{\Omega} u(x) \frac{\partial \phi}{\partial x_i}(x) dx = - \int_{\Omega} w_i(x) \phi(x) dx,$$

ce qui prouve que u est dérivable au sens faible et que w_i est la i -ème dérivée partielle faible de u , $\frac{\partial u}{\partial x_i}$. Donc, u appartient bien à $H^1(\Omega)$ et $(u_n)_{n \geq 1}$ converge vers u dans $H^1(\Omega)$. \square

Exercice 2.3.1 Montrer que les fonctions continues, C^1 par morceaux et à support borné dans $\overline{\Omega}$, appartiennent à $H^1(\Omega)$.

En dimension $N \geq 2$, les fonctions de $H^1(\Omega)$ ne sont en général **ni continues ni bornées**, comme le montre le contre-exemple suivant.

Exercice 2.3.2 Soit B la boule unité ouverte de \mathbb{R}^N . Si $N = 2$, montrer que la fonction $u(x) = |\log(|x|)|^\alpha$ appartient à $H^1(B)$ pour $0 < \alpha < 1/2$, mais n'est pas bornée au voisinage de l'origine. Si $N \geq 3$, montrer que la fonction $u(x) = |x|^{-\beta}$ appartient à $H^1(B)$ pour $0 < \beta < (N - 2)/2$, mais n'est pas bornée au voisinage de l'origine.

La dimension $N = 1$ fait "exception" à la non-continuité des fonctions de $H^1(\Omega)$ comme l'affirme le lemme suivant où, sans perte de généralité, on prend $\Omega = (0, 1)$.

Lemme 2.3.3 Pour toute fonction $v \in H^1(0, 1)$ et pour tout $x, y \in [0, 1]$, on a

$$v(y) = v(x) + \int_x^y v'(s) ds. \quad (2.8)$$

Plus généralement, pour tout $x \in [0, 1]$, l'application $v \rightarrow v(x)$, définie de $H^1(0, 1)$ dans \mathbb{R} , est une forme linéaire continue sur $H^1(0, 1)$. En particulier, toute fonction $v \in H^1(0, 1)$ est continue sur $[0, 1]$.

Il est très important en pratique de savoir si **les fonctions régulières sont denses dans l'espace de Sobolev $H^1(\Omega)$** . Cela justifie en partie la notion d'espace de Sobolev qui apparaît ainsi très simplement comme l'ensemble des fonctions régulières complété par les limites de suites de fonctions régulières dans la norme de l'énergie $\|u\|_{H^1(\Omega)}$. Cela permet de démontrer facilement de nombreuses propriétés en les établissant d'abord sur les fonctions régulières puis en utilisant un argument de "densité".

Théorème 2.3.5 (de densité) *Si Ω est un ouvert borné régulier de classe C^1 , ou bien si $\Omega = \mathbb{R}_+^N$, ou encore si $\Omega = \mathbb{R}^N$, alors $C_c^\infty(\overline{\Omega})$ est dense dans $H^1(\Omega)$.*

Rappelons que la notation \mathbb{R}_+^N désigne le demi-espace $\{x \in \mathbb{R}^N \text{ tel que } x_N > 0\}$.

Remarque 2.3.6 L'espace $C_c^\infty(\overline{\Omega})$ qui est dense dans $H^1(\Omega)$ est constitué des fonctions régulières de classe C^∞ à support borné (ou compact) dans le fermé $\overline{\Omega}$. En particulier, si Ω est borné, toutes les fonctions de $C^\infty(\overline{\Omega})$ ont nécessairement un support borné, et donc $C_c^\infty(\overline{\Omega}) = C^\infty(\overline{\Omega})$. Précisons que les fonctions de $C_c^\infty(\overline{\Omega})$ ne s'annulent pas nécessairement sur le bord de l'ouvert Ω , ce qui différencie cet espace de $C_c^\infty(\Omega)$ (voir la Remarque 1.2.2). Par contre, si Ω n'est pas borné, les fonctions de $C_c^\infty(\overline{\Omega})$ s'annulent "à l'infini". •

Remarque 2.3.7 La notion de régularité d'un ouvert a été introduite dans la Définition 1.2.5. Il n'est pas nécessaire de connaître précisément cette définition de la régularité d'un ouvert. Il suffit de savoir *grosso modo* que l'on demande que le bord de l'ouvert soit une surface régulière et que l'on exclut certaines "pathologies" (voir la Remarque 1.2.6). •

2.3.2 Espace $H_0^1(\Omega)$

Définissons maintenant un autre espace de Sobolev qui est un sous-espace de $H^1(\Omega)$ et qui nous sera très utile pour les problèmes avec conditions aux limites de Dirichlet.

Définition 2.3.8 *Soit $C_c^\infty(\Omega)$ l'espace des fonctions de classe C^∞ à support compact dans Ω . L'espace de Sobolev $H_0^1(\Omega)$ est défini comme l'adhérence de $C_c^\infty(\Omega)$ dans $H^1(\Omega)$.*

On verra un peu plus loin (voir le Corollaire 2.3.16) que $H_0^1(\Omega)$ est en fait le sous-espace de $H^1(\Omega)$ constitué des **fonctions qui s'annulent sur le bord** $\partial\Omega$ puisque tel est le cas des fonctions de $C_c^\infty(\Omega)$. En général, $H_0^1(\Omega)$ est **strictement plus petit** que $H^1(\Omega)$ car $C_c^\infty(\Omega)$ est un sous-espace **strict** de $C_c^\infty(\overline{\Omega})$ (voir le Théorème 2.3.5 et la Remarque 2.3.6). Une exception importante est le cas où $\Omega = \mathbb{R}^N$: en effet, dans ce cas $\overline{\Omega} = \mathbb{R}^N = \Omega$ et le Théorème 2.3.5 affirme que $C_c^\infty(\mathbb{R}^N)$ est dense dans $H^1(\mathbb{R}^N)$, donc on a $H_0^1(\mathbb{R}^N) = H^1(\mathbb{R}^N)$. Cette exception se comprend aisément puisque l'espace entier \mathbb{R}^N n'a pas de bord.

Proposition 2.3.9 *Muni du produit scalaire (2.6) de $H^1(\Omega)$, l'espace de Sobolev $H_0^1(\Omega)$ est un espace de Hilbert.*

Démonstration. Par définition $H_0^1(\Omega)$ est un sous-espace fermé de $H^1(\Omega)$ (qui est un espace de Hilbert), donc c'est aussi un espace de Hilbert. □

Un résultat essentiel pour les applications du prochain chapitre est l'inégalité suivante.

Proposition 2.3.10 (Inégalité de Poincaré) *Soit Ω un ouvert de \mathbb{R}^N borné dans au moins une direction de l'espace. Il existe une constante $C > 0$ telle que, pour toute fonction $v \in H_0^1(\Omega)$,*

$$\int_{\Omega} |v(x)|^2 dx \leq C \int_{\Omega} |\nabla v(x)|^2 dx. \quad (2.9)$$

Démonstration. Pour les fonctions $v \in C_c^\infty(\Omega)$ on a déjà démontré l'inégalité de Poincaré (2.9) dans le Lemme 1.3.5. Par un argument de densité le résultat reste vrai pour toute fonction $v \in H_0^1(\Omega)$. En effet, comme $C_c^\infty(\Omega)$ est dense dans $H_0^1(\Omega)$ (par sa Définition 2.3.8), il existe une suite $v_n \in C_c^\infty(\Omega)$ telle que

$$\lim_{n \rightarrow +\infty} \|v - v_n\|_{H^1(\Omega)}^2 = \lim_{n \rightarrow +\infty} \int_{\Omega} (|v - v_n|^2 + |\nabla(v - v_n)|^2) dx = 0.$$

En particulier, on en déduit que

$$\lim_{n \rightarrow +\infty} \int_{\Omega} |v_n|^2 dx = \int_{\Omega} |v|^2 dx \text{ et } \lim_{n \rightarrow +\infty} \int_{\Omega} |\nabla v_n|^2 dx = \int_{\Omega} |\nabla v|^2 dx.$$

Par application du Lemme 1.3.5, on a

$$\int_{\Omega} |v_n(x)|^2 dx \leq C \int_{\Omega} |\nabla v_n(x)|^2 dx. \quad (2.10)$$

On passe alors à la limite $n \rightarrow +\infty$ dans chacun des deux termes de l'inégalité (2.10) pour obtenir le résultat recherché. Ce type d'argument "par densité" sera très souvent repris par la suite. \square

Remarque 2.3.11 L'inégalité de Poincaré (2.9) n'est pas vraie pour les fonctions de $H^1(\Omega)$. En effet, les fonctions constantes (non nulles) annulent le terme de droite dans (2.9) mais pas le terme de gauche. L'hypothèse sous-jacente essentielle dans l'inégalité de Poincaré est que les fonctions de $H_0^1(\Omega)$ s'annulent sur le bord $\partial\Omega$ de l'ouvert Ω (voir la Remarque 2.3.18 pour des variantes de cette hypothèse). \bullet

Un corollaire important de l'inégalité de Poincaré est le résultat suivant qui fournit une norme équivalente plus simple dans $H_0^1(\Omega)$.

Corollaire 2.3.12 *Soit Ω un ouvert de \mathbb{R}^N borné dans au moins une direction de l'espace. Alors la semi-norme*

$$|v|_{H_0^1(\Omega)} = \left(\int_{\Omega} |\nabla v(x)|^2 dx \right)^{1/2}$$

est une norme sur $H_0^1(\Omega)$ équivalente à la norme usuelle induite par celle de $H^1(\Omega)$.

Démonstration. Soit $v \in H_0^1(\Omega)$. La première inégalité

$$|v|_{H_0^1(\Omega)} \leq \|v\|_{H^1(\Omega)} = \left(\int_{\Omega} (|v|^2 + |\nabla v|^2) dx \right)^{1/2}$$

est évidente. D'autre part, l'inégalité de Poincaré du Lemme 1.3.5 conduit à

$$\|v\|_{H^1(\Omega)}^2 \leq (C + 1) \int_{\Omega} |\nabla v|^2 dx = (C + 1) |v|_{H_0^1(\Omega)}^2,$$

ce qui prouve que $|v|_{H_0^1(\Omega)}$ est une norme équivalente à $\|v\|_{H^1(\Omega)}$. \square

2.3.3 Traces et formules de Green

Nous avons vu qu'en dimension $N \geq 2$ les fonctions de $H^1(\Omega)$ ne sont pas continues (voir le contre-exemple de l'Exercice 2.3.2). Comme pour toute fonction mesurable, on ne peut donc parler de la valeur ponctuelle d'une fonction $v \in H^1(\Omega)$ que "presque partout" dans Ω . En particulier, il n'est pas clair de savoir si on peut définir la "valeur au bord", ou "trace" de v sur le bord $\partial\Omega$ car $\partial\Omega$ est un ensemble négligeable ou de mesure nulle. Fort heureusement pour les problèmes aux limites que nous étudions, il y a tout de même un moyen pour définir la trace $v|_{\partial\Omega}$ d'une fonction de $H^1(\Omega)$. Ce résultat essentiel, appelé théorème de trace, est le suivant.

Théorème 2.3.13 (de trace) *Soit Ω un ouvert borné régulier de classe \mathcal{C}^1 , ou bien $\Omega = \mathbb{R}_+^N$. On définit l'application trace γ_0*

$$\begin{aligned} H^1(\Omega) \cap C(\overline{\Omega}) &\rightarrow L^2(\partial\Omega) \cap C(\overline{\partial\Omega}) \\ v &\rightarrow \gamma_0(v) = v|_{\partial\Omega}. \end{aligned} \quad (2.11)$$

Cette application γ_0 se prolonge par continuité en une application linéaire continue de $H^1(\Omega)$ dans $L^2(\partial\Omega)$, notée encore γ_0 . En particulier, il existe une constante $C > 0$ telle que, pour toute fonction $v \in H^1(\Omega)$, on a

$$\|v\|_{L^2(\partial\Omega)} \leq C\|v\|_{H^1(\Omega)}. \quad (2.12)$$

Remarque 2.3.14 Grâce au Théorème de trace 2.3.13 on peut donc parler de la valeur d'une fonction de $H^1(\Omega)$ sur le bord $\partial\Omega$. Ce résultat est remarquable car il n'est pas vrai pour une fonction de $L^2(\Omega)$. •

Démonstration. Démontrons le résultat pour le demi-espace $\Omega = \mathbb{R}_+^N = \{x \in \mathbb{R}^N, x_N > 0\}$. Soit $v \in C_c^\infty(\overline{\mathbb{R}_+^N})$. Avec la notation $x = (x', x_N)$, on a

$$|v(x', 0)|^2 = -2 \int_0^{+\infty} v(x', x_N) \frac{\partial v}{\partial x_N}(x', x_N) dx_N,$$

et, en utilisant l'inégalité $2ab \leq a^2 + b^2$,

$$|v(x', 0)|^2 \leq \int_0^{+\infty} \left(|v(x', x_N)|^2 + \left| \frac{\partial v}{\partial x_N}(x', x_N) \right|^2 \right) dx_N.$$

Par intégration en x' , on en déduit

$$\int_{\mathbb{R}^{N-1}} |v(x', 0)|^2 dx' \leq \int_{\mathbb{R}_+^N} \left(|v(x)|^2 + \left| \frac{\partial v}{\partial x_N}(x) \right|^2 \right) dx,$$

c'est-à-dire $\|v\|_{L^2(\partial\mathbb{R}_+^N)} \leq \|v\|_{H^1(\mathbb{R}_+^N)}$. Par densité de $C_c^\infty(\overline{\mathbb{R}_+^N})$ dans $H^1(\mathbb{R}_+^N)$, on obtient ainsi le résultat.

Pour un ouvert borné régulier de classe \mathcal{C}^1 , on utilise un argument de cartes locales du bord qui permet de se ramener au cas de $\Omega = \mathbb{R}_+^N$. Nous ne détaillons pas

cet argument de cartes locales (assez technique) et nous renvoyons le lecteur curieux à [1]. \square

Le Théorème de trace 2.3.13 permet de généraliser aux fonctions de $H^1(\Omega)$ la formule de Green précédemment établie pour des fonctions de classe C^1 au Corollaire 1.2.3.

Théorème 2.3.15 (Formule de Green) *Soit Ω un ouvert borné régulier de classe C^1 . Si u et v sont des fonctions de $H^1(\Omega)$, elles vérifient*

$$\int_{\Omega} u(x) \frac{\partial v}{\partial x_i}(x) dx = - \int_{\Omega} v(x) \frac{\partial u}{\partial x_i}(x) dx + \int_{\partial\Omega} u(x)v(x)n_i(x) ds, \quad (2.13)$$

où $n = (n_i)_{1 \leq i \leq N}$ est la normale unité extérieure à $\partial\Omega$.

Démonstration. Rappelons que la formule (2.13) a été établie pour des fonctions de classe C^1 dans le Corollaire 1.2.3. On utilise à nouveau un argument de densité. Par densité de $C_c^\infty(\overline{\Omega})$ dans $H^1(\Omega)$ (voir le Théorème 2.3.5), il existe des suites $(u_n)_{n \geq 1}$ et $(v_n)_{n \geq 1}$ dans $C_c^\infty(\overline{\Omega})$ qui convergent dans $H^1(\Omega)$ vers u et v , respectivement. En vertu du Corollaire 1.2.3 on a

$$\int_{\Omega} u_n \frac{\partial v_n}{\partial x_i} dx = - \int_{\Omega} v_n \frac{\partial u_n}{\partial x_i} dx + \int_{\partial\Omega} u_n v_n n_i ds. \quad (2.14)$$

On peut passer à la limite $n \rightarrow +\infty$ dans les deux premiers termes de (2.14) car u_n et $\frac{\partial u_n}{\partial x_i}$ (respectivement, v_n et $\frac{\partial v_n}{\partial x_i}$) convergent vers u et $\frac{\partial u}{\partial x_i}$ (respectivement, v et $\frac{\partial v}{\partial x_i}$) dans $L^2(\Omega)$. Pour passer à la limite dans la dernière intégrale de (2.14), on utilise la continuité de l'application trace γ_0 , c'est-à-dire l'inégalité (2.12), qui permet d'affirmer que $\gamma_0(u_n)$ (respectivement, $\gamma_0(v_n)$) converge vers $\gamma_0(u)$ (respectivement, $\gamma_0(v)$) dans $L^2(\partial\Omega)$. On obtient ainsi la formule (2.13) pour des fonctions u et v de $H^1(\Omega)$. \square

Comme conséquence du Théorème de trace 2.3.13 on obtient une caractérisation très simple de l'espace $H_0^1(\Omega)$.

Corollaire 2.3.16 *Soit Ω un ouvert borné régulier de classe C^1 . L'espace $H_0^1(\Omega)$ coïncide avec le sous-espace de $H^1(\Omega)$ constitué des fonctions qui s'annulent sur le bord $\partial\Omega$.*

Démonstration. Comme toute fonction de $H_0^1(\Omega)$ est limite d'une suite de fonctions appartenant à $C_c^\infty(\Omega)$ qui ont bien sûr une trace nulle, la continuité de l'application trace γ_0 implique que la trace de la limite est aussi nulle. On en déduit que $H_0^1(\Omega)$ est contenu dans le sous-espace de $H^1(\Omega)$ des fonctions qui s'annulent sur le bord $\partial\Omega$. La réciproque est plus technique et découle d'un double procédé de cartes locales puis de régularisation et de translation. Nous renvoyons à [1], [7], [24] pour plus de détails. \square

Grâce au Corollaire 2.3.16 nous pouvons donner une autre démonstration de la Proposition 2.3.10 à propos de l'inégalité de Poincaré. Cette nouvelle démonstration n'est plus "constructive" mais est basée sur un argument de contradiction qui possède

le mérite de se généraliser très facilement. En effet, il existe de nombreuses variantes de l'inégalité de Poincaré, adaptées aux différents modèles d'équations aux dérivées partielles. Au vu de l'importance de cette inégalité pour la suite, il n'est donc pas inutile d'en donner une démonstration aisément adaptable à tous les cas de figure.

Autre démonstration de la Proposition 2.3.10. On procède par contradiction. S'il n'existe pas de constante $C > 0$ telle que, pour toute fonction $v \in H_0^1(\Omega)$,

$$\int_{\Omega} |v(x)|^2 dx \leq C \int_{\Omega} |\nabla v(x)|^2 dx,$$

cela veut dire qu'il existe une suite $v_n \in H_0^1(\Omega)$ telle que

$$1 = \int_{\Omega} |v_n(x)|^2 dx > n \int_{\Omega} |\nabla v_n(x)|^2 dx. \quad (2.15)$$

En particulier, (2.15) implique que la suite v_n est bornée dans $H_0^1(\Omega)$. Par application du Théorème de Rellich 2.3.21 ci-dessous, il existe une sous-suite $v_{n'}$ qui converge dans $L^2(\Omega)$. De plus, (2.15) montre que la suite $\nabla v_{n'}$ converge vers zéro dans $L^2(\Omega)$ (composante par composante). Par conséquent, $v_{n'}$ est une suite de Cauchy dans $H_0^1(\Omega)$, qui est un espace de Hilbert, donc elle converge dans $H_0^1(\Omega)$ vers une limite v . Comme on a

$$\int_{\Omega} |\nabla v(x)|^2 dx = \lim_{n \rightarrow +\infty} \int_{\Omega} |\nabla v_n(x)|^2 dx \leq \lim_{n \rightarrow +\infty} \frac{1}{n} = 0,$$

on en déduit, en vertu du Lemme 2.2.5, que v est une constante dans chaque composante connexe de Ω . Mais comme v est nulle sur le bord $\partial\Omega$ (en vertu du Corollaire 2.3.16), v est identiquement nulle dans tout Ω . Par ailleurs,

$$\int_{\Omega} |v(x)|^2 dx = \lim_{n \rightarrow +\infty} \int_{\Omega} |v_n(x)|^2 dx = 1,$$

ce qui est une contradiction avec le fait que $v = 0$. □

Remarque 2.3.18 La démonstration par contradiction de la Proposition 2.3.10 se généralise facilement. Prenons par exemple, le cas d'un ouvert Ω , borné connexe et régulier de classe C^1 , dont le bord $\partial\Omega$ se décompose en deux parties disjointes régulières $\partial\Omega_N$ et $\partial\Omega_D$ dont les mesures superficielles sont non nulles. On définit un espace V par

$$V = \{v \in H^1(\Omega) \text{ tel que } v = 0 \text{ sur } \partial\Omega_D\}.$$

Par application du Théorème de trace 2.3.13, il est facile de voir que V est un sous-espace fermé de $H^1(\Omega)$, donc est un espace de Hilbert pour le produit scalaire de $H^1(\Omega)$. Comme pour $H_0^1(\Omega)$, l'argument de contradiction permet de démontrer l'existence d'une constante $C > 0$ telle que toute fonction $v \in V$ vérifie l'inégalité de Poincaré (2.9). •

2.3.4 Un résultat de compacité

Nous consacrons cette sous-section à l'étude d'une propriété de compacité connue sous le nom de théorème de Rellich qui jouera un rôle essentiel dans la théorie spectrale des problèmes aux limites (voir le Chapitre 5) que nous utiliserons pour résoudre les problèmes d'évolution en temps. Rappelons tout d'abord que, dans un espace de Hilbert de dimension infinie, il n'est pas vrai que, de toute suite bornée, on puisse extraire une sous-suite convergente (au contraire de ce qui se passe en dimension finie, voir l'Exercice 2.3.5).

Théorème 2.3.21 (de Rellich) *Si Ω est un ouvert borné régulier de classe C^1 , alors de toute suite bornée de $H^1(\Omega)$ on peut extraire une sous-suite convergente dans $L^2(\Omega)$ (on dit que l'injection canonique de $H^1(\Omega)$ dans $L^2(\Omega)$ est compacte).*

Le Théorème 2.3.21 peut être faux si l'ouvert Ω n'est pas borné. Par exemple, si $\Omega = \mathbb{R}^N$, l'injection canonique de $H^1(\mathbb{R}^N)$ dans $L^2(\mathbb{R}^N)$ n'est pas compacte. Pour s'en convaincre il suffit de considérer la suite $u_n(x) = u(x + ne)$ où e est un vecteur non nul et u une fonction de $H^1(\mathbb{R}^N)$ (on translate u dans la direction e). Il est clair qu'aucune sous-suite de u_n ne converge dans $L^2(\mathbb{R}^N)$.

Remarque 2.3.22 Si l'on remplace $H^1(\Omega)$ par $H_0^1(\Omega)$, alors non seulement le Théorème 2.3.21 de Rellich reste vrai, mais en plus il n'est pas nécessaire de supposer que l'ouvert Ω est régulier. •

Exercice 2.3.5 Soit $\Omega = (0, 1)$ et $u_n(x) = \sin(2\pi nx)$. Montrer que la suite u_n est uniformément bornée dans $L^2(\Omega)$, mais qu'il n'existe aucune sous-suite convergente. Pour cela on montrera, grâce à une intégration par parties, que, pour toute fonction $\phi \in C_c^\infty(\Omega)$, on a

$$\lim_{n \rightarrow +\infty} \int_0^1 u_n(x) \phi(x) dx = 0,$$

et on en déduira une contradiction si une sous-suite de u_n converge dans $L^2(\Omega)$. Généraliser ce contre exemple à $H^1(\Omega)$ en considérant une primitive de u_n .

2.3.5 Espaces $H^m(\Omega)$

On peut aisément généraliser la Définition 2.3.1 de l'espace de Sobolev $H^1(\Omega)$ aux fonctions qui sont $m \geq 0$ fois dérivables au sens faible. Commençons par donner une convention d'écriture bien utile. Soit $\alpha = (\alpha_1, \dots, \alpha_N)$ un **multi-indice**, c'est-à-dire un vecteur à N composantes entières positives $\alpha_i \geq 0$. On note $|\alpha| = \sum_{i=1}^N \alpha_i$ et, pour une fonction v ,

$$\partial^\alpha v(x) = \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \dots \partial x_N^{\alpha_N}}(x).$$

A partir de la Définition 2.2.3 de la dérivée première faible, on définit par récurrence sur m la dérivée d'ordre m faible : on dit qu'une fonction $v \in L^2(\Omega)$ est m fois dérivable au sens faible si toutes ses dérivées partielles faibles d'ordre $m - 1$

sont dérivables faiblement au sens de la Définition 2.2.3. Remarquons que, dans la définition d'une dérivée croisée, l'ordre de dérivation n'est pas important, à cause du théorème de Schwarz $\frac{\partial^2 v}{\partial x_i \partial x_j} = \frac{\partial^2 v}{\partial x_j \partial x_i}$, ce qui justifie la notation $\partial^\alpha v$ où l'ordre de dérivation n'est pas indiqué.

Définition 2.3.23 *Pour un entier $m \geq 0$, l'espace de Sobolev $H^m(\Omega)$ est défini par*

$$H^m(\Omega) = \left\{ v \in L^2(\Omega) \text{ tel que, } \forall \alpha \text{ avec } |\alpha| \leq m, \partial^\alpha v \in L^2(\Omega) \right\}, \quad (2.17)$$

où la dérivée partielle $\partial^\alpha v$ est à prendre au sens faible.

Nous laissons au lecteur le soin de vérifier le résultat facile suivant.

Proposition 2.3.24 *Muni du produit scalaire*

$$\langle u, v \rangle = \int_{\Omega} \sum_{|\alpha| \leq m} \partial^\alpha u(x) \partial^\alpha v(x) dx \quad (2.18)$$

et de la norme $\|u\|_{H^m(\Omega)} = \sqrt{\langle u, u \rangle}$, l'espace de Sobolev $H^m(\Omega)$ est un espace de Hilbert.

Les fonctions de $H^m(\Omega)$ ne sont pas toujours continues ou régulières (cela dépend de m et de la dimension N), mais si m est suffisamment grand alors toute fonction de $H^m(\Omega)$ est continue. Rappelons qu'en vertu du Lemme 2.3.3, en dimension d'espace $N = 1$, les fonctions de $H^1(\Omega)$ sont continues. Nous admettrons le résultat suivant qui généralise le Lemme 2.3.3 aux dimensions supérieures.

Théorème 2.3.25 *Si Ω est un ouvert borné régulier de classe C^1 , et si $m > N/2$, alors $H^m(\Omega)$ est un sous-espace de l'ensemble $C(\overline{\Omega})$ des fonctions continues sur $\overline{\Omega}$.*

Remarque 2.3.26 Par application réitérée du Théorème 2.3.25 à une fonction et à ses dérivées, on peut en fait améliorer sa conclusion. S'il existe un entier $k \geq 0$ tel que $m - N/2 > k$, alors $H^m(\Omega)$ est un sous-espace de l'ensemble $C^k(\overline{\Omega})$ des fonctions k fois différentiables sur $\overline{\Omega}$. •

La "morale" du Théorème 2.3.25 est que plus m est grand, plus les fonctions de $H^m(\Omega)$ sont régulières, c'est-à-dire dérivables au sens usuel (il suffit d'appliquer successivement le Théorème 2.3.25 à une fonction $v \in H^m(\Omega)$ et à ses dérivées $\partial^\alpha v \in H^{m-|\alpha|}(\Omega)$).

Comme pour $H^1(\Omega)$, les fonctions régulières sont denses dans $H^m(\Omega)$ (si du moins l'ouvert Ω est régulier ; voir la Définition 1.2.5). La démonstration du Théorème de densité 2.3.5 se généralise très facilement à $H^m(\Omega)$. Nous ne la répéterons pas et nous énonçons seulement le résultat de densité suivant.

Théorème 2.3.27 *Si Ω est un ouvert borné régulier de classe C^m , ou bien si $\Omega = \mathbb{R}_+^N$, alors $C_c^\infty(\overline{\Omega})$ est dense dans $H^m(\Omega)$.*

On peut aussi obtenir des résultats de trace et des formules de Green d'ordre plus élevés pour l'espace $H^m(\Omega)$. Par souci de simplicité, nous nous contentons de traiter le cas $m = 2$ (qui est le seul que nous utiliserons par la suite).

Théorème 2.3.28 *Soit Ω un ouvert borné régulier de classe C^1 . On définit l'application trace γ_1*

$$\begin{aligned} H^2(\Omega) \cap C^1(\overline{\Omega}) &\rightarrow L^2(\partial\Omega) \cap C(\overline{\partial\Omega}) \\ v &\rightarrow \gamma_1(v) = \frac{\partial v}{\partial n} \Big|_{\partial\Omega}, \end{aligned} \quad (2.19)$$

avec $\frac{\partial v}{\partial n} = \nabla u \cdot n$. Cette application γ_1 se prolonge par continuité en une application linéaire continue de $H^2(\Omega)$ dans $L^2(\partial\Omega)$. En particulier, il existe une constante $C > 0$ telle que, pour toute fonction $v \in H^2(\Omega)$, on a

$$\left\| \frac{\partial v}{\partial n} \right\|_{L^2(\partial\Omega)} \leq C \|v\|_{H^2(\Omega)}. \quad (2.20)$$

Démonstration. L'existence de l'application trace γ_1 (et ses propriétés) est une simple conséquence du précédent Théorème de trace 2.3.13 pour les fonctions de $H^1(\Omega)$. En effet, si $v \in H^2(\Omega)$, alors $\nabla v \in H^1(\Omega)^N$ et on peut donc définir la trace de ∇v sur $\partial\Omega$ comme une fonction de $L^2(\partial\Omega)^N$. Comme la normale est une fonction continue bornée sur $\partial\Omega$, on en déduit bien que $\frac{\partial v}{\partial n} \in L^2(\partial\Omega)$. \square

Le Théorème de trace 2.3.28 permet de généraliser aux fonctions de $H^2(\Omega)$ une formule de Green précédemment établie pour des fonctions de classe C^2 au Corollaire 1.2.4.

Théorème 2.3.30 *Soit Ω un ouvert borné régulier de classe C^2 . Si $u \in H^2(\Omega)$ et $v \in H^1(\Omega)$, on a*

$$\int_{\Omega} \Delta u(x)v(x) dx = - \int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx + \int_{\partial\Omega} \frac{\partial u}{\partial n}(x)v(x) ds. \quad (2.22)$$

Démonstration. Comme (2.22) est vraie pour des fonctions de classe C^2 et que les fonctions régulières sont denses dans $H^2(\Omega)$ et $H^1(\Omega)$, on utilise un argument de densité. Nous renvoyons à la démonstration du Théorème 2.3.15 pour plus de détails. Le seul argument nouveau ici est qu'il faut utiliser la continuité de l'application trace γ_1 , c'est-à-dire l'inégalité (2.20). \square

Lemme 2.2.4 (dérivation faible)	$u \in L^2(\Omega)$ est dérivable au sens faible si, $\forall i$, $\left \int_{\Omega} u \frac{\partial \phi}{\partial x_i} dx \right \leq C \ \phi\ _{L^2(\Omega)} \quad \forall \phi \in C_c^\infty(\Omega)$
Proposition 2.3.2	$H^1(\Omega)$ est un espace de Hilbert pour le produit scalaire $\langle u, v \rangle = \int_{\Omega} (\nabla u \cdot \nabla v + uv) dx$
Théorème 2.3.5 (théorème de densité)	$C_c^\infty(\bar{\Omega})$ est dense dans $H^1(\Omega)$
Proposition 2.3.10 (inégalité de Poincaré)	$\forall u \in H_0^1(\Omega)$ (Ω borné) $\ u\ _{L^2(\Omega)} \leq C \ \nabla u\ _{L^2(\Omega)}$
Théorème 2.3.13 (théorème de trace)	$u \rightarrow u _{\partial\Omega}$ application continue de $H^1(\Omega)$ dans $L^2(\partial\Omega)$
Théorème 2.3.15 (formule de Green)	$\forall u, v \in H^1(\Omega)$ $\int_{\Omega} u \frac{\partial v}{\partial x_i} dx = - \int_{\Omega} v \frac{\partial u}{\partial x_i} dx + \int_{\partial\Omega} uv n_i ds$
Corollaire 2.3.16 (caractérisation de $H_0^1(\Omega)$)	$H_0^1(\Omega)$ est le sous-espace des fonctions de $H^1(\Omega)$ qui s'annulent sur $\partial\Omega$
Théorème 2.3.21 (théorème de Rellich)	l'injection de $H^1(\Omega)$ dans $L^2(\Omega)$ est compacte (Ω borné régulier)
Théorème 2.3.30 (formule de Green)	$\forall u \in H^2(\Omega), v \in H^1(\Omega)$ $\int_{\Omega} v \Delta u dx = - \int_{\Omega} \nabla u \cdot \nabla v dx + \int_{\partial\Omega} \frac{\partial u}{\partial n} v ds$

TABLE 2.1 – Principaux résultats sur les espaces de Sobolev qu'il faut absolument connaître.

Chapitre 3

ÉTUDE MATHÉMATIQUE DES PROBLÈMES ELLIPTIQUES

3.1 Introduction

Dans ce chapitre nous terminons l'analyse mathématique des équations aux dérivées partielles de type elliptique commencée au Chapitre 1. Pour montrer que les problèmes aux limites sont bien posés pour ces e.d.p. elliptiques, c'est-à-dire qu'elles admettent une solution, unique, et dépendant continûment des données, nous suivons **l'approche variationnelle** présentée au Chapitre 1 et nous utilisons les **espaces de Sobolev** introduits au Chapitre 2.

Le plan de ce chapitre est le suivant. Dans la Section 3.2 nous expliquons en détail le fonctionnement de l'approche variationnelle pour le Laplacien avec divers types de conditions aux limites. Nous démontrons des **résultats d'existence et d'unicité des solutions**. Nous montrons aussi que ces solutions **minimisent une énergie** et qu'elles vérifient un certain nombre de **propriétés qualitatives** très naturelles et importantes du point de vue des applications (principe du maximum, régularité). La Section 3.3 reprend le même programme mais pour d'autres modèles plus compliqués comme celui de **l'élasticité linéarisée** ou celui des **équations de Stokes**. Si la théorie d'existence et d'unicité est très semblable au cas précédent, il n'en est pas de même de toutes les propriétés qualitatives.

3.2 Étude du Laplacien

3.2.1 Conditions aux limites de Dirichlet

Nous considérons le problème aux limites suivant

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \quad (3.1)$$

où Ω est un ouvert borné de l'espace \mathbb{R}^N , et f est un second membre qui appartient à l'espace $L^2(\Omega)$. L'approche variationnelle pour étudier (3.1) est constituée de trois étapes que nous détaillons.

Étape 1 : Établissement d'une formulation variationnelle.

Dans une première étape il faut proposer une formulation variationnelle du problème aux limites (3.1), c'est-à-dire qu'il faut trouver une forme bilinéaire $a(\cdot, \cdot)$, une forme linéaire $L(\cdot)$, et un espace de Hilbert V tels que (3.1) soit équivalent à :

$$\text{Trouver } u \in V \text{ tel que } a(u, v) = L(v) \text{ pour tout } v \in V. \quad (3.2)$$

Le but de cette première étape est seulement de trouver la formulation variationnelle (3.2) ; on vérifiera l'équivalence précise avec (3.1) plus tard au cours de la troisième étape.

Pour trouver la formulation variationnelle on multiplie l'équation (3.1) par une fonction test régulière v et on intègre par parties. Ce calcul est principalement formel au sens où l'on suppose l'existence et la régularité de la solution u afin que tous les calculs effectués soient licites. A l'aide de la formule de Green (2.22) (voir aussi (1.7)) on trouve

$$\int_{\Omega} f v \, dx = - \int_{\Omega} \Delta u v \, dx = \int_{\Omega} \nabla u \cdot \nabla v \, dx - \int_{\partial\Omega} \frac{\partial u}{\partial n} v \, ds. \quad (3.3)$$

Comme u doit satisfaire une condition aux limites de Dirichlet, $u = 0$ sur $\partial\Omega$, on choisit un espace de Hilbert V tel que toute fonction $v \in V$ vérifie aussi $v = 0$ sur $\partial\Omega$. Dans ce cas, l'égalité (3.3) devient

$$\int_{\Omega} \nabla u(x) \cdot \nabla v(x) \, dx = \int_{\Omega} f(x)v(x) \, dx. \quad (3.4)$$

Pour que le terme de gauche de (3.4) ait un sens il suffit que ∇u et ∇v appartiennent à $L^2(\Omega)$ (composante par composante), et pour que le terme de droite de (3.4) ait aussi un sens il suffit que v appartienne à $L^2(\Omega)$ (on a supposé que $f \in L^2(\Omega)$). Par conséquent, un choix raisonnable pour l'espace de Hilbert est $V = H_0^1(\Omega)$, le sous-espace de $H^1(\Omega)$ dont les éléments s'annulent sur le bord $\partial\Omega$.

En conclusion, la formulation variationnelle proposée pour (3.1) est :

$$\text{trouver } u \in H_0^1(\Omega) \text{ tel que } \int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H_0^1(\Omega). \quad (3.5)$$

Évidemment, nous avons fait un certain nombre de choix pour arriver à (3.5) ; d'autres choix nous auraient conduit à d'autres formulations variationnelles possibles. La justification de (3.5) s'effectuera donc a posteriori : tout d'abord, la deuxième étape consiste à vérifier que (3.5) admet bien une unique solution, puis la troisième étape que la solution de (3.5) est aussi une solution du problème aux limites (3.1) (dans un sens à préciser).

Étape 2 : Résolution de la formulation variationnelle.

Dans cette deuxième étape nous vérifions que la formulation variationnelle (3.5) admet une solution unique. Pour cela nous utilisons le Théorème de Lax-Milgram 1.3.1 dont nous vérifions les hypothèses avec les notations

$$a(u, v) = \int_{\Omega} \nabla u(x) \cdot \nabla v(x) \, dx \text{ et } L(v) = \int_{\Omega} f(x)v(x) \, dx.$$

On voit facilement en utilisant l'inégalité de Cauchy-Schwarz que a est une forme bilinéaire continue sur $H_0^1(\Omega)$ et que L est une forme linéaire continue sur $H_0^1(\Omega)$. De plus, en vertu de l'inégalité de Poincaré (voir le Corollaire 2.3.12 ; on utilise ici le caractère borné de l'ouvert Ω), la forme bilinéaire a est coercive, c'est-à-dire qu'il existe $\nu > 0$ tel que

$$a(v, v) = \int_{\Omega} |\nabla v(x)|^2 dx \geq \nu \|v\|_{H_0^1(\Omega)}^2 \quad \forall v \in H_0^1(\Omega).$$

Comme $H_0^1(\Omega)$ est un espace de Hilbert (voir la Proposition 2.3.9), toutes les hypothèses du Théorème de Lax-Milgram 1.3.1 sont satisfaites et on peut donc conclure qu'il existe une unique solution $u \in H_0^1(\Omega)$ de la formulation variationnelle (3.5).

Étape 3 : Équivalence avec l'équation.

La troisième étape (la dernière et la plus délicate) consiste à vérifier qu'en résolvant la formulation variationnelle (3.5) on a bien résolu le problème aux limites (3.1), et à préciser dans quel sens la solution de (3.5) est aussi une solution de (3.1). En d'autres termes, il s'agit d'interpréter la formulation variationnelle et de retourner à l'équation. Pour cela on procède aux mêmes intégrations par parties qui ont conduit à la formulation variationnelle, mais en sens inverse, et en les justifiant soigneusement.

Cette justification est très facile si l'on suppose que la solution u de la formulation variationnelle (3.5) est régulière (précisément si $u \in H^2(\Omega)$) et que l'ouvert Ω est aussi régulier, ce que nous faisons dans un premier temps. En effet, il suffit d'invoquer la formule de Green (2.22) qui nous donne, pour $v \in H_0^1(\Omega)$,

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = - \int_{\Omega} v \Delta u \, dx$$

puisque $v = 0$ sur le bord $\partial\Omega$. On en déduit alors

$$\int_{\Omega} (\Delta u + f) v \, dx = 0 \quad \forall v \in C_c^\infty(\Omega),$$

ce qui implique, en vertu du Corollaire 2.2.2, que $-\Delta u = f$ dans $L^2(\Omega)$, et on a l'égalité

$$-\Delta u = f \text{ presque partout dans } \Omega. \quad (3.6)$$

De plus, si Ω est un ouvert borné régulier de classe \mathcal{C}^1 , alors le Théorème de trace 2.3.13 (ou plus précisément son Corollaire 2.3.16) affirme que toute fonction de $H_0^1(\Omega)$ a une trace sur $\partial\Omega$ nulle dans $L^2(\Omega)$. On en déduit, en particulier, que

$$u = 0 \text{ presque partout sur } \partial\Omega. \quad (3.7)$$

On a donc bien retrouvé l'équation et la condition aux limites de (3.1).

Si l'on ne suppose plus que la solution u de (3.5) et l'ouvert Ω sont réguliers, il faut travailler davantage (on ne peut plus utiliser la formule de Green (2.22) qui

nécessite que $u \in H^2(\Omega)$). On note $\sigma = \nabla u$ qui est une fonction à valeurs vectorielles dans $L^2(\Omega)^N$. Par l'inégalité de Cauchy-Schwarz, on déduit de la formulation variationnelle (3.5) que, pour tout $v \in H_0^1(\Omega)$,

$$\left| \int_{\Omega} \sigma \cdot \nabla v \, dx \right| = \left| \int_{\Omega} f v \, dx \right| \leq C \|v\|_{L^2(\Omega)}. \quad (3.8)$$

Comme $C_c^\infty(\Omega) \subset H_0^1(\Omega)$, (3.8) n'est rien d'autre que le critère d'existence d'une divergence faible de σ dans $L^2(\Omega)$ (voir la Définition 2.2.6 et le Lemme 2.2.7) qui vérifie, pour tout $v \in C_c^\infty(\Omega)$,

$$\int_{\Omega} \sigma \cdot \nabla v \, dx = - \int_{\Omega} \operatorname{div} \sigma v \, dx.$$

On en déduit donc que

$$\int_{\Omega} (\operatorname{div} \sigma + f) v \, dx = 0 \quad \forall v \in C_c^\infty(\Omega),$$

ce qui implique, en vertu du Corollaire 2.2.2, que $-\operatorname{div} \sigma = f$ dans $L^2(\Omega)$. Par conséquent $\operatorname{div} \sigma = \Delta u$ appartient à $L^2(\Omega)$ (rappelons que $\operatorname{div} \nabla = \Delta$), et on retrouve l'équation (3.6). On retrouve la condition aux limites (3.7) comme précédemment si l'ouvert Ω est régulier de classe C^1 . Si Ω n'est pas régulier, alors on ne peut pas invoquer le Théorème de trace 2.3.13 pour obtenir (3.7). Néanmoins, le simple fait d'appartenir à $H_0^1(\Omega)$ est une généralisation de la condition aux limites de Dirichlet pour un ouvert non régulier, et on continuera à écrire **formellement** que $u = 0$ sur $\partial\Omega$.

En conclusion nous avons démontré le résultat suivant.

Théorème 3.2.2 *Soit Ω un ouvert borné de \mathbb{R}^N . Soit $f \in L^2(\Omega)$. Il existe une unique solution $u \in H_0^1(\Omega)$ de la formulation variationnelle (3.5). De plus, u vérifie*

$$-\Delta u = f \text{ presque partout dans } \Omega, \quad \text{et } u \in H_0^1(\Omega). \quad (3.9)$$

Si on suppose en plus que Ω est régulier de classe C^1 , alors u est solution du problème aux limites (3.1) au sens où

$$-\Delta u = f \text{ presque partout dans } \Omega, \quad u = 0 \text{ presque partout sur } \partial\Omega.$$

On appelle la solution $u \in H_0^1(\Omega)$ de la formulation variationnelle (3.5) **solution variationnelle** du problème aux limites (3.1). Par un raccourci de langage bien commode, **on dira que l'unique solution $u \in H_0^1(\Omega)$ de la formulation variationnelle (3.5) est l'unique solution du problème aux limites (3.1)**. Cette appellation est bien sûr justifiée par le Théorème 3.2.2.

La solution de (3.1), que nous venons d'obtenir, ne vérifie *a priori* l'équation et la condition aux limites que dans un sens "faible", c'est-à-dire presque partout (ou même pire pour la condition aux limites si l'ouvert n'est pas régulier). On parle alors de **solution faible** par opposition aux solutions fortes qu'on aurait pu espérer obtenir dans une formulation classique de (3.1). De même, on appelle parfois la formulation variationnelle **formulation faible** de l'équation.

Remarque 3.2.3 En fait, la solution faible peut être une solution forte si le second membre f est plus régulier. Autrement dit, l'équation et la condition aux limites de (3.1) peuvent être vérifiées en un sens classique, c'est-à-dire pour tout $x \in \Omega$, et tout $x \in \partial\Omega$, respectivement. C'est ce qu'on appelle un résultat de régularité pour la solution (voir plus loin le Corollaire 3.2.27). •

Pour que le problème aux limites (3.1) soit bien posé (au sens de Hadamard), il faut en plus de l'existence et de l'unicité de sa solution, montrer que la solution dépend continûment des données. C'est une conséquence immédiate du Théorème de Lax-Milgram 1.3.1 mais nous en donnons un nouvel énoncé et une nouvelle démonstration.

Proposition 3.2.5 Soit Ω un ouvert borné de \mathbb{R}^N , et soit $f \in L^2(\Omega)$. L'application qui à $f \in L^2(\Omega)$ fait correspondre la solution unique $u \in H_0^1(\Omega)$ de la formulation variationnelle de (3.1) est linéaire et continue de $L^2(\Omega)$ dans $H^1(\Omega)$. En particulier, il existe une constante $C > 0$ telle que, pour tout $f \in L^2(\Omega)$, on a

$$\|u\|_{H^1(\Omega)} \leq C \|f\|_{L^2(\Omega)}. \quad (3.10)$$

Remarque 3.2.6 L'inégalité (3.10) est ce qu'on appelle une **estimation d'énergie**. Elle garantit que l'énergie de la solution est contrôlée par celle de la donnée. Les estimations d'énergie sont très naturelles d'un point de vue physique et très utiles d'un point de vue mathématique. •

Démonstration. La linéarité de $f \rightarrow u$ est évidente. Pour obtenir la continuité on prend $v = u$ dans la formulation variationnelle (3.5)

$$\int_{\Omega} |\nabla u|^2 dx = \int_{\Omega} f u dx.$$

On majore le terme de droite à l'aide de l'inégalité de Cauchy-Schwarz, et on minore celui de gauche par la coercivité de la forme bilinéaire

$$\nu \|u\|_{H^1(\Omega)}^2 \leq \|f\|_{L^2(\Omega)} \|u\|_{L^2(\Omega)} \leq \|f\|_{L^2(\Omega)} \|u\|_{H^1(\Omega)},$$

d'où l'on déduit le résultat. □

Nous avons déjà dit que la formulation variationnelle possède souvent une interprétation physique (c'est, par exemple, le principe des travaux virtuels en mécanique). En fait, la solution de la formulation variationnelle (3.5) réalise le minimum d'une énergie (très naturelle en physique ou en mécanique). Le résultat suivant est une application immédiate de la Proposition 1.3.4.

Proposition 3.2.7 Soit $J(v)$ l'énergie définie pour $v \in H_0^1(\Omega)$ par

$$J(v) = \frac{1}{2} \int_{\Omega} |\nabla v|^2 dx - \int_{\Omega} f v dx. \quad (3.11)$$

Soit $u \in H_0^1(\Omega)$ la solution unique de la formulation variationnelle (3.5). Alors u est aussi l'unique point de minimum de l'énergie, c'est-à-dire que

$$J(u) = \min_{v \in H_0^1(\Omega)} J(v).$$

Réciproquement, si $u \in H_0^1(\Omega)$ est un point de minimum de l'énergie $J(v)$, alors u est la solution unique de la formulation variationnelle (3.5).

Remarque 3.2.8 La Proposition 3.2.7 repose de manière cruciale sur le fait que la forme bilinéaire de la formulation variationnelle est symétrique. Si cela n'est pas le cas, la solution de la formulation variationnelle ne minimise pas l'énergie (voir le contre-exemple de l'Exercice 3.2.3).

Souvent l'origine physique du Laplacien est en fait la recherche des minima de l'énergie $J(v)$. Il est remarquable que ce problème de minimisation nécessite de la part de la solution u moins de régularité que l'équation aux dérivées partielles (une seule dérivée permet de définir $J(u)$ tandis qu'il en faut deux pour Δu). Cette constatation confirme le caractère "naturel" de la formulation variationnelle pour analyser une équation aux dérivées partielles. •

Exercice 3.2.1 A l'aide de l'approche variationnelle démontrer l'existence et l'unicité de la solution de

$$\begin{cases} -\Delta u + u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \quad (3.12)$$

où Ω est un ouvert quelconque de l'espace \mathbb{R}^N , et $f \in L^2(\Omega)$. Montrer en particulier que l'ajout d'un terme d'ordre zéro au Laplacien permet de ne pas avoir besoin de l'hypothèse que Ω est borné.

Exercice 3.2.2 Soit Ω un ouvert borné de \mathbb{R}^N . A l'aide de l'approche variationnelle démontrer l'existence et l'unicité de la solution du problème suivant de convection-diffusion

$$\begin{cases} V \cdot \nabla u - \Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \quad (3.13)$$

où $f \in L^2(\Omega)$ et V est une fonction régulière à valeurs vectorielles telle que $\operatorname{div} V = 0$ dans Ω .

Exercice 3.2.3 On reprend les notations et hypothèses de l'Exercice 3.2.2. Montrer que tout $v \in H_0^1(\Omega)$ vérifie

$$\int_{\Omega} v V \cdot \nabla v \, dx = 0.$$

Montrer que la solution de la formulation variationnelle du problème de convection-diffusion ne minimise pas dans $H_0^1(\Omega)$ l'énergie

$$J(v) = \frac{1}{2} \int_{\Omega} (|\nabla v|^2 + v V \cdot \nabla v) \, dx - \int_{\Omega} f v \, dx.$$

3.2.2 Conditions aux limites de Neumann

Nous considérons le problème aux limites suivant

$$\begin{cases} -\Delta u + u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} = g & \text{sur } \partial\Omega \end{cases} \quad (3.16)$$

où Ω est un ouvert (non nécessairement borné) de l'espace \mathbb{R}^N , $f \in L^2(\Omega)$ et $g \in L^2(\partial\Omega)$. L'équation de (3.16) est une variante du Laplacien où nous avons ajouté un terme d'ordre zéro afin d'éviter (en premier abord) une difficulté que nous réglerons plus loin au Théorème 3.2.18. L'approche variationnelle pour étudier (3.16) est sensiblement différente de celle présentée à la sous-section précédente dans le traitement des conditions aux limites. C'est pourquoi nous détaillons à nouveau les trois étapes de l'approche.

Étape 1 : Établissement d'une formulation variationnelle.

Pour trouver la formulation variationnelle on multiplie l'équation (3.16) par une fonction test régulière v et on intègre par parties en admettant que la solution u est suffisamment régulière afin que tous les calculs effectués soient licites. La formule de Green (2.22) (voir aussi (1.7)) donne

$$\begin{aligned} \int_{\Omega} f(x)v(x) dx &= \int_{\Omega} (-\Delta u(x) + u(x))v(x) dx \\ &= \int_{\Omega} (\nabla u(x) \cdot \nabla v(x) + u(x)v(x)) dx - \int_{\partial\Omega} \frac{\partial u}{\partial n}(x)v(x) ds \quad (3.17) \\ &= \int_{\Omega} (\nabla u(x) \cdot \nabla v(x) + u(x)v(x)) dx - \int_{\partial\Omega} g(x)v(x) ds. \end{aligned}$$

Nous avons utilisé la condition aux limites de Neumann dans (3.17) et il n'est pas nécessaire de l'inscrire dans le choix de l'espace de Hilbert V . Pour que le premier et les deux derniers termes de (3.17) aient un sens il suffit de prendre $V = H^1(\Omega)$ (on utilise le Théorème de trace 2.3.13 pour justifier l'intégrale de bord).

En conclusion, la formulation variationnelle proposée pour (3.16) est : trouver $u \in H^1(\Omega)$ tel que

$$\int_{\Omega} (\nabla u \cdot \nabla v + uv) dx = \int_{\partial\Omega} gv ds + \int_{\Omega} fv dx \quad \forall v \in H^1(\Omega). \quad (3.18)$$

Les étapes suivantes justifieront le choix de (3.18).

Remarque 3.2.11 La principale différence entre la formulation variationnelle (3.18) pour une condition aux limites de Neumann et celle (3.5) pour une condition aux limites de Dirichlet vient de ce que la condition de Dirichlet est inscrite dans le choix de l'espace alors que la condition de Neumann apparaît dans la forme linéaire mais pas dans l'espace. La condition de Dirichlet est dite **essentielle** (ou explicite) car elle est forcée par l'appartenance à un espace, tandis que la condition de Neumann est dite **naturelle** (ou implicite) car elle découle de l'intégration par parties qui conduit à la formulation variationnelle. •

Étape 2 : Résolution de la formulation variationnelle.

Dans cette deuxième étape nous vérifions que la formulation variationnelle (3.18) admet une solution unique. Pour cela nous utilisons le Théorème de Lax-Milgram 1.3.1 dont nous vérifions les hypothèses avec les notations

$$a(u, v) = \int_{\Omega} (\nabla u \cdot \nabla v + uv) dx \text{ et } L(v) = \int_{\partial\Omega} gv ds + \int_{\Omega} fv dx.$$

En utilisant l'inégalité de Cauchy-Schwarz et à l'aide du Théorème de trace 2.3.13, on voit clairement que a est une forme bilinéaire continue sur $H^1(\Omega)$ et que L est une forme linéaire continue sur $H^1(\Omega)$. Par ailleurs, la forme bilinéaire a est manifestement coercive (c'est pour cela qu'on a ajouté un terme d'ordre zéro au Laplacien) car

$$a(v, v) = \|v\|_{H^1(\Omega)}^2 \quad \forall v \in H^1(\Omega).$$

Comme $H^1(\Omega)$ est un espace de Hilbert (voir la Proposition 2.3.2), toutes les hypothèses du Théorème de Lax-Milgram 1.3.1 sont satisfaites et on peut donc conclure qu'il existe une unique solution $u \in H^1(\Omega)$ de la formulation variationnelle (3.18).

Étape 3 : Équivalence avec l'équation.

Nous interprétons maintenant la formulation variationnelle (3.18) pour vérifier qu'on a bien résolu le problème aux limites (3.16), dans un sens à préciser. Nous allons supposer que les données sont régulières. Plus précisément nous allons supposer que nous sommes dans les conditions d'application du lemme suivant de régularité que nous admettrons (voir la sous-section 3.2.4 pour des résultats similaires).

Lemme 3.2.13 *Soit Ω un ouvert régulier de classe C^1 de \mathbb{R}^N . Soit $f \in L^2(\Omega)$ et g la trace sur $\partial\Omega$ d'une fonction de $H^1(\Omega)$. Alors la solution u de la formulation variationnelle (3.18) appartient à $H^2(\Omega)$.*

Grâce au Lemme 3.2.13 on peut utiliser la formule de Green du Théorème 2.3.30

$$\int_{\Omega} \Delta u(x)v(x) dx = - \int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx + \int_{\partial\Omega} \frac{\partial u}{\partial n}(x)v(x) ds. \quad (3.19)$$

qui est valable pour $u \in H^2(\Omega)$ et $v \in H^1(\Omega)$. Rappelons que l'intégrale de bord dans (3.19) a bien un sens à cause du Théorème de trace 2.3.28 qui affirme que pour $u \in H^2(\Omega)$ la dérivée normale $\frac{\partial u}{\partial n}$ a un sens dans $L^2(\partial\Omega)$. On déduit alors de (3.18) et (3.19) que, pour tout $v \in H^1(\Omega)$,

$$\int_{\Omega} (\Delta u - u + f)v dx = \int_{\partial\Omega} \left(\frac{\partial u}{\partial n} - g \right) v ds. \quad (3.20)$$

Si l'on prend $v \in C_c^\infty(\Omega) \subset H^1(\Omega)$ dans (3.20), le terme de bord disparaît et l'on déduit, en vertu du Corollaire 2.2.2, que $\Delta u - u + f = 0$ dans $L^2(\Omega)$, donc presque partout dans Ω . Par conséquent, le membre de gauche de (3.20) est nul, donc

$$\int_{\partial\Omega} \left(\frac{\partial u}{\partial n} - g \right) v ds = 0 \quad \forall v \in H^1(\Omega).$$

D'après le Théorème 2.3.5 les fonctions régulières sont denses dans $H^1(\Omega)$. Par conséquent, l'égalité ci-dessus est vraie pour toute fonction v régulière sur $\partial\Omega$, et on déduit d'une version du Corollaire 2.2.2, adaptée au bord $\partial\Omega$, que $\frac{\partial u}{\partial n} - g = 0$ dans $L^2(\partial\Omega)$, et donc presque partout sur $\partial\Omega$. En conclusion nous avons démontré le résultat suivant.

Théorème 3.2.14 *Soit Ω un ouvert régulier de classe C^1 de \mathbb{R}^N . Soit $f \in L^2(\Omega)$ et g la trace sur $\partial\Omega$ d'une fonction de $H^1(\Omega)$. Il existe une unique solution $u \in H^1(\Omega)$ de la formulation variationnelle (3.18). De plus, u appartient à $H^2(\Omega)$ et est solution de (3.16) au sens où*

$$-\Delta u + u = f \text{ presque partout dans } \Omega, \quad \frac{\partial u}{\partial n} = g \text{ presque partout sur } \partial\Omega.$$

Exercice 3.2.5 Démontrer que l'unique solution $u \in H^1(\Omega)$ de la formulation variationnelle (3.18) vérifie l'estimation d'énergie suivante

$$\|u\|_{H^1(\Omega)} \leq C (\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)}),$$

où $C > 0$ est une constante qui ne dépend pas de u , f et g .

Comme dans la sous-section précédente, on peut montrer que la solution de (3.16) minimise une énergie. Remarquons que, si $g = 0$, alors l'énergie (3.22) est la même que celle (3.11) définie pour le Laplacien avec conditions aux limites de Dirichlet (modulo le terme d'ordre zéro). Néanmoins, leurs minima ne sont en général pas les mêmes car on minimise sur deux espaces différents, à savoir $H_0^1(\Omega)$ et $H^1(\Omega)$. La Proposition suivante est une application immédiate de la Proposition 1.3.4.

Proposition 3.2.16 *Soit $J(v)$ l'énergie définie pour $v \in H^1(\Omega)$ par*

$$J(v) = \frac{1}{2} \int_{\Omega} (|\nabla v|^2 + |v|^2) dx - \int_{\Omega} f v dx - \int_{\partial\Omega} g v ds. \quad (3.22)$$

Soit $u \in H^1(\Omega)$ la solution unique de la formulation variationnelle (3.18). Alors u est aussi l'unique point de minimum de l'énergie, c'est-à-dire que

$$J(u) = \min_{v \in H^1(\Omega)} J(v).$$

Réciproquement, si $u \in H^1(\Omega)$ est un point de minimum de l'énergie $J(v)$, alors u est la solution unique de la formulation variationnelle (3.18).

Exercice 3.2.6 On suppose que Ω est un ouvert borné régulier de classe C^1 . A l'aide de l'approche variationnelle démontrer l'existence et l'unicité de la solution du Laplacien avec une condition aux limites de Fourier

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} + u = g & \text{sur } \partial\Omega \end{cases} \quad (3.23)$$

où $f \in L^2(\Omega)$ et g est la trace sur $\partial\Omega$ d'une fonction de $H^1(\Omega)$. On démontrera l'inégalité suivante (qui généralise celle de Poincaré)

$$\|v\|_{L^2(\Omega)} \leq C (\|v\|_{L^2(\partial\Omega)} + \|\nabla v\|_{L^2(\Omega)}) \quad \forall v \in H^1(\Omega).$$

Exercice 3.2.7 On suppose que Ω est un ouvert borné connexe. A l'aide de l'approche variationnelle démontrer l'existence et l'unicité de la solution du Laplacien avec des conditions aux limites mêlées

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} = 0 & \text{sur } \partial\Omega_N \\ u = 0 & \text{sur } \partial\Omega_D \end{cases} \quad (3.24)$$

où $f \in L^2(\Omega)$, et $(\partial\Omega_N, \partial\Omega_D)$ est une partition de $\partial\Omega$ telle que les mesures superficielles de $\partial\Omega_N$ et $\partial\Omega_D$ sont non nulles. (Utiliser la Remarque 2.3.18.)

Nous revenons maintenant au véritable opérateur Laplacien (sans ajout d'un terme d'ordre zéro comme dans (3.16)) et nous considérons le problème aux limites

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ \frac{\partial u}{\partial n} = g & \text{sur } \partial\Omega \end{cases} \quad (3.25)$$

où Ω est un ouvert borné connexe de l'espace \mathbb{R}^N , $f \in L^2(\Omega)$ et $g \in L^2(\partial\Omega)$. La difficulté nouvelle dans (3.25) par rapport à (3.16) est qu'il n'existe une solution que si les données f et g vérifient une **condition de compatibilité**. En effet, il est facile de voir que s'il existe une solution $u \in H^2(\Omega)$, alors intégrant l'équation sur Ω (ou bien utilisant la formule de Green (2.22)) on a nécessairement

$$\int_{\Omega} f(x) dx + \int_{\partial\Omega} g(x) ds = 0. \quad (3.26)$$

Remarquons aussi que si u est solution alors $u + C$, avec $C \in \mathbb{R}$, est aussi solution. En fait, (3.26) est une condition nécessaire et suffisante d'existence d'une solution dans $H^1(\Omega)$, unique à l'addition d'une constante près. Remarquons que, si l'ouvert Ω n'est pas connexe, alors il faut écrire (3.26) pour chaque composante connexe de Ω et l'unicité de la solution vaudra à l'addition près d'une constante par composante connexe (avec ces modifications tous les résultats qui suivent restent valables).

Remarque 3.2.17 Physiquement, la condition de compatibilité (3.26) s'interprète comme une **condition d'équilibre** : f correspond à une source volumique, et g à un flux entrant au bord. Pour qu'il existe un état stationnaire ou d'équilibre (c'est-à-dire une solution de (3.25)), il faut que ces deux termes se balancent parfaitement. De même, l'unicité "à une constante additive près" correspond à l'absence d'origine de référence sur l'échelle qui mesure les valeurs de u (comme pour la température, par exemple). •

Théorème 3.2.18 Soit Ω un ouvert borné connexe régulier de classe C^1 de \mathbb{R}^N . Soit $f \in L^2(\Omega)$ et $g \in L^2(\partial\Omega)$ qui vérifient la condition de compatibilité (3.26). Il existe une solution faible $u \in H^1(\Omega)$ de (3.25), unique à l'addition d'une constante près.

Démonstration. Pour trouver la formulation variationnelle on procède comme pour l'équation (3.16). Un calcul similaire conduit à

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\partial\Omega} gv \, ds + \int_{\Omega} fv \, dx$$

pour toute fonction test régulière v . Pour donner un sens à tous les termes de cette égalité, on pourrait choisir $H^1(\Omega)$ comme espace de Hilbert V , mais nous ne pourrions montrer la coercivité de la forme bilinéaire. Cette difficulté est intimement liée au fait que, si u est solution, alors $u + C$ est aussi solution. Pour éviter cet inconvénient, on lève l'indétermination de cette constante additive en ne travaillant qu'avec des fonctions de moyenne nulle. Autrement dit, on pose

$$V = \left\{ v \in H^1(\Omega), \int_{\Omega} v(x) \, dx = 0 \right\}$$

et la formulation variationnelle de (3.25) est :

$$\text{trouver } u \in V \text{ tel que } \int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\partial\Omega} gv \, ds + \int_{\Omega} fv \, dx \quad \forall v \in V. \quad (3.27)$$

Pour pouvoir appliquer le Théorème de Lax-Milgram à la formulation variationnelle (3.27), la seule hypothèse délicate à vérifier est la coercivité de la forme bilinéaire. Celle-ci s'obtient grâce à une généralisation de l'inégalité de Poincaré, connue sous le nom d'inégalité de Poincaré-Wirtinger : si Ω est borné et connexe, il existe une constante $C > 0$ telle que, pour tout $v \in H^1(\Omega)$,

$$\|v - m(v)\|_{L^2(\Omega)} \leq C \|\nabla v\|_{L^2(\Omega)} \text{ avec } m(v) = \frac{\int_{\Omega} v \, dx}{\int_{\Omega} dx}. \quad (3.28)$$

L'inégalité (3.28) se démontre par contradiction comme dans la deuxième démonstration de la Proposition 2.3.10 (nous laissons cela au lecteur en exercice). Comme $m(v) = 0$ pour tout $v \in V$, (3.28) prouve que $\|\nabla v\|_{L^2(\Omega)}$ est une norme dans V , équivalente à la norme usuelle $\|v\|_{H^1(\Omega)}$, et donc que la forme bilinéaire est coercive sur V . Finalement, pour montrer que la solution unique de (3.27) est bien une solution du problème aux limites (3.25), on procède comme précédemment lors de la démonstration du Théorème 3.2.14. On obtient ainsi pour tout $v \in V$,

$$\int_{\Omega} (\Delta u + f)v \, dx = \int_{\partial\Omega} \left(\frac{\partial u}{\partial n} - g \right) v \, ds. \quad (3.29)$$

Or, quelque soit $w \in H^1(\Omega)$, la fonction $v = w - m(w)$ appartient à V . En choisissant une telle fonction dans (3.29), en regroupant les termes en facteur de la constante $m(w)$ et en utilisant la condition de compatibilité (3.26) ainsi que l'égalité $\int_{\Omega} \Delta u \, dx = \int_{\partial\Omega} \frac{\partial u}{\partial n} \, ds$, on déduit donc de (3.29)

$$\int_{\Omega} (\Delta u + f)w \, dx = \int_{\partial\Omega} \left(\frac{\partial u}{\partial n} - g \right) w \, ds \quad \forall w \in H^1(\Omega).$$

On peut donc conclure comme d'habitude que u vérifie bien le problème aux limites (3.25). \square

Exercice 3.2.8 Démontrer l'inégalité de Poincaré-Wirtinger (3.28).

Exercice 3.2.9 On suppose que Ω est un ouvert borné connexe régulier. Soit $f \in L^2(\Omega)$. On considère la formulation variationnelle suivante : trouver $u \in H^1(\Omega)$ tel que

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx + \left(\int_{\Omega} u \, dx \right) \left(\int_{\Omega} v \, dx \right) = \int_{\Omega} f v \, dx \quad \forall v \in H^1(\Omega).$$

Démontrer l'existence et l'unicité de la solution de cette formulation variationnelle. Quel problème aux limites a-t-on ainsi résolu ? En particulier, si on suppose que $\int_{\Omega} f \, dx = 0$, quel problème déjà étudié retrouve-t-on ?

3.2.3 Coefficients variables

Dans les deux sous-sections précédentes nous avons considéré des problèmes aux limites pour l'opérateur Laplacien. On peut facilement généraliser les résultats obtenus à des opérateurs plus généraux, dits elliptiques du deuxième ordre à **coefficients variables**. Ce type de problème est issu de la modélisation des milieux hétérogènes. Si l'on reprend l'exemple de la conduction de la chaleur, dans un milieu hétérogène la conductivité $k(x)$ est une fonction variable d'un point à l'autre du domaine. Dans ce cas, on considère le problème aux limites

$$\begin{cases} -\operatorname{div}(k\nabla u) = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \quad (3.30)$$

où Ω est un ouvert borné de l'espace \mathbb{R}^N , et $f \in L^2(\Omega)$. Bien sûr, si $k(x) \equiv 1$, on retrouve le Laplacien. Il est facile de généraliser le Théorème 3.2.2.

Proposition 3.2.19 Soit Ω un ouvert borné de \mathbb{R}^N . Soit $f \in L^2(\Omega)$. On suppose que le coefficient $k(x)$ est une fonction mesurable et qu'il existe deux constantes strictement positives $0 < k^- \leq k^+$ telles que

$$0 < k^- \leq k(x) \leq k^+ \text{ presque partout } x \in \Omega. \quad (3.31)$$

Alors, il existe une unique solution (faible) $u \in H_0^1(\Omega)$ de (3.30).

Démonstration. Pour trouver la formulation variationnelle on multiplie l'équation (3.30) par une fonction test v et on intègre par parties en utilisant la formule de Green de l'Exercice 1.2.1

$$\int_{\Omega} \operatorname{div} \sigma(x) v(x) \, dx = - \int_{\Omega} \sigma(x) \cdot \nabla v(x) \, dx + \int_{\partial\Omega} \sigma(x) \cdot n(x) v(x) \, ds,$$

avec $\sigma = k\nabla u$. A cause de la condition aux limites de Dirichlet on choisit $H_0^1(\Omega)$ comme espace de Hilbert, et on trouve la formulation variationnelle de (3.30) :

$$\text{trouver } u \in H_0^1(\Omega) \text{ tel que } \int_{\Omega} k\nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H_0^1(\Omega). \quad (3.32)$$

Grâce à l'hypothèse (3.31) on vérifie que la forme bilinéaire de (3.32) est continue

$$\left| \int_{\Omega} k \nabla u \cdot \nabla v \, dx \right| \leq k^+ \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)},$$

et qu'elle est coercive

$$\int_{\Omega} k \nabla u \cdot \nabla u \, dx \geq k^- \int_{\Omega} |\nabla u|^2 \, dx \geq \nu \|u\|_{H_0^1(\Omega)},$$

avec $\nu > 0$ grâce à l'inégalité de Poincaré. On peut donc appliquer le Théorème de Lax-Milgram, ce qui démontre l'existence et l'unicité de la solution de la formulation variationnelle (3.32). Pour montrer que cette solution variationnelle est bien une solution du problème aux limites (3.30), on procède comme précédemment lors de la démonstration du Théorème 3.2.2. \square

Il est très important de comprendre que l'équation (3.30) admet une solution même si k est discontinu! En fait, son interprétation "au sens faible" contient plus d'informations que son écriture classique. Plus précisément, la notion de divergence faible contient implicitement ce qu'on appelle les **conditions aux limites de transmission** entre deux sous-domaines occupés par deux matériaux de conductivité différente. Considérons un exemple où (Ω_1, Ω_2) est une partition de Ω sur laquelle $k(x)$ est constant par morceaux

$$k(x) = k_i > 0 \text{ pour } x \in \Omega_i, \quad i = 1, 2. \quad (3.33)$$

On note $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$ l'interface (supposée régulière et incluse dans Ω) entre Ω_1 et Ω_2 , et $u_i = u|_{\Omega_i}$ la restriction de la solution u à Ω_i .

Lemme 3.2.21 *Sous l'hypothèse (3.33) le problème (3.30) est équivalent à*

$$\begin{cases} -k_i \Delta u_i = f & \text{dans } \Omega_i, \quad i = 1, 2, \\ u_1 = 0 & \text{sur } \partial\Omega, \\ u_1 = u_2 & \text{sur } \Gamma, \\ k_1 \nabla u_1 \cdot n = k_2 \nabla u_2 \cdot n & \text{sur } \Gamma. \end{cases} \quad (3.34)$$

Les deux dernières lignes de (3.34) sont appelées **conditions aux limites de transmission** sur l'interface Γ .

Démonstration. Si $u \in H_0^1(\Omega)$ est solution de (3.30), par application du Théorème de trace 2.3.13, on doit avoir $u_1 = u_2$ sur Γ . Si on pose $\sigma = k \nabla u$ et $\sigma_i = \sigma|_{\Omega_i} = k_i \nabla u_i$ sa restriction à Ω_i , on sait que σ , ainsi que sa divergence, appartiennent à $L^2(\Omega)$. Alors, un théorème de trace permet d'affirmer que la composante normale $\sigma \cdot n$ a un sens sur Γ et on doit donc avoir $\sigma_1 \cdot n = \sigma_2 \cdot n$ sur Γ (voir l'Exercice 2.2.4).

Réciproquement, on construit une formulation variationnelle de (3.34) pour montrer qu'elle admet une unique solution qui coïncide avec la solution u de (3.30). On cherche $u_i \in H^1(\Omega_i)$ et on multiplie chaque équation dans Ω_i par la même fonction test $v \in H_0^1(\Omega)$. En intégrant par parties et en sommant on obtient

$$\int_{\Omega_1} \nabla u_1 \cdot \nabla v \, dx + \int_{\Omega_2} \nabla u_2 \cdot \nabla v \, dx + \int_{\Gamma} \left(k_1 \frac{\partial u_1}{\partial n_1} + k_2 \frac{\partial u_2}{\partial n_2} \right) v \, ds = \int_{\Omega_1} f v \, dx + \int_{\Omega_2} f v \, dx. \quad (3.35)$$

Comme $n_1 = -n_2$ l'intégrale sur l'interface Γ disparaît à cause de la condition aux limites de transmission. D'autre part, si u est défini comme u_1 dans Ω_1 et u_2 dans Ω_2 , la condition de transmission $u_1 = u_2$ sur Γ implique que $u \in H^1(\Omega)$. Par conséquent, (3.35) n'est rien d'autre que la formulation variationnelle (3.32). \square

3.2.4 Propriétés qualitatives

Dans cette sous-section nous étudions quelques propriétés qualitatives des solutions du Laplacien avec conditions aux limites de Dirichlet. Dans toute cette sous-section, Ω est un ouvert borné de \mathbb{R}^N et $f \in L^2(\Omega)$.

Principe du maximum

Théorème 3.2.22 (Principe du maximum) *Si $f \geq 0$ presque partout dans Ω , alors la solution de (3.1) vérifie $u \geq 0$ presque partout dans Ω .*

Remarque 3.2.23 Le principe du maximum ne fait que traduire une propriété parfaitement naturelle du point de vue physique : par exemple dans le contexte de l'équation stationnaire de la chaleur, si on chauffe ($f \geq 0$), la température intérieure est toujours plus grande que la température au bord ($u \geq 0$). Le principe du maximum reste valable si on remplace le Laplacien par l'opérateur plus général (3.30) à coefficients variables dans $L^\infty(\Omega)$ ou bien si l'on considère le problème (3.16) avec condition aux limites de Neumann. La validité du principe du maximum est fondamentalement liée au caractère "scalaire" de l'équation (c'est-à-dire que l'inconnue u est à valeurs dans \mathbb{R}). Ce principe du maximum tombe généralement en défaut si l'inconnue u est à valeurs vectorielles (par exemple pour le système (3.56) de l'élasticité). \bullet

Démonstration. On utilise la formulation variationnelle (3.5) de (3.1) avec $v = u^- = \min(u, 0)$ qui appartient bien à $H_0^1(\Omega)$ en vertu du Lemme 3.2.24 (car $u = u^+ + u^-$). On a

$$\int_{\Omega} f u^- dx = \int_{\Omega} \nabla u \cdot \nabla u^- dx = \int_{\Omega} 1_{u < 0} \nabla u \cdot \nabla u dx = \int_{\Omega} |\nabla u^-|^2 dx \geq 0. \quad (3.41)$$

Mais $u^- \leq 0$ et $f \geq 0$ presque partout dans Ω . Par conséquent, tous les termes de (3.41) sont nuls, et comme $u^- \in H_0^1(\Omega)$ on en déduit que $u^- = 0$, c'est-à-dire que $u \geq 0$ presque partout dans Ω . \square

Lemme 3.2.24 *Si $v \in H_0^1(\Omega)$, alors $v^+ = \max(v, 0)$ appartient à $H_0^1(\Omega)$ et*

$$\nabla v^+ = 1_{v > 0} \nabla v \text{ presque partout dans } \Omega,$$

où $1_{v > 0}(x)$ est la fonction qui vaut 1 là où $v(x) > 0$ et 0 ailleurs.

Régularité

Théorème 3.2.26 (de régularité) *Soit un entier $m \geq 0$. Soit Ω un ouvert borné de \mathbb{R}^N de classe C^{m+2} . Soit $f \in H^m(\Omega)$. Alors, l'unique solution $u \in H_0^1(\Omega)$ de (3.1) appartient à $H^{m+2}(\Omega)$. De plus, l'application $f \rightarrow u$ est linéaire continue de $H^m(\Omega)$ dans $H^{m+2}(\Omega)$, c'est-à-dire qu'il existe une constante $C > 0$ telle que*

$$\|u\|_{H^{m+2}(\Omega)} \leq C \|f\|_{H^m(\Omega)}.$$

Par application immédiate du Théorème de régularité 3.2.26 et du Théorème 2.3.25 (sur la continuité des fonctions de $H^m(\Omega)$), on montre que les solutions faibles d'équations aux dérivées partielles elliptiques sont en fait des solutions fortes, ou classiques, si les données sont régulières (comme annoncé dans la Remarque 3.2.3).

Corollaire 3.2.27 *Si Ω est un ouvert borné de \mathbb{R}^N de classe C^{m+2} , si $f \in H^m(\Omega)$, et si $m > N/2$, alors la solution variationnelle $u \in H_0^1(\Omega)$ de (3.1) est une solution forte car elle appartient à $C^2(\overline{\Omega})$.*

En particulier, si Ω est un ouvert borné de \mathbb{R}^N de classe C^∞ , et si $f \in C^\infty(\overline{\Omega})$, alors la solution $u \in H_0^1(\Omega)$ de (3.1) est aussi dans $C^\infty(\overline{\Omega})$.

Remarque 3.2.28 Le Théorème de régularité 3.2.26 et son Corollaire 3.2.27 restent valables pour des conditions aux limites de Neumann. Ils se généralisent aussi au cas des opérateurs elliptiques à coefficients variables (comme dans la Sous-section 3.2.3). Dans ce dernier cas, il faut ajouter à l'hypothèse habituelle de coercivité des coefficients, l'hypothèse que les coefficients sont de classe C^{m+1} dans Ω (tandis que Ω est borné régulier de classe C^{m+2} et que $f \in H^m(\Omega)$). •

Exemple de singularité

Voyons maintenant un exemple de solutions **singulières**, c'est-à-dire non régulières. Il s'agit d'un problème posé dans un ouvert non régulier pour lequel le Théorème de régularité 3.2.26 et son Corollaire 3.2.27 sont faux. En particulier, bien qu'il existe des solutions faibles (c'est-à-dire appartenant à l'espace de Sobolev $H^1(\Omega)$) il n'y a pas de solutions fortes (c'est-à-dire deux fois dérivables) pour ce problème. Nous nous plaçons en dimension $N = 2$ d'espace, et nous considérons un secteur angulaire Ω défini en coordonnées radiales par (voir la Figure 3.2)

$$\Omega = \{(r, \theta) \text{ tel que } 0 \leq r < R \text{ et } 0 < \theta < \Theta\}$$

avec $0 < R < +\infty$ et $0 < \Theta \leq 2\pi$ (rappelons que $x_1 = r \cos \theta$ et $x_2 = r \sin \theta$). On note Γ_0 la partie du bord de Ω où $r = R$, Γ_1 celle où $\theta = 0$ et Γ_2 celle où $\theta = \Theta$. Cet ouvert Ω présente trois "coins", mais seule l'origine (le coin entre les bords Γ_1 et Γ_2) peut poser problème pour la régularité dans les exemples ci-dessous. Physiquement, le cas d'un angle $\Theta < \pi$ est représentatif d'un effet de pointe, tandis que le cas $\Theta > \pi$ correspond à une encoche (ou même à une **fissure** si $\Theta = 2\pi$).

Pour un entier $k \geq 1$, on étudie le problème aux limites suivant

$$\begin{cases} -\Delta u = 0 & \text{dans } \Omega \\ u = \cos\left(\frac{k\pi\theta}{\Theta}\right) & \text{sur } \Gamma_0 \\ \frac{\partial u}{\partial n} = 0 & \text{sur } \Gamma_1 \cup \Gamma_2 \end{cases} \quad (3.50)$$

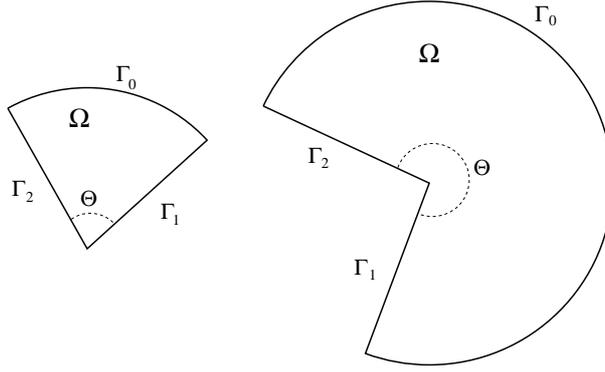


FIGURE 3.2 – Secteur angulaire Ω d'angle Θ (plus petit ou plus grand que π).

Lemme 3.2.33 *Il existe une unique solution faible de (3.50) dans $H^1(\Omega)$, donnée par la formule*

$$u(r, \theta) = \left(\frac{r}{R}\right)^{\frac{k\pi}{\Theta}} \cos\left(\frac{k\pi\theta}{\Theta}\right). \quad (3.52)$$

Si $k = 1$ et $\pi < \Theta$, alors le gradient ∇u n'est pas borné à l'origine, tandis que si $k \geq 2$ ou $\pi \geq \Theta$, alors le gradient ∇u est continu à l'origine.

Remarque 3.2.34 Puisque le gradient n'est pas borné, la solution de (3.50) ne peut pas être régulière au sens du Corollaire 3.2.27. D'un point de vue physique ou mécanique, ce gradient correspond à un flux de chaleur, au champ électrique, ou à un champ de contraintes : il importe de savoir si cette quantité est bornée continue ou non à l'origine. Physiquement, on interprète ce résultat de régularité en disant qu'une encoche ($\Theta > \pi$) engendre une singularité, au contraire d'une pointe ($\Theta \leq \pi$).

Dans le cas où $\Theta = 2\pi$, l'ouvert Ω est un domaine **fissuré** et le Lemme 3.2.33 a une interprétation mécanique très importante car le problème (3.50) modélise le cisaillement anti-plan d'un cylindre de base Ω et sa solution permet de calculer le **facteur d'intensité des contraintes** qui est souvent utilisé dans des modèles de propagation de fissures ou de rupture (voir par exemple [19]). •

Démonstration. On commence par “relever” la condition aux limites non-homogène en définissant une fonction $u_0 \in H^1(\Omega)$ dont la trace sur le bord coïncide avec cette condition aux limites. On vérifie aisément que

$$u_0(r, \theta) = \frac{r^2}{R^2} \cos\left(\frac{k\pi\theta}{\Theta}\right)$$

répond à la question, c'est-à-dire que $u = u_0 + v$ où v est la solution d'un problème homogène

$$\begin{cases} -\Delta v = \Delta u_0 & \text{dans } \Omega \\ v = 0 & \text{sur } \Gamma_0 \\ \frac{\partial v}{\partial n} = 0 & \text{sur } \Gamma_1 \cup \Gamma_2. \end{cases} \quad (3.54)$$

Remarquons que ce relèvement est possible car la donnée sur Γ_0 est compatible avec les conditions aux limites sur Γ_1 et Γ_2 , c'est-à-dire dans le cas présent que sa dérivée

en θ (i.e. sa dérivée normale) s'annule en $\theta = 0$, ou Θ . Comme Δu_0 appartient à $L^2(\Omega)$, il existe une unique solution de (3.54) dans $H^1(\Omega)$, et par conséquent (3.50) admet aussi une unique solution dans $H^1(\Omega)$. Vérifions que (3.52) est précisément cette unique solution. Rappelons que

$$\Delta\phi(r, \theta) = \frac{\partial^2\phi}{\partial r^2} + \frac{1}{r} \frac{\partial\phi}{\partial r} + \frac{1}{r^2} \frac{\partial^2\phi}{\partial\theta^2}.$$

Un calcul simple montre que

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} \right) \left(\frac{r}{R} \right)^{\frac{k\pi}{\Theta}} = \left(\frac{k\pi}{\Theta} \right)^2 \frac{1}{r^2} \left(\frac{r}{R} \right)^{\frac{k\pi}{\Theta}},$$

donc (3.52) est bien solution de (3.50). On vérifie aussi facilement que (3.52) appartient à $H^1(\Omega)$. Finalement, la formule du gradient en coordonnées radiales dans la base (e_r, e_θ)

$$\nabla\phi(r, \theta) = \frac{\partial\phi}{\partial r} e_r + \frac{1}{r} \frac{\partial\phi}{\partial\theta} e_\theta$$

nous donne

$$\nabla u = \left(\frac{r}{R} \right)^{\frac{k\pi}{\Theta}-1} \frac{k\pi}{R\Theta} \left(\cos\left(\frac{k\pi\theta}{\Theta} \right) e_r - \sin\left(\frac{k\pi\theta}{\Theta} \right) e_\theta \right)$$

qui, clairement, n'est pas borné à l'origine si $0 < k\pi < \Theta$ et est continu à l'origine si $k\pi > \Theta$. Le cas limite $k\pi = \Theta$ correspond aussi à un gradient continu car en fait $u = x_1/R$! \square

3.3 Résolution d'autres modèles

3.3.1 Système de l'élasticité linéarisée

Nous appliquons l'approche variationnelle à la résolution du système d'équations de l'élasticité linéarisée. Ces équations modélisent les déformations d'un solide sous l'hypothèse de petites déformations et de petits déplacements (hypothèse qui permet d'obtenir des équations linéaires; d'où le nom d'élasticité **linéarisée**, voir par exemple [25]). On considère les équations stationnaires de l'élasticité, c'est-à-dire indépendantes du temps. Soit Ω un ouvert borné de \mathbb{R}^N . Soit une force $f(x)$, une fonction de Ω dans \mathbb{R}^N . L'inconnue u (le déplacement) est aussi une fonction de Ω dans \mathbb{R}^N . La modélisation mécanique fait intervenir le tenseur des déformations, noté $e(u)$, qui est une fonction à valeurs dans l'ensemble des matrices symétriques

$$e(u) = \frac{1}{2} \left(\nabla u + (\nabla u)^t \right) = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)_{1 \leq i, j \leq N},$$

ainsi que le tenseur des contraintes σ (une autre fonction à valeurs dans l'ensemble des matrices symétriques) qui est relié à $e(u)$ par la loi de Hooke

$$\sigma = 2\mu e(u) + \lambda \operatorname{tr}(e(u)) \operatorname{Id},$$

où λ et μ sont les coefficients de Lamé du matériau homogène isotrope qui occupe Ω . Pour des raisons de thermodynamique les coefficients de Lamé vérifient

$$\mu > 0 \text{ et } 2\mu + N\lambda > 0.$$

On ajoute à cette loi constitutive le bilan des forces dans le solide

$$-\operatorname{div}\sigma = f \text{ dans } \Omega$$

où, par définition, la divergence de σ est le vecteur de composantes

$$\operatorname{div}\sigma = \left(\sum_{j=1}^N \frac{\partial \sigma_{ij}}{\partial x_j} \right)_{1 \leq i \leq N}.$$

Utilisant le fait que $\operatorname{tr}(e(u)) = \operatorname{div}u$, on en déduit les équations pour $1 \leq i \leq N$

$$-\sum_{j=1}^N \frac{\partial}{\partial x_j} \left(\mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \lambda(\operatorname{div}u)\delta_{ij} \right) = f_i \text{ dans } \Omega \quad (3.55)$$

avec f_i et u_i , pour $1 \leq i \leq N$, les composantes de f et u dans la base canonique de \mathbb{R}^N . En ajoutant une condition aux limites de Dirichlet, et en utilisant des notations vectorielles, le problème aux limites considéré est

$$\begin{cases} -\operatorname{div}(2\mu e(u) + \lambda \operatorname{tr}(e(u)) \operatorname{Id}) = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega. \end{cases} \quad (3.56)$$

Nous pouvons énoncer et démontrer un premier résultat d'existence et d'unicité du système de l'élasticité linéarisé.

Théorème 3.3.1 *Soit Ω un ouvert borné de \mathbb{R}^N . Soit $f \in L^2(\Omega)^N$. Il existe une unique solution (faible) $u \in H_0^1(\Omega)^N$ de (3.56).*

Démonstration. Pour trouver la formulation variationnelle on multiplie chaque équation (3.55) par une fonction test v_i (qui s'annule sur le bord $\partial\Omega$ pour prendre en compte la condition aux limites de Dirichlet) et on intègre par parties pour obtenir

$$\int_{\Omega} \mu \sum_{j=1}^N \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \frac{\partial v_i}{\partial x_j} dx + \int_{\Omega} \lambda \operatorname{div}u \frac{\partial v_i}{\partial x_i} dx = \int_{\Omega} f_i v_i dx.$$

On somme alors ces équations, pour i allant de 1 à N , afin de faire apparaître la divergence de la fonction $v = (v_1, \dots, v_N)$ et de simplifier la première intégrale car

$$\sum_{i,j=1}^N \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \frac{\partial v_i}{\partial x_j} = \frac{1}{2} \sum_{i,j=1}^N \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) = 2e(u) \cdot e(v).$$

Choisissant $H_0^1(\Omega)^N$ comme espace de Hilbert, on obtient alors la formulation variationnelle : trouver $u \in H_0^1(\Omega)^N$ tel que

$$\int_{\Omega} 2\mu e(u) \cdot e(v) dx + \int_{\Omega} \lambda \operatorname{div}u \operatorname{div}v dx = \int_{\Omega} f \cdot v dx \quad \forall v \in H_0^1(\Omega)^N. \quad (3.57)$$

On vérifie aisément que chaque terme de (3.57) a bien un sens.

Pour pouvoir appliquer le Théorème de Lax-Milgram 1.3.1 à la formulation variationnelle (3.57), la seule hypothèse délicate à vérifier est la coercivité de la forme bilinéaire. On procède en trois étapes. Premièrement, montrons que

$$\int_{\Omega} 2\mu|e(v)|^2 dx + \int_{\Omega} \lambda|\operatorname{div}v|^2 dx \geq \nu \int_{\Omega} |e(v)|^2 dx,$$

avec $\nu = \min(2\mu, (2\mu + N\lambda)) > 0$. Pour cela, on utilise une inégalité algébrique : si on note $A \cdot B = \sum_{i,j=1}^N a_{ij}b_{ij}$ le produit scalaire usuel des matrices symétriques, on peut décomposer toute matrice réelle symétrique A sous la forme

$$A = A^d + A^h \text{ avec } A^d = A - \frac{1}{N} \operatorname{tr}A \operatorname{Id} \text{ et } A^h = \frac{1}{N} \operatorname{tr}A \operatorname{Id},$$

de telle manière que $A^d \cdot A^h = 0$ et $|A|^2 = |A^d|^2 + |A^h|^2$. On a alors

$$2\mu|A|^2 + \lambda(\operatorname{tr}A)^2 = 2\mu|A^d|^2 + (2\mu + N\lambda)|A^h|^2 \geq \nu|A|^2$$

avec $\nu = \min(2\mu, (2\mu + N\lambda))$, ce qui donne le résultat pour $A = e(u)$. Le fait que $\nu > 0$ n'est pas un hasard : les arguments mécaniques et thermodynamiques qui conduisent aux inégalités $\mu > 0$ et $(2\mu + N\lambda) > 0$ sont précisément les mêmes. Deuxièmement, on utilise l'inégalité de Korn (ou plutôt un cas particulièrement simple de cette inégalité, voir le Lemme 3.3.2 ci-dessous) qui donne une constante $C > 0$ telle que

$$\int_{\Omega} |e(v)|^2 dx \geq C \int_{\Omega} |\nabla v|^2 dx$$

pour tout $v \in H_0^1(\Omega)^N$. Troisièmement, on utilise l'inégalité de Poincaré (composante par composante, voir la Proposition 2.3.10) qui donne une constante $C > 0$ telle que, pour tout $v \in H_0^1(\Omega)^N$,

$$\int_{\Omega} |v|^2 dx \leq C \int_{\Omega} |\nabla v|^2 dx.$$

Au total, ces trois inégalités conduisent à la coercivité

$$\int_{\Omega} 2\mu|e(v)|^2 dx + \int_{\Omega} \lambda|\operatorname{div}v|^2 dx \geq C\|v\|_{H^1(\Omega)}^2.$$

Le Théorème de Lax-Milgram 1.3.1 donne donc l'existence et l'unicité de la solution de la formulation variationnelle (3.57). Finalement, pour montrer que la solution unique de (3.57) est bien une solution du problème aux limites (3.56), on procède comme lors de la démonstration du Théorème 3.2.2 pour le Laplacien. \square

Lemme 3.3.2 *Soit Ω un ouvert de \mathbb{R}^N . Pour toute fonction $v \in H_0^1(\Omega)^N$, on a*

$$\|\nabla v\|_{L^2(\Omega)} \leq \sqrt{2}\|e(v)\|_{L^2(\Omega)}. \quad (3.58)$$

Démonstration. Soit $v \in C_c^\infty(\Omega)^N$. Par intégration par parties on obtient

$$2 \int_{\Omega} |e(v)|^2 dx = \int_{\Omega} |\nabla v|^2 dx + \int_{\Omega} \nabla v \cdot (\nabla v)^t dx = \int_{\Omega} |\nabla v|^2 dx + \int_{\Omega} |\operatorname{div}v|^2 dx.$$

Par densité de $C_c^\infty(\Omega)$ dans $H_0^1(\Omega)$, on en déduit (3.58). \square

Exercice 3.3.1 Montrer que l'application de $L^2(\Omega)^N$ dans $H_0^1(\Omega)^N$ qui à f fait correspondre u , l'unique solution faible de (3.56), est linéaire continue.

L'analyse du problème aux limites (3.56) avec une condition aux limites de Dirichlet **sur tout le bord** $\partial\Omega$ est un peu trompeuse par sa simplicité. En effet, dès que l'on introduit une autre condition aux limites (par exemple, de Neumann) sur une partie du bord, la démonstration de la coercivité de la formulation variationnelle devient beaucoup plus difficile car on doit remplacer l'inégalité élémentaire du Lemme 3.3.2 par sa généralisation, nettement plus technique, dite inégalité de Korn (voir le Lemme 3.3.3 ci-dessous). Rappelons qu'on ne peut pas, en général, se contenter d'une condition aux limites de Dirichlet sur l'intégralité du bord $\partial\Omega$ car elle s'interprète comme le fait que le solide est fixé et immobile sur son bord. En pratique, tout le bord n'est pas bloqué et souvent une partie du bord est libre de bouger, ou bien des forces surfaciques sont appliquées sur une autre partie. Ces deux cas de figures sont modélisés par des conditions aux limites de Neumann qui s'écrivent ici

$$\sigma n = g \text{ sur } \partial\Omega, \quad (3.59)$$

où g est une fonction à valeurs vectorielles. La condition de Neumann (3.59) s'interprète en disant que g est une force appliquée sur le bord (plus précisément, g est une densité de forces surfacique, homogène à une pression, tandis que f est une densité de forces volumique). Si $g = 0$, aucune force ne s'applique et le bord peut bouger sans restriction : on dit que le bord est libre.

Nous allons maintenant considérer le système de l'élasticité avec des conditions aux limites mêlées (un mélange de Dirichlet et de Neumann), c'est-à-dire

$$\begin{cases} -\operatorname{div}(2\mu e(u) + \lambda \operatorname{tr}(e(u)) \operatorname{Id}) = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega_D \\ \sigma n = g & \text{sur } \partial\Omega_N, \end{cases} \quad (3.60)$$

où $(\partial\Omega_N, \partial\Omega_D)$ est une partition de $\partial\Omega$ telle que les mesures superficielles de $\partial\Omega_N$ et $\partial\Omega_D$ sont non nulles. L'analyse de ce nouveau problème aux limites est plus compliquée que dans le cas de la condition aux limites de Dirichlet : il faudra utiliser l'inégalité de Korn ci-dessous.

Lemme 3.3.3 (Inégalité de Korn) *Soit Ω un ouvert borné régulier de classe C^1 de \mathbb{R}^N . Il existe une constante $C > 0$ telle que, pour toute fonction $v \in H^1(\Omega)^N$, on a*

$$\|v\|_{H^1(\Omega)} \leq C \left(\|v\|_{L^2(\Omega)}^2 + \|e(v)\|_{L^2(\Omega)}^2 \right)^{1/2}. \quad (3.61)$$

L'inégalité (3.61) n'est pas une banalité : en effet, son membre de gauche contient toutes les dérivées partielles de v alors que son membre de droite fait intervenir seulement certaines combinaisons linéaires des dérivées partielles. Comme l'inégalité inverse de (3.61) est évidente, on en déduit que les deux membres de (3.61) sont des normes équivalentes. La démonstration du Lemme 3.3.3 est compliquée et dépasse le cadre de ce cours. Nous l'admettons donc en remarquant que nous avons bien démontré l'inégalité de Korn (3.61) lorsque v appartient à $H_0^1(\Omega)^N$. En effet, dans

ce cas une combinaison du Lemme 3.3.2 et de l'inégalité de Poincaré (pour un ouvert borné) donne bien l'inégalité (3.61).

L'interprétation mécanique de l'inégalité de Korn est la suivante. L'énergie élastique, proportionnelle à la norme du tenseur des déformations $e(u)$ dans $L^2(\Omega)$, contrôle la norme du déplacement u dans $H^1(\Omega)^N$, à l'addition près de la norme de u dans $L^2(\Omega)$. Comme nous allons le voir dans l'Exercice 3.3.2, ce dernier ajout est destiné à prendre en compte les **mouvements de corps rigides** c'est-à-dire les déplacements u non nuls mais d'énergie élastique nulle.

Exercice 3.3.2 Soit Ω un ouvert connexe de \mathbb{R}^N . Soit l'ensemble \mathcal{R} des "mouvements rigides" de Ω défini par

$$\mathcal{R} = \{v(x) = b + Mx \text{ avec } b \in \mathbb{R}^N, M = -M^t \text{ matrice antisymétrique}\}. \quad (3.62)$$

Montrer que $v \in H^1(\Omega)^N$ vérifie $e(v) = 0$ dans Ω si et seulement si $v \in \mathcal{R}$.

Nous pouvons maintenant énoncer un deuxième résultat d'existence et d'unicité pour le système de l'élasticité linéarisé avec conditions aux limites mêlées.

Théorème 3.3.4 Soit Ω un ouvert borné connexe régulier de classe \mathcal{C}^1 de \mathbb{R}^N . Soit $f \in L^2(\Omega)^N$ et $g \in L^2(\partial\Omega_N)^N$. On définit l'espace

$$V = \{v \in H^1(\Omega)^N \text{ tel que } v = 0 \text{ sur } \partial\Omega_D\}. \quad (3.63)$$

Il existe une unique solution (faible) $u \in V$ de (3.60) qui dépend linéairement et continûment des données f et g .

Démonstration. La formulation variationnelle de (3.60) s'obtient comme dans la démonstration du Théorème 3.3.1. L'espace V , défini par (3.63), contient la condition aux limites de Dirichlet sur $\partial\Omega_D$ et est bien un espace de Hilbert comme sous-espace fermé de $H^1(\Omega)^N$ (par application du Théorème de trace 2.3.13). On obtient alors la formulation variationnelle : trouver $u \in V$ tel que

$$\int_{\Omega} 2\mu e(u) \cdot e(v) dx + \int_{\Omega} \lambda \operatorname{div} u \operatorname{div} v dx = \int_{\Omega} f \cdot v dx + \int_{\partial\Omega_N} g \cdot v ds \quad \forall v \in V. \quad (3.64)$$

Pour pouvoir appliquer le Théorème de Lax-Milgram 1.3.1 à la formulation variationnelle (3.64), la seule hypothèse délicate à vérifier est encore une fois la coercivité de la forme bilinéaire. Autrement dit, il faut montrer qu'il existe une constante $C > 0$ telle que, pour toute fonction $v \in V$, on a

$$\|v\|_{H^1(\Omega)} \leq C \|e(v)\|_{L^2(\Omega)}. \quad (3.65)$$

Tout d'abord, on note que $\|e(v)\|_{L^2(\Omega)}$ est une norme sur V . Le seul point à vérifier qui mérite que l'on s'y attarde est que $\|e(v)\|_{L^2(\Omega)} = 0$ implique $v = 0$. Supposons donc que $\|e(v)\|_{L^2(\Omega)} = 0$: alors l'Exercice 3.3.2 montre que v est un déplacement rigide, c'est-à-dire que $v(x) = b + Mx$ avec $M = -M^t$. On vérifie facilement que, si $M \neq 0$, alors les points x , solutions de $b + Mx = 0$, forment une droite dans \mathbb{R}^3 et un

point dans \mathbb{R}^2 . Or $v(x) = 0$ sur $\partial\Omega_D$, qui est de mesure surfacique non nulle, donc nécessairement $M = 0$ et $b = 0$. Démontrons maintenant (3.65) par un argument de contradiction (voir la Remarque 2.3.18). Si (3.65) est faux, il existe une suite $v_n \in V$ telle que

$$\|v_n\|_{H^1(\Omega)} = 1 > n\|e(v_n)\|_{L^2(\Omega)}.$$

En particulier, la suite $e(v_n)$ tend vers 0 dans $L^2(\Omega)^{N^2}$. D'autre part, comme v_n est bornée dans $H^1(\Omega)^N$, par application du Théorème de Rellich 2.3.21, il existe une sous-suite $v_{n'}$ qui converge dans $L^2(\Omega)^N$. L'inégalité de Korn du Lemme 3.3.3 implique que

$$\|v_{n'} - v_{p'}\|_{H^1(\Omega)}^2 \leq C\|v_{n'} - v_{p'}\|_{L^2(\Omega)}^2 + \|e(v_{n'}) - e(v_{p'})\|_{L^2(\Omega)}^2,$$

d'où l'on déduit que la suite $v_{n'}$ est de Cauchy dans $H^1(\Omega)^N$, donc convergente vers une limite v_∞ qui vérifie $\|e(v_\infty)\|_{L^2(\Omega)} = 0$. Comme il s'agit d'une norme on en déduit que la limite est nulle $v_\infty = 0$, ce qui est une contradiction avec le fait que $\|v_{n'}\|_{H^1(\Omega)} = 1$.

L'interprétation de la formulation variationnelle (3.64) pour retrouver l'équation (3.60) est semblable à celle que nous avons fait lors de la démonstration du Théorème 3.2.14 sur le Laplacien avec condition aux limites de Neumann. Finalement, l'application $(f, g) \rightarrow u$ est linéaire. Pour montrer qu'elle est continue de $L^2(\Omega)^N \times L^2(\partial\Omega)^N$ dans $H^1(\Omega)^N$, nous prenons $v = u$ dans la formulation variationnelle (3.64). En utilisant la coercivité de la forme bilinéaire et en majorant la forme linéaire, on obtient **l'estimation d'énergie**

$$C\|u\|_{H^1(\Omega)}^2 \leq \|f\|_{L^2(\Omega)}\|u\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega_N)}\|u\|_{L^2(\partial\Omega)}. \quad (3.66)$$

Grâce à l'inégalité de Poincaré et au théorème de trace, on peut majorer le terme de droite de (3.66) par $C(\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega_N)})\|u\|_{H^1(\Omega)}$, ce qui prouve la continuité. \square

Remarque 3.3.5 Pour le système de l'élasticité avec des conditions aux limites de Dirichlet (ou de Neumann) on a les mêmes résultats de régularité qu'avec le Laplacien (voir le Théorème 3.2.26). Par contre, à la différence du Laplacien, il n'y a pas de principe du maximum pour le système de l'élasticité (comme pour la plupart des systèmes de plusieurs équations). Nous donnons un contre-exemple numérique à la Figure 3.3 : en l'absence de forces, $f = 0$, les conditions aux limites sont de Neumann sur les faces supérieure et inférieure du domaine, de Dirichlet $u = 0$ à droite et $u = e_1$ à gauche. Cela revient à étirer le domaine qui, du coup, s'amincit, conduisant ainsi à un déplacement vertical qui change de signe. \bullet

Nous avons déjà dit que la formulation variationnelle n'est rien d'autre que le **principe des travaux virtuels** en mécanique. Poursuivant cette analogie, l'espace V est l'espace des **déplacements v cinématiquement admissibles**, et l'espace des tenseurs symétriques $\sigma \in L^2(\Omega)^{N^2}$, tels que $-\operatorname{div}\sigma = f$ dans Ω et $\sigma n = g$ sur $\partial\Omega_N$, est celui des tenseurs de **contraintes statiquement admissibles**. Comme

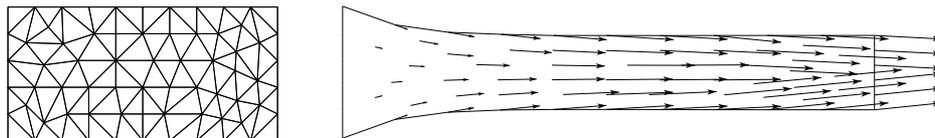


FIGURE 3.3 – Contre-exemple numérique pour le principe du maximum en élasticité. A gauche le domaine maillé au repos, et à droite le domaine déformé où les flèches représentent le déplacement (sa composante verticale change de signe).

pour le Laplacien, la solution de la formulation variationnelle (3.64) réalise le minimum d'une énergie mécanique définie pour $v \in V$ par

$$J(v) = \frac{1}{2} \int_{\Omega} (2\mu|e(v)|^2 + \lambda|\operatorname{div}v|^2) dx - \int_{\Omega} f \cdot v dx - \int_{\partial\Omega_N} g \cdot v ds. \quad (3.67)$$

En termes mécaniques $J(v)$ est la somme de **l'énergie de déformation**

$$\frac{1}{2} \int_{\Omega} (2\mu|e(v)|^2 + \lambda|\operatorname{div}v|^2) dx$$

et de **l'énergie potentielle des forces extérieures** (ou travail des forces extérieures au signe près)

$$- \int_{\Omega} f \cdot v dx - \int_{\partial\Omega_N} g \cdot v ds.$$

Exercice 3.3.3 Montrer que $u \in V$ est l'unique solution de la formulation variationnelle (3.64) si et seulement si u réalise le minimum sur V de l'énergie $J(v)$ définie par (3.67). (Indication : on pourra s'inspirer de la Proposition 1.3.4).

Exercice 3.3.4 Soit Ω un ouvert borné connexe de \mathbb{R}^N . On considère le système de l'élasticité avec la condition de Neumann (3.59) sur tout le bord $\partial\Omega$. Montrer que la condition d'équilibre

$$\int_{\Omega} f \cdot (Mx + b) dx + \int_{\partial\Omega} g \cdot (Mx + b) ds = 0 \quad \forall b \in \mathbb{R}^N, \forall M = -M^t \in \mathbb{R}^{N \times N}$$

est une condition nécessaire et suffisante d'existence et d'unicité d'une solution dans $H^1(\Omega)^N$ (l'unicité étant obtenue à l'addition d'un "mouvement de corps rigide" près, voir (3.62)).

3.3.2 Équations de Stokes

Nous appliquons l'approche variationnelle à la résolution du système d'équations de Stokes. Soit Ω un ouvert borné connexe de \mathbb{R}^N . Soit une force $f(x)$, une fonction de Ω dans \mathbb{R}^N . Il y a deux inconnues : la vitesse u qui est une fonction vectorielle, et la pression p qui est une fonction scalaire. En notation vectorielle, le problème aux limites considéré est

$$\begin{cases} \nabla p - \mu\Delta u = f & \text{dans } \Omega \\ \operatorname{div}u = 0 & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \quad (3.71)$$

où $\mu > 0$ est la viscosité du fluide. La deuxième équation de (3.71) est la contrainte d'incompressibilité du fluide, tandis que la première est le bilan des forces. La condition aux limites de Dirichlet modélise l'adhérence du fluide sur les parois.

Remarque 3.3.7 Le problème de Stokes (3.71) est un modèle **simplifié** d'écoulement d'un fluide visqueux incompressible (en régime stationnaire). En effet, les "vraies" équations du mouvement d'un tel fluide sont les équations de Navier-Stokes stationnaires (voir par exemple [26])

$$\begin{cases} (u \cdot \nabla)u + \nabla p - \mu \Delta u = f & \text{dans } \Omega \\ \operatorname{div} u = 0 & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega. \end{cases} \quad (3.72)$$

Lorsque la vitesse du fluide u est faible, le terme non-linéaire $(u \cdot \nabla)u$ étant quadratique en u devient négligeable. On obtient alors les équations de Stokes. Il faut donc avoir à l'esprit que le domaine de validité de ce modèle est limité par cette hypothèse de petite vitesse. •

Théorème 3.3.8 Soit Ω un ouvert borné connexe régulier de classe \mathcal{C}^1 de \mathbb{R}^N . Soit $f \in L^2(\Omega)^N$. Il existe une unique solution (faible) $u \in H_0^1(\Omega)^N$ et $p \in L^2(\Omega)/\mathbb{R}$ de (3.71) (la pression est unique à une constante additive près dans Ω).

Démonstration. Pour trouver la formulation variationnelle on multiplie chaque équation du système (3.71) par une fonction test v_i (qui s'annule sur le bord $\partial\Omega$ pour prendre en compte la condition aux limites de Dirichlet), on intègre par parties et on somme pour i allant de 1 à N (pour faire apparaître la divergence de la fonction $v = (v_1, \dots, v_N)$)

$$\int_{\Omega} \mu \nabla u \cdot \nabla v \, dx - \int_{\Omega} p \operatorname{div} v \, dx = \int_{\Omega} f \cdot v \, dx,$$

avec la notation $\nabla u \cdot \nabla v = \sum_{i=1}^N \nabla u_i \cdot \nabla v_i$. Afin de tenir compte de la condition d'incompressibilité $\operatorname{div} u = 0$, on choisit comme espace de Hilbert le sous-espace suivant de $H_0^1(\Omega)^N$

$$V = \{v \in H_0^1(\Omega)^N \text{ tel que } \operatorname{div} v = 0 \text{ p.p. dans } \Omega\}, \quad (3.73)$$

qui est bien un espace de Hilbert comme sous-espace fermé de $H_0^1(\Omega)^N$. On trouve alors la formulation variationnelle :

$$\text{trouver } u \in V \text{ tel que } \int_{\Omega} \mu \nabla u \cdot \nabla v \, dx = \int_{\Omega} f \cdot v \, dx \quad \forall v \in V, \quad (3.74)$$

dans laquelle la pression a disparu ! On vérifie aisément que chaque terme de (3.74) a bien un sens.

L'application du Théorème de Lax-Milgram à la formulation variationnelle (3.74) ne pose pas de problème. En particulier, la coercivité de la forme bilinéaire est évidente dans $H_0^1(\Omega)^N$ (grâce à l'inégalité de Poincaré) donc dans V qui est un sous-espace de $H_0^1(\Omega)^N$.

Le point le plus délicat est ici de montrer que la solution unique de (3.74) est bien une solution du problème aux limites (3.71). Expliquons la difficulté en supposant même momentanément que u est régulière, c'est-à-dire appartient à $H^2(\Omega)^N$. Par intégration par parties, (3.74) implique que

$$\int_{\Omega} (\mu\Delta u + f) \cdot v \, dx = 0 \quad \forall v \in V,$$

mais on ne peut pas en déduire que $(\mu\Delta u + f) = 0$ car l'orthogonal de V dans $L^2(\Omega)^N$ n'est pas réduit au seul vecteur nul! En effet, on voit facilement que si ϕ est une fonction régulière, alors, pour tout $v \in V$, on a

$$\int_{\Omega} \nabla \phi \cdot v \, dx = - \int_{\Omega} \phi \operatorname{div} v \, dx = 0,$$

c'est-à-dire que l'orthogonal de V contient au moins tous les gradients. En fait le Théorème de de Rham 3.3.9 nous dit que l'orthogonal de V coïncide exactement avec l'espace des gradients.

Grâce au Théorème de de Rham 3.3.9 nous pouvons conclure comme suit. On pose

$$L(v) = \int_{\Omega} \mu \nabla u \cdot \nabla v \, dx - \int_{\Omega} f \cdot v \, dx,$$

qui est bien une forme linéaire continue sur $H_0^1(\Omega)^N$ et nulle sur V . Par conséquent, il existe $p \in L^2(\Omega)$, unique à une constante additive près, tel que

$$L(v) = \int_{\Omega} p \operatorname{div} v \, dx \quad \forall v \in H_0^1(\Omega)^N.$$

Si on pose $\sigma = \mu \nabla u - p \operatorname{Id}$ qui appartient à $L^2(\Omega)^{N^2}$, on a donc

$$\left| \int_{\Omega} \sigma \cdot \nabla v \, dx \right| = \left| \int_{\Omega} f \cdot v \, dx \right| \leq C \|v\|_{L^2(\Omega)},$$

ce qui prouve que σ admet une divergence faible dans $L^2(\Omega)^N$, et on a $-\operatorname{div} \sigma = f$. Par conséquent, on en déduit que

$$\nabla p - \mu \Delta u = f \quad \text{presque partout dans } \Omega. \quad (3.75)$$

D'autre part, comme $u \in V$, $\operatorname{div} u$ est nul dans $L^2(\Omega)$, ce qui implique

$$\operatorname{div} u = 0 \quad \text{presque partout dans } \Omega.$$

De même, la condition aux limites s'interprète par le théorème de trace et on obtient que $u = 0$ presque partout sur $\partial\Omega$. \square

Nous énonçons maintenant le résultat très profond et difficile qui nous a permis de retrouver la pression à partir de la formulation variationnelle (3.74) du système de Stokes où elle avait disparu! Sa démonstration dépasse très largement le cadre de ce cours (signalons qu'il n'est pas facile de trouver dans la littérature une démonstration "élémentaire" et auto-contenue; voir néanmoins [18]).

Théorème 3.3.9 (de de Rham) Soit Ω un ouvert borné connexe régulier de classe \mathcal{C}^1 de \mathbb{R}^N . Soit L une forme linéaire continue sur $H_0^1(\Omega)^N$. Alors L s'annule sur V si et seulement si il existe une fonction $p \in L^2(\Omega)$ telle que

$$L(v) = \int_{\Omega} p \operatorname{div} v \, dx \quad \forall v \in H_0^1(\Omega)^N. \quad (3.76)$$

De plus p est unique à une constante additive près.

Comme pour le système de l'élasticité il existe un principe de minimisation de l'énergie (dite de **dissipation visqueuse**) pour les équations de Stokes.

Exercice 3.3.10 Soit V l'espace des champs de vitesse à divergence nulle défini par (3.73). Soit $J(v)$ l'énergie définie pour $v \in V$ par

$$J(v) = \frac{1}{2} \int_{\Omega} \mu |\nabla v|^2 \, dx - \int_{\Omega} f \cdot v \, dx. \quad (3.78)$$

Soit $u \in V$ la solution unique de la formulation variationnelle (3.74). Montrer que u est aussi l'unique point de minimum de l'énergie, c'est-à-dire que $J(u) = \min_{v \in V} J(v)$. Réciproquement, montrer que, si $u \in V$ est un point de minimum de l'énergie $J(v)$, alors u est la solution unique de la formulation variationnelle (3.74).

Chapitre 4

MÉTHODE DES ÉLÉMENTS FINIS

4.1 Approximation variationnelle

4.1.1 Introduction

Dans ce chapitre nous présentons la méthode des **éléments finis** qui est la méthode numérique de référence pour le calcul des solutions de problèmes aux limites. Le principe de cette méthode est directement issu de l'**approche variationnelle** que nous avons étudiée en détail dans les chapitres précédents.

L'idée de base de la méthode des éléments finis est de remplacer l'espace de Hilbert V sur lequel est posée la formulation variationnelle par un sous-espace V_h de dimension finie. Le problème "approché" posé sur V_h se ramène à la simple résolution d'un système linéaire, dont la matrice est appelée **matrice de rigidité**. Par ailleurs, on peut choisir le mode de construction de V_h de manière à ce que le sous-espace V_h soit une bonne approximation de V et que la solution u_h dans V_h de la formulation variationnelle soit "**proche**" de la solution exacte u dans V .

Le plan de ce chapitre est le suivant. Dans la suite de cette section nous détaillons le processus **d'approximation variationnelle interne**. La Section 4.2 présente les éléments finis en une dimension d'espace où, sans trahir les idées générales valables en dimensions supérieures, les aspects techniques sont nettement plus simples. On discute des aspects pratiques (assemblage de la matrice de rigidité, formules de quadrature, etc.) autant que théoriques (convergence de la méthode, interpolation et estimation d'erreur). La Section 4.3 est dédiée aux éléments finis en dimension supérieure ($N \geq 2$). On introduit les notions de **maillage** (triangulaire ou quadrangulaire) et de **degrés de liberté** qui permettent de construire plusieurs familles de méthodes d'éléments finis. Pour plus de détails sur la méthode des éléments finis nous renvoyons à [3], [10], [16], [18], [23], [24] (voir aussi [13], [14], [21] pour des aspects pratiques de programmation informatique).

4.1.2 Approximation interne générale

Nous considérons à nouveau le cadre général du formalisme variationnel introduit au Chapitre 1. Étant donné un espace de Hilbert V , une forme bilinéaire

continue et coercive $a(u, v)$, et une forme linéaire continue $L(v)$, on considère la formulation variationnelle :

$$\text{trouver } u \in V \text{ tel que } a(u, v) = L(v) \quad \forall v \in V, \quad (4.1)$$

dont on sait qu'elle admet une unique solution par le Théorème 1.3.1 de Lax-Milgram. **L'approximation interne** de (4.1) consiste à remplacer l'espace de Hilbert V par un sous-espace de dimension finie V_h , c'est-à-dire à chercher la solution de :

$$\text{trouver } u_h \in V_h \text{ tel que } a(u_h, v_h) = L(v_h) \quad \forall v_h \in V_h. \quad (4.2)$$

La résolution de l'approximation interne (4.2) est facile comme le montre le lemme suivant.

Lemme 4.1.1 *Soit V un espace de Hilbert réel, et V_h un sous-espace de dimension finie. Soit $a(u, v)$ une forme bilinéaire continue et coercive sur V , et $L(v)$ une forme linéaire continue sur V . Alors l'approximation interne (4.2) admet une unique solution. Par ailleurs cette solution peut s'obtenir en résolvant un système linéaire de matrice définie positive (et symétrique si $a(u, v)$ est symétrique).*

Démonstration. L'existence et l'unicité de $u_h \in V_h$, solution de (4.2), découle du Théorème 1.3.1 de Lax-Milgram appliqué à V_h . Pour mettre le problème sous une forme plus simple, on introduit une base $(\phi_j)_{1 \leq j \leq N_h}$ de V_h . Si $u_h = \sum_{j=1}^{N_h} u_j \phi_j$, on pose $U_h = (u_1, \dots, u_{N_h})$ le vecteur dans \mathbb{R}^{N_h} des coordonnées de u_h . Le problème (4.2) est équivalent à :

$$\text{trouver } U_h \in \mathbb{R}^{N_h} \text{ tel que } a \left(\sum_{j=1}^{N_h} u_j \phi_j, \phi_i \right) = L(\phi_i) \quad \forall 1 \leq i \leq N_h,$$

ce qui s'écrit sous la forme d'un système linéaire

$$\mathcal{K}_h U_h = b_h, \quad (4.3)$$

avec, pour $1 \leq i, j \leq N_h$,

$$(\mathcal{K}_h)_{ij} = a(\phi_j, \phi_i), \quad (b_h)_i = L(\phi_i).$$

La coercivité de la forme bilinéaire $a(u, v)$ entraîne le caractère défini positif de la matrice \mathcal{K}_h , et donc son inversibilité. En effet, pour tout vecteur $U_h \in \mathbb{R}^{N_h}$, on a

$$\mathcal{K}_h U_h \cdot U_h \geq \nu \left\| \sum_{j=1}^{N_h} u_j \phi_j \right\|^2 \geq C |U_h|^2 \quad \text{avec } C > 0,$$

car toutes les normes sont équivalentes en dimension finie ($|\cdot|$ désigne la norme euclidienne dans \mathbb{R}^{N_h}). De même, la symétrie de $a(u, v)$ implique celle de \mathcal{K}_h . Dans les applications mécaniques la matrice \mathcal{K}_h est appelée **matrice de rigidité**. \square

Nous allons maintenant comparer l'erreur commise en remplaçant l'espace V par son sous-espace V_h . Plus précisément, nous allons majorer la différence $\|u - u_h\|$ où u est la solution dans V de (4.1) et u_h celle dans V_h de (4.2). Précisons auparavant quelques notations : on note $\nu > 0$ la constante de coercivité et $M > 0$ la constante de continuité de la forme bilinéaire $a(u, v)$ qui vérifient

$$\begin{aligned} a(u, u) &\geq \nu \|u\|^2 \quad \forall u \in V, \\ |a(u, v)| &\leq M \|u\| \|v\| \quad \forall u, v \in V \end{aligned}$$

Le lemme suivant, dû à Jean Céa, montre que la distance entre la solution exacte u et la solution approchée u_h est majorée **uniformément par rapport au sous-espace** V_h par la distance entre u et V_h .

Lemme 4.1.2 (de Céa) *On se place sous les hypothèses du Lemme 4.1.1. Soit u la solution de (4.1) et u_h celle de (4.2). On a*

$$\|u - u_h\| \leq \frac{M}{\nu} \inf_{v_h \in V_h} \|u - v_h\|. \quad (4.4)$$

Démonstration. Puisque $V_h \subset V$, on déduit, par soustraction des formulations variationnelles (4.1) et (4.2), que

$$a(u - u_h, w_h) = 0 \quad \forall w_h \in V_h.$$

En choisissant $w_h = u_h - v_h$ on obtient

$$\nu \|u - u_h\|^2 \leq a(u - u_h, u - u_h) = a(u - u_h, u - v_h) \leq M \|u - u_h\| \|u - v_h\|,$$

d'où l'on déduit (4.4). □

Exercice 4.1.1 Dans le cadre du Lemme de Céa 4.1.2, démontrer que, si la forme bilinéaire $a(u, v)$ est symétrique, alors on améliore (4.4) en

$$\|u - u_h\| \leq \sqrt{\frac{M}{\nu}} \inf_{v_h \in V_h} \|u - v_h\|.$$

Indication : on utilisera le fait que la solution u_h de (4.2) réalise aussi le minimum d'une énergie.

Finalement, pour démontrer la convergence de cette approximation variationnelle, nous donnons un dernier lemme général. Rappelons que dans la notation V_h le paramètre $h > 0$ n'a pas encore de signification pratique. Néanmoins, nous supposons que c'est dans la limite $h \rightarrow 0$ que l'approximation interne (4.2) "converge" vers la formulation variationnelle (4.1).

Lemme 4.1.3 *On se place sous les hypothèses du Lemme 4.1.1. On suppose qu'il existe un sous-espace $\mathcal{V} \subset V$ dense dans V et une application r_h de \mathcal{V} dans V_h (appelée **opérateur d'interpolation**) tels que*

$$\lim_{h \rightarrow 0} \|v - r_h(v)\| = 0 \quad \forall v \in \mathcal{V}. \quad (4.5)$$

Alors la méthode d'approximation variationnelle interne converge, c'est-à-dire que

$$\lim_{h \rightarrow 0} \|u - u_h\| = 0. \quad (4.6)$$

Démonstration. Soit $\epsilon > 0$. Par densité de \mathcal{V} , il existe $v \in \mathcal{V}$ tel que $\|u - v\| \leq \epsilon$. Par ailleurs, il existe un $h_0 > 0$ (dépendant de ϵ) tel que, pour cet élément $v \in \mathcal{V}$, on a

$$\|v - r_h(v)\| \leq \epsilon \quad \forall h \leq h_0.$$

En vertu du Lemme 4.1.2, on a

$$\|u - u_h\| \leq C\|u - r_h(v)\| \leq C(\|u - v\| + \|v - r_h(v)\|) \leq 2C\epsilon,$$

d'où l'on déduit le résultat. \square

4.1.4 Méthode des éléments finis (principes généraux)

Le principe de la méthode des éléments finis est de construire des espaces d'approximation interne V_h dont la définition est basée sur la notion géométrique de **maillage** du domaine Ω . Un maillage est un pavage de l'espace en volumes élémentaires très simples : triangles, tétraèdres, parallélépipèdes (voir, par exemple, la Figure 4.7). Dans ce contexte le paramètre h de V_h correspond à la **taille maximale des mailles** ou cellules qui composent le maillage. Typiquement une base de V_h sera constituée de fonctions dont le support est **localisé** sur une ou quelques mailles. Ceci aura deux conséquences importantes : d'une part, dans la limite $h \rightarrow 0$, l'espace V_h sera de plus en plus "gros" et approchera de mieux en mieux l'espace V tout entier, et d'autre part, la matrice de rigidité \mathcal{K}_h du système linéaire (4.3) sera **creuse**, c'est-à-dire que la plupart de ses coefficients seront nuls (ce qui limitera le coût de la résolution numérique).

4.2 Éléments finis en dimension $N = 1$

Pour simplifier l'exposition nous commençons par présenter la méthode des éléments finis en une dimension d'espace. Sans perte de généralité nous choisissons le domaine $\Omega =]0, 1[$. En dimension 1 un maillage est simplement constitué d'une collection de points $(x_j)_{0 \leq j \leq n+1}$ (comme pour la méthode des différences finies) tels que

$$x_0 = 0 < x_1 < \dots < x_n < x_{n+1} = 1.$$

Le maillage sera dit **uniforme** si les points x_j sont équidistants, c'est-à-dire que

$$x_j = jh \quad \text{avec} \quad h = \frac{1}{n+1}, \quad 0 \leq j \leq n+1.$$

Les points x_j sont aussi appelés les **sommets** (ou noeuds) du maillage. Par souci de simplicité nous considérons, pour l'instant, le problème modèle suivant

$$\begin{cases} -u'' = f & \text{dans }]0, 1[\\ u(0) = u(1) = 0, \end{cases} \quad (4.7)$$

dont nous savons qu'il admet une solution unique dans $H_0^1(\Omega)$ si $f \in L^2(\Omega)$ (voir le Chapitre 3). Dans tout ce qui suit on notera \mathbb{P}_k l'ensemble des polynômes à coefficients réels d'une variable réelle de degré inférieur ou égal à k .

4.2.1 Éléments finis \mathbb{P}_1

La méthode des éléments finis \mathbb{P}_1 repose sur l'espace discret des fonctions globalement continues et affines sur chaque maille

$$V_h = \{v \in C([0, 1]) \text{ tel que } v|_{[x_j, x_{j+1}]} \in \mathbb{P}_1 \text{ pour tout } 0 \leq j \leq n\}, \quad (4.8)$$

et sur son sous-espace

$$V_{0h} = \{v \in V_h \text{ tel que } v(0) = v(1) = 0\}. \quad (4.9)$$

La méthode des éléments finis \mathbb{P}_1 est alors simplement la méthode d'approximation variationnelle interne de la Sous-section 4.1.2 appliquée aux espaces V_h ou V_{0h} définis par (4.8) ou (4.9).

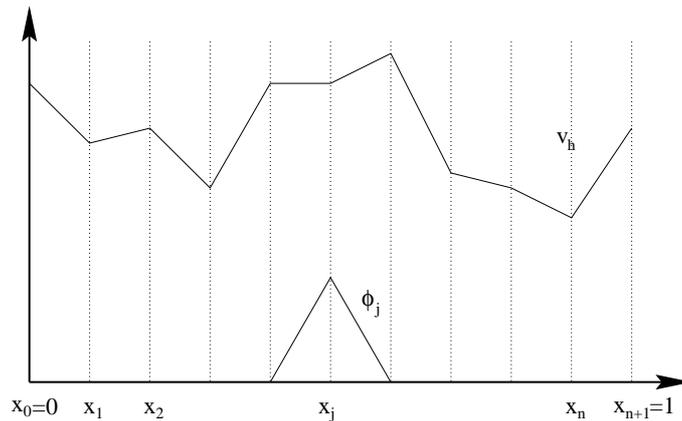


FIGURE 4.1 – Maillage de $\Omega =]0, 1[$ et fonction de base en éléments finis \mathbb{P}_1 .

On peut représenter les fonctions de V_h ou V_{0h} , affines par morceaux, à l'aide de fonctions de base très simples. Introduisons la “fonction chapeau” ϕ définie par

$$\phi(x) = \begin{cases} 1 - |x| & \text{si } |x| \leq 1, \\ 0 & \text{si } |x| > 1. \end{cases}$$

Si le maillage est uniforme, pour $0 \leq j \leq n + 1$ on définit les fonctions de base (voir la Figure 4.1)

$$\phi_j(x) = \phi\left(\frac{x - x_j}{h}\right). \quad (4.10)$$

Lemme 4.2.1 *L'espace V_h , défini par (4.8), est un sous-espace de $H^1(0, 1)$ de dimension $n + 2$, et toute fonction $v_h \in V_h$ est définie de manière unique par ses*

valeurs aux sommets $(x_j)_{0 \leq j \leq n+1}$

$$v_h(x) = \sum_{j=0}^{n+1} v_h(x_j) \phi_j(x) \quad \forall x \in [0, 1].$$

De même, V_{0h} , défini par (4.9), est un sous-espace de $H_0^1(0, 1)$ de dimension n , et toute fonction $v_h \in V_{0h}$ est définie de manière unique par ses valeurs aux sommets $(x_j)_{1 \leq j \leq n}$

$$v_h(x) = \sum_{j=1}^n v_h(x_j) \phi_j(x) \quad \forall x \in [0, 1].$$

Démonstration. Rappelons que les fonctions continues et de classe C^1 par morceaux appartiennent à $H^1(\Omega)$. Donc V_h et V_{0h} sont bien des sous-espaces de $H^1(0, 1)$. Le reste de la preuve est immédiat en remarquant que $\phi_j(x_i) = \delta_{ij}$, où δ_{ij} est le symbole de Kronecker qui vaut 1 si $i = j$ et 0 sinon (voir la Figure 4.1). \square

Remarque 4.2.2 La base (ϕ_j) , définie par (4.10), permet de caractériser une fonction de V_h par ses valeurs aux noeuds du maillage. Dans ce cas on parle **d'éléments finis de Lagrange**. Par ailleurs, comme les fonctions sont localement \mathbb{P}_1 , on dit que l'espace V_h , défini par (4.8), est l'espace des éléments finis de Lagrange d'ordre 1.

Cet exemple des éléments finis \mathbb{P}_1 permet à nouveau de comprendre l'intérêt de la formulation variationnelle. En effet, les fonctions de V_h ne sont pas deux fois dérivables sur le segment $[0, 1]$ et cela n'a pas de sens de résoudre, même de manière approchée, l'équation (4.7) (en fait la dérivée seconde d'une fonction de V_h est une somme de masses de Dirac aux noeuds du maillage!). Au contraire, il est parfaitement légitime d'utiliser des fonctions de V_h dans la formulation variationnelle (4.2) qui ne requiert qu'une seule dérivée. \bullet

Décrivons la **résolution pratique** du problème de Dirichlet (4.7) par la méthode des éléments finis \mathbb{P}_1 . La formulation variationnelle (4.2) de l'approximation interne devient ici :

$$\text{trouver } u_h \in V_{0h} \text{ tel que } \int_0^1 u_h'(x) v_h'(x) dx = \int_0^1 f(x) v_h(x) dx \quad \forall v_h \in V_{0h}. \quad (4.11)$$

On décompose u_h sur la base des $(\phi_j)_{1 \leq j \leq n}$ et on prend $v_h = \phi_i$ ce qui donne

$$\sum_{j=1}^n u_h(x_j) \int_0^1 \phi_j'(x) \phi_i'(x) dx = \int_0^1 f(x) \phi_i(x) dx.$$

En notant $U_h = (u_h(x_j))_{1 \leq j \leq n}$, $b_h = \left(\int_0^1 f(x) \phi_i(x) dx \right)_{1 \leq i \leq n}$, et en introduisant la **matrice de rigidité**

$$\mathcal{K}_h = \left(\int_0^1 \phi_j'(x) \phi_i'(x) dx \right)_{1 \leq i, j \leq n},$$

la formulation variationnelle dans V_{0h} revient à résoudre dans \mathbb{R}^n le système linéaire

$$\mathcal{K}_h U_h = b_h.$$

Comme les fonctions de base ϕ_j ont un “petit” support, l’intersection des supports de ϕ_j et ϕ_i est souvent vide et la plupart des coefficients de \mathcal{K}_h sont nuls. Un calcul simple montre que

$$\int_0^1 \phi_j'(x) \phi_i'(x) dx = \begin{cases} -h^{-1} & \text{si } j = i - 1 \\ 2h^{-1} & \text{si } j = i \\ -h^{-1} & \text{si } j = i + 1 \\ 0 & \text{sinon} \end{cases}$$

et la matrice \mathcal{K}_h est tridiagonale

$$\mathcal{K}_h = h^{-1} \begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix}. \quad (4.12)$$

Pour obtenir le second membre b_h il faut calculer les intégrales

$$(b_h)_i = \int_{x_{i-1}}^{x_{i+1}} f(x) \phi_i(x) dx \quad \text{pour tout } 1 \leq i \leq n.$$

L’évaluation exacte du second membre b_h peut être difficile ou impossible si la fonction f est compliquée. En pratique on a recours à des **formules de quadrature** (ou formules d’intégration numérique) qui donnent une approximation des intégrales définissant b_h . Par exemple, on peut utiliser la formule du “point milieu”

$$\frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} \psi(x) dx \approx \psi\left(\frac{x_{i+1} + x_i}{2}\right),$$

ou la formule des “trapèzes”

$$\frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} \psi(x) dx \approx \frac{1}{2} (\psi(x_{i+1}) + \psi(x_i)).$$

Ces deux formules sont exactes pour les fonctions ψ affines. Si la fonction ψ est régulière quelconque, alors ces formules sont simplement approchées avec un reste de l’ordre de $\mathcal{O}(h^2)$.

La résolution du système linéaire $\mathcal{K}_h U_h = b_h$ est la partie la plus coûteuse de la méthode en terme de temps de calcul. C’est pourquoi nous présentons dans la Section 4.4 des méthodes performantes de résolution. Rappelons que la matrice \mathcal{K}_h est nécessairement inversible par application du Lemme 4.1.1.

Remarque 4.2.3 La matrice de rigidité \mathcal{K}_h est très similaire à des matrices déjà rencontrées lors de l’étude des méthodes de différences finies. •

Problème de Neumann. La mise en oeuvre de la méthode des éléments finis \mathbb{P}_1 pour le problème de Neumann suivant est très similaire

$$\begin{cases} -u'' + au = f \text{ dans }]0, 1[\\ u'(0) = \alpha, u'(1) = \beta. \end{cases} \quad (4.13)$$

Rappelons que (4.13) admet une solution unique dans $H^1(\Omega)$ si $f \in L^2(\Omega)$, $\alpha, \beta \in \mathbb{R}$, et $a \in L^\infty(\Omega)$ tel que $a(x) \geq a_0 > 0$ p.p. dans Ω (voir le Chapitre 3). La formulation variationnelle (4.2) de l'approximation interne devient ici : trouver $u_h \in V_h$ tel que

$$\int_0^1 (u_h'(x)v_h'(x) + a(x)u_h(x)v_h(x)) dx = \int_0^1 f(x)v_h(x) dx - \alpha v_h(0) + \beta v_h(1),$$

pour tout $v_h \in V_h$. En décomposant u_h sur la base des $(\phi_j)_{0 \leq j \leq n+1}$, la formulation variationnelle dans V_h revient à résoudre dans \mathbb{R}^{n+2} le système linéaire

$$\mathcal{K}_h U_h = b_h,$$

avec $U_h = (u_h(x_j))_{0 \leq j \leq n+1}$, et une nouvelle matrice de rigidité

$$\mathcal{K}_h = \left(\int_0^1 (\phi_j'(x)\phi_i'(x) + a(x)\phi_j(x)\phi_i(x)) dx \right)_{0 \leq i, j \leq n+1},$$

et

$$\begin{aligned} (b_h)_i &= \int_0^1 f(x)\phi_i(x) dx \quad \text{si } 1 \leq i \leq n, \\ (b_h)_0 &= \int_0^1 f(x)\phi_0(x) dx - \alpha, \\ (b_h)_{n+1} &= \int_0^1 f(x)\phi_{n+1}(x) dx + \beta. \end{aligned}$$

Lorsque $a(x)$ n'est pas une fonction constante, il est aussi nécessaire en pratique d'utiliser des formules de quadrature pour évaluer les coefficients de la matrice \mathcal{K}_h (comme nous l'avons fait dans l'exemple précédent pour le second membre b_h).

Exercice 4.2.1 Appliquer la méthode des éléments finis \mathbb{P}_1 au problème

$$\begin{cases} -u'' = f \text{ dans }]0, 1[\\ u(0) = \alpha, u(1) = \beta, \end{cases}$$

Vérifier que les conditions aux limites de Dirichlet non-homogènes apparaissent dans le second membre du système linéaire qui en découle.

Exercice 4.2.2 On reprend le problème de Neumann (4.13) en supposant que la fonction $a(x) = 0$ dans Ω . Montrer que la matrice du système linéaire issu de la méthode des éléments finis \mathbb{P}_1 est singulière. Montrer qu'on peut néanmoins résoudre le système linéaire si les données vérifient la condition de compatibilité

$$\int_0^1 f(x) dx = \alpha - \beta.$$

Comparer ce résultat avec le Théorème 3.2.18.

Exercice 4.2.3 Appliquer la méthode des différences finies au problème de Dirichlet (4.7). Vérifier qu'avec un schéma centré d'ordre deux, on obtient un système linéaire à résoudre avec la même matrice \mathcal{K}_h (à un coefficient multiplicatif près) mais avec un second membre b_h différent. Même question pour le problème de Neumann (4.13).

4.2.2 Convergence et estimation d'erreur

Pour démontrer la convergence de la méthode des éléments finis \mathbb{P}_1 en une dimension d'espace nous suivons la démarche esquissée dans la Sous-section 4.1.2. Nous définissons tout d'abord un **opérateur d'interpolation** r_h (comme dans le Lemme 4.1.3).

Définition 4.2.4 On appelle opérateur d'interpolation \mathbb{P}_1 l'application linéaire r_h de $H^1(0, 1)$ dans V_h définie, pour tout $v \in H^1(0, 1)$, par

$$(r_h v)(x) = \sum_{j=0}^{n+1} v(x_j) \phi_j(x).$$

Cette définition a bien un sens car, en vertu du Lemme 2.3.3, les fonctions de $H^1(0, 1)$ sont continues et leurs valeurs ponctuelles sont donc bien définies. L'interpolée $r_h v$ d'une fonction v est simplement la fonction affine par morceaux qui coïncide avec v sur les sommets du maillage x_j (voir la Figure 4.2). Remarquons qu'en une dimension d'espace l'interpolée est définie pour toute fonction de $H^1(0, 1)$, et non pas seulement pour les fonctions régulières de $H^1(0, 1)$ (ce qui sera le cas en dimension supérieure).

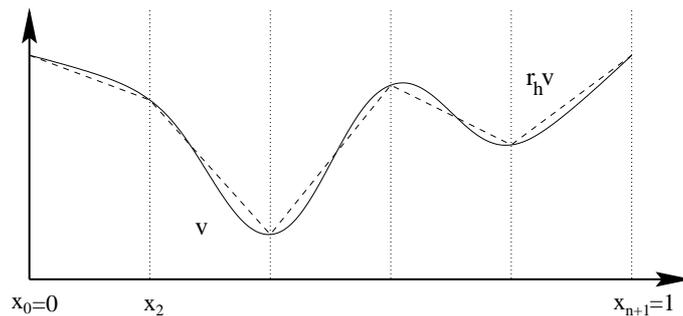


FIGURE 4.2 – Interpolation \mathbb{P}_1 d'une fonction de $H^1(0, 1)$.

La convergence de la méthode des éléments finis \mathbb{P}_1 repose sur le lemme suivant.

Lemme 4.2.5 (d'interpolation) Soit r_h l'opérateur d'interpolation \mathbb{P}_1 . Pour tout $v \in H^1(0, 1)$, il vérifie

$$\lim_{h \rightarrow 0} \|v - r_h v\|_{H^1(0,1)} = 0.$$

De plus, si $v \in H^2(0, 1)$, alors il existe une constante C indépendante de h telle que

$$\|v - r_h v\|_{H^1(0,1)} \leq Ch \|v''\|_{L^2(0,1)}.$$

Nous repoussons momentanément la démonstration de ce lemme pour énoncer tout de suite le résultat principal de cette sous-section qui établit la convergence de la méthode des éléments finis \mathbb{P}_1 pour le problème de Dirichlet.

Théorème 4.2.6 Soit $u \in H_0^1(0,1)$ et $u_h \in V_{0h}$ les solutions de (4.7) et (4.11), respectivement. Alors, la méthode des éléments finis \mathbb{P}_1 converge, c'est-à-dire que

$$\lim_{h \rightarrow 0} \|u - u_h\|_{H^1(0,1)} = 0. \quad (4.14)$$

De plus, si $u \in H^2(0,1)$ (ce qui est vrai si $f \in L^2(0,1)$), alors il existe une constante C indépendante de h telle que

$$\|u - u_h\|_{H^1(0,1)} \leq Ch \|u''\|_{L^2(0,1)} = Ch \|f\|_{L^2(0,1)}. \quad (4.15)$$

Remarque 4.2.7 On peut faire une analogie entre la convergence d'une méthode d'éléments finis et la convergence d'une méthode de différences finies. Rappelons que, d'après le Théorème de Lax, la convergence d'un schéma aux différences finies découle de sa stabilité et de sa consistance. Indiquons quels sont les équivalents (formels) de ces ingrédients dans le contexte des éléments finis. Le rôle de la consistance pour les éléments finis est joué par la propriété d'interpolation du Lemme 4.2.5, tandis que le rôle de la stabilité est tenu par la propriété de coercivité de la forme bilinéaire qui assure la résolution (stable) de toute approximation interne. •

Démonstration. Le Lemme 4.2.5 permet d'appliquer le résultat de convergence du Lemme 4.1.3 qui entraîne immédiatement (4.14). Pour obtenir (4.15), on majore l'estimation du Lemme 4.1.2 de Céa

$$\|u - u_h\|_{H^1(0,1)} \leq C \inf_{v_h \in V_h} \|u - v_h\|_{H^1(0,1)} \leq C \|u - r_h u\|_{H^1(0,1)},$$

ce qui permet de conclure grâce au Lemme 4.2.5. \square

Nous donnons maintenant la démonstration du Lemme 4.2.5 sous la forme de deux autres lemmes techniques.

Lemme 4.2.10 Il existe une constante C indépendante de h telle que, pour tout $v \in H^2(0,1)$,

$$\|v - r_h v\|_{L^2(0,1)} \leq Ch^2 \|v''\|_{L^2(0,1)}, \quad (4.16)$$

et

$$\|v' - (r_h v)'\|_{L^2(0,1)} \leq Ch \|v''\|_{L^2(0,1)}. \quad (4.17)$$

Démonstration. Soit $v \in C^\infty([0,1])$. Par définition, l'interpolée $r_h v$ est une fonction affine et, pour tout $x \in]x_j, x_{j+1}[$, on a

$$\begin{aligned} v(x) - r_h v(x) &= v(x) - \left(v(x_j) + \frac{v(x_{j+1}) - v(x_j)}{x_{j+1} - x_j} (x - x_j) \right) \\ &= \int_{x_j}^x v'(t) dt - \frac{x - x_j}{x_{j+1} - x_j} \int_{x_j}^{x_{j+1}} v'(t) dt \\ &= (x - x_j) v'(x_j + \theta_x) - (x - x_j) v'(x_j + \theta_j) \\ &= (x - x_j) \int_{x_j + \theta_j}^{x_j + \theta_x} v''(t) dt, \end{aligned} \quad (4.18)$$

par application de la formule des accroissements finis avec $0 \leq \theta_x \leq x - x_j$ et $0 \leq \theta_j \leq h$. On en déduit en utilisant l'inégalité de Cauchy-Schwarz

$$|v(x) - r_h v(x)|^2 \leq h^2 \left(\int_{x_j}^{x_{j+1}} |v''(t)| dt \right)^2 \leq h^3 \int_{x_j}^{x_{j+1}} |v''(t)|^2 dt. \quad (4.19)$$

En intégrant (4.19) par rapport à x sur l'intervalle $[x_j, x_{j+1}]$, on obtient

$$\int_{x_j}^{x_{j+1}} |v(x) - r_h v(x)|^2 dx \leq h^4 \int_{x_j}^{x_{j+1}} |v''(t)|^2 dt,$$

ce qui, par sommation en j , donne exactement (4.16). Par densité ce résultat est encore vrai pour tout $v \in H^2(0, 1)$. La démonstration de (4.17) est tout à fait similaire : pour $v \in C^\infty([0, 1])$ et $x \in]x_j, x_{j+1}[$ on écrit

$$\begin{aligned} v'(x) - (r_h v)'(x) &= v'(x) - \frac{v(x_{j+1}) - v(x_j)}{h} = \frac{1}{h} \int_{x_j}^{x_{j+1}} (v'(x) - v'(t)) dt \\ &= \frac{1}{h} \int_{x_j}^{x_{j+1}} \int_t^x v''(y) dy. \end{aligned}$$

Élevant au carré cette inégalité, appliquant Cauchy-Schwarz deux fois et sommant en j on obtient (4.17), qui est aussi valide pour tout $v \in H^2(0, 1)$ par densité. \square

Lemme 4.2.11 *Il existe une constante C indépendante de h telle que, pour tout $v \in H^1(0, 1)$,*

$$\|r_h v\|_{H^1(0,1)} \leq C \|v\|_{H^1(0,1)}, \quad (4.20)$$

et

$$\|v - r_h v\|_{L^2(0,1)} \leq Ch \|v'\|_{L^2(0,1)}. \quad (4.21)$$

De plus, pour tout $v \in H^1(0, 1)$, on a

$$\lim_{h \rightarrow 0} \|v' - (r_h v)'\|_{L^2(0,1)} = 0. \quad (4.22)$$

Démonstration. Les preuves de (4.20) et (4.21) sont dans le même esprit que celles du lemme précédent. Soit $v \in H^1(0, 1)$. Tout d'abord on a

$$\|r_h v\|_{L^2(0,1)} \leq \max_{x \in [0,1]} |r_h v(x)| \leq \max_{x \in [0,1]} |v(x)| \leq C \|v\|_{H^1(0,1)},$$

en vertu du Lemme 2.3.3. D'autre part, comme $r_h v$ est affine, et grâce à la propriété (2.8) du Lemme 2.3.3 qui affirme que v est bien la primitive de v' , on a

$$\begin{aligned} \int_{x_j}^{x_{j+1}} |(r_h v)'(x)|^2 dx &= \frac{(v(x_{j+1}) - v(x_j))^2}{h} \\ &= \frac{1}{h} \left(\int_{x_j}^{x_{j+1}} v'(x) dx \right)^2 \leq \int_{x_j}^{x_{j+1}} |v'(x)|^2 dx, \end{aligned}$$

par Cauchy-Schwarz, ce qui, par sommation en j , conduit à (4.20). Pour obtenir (4.21) on reprend la deuxième égalité de (4.18) d'où l'on déduit

$$|v(x) - r_h v(x)| \leq 2 \int_{x_j}^{x_{j+1}} |v'(t)| dt.$$

En élevant au carré, en utilisant Cauchy-Schwarz, en intégrant par rapport à x , puis en sommant en j , on obtient bien (4.21).

Passons à la démonstration de (4.22). Soit $\epsilon > 0$. Comme $C^\infty([0, 1])$ est dense dans $H^1(0, 1)$, pour tout $v \in H^1(0, 1)$ il existe $\phi \in C^\infty([0, 1])$ tel que

$$\|v' - \phi'\|_{L^2(0,1)} \leq \epsilon.$$

Or r_h est une application linéaire qui vérifie (4.20), donc on en déduit

$$\|(r_h v)' - (r_h \phi)'\|_{L^2(0,1)} \leq C \|v' - \phi'\|_{L^2(0,1)} \leq C\epsilon.$$

Le choix de ϕ et de ϵ étant fixé, on déduit de (4.17) appliqué à ϕ que, pour h suffisamment petit,

$$\|\phi' - (r_h \phi)'\|_{L^2(0,1)} \leq \epsilon.$$

Par conséquent, en sommant ces trois dernières inégalités on obtient

$$\|v' - (r_h v)'\|_{L^2(0,1)} \leq \|v' - \phi'\|_{L^2} + \|\phi' - (r_h \phi)'\|_{L^2} + \|(r_h v)' - (r_h \phi)'\|_{L^2} \leq C\epsilon,$$

ce qui implique (4.22). \square

4.2.3 Éléments finis \mathbb{P}_2

La méthode des éléments finis \mathbb{P}_2 repose sur l'espace discret

$$V_h = \{v \in C([0, 1]) \text{ tel que } v|_{[x_j, x_{j+1}]} \in \mathbb{P}_2 \text{ pour tout } 0 \leq j \leq n\}, \quad (4.23)$$

et sur son sous-espace

$$V_{0h} = \{v \in V_h \text{ tel que } v(0) = v(1) = 0\}. \quad (4.24)$$

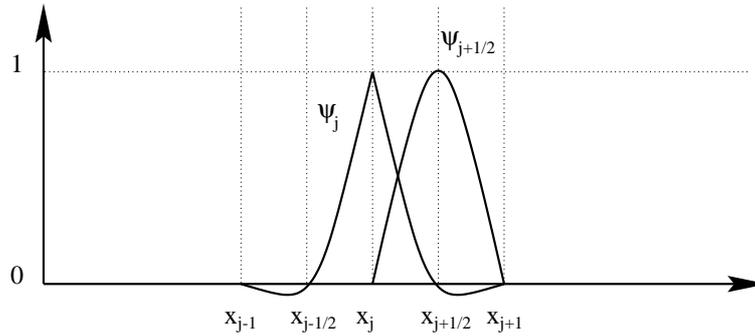
La méthode des éléments finis \mathbb{P}_2 est la méthode d'approximation variationnelle interne de la Sous-section 4.1.2 appliquée à ces espaces V_h ou V_{0h} . Ceux-ci sont composés de fonctions continues, paraboliques par morceaux qu'on peut représenter à l'aide de fonctions de base très simples.

Introduisons tout d'abord les points milieux des segments $[x_j, x_{j+1}]$ définis par $x_{j+1/2} = x_j + h/2$ pour $0 \leq j \leq n$. On définit aussi deux fonctions "mères"

$$\phi(x) = \begin{cases} (1+x)(1+2x) & \text{si } -1 \leq x \leq 0, \\ (1-x)(1-2x) & \text{si } 0 \leq x \leq 1, \\ 0 & \text{si } |x| > 1, \end{cases}$$

et

$$\psi(x) = \begin{cases} 1 - 4x^2 & \text{si } |x| \leq 1/2, \\ 0 & \text{si } |x| > 1/2. \end{cases}$$

FIGURE 4.3 – Les fonctions de base des éléments finis \mathbb{P}_2 .

Si le maillage est uniforme, pour $0 \leq j \leq n+1$ on définit les fonctions de base (voir la Figure 4.3)

$$\psi_j(x) = \phi\left(\frac{x-x_j}{h}\right), \quad 0 \leq j \leq n+1, \quad \text{et} \quad \psi_{j+1/2}(x) = \psi\left(\frac{x-x_{j+1/2}}{h}\right), \quad 0 \leq j \leq n.$$

Lemme 4.2.12 *L'espace V_h , défini par (4.23), est un sous-espace de $H^1(0,1)$ de dimension $2n+3$, et toute fonction $v_h \in V_h$ est définie de manière unique par ses valeurs aux sommets $(x_j)_{0 \leq j \leq n+1}$ et aux milieux $(x_{j+1/2})_{0 \leq j \leq n}$*

$$v_h(x) = \sum_{j=0}^{n+1} v_h(x_j) \psi_j(x) + \sum_{j=0}^n v_h(x_{j+1/2}) \psi_{j+1/2}(x) \quad \forall x \in [0, 1].$$

De même, V_{0h} , défini par (4.24), est un sous-espace de $H_0^1(0,1)$ de dimension $2n+1$, et toute fonction $v_h \in V_{0h}$ est définie de manière unique par ses valeurs aux sommets $(x_j)_{1 \leq j \leq n}$ et aux milieux $(x_{j+1/2})_{0 \leq j \leq n}$

$$v_h(x) = \sum_{j=1}^n v_h(x_j) \psi_j(x) + \sum_{j=0}^n v_h(x_{j+1/2}) \psi_{j+1/2}(x) \quad \forall x \in [0, 1].$$

Remarque 4.2.13 Ici encore, V_h est un espace d'éléments finis de Lagrange (cf. la Remarque 4.2.2). Comme les fonctions sont localement \mathbb{P}_2 , on dit que l'espace V_h , défini par (4.23), est l'espace des éléments finis de Lagrange d'ordre 2. •

Démonstration. Clairement V_h et V_{0h} sont bien des sous-espaces de $H^1(0,1)$. Leur dimension et les bases proposées se trouvent facilement en remarquant que $\psi_j(x_i) = \delta_{ij}$, $\psi_{j+1/2}(x_{i+1/2}) = \delta_{ij}$, $\psi_j(x_{i+1/2}) = 0$, $\psi_{j+1/2}(x_i) = 0$ (voir la Figure 4.3). □

Décrivons la **résolution pratique** du problème de Dirichlet (4.7) par la méthode des éléments finis \mathbb{P}_2 . La formulation variationnelle (4.2) de l'approximation interne revient à résoudre dans \mathbb{R}^{2n+1} un système linéaire $\mathcal{K}_h U_h = b_h$. On note $(x_{k/2})_{1 \leq k \leq 2n+1}$ les points du maillage et $(\psi_{k/2})_{1 \leq k \leq 2n+1}$ les fonctions de base correspondantes dans V_{0h} . Dans cette base $U_h \in \mathbb{R}^{2n+1}$ est le vecteur des coordonnées de

Démonstration. Soit u_h la solution de (4.11). En vertu du Lemme 4.2.1, on a

$$u_h(x) = \sum_{j=1}^n u_h(x_j) \phi_j(x),$$

où les fonctions ϕ_j sont les fonctions de base des éléments finis \mathbb{P}_1 dans V_{0h} et $U_h = (u_h(x_j))_{1 \leq j \leq n}$ est solution du système linéaire

$$\mathcal{K}_h U_h = b_h. \quad (4.27)$$

Les fonctions ϕ_j sont des fonctions “chapeaux” (voir la Figure 4.1) qui sont positives : il suffit donc de montrer que toutes les composantes du vecteur $U_h = (U_h^j)_{1 \leq j \leq n}$ sont positives pour prouver que la fonction u_h est positive sur $[0, 1]$. Rappelons que, en posant $U_h^0 = U_h^{n+1} = 0$, le système linéaire (4.27) est équivalent à

$$-U_h^{j-1} + 2U_h^j - U_h^{j+1} = hb_h^j \text{ pour tout } 1 \leq j \leq n. \quad (4.28)$$

Soit $U_h^{j_0} = \min_j U_h^j$ la plus petite composante de U_h : s’il y a plusieurs plus petites composantes, on choisit celle de plus petit indice j_0 . Si $j_0 = 0$, alors $U_h^{j_0} \geq U_h^0 = 0$ pour tout j , ce qui est le résultat recherché. Si $j_0 \geq 1$, alors $U_h^{j_0} < U_h^0 = 0$, et comme $U_h^{n+1} = 0$ on en déduit que $j_0 \leq n$. Comme $b_h^j = \int_0^1 f \psi_j dx \geq 0$ par hypothèse sur f , on peut alors déduire de la relation (4.28) pour j_0 que

$$(U_h^{j_0} - U_h^{j_0-1}) + (U_h^{j_0} - U_h^{j_0+1}) \geq 0,$$

ce qui est une contradiction avec le caractère minimal (strict) de $U_h^{j_0}$. Par conséquent la méthode des éléments finis \mathbb{P}_1 vérifie le principe du maximum discret. \square

4.3 Éléments finis en dimension $N \geq 2$

Nous nous plaçons maintenant en dimension d’espace $N \geq 2$ (en pratique $N = 2, 3$). Pour simplifier l’exposé, certains résultats ne seront démontrés qu’en dimension $N = 2$, mais ils s’étendent à la dimension $N = 3$ (au prix, parfois, de complications techniques et pratiques importantes).

Nous considérons le problème modèle de Dirichlet

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega, \end{cases} \quad (4.36)$$

dont nous savons qu’il admet une solution unique dans $H_0^1(\Omega)$, si $f \in L^2(\Omega)$ (voir le Chapitre 3).

Dans tout ce qui suit nous supposons que le domaine Ω est **polyédrique** (polygonal si $N = 2$), c’est-à-dire que $\overline{\Omega}$ est une réunion finie de polyèdres de \mathbb{R}^N . Rappelons qu’un polyèdre est une intersection finie de demi-espaces de \mathbb{R}^N et que les parties de son bord qui appartiennent à un seul hyperplan sont appelées ses faces. La raison de cette hypothèse est qu’il n’est possible de mailler exactement que de tels ouverts. Nous dirons plus loin ce qui se passe pour des domaines généraux à bords “courbes”.

4.3.1 Éléments finis triangulaires

Tout commence par la définition d'un maillage du domaine Ω par des triangles en dimension $N = 2$ et des tétraèdres en dimension $N = 3$. On regroupe les triangles et les tétraèdres dans la famille plus générale des N -simplexes. On appelle N -simplexe K de \mathbb{R}^N l'enveloppe convexe de $(N + 1)$ points $(a_j)_{1 \leq j \leq N+1}$ de \mathbb{R}^N , appelés sommets de K . Bien sûr un 2-simplexe est simplement un triangle et un 3-simplexe un tétraèdre (voir la Figure 4.9). On dit que le N -simplexe K est non dégénéré si les points $(a_j)_{1 \leq j \leq N+1}$ n'appartiennent pas à un même hyperplan de \mathbb{R}^N (le triangle ou le tétraèdre est non "plat"). Si on note $(a_{i,j})_{1 \leq i \leq N}$ les coordonnées du vecteur a_j , la condition de non dégénérescence de K est que la matrice

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N+1} \\ a_{2,1} & a_{2,2} & \dots & a_{2,N+1} \\ \vdots & \vdots & & \vdots \\ a_{N,1} & a_{N,2} & \dots & a_{N,N+1} \\ 1 & 1 & \dots & 1 \end{pmatrix} \quad (4.37)$$

soit inversible (ce que l'on supposera toujours par la suite). Un N -simplexe a autant de faces que de sommets, qui sont elles-mêmes des $(N - 1)$ -simplexes.

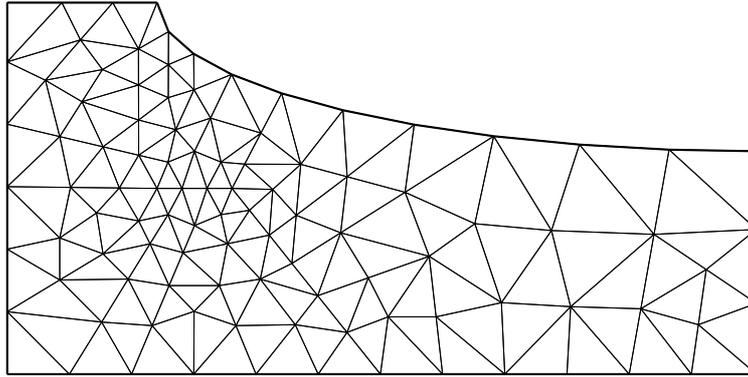


FIGURE 4.7 – Exemple de maillage triangulaire en dimension $N = 2$.

Définition 4.3.1 Soit Ω un ouvert connexe polyédrique de \mathbb{R}^N . Un **maillage triangulaire** ou une **triangulation** de $\overline{\Omega}$ est un ensemble \mathcal{T}_h de N -simplexes (non dégénérés) $(K_i)_{1 \leq i \leq n}$ qui vérifient

1. $K_i \subset \overline{\Omega}$ et $\overline{\Omega} = \cup_{i=1}^n K_i$,
2. l'intersection $K_i \cap K_j$ de deux N -simplexes distincts est un m -simplexe, avec $0 \leq m \leq N - 1$, dont tous les sommets sont aussi des sommets de K_i et K_j . (En dimension $N = 2$, l'intersection de deux triangles est soit vide, soit réduite à un sommet commun, soit une arête commune **entière**; en dimension $N = 3$, l'intersection de deux tétraèdres est soit vide, soit un sommet commun, soit une arête commune **entière**, soit une face commune **entière**.)

Les **sommets** ou **noeuds** du maillage \mathcal{T}_h sont les sommets des N -simplexes K_i qui le composent. Par convention, le paramètre h désigne le maximum des diamètres des N -simplexes K_i .

Il est clair que la Définition 4.3.1 ne peut s'appliquer qu'à un ouvert polyédrique et pas à un ouvert quelconque. La Définition 4.3.1 contient un certain nombre de restrictions sur le maillage : dans ce cas on parle souvent de **maillage conforme**. Un exemple de maillage conforme est donné à la Figure 4.7, tandis que la Figure 4.8 présente des situations interdites par la Définition 4.3.1.

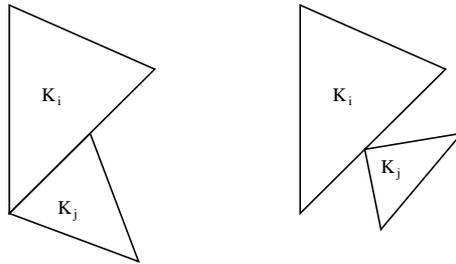


FIGURE 4.8 – Exemples de situations interdites pour un maillage triangulaire.

Remarque 4.3.2 Nous ne disons rien ici des algorithmes qui permettent de construire un maillage triangulaire. Contentons nous de dire que, s'il est relativement facile de mailler des domaines plans (il existe de nombreux logiciels libres qui permettent de le faire), il est encore assez compliqué de mailler des domaines tridimensionnels. Nous renvoyons à l'ouvrage [17] le lecteur intéressé par ce sujet. •

Dans un N -simplexe K il est commode d'utiliser des coordonnées barycentriques au lieu des coordonnées cartésiennes usuelles. Rappelons que, si K est un N -simplexe non dégénéré de sommets $(a_j)_{1 \leq j \leq N+1}$, les **coordonnées barycentriques** $(\lambda_j)_{1 \leq j \leq N+1}$ de $x \in \mathbb{R}^N$ sont définies par

$$\sum_{j=1}^{N+1} \lambda_j = 1, \quad \sum_{j=1}^{N+1} a_{i,j} \lambda_j = x_i \quad \text{pour } 1 \leq i \leq N, \quad (4.38)$$

qui admet bien une unique solution car la matrice A , définie par (4.37), est inversible. Remarquons que les λ_j sont des fonctions affines de x . On vérifie alors que

$$K = \left\{ x \in \mathbb{R}^N \text{ tel que } \lambda_j(x) \geq 0 \text{ pour } 1 \leq j \leq N+1 \right\},$$

et que les $(N+1)$ faces de K sont les intersections de K et des hyperplans $\lambda_j(x) = 0$, $1 \leq j \leq N+1$. On peut alors définir un ensemble de points de K qui vont jouer un rôle particulier pour la suite : pour tout entier $k \geq 1$ on appelle **treillis d'ordre k** l'ensemble

$$\Sigma_k = \left\{ x \in K \text{ tel que } \lambda_j(x) \in \left\{ 0, \frac{1}{k}, \dots, \frac{k-1}{k}, 1 \right\} \text{ pour } 1 \leq j \leq N \right\}. \quad (4.39)$$

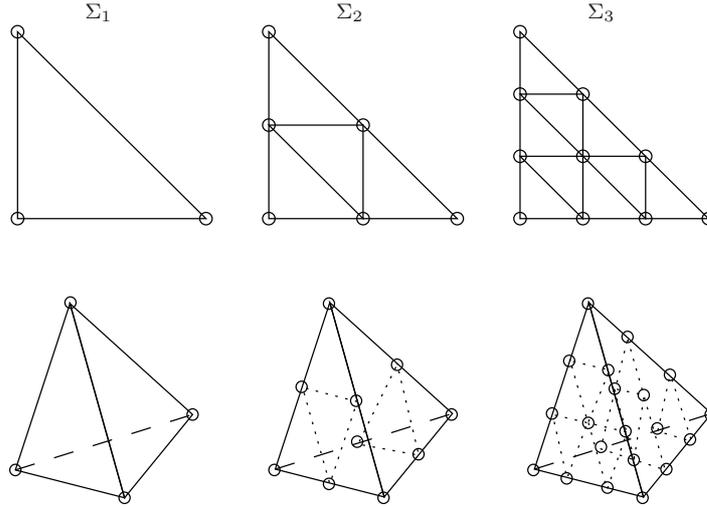


FIGURE 4.9 – Treillis d'ordre 1, 2, et 3 pour un triangle (en haut) et un tétraèdre (en bas). Les ronds représentent les points du treillis.

Pour $k = 1$ il s'agit de l'ensemble des sommets de K , et pour $k = 2$ des sommets et des points milieux des arêtes reliant deux sommets (voir la Figure 4.9). Dans le cas général, Σ_k est un ensemble fini de points $(\sigma_j)_{1 \leq j \leq n_k}$.

Nous définissons maintenant l'ensemble \mathbb{P}_k des polynômes à coefficients réels de \mathbb{R}^N dans \mathbb{R} de degré inférieur ou égal à k , c'est-à-dire que tout $p \in \mathbb{P}_k$ s'écrit sous la forme

$$p(x) = \sum_{\substack{i_1, \dots, i_N \geq 0 \\ i_1 + \dots + i_N \leq k}} \alpha_{i_1, \dots, i_N} x_1^{i_1} \cdots x_N^{i_N} \text{ avec } x = (x_1, \dots, x_N).$$

L'intérêt de la notion de treillis Σ_k d'un N -simplexe K est qu'il permet de caractériser tous les polynômes de \mathbb{P}_k (on dit que Σ_k est **unisolvant** pour \mathbb{P}_k).

Lemme 4.3.3 *Soit K un N -simplexe. Pour un entier $k \geq 1$, soit Σ_k le treillis d'ordre k , défini par (4.39), dont les points sont notés $(\sigma_j)_{1 \leq j \leq n_k}$. Alors, tout polynôme de \mathbb{P}_k est déterminé de manière unique par ses valeurs aux points $(\sigma_j)_{1 \leq j \leq n_k}$. Autrement dit, il existe une base $(\psi_j)_{1 \leq j \leq n_k}$ de \mathbb{P}_k telle que*

$$\psi_j(\sigma_i) = \delta_{ij} \quad 1 \leq i, j \leq n_k.$$

Démonstration. Le cardinal de Σ_k et la dimension de \mathbb{P}_k coïncident

$$\text{card}(\Sigma_k) = \dim(\mathbb{P}_k) = \frac{(N+k)!}{N! k!}$$

(le vérifier en guise d'exercice, au moins pour $k = 1, 2$). Comme l'application qui, à tout polynôme de \mathbb{P}_k , fait correspondre ses valeurs sur le treillis Σ_k est linéaire, il suffit de montrer qu'elle est injective pour montrer qu'elle est bijective. Soit donc un polynôme $p \in \mathbb{P}_k$ qui s'annule sur Σ_k . Montrons, par récurrence sur la dimension N , que p est identiquement nul sur \mathbb{R}^N . Pour $N = 1$, il est clair qu'un polynôme de degré k qui s'annule en $(k+1)$ points distincts est nul. Supposons le résultat

vrai à l'ordre $N - 1$. Comme x dépend linéairement des coordonnées barycentriques $(\lambda_j(x))_{1 \leq j \leq N+1}$, on peut définir un polynôme $q(\lambda) = p(x)$ de degré au plus k en la variable $\lambda \in \mathbb{R}^{N+1}$. Si l'on fixe une coordonnée λ_j dans l'ensemble $\{0, 1/k, \dots, (k-1)/k, 1\}$ et que l'on pose $\lambda = (\lambda', \lambda_j)$, on obtient un polynôme $q_j(\lambda') = q(\lambda)$ qui dépend de $N - 1$ variables indépendantes (car on a la relation $\sum_{j=1}^{N+1} \lambda_j = 1$) et qui est nul sur la section du treillis Σ_k correspondant à la valeur fixée de λ_j . Comme cette section est aussi le treillis d'ordre k d'un $(N - 1)$ -simplexe dans l'hyperplan λ_j fixé, on peut appliquer l'hypothèse de récurrence et en déduire que $q_j = 0$. Autrement dit, le facteur $\lambda_j(\lambda_j - 1/k) \cdots (\lambda_j - (k-1)/k)(\lambda_j - 1)$ divise q , ce qui est une contradiction avec le fait que le degré de $q(\lambda)$ est inférieur ou égal à k , sauf si $q = 0$, ce qui est le résultat désiré. \square

Lemme 4.3.4 *Soit K et K' deux N -simplexes ayant une face commune $\Gamma = \partial K \cap \partial K'$. Soit un entier $k \geq 1$. Alors, leurs treillis d'ordre k , Σ_k et Σ'_k coïncident sur cette face Γ . De plus, étant donné p_K et $p_{K'}$ deux polynômes de \mathbb{P}_k , la fonction v définie par*

$$v(x) = \begin{cases} p_K(x) & \text{si } x \in K \\ p_{K'}(x) & \text{si } x \in K' \end{cases}$$

est continue sur $K \cup K'$, si et seulement si p_K et $p_{K'}$ ont des valeurs qui coïncident aux points du treillis sur la face commune Γ .

Démonstration. Il est clair que la restriction à une face de K de son treillis d'ordre Σ_k est aussi un treillis d'ordre k dans l'hyperplan contenant cette face, qui ne dépend que des sommets de cette face. Par conséquent, les treillis Σ_k et Σ'_k coïncident sur leur face commune Γ . Si les polynômes p_K et $p_{K'}$ coïncident aux points de $\Sigma_k \cap \Gamma$, alors par application du Lemme 4.3.3 ils sont égaux sur Γ , ce qui prouve la continuité de v . \square

En pratique, on utilise surtout des polynômes de degré 1 ou 2. Dans ce cas on a les caractérisations suivantes de \mathbb{P}_1 et \mathbb{P}_2 dans un N -simplexe K .

Exercice 4.3.2 Soit K un N -simplexe de sommets $(a_j)_{1 \leq j \leq N+1}$. Montrer que tout polynôme $p \in \mathbb{P}_1$ se met sous la forme

$$p(x) = \sum_{j=1}^{N+1} p(a_j) \lambda_j(x),$$

où les $(\lambda_j(x))_{1 \leq j \leq N+1}$ sont les coordonnées barycentriques de $x \in \mathbb{R}^N$.

Exercice 4.3.3 Soit K un N -simplexe de sommets $(a_j)_{1 \leq j \leq N+1}$. On définit les points milieux $(a_{jj'})_{1 \leq j < j' \leq N+1}$ des arêtes de K par leur coordonnées barycentriques

$$\lambda_j(a_{jj'}) = \lambda_{j'}(a_{jj'}) = \frac{1}{2}, \quad \lambda_l(a_{jj'}) = 0 \text{ pour } l \neq j, j'.$$

Vérifier que Σ_2 est précisément constitué des sommets et des points milieux des arêtes et que tout polynôme $p \in \mathbb{P}_2$ se met sous la forme

$$p(x) = \sum_{j=1}^{N+1} p(a_j) \lambda_j(x) (2\lambda_j(x) - 1) + \sum_{1 \leq j < j' \leq N+1} 4p(a_{jj'}) \lambda_j(x) \lambda_{j'}(x),$$

où les $(\lambda_j(x))_{1 \leq j \leq N+1}$ sont les coordonnées barycentriques de $x \in \mathbb{R}^N$.

Nous avons maintenant tous les outils pour définir la méthode des éléments finis \mathbb{P}_k .

Définition 4.3.5 *Étant donné un maillage \mathcal{T}_h d'un ouvert connexe polyédrique Ω , la méthode des éléments finis \mathbb{P}_k , ou **éléments finis triangulaires de Lagrange d'ordre k** , associée à ce maillage, est définie par l'espace discret*

$$V_h = \{v \in C(\overline{\Omega}) \text{ tel que } v|_{K_i} \in \mathbb{P}_k \text{ pour tout } K_i \in \mathcal{T}_h\}. \quad (4.40)$$

On appelle **noeuds des degrés de liberté** l'ensemble des points $(\hat{a}_i)_{1 \leq i \leq n_{dl}}$ des treillis d'ordre k de chacun des N -simplexes $K_i \in \mathcal{T}_h$. On ne compte qu'une seule fois les points qui coïncident et n_{dl} est le nombre de degrés de liberté de la méthode des éléments finis \mathbb{P}_k . On appelle **degrés de liberté** d'une fonction $v \in V_h$ l'ensemble des valeurs de v en ces noeuds $(\hat{a}_i)_{1 \leq i \leq n_{dl}}$. On définit aussi le sous-espace V_{0h} par

$$V_{0h} = \{v \in V_h \text{ tel que } v = 0 \text{ sur } \partial\Omega\}. \quad (4.41)$$

Lorsque $k = 1$ les noeuds des degrés de liberté coïncident avec les sommets du maillage. Lorsque $k = 2$ ces noeuds sont constitués d'une part des sommets du maillage et d'autre part des points milieux des arêtes reliant deux sommets.

Remarque 4.3.6 L'appellation "éléments finis de Lagrange" correspond aux éléments finis dont les degrés de liberté sont des valeurs ponctuelles des fonctions de l'espace V_h . On peut définir d'autres types d'éléments finis, par exemple les éléments finis de Hermite pour lesquels les degrés de liberté sont les valeurs ponctuelles de la fonction et de ses dérivées. •

Proposition 4.3.7 *L'espace V_h , défini par (4.40), est un sous-espace de $H^1(\Omega)$ dont la dimension est finie, égale au nombre de degrés de liberté. De plus, il existe une base de V_h $(\phi_i)_{1 \leq i \leq n_{dl}}$ définie par*

$$\phi_i(\hat{a}_j) = \delta_{ij} \quad 1 \leq i, j \leq n_{dl},$$

telle que

$$v(x) = \sum_{i=1}^{n_{dl}} v(\hat{a}_i) \phi_i(x).$$

Démonstration. Les éléments de V_h , étant réguliers sur chaque maille K_i et continus sur $\overline{\Omega}$, appartiennent à $H^1(\Omega)$. Grâce au Lemme 4.3.4 les éléments de V_h sont exactement obtenus en assemblant sur chaque $K_i \in \mathcal{T}_h$ des polynômes de \mathbb{P}_k qui

coïncident sur les degrés de liberté des faces (ce qui prouve au passage que V_h n'est pas réduit aux seules fonctions constantes). Enfin, en assemblant les bases $(\psi_j)_{1 \leq j \leq n_k}$ de \mathbb{P}_k sur chaque maille K_i (fournies par le Lemme 4.3.3) on obtient la base annoncée $(\phi_i)_{1 \leq i \leq n_{al}}$ de V_h . \square

Remarque 4.3.8 On obtient un résultat semblable pour le sous-espace V_{0h} , défini par (4.41), qui est un sous-espace de $H_0^1(\Omega)$ de dimension finie égale au nombre de degrés de liberté intérieurs (on ne compte pas les noeuds sur le bord $\partial\Omega$). \bullet

Décrivons la **résolution pratique** du problème de Dirichlet (4.36) par la méthode des éléments finis \mathbb{P}_k . La formulation variationnelle (4.2) de l'approximation interne devient ici :

$$\text{trouver } u_h \in V_{0h} \text{ tel que } \int_{\Omega} \nabla u_h \cdot \nabla v_h \, dx = \int_{\Omega} f v_h \, dx \quad \forall v_h \in V_{0h}. \quad (4.42)$$

On décompose u_h sur la base des $(\phi_j)_{1 \leq j \leq n_{al}}$ et on prend $v_h = \phi_i$ ce qui donne

$$\sum_{j=1}^{n_{al}} u_h(\hat{a}_j) \int_{\Omega} \nabla \phi_j \cdot \nabla \phi_i \, dx = \int_{\Omega} f \phi_i \, dx.$$

En notant $U_h = (u_h(\hat{a}_j))_{1 \leq j \leq n_{al}}$, $b_h = (\int_{\Omega} f \phi_i \, dx)_{1 \leq i \leq n_{al}}$, et en introduisant la **matrice de rigidité**

$$\mathcal{K}_h = \left(\int_{\Omega} \nabla \phi_j \cdot \nabla \phi_i \, dx \right)_{1 \leq i, j \leq n_{al}},$$

la formulation variationnelle dans V_{0h} revient à résoudre dans $\mathbb{R}^{n_{al}}$ le système linéaire

$$\mathcal{K}_h U_h = b_h.$$

Comme les fonctions de base ϕ_j ont un “petit” support autour du noeud \hat{a}_i (voir la Figure 4.10), l'intersection des supports de ϕ_j et ϕ_i est souvent vide et la plupart des coefficients de \mathcal{K}_h sont nuls. On dit que la matrice \mathcal{K}_h est **creuse**.

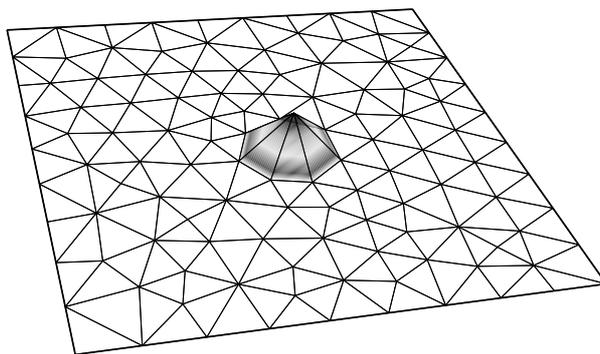


FIGURE 4.10 – Fonction de base \mathbb{P}_1 en dimension $N = 2$.

Pour calculer les coefficients de \mathcal{K}_h , on peut utiliser la formule d'intégration exacte suivante. On note $(\lambda_i(x))_{1 \leq i \leq N+1}$ les coordonnées barycentriques du point

courant x d'un N -simplexe K . Pour tout $\alpha_1, \dots, \alpha_{N+1} \in \mathbb{N}$, on a

$$\int_K \lambda_1(x)^{\alpha_1} \cdots \lambda_{N+1}(x)^{\alpha_{N+1}} dx = \text{Volume}(K) \frac{\alpha_1! \cdots \alpha_{N+1}! N!}{(\alpha_1 + \cdots + \alpha_{N+1} + N)!}. \quad (4.43)$$

Pour calculer le second membre b_h (et même éventuellement la matrice \mathcal{K}_h), on utilise des **formules de quadrature** (ou formules d'intégration numérique) qui donnent une approximation des intégrales sur chaque N -simplexe $K_i \in \mathcal{T}_h$. Par exemple, si K est un N -simplexe de sommets $(a_i)_{1 \leq i \leq N+1}$, les formules suivantes généralisent les formules en dimension 1, dites du "point milieu" et des "trapèzes" :

$$\int_K \psi(x) dx \approx \text{Volume}(K) \psi(a_0), \quad (4.44)$$

avec $a_0 = (N+1)^{-1} \sum_{i=1}^{N+1} a_i$, le barycentre de K , et

$$\int_K \psi(x) dx \approx \frac{\text{Volume}(K)}{N+1} \sum_{i=1}^{N+1} \psi(a_i). \quad (4.45)$$

Comme le montre l'Exercice 4.3.6, ces formules sont exactes pour des fonctions affines et sont donc approchées à l'ordre 2 en h pour des fonctions régulières.

La construction de la matrice \mathcal{K}_h est appelée **assemblage de la matrice**. La mise en oeuvre informatique de cette étape du calcul peut être assez compliquée, mais son coût en terme de temps de calcul est faible. Ce n'est pas le cas de la résolution du système linéaire $\mathcal{K}_h U_h = b_h$ qui est l'étape la **plus coûteuse** de la méthode en temps de calcul (et en place mémoire). En particulier, les calculs tridimensionnels sont encore très chers de nos jours dès que l'on utilise des maillages fins. L'Exercice 4.3.11 permet de s'en rendre compte. Heureusement, la matrice de rigidité \mathcal{K}_h est **creuse** (c'est-à-dire que la plupart de ses éléments sont nuls), ce qui permet de minimiser les calculs (pour plus de détails voir les algorithmes de résolution de systèmes linéaires dans la Section 4.4 de l'annexe). Rappelons que la matrice \mathcal{K}_h est nécessairement inversible par application du Lemme 4.1.1 et qu'elle est symétrique.

Exercice 4.3.5 Démontrer la formule (4.43) en dimension $N = 2$.

Exercice 4.3.6 Montrer que les formules (4.44) et (4.45) sont exactes pour $\psi \in \mathbb{P}_1$.

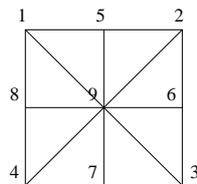


FIGURE 4.11 – Exemple de maillage et de numérotation des noeuds.

Exercice 4.3.9 On considère le carré $\Omega =]-1, +1[^2$ maillé suivant la Figure 4.11. Calculer la matrice de rigidité \mathcal{K}_h des éléments finis \mathbb{P}_1 appliqués au Laplacien avec condition aux limites de Neumann (on utilisera les symétries du maillage).

Exercice 4.3.10 Appliquer la méthode des éléments finis \mathbb{P}_1 au problème de Dirichlet (4.36) dans le carré $\Omega =]0, 1[^2$ avec le maillage triangulaire uniforme de la Figure 4.12. Montrer que la matrice de rigidité \mathcal{K}_h est la même matrice que celle que l'on obtiendrait par application de la méthode des différences finies (à un facteur multiplicatif h^2 près), mais que le second membre b_h est différent.

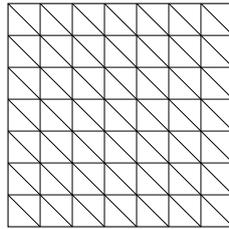


FIGURE 4.12 – Maillage triangulaire uniforme d'un carré.

Exercice 4.3.11 On reprend les notations de l'Exercice 4.3.10. On note n le nombre de points du maillage sur un côté du carré (supposé être le même pour chaque côté). On numérote "ligne par ligne" les noeuds du maillage (ou les degrés de liberté). Montrer que la matrice de rigidité \mathcal{K}_h des éléments finis \mathbb{P}_1 est de taille de l'ordre de n^2 et de largeur de bande de l'ordre de $2n$ (pour n grand).

Montrer que la même méthode et le même type de maillage pour le cube $\Omega =]0, 1[^3$ conduisent à une matrice de taille de l'ordre de n^3 et de largeur de bande de l'ordre de $2n^2$ (où n est le nombre de noeuds le long d'une arête du cube Ω).

L'exercice suivant montre que la méthode des éléments finis \mathbb{P}_1 vérifie le principe du maximum.

Exercice 4.3.12 On dit qu'une matrice carrée réelle $B = (b_{ij})_{1 \leq i, j \leq n}$ est une M-matrice si, pour tout i ,

$$b_{ii} > 0, \quad \sum_{k=1}^n b_{ik} > 0, \quad b_{ij} \leq 0 \quad \forall j \neq i.$$

Montrer que toute M-matrice est inversible et que tous les coefficients de son inverse sont positifs ou nuls.

Exercice 4.3.13 On se place en dimension $N = 2$. Soit u_h la solution approchée du problème de Dirichlet (4.36) obtenue par la méthode des éléments finis \mathbb{P}_1 . On suppose que tous les angles des triangles $K_i \in \mathcal{T}_h$ sont inférieurs ou égaux à $\pi/2$. Montrer que $u_h(x) \geq 0$ dans Ω si $f(x) \geq 0$ dans Ω . Indication : on montrera que, pour tout $\epsilon > 0$, $\mathcal{K}_h + \epsilon \text{Id}$ est une M-matrice, où \mathcal{K}_h est la matrice de rigidité.

4.3.2 Convergence et estimation d'erreur

Nous démontrons la convergence des méthodes d'éléments finis \mathbb{P}_k pour le problème de Dirichlet (4.36). Insistons sur le fait qu'il s'agit seulement d'un problème modèle, et que ces méthodes convergent pour d'autres problèmes, comme celui de Neumann. Nous allons avoir besoin d'hypothèses géométriques sur la qualité du maillage. Pour tout N -simplexe K on introduit deux paramètres géométriques : le **diamètre** $\text{diam}(K)$ et la **rondeur** $\rho(K)$, définie comme le diamètre de la plus grande boule contenue dans K ,

$$\text{diam}(K) = \max_{x,y \in K} \|x - y\|, \quad \rho(K) = \max_{B_r \subset K} (2r).$$

Bien sûr, on a toujours $\text{diam}(K)/\rho(K) > 1$. Ce rapport est d'autant plus grand que K est "aplati" : il mesure en quelque sorte la tendance à la dégénérescence de K . En pratique, comme en théorie, il faut éviter d'utiliser des N -simplexes K trop aplatis.

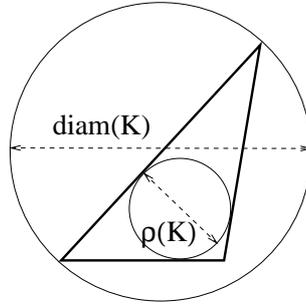


FIGURE 4.17 – Diamètre $\text{diam}(K)$ et rondeur $\rho(K)$ d'un triangle K .

Définition 4.3.11 Soit $(\mathcal{T}_h)_{h>0}$ une suite de maillages de Ω . On dit qu'il s'agit d'une suite de **maillages réguliers** si

1. la suite $h = \max_{K_i \in \mathcal{T}_h} \text{diam}(K_i)$ tend vers 0,
2. il existe une constante C telle que, pour tout $h > 0$ et tout $K \in \mathcal{T}_h$,

$$\frac{\text{diam}(K)}{\rho(K)} \leq C. \quad (4.46)$$

Remarque 4.3.12 En dimension $N = 2$ la condition (4.46) est équivalente à la condition suivante sur les angles du triangle K : il existe un angle minimum $\theta_0 > 0$ qui minore (uniformément en h) tous les angles de tout $K \in \mathcal{T}_h$. Insistons sur le fait que la condition (4.46) est tout aussi importante en pratique que pour l'analyse de convergence qui va suivre. •

Nous pouvons maintenant énoncer le résultat principal de cette sous-section qui affirme la convergence de la méthode des éléments finis \mathbb{P}_k et qui donne une estimation de la vitesse de convergence si la solution est régulière.

Théorème 4.3.13 Soit $(\mathcal{T}_h)_{h>0}$ une suite de maillages réguliers de Ω . Soit $u \in H_0^1(\Omega)$, la solution du problème de Dirichlet (4.36), et $u_h \in V_{0h}$, celle de son approximation interne (4.42) par la méthode des éléments finis \mathbb{P}_k . Alors la méthode des éléments finis \mathbb{P}_k converge, c'est-à-dire que

$$\lim_{h \rightarrow 0} \|u - u_h\|_{H^1(\Omega)} = 0. \quad (4.47)$$

De plus, si $u \in H^{k+1}(\Omega)$ et si $k + 1 > N/2$, alors on a l'estimation d'erreur

$$\|u - u_h\|_{H^1(\Omega)} \leq Ch^k \|u\|_{H^{k+1}(\Omega)}, \quad (4.48)$$

où C est une constante indépendante de h et de u .

Remarque 4.3.14 Le Théorème 4.3.13 s'applique en fait à toute méthode d'éléments finis de type Lagrange (par exemple, les éléments finis rectangulaires de la Sous-section 4.3.3). En effet, le seul argument utilisé est la construction d'un opérateur d'interpolation basé sur la caractérisation des fonctions de V_h par leurs valeurs aux noeuds des degrés de liberté, ce qui est toujours possible pour des éléments finis de type Lagrange (voir la Remarque 4.3.6). Remarquons que, pour les cas physiquement pertinents $N = 2$ ou $N = 3$, la condition $k + 1 > N/2$ est toujours satisfaite dès que $k \geq 1$. •

La démonstration du Théorème 4.3.13 repose sur la définition suivante d'un **opérateur d'interpolation** r_h et sur le résultat d'interpolation de la Proposition 4.3.16. Rappelons que nous avons noté $(\hat{a}_i)_{1 \leq i \leq n_{dl}}$ la famille des noeuds des degrés de liberté et $(\phi_i)_{1 \leq i \leq n_{dl}}$ la base de V_{0h} de la méthode des éléments finis \mathbb{P}_k (voir la Proposition 4.3.7). Pour toute fonction continue v , on définit son interpolée

$$r_h v(x) = \sum_{i=1}^{n_{dl}} v(\hat{a}_i) \phi_i(x). \quad (4.49)$$

La différence principale avec l'étude faite en dimension $N = 1$ est que, les fonctions de $H^1(\Omega)$ n'étant pas continues lorsque $N \geq 2$, l'opérateur d'interpolation r_h n'est pas défini sur $H^1(\Omega)$ (les valeurs ponctuelles d'une fonction de $H^1(\Omega)$ n'ont a priori pas de sens). Néanmoins, et c'est la raison de l'hypothèse $k + 1 > N/2$, r_h **est bien défini sur** $H^{k+1}(\Omega)$ car les fonctions de $H^{k+1}(\Omega)$ sont continues ($H^{k+1}(\Omega) \subset C(\bar{\Omega})$) d'après le Théorème 2.3.25).

Proposition 4.3.16 Soit $(\mathcal{T}_h)_{h>0}$ une suite de maillages réguliers de Ω . On suppose que $k + 1 > N/2$. Alors, pour tout $v \in H^{k+1}(\Omega)$ l'interpolée $r_h v$ est bien définie, et il existe une constante C , indépendante de h et de v , telle que

$$\|v - r_h v\|_{H^1(\Omega)} \leq Ch^k \|v\|_{H^{k+1}(\Omega)}. \quad (4.50)$$

En admettant la Proposition 4.3.16 nous pouvons conclure quant à la convergence de la méthode des éléments finis \mathbb{P}_k .

Démonstration du Théorème 4.3.13. On applique le cadre abstrait de la Sous-section 4.1.2. Pour démontrer (4.47) on utilise le Lemme 4.1.3 avec $\mathcal{V} = C_c^\infty(\Omega)$ qui est bien dense dans $H_0^1(\Omega)$. Comme $C_c^\infty(\Omega) \subset H^{k+1}(\Omega)$, l'estimation (4.50) de la Proposition 4.3.16 permet de vérifier l'hypothèse (4.5) du Lemme 4.1.3 (pour des fonctions régulières on n'a pas besoin de la condition $k+1 > N/2$ dans la Proposition 4.3.16).

Pour obtenir l'estimation d'erreur (4.48) on utilise le Lemme de Céa 4.1.2 qui nous dit que

$$\|u - u_h\|_{H^1(\Omega)} \leq C \inf_{v_h \in V_{0h}} \|u - v_h\|_{H^1(\Omega)} \leq C \|u - r_h u\|_{H^1(\Omega)},$$

si $r_h u$ appartient bien à $H^1(\Omega)$. Par application de la Proposition 4.3.16 à u on obtient (4.48). \square

4.3.3 Éléments finis rectangulaires

Si le domaine Ω est de type rectangulaire (c'est-à-dire que Ω est un ouvert polyédrique dont les faces sont perpendiculaires aux axes), on peut le mailler par des rectangles (voir la Figure 4.20) et utiliser une méthode d'éléments finis adaptée. Nous allons définir des éléments finis de type Lagrange (c'est-à-dire dont les degrés de liberté sont des valeurs ponctuelles de fonctions), dits éléments finis \mathbb{Q}_k . Commençons par définir un N -rectangle K de \mathbb{R}^N comme le pavé (non-dégénéré) $\prod_{i=1}^N [l_i, L_i]$ avec $-\infty < l_i < L_i < +\infty$. On note $(a_j)_{1 \leq j \leq 2^N}$ les sommets de K .

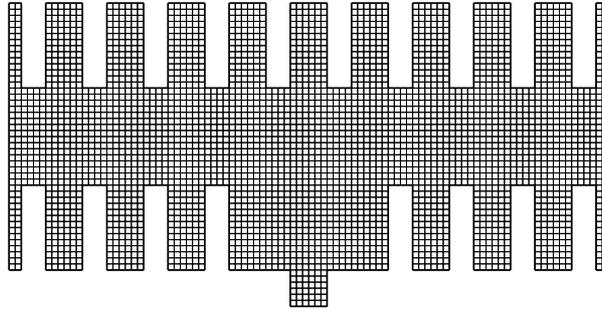


FIGURE 4.20 – Exemple de maillage rectangulaire en dimension $N = 2$.

Définition 4.3.21 Soit Ω un ouvert connexe polyédrique de \mathbb{R}^N . Un **maillage rectangulaire** de $\bar{\Omega}$ est un ensemble \mathcal{T}_h de N -rectangles (non dégénérés) $(K_i)_{1 \leq i \leq n}$ qui vérifient

1. $K_i \subset \bar{\Omega}$ et $\bar{\Omega} = \cup_{i=1}^n K_i$,
2. l'intersection $K_i \cap K_j$ de deux N -rectangles distincts est un m -rectangle, avec $0 \leq m \leq N - 1$, dont tous les sommets sont aussi des sommets de K_i et K_j . (En dimension $N = 2$, l'intersection de deux rectangles est soit vide, soit un sommet commun, soit une face commune entière.)

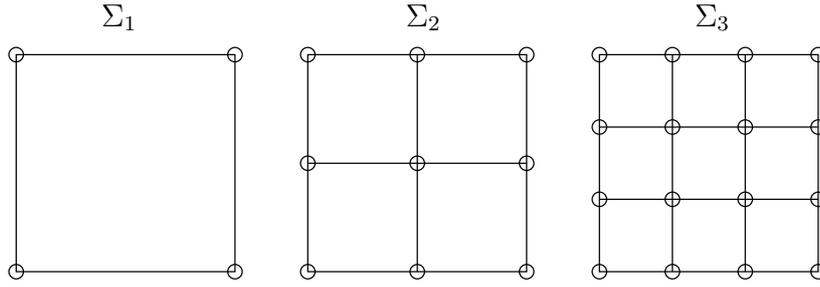


FIGURE 4.21 – Treillis d'ordre 1, 2, et 3 pour un rectangle (les ronds représentent les points du treillis).

Les **sommets** ou **noeuds** du maillage \mathcal{T}_h sont les sommets des N -rectangles K_i qui le composent. Par convention, le paramètre h désigne le maximum des diamètres des N -rectangles K_i .

Nous définissons l'ensemble \mathbb{Q}_k des polynômes à coefficients réels de \mathbb{R}^N dans \mathbb{R} de degré inférieur ou égal à k **par rapport à chaque variable**, c'est-à-dire que tout $p \in \mathbb{Q}_k$ s'écrit sous la forme

$$p(x) = \sum_{0 \leq i_1 \leq k, \dots, 0 \leq i_N \leq k} \alpha_{i_1, \dots, i_N} x_1^{i_1} \cdots x_N^{i_N} \text{ avec } x = (x_1, \dots, x_N).$$

Remarquons que le degré total de p peut être supérieur à k , ce qui différencie l'espace \mathbb{Q}_k de \mathbb{P}_k .

Pour tout entier $k \geq 1$ on définit le **treillis d'ordre k** du N -rectangle K comme l'ensemble

$$\Sigma_k = \left\{ x \in K \text{ tel que } \frac{x_j - l_j}{L_j - l_j} \in \left\{ 0, \frac{1}{k}, \dots, \frac{k-1}{k}, 1 \right\} \text{ pour } 1 \leq j \leq N \right\}. \quad (4.51)$$

Pour $k = 1$ il s'agit de l'ensemble des sommets de K , et pour $k = 2$ et $N = 2$ des sommets, des points milieux des arêtes reliant deux sommets, et du barycentre (voir la Figure 4.21).

Le treillis Σ_k d'un N -rectangle K est **unisolvant** pour \mathbb{Q}_k , c'est-à-dire qu'il permet de caractériser tous les polynômes de \mathbb{Q}_k .

Lemme 4.3.22 *Soit K un N -rectangle. Soit un entier $k \geq 1$. Alors, tout polynôme de \mathbb{Q}_k est déterminé de manière unique par ses valeurs aux points du treillis d'ordre k , Σ_k , défini par (4.51).*

Démonstration. On vérifie que le cardinal de Σ_k et la dimension de \mathbb{Q}_k coïncident

$$\text{card}(\Sigma_k) = \dim(\mathbb{Q}_k) = (k+1)^N.$$

Comme l'application qui, à tout polynôme de \mathbb{Q}_k , fait correspondre ses valeurs sur le treillis Σ_k est linéaire, il suffit d'exhiber une base de \mathbb{Q}_k dont les éléments valent

1 en un point du treillis et 0 ailleurs pour démontrer le résultat. Soit un point x^μ de Σ_k défini par

$$\frac{x_j^\mu - l_j}{L_j - l_j} = \frac{\mu_j}{k} \quad \text{avec } 0 \leq \mu_j \leq k, \quad \forall j \in \{1, \dots, N\}.$$

On définit le polynôme $p \in \mathbb{Q}_k$ par

$$p(x) = \prod_{j=1}^N \left(\prod_{\substack{i=0 \\ i \neq \mu_j}}^k \frac{k(x_j - l_j) - i(L_j - l_j)}{(\mu_j - i)(L_j - l_j)} \right) \quad \text{avec } x = (x_1, \dots, x_N).$$

On vérifie facilement que $p(x^\mu) = 1$ tandis que p s'annule sur tous les autres points de Σ_k , ce qui est le résultat désiré. \square

Comme dans le cas triangulaire nous avons la condition suivante de continuité à travers une face (nous laissons la démonstration, tout à fait similaire à celle du Lemme 4.3.4, au lecteur).

Lemme 4.3.23 *Soit K et K' deux N -rectangles ayant une face commune $\Gamma = \partial K \cap \partial K'$. Soit un entier $k \geq 1$. Alors, leur treillis d'ordre k Σ_k et Σ'_k coïncident sur cette face Γ . De plus, étant donné p_K et $p_{K'}$ deux polynômes de \mathbb{Q}_k , la fonction v définie par*

$$v(x) = \begin{cases} p_K(x) & \text{si } x \in K \\ p_{K'}(x) & \text{si } x \in K' \end{cases}$$

est continue sur $K \cup K'$, si et seulement si p_K et $p_{K'}$ ont des valeurs qui coïncident aux points du treillis sur la face commune Γ .

En pratique, on utilise surtout les espaces \mathbb{Q}_1 et \mathbb{Q}_2 . La Figure 4.22 montre une fonction de base \mathbb{Q}_1 en dimension $N = 2$ (on peut y vérifier que les fonctions de \mathbb{Q}_1 ne sont pas affines par morceaux comme celles de \mathbb{P}_1).

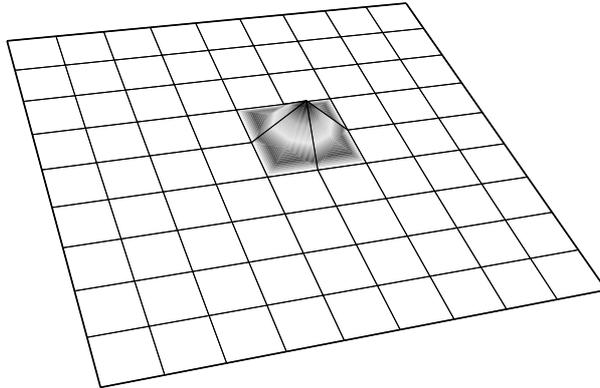


FIGURE 4.22 – Fonction de base \mathbb{Q}_1 en dimension $N = 2$.

Définition 4.3.25 *Étant donné un maillage rectangulaire \mathcal{T}_h d'un ouvert Ω , la méthode des éléments finis \mathbb{Q}_k est définie par l'espace discret*

$$V_h = \{v \in C(\overline{\Omega}) \text{ tel que } v|_{K_i} \in \mathbb{Q}_k \text{ pour tout } K_i \in \mathcal{T}_h\}. \quad (4.52)$$

On appelle noeuds des **degrés de liberté** l'ensemble des points $(\hat{a}_i)_{1 \leq i \leq n_{dl}}$ des treillis d'ordre k de chacun des N -rectangles $K_i \in \mathcal{T}_h$.

Comme dans le cas triangulaire, la Définition 4.3.25 a un sens grâce à la proposition suivante (dont nous laissons la démonstration au lecteur en guise d'exercice).

Proposition 4.3.26 *L'espace V_h , défini par (4.52), est un sous-espace de $H^1(\Omega)$ dont la dimension est le nombre de degrés de liberté n_{dl} . De plus, il existe une base de V_h $(\phi_i)_{1 \leq i \leq n_{dl}}$ définie par*

$$\phi_i(\hat{a}_j) = \delta_{ij} \quad 1 \leq i, j \leq n_{dl},$$

telle que

$$v(x) = \sum_{i=1}^{n_{dl}} v(\hat{a}_i) \phi_i(x).$$

Comme les éléments finis \mathbb{Q}_k sont des éléments finis de type Lagrange, on peut démontrer les mêmes résultats de convergence que pour la méthode des éléments finis \mathbb{P}_k .

4.4 Résolution des systèmes linéaires

Cette section est consacrée à la résolution des systèmes linéaires. Pour plus de détails nous renvoyons aux ouvrages [2] et [9]. On appelle système linéaire le problème qui consiste à trouver la ou les solutions $x \in \mathbb{R}^n$ (si elle existe) de l'équation algébrique suivante

$$Ax = b, \quad (4.53)$$

où A appartient à l'ensemble $\mathcal{M}_n(\mathbb{R})$ des matrices réelles carrées d'ordre n , et $b \in \mathbb{R}^n$ est un vecteur appelé second membre. Nous allons voir deux types de méthodes de résolution de systèmes linéaires : celles dites directes, c'est-à-dire qui permettent de calculer la solution exacte en un nombre fini d'opérations, et celles dites **itératives**, c'est-à-dire qui calculent une suite de solutions approchées qui converge vers la solution exacte.

4.4.1 Rappels sur les normes matricielles

Nous commençons par rappeler la notion de **norme subordonnée** pour les matrices. Même si l'on considère des matrices réelles, il est nécessaire, pour des raisons techniques qui seront exposées à la Remarque 4.4.4, de les traiter comme des matrices complexes.

Définition 4.4.1 Soit $\|\cdot\|$ une norme vectorielle sur \mathbb{C}^n . On lui associe une norme matricielle, dite subordonnée à cette norme vectorielle, définie par

$$\|A\| = \sup_{x \in \mathbb{C}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Par abus de langage on note de la même façon les normes vectorielle et matricielle subordonnée. On vérifie aisément qu'une norme subordonnée ainsi définie est bien une norme matricielle et qu'elle vérifie le résultat suivant.

Lemme 4.4.2 Soit $\|\cdot\|$ une norme matricielle subordonnée sur $\mathcal{M}_n(\mathbb{C})$.

1. Pour toute matrice A , la norme $\|A\|$ est aussi définie par

$$\|A\| = \sup_{x \in \mathbb{C}^n, \|x\|=1} \|Ax\| = \sup_{x \in \mathbb{C}^n, \|x\| \leq 1} \|Ax\|.$$

2. Il existe $x_A \in \mathbb{C}^n, x_A \neq 0$ tel que $\|A\| = \frac{\|Ax_A\|}{\|x_A\|}$.

3. La matrice identité vérifie $\|\text{Id}\| = 1$.

4. Soient A et B deux matrices. On a $\|AB\| \leq \|A\| \|B\|$.

On note $\|A\|_p$ la norme matricielle subordonnée à la norme vectorielle sur \mathbb{C}^n définie pour $p \geq 1$ par $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$, et pour $p = +\infty$ par $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$. On peut calculer explicitement certaines de ces normes subordonnées. (Dans tout ce qui suit on note A^* la matrice adjointe de A .)

Exercice 4.4.1 Montrer que

1. $\|A\|_2 = \|A^*\|_2 = \text{maximum des valeurs singulières de } A$,
2. $\|A\|_1 = \max_{1 \leq j \leq n} (\sum_{i=1}^n |a_{ij}|)$,
3. $\|A\|_\infty = \max_{1 \leq i \leq n} (\sum_{j=1}^n |a_{ij}|)$.

Remarque 4.4.4 Une matrice réelle peut être considérée soit comme une matrice de $\mathcal{M}_n(\mathbb{R})$, soit comme une matrice de $\mathcal{M}_n(\mathbb{C})$ car $\mathbb{R} \subset \mathbb{C}$. Si $\|\cdot\|_{\mathbb{C}}$ est une norme vectorielle dans \mathbb{C}^n , on peut définir sa restriction $\|\cdot\|_{\mathbb{R}}$ à \mathbb{R}^n qui est aussi une norme vectorielle dans \mathbb{R}^n . Pour une matrice réelle $A \in \mathcal{M}_n(\mathbb{R})$, on peut donc définir deux normes matricielles subordonnées $\|A\|_{\mathbb{C}}$ et $\|A\|_{\mathbb{R}}$ par

$$\|A\|_{\mathbb{C}} = \sup_{x \in \mathbb{C}^n, x \neq 0} \frac{\|Ax\|_{\mathbb{C}}}{\|x\|_{\mathbb{C}}} \text{ et } \|A\|_{\mathbb{R}} = \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|_{\mathbb{R}}}{\|x\|_{\mathbb{R}}}.$$

A priori ces deux définitions peuvent être distinctes. Grâce aux formules explicites de l'Exercice 4.4.1, on sait qu'elles coïncident si $\|x\|_{\mathbb{C}}$ est une des normes $\|x\|_1, \|x\|_2$, ou $\|x\|_\infty$. Cependant, pour d'autres normes vectorielles on peut avoir $\|A\|_{\mathbb{C}} > \|A\|_{\mathbb{R}}$. Par ailleurs, dans la preuve de la Proposition 4.4.7 on a besoin de la définition sur \mathbb{C} de la norme subordonnée même si la matrice est réelle. C'est pourquoi on utilise \mathbb{C} dans la Définition 4.4.1 de la norme subordonnée. •

Définition 4.4.5 Soit A une matrice dans $\mathcal{M}_n(\mathbb{C})$. On appelle rayon spectral de A , et on note $\rho(A)$, le maximum des modules des valeurs propres de A .

Proposition 4.4.7 Soit $\|\cdot\|$ une norme subordonnée sur $\mathcal{M}_n(\mathbb{C})$. On a

$$\rho(A) \leq \|A\|.$$

Réciproquement, pour toute matrice A et pour tout réel $\epsilon > 0$, il existe une norme subordonnée $\|\cdot\|$ (qui dépend de A et ϵ) telle que

$$\|A\| \leq \rho(A) + \epsilon. \quad (4.54)$$

Lemme 4.4.8 Soit A une matrice de $\mathcal{M}_n(\mathbb{C})$. Les quatre conditions suivantes sont équivalentes

1. $\lim_{i \rightarrow +\infty} A^i = 0$,
2. $\lim_{i \rightarrow +\infty} A^i x = 0$ pour tout vecteur $x \in \mathbb{C}^n$,
3. $\rho(A) < 1$,
4. il existe au moins une norme matricielle subordonnée telle que $\|A\| < 1$.

4.4.2 Conditionnement et stabilité

Avant de décrire les algorithmes de résolution de systèmes linéaires, il nous faut évoquer les problèmes de précision et de stabilité dus aux erreurs d'arrondi. En effet, dans un ordinateur il n'y a pas de calculs exacts, et la précision est limitée à cause du nombre de bits utilisés pour représenter les nombres réels : d'habitude 32 ou 64 bits (ce qui fait à peu près 8 ou 16 chiffres significatifs). Il faut donc faire très attention aux inévitables erreurs d'arrondi et à leur propagation au cours d'un calcul. Les méthodes numériques de résolution de systèmes linéaires qui n'amplifient pas ces erreurs sont dites stables. En pratique, on utilisera donc des algorithmes qui sont à la fois **efficaces et stables**. Cette amplification des erreurs dépend de la matrice considérée. Pour quantifier ce phénomène, on introduit la notion de conditionnement d'une matrice.

Définition 4.4.9 Soit une norme matricielle subordonnée $\|A\|$. On appelle conditionnement d'une matrice $A \in \mathcal{M}_n(\mathbb{C})$, relatif à cette norme, la valeur définie par

$$\text{cond}(A) = \|A\| \cdot \|A^{-1}\|.$$

Cette notion de conditionnement va permettre de mesurer l'amplification des erreurs des données (second membre ou matrice) au résultat.

Proposition 4.4.10 Soit A une matrice inversible. Soit $b \neq 0$ un vecteur non nul.

1. Soit x et $x + \delta x$ les solutions respectives des systèmes

$$Ax = b, \text{ et } A(x + \delta x) = b + \delta b.$$

Alors on a

$$\frac{\|\delta x\|}{\|x\|} \leq \text{cond}(A) \frac{\|\delta b\|}{\|b\|}. \quad (4.55)$$

2. Soit x et $x + \delta x$ les solutions respectives des systèmes

$$Ax = b, \text{ et } (A + \delta A)(x + \delta x) = b.$$

Alors on a

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq \text{cond}(A) \frac{\|\delta A\|}{\|A\|}. \quad (4.56)$$

De plus, ces inégalités sont optimales.

Remarque 4.4.11 On dira qu'une matrice est bien conditionnée si son conditionnement est proche de 1 (sa valeur minimale) et qu'elle est mal conditionnée si son conditionnement est grand. A cause des résultats de la Proposition 4.4.10, en pratique il faudra faire attention aux erreurs d'arrondi si on résout un système linéaire pour une matrice mal conditionnée. •

Démonstration. Pour montrer le premier résultat, on remarque que $A\delta x = \delta b$, et donc $\|\delta x\| \leq \|A^{-1}\| \|\delta b\|$. Or, on a aussi $\|b\| \leq \|A\| \|x\|$, ce qui donne (4.55). Cette inégalité est optimale au sens suivant : pour toute matrice A , il existe δb et x (qui dépendent de A) tels que (4.55) est en fait une égalité. En effet, d'après une propriété des normes matricielles subordonnées (voir le Lemme 4.4.2) il existe x tel que $\|b\| = \|A\| \|x\|$ et il existe δb tel que $\|\delta x\| = \|A^{-1}\| \|\delta b\|$.

Pour obtenir (4.56) on remarque que $A\delta x + \delta A(x + \delta x) = 0$, et donc $\|\delta x\| \leq \|A^{-1}\| \|\delta A\| \|x + \delta x\|$, ce qui implique (4.56). Pour en démontrer l'optimalité, on va montrer que pour toute matrice A il existe une perturbation δA et un second membre b pour lesquels il y a égalité. Grâce au Lemme 4.4.2 il existe $y \neq 0$ tel que $\|A^{-1}y\| = \|A^{-1}\| \|y\|$. Soit ϵ un scalaire non nul. On pose $\delta A = \epsilon Id$ et $b = (A + \delta A)y$. On vérifie alors que $y = y + \delta x$ et $\delta x = -\epsilon A^{-1}y$, et comme $\|\delta A\| = |\epsilon|$ on obtient l'égalité dans (4.56). □

Les conditionnements les plus utilisés en pratique sont ceux associés aux normes $\|A\|_p$ avec $p = 1, 2, +\infty$.

Exercice 4.4.2 Soit une matrice $A \in \mathcal{M}_n(\mathbb{C})$. Vérifier que

1. $\text{cond}(A) = \text{cond}(A^{-1}) \geq 1$, $\text{cond}(\alpha A) = \text{cond}(A) \forall \alpha \neq 0$,
2. pour une matrice quelconque, $\text{cond}_2(A) = \frac{\mu_n(A)}{\mu_1(A)}$, où $\mu_1(A), \mu_n(A)$ sont respectivement la plus petite et la plus grande valeur singulière de A ,
3. pour une matrice normale, $\text{cond}_2(A) = \frac{|\lambda_n(A)|}{|\lambda_1(A)|}$, où $|\lambda_1(A)|, |\lambda_n(A)|$ sont respectivement la plus petite et la plus grande valeur propre en module de A ,
4. pour toute matrice unitaire U , $\text{cond}_2(U) = 1$,
5. pour toute matrice unitaire U , $\text{cond}_2(AU) = \text{cond}_2(UA) = \text{cond}_2(A)$.

4.4.3 Méthodes directes

Méthode d'élimination de Gauss

L'idée principale de cette méthode est de se ramener à la résolution d'un système linéaire dont la matrice est triangulaire. En effet, la résolution d'un système linéaire, $Tx = b$, où la matrice T est triangulaire et inversible, est très facile par simple substitution récursive. On appelle ce procédé **remontée** dans le cas d'une matrice triangulaire supérieure et **descente** dans le cas d'une matrice triangulaire inférieure. Remarquons que l'on résout ainsi le système $Tx = b$ sans inverser la matrice T . De la même manière, la méthode d'élimination de Gauss va résoudre le système $Ax = b$ sans calculer l'inverse de la matrice A .

La méthode d'élimination de Gauss se décompose en trois étapes :

- (i) élimination : calcul d'une matrice M inversible telle que $MA = T$ soit triangulaire supérieure,
- (ii) mise à jour du second membre : calcul simultané de Mb ,
- (iii) substitution : résolution du système triangulaire $Tx = Mb$ par simple remontée.

L'existence d'une telle matrice M est garantie par le résultat suivant dont on va donner une démonstration constructive qui n'est rien d'autre que la méthode d'élimination de Gauss.

Proposition 4.4.13 *Soit A une matrice carrée (inversible ou non). Il existe au moins une matrice inversible M telle que la matrice $T = MA$ soit triangulaire supérieure.*

Démonstration. Le principe est de construire une suite de matrices A^k , $1 \leq k \leq n$, dont les $(k - 1)$ premières colonnes sont remplies de zéros sous la diagonale. Par modifications successives, on passe de $A^1 = A$ à $A^n = T$ qui est triangulaire supérieure. On note $(a_{ij}^k)_{1 \leq i, j \leq n}$ les éléments de la matrice A^k , et on appelle pivot de A^k l'élément a_{kk}^k . Pour passer de la matrice A^k à la matrice A^{k+1} , on s'assure tout d'abord que le pivot a_{kk}^k n'est pas nul. S'il l'est, on permute la k -ème ligne avec une autre ligne pour amener en position de pivot un élément non nul. Puis on procède à l'élimination de tous les éléments de la k -ème colonne en dessous de la k -ème ligne en faisant des combinaisons linéaires de la ligne courante avec la k -ème ligne. \square

Méthode de la factorisation LU

La méthode LU consiste à factoriser la matrice A en un produit de deux matrices triangulaires $A = LU$, où L est triangulaire inférieure (L pour "lower" en anglais) et U est triangulaire supérieure (U pour "upper" en anglais). Il s'agit en fait du même algorithme que celui de l'élimination de Gauss dans le cas particulier où **on ne pivote jamais**. Une fois établie la factorisation LU de A , la résolution du système linéaire $Ax = b$ est équivalente à la simple résolution de deux systèmes triangulaires $Ly = b$ puis $Ux = y$.

Proposition 4.4.15 Soit une matrice $A = (a_{ij})_{1 \leq i, j \leq n}$ d'ordre n telle que toutes les sous-matrices diagonales d'ordre k , définies par

$$\Delta^k = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix},$$

soient inversibles. Il existe un unique couple de matrices (L, U) , avec U triangulaire supérieure, et L triangulaire inférieure ayant une diagonale de 1, tel que

$$A = LU.$$

Calcul pratique de la factorisation LU. On peut calculer la factorisation LU (si elle existe) d'une matrice A par identification de A au produit LU . En posant $A = (a_{ij})_{1 \leq i, j \leq n}$, et

$$L = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ l_{2,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ l_{n,1} & \cdots & l_{n,n-1} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} u_{1,1} & \cdots & \cdots & u_{1,n} \\ 0 & u_{2,2} & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{n,n} \end{pmatrix},$$

comme L est triangulaire inférieure et U triangulaire supérieure, pour $1 \leq i, j \leq n$ il vient

$$a_{i,j} = \sum_{k=1}^n l_{i,k} u_{k,j} = \sum_{k=1}^{\min(i,j)} l_{i,k} u_{k,j}.$$

En identifiant par ordre croissant les colonnes de A on en déduit les colonnes de L et de U . Ainsi, après avoir calculé les $(j-1)$ premières colonnes de L et de U en fonction des $(j-1)$ premières colonnes de A , on lit la j -ème colonne de A

$$\begin{aligned} a_{i,j} &= \sum_{k=1}^i l_{i,k} u_{k,j} \Rightarrow u_{i,j} = a_{i,j} - \sum_{k=1}^{i-1} l_{i,k} u_{k,j} \quad \text{pour } 1 \leq i \leq j, \\ a_{i,j} &= \sum_{k=1}^j l_{i,k} u_{k,j} \Rightarrow l_{i,j} = \frac{a_{i,j} - \sum_{k=1}^{j-1} l_{i,k} u_{k,j}}{u_{j,j}} \quad \text{pour } j+1 \leq i \leq n. \end{aligned}$$

On calcule donc les j premières composantes de la j -ème colonne de U et les $n-j$ dernières composantes de la j -ème colonne de L en fonction de leurs $(j-1)$ premières colonnes. On divise par le pivot u_{jj} qui doit donc être non nul!

Compte d'opérations. Pour mesurer l'efficacité de l'algorithme de la décomposition LU on compte le nombre d'opérations nécessaires à son accomplissement (qui sera proportionnel à son temps d'exécution sur un ordinateur). On ne calcule pas exactement ce nombre d'opérations, et on se contente du premier terme de son développement asymptotique lorsque la dimension n est grande. De plus, pour simplifier on ne compte que les multiplications et divisions (et pas les additions dont le nombre est en général du même ordre de grandeur).

– Élimination ou décomposition LU : le nombre d'opérations N_{op} est

$$N_{op} = \sum_{j=1}^{n-1} \sum_{i=j+1}^n (1 + \sum_{k=j+1}^n 1),$$

qui, au premier ordre, donne $N_{op} \approx n^3/3$.

– Substitution (ou remontée-descente sur les deux systèmes triangulaires) : le nombre d'opérations N_{op} est donné par la formule

$$N_{op} = 2 \sum_{j=1}^n j,$$

qui, au premier ordre, donne $N_{op} \approx n^2$.

Au total la résolution d'un système linéaire $Ax = b$ par la méthode de la factorisation LU demande $N_{op} \approx n^3/3$ opérations car n^2 est négligeable devant n^3 quand n est grand.

Méthode de Cholesky

C'est une méthode qui ne s'applique qu'aux matrices symétriques réelles, définies positives. Elle consiste à factoriser une matrice A sous la forme $A = BB^*$ où B est une matrice triangulaire inférieure (et B^* son adjointe ou transposée).

Proposition 4.4.19 *Soit A une matrice symétrique réelle, définie positive. Il existe une unique matrice réelle B triangulaire inférieure, telle que tous ses éléments diagonaux soient positifs, et qui vérifie*

$$A = BB^*.$$

Calcul pratique de la factorisation de Cholesky. En pratique, on calcule le facteur de Cholesky B par identification dans l'égalité $A = BB^*$. Soit $A = (a_{ij})_{1 \leq i, j \leq n}$, $B = (b_{ij})_{1 \leq i, j \leq n}$ avec $b_{ij} = 0$ si $i < j$. Pour $1 \leq i, j \leq n$, il vient

$$a_{ij} = \sum_{k=1}^n b_{ik}b_{jk} = \sum_{k=1}^{\min(i,j)} b_{ik}b_{jk}.$$

En identifiant par ordre croissant les colonnes de A (ou ses lignes, ce qui revient au même puisque A est symétrique) on en déduit les colonnes de B . Ainsi, après avoir calculé les $(j-1)$ premières colonnes de B en fonction des $(j-1)$ premières colonnes de A , on lit la j -ème colonne de A en dessous de la diagonale

$$\begin{aligned} a_{jj} &= \sum_{k=1}^j (b_{jk})^2 \Rightarrow b_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} (b_{jk})^2} \\ a_{i,j} &= \sum_{k=1}^j b_{jk}b_{i,k} \Rightarrow b_{i,j} = \frac{a_{i,j} - \sum_{k=1}^{j-1} b_{jk}b_{i,k}}{b_{jj}} \text{ pour } j+1 \leq i \leq n. \end{aligned}$$

On calcule donc la j -ème colonne de B en fonction de ses $(j-1)$ premières colonnes. A cause du théorème précédent, on est sûr que, si A est symétrique définie positive, les termes sous les racines carrées sont strictement positifs. Au contraire, si A n'est pas définie positive, on trouvera que $a_{jj} - \sum_{k=1}^{j-1} (b_{jk})^2 \leq 0$ pour un certain rang j , ce qui empêche de terminer l'algorithme.

Compte d'opérations. Pour mesurer l'efficacité de la méthode de Cholesky on compte le nombre d'opérations (uniquement les multiplications) nécessaires à son accomplissement. Le nombre de racines carrées est n qui est négligeable dans ce compte d'opérations.

– Factorisation de Cholesky : le nombre d'opérations N_{op} est

$$N_{op} = \sum_{j=1}^n \left((j-1) + \sum_{i=j+1}^n j \right),$$

qui, au premier ordre, donne $N_{op} \approx n^3/6$.

– Substitution : il faut effectuer une remontée et une descente sur les systèmes triangulaire associés à B et B^* . Le nombre d'opérations est au premier ordre $N_{op} \approx n^2$.

La méthode de Cholesky est donc approximativement **deux fois plus rapide** que celle de Gauss pour une matrice symétrique définie positive.

Matrices bandes et matrices creuses

Lorsqu'une matrice a beaucoup de coefficients nuls, on dit qu'elle est **creuse**. Si les éléments non nuls sont répartis à proximité de la diagonale, on dit que la matrice a une structure **bande**. Pour ces deux types de matrices (qui apparaissent naturellement dans la méthode des éléments finis comme dans la plupart des autres méthodes), on peut améliorer le compte d'opérations et la taille de stockage nécessaire pour résoudre un système linéaire. Ce gain est très important en pratique.

Définition 4.4.20 Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est dite matrice bande, de demi largeur de bande (hors diagonale) $p \in \mathbb{N}$ si ses éléments vérifient $a_{i,j} = 0$ pour $|i-j| > p$. La largeur de la bande est alors $2p+1$.

L'intérêt des matrices bandes vient de la propriété suivante.

Exercice 4.4.4 Montrer que les factorisations LU et de Cholesky conservent la structure bande des matrices.

Remarque 4.4.21 Si les factorisations LU et de Cholesky préservent la structure bande des matrices, il n'en est pas de même de leur structure creuse. En général, si A est creuse (même à l'intérieur d'une bande), les facteurs L et U , ou B et B^* sont "pleins" (le contraire de creux) à l'intérieur de la même bande. •

L'exercice suivant permet de quantifier le gain qu'il y a à utiliser des matrices bandes.

Exercice 4.4.5 Montrer que, pour une matrice bande d'ordre n et de demie largeur de bande p , le compte d'opérations de la factorisation LU est $\mathcal{O}(np^2/3)$ et celui de la factorisation de Cholesky est $\mathcal{O}(np^2/6)$.

4.4.4 Méthodes itératives

Les méthodes itératives sont particulièrement intéressantes pour les très grandes matrices ou les matrices creuses. En effet, dans ce cas les méthodes directes peuvent avoir un coût de calcul et de stockage en mémoire prohibitif (se rappeler que la factorisation LU ou de Cholesky demande de l'ordre de n^3 opérations). Commençons par une classe très simple de méthodes itératives.

Définition 4.4.24 Soit A une matrice inversible. On introduit une décomposition régulière de A (en anglais "splitting"), c'est-à-dire un couple de matrices (M, N) avec M inversible (et facile à inverser dans la pratique) tel que $A = M - N$. La méthode itérative basée sur le splitting (M, N) est définie par

$$\begin{cases} x_0 \text{ donné dans } \mathbb{R}^n, \\ Mx_{k+1} = Nx_k + b \quad \forall k \geq 1. \end{cases} \quad (4.60)$$

Si la suite de solutions approchées x_k converge vers une limite x quand k tend vers l'infini, alors, par passage à la limite dans la relation de récurrence (4.60), on obtient

$$(M - N)x = Ax = b.$$

Par conséquent, si la suite de solutions approchées converge, sa limite est forcément la solution du système linéaire.

D'un point de vue pratique, il faut savoir quand on peut arrêter les itérations, c'est-à-dire à quel moment x_k est suffisamment proche de la solution inconnue x . Comme on ne connaît pas x , on ne peut pas décider d'arrêter le calcul dès que $\|x - x_k\| \leq \epsilon$ où ϵ est la précision désirée. Par contre on connaît Ax (qui vaut b), et un critère d'arrêt fréquemment utilisé est $\|b - Ax_k\| \leq \epsilon$. Cependant, si la norme de A^{-1} est grande ce critère peut être trompeur car

$$\|x - x_k\| \leq \|A^{-1}\| \|b - Ax_k\| \leq \epsilon \|A^{-1}\|$$

qui peut ne pas être petit.

Définition 4.4.25 On dit qu'une méthode itérative est convergente si, quel que soit le choix du vecteur initial $x_0 \in \mathbb{R}^n$, la suite de solutions approchées x_k converge vers la solution exacte x .

On commence par donner une condition nécessaire et suffisante de convergence d'une méthode itérative à l'aide du rayon spectral de la matrice d'itération (voir la Définition 4.4.5 pour la notion de rayon spectral).

Lemme 4.4.26 La méthode itérative définie par (4.60) converge si et seulement si le rayon spectral de la matrice d'itération $M^{-1}N$ vérifie $\rho(M^{-1}N) < 1$.

Démonstration. On définit l'erreur $e_k = x_k - x$. On a

$$e_k = (M^{-1}Nx_{k-1} + M^{-1}b) - (M^{-1}Nx + M^{-1}b) = M^{-1}Ne_{k-1} = (M^{-1}N)^k e_0.$$

Par application du Lemme 4.4.8, on en déduit que e_k tend vers 0, quel que soit e_0 , si et seulement si $\rho(M^{-1}N) < 1$. \square

Définition 4.4.29 (méthode de Jacobi) Soit $A = (a_{ij})_{1 \leq i, j \leq n}$. On note $D = \text{diag}(a_{ii})$ la diagonale de A . On appelle méthode de Jacobi la méthode itérative associée à la décomposition

$$M = D, \quad N = D - A.$$

Définition 4.4.30 (méthode de Gauss-Seidel) Soit $A = (a_{ij})_{1 \leq i, j \leq n}$. On décompose A sous la forme $A = D - E - F$ où $D = \text{diag}(a_{ii})$ est la diagonale, $-E$ est la partie triangulaire inférieure (strictement), et $-F$ est la partie triangulaire supérieure (strictement) de A . On appelle méthode de Gauss-Seidel la méthode itérative associée à la décomposition

$$M = D - E, \quad N = F.$$

Définition 4.4.31 (méthode de relaxation (SOR)) Soit $\omega \in \mathbb{R}^+$. On appelle méthode de relaxation (SOR en anglais pour "Successive Over Relaxation"), pour le paramètre ω , la méthode itérative associée à la décomposition

$$M = \frac{D}{\omega} - E, \quad N = \frac{1 - \omega}{\omega}D + F$$

Définition 4.4.32 (méthode du gradient) Soit un paramètre réel $\alpha \neq 0$. On appelle méthode du gradient la méthode itérative associée à la décomposition

$$M = \frac{1}{\alpha} \text{Id} \quad \text{et} \quad N = \left(\frac{1}{\alpha} \text{Id} - A \right).$$

La méthode du gradient semble encore plus primitive que les méthodes précédentes, mais elle a une interprétation en tant que méthode de minimisation de la fonction $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$ qui lui donne une plus grande applicabilité.

4.4.5 Méthode du gradient conjugué

La méthode du gradient conjugué est la méthode itérative de choix pour résoudre des systèmes linéaires dont la matrice est symétrique réelle définie positive. Il s'agit d'une amélioration spectaculaire de la méthode du gradient (surtout si elle combinée avec un préconditionnement, voir la Définition 4.4.43).

Proposition 4.4.39 Soit A une matrice symétrique définie positive, et $x_0 \in \mathbb{R}^n$. Soit (x_k, r_k, p_k) trois suites définies par les relations de récurrence

$$p_0 = r_0 = b - Ax_0, \quad \text{et pour } 0 \leq k \quad \begin{cases} x_{k+1} = x_k + \alpha_k p_k \\ r_{k+1} = r_k - \alpha_k A p_k \\ p_{k+1} = r_{k+1} + \beta_k p_k \end{cases} \quad (4.63)$$

avec

$$\alpha_k = \frac{\|r_k\|^2}{Ap_k \cdot p_k} \text{ et } \beta_k = \frac{\|r_{k+1}\|^2}{\|r_k\|^2}.$$

Alors, la suite (x_k) de la méthode du gradient conjugué converge en moins de n itérations vers la solution du système linéaire $Ax = b$.

On peut montrer que r_k est la suite des résidus, c'est-à-dire $r_k = b - Ax_k$. Par conséquent, dès que $r_k = 0$, l'algorithme a convergé, c'est-à-dire que x_k est la solution du système $Ax = b$. On sait que la convergence est atteinte en moins de n itérations. Cependant dans la pratique, les calculs sur ordinateurs sont toujours sujets à des erreurs d'arrondi, et on ne trouve pas exactement $r_n = 0$. C'est pourquoi, on introduit un "petit" paramètre ϵ (typiquement 10^{-4} ou 10^{-8} selon la précision désirée), et on décide que l'algorithme a convergé dès que

$$\frac{\|r_k\|}{\|r_0\|} \leq \epsilon.$$

Par ailleurs, pour des systèmes de grande taille (pour lesquels n est "grand", de l'ordre de 10^4 à 10^6), la méthode du gradient conjugué est utilisée comme une méthode itérative, c'est-à-dire qu'elle converge, au sens du critère ci-dessus, en un nombre d'itérations bien inférieur à n .

Remarque 4.4.41

1. En général, si on n'a pas d'indications sur la solution, on choisit d'initialiser la méthode du gradient conjugué par $x_0 = 0$. Si on résout une suite de problèmes peu différents les uns des autres, on peut initialiser x_0 par la solution précédente.
2. A chaque itération on n'a besoin de faire qu'un seul produit matrice-vecteur, à savoir Ap_k , car r_k est calculé par la formule de récurrence et non par la relation $r_k = b - Ax_k$.
3. Pour mettre en oeuvre la méthode du gradient conjugué, il n'est pas nécessaire de stocker la matrice A dans un tableau si on sait calculer le produit matrice vecteur Ay pour tout vecteur y .
4. La méthode du gradient conjugué est très efficace et très utilisée. Elle a beaucoup de variantes ou de généralisations, notamment au cas des matrices non symétriques définies positives.

•

La vitesse de convergence de la méthode du gradient conjugué dépend du conditionnement de la matrice A , l'idée du préconditionnement est de pré-multiplier le système linéaire $Ax = b$ par une matrice C^{-1} telle que le conditionnement de $(C^{-1}A)$ soit plus petit que celui de A . En pratique on choisit une matrice C "proche" de A mais plus facile à inverser.

Définition 4.4.43 Soit à résoudre le système linéaire $Ax = b$. On appelle préconditionnement de A , une matrice C (facile à inverser) telle que $\text{cond}_2(C^{-1}A)$ soit

plus petit que $\text{cond}_2(A)$. On appelle système préconditionné le système équivalent $C^{-1}Ax = C^{-1}b$.

La technique du préconditionnement est très efficace et essentielle en pratique pour converger rapidement. Nous indiquons trois choix possibles de C du plus simple au plus compliqué. Le préconditionnement le plus simple est le “préconditionnement diagonal” : il consiste à prendre $C = \text{diag}(A)$. Il est malheureusement peu efficace, et on lui préfère souvent le “préconditionnement SSOR” (pour Symmetric SOR). En notant $D = \text{diag}(A)$ la diagonale d’une matrice symétrique A et $-E$ sa partie strictement inférieure telle que $A = D - E - E^*$, pour $\omega \in]0, 2[$, on pose

$$C_\omega = \frac{\omega}{2 - \omega} \left(\frac{D}{\omega} - E \right) D^{-1} \left(\frac{D}{\omega} - E^* \right).$$

On vérifie que, si A est définie positive, alors C l’est aussi. Le système $Cz = r$ est facile à résoudre car C est déjà sous une forme factorisée en produit de matrices triangulaires. Le nom de ce préconditionnement vient du fait qu’inverser C revient à effectuer deux itérations successives de la méthode itérative de relaxation (SOR), avec deux matrices d’itérations symétriques l’une de l’autre.

Un dernier exemple est le “préconditionnement de Cholesky incomplet”. La matrice C est cherchée sous la forme BB^* où B est le facteur “incomplet” de la factorisation de Cholesky de A (voir la Proposition 4.4.19). Cette matrice triangulaire inférieure B est obtenue en appliquant l’algorithme de factorisation de Cholesky à A en forçant l’égalité $b_{ij} = 0$ si $a_{ij} = 0$. Cette modification de l’algorithme assure, d’une part que le facteur B sera aussi creux que la matrice A , et d’autre part que le calcul de ce facteur incomplet sera beaucoup moins cher (en temps de calcul) que le calcul du facteur exact si A est creuse (ce qui est le cas pour des matrices de discrétisation par éléments finis). Le préconditionnement de Cholesky incomplet est souvent le préconditionnement le plus efficace en pratique.

Chapitre 5

PROBLÈMES AUX VALEURS PROPRES

5.1 Motivation et exemples

5.1.1 Introduction

Ce chapitre est consacré à la théorie spectrale des équations aux dérivées partielles, c'est-à-dire à l'étude des valeurs propres et des fonctions propres de ces équations. La motivation de cette étude est double. D'une part, cela va nous permettre d'étudier des solutions particulières, dites oscillantes en temps (ou vibrantes), des problèmes d'évolution associés à ces équations. D'autre part, cela permet de déduire une méthode de résolution générale de ces mêmes problèmes d'évolution qui dépasse l'objet de ce cours.

Donnons tout de suite un exemple de **problème aux valeurs propres** pour le Laplacien avec condition aux limites de Dirichlet. Si Ω est un ouvert borné de \mathbb{R}^N on cherche les couples $(\lambda, u) \in \mathbb{R} \times H_0^1(\Omega)$, avec $u \neq 0$, solutions de

$$\begin{cases} -\Delta u = \lambda u & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega. \end{cases} \quad (5.1)$$

Le réel λ est appelé **valeur propre**, et la fonction $u(x)$ **mode propre ou fonction propre**. L'ensemble des valeurs propres est appelé le spectre de (5.1). On peut faire l'analogie entre (5.1) et le problème plus simple de détermination des valeurs et vecteurs propres d'une matrice A d'ordre n ,

$$Au = \lambda u \quad \text{avec} \quad (\lambda, u) \in \mathbb{R} \times \mathbb{R}^n, \quad (5.2)$$

en affirmant que l'opérateur $-\Delta$ est une "généralisation" en dimension infinie d'une matrice A en dimension finie. La résolution de (5.1) sera utile pour résoudre les problèmes d'évolution, de type parabolique ou hyperbolique, associés au Laplacien, c'est-à-dire l'équation de la chaleur (5.5) ou l'équation des ondes (5.7). Néanmoins, les solutions de (5.1) ont aussi une interprétation physique qui leur est propre, par exemple comme modes propres de vibration.

Le plan de ce chapitre est le suivant. Après avoir motivé plus amplement le problème aux valeurs propres (5.1), nous développons dans la Section 5.3 une **théorie spectrale abstraite** dans les espaces de Hilbert. Le but de cette section est de généraliser en dimension infinie le résultat bien connu en dimension finie qui affirme que toute matrice symétrique réelle est diagonalisable dans une base orthonormée. Cette section relève en partie d'un cours de mathématiques "pures", aussi nous insistons à nouveau sur le fait que c'est l'esprit des résultats plus que la lettre des démonstrations qui importe ici. Nous appliquons cette théorie spectrale aux équations aux dérivées partielles elliptiques dans la Section 5.3. En particulier, nous démontrons que le problème spectral (5.1) **admet une infinité dénombrable de solutions**. Enfin, la Section 5.4 est consacrée aux questions d'**approximation numérique** des valeurs propres et fonctions propres d'une équation aux dérivées partielles. En particulier, nous introduisons la notion de **matrice de masse** \mathcal{M} qui vient compléter celle de matrice de rigidité \mathcal{K} , et nous montrons que des valeurs propres approchées de (5.1) se calculent comme les valeurs propres du système $\mathcal{K}u = \lambda\mathcal{M}u$, ce qui confirme l'analogie entre (5.1) et sa version discrète (5.2).

5.1.2 Résolution des problèmes instationnaires

Avant de nous lancer dans les développements abstraits de la prochaine section, montrons en quoi la résolution d'un problème aux valeurs propres permet de résoudre aussi un problème d'évolution. Pour cela nous allons faire une analogie avec la résolution de systèmes différentiels en dimension finie. Dans tout ce qui suit A désigne une matrice symétrique réelle, définie positive, d'ordre n . On note λ_k ses valeurs propres et r_k ses vecteurs propres, $1 \leq k \leq n$, tels que $Ar_k = \lambda_k r_k$.

On commence par un système différentiel du premier ordre

$$\begin{cases} \frac{\partial u}{\partial t} + Au = 0 & \text{pour } t \geq 0 \\ u(t=0) = u_0, \end{cases} \quad (5.3)$$

où $u(t)$ est une fonction de classe C^1 de \mathbb{R}^+ dans \mathbb{R}^n , et $u_0 \in \mathbb{R}^n$. Il est bien connu que (5.3) admet une solution unique obtenue en diagonalisant la matrice A . Plus précisément, la donnée initiale se décompose sous la forme $u_0 = \sum_{k=1}^n u_k^0 r_k$, ce qui donne

$$u(t) = \sum_{k=1}^n u_k^0 e^{-\lambda_k t} r_k.$$

Un deuxième exemple est le système différentiel du deuxième ordre

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} + Au = 0 & \text{pour } t \geq 0 \\ u(t=0) = u_0, \\ \frac{\partial u}{\partial t}(t=0) = u_1, \end{cases} \quad (5.4)$$

où $u(t)$ est une fonction de classe C^2 de \mathbb{R}^+ dans \mathbb{R}^n , et $u_0, u_1 \in \mathbb{R}^n$. En décomposant les données initiales sous la forme $u_0 = \sum_{k=1}^n u_k^0 r_k$ et $u_1 = \sum_{k=1}^n u_k^1 r_k$, (5.4) admet

comme solution unique

$$u(t) = \sum_{k=1}^n \left(u_k^0 \cos(\sqrt{\lambda_k} t) + \frac{u_k^1}{\sqrt{\lambda_k}} \sin(\sqrt{\lambda_k} t) \right) r_k.$$

Il est clair sur ces deux exemples que la connaissance du spectre de la matrice A permet de résoudre les problèmes d'évolution (5.3) et (5.4). Aussi évident soient-ils, ces exemples sont tout à fait représentatifs de la démarche que nous allons suivre dans la suite. Nous allons remplacer la matrice A par l'opérateur $-\Delta$, l'espace \mathbb{R}^n par l'espace de Hilbert $L^2(\Omega)$, et nous allons "diagonaliser" le Laplacien pour résoudre l'équation de la chaleur ou l'équation des ondes.

Afin de se convaincre que (5.1) est bien la "bonne" formulation du problème aux valeurs propres pour le Laplacien, on peut passer par un argument de "séparation des variables" dans l'équation de la chaleur ou l'équation des ondes que nous décrivons formellement. En l'absence de terme source, et en "oubliant" (provisoirement) la condition initiale et les conditions aux limites, nous cherchons une solution \mathbf{u} de ces équations qui s'écrive sous la forme

$$\mathbf{u}(x, t) = \phi(t)u(x),$$

c'est-à-dire que l'on sépare les variables de temps et d'espace. Si \mathbf{u} est solution de l'équation de la chaleur

$$\frac{\partial \mathbf{u}}{\partial t} - \Delta \mathbf{u} = 0, \quad (5.5)$$

on trouve (au moins formellement) que

$$\frac{\phi'(t)}{\phi(t)} = \frac{\Delta u(x)}{u(x)} = -\lambda$$

où $\lambda \in \mathbb{R}$ est une constante indépendante de t et de x . On en déduit que $\phi(t) = e^{-\lambda t}$ et que u doit être solution du problème aux valeurs propres

$$-\Delta u = \lambda u \quad (5.6)$$

muni de conditions aux limites adéquates.

De la même manière, si \mathbf{u} est solution de l'équation des ondes

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} - \Delta \mathbf{u} = 0, \quad (5.7)$$

on trouve que

$$\frac{\phi''(t)}{\phi(t)} = \frac{\Delta u(x)}{u(x)} = -\lambda$$

où $\lambda \in \mathbb{R}$ est une constante. Cette fois-ci on en déduit que, si $\lambda > 0$ (ce qui sera effectivement le cas), alors $\phi(t) = a \cos(\sqrt{\lambda} t) + b \sin(\sqrt{\lambda} t)$ et que u doit encore être solution de (5.6). Remarquons que, si le comportement en espace de la solution \mathbf{u} est le même pour l'équation de la chaleur et pour l'équation des ondes, il n'en est pas de même pour son comportement en temps : elle oscille en temps pour les ondes alors qu'elle décroît exponentiellement en temps (car $\lambda > 0$) pour la chaleur.

Exercice 5.1.1 Soit $\Omega = \mathbb{R}^N$. Montrer que $u(x) = \exp(ik \cdot x)$ est une solution de (5.6) si $|k|^2 = \lambda$. Une telle solution est appelée onde plane.

Exercice 5.1.2 Soit un potentiel régulier $V(x)$. Montrer que, si $\mathbf{u}(x, t) = e^{-i\omega t}u(x)$ est solution de

$$i \frac{\partial \mathbf{u}}{\partial t} + \Delta \mathbf{u} - V \mathbf{u} = 0 \quad \text{dans } \mathbb{R}^N \times \mathbb{R}_*^+, \quad (5.8)$$

alors $u(x)$ est solution de

$$-\Delta u + Vu = \omega u \quad \text{dans } \mathbb{R}^N. \quad (5.9)$$

On retrouve le même type de problème spectral que (5.6), à l'addition d'un terme d'ordre zéro près. Pour l'équation de Schrödinger la valeur propre ω s'interprète comme une énergie. La plus petite valeur possible de cette énergie correspond à l'énergie de l'état fondamental du système décrit par (5.8). Les autres valeurs, plus grandes, donnent les énergies des états excités. Sous des conditions "raisonnables" sur le potentiel V , ces niveaux d'énergie sont discrets en nombre infini dénombrable (ce qui est cohérent avec la vision physique des *quanta*).

Exercice 5.1.3 Soit $V(x) = Ax \cdot x$ avec A matrice symétrique réelle définie positive. Montrer que $u(x) = \exp(-A^{1/2}x \cdot x/2)$ est une solution de (5.9) si $\omega = \text{tr}(A^{1/2})$. Une telle solution est appelée état fondamental.

5.3 Valeurs propres d'un problème elliptique

5.3.1 Problème variationnel

Nous revenons au cadre variationnel introduit au Chapitre 1. L'intérêt de ce cadre assez général est qu'il s'appliquera à de nombreux modèles différents. Dans un espace de Hilbert V nous considérons une forme bilinéaire $a(\cdot, \cdot)$, **symétrique**, continue et coercive, c'est-à-dire que $a(w, v) = a(v, w)$, et il existe $M > 0$ et $\nu > 0$ tels que

$$|a(w, v)| \leq M \|w\|_V \|v\|_V \quad \text{pour tout } w, v \in V$$

et

$$a(v, v) \geq \nu \|v\|_V^2 \quad \text{pour tout } v \in V.$$

Remarquons que l'hypothèse de symétrie de la forme bilinéaire $a(\cdot, \cdot)$ est nouvelle puisqu'elle n'est pas nécessaire dans le cadre du théorème de Lax-Milgram. De plus, nous introduisons un nouvel ingrédient, à savoir un autre espace de Hilbert H sur lequel nous faisons l'hypothèse fondamentale suivante

$$\left\{ \begin{array}{l} V \subset H \text{ avec injection compacte} \\ V \text{ est dense dans } H. \end{array} \right. \quad (5.11)$$

L'expression "injection compacte" veut dire que de toute suite bornée de V on peut extraire une sous-suite convergente dans H . Les espaces H et V ne partagent pas

le même produit scalaire, et nous les noterons $\langle \cdot, \cdot \rangle_H$ et $\langle \cdot, \cdot \rangle_V$ pour éviter toute confusion.

Nous considérons le problème variationnel de valeurs propres suivant (ou problème spectral) : trouver $\lambda \in \mathbb{R}$ et $u \in V \setminus \{0\}$ tels que

$$a(u, v) = \lambda \langle u, v \rangle_H \quad \forall v \in V. \quad (5.12)$$

On dira que λ est une valeur propre du problème variationnel (5.12) (ou de la forme bilinéaire a) et que u est le vecteur propre associé.

Les solutions de (5.12) sont données par le résultat suivant que nous admettrons.

Théorème 5.3.2 *Soit V et H deux espaces de Hilbert réels de dimension infinie. On suppose que $V \subset H$ avec injection compacte et que V est dense dans H . Soit $a(\cdot, \cdot)$ une forme bilinéaire symétrique continue et coercive sur V . Alors les valeurs propres de (5.12) forment une suite croissante $(\lambda_k)_{k \geq 1}$ de réels positifs qui tend vers l'infini, et il existe une base hilbertienne de H $(u_k)_{k \geq 1}$ de vecteurs propres associés, c'est-à-dire que*

$$u_k \in V, \quad \text{et } a(u_k, v) = \lambda_k \langle u_k, v \rangle_H \quad \forall v \in V.$$

De plus, $(u_k / \sqrt{\lambda_k})_{k \geq 1}$ est une base hilbertienne de V pour le produit scalaire $a(\cdot, \cdot)$.

Remarque 5.3.3 Dans le Théorème 5.3.2 on peut remplacer l'hypothèse de coercivité de la forme bilinéaire $a(\cdot, \cdot)$ par l'hypothèse plus faible qu'il existe deux constantes positives $\eta > 0$ et $\nu > 0$ telles que

$$a(v, v) + \eta \|v\|_H^2 \geq \nu \|v\|_V^2 \quad \text{pour tout } v \in V.$$

Dans ce cas les valeurs propres $(\lambda_k)_{k \geq 1}$ ne sont pas forcément positives, mais vérifient seulement $\lambda_k + \eta > 0$. •

Nous donnons au passage une caractérisation très utile des valeurs propres du problème variationnel (5.12), appelée **principe du min-max ou de Courant-Fisher**. Pour cela on introduit le quotient de Rayleigh défini, pour chaque fonction $v \in V \setminus \{0\}$, par

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

Proposition 5.3.4 (Courant-Fisher) *Soit V et H deux espaces de Hilbert réels de dimension infinie. On suppose que $V \subset H$ avec injection compacte et que V est dense dans H . Soit $a(\cdot, \cdot)$ une forme bilinéaire symétrique continue et coercive sur V . Pour $k \geq 0$ on note \mathcal{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 (5.12). Alors, pour tout $k \geq 1$, la k -ème valeur propre est donnée par*

$$\lambda_k = \min_{W \in \mathcal{E}_k} \left(\max_{v \in W \setminus \{0\}} R(v) \right) = \max_{W \in \mathcal{E}_{k-1}} \left(\min_{v \in W^\perp \setminus \{0\}} R(v) \right). \quad (5.15)$$

En particulier, la première valeur propre vérifie

$$\lambda_1 = \min_{v \in V \setminus \{0\}} R(v), \quad (5.16)$$

et tout point de minimum dans (5.16) est un vecteur propre associé à λ_1 .

Démonstration. Soit $(u_k)_{k \geq 1}$ la base hilbertienne de H formée des vecteurs propres de (5.12). D'après le Théorème 5.3.2, $(u_k/\sqrt{\lambda_k})_{k \geq 1}$ est une base hilbertienne de V . On peut donc caractériser les espaces H et V à partir de leur décomposition spectrale

$$H = \left\{ v = \sum_{k=1}^{+\infty} \alpha_k u_k, \quad \|v\|_H^2 = \sum_{k=1}^{+\infty} \alpha_k^2 < +\infty \right\},$$

$$V = \left\{ v = \sum_{k=1}^{+\infty} \alpha_k u_k, \quad \|v\|_V^2 = \sum_{k=1}^{+\infty} \lambda_k \alpha_k^2 < +\infty \right\}.$$

On remarque au passage que, comme les valeurs propres λ_k sont minorées par $\lambda_1 > 0$, cette caractérisation fait bien apparaître V comme un sous-espace de H . On peut alors réécrire le quotient de Rayleigh

$$R(v) = \frac{\sum_{k=1}^{+\infty} \lambda_k \alpha_k^2}{\sum_{k=1}^{+\infty} \alpha_k^2},$$

ce qui démontre immédiatement le résultat pour la première valeur propre. Introduisons le sous-espace $W_k \in \mathcal{E}_k$ engendré par (u_1, u_2, \dots, u_k) . On a

$$R(v) = \frac{\sum_{j=1}^k \lambda_j \alpha_j^2}{\sum_{j=1}^k \alpha_j^2} \quad \forall v \in W_k \quad \text{et} \quad R(v) = \frac{\sum_{j=k}^{+\infty} \lambda_j \alpha_j^2}{\sum_{j=k}^{+\infty} \alpha_j^2} \quad \forall v \in W_{k-1}^\perp,$$

d'où l'on déduit

$$\lambda_k = \max_{v \in W_k \setminus \{0\}} R(v) = \min_{v \in W_{k-1}^\perp \setminus \{0\}} R(v).$$

Soit W un sous-espace quelconque dans \mathcal{E}_k . Comme W est de dimension k et W_{k-1} de dimension $k-1$, l'intersection $W \cap W_{k-1}^\perp$ n'est pas réduite à $\{0\}$. Par conséquent,

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

ce qui prouve la première égalité dans (5.15). De même, si W est un sous-espace dans \mathcal{E}_{k-1} , alors $W^\perp \cap W_k$ n'est pas réduit à $\{0\}$, et

$$\min_{v \in W^\perp \setminus \{0\}} R(v) \leq \min_{v \in W^\perp \cap W_k \setminus \{0\}} R(v) \leq \max_{v \in W^\perp \cap W_k \setminus \{0\}} R(v) \leq \max_{v \in W_k \setminus \{0\}} R(v) = \lambda_k,$$

ce qui prouve la deuxième égalité dans (5.15). Soit maintenant u un point de minimum dans (5.16). Pour $v \in V$, on introduit la fonction $f(t) = R(u + tv)$ d'une variable réelle $t \in \mathbb{R}$ qui admet un minimum en $t = 0$. Par conséquent sa dérivée

s'annule en $t = 0$. En tenant compte de ce que $f(0) = \lambda_1$, un simple calcul montre que

$$f'(0) = 2 \frac{a(u, v) - \lambda_1 \langle u, v \rangle_H}{\|u\|_H^2}.$$

Comme v est quelconque dans V , la condition $f'(0) = 0$ n'est rien d'autre que la formulation variationnelle (5.12), c'est-à-dire que u est un vecteur propre associé à la valeur propre λ_1 . \square

5.3.2 Valeurs propres du Laplacien

On peut immédiatement appliquer le Théorème 5.3.2 à la formulation variationnelle du Laplacien avec conditions aux limites de Dirichlet, ce qui nous donne le résultat suivant.

Théorème 5.3.5 *Soit Ω un ouvert borné régulier de classe C^1 de \mathbb{R}^N . Il existe une suite croissante $(\lambda_k)_{k \geq 1}$ de réels positifs qui tend vers l'infini, et il existe une base hilbertienne de $L^2(\Omega)$ $(u_k)_{k \geq 1}$, telle que chaque u_k appartient à $H_0^1(\Omega)$ et vérifie*

$$\begin{cases} -\Delta u_k = \lambda_k u_k & \text{p.p. dans } \Omega \\ u_k = 0 & \text{p.p. sur } \partial\Omega. \end{cases} \quad (5.17)$$

Démonstration. Pour le Laplacien avec conditions aux limites de Dirichlet, on choisit $V = H_0^1(\Omega)$, $H = L^2(\Omega)$, et la forme bilinéaire symétrique est définie par

$$a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx,$$

et le produit scalaire sur $L^2(\Omega)$ est bien sûr

$$\langle u, v \rangle_H = \int_{\Omega} uv \, dx.$$

On vérifie aisément les hypothèses du Théorème 5.3.2. Grâce au Théorème 2.3.21 de Rellich, V est bien compactement inclus dans H . Comme $C_c^\infty(\Omega)$ est dense à la fois dans H et dans V , V est bien dense dans H . Enfin, on a vu au Chapitre 3 que la forme bilinéaire a est bien continue et coercive sur V . Par conséquent, il existe une suite croissante $(\lambda_k)_{k \geq 1}$ de réels positifs qui tend vers l'infini, et il existe une base hilbertienne de $L^2(\Omega)$ $(u_k)_{k \geq 1}$, tels que $u_k \in H_0^1(\Omega)$ et

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

Par une simple intégration par parties (du même type que celle pratiquée dans la démonstration du Théorème 3.2.2) on obtient (5.17). Remarquons que nous n'utilisons la régularité de Ω que pour pouvoir appliquer le Théorème de trace 2.3.13 et donner un sens "presque partout" à la condition aux limites de Dirichlet. \square

Remarque 5.3.6 L'hypothèse sur le caractère borné de l'ouvert Ω est absolument fondamentale dans le Théorème 5.3.5. Si elle n'est pas satisfaite, le Théorème 2.3.21 de Rellich (sur l'injection compacte de $H^1(\Omega)$ dans $L^2(\Omega)$) est en général faux, et on peut montrer que le Théorème 5.3.5 n'a pas lieu. En fait, il se peut qu'il existe une infinité (non dénombrable) de valeurs propres "généralisées" au sens où les fonctions propres n'appartiennent pas à $L^2(\Omega)$. A la lumière de l'Exercice 5.1.1 on méditera le cas du Laplacien dans $\Omega = \mathbb{R}^N$. •

Exercice 5.3.2 En dimension $N = 1$, on considère $\Omega =]0, 1[$. Calculer explicitement toutes les valeurs propres et les fonctions propres du Laplacien avec conditions aux limites de Dirichlet (5.17). A l'aide de la décomposition spectrale de ce problème, montrer que la série

$$\sum_{k=1}^{+\infty} a_k \sin(k\pi x)$$

converge dans $L^2(0, 1)$ si et seulement si $\sum_{k=1}^{+\infty} a_k^2 < +\infty$, et dans $H^1(0, 1)$ si et seulement si $\sum_{k=1}^{+\infty} k^2 a_k^2 < +\infty$.

Exercice 5.3.3 On considère un parallélépipède $\Omega =]0, L_1[\times]0, L_2[\times \cdots \times]0, L_N[$, où les $(L_i > 0)_{1 \leq i \leq N}$ sont des constantes positives. Calculer explicitement toutes les valeurs propres et les fonctions propres du Laplacien avec conditions aux limites de Dirichlet (5.17).

Exercice 5.3.4 On considère à nouveau un ouvert Ω parallélépipédique comme dans l'Exercice 5.3.3. Calculer explicitement toutes les valeurs propres et les fonctions propres du Laplacien avec conditions aux limites de Neumann sur tout le bord $\partial\Omega$.

On peut montrer que les fonctions propres du Laplacien, avec conditions aux limites de Dirichlet ou de Neumann, sont régulières.

Proposition 5.3.9 Soit Ω un ouvert borné régulier de classe C^∞ . Alors les fonctions propres solutions de (5.17) appartiennent à $C^\infty(\overline{\Omega})$.

Démonstration. Soit u_k la k -ème fonction propre solution dans $H_0^1(\Omega)$ de (5.17). On peut considérer que u_k est solution du problème aux limites suivant

$$\begin{cases} -\Delta u_k = f_k & \text{dans } \Omega \\ u_k = 0 & \text{sur } \partial\Omega, \end{cases}$$

avec $f_k = \lambda_k u_k$. Comme f_k appartient à $H^1(\Omega)$, par application du Théorème 3.2.26 de régularité on en déduit que la solution u_k appartient à $H^3(\Omega)$. Du coup, le second membre f_k est plus régulier ce qui permet d'augmenter encore la régularité de u_k . Par une récurrence facile on montre ainsi que u_k appartient à $H^m(\Omega)$ pour tout $m \geq 1$. En vertu du Théorème 2.3.25 sur la continuité des fonctions de $H^m(\Omega)$ (voir aussi la Remarque 2.3.26), on en déduit que u_k appartient donc $C^\infty(\overline{\Omega})$. □

Nous admettons un résultat qualitatif important à propos de la première valeur propre.

Théorème 5.3.10 (de Krein-Rutman) *On reprend les notations et les hypothèses du Théorème 5.3.5. On suppose que l'ouvert Ω est connexe. Alors la première valeur propre λ_1 est simple (i.e. le sous-espace propre correspondant est de dimension 1) et le premier vecteur propre peut être choisi positif presque partout dans Ω .*

Remarque 5.3.11 Le Théorème 5.3.10 de Krein-Rutman est spécifique au cas des équations “scalaires” (c'est-à-dire que l'inconnue u est à valeurs dans \mathbb{R}). Ce résultat est faux en général si l'inconnue u est à valeurs vectorielles (voir plus loin l'exemple du système de l'élasticité). La raison de cette différence entre le cas scalaire et vectoriel est que ce théorème s'appuie sur le principe du maximum (voir le Théorème 3.2.22) qui n'est valable que dans le cas scalaire. •

Exercice 5.3.6 Soit Ω un ouvert borné régulier et connexe. Montrer que la première valeur propre du Laplacien dans Ω avec condition aux limites de Neumann est nulle et qu'elle est simple.

Remarque 5.3.12 L'ensemble des résultats de cette sous-section se généralise sans difficulté à d'autres conditions aux limites et à des opérateurs elliptiques généraux du second ordre (voir la Sous-section 3.2.3). •

5.3.3 Autres modèles

L'extension des résultats de la sous-section précédente à des équations aux dérivées partielles elliptiques plus compliquées que le Laplacien ne pose pas de problèmes conceptuels nouveaux. Nous décrivons brièvement cette généralisation pour un exemple significatif : le système de l'élasticité linéarisée.

Les équations (3.56) de l'élasticité linéarisée décrivent en fait le régime stationnaire des équations dynamiques suivantes (très semblables à l'équation des ondes)

$$\begin{cases} \rho \frac{\partial^2 u}{\partial t^2} - \operatorname{div} (2\mu e(u) + \lambda \operatorname{tr}(e(u)) \operatorname{Id}) = f & \text{dans } \Omega \times \mathbb{R}_*^+ \\ u = 0 & \text{sur } \partial\Omega \times \mathbb{R}_*^+, \end{cases} \quad (5.18)$$

où $\rho > 0$ est la densité volumique du matériau et $e(u) = (\nabla u + (\nabla u)^t)/2$. Rappelons que les coefficients de Lamé du matériau vérifient $\mu > 0$ et $2\mu + N\lambda > 0$. En l'absence de forces extérieures f (et en ne tenant pas compte d'éventuelles conditions initiales) on peut aussi chercher des solutions oscillantes en temps de (5.18) comme nous l'avons décrit pour l'équation des ondes dans Sous-section 5.1.2. Cela conduit à chercher des solutions (ℓ, u) du problème aux valeurs propres suivant

$$\begin{cases} -\operatorname{div} (2\mu e(u) + \lambda \operatorname{tr}(e(u)) \operatorname{Id}) = \ell u & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega, \end{cases} \quad (5.19)$$

où $\ell = \omega^2$ est le carré de la fréquence de vibration (nous avons changé la notation de la valeur propre pour éviter une confusion avec le coefficient de Lamé λ). En mécanique, la fonction propre u est aussi appelée **mode propre de vibration**.

En suivant la méthode appliquée ci-dessus au Laplacien on peut démontrer le résultat suivant (nous laissons les détails au lecteur en guise d'exercice).

Proposition 5.3.13 Soit Ω un ouvert borné régulier de classe \mathcal{C}^1 de \mathbb{R}^N . Il existe une suite croissante $(\ell_k)_{k \geq 1}$ de réels positifs qui tend vers l'infini, et il existe une base hilbertienne de $L^2(\Omega)^N$ $(u_k)_{k \geq 1}$, telle que chaque u_k appartient à $H_0^1(\Omega)^N$ et vérifie

$$\begin{cases} -\operatorname{div}(2\mu e(u_k) + \lambda \operatorname{tr}(e(u_k)) \operatorname{Id}) = \ell_k u_k & \text{p.p. dans } \Omega \\ u_k = 0 & \text{p.p. sur } \partial\Omega. \end{cases}$$

Rang du mode propre	1	2	3	4
Valeur propre	102.54	102.54	1885.2	2961.2

TABLE 5.1 – Valeurs propres correspondant aux modes propres de la Figure 5.1.

Le résultat de régularité sur les fonctions propres u_k de la Proposition 5.3.9 s'étend aussi facilement au cas de l'élasticité et du problème (5.19). Par contre le Théorème 5.3.10 sur la simplicité de la première valeur propre et la positivité de la première fonction propre est faux en général (comme est faux le principe du maximum). Comme exemple nous calculons par la méthode des éléments finis Q_1 les 4 premiers modes propres d'une "tour" dont la base est fixée (condition aux limites de Dirichlet) et dont les autres parois sont libres (condition aux limites de Neumann). Les 2 premiers modes, dits de battement, correspondent à la même valeur propre (ils sont indépendants mais symétriques par rotation de 90° suivant l'axe z) (voir la Figure 5.1 et le Tableau 5.3.3). La première valeur propre est donc "double".

Exercice 5.3.8 On considère le problème aux valeurs propres pour l'équation de Schrödinger avec un potentiel quadratique $V(x) = Ax \cdot x$ où A est une matrice symétrique définie positive (modèle de l'oscillateur harmonique)

$$-\Delta u + Vu = \lambda u \quad \text{dans } \mathbb{R}^N. \quad (5.21)$$

On définit les espaces $H = L^2(\mathbb{R}^N)$ et

$$V = \{v \in H^1(\mathbb{R}^N) \text{ tel que } |x|v(x) \in L^2(\mathbb{R}^N)\}.$$

Montrer que V est un espace de Hilbert pour le produit scalaire

$$\langle u, v \rangle_V = \int_{\mathbb{R}^N} \nabla u(x) \cdot \nabla v(x) dx + \int_{\mathbb{R}^N} |x|^2 u(x)v(x) dx,$$

et que l'injection de V dans H est compacte. En déduire qu'il existe une suite croissante $(\lambda_k)_{k \geq 1}$ de réels positifs qui tend vers l'infini et une base hilbertienne de $L^2(\mathbb{R}^N)$ $(u_k)_{k \geq 1}$ qui sont les valeurs propres et les fonctions propres de (5.21). Calculer explicitement ses valeurs et fonctions propres (on cherchera u_k sous la forme $p_k(x) \exp(-Ax \cdot x/2)$ où p_k est un polynôme de degré $k - 1$). Interpréter physiquement les résultats.

Exercice 5.3.9 Soit Ω un ouvert borné régulier de \mathbb{R}^N . On considère le problème de vibrations pour l'équation des plaques avec condition aux limites d'encastrement

$$\begin{cases} \Delta(\Delta u) = \lambda u & \text{dans } \Omega \\ \frac{\partial u}{\partial n} = u = 0 & \text{sur } \partial\Omega. \end{cases}$$

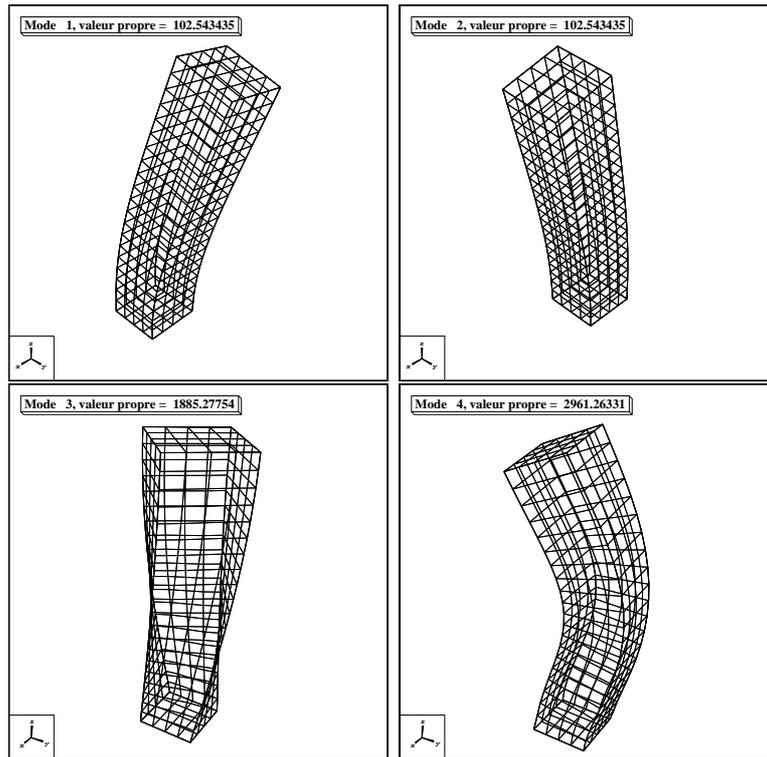


FIGURE 5.1 – Les 4 premiers modes propres d’une “tour” en élasticité.

Montrer qu’il existe une suite croissante $(\lambda_k)_{k \geq 1}$ de valeurs propres positives qui tend vers l’infini et une base hilbertienne dans $L^2(\Omega)$ de fonctions propres $(u_k)_{k \geq 1}$ qui appartiennent à $H_0^2(\Omega)$.

5.4 Méthodes numériques

5.4.1 Discrétisation par éléments finis

On va considérer une approximation interne de la formulation variationnelle introduite à la Sous-section 5.3.1. Étant donné un sous-espace V_h de l’espace de Hilbert V , de dimension finie, on cherche les solutions $(\lambda_h, u_h) \in \mathbb{R} \times V_h$ de

$$a(u_h, v_h) = \lambda_h \langle u_h, v_h \rangle_H \quad \forall v_h \in V_h. \quad (5.22)$$

Typiquement, V_h est un espace d’éléments finis comme ceux introduits par les Définitions 4.3.5 et 4.3.25, et H est l’espace $L^2(\Omega)$. La résolution de l’approximation interne (5.22) est facile comme le montre le lemme suivant.

Lemme 5.4.1 *On se place sous les hypothèses du Théorème 5.3.2. Alors les valeurs propres de (5.22) forment une suite croissante finie*

$$0 < \lambda_1 \leq \dots \leq \lambda_{n_{dl}} \quad \text{avec } n_{dl} = \dim V_h,$$

et il existe une base de V_h , orthonormale dans H , $(u_{k,h})_{1 \leq k \leq n_{dl}}$ de vecteurs propres associés, c'est-à-dire que

$$u_{k,h} \in V_h, \quad \text{et } a(u_{k,h}, v_h) = \lambda_k \langle u_{k,h}, v_h \rangle_H \quad \forall v_h \in V_h.$$

Démonstration. Ce lemme peut être considéré comme une variante évidente du Théorème 5.3.2 (à la différence près qu'en dimension finie il existe un nombre fini de valeurs propres). Néanmoins nous en donnons une démonstration différente, purement algébrique, qui correspond plus à la démarche suivie en pratique. Soit $(\phi_i)_{1 \leq i \leq n_{dl}}$ une base de V_h (par exemple, les fonctions de base d'une méthode d'éléments finis, voir la Proposition 4.3.7). On cherche u_h solution de (5.22) sous la forme

$$u_h(x) = \sum_{i=1}^{n_{dl}} U_i^h \phi_i(x).$$

Introduisant la **matrice de masse** \mathcal{M}_h définie par

$$(\mathcal{M}_h)_{ij} = \langle \phi_i, \phi_j \rangle_H \quad 1 \leq i, j \leq n_{dl},$$

et la **matrice de rigidité** \mathcal{K}_h définie par

$$(\mathcal{K}_h)_{ij} = a(\phi_i, \phi_j) \quad 1 \leq i, j \leq n_{dl},$$

le problème (5.22) est équivalent à trouver $(\lambda_h, U_h) \in \mathbb{R} \times \mathbb{R}^{n_{dl}}$ solution de

$$\mathcal{K}_h U_h = \lambda_h \mathcal{M}_h U_h. \quad (5.23)$$

Les appellations “matrices de masse et de rigidité” proviennent des applications en mécanique des solides. Remarquons que, dans le cas où V_h est un espace d'éléments finis, la matrice de rigidité \mathcal{K}_h est exactement la même matrice que celle rencontrée au Chapitre 4 dans l'application de la méthode des éléments finis aux problèmes elliptiques. On vérifie immédiatement que les matrices \mathcal{M}_h et \mathcal{K}_h sont symétriques et définies positives. Le système (5.23) est un problème matriciel aux valeurs propres “généralisé”. Le théorème de réduction simultanée (voir par exemple le théorème 2.3.6 dans [2]) affirme qu'il existe une matrice inversible P_h telle que

$$\mathcal{M}_h = P_h P_h^*, \quad \text{et } \mathcal{K}_h = P_h \text{diag}(\lambda_k) P_h^*.$$

Par conséquent, les solutions de (5.23) sont les valeurs propres (λ_k) et les vecteurs propres $(U_{k,h})_{1 \leq k \leq n_{dl}}$ qui sont les vecteurs colonnes de l'inverse de P_h^* . Ces vecteurs colonnes forment donc une base, orthogonale pour \mathcal{K}_h et orthonormale pour \mathcal{M}_h (nous indiquerons brièvement à la Remarque 5.4.3 comment calculer cette base). Finalement, les vecteurs $U_{k,h}$ sont simplement les vecteurs des coordonnées dans la base $(\phi_i)_{1 \leq i \leq n_{dl}}$ des fonctions $u_{k,h}$ qui forment une base orthonormale de V_h pour le produit scalaire de H . \square

Remarque 5.4.2 Dans le Lemme 5.4.1 on a repris les hypothèses du Théorème 5.3.2 : en particulier, la forme bilinéaire $a(u, v)$ est supposée **symétrique**. On voit bien l'importance de cette hypothèse dans la démonstration. En effet, si elle n'était pas symétrique, on ne saurait pas si le système (5.23) est diagonalisable, c'est-à-dire s'il existe des solutions du problème aux valeurs propres (5.22). \bullet

L'application du Lemme 5.4.1 à l'approximation variationnelle par éléments finis du problème de Dirichlet (5.17) est immédiate. On prend $V = H_0^1(\Omega)$, $H = L^2(\Omega)$, et l'espace discret V_{0h} de la Définition 4.3.5 (rappelons que V_{0h} contient la condition aux limites de Dirichlet).

Exercice 5.4.1 On considère le problème aux valeurs propres en dimension $N = 1$

$$\begin{cases} -u_k'' = \lambda_k u_k & \text{pour } 0 < x < 1 \\ u_k(0) = u_k(1) = 0. \end{cases}$$

On se propose de calculer la matrice de masse pour la méthode des éléments finis P_1 . On reprend les notations de la Section 4.2. Montrer que la matrice de masse \mathcal{M}_h est donnée par

$$\mathcal{M}_h = h \begin{pmatrix} 2/3 & 1/6 & & & 0 \\ 1/6 & 2/3 & 1/6 & & \\ & \ddots & \ddots & \ddots & \\ & & 1/6 & 2/3 & 1/6 \\ 0 & & & 1/6 & 2/3 \end{pmatrix},$$

et que ses valeurs propres sont

$$\lambda_k(\mathcal{M}_h) = \frac{h}{3} (2 + \cos(k\pi h)) \quad \text{pour } 1 \leq k \leq n.$$

Montrer que, si on utilise la formule de quadrature (4.45), alors on trouve que $\mathcal{M}_h = h \text{Id}$. Dans ce dernier cas, calculer les valeurs propres du problème spectral discret.

Remarque 5.4.3 Pour calculer les valeurs et vecteurs propres du problème spectral matriciel (5.23) il faut, en général, commencer par calculer la factorisation de Cholesky de la matrice de masse $\mathcal{M}_h = \mathcal{L}_h \mathcal{L}_h^*$, pour se ramener au cas classique

$$\tilde{\mathcal{K}}_h \tilde{U}_h = \lambda_h \tilde{U}_h \quad \text{avec } \tilde{\mathcal{K}}_h = \mathcal{L}_h^{-1} \mathcal{K}_h (\mathcal{L}_h^*)^{-1} \text{ et } \tilde{U}_h = \mathcal{L}_h^* U_h,$$

pour lequel on dispose d'algorithmes de calcul de valeurs et vecteurs propres. Nous renvoyons à la Section 13.2 pour plus de détails sur ces algorithmes : disons seulement que c'est l'étape la plus coûteuse en temps de calcul.

On peut éviter de construire la matrice $\tilde{\mathcal{K}}_h$ et faire l'économie de la factorisation de Cholesky de \mathcal{M}_h si on utilise une formule de quadrature pour évaluer les coefficients de la matrice \mathcal{M}_h qui la rende **diagonale**. Ce procédé d'intégration numérique est appelé **condensation de masse** (ou "mass lumping" en anglais) et est fréquemment utilisé. Par exemple, si on utilise la formule de quadrature (4.45) (qui utilise uniquement les valeurs aux noeuds d'une fonction pour calculer une intégrale), on voit facilement que la matrice de masse \mathcal{M}_h ainsi obtenue est diagonale (voir l'Exercice 5.4.1). •

Une analyse numérique précise de la méthode des éléments finis montre que **seules les premières valeurs propres** discrètes $\lambda_{k,h}$ (les plus petites) sont des

approximations correctes des valeurs propres exactes λ_k (même chose pour les vecteurs propres). Il faut donc faire attention au fait que les dernières valeurs propres (les plus grandes) du problème discret (5.23) n'ont aucune signification physique! Par conséquent, si on est intéressé par la millièmes valeur propre du problème de Dirichlet (5.17), il faut prendre un maillage suffisamment fin du domaine Ω afin que la dimension de l'espace d'éléments finis V_h soit bien plus grande que mille.

13.2 Calcul de valeurs et vecteurs propres

Dans cette section nous expliquons comment calculer les valeurs propres et les vecteurs propres d'une matrice symétrique réelle. Pour plus de détails nous renvoyons aux ouvrages [2] et [9].

Puisque les valeurs propres d'une matrice A sont les racines de son polynôme caractéristique $\det(A - \lambda \text{Id})$, on pourrait penser naïvement que, pour les calculer, il "suffit" de factoriser son polynôme caractéristique. Il n'en est rien : on sait depuis Galois et Abel qu'on ne peut pas calculer par opérations élémentaires (addition, multiplication, extraction de racines) les racines d'un polynôme quelconque de degré supérieur ou égal à 5. Pour s'en convaincre, on peut remarquer que n'importe quel polynôme de degré n ,

$$P(\lambda) = (-1)^n (\lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \cdots + a_{n-1} \lambda + a_n),$$

est le polynôme caractéristique (le développer par rapport à la dernière colonne) de la matrice

$$A = \begin{pmatrix} -a_1 & -a_2 & \cdots & \cdots & -a_n \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

Par conséquent, il ne peut pas exister de méthodes directes (c'est-à-dire qui donnent le résultat en un nombre fini d'opérations) pour le calcul des valeurs propres! Il n'existe donc que des méthodes itératives pour calculer des valeurs propres (et des vecteurs propres). Il se trouve que le calcul pratique des valeurs et vecteurs propres d'une matrice est une tâche beaucoup plus difficile que la résolution d'un système linéaire. Fort heureusement, le cas des matrices symétriques réelles (auquel nous nous limitons puisqu'il suffit pour nos applications) est bien plus simple que le cas des matrices non auto-adjointes.

13.2.1 Méthode de la puissance

Une méthode très simple pour calculer la plus grande ou la plus petite valeur propre (en module) d'une matrice (et un vecteur propre associé) est la méthode de la puissance. Une limitation de la méthode est que la valeur propre extrême que l'on calcule doit être simple (ou de multiplicité égale à 1, c'est-à-dire que la dimension

du sous-espace propre correspondant est 1). Soit A une matrice symétrique réelle d'ordre n , de valeurs propres $(\lambda_1, \dots, \lambda_n)$ avec $\lambda_n > |\lambda_i|$ pour tout $1 \leq i \leq n-1$. La méthode de la puissance pour calculer la plus grande valeur propre λ_n est définie par l'algorithme ci-dessous.

1. Initialisation: $x_0 \in \mathbb{R}^n$ tel que $\|x_0\| = 1$.
2. Itérations: pour $k \geq 1$
 1. $y_k = Ax_{k-1}$
 2. $x_k = y_k / \|y_k\|$
 3. test de convergence: si $\|x_k - x_{k-1}\| \leq \varepsilon$, on arrête.

Dans le test de convergence ε est un petit nombre réel, typiquement égal à 10^{-6} . Si $\delta_k = x_k - x_{k-1}$ est petit, alors x_k est un vecteur propre approché de A de valeur propre approchée $\|y_k\|$ car $Ax_k - \|y_k\|x_k = A\delta_k$.

Proposition 13.2.1 *On suppose que la matrice A est symétrique réelle, de valeurs propres $(\lambda_1, \dots, \lambda_n)$, associées à une base orthonormée de vecteurs propres (e_1, \dots, e_n) , et que la valeur propre de plus grand module λ_n est simple et positive, c'est-à-dire que $|\lambda_1|, \dots, |\lambda_{n-1}| < \lambda_n$. On suppose aussi que le vecteur initial x_0 n'est pas orthogonal à e_n . Alors la méthode de la puissance converge, c'est-à-dire que*

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

La vitesse de convergence est proportionnelle au rapport $|\lambda_{n-1}|/|\lambda_n|$

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

Démonstration. Soit $x_0 = \sum_{i=1}^n \beta_i e_i$ le vecteur initial, avec $\beta_n \neq 0$. Le vecteur x_k est proportionnel à $A^k x_0 = \sum_{i=1}^n \beta_i \lambda_i^k e_i$, d'où il vient

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

Comme $|\lambda_i| < \lambda_n$ on en déduit que x_k converge vers $\text{sign}(\beta_n)e_n$. De même, on a

$$\|y_{k+1}\| = \lambda_n \frac{\left(\beta_n^2 + \sum_{i=1}^{n-1} \beta_i^2 \left(\frac{\lambda_i}{\lambda_n} \right)^{2(k+1)} \right)^{1/2}}{\left(\beta_n^2 + \sum_{i=1}^{n-1} \beta_i^2 \left(\frac{\lambda_i}{\lambda_n} \right)^{2k} \right)^{1/2}},$$

qui converge vers λ_n . □

En pratique (et notamment pour le calcul des valeurs propres de la discrétisation d'un problème aux limites elliptiques), on est surtout intéressé par la **plus**

petite valeur propre, en module, de A . On peut adapter les idées précédentes, ce qui donne la méthode de la puissance inverse dont l'algorithme est écrit ci-dessous. On considère une matrice symétrique réelle A dont la plus petite valeur propre en module est simple et strictement positive $0 < \lambda_1 < |\lambda_i|$ pour tout $2 \leq i \leq n$.

1. Initialisation: $x_0 \in \mathbb{R}^n$ tel que $\|x_0\| = 1$.
2. Itérations: pour $k \geq 1$
 1. résoudre $Ay_k = x_{k-1}$
 2. $x_k = y_k / \|y_k\|$
 3. test de convergence: si $\|x_k - x_{k-1}\| \leq \varepsilon$, on arrête.

Si $\delta_k = x_k - x_{k-1}$ est petit, alors x_{k-1} est un vecteur propre approché de valeur propre approchée $1/\|y_k\|$ car $Ax_{k-1} - \frac{x_{k-1}}{\|y_k\|} = -A\delta_k$.

Proposition 13.2.3 *On suppose que la matrice A est symétrique réelle, de valeurs propres $(\lambda_1, \dots, \lambda_n)$, associées à une base orthonormée de vecteurs propres (e_1, \dots, e_n) , et que la valeur propre de plus petit module λ_1 est simple et strictement positive, c'est-à-dire que $0 < \lambda_1 < |\lambda_2|, \dots, |\lambda_n|$. On suppose aussi que le vecteur initial x_0 n'est pas orthogonal à e_1 . Alors la méthode de la puissance inverse converge, c'est-à-dire que*

$$\lim_{k \rightarrow +\infty} \frac{1}{\|y_k\|} = |\lambda_1|, \quad \lim_{k \rightarrow +\infty} x_k = x_\infty \text{ avec } x_\infty = \pm e_1.$$

La vitesse de convergence est proportionnelle au rapport $\lambda_1/|\lambda_2|$

$$\left| \|y_k\|^{-1} - \lambda_1 \right| \leq C \left| \frac{\lambda_1}{\lambda_2} \right|^{2k}, \quad \|x_k - x_\infty\| \leq C \left| \frac{\lambda_1}{\lambda_2} \right|^k.$$

La démonstration est similaire à celle de la Proposition 13.2.1 et nous la laissons au lecteur en guise d'exercice.

Remarque 13.2.4 Pour accélérer la convergence, on peut toujours procéder à une translation de la matrice A qu'on remplace par $A - \sigma \text{Id}$ avec σ une approximation de λ_1 . •

13.2.2 Méthode de Givens-Householder

La méthode de Givens-Householder se compose de deux étapes successives : tout d'abord l'algorithme de Householder qui réduit une matrice symétrique A en une matrice tridiagonale (cette étape s'effectue en un nombre fini d'opérations), ensuite l'algorithme de bisection de Givens qui fournit (de manière itérative) les valeurs propres d'une matrice tridiagonale.

Lemme 13.2.5 (de Householder) *Soit A une matrice symétrique réelle d'ordre n . Il existe $(n - 2)$ matrices orthogonales H_k telles que*

$$T = (H_1 H_2 \cdots H_{n-2})^* A (H_1 H_2 \cdots H_{n-2})$$

soit tridiagonale. Bien sûr, A et T ont les mêmes valeurs propres.

Démonstration. A partir de A , on construit une suite de matrices $(A_k)_{1 \leq k \leq n-1}$ telle que $A_1 = A$ et $A_{k+1} = H_k^* A_k H_k$ avec H_k une matrice orthogonale choisie de telle manière que A_k ait la structure par blocs suivante

$$A_k = \begin{pmatrix} T_k & E_k^* \\ E_k & M_k \end{pmatrix}$$

où T_k est une matrice carrée tridiagonale de taille k , M_k est une matrice carrée de taille $n - k$, et E_k est une matrice rectangulaire à $(n - k)$ lignes et k colonnes dont seule la dernière colonne, notée $a_k \in \mathbb{R}^{n-k}$ est non nulle. Il est clair qu'ainsi A_{n-1} sera sous forme tridiagonale. On remarque que A est bien sous cette forme pour $k = 1$. Soit la matrice H_k définie par

$$H_k = \begin{pmatrix} \text{Id}_k & 0 \\ 0 & \tilde{H}_k \end{pmatrix},$$

avec Id_k la matrice identité d'ordre k et \tilde{H}_k la matrice, dite de Householder, d'ordre $n - k$ définie par

$$\tilde{H}_k = \text{Id}_{n-k} - 2 \frac{v_k (v_k)^*}{\|v_k\|^2}, \quad \text{avec } v_k = a_k + \|a_k\| e_1, \quad (13.13)$$

où e_1 est le premier vecteur de la base canonique de \mathbb{R}^{n-k} . Remarquons que $\tilde{H}_k a_k = -\|a_k\| e_1$, et que H_k est orthogonale et symétrique. Notons que \tilde{H}_k n'est bien définie que si $v_k \neq 0$, mais si ça n'est pas le cas alors la k -ème colonne de A_k est déjà du type désiré, et il suffit de prendre $H_k = \text{Id}_n$. Un simple calcul montre que $A_{k+1} = H_k^* A_k H_k$ se met bien sous la forme désirée. \square

Étudions maintenant l'algorithme de bisection de Givens pour une matrice tridiagonale symétrique réelle dont on suppose, sans perte de généralité, que les coefficients extra-diagonaux sont non-nuls. Nous admettons les deux lemmes suivants.

Lemme 13.2.6 *Pour $1 \leq i \leq n$, on définit une matrice A_i de taille i par*

$$A_i = \begin{pmatrix} b_1 & c_1 & & 0 \\ c_1 & \ddots & \ddots & \\ & \ddots & \ddots & c_{i-1} \\ 0 & & c_{i-1} & b_i \end{pmatrix}.$$

Soit $p_i(\lambda) = \det(A_i - \lambda I)$ son polynôme caractéristique. La suite p_i vérifie la formule de récurrence

$$p_i(\lambda) = (b_i - \lambda)p_{i-1}(\lambda) - c_{i-1}^2 p_{i-2}(\lambda) \quad \forall i \geq 2,$$

avec $p_1(\lambda) = b_1 - \lambda$ et $p_0(\lambda) = 1$. De plus, pour tout $i \geq 1$, le polynôme p_i possède les propriétés suivantes

1. $\lim_{\lambda \rightarrow -\infty} p_i(\lambda) = +\infty$,
2. si $p_i(\lambda_0) = 0$, alors $p_{i-1}(\lambda_0)p_{i+1}(\lambda_0) < 0$,

3. p_i admet i racines réelles distinctes qui séparent strictement les $(i+1)$ racines de p_{i+1} .

Lemme 13.2.7 Pour tout $\mu \in \mathbb{R}$, on définit

$$\operatorname{sgn}p_i(\mu) = \begin{cases} \text{signe de } p_i(\mu) & \text{si } p_i(\mu) \neq 0, \\ \text{signe de } p_{i-1}(\mu) & \text{si } p_i(\mu) = 0. \end{cases}$$

Soit $N(i, \mu)$ le nombre de changements de signe entre éléments consécutifs de la famille ordonnée $E(i, \mu) = \{+1, \operatorname{sgn}p_1(\mu), \operatorname{sgn}p_2(\mu), \dots, \operatorname{sgn}p_i(\mu)\}$. Alors, $N(i, \mu)$ est le nombre de racines de p_i qui sont strictement inférieures à μ .

Algorithme pratique de Givens. On note $\lambda_1 \leq \dots \leq \lambda_n$ les valeurs propres de A rangées par ordre croissant. Pour calculer numériquement la i -ème valeur propre λ_i , on prend un intervalle $[a_0, b_0]$ dans lequel on est sûr que λ_i se trouve (par exemple $-a_0 = b_0 = \|A\|_2$). On calcule alors le nombre $N(n, \frac{a_0+b_0}{2})$ défini dans le Lemme 13.2.7 (les valeurs de la suite $p_j(\frac{a_0+b_0}{2})$, pour $1 \leq j \leq n$, sont calculées par la formule de récurrence du Lemme 13.2.6). S'il se trouve que $N(n, \frac{a_0+b_0}{2}) \geq i$, alors on en déduit que λ_i appartient à l'intervalle $[a_0, \frac{a_0+b_0}{2}[$. Si au contraire on trouve que $N(n, \frac{a_0+b_0}{2}) < i$, alors λ_i appartient à l'autre intervalle $[\frac{a_0+b_0}{2}, b_0]$. Dans tous les cas on a divisé par deux l'intervalle initial qui contient λ_i . Par dichotomie, c'est-à-dire en répétant cette procédure de division de l'intervalle contenant λ_i , on approche la valeur exacte de λ_i avec la précision désirée.

Bibliographie

- [1] ALLAIRE G., *Analyse numérique et optimisation*, Editions de l'Ecole Polytechnique, Palaiseau (2005).
- [2] ALLAIRE G., KABER S. M., *Algèbre linéaire numérique. Cours et exercices*, Éditions Ellipses, Paris (2002).
- [3] BERNARDI Ch., MADAY Y., RAPETTI F., *Discrétisation variationnelle de problèmes aux limites elliptiques*, Mathématiques et Applications 45, Springer, Paris (2004).
- [4] BONNANS J., *Optimisation continue*, Mathématiques appliquées pour le Master / SMAI, Dunod, Paris (2006).
- [5] BONNANS J., GILBERT J.-C., LEMARECHAL C., SAGASTIZABAL C., *Optimisation numérique*, Mathématiques et Applications 27, Springer, Paris (1997).
- [6] BONY J.-M., *Cours d'analyse. Théorie des distributions et analyse de Fourier*, Éditions de l'École Polytechnique, Palaiseau (2001).
- [7] BREZIS H., *Analyse fonctionnelle*, Masson, Paris (1983).
- [8] CHVÁTAL V., *Linear programming*, Freeman and Co., New York (1983).
- [9] CIARLET P.G., *Introduction à l'analyse numérique matricielle et à l'optimisation*, Masson, Paris (1982).
- [10] CIARLET P.G., *The finite element methods for elliptic problems*, North-Holland, Amsterdam (1978).
- [11] CIARLET P.G., LIONS J.-L., *Handbook of numerical analysis*, North-Holland, Amsterdam (1990).
- [12] CULIOLI J.-C., *Introduction à l'optimisation*, Éditions Ellipses, Paris (1994).
- [13] DANAILA I., HECHT F., PIRONNEAU O., *Simulation numérique en C++*, Dunod, Paris (2003).
- [14] DANAILA I., JOLY P., KABER S. M., POSTEL M., *Introduction au calcul scientifique par la pratique*, Dunod, Paris (2005).
- [15] DAUTRAY R., LIONS J.-L., *Analyse mathématique et calcul numérique pour les sciences et les techniques*, Masson, Paris (1988).
- [16] ERN A., GUERMOND J.-L., *Éléments finis : théorie, applications, mise en oeuvre*, Mathématiques et Applications 36, Springer, Paris (2002).

- [17] GEORGE P.L., *Génération automatique de maillages, application aux méthodes d'éléments finis*, Masson, Paris (1991).
- [18] GIRAULT V., RAVIART P.-A., *Finite element methods for Navier-Stokes equations : theory and algorithms*, Springer, Berlin (1986).
- [19] LEMAITRE J., CHABOCHE J.-L., *Mécanique des matériaux solides*, Dunod, Paris (1985).
- [20] LIONS J.-L., MAGENES E., *Problèmes aux limites non homogènes et applications*, Dunod, Paris (1970).
- [21] LUCQUIN B., PIRONNEAU O., *Introduction au calcul scientifique*, Masson, Paris (1996).
- [22] MOHAMMADI B., SAÏAC J.-H., *Pratique de la simulation numérique*, Dunod, Paris (2003).
- [23] PIRONNEAU O., *Méthodes des éléments finis pour les fluides*, Masson, Paris (1988).
- [24] RAVIART P.-A., THOMAS J.-M., *Introduction à l'analyse numérique des équations aux dérivées partielles*, Masson, Paris (1983).
- [25] SALENÇON J., *Mécanique des milieux continus*, Éditions de l'École Polytechnique, Palaiseau (2002).
- [26] TEMAM R., *Navier-Stokes equations. Theory and numerical analysis*, AMS Chelsea Publishing, Providence, RI (2001).