REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE AMAR TELIDJI LAGHOUAT

FACULTE DES SCIENCES

DEPARTEMENT DE MATHEMATIQUES ET INFORMATIQUE

Mémoire de Master

Domaine : Mathématique et Informatique
Filière : Mathématiques
Option : Analyse Mathématique

Présenté par :

Sebai Meriem

Thème

Résolution d'un problème aux limites par la programmation quadratique

---

Soutenu publiquement devant le jury composé de :

Mr. A. Mokhtari	Prof. Université de Laghouat	**Président**
Mr. A. Belacel	M.C.A. Université de Laghouat	**Examinateur**
Mr. A. Bougoutaia	M.A.A. Université de Laghouat	**Examinateur**
Mr. M. Bentobache	M.C.A. Université de Laghouat	**Encadreur**
Mr. S. Achour	Doctorant, Université de Laghouat	**Co-encadreur**

Année universitaire 2018/2019

Remerciements

*Je remercie **Allah** le tout puissant de m'avoir donné le courage, la volonté et la patience de mener à terme ce présent travail.*

*Je tiens tout d'abord à exprimer mes sincères remerciements à mon encadreur Dr. **Mohand Bentobache** pour ces orientations et tous ses efforts consentis dans la préparation de ce travail.*

*Je remercie vivement mon co-encadreur Mr. **Saadi Achour** pour toute la documentation qu'il m'a fournie, ses conseils judicieux, son aide précieuse et les efforts qu'il a déployés pour l'accomplissement de ce travail.*

*Je remercie vivement Professeur **Abdelkader Mokhtari** pour avoir accepté de présider ce travail et aussi pour ses encouragements et ses précieux conseils. Je tiens également à remercier Dr. **Amar Belacel** et Mr. **Amar Bougoutaia** pour leurs acceptations d'examiner ce travail.*

Enfin, mes meilleurs et vifs remerciements s'adressent à tous mes chers enseignants pour les efforts qu'ils ont fournis pour avoir participer à notre formation, et tous ceux qui m'on aidé de près ou de loin à la réalisation de ce modeste travail.

SEBAI Meriem.

Dédicaces

Mes dédicaces s'adressent d'abord à ma mère et mon père pour leurs confiance et leurs encouragements.

A mes soeurs et mes frères pour leurs douceur et leurs gentillesse.

A toute ma famille ainsi qu'à mes amies.

ملخص

في هذه المذكرة، تعاملنا مع مسألة حدية التي يتم تطبيقها في العديد من المجالات (الفيزياء، الكيمياء، الهندسة المعمارية، الهندسة المدنية، الميكانيك،...). و هي مسالة حدود Dirichlet احادية البعد. ثم تطرقنا إلى حل هذا الأخير باستخدام طريقة العناصر المحدودة. أخيرا، من خلال إجراء محاكاة رقمية باستخدام ماتلاب لحل معادلة بواسون احادية البعد، أظهرنا أنه من المثير الاهتمام من حيث وقت التنفيذ، تحويل نظام المعادلات الذي تم الحصول عليه بواسطة طريقة العناصر المحدودة إلى مسالة في البرمجة التربيعية المحدبة دون قيود، ومن ثم حلها باستخدام طريقة التدرج المرافق.

الكلمات الدلالية:

مشكلة حدود، صياغة الاختلافات، طريقة العناصر المحدودة، معادلة بواسون، البرمجة التربيعية، المحاكاة الرقمية.

Abstract

In this memory, we have studied the boundary value problem of Dirichlet in one dimension which arise in many applications in various fields (physics, chemistry, architecture, civil engineering, mechanics, etc.). Then, we have solved this latter using the finite elements method (FEM). In order to solve the linear system obtained by the FEM, we have transformed it to an equivalent unconstrained convex quadratic program. Finally, we have conducted numerical simulation with Matlab for solving the Poisson equation in one dimension. The numerical results have shown that solving the quadratic programming problem with the conjugate gradient method is very efficient.

Key words : Boundary value problem, Finite element method, Numerical simulation, Poisson equation, Unconstrained quadratic programming, Variational formulation.

Résumé

Dans ce mémoire, nous avons abordé un problème aux limites qui est appliqué dans de nombreux domaines (physique, chimie, architecture, génie civil, mécanique, etc.), à savoir le problème aux limites de Dirichlet unidimentionnel. Puis, nous nous somme intéressés à la résolution de ce dernier en utilisant la méthode des éléments finis. Enfin, en effectuant des simulations numériques avec Matlab pour la résolution de l'équation de poisson 1D, nous avons montré qu'il est intéressant en terme de temps d'exécution de transformer le système d'équations obtenus par la méthode des éléments finis en un problème de programmation quadratique convexe sans contraintes et de le résoudre ensuite par la méthode du gradient conjugé.

Mots clés : Equation de Poisson 1D, Formulation variationnelle, Méthode des éléments finis, Problème aux limites, Programmation quadratique, Simulation numérique.

Table des matières

Notations

Soit Ω un ouvert borné de $\mathbb{R}^n$, les notations qu'on a utilisé dans ce mémoire sont les suivantes :

$\bar{\Omega}$	L'adhérence de Ω.
Γ	La frontière de Ω.
$D(\Omega) = C_0^\infty(\Omega)$	L'espace des fonctions réelles indéfiniment différentiables .
	à support compact contenu dans Ω.
$\langle .,. \rangle, (.,.)$	Le produit scalaire d'un espace de Hilbert.
E'	Le dual topologique de E.
p.p.	Presque partout.
$\|.\|_p$	La norme associée à l'espace de Lebesgue $L^p(\Omega)$.
$\|.\|_{H_0^1(\Omega)}$	La norme associée à l'espace de Sobolev $H_0^1(\Omega)$.
$\mathcal{C}^0(\mathbb{R})$	L'espace des fonctions continues.
$L^p(\Omega)$	L'espace de Lebesgue, $1 \leq p \leq \infty$.
$W^{1,p}(\Omega)$	L'espace de Sobolev, $1 \leq p \leq \infty$.
$W_0^{1,p}(\Omega)$	La fermeture de $C_0^\infty(\Omega)$ dans $W^{1,p}(\Omega)$, $1 \leq p < \infty$.
$H^1(\Omega) = W^{1,2}(\Omega)$	L'espace de Sobolev.

Introduction

Les mathématiques utilisent couramment les notions d'infini et de continu. La solution exacte d'un problème d'équations différentielles ordinaires ou aux dérivées partielles est une fonction continue. Les ordinateurs ne connaissent que le fini et le discret. Les solutions approchées seront calculées en définitive comme des collections de valeurs discrètes sous la forme de composantes d'un vecteur solution d'un problème matriciel.

En vue du passage d'un problème exact (continu) au problème approchée (discret), on dispose de plusieurs techniques concurrentes : les différences finies, les éléments finis et les volumes finis. Chacune de ces trois méthodes correspond à une formulation différente des équations de la physique :

- équilibre des forces en chaque point pour les différences finies.
- minimisation de l'énergie ou principe des travaux virtuels pour les éléments finis.
- loi de conservation et calcul des flux pour la méthode des volumes finis.

Examinons rapidement les avantages et les inconvénients de chacune de ces trois méthodes.

1- Différences finies

La méthode des différences finies consiste à remplacer les dérivées apparaissant dans le problème continu par des différences divisées ou combinaisons de valeurs ponctuelles de la fonction en un nombre fini de points discrets ou naeuds du maillage.
Avantages : grande simplicité d'écriture et faible cout de calcul
Inconvénients : Limitation de la géométrie des domaines de calculs, difficultés de prise en compte des conditions aux limites et en général absence de résultats de majoration d'erreurs.

2- Éléments finis

La méthode des éléments finis consiste à approcher, dans un sous-espace de dimension finie, un problème écrit sous forme variationnelle (comme minimisation de l'énergie, en général) dans un espace de dimension infinie. La solution approchée est dans ce cas une fonction déterminée par un nombre fini de paramètres comme, par exemple, ses valeurs en certains points (les naeuds du maillage).
Avantages : Traitement possible de géométries complexes, détermination plus naturelle des conditions aux limites, possibilité de démonstrations mathématiques de convergence et de majoration d'erreurs.
Inconvénients : Complexité de mise en ouvre et coût en temps de calcul et en mémoire.

3- Volumes finis

La méthode des volumes finis intègre, sur des volumes élémentaires de forme simple, les équations écrites sous forme de loi de conservation. Elle fournit ainsi de manière naturelle des formulations discrètes conservative et est donc particulièrement adaptée aux équations de la mécanique des fluides : équation de conservation de la masse, équation de conservation de la quantité de mouvement, équation de conservation de l'énergie.
Sa mise en ouvere est simple si les volumes élémentaires sont des rectangles (ou des parallélépipèdes rectangles en dimension 3). Cependant la méthode des volumes finis permet d'utiliser des volumes élémentaires de forme quelconque, donc de traiter des géométries complexes, ce qui est un avantage sur les différences finies. Il existe une grande variété de méthodes selon le choix de la géométrie des volumes élémentaires et des formules de calcul des flux. Par contre, on dispose de peu de résultats théoriques de convergence.

Un exemple intéressant des problèmes aux limites en dimension 1 est l'équation de poisson 1D donnée par :

$$\begin{cases} -u''(x) = f(x), & a < x < b; \\ u(a) = u(b) = 0. \end{cases} \tag{1}$$

Cette équation modélise plusieurs problèmes physiques. On cite par exemple :

- Barre élastique sous un chargement axial.

$$\begin{cases} -E.u''(x) = f(x), & a < x < b; \\ u(a) = u(b) = 0, \end{cases} \tag{2}$$

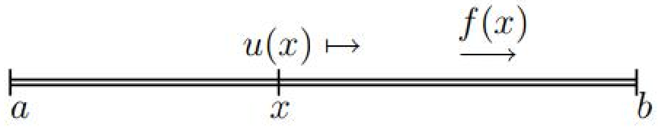

FIGURE 1 – Barre sous chargement axial

où E est le module d'élasticité.

- Corde élastique soumise à un chargement normal.

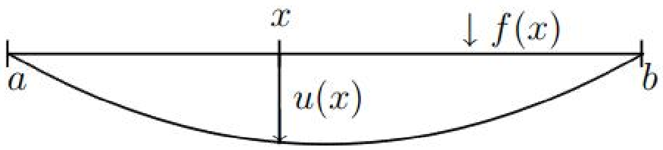

FIGURE 2 – Corde élastique

- Conduction thermique dans une barre.

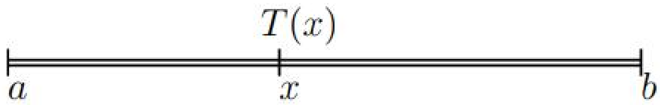

FIGURE 3 – Barre en équilibre thermique

$$\begin{cases} -q(x) = k.T'(x), & a < x < b; \\ q'(x) = f(x), & \\ T(a) = T(b) = 0. & \end{cases} \tag{3}$$

D'où

$$\begin{cases} -k.T''(x) = f(x), & a < x < b; \\ T(a) = T(b) = 0. & \end{cases} \tag{4}$$

La programmation quadratique est une discipline très importante vu le grand nombre d'applications qu'elle possède. Dans ce travail, nous allons montrer qu'il intéressant en terme de temps d'exécution de transformer les systèmes d'équations obtenus par la méthode des éléments finis appliquée à l'équation de Poisson 1D en un problème de programmation quadratique convexe sans contraintes et de le résoudre ensuite par la méthode du gradient conjugé.

Organisation du mémoire

Notre mémoire est organisé de la façon suivante :

1. Dans le premier chapitre, nous allons rappeler les principaux outils mathématiques. Ce chapitre est divisé en deux parties :
   (a) Outils des espaces hilbertiens ; où on présente quelques définitions et quelques notions générales sur ces espaces qui sont très importants dans notre étude.
   (b) Outils des espaces fonctionnels, qui présente les outils nécessaire pour le deuxième chapitre.
2. Dans le deuxième chapitre, nous allons donner la formulation variationnelle du problème aux limites de Dirichlet en dimension un. On commence par étudier un exemple de problèmes aux limites avec conditions aux limites de Dirichlet, justifiant ainsi l'introduction du théorème de $Lax - Milgram$.
3. Dans le troisième chapitre, on va s'intéresser à la résolution d'un problème aux limites de Dirichlet par la méthode des éléments finis.
4. Enfin dans le dernier chapitre, nous allons montrer comment transformer le système d'équations obtenu par la méthode des éléments finis P1 en un problème de programmation quadratique convexe sans contraintes et puis le résoudre d'une façon efficace en utilisant la méthode du gradient conjugés.

Chapitre 1

Rappels sur des outils mathématiques

Les espaces de *Hilbert*

1.1 Introduction

Le mathématicien *allemand David Hilbert* (1862 – 1943), le premier à avoir introduit de manière systématique les espaces qui maintenant portent son nom, est connu pour les 23 fameux problèmes qu'il proposa au congrès international des mathématiciens en 1900. Ses idées ont profondément marqué l'ensemble des mathématiciens jusqu'à l'heure actuelle, en particulier en introduisant des méthodes géométriques en analyse, donnant ainsi naissance à un domaine important des mathématiques : l'analyse fonctionnelle. Les méthodes géométriques présentées dans ce mémoire sont basées sur la notion de produit scalaire et d'orthogonalité. Dans ce paragraphe, nous introduisons l'espace de *Hilbert* avec ses propriétés élémentaires. Nous caractériserons la norme qui est déduite d'un produit scalaire. S'il n'y a pas de spécification le corps $\mathbb{K}$ est, comme auparavant, $\mathbb{R}$ ou $\mathbb{C}$.

Dans cette partie $\mathbb{K}$ désigne $\mathbb{R}$ ou $\mathbb{C}$ et E est un $\mathbb{K}$-espace vectoriel.

1.2 Généralités sur les formes bilinéaires et quadratiques

1.2.1 Formes bilinéaires symétriques et antisymétriques

Définition 1.1. *(les formes bilinéaires) On appelle forme bilinéaire sur $E \times E$ (E est un $\mathbb{K}$-espace vectoriel) toute application $\varphi : E \times E \to \mathbb{K}$ telle que :*

$$\forall \alpha \in \mathbb{K}, \forall (x, x\prime, y) \in E^3, \varphi(\alpha x + x\prime, y) = \alpha\varphi(x, y) + \varphi(x\prime, y)$$

φ est linéaire par rapport à la $1^{ère}$ place.

$$\forall \beta \in \mathbb{K}, \forall (x, y, y\prime) \in E^3, \varphi(x, \beta y + y\prime) = \beta\varphi(x, y) + \varphi(x, y\prime)$$

φ est linéaire par rapport à la $2^{\text{ème}}$ place.

Notation 1.1. *Nous notons $L(E, E; \mathbb{K})$ l'ensemble des formes bilinéaires sur $E \times E$. Il est immédiat que $L(E, E; \mathbb{K})$ est un $\mathbb{K}$-espace vectoriel.*

Proposition. 1.1. *Soient φ une forme bilinéaire sur $E \times E$, $(n, p) \in (\mathbb{N}^*)^2$, $\alpha_1, ..., \alpha_n, \beta_1, ..., \beta_p \in \mathbb{K}$, $x_1, ..., x_n$, $y_1, ..., y_p \in E$. On a alors :*

$$\varphi\left(\sum_{i=1}^{n} \alpha_i x_i, \sum_{j=1}^{p} \beta_j y_j\right) = \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}^{n} \alpha_i \beta_j \varphi(x_i, y_j).$$

Démonstration. On a, par récurrence immédiate sur n :

$$\forall y \in E, \varphi\left(\sum_{i=1}^{n} \alpha_i x_i, y\right) = \sum_{i=1}^{n} \alpha_i \varphi(x_i, y),$$

d'où

$$\varphi\left(\sum_{i=1}^{n} \alpha_i x_i, \sum_{j=1}^{p} \beta_j y_j\right) = \sum_{i=1}^{n} \alpha_i \varphi\left(x_i, \sum_{j=1}^{p} \beta_j y_j\right) = \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{p} \beta_j \varphi(x_i, y_j)\right),$$

$$= \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}^{n} \alpha_i \beta_j \varphi(x_i, y_j).$$

□

Définition 1.2. *Une application $\varphi : E \times E \to \mathbb{K}$ est dite symétrique si et seulement si : $\forall (x, y) \in E^2, \varphi(y, x) = \varphi(x, y)$. On abrégera forme bilinéaire symétrique en : fbs.*

Notation 1.2. *Nous notons $S(E; \mathbb{K})$ l'ensemble des formes bilinéaires symétriques sur $E \times E$. Il est immédiat que $S(E; \mathbb{K})$ est un $\mathbb{K}$-sous espace vectoriel (sev) de $L(E, E; \mathbb{K})$.*

Proposition. 1.2. *Pour qu'une application $\varphi : E \times E \longrightarrow \mathbb{K}$ soit une fbs, il faut et il suffit que l'on ait :*

- *φ est symétrique ;*
- *φ est linéaire par rapport à la 2 ème place.*

1.2.2 Formes quadratiques

Définition 1.3. *Soit φ une fbs sur $E \times E$. On appelle forme quadratique associée à φ l'application, souvent notée φ, de E dans $\mathbb{K}$ définie par :*

$$\forall x \in E, \quad \varphi(x, x) = \varphi(x).$$

On abrègera forme quadratique en : fq.

Proposition. 1.3. *Soient φ une fbs sur $E \times E$, φ la fq associée φ. On a :*

1. $\forall n \in \mathbb{N}^*, \quad \alpha_1, ..., \alpha_n \in \mathbb{K}, \quad \forall x_1, ..., x_n \in E,$
$$\varphi\left(\sum_{i=1}^{n} \alpha_i x_i\right) = \sum_{i=1}^{n} \alpha_i^2 \varphi(x_i) + 2 \sum_{1 \leq i < j \leq n} \alpha_i \alpha_j \varphi(x_i, y_j);$$
2. $\forall (\alpha, \beta) \in \mathbb{K}^2, \forall (x, y) \in E^2, \ \varphi(\alpha x + \beta y) = \alpha^2 \varphi(x) + 2\alpha\beta\varphi(x, y) + \beta^2 \varphi(y);$
3. $\forall (x, y) \in E^2, \varphi(x + y) = \varphi(x) + 2\varphi(x, y) + \varphi(y);$
4. $\forall (x, y) \in E^2, \varphi(x, y) = 1/4(\varphi(x + y) - \varphi(x - y));$
5. $\forall (x, y) \in E^2, \varphi(x + y) + \varphi(x - y) = 2\varphi(x) + \varphi(y).$

Notation 1.3. *Notons $Q(E)$ l'ensemble des fq sur E.*

Remarque 1.1. *Il est clair que l'application $U : S(E; \mathbb{K}) \to Q(E)$ qui, à toute fbs φ sur $E \times E$ fait correspondre la fq associée à φ, et l'application $V : Q(E) \to S(E; K)$ qui, à toute fq φ sur E associe la forme polaire de φ, sont des bijections réciproques l'une de l'autre.*

1.3 Rappels sur les espaces euclidiens

1.3.1 Produit scalaire

Définition 1.4. *(produit scalaire) Soit E un espace vectoriel sur le corps $\mathbb{K}$. Un produit scalaire sur E est une application $\varphi : E \times E \to \mathbb{K}$ telle que pour tout $x, x_1, x_2, y \in E$, et $\lambda \in \mathbb{K}$, on a :*

1. $\varphi(x, x) \geq 0$ *et* $\varphi(x, x) = 0$ *si et seulement si* $x = 0$ *;*
2. $\varphi(x, y) = \overline{(\varphi(y, x))};$
3. $\varphi(x_1 + x_2, y) = \varphi(x_1, y) + \varphi(x_2, y);$
4. $\varphi(\lambda x, y) = \lambda\varphi(x, y).$

Remarque 1.2. *Les propriétés suivantes d'un produit scalaire induit par les axiomes* 1. *à* 4. *:*

i) $\langle 0, x \rangle = \langle x, 0 \rangle = 0$, *pour tout* $x \in E$, *ii)* $\langle \sum_{i=1}^{n} \lambda_i x_i, \sum_{j=1}^{m} \mu_j y_j \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_i \overline{\mu_j} \langle x_i, y_j \rangle$, *pour tout* $x_i, y_j \in E$, $\lambda_i,\ \mu_j \in \mathbb{K}$.

Définition 1.5. *(définition de produit scalaire dans le cas réel) On appelle produit scalaire sur* E *toute fbs* φ *sur* $E \times E$ *telle qu'en notant* φ *la fq associée à* φ*, on ait :*

i) $\forall x \in E : \varphi(x) \geq 0$;

ii) $\forall x \in E : (\varphi(x) = 0) \Rightarrow (x = 0)$.

Notation 1.4. *Lorsque* φ *est un produit scalaire, on note souvent* $(x \backslash y)$ *ou* $\langle x, y \rangle$ *ou* $x.y$ *à la place de* $\varphi(x, y)$.

Définition 1.6. *On appelle espace euclidien tout couple* (E, φ) *où* E *est un* $\mathbb{R}$*-espace vectoriel de dimension finie et* φ *un produit scalaire sur* E*. On abrégera espace vectoriel euclidien en : eve.*

Théorème 1.1. *(Inégalité de Cauchy-Schwarz) Soient* (E, φ) *un eve,* φ *la fq associée à* φ*. On a alors :*

$$\forall x, y \in E, \quad \varphi(x, y)^2 \leq \varphi(x)\varphi(y).$$

Ou

$$|\langle x, y \rangle| \leq \langle x, x \rangle^{1/2} \cdot \langle y, y \rangle^{1/2}.$$

Démonstration. On procède comme d'habitude, pour cette inégalité, on introduit un paramètre $\lambda \in \mathbb{R}$. De la positivité du produit scalaire, on déduit

$$\langle x - \lambda y, x - \lambda y \rangle \geq 0, \quad \forall x, y \in E, \quad \lambda \in \mathbb{R}$$

Mais, de la bilinéarité, on déduit :

$$\langle x - \lambda y, x - \lambda y \rangle = \langle x, x \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle \geq 0, \quad \forall x, y \in E, \quad \lambda \in \mathbb{R}.$$

Ce trinôme du second degré en λ reste toujours positif. Nécessairement, son discriminant est négatif;

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \cdot \langle y, y \rangle.$$

Ceci conduit à l'inégalité de Cauchy-Schwarz. $\square$

Proposition. 1.4. *Soient* (E, φ) *un eve,* φ *la fq associée à* φ*. L'application* $\|.\| : E \underset{x \longrightarrow (\varphi(x))^{1/2}}{\longrightarrow} \mathbb{K}$ *est une norme sur* E*, appelée norme euclidienne associée à* φ*.*

Remarque 1.3. *Soient $\langle .,. \rangle$ un produit scalaire sur E, $||.||$ la norme euclidienne associée. Les formules obtenues en 1), 2) proposition (1.3), 3), 4), 5) peuvent être récrites sous la forme suivante, pour tout (x, y) de E^2 :*

- $||x + y||^2 = ||x||^2 + 2\langle x, y\rangle + ||y||^2$;
- $||x + y||^2 - ||x - y||^2 = 4\langle x, y\rangle$;
- $||x + y||^2 + ||x - y||^2 = 2(||x||^2 + ||y||^2)$.
  L'inégalité de Cauchy-Schwarz peut être récrite :
- $(\langle x, y\rangle)^2 \leq ||x||^2||y||^2$.

1.3.2 Orthogonalité

Définition 1.7. *Soit $(E, \langle .,. \rangle)$ un espace vectoriel euclidien.*

1. *Soit $(x, y) \in E^2$; on dit que x est orthogonal à y, et on note $x \perp y$, si :*
$$\langle x, y\rangle = 0.$$
2. *Soient $x \in E$, $A \in P(E) :=$ { l'ensemble de tous les parties de E} ; on dit que x est orthogonal à A, et on note $x \perp A$, si :*
$$\forall a \in A, \langle x, a\rangle = 0.$$
3. *Pour toute partie A de E, on définit l'orthogonal de A, noté $A^\perp$:*
$$A^\perp = \{x \in E ; \forall a \in A : \langle x, a\rangle = 0\}.$$
4. *Une famille $(x_i)_{i\in I}$ d'éléments de E est dite orthogonale si :*
$$\forall (i, j) \in I^2 : (i \neq j \Rightarrow \langle x_i, x_j\rangle = 0).$$
5. *Une famille $(x_i)_{i\in I}$ d'éléments de E est dite orthonormale si :*
$$(x_i)_{i\in I} \quad \textit{est} \quad \textit{orthogonale} \quad \textit{et} \quad \forall i \in I : \|x_i\| = 1.$$

Remarque 1.4.

1. *$0 \perp x$ pour tout $x \in H$;*
2. *$x \perp y \Longleftrightarrow y \perp x$, symétrie ;*

3. $x \perp x \Rightarrow x = 0$;
4. $x \perp x_i$ *et* $\lambda_i \in \mathbb{K}, i = 1, \ldots, n \Rightarrow x \perp \sum_{i=1}^{n} \lambda_i x_i$;
5. $x \perp x_n$ *et* $x_n \xrightarrow{\|\cdot\|} x_0 \Rightarrow x \perp x_0$;
6. *Soit* $\langle x_i, x_j \rangle = 0; i \neq j; i, j = 1, \ldots, n$ *alors*

$$\|x_1 + \cdots + x_n\|^2 = \|x_1\|^2 + \cdots + \|x_n\|^2.$$

1.4 Espaces préhilbertiens, hilbertiens et théorème de projection et de Riesz

1.4.1 Généralités (Espaces préhilbertiens, espaces de Hilbert)

Définition 1.8. *(espace préhilbertien) Un espace vectoriel muni d'un produit scalaire est appelé un espace préhilbertien.*

Théorème 1.2. *Dans un espace préhilbertien E, l'application $\|.\| : E \to \mathbb{R}$, donnée par $\|x\| = \langle x, x \rangle^{1/2}$, pour tout $x \in E$, est une norme pour E.*

Démonstration. Nous avons $\|x\| \geq 0$ et $\|x\| = 0$ si et seulement si $x = 0$, de plus

$$\|\lambda x\| = \langle \lambda x, \lambda x \rangle^{1/2} = (\lambda \overline{\lambda} \langle x, x \rangle)^{1/2} = (|\lambda|^2 \langle x, x \rangle)^{1/2} = |\lambda| \langle x, x \rangle^{1/2} = |\lambda| \|x\|.$$

Enfin

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle, \\
&= \|x\|^2 + \langle x, y \rangle + \overline{\langle x, y \rangle} + \|y\|^2, \\
&= \|x\|^2 + 2Re\langle x, y \rangle + \|y\|^2, \\
&\leq \|x\|^2 + 2|\langle x, y \rangle| + \|y\|^2, \\
&\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2, \\
&= (\|x\| + \|y\|)^2.
\end{aligned}$$

D'où

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

La norme $\|.\|$ ainsi définie s'appelle la norme induite par le produit scalaire. □

Théorème 1.3. *(Loi du parallélogramme) La norme induite par le produit scalaire satisfait l'égalité*

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

Démonstration. En effet,

$$\|x+y\|^2 + \|x-y\|^2 = \langle x+y, x+y\rangle + \langle x-y, x-y\rangle,$$

$$= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle + \langle x,x\rangle - \langle x,y\rangle - \langle y,x\rangle + \langle y,y\rangle,$$

$$= 2(\|x\|^2 + \|y\|^2).$$

Les égalités

$$\|x+y\|^2 + \|x-y\|^2 = 2(\langle x,y\rangle + \langle y,x\rangle) = 4Re\langle x,y\rangle,$$

$$\|x+iy\|^2 + \|x-iy\|^2 = 4Im\langle x,y\rangle,$$

Nous amènent à

$$4\langle x,y\rangle = \|x+y\|^2 - \|x-y\|^2 + i(\|x+iy\|^2 - \|x-iy\|^2).$$

Cette égalité nous montre que deux produits scalaires différents sur E entraînent deux normes induites différentes.

□

Théorème 1.4. *Un produit scalaire est une fonction continue sur $E \times E$, par rapport à la norme induite.*

Démonstration. Soient x_0, $y_0 \in E$ et les suites $\{x_n\}$ et $\{y_n\}$ de E telles que $x_n \xrightarrow{\|\cdot\|} x_0$ et $y_n \xrightarrow{\|\cdot\|} y_0$.

Alors $|\langle x_n, y_n\rangle - \langle x_0, y_0\rangle| = |\langle x_n - x_0, y_n\rangle + \langle x_0, y_n - y_0\rangle|$,

$$\leq |\langle x_n - x_0, y_n\rangle| + |\langle x_0, y_n - y_0\rangle|,$$

$$\leq \|x_n - x_0\|\|y_n\| + \|x_0\|\|y_n - y_0\|.$$

D'où

$$\lim_{n \longrightarrow +\infty} \langle x_n, y_n \rangle = \langle x_0, y_0 \rangle.$$

□

Définition 1.9 (Espace de Hilbert)**.** *Un espace de Hilbert (ou hilbertien) est un espace vectoriel normé complet par rapport à la norme induite par un produit scalaire. En d'autres mots, un espace de Hilbert est un espace de Banach dont la norme est induite par un produit scalaire (généralement, le corps de base est $\mathbb{R}$ ou $\mathbb{C}$)).*

Dans toute la suite de ce mémoire, H sera un espace hilbertien de dimension infinie, et $(x, y) \in H^2 \longrightarrow \langle x, y \rangle$ un produit scalaire sur H.

Définition 1.10. *On dit q'une suite (x_n) est de Cauchy si :*

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} / \forall p > q \geq n_0 \in \mathbb{N} : \mid x_p - x_q \mid < \varepsilon.$$

On dit que E est complet si et seulement si toute suite de Cauchy de E converge dans E.

1.4.2 Projection d'un point sur un convexe fermé (Théorème de projection)

Définition 1.11. *(la convexité) Soit E un e.v. réel et $K \subset E$,*

- *On dit que K est convexe si*

$$\forall (x, y) \in K^2, \forall t \in [0, 1] : tx + (1 - t)y \in K.$$

(On note encore $[x, y] \subset K$).

- *On dit que K est dite strictement convexe si*

$$\forall (x, y) \in K^2, \forall t \in]0, 1[: tx + (1 - t)y \in K.$$

Ou un ensemble K est convexe si, pour toute paire de points (a, b) de K, K contient aussi le segment $[a, b]$.

Remarque 1.5. *La notion d'ensemble concave n'existe pas.*

Théorème 1.5 (Théorème de projection)**.** *Soit $C \subset H$ une partie convexe fermée non vide. Pour tout $x \in H$, il existe un unique $y \in C$ tel que $d(x, C) = \|x - y\|$, $\quad y = x_C$ est appelé projeté orthogonal de x sur C.*

Démonstration. Pour tout $z \in C$ on a $||x - z|| \geq d(x, C)$ (qui existe, puisque c'est l'inf d'une partie non vide de $\mathbb{R}$ minorée par 0). Par définition, il existe $(x_n)_{n\in\mathbb{N}} \subset C$ tel que :

$$d(x, C) \leq ||x - x_n|| \leq d(x, C) + 1/n, \quad \forall n \in \mathbb{N}.$$

Montrons que (x_n) est de Cauchy ; ainsi, H étant hilbertien c'est un espace complet, donc on pourra conclure quant à la convergence de la suite (x_n) :
Soit $p, q \in \mathbb{N}$, et notons $\delta = d(x, C)$. H étant un espace de Hilbert, l'identité du parallélogramme est vérifiée, donc

$$2(||x - x_p||^2 + ||x - x_q||^2) = ||(x - x_p) - (x - x_q)||^2 + ||(x - x_p) + (x - x_q)||^2,$$

et on a

$$2(||x - x_p||^2 + ||x - x_q||^2) = ||x_p - x_q||^2 + 4||x - (x_p + x_q)/2||^2.$$

Or C est convexe, donc $(x_p + x_q)/2 \in C$, donc

$$||x - (x_p + x_q)/2||^2 \geq \delta^2.$$

En réorganisant, on a donc

$$||x_p - x_q||^2 \leq 2(||x - x_p||^2 - \delta^2 + ||x - x_q||^2 - \delta^2).$$

Or il apparaît par comparaison que

$$\underset{p\to+\infty}{lim} ||x - x_p|| = \delta,$$

donc

$$\underset{p\to+\infty}{lim} ||x - x_p||^2 = \delta^2.$$

D'où pour tout $\varepsilon > 0$ il existe $n_0 \in \mathbb{N}$ tel que

$$\forall p \geq n_0 : |(||x - x_p||^2 - \delta^2)| \leq \varepsilon^2/2.$$

D'où pour $p, q \geq n_0$ on a $||x_p - x_q|| \leq \varepsilon$, donc (x_n) est bien une suite de Cauchy, donc elle converge vers $y \in \overline{C}$ car H est complet. Mais C est fermé, donc $C = \overline{C}$ et $y \in C$. On a

$$\underset{n\to+\infty}{lim} ||x - x_n|| = ||x - y||.$$

par continuité de l'application norme, donc par unicité de la limite,

$$||x - y|| = \delta = d(x, C).$$

Enfin, si on prend $y, z \in C$ vérifiant $||x-y|| = ||x-z|| = d(x,C)$: on construit une suite $(c_n)_{n\in\mathbb{N}}$ telle que $c_n = y$ si n est pair, $c_n = z$ si n est impair. Pour tout $n \in \mathbb{N}$ on a $||x-c_n|| = \delta$ donc avec le même raisonnement que ci-dessus, cela implique la convergence de la suite (c_n), et donc $y = z$. Il y a donc unicité du projeté orthogonal, ce que l'on souhaitait démontrer. □

Proposition. 1.5. *Soit C un convexe fermé non vide de H, $x \in H$ quelconque. Alors il existe un unique $y \in C$ vérifiant*

$$\forall c \in C, \langle c-y, x-y\rangle \leq 0,$$

et $y = x_C$ projection orthogonale de x sur C (c'est une caractérisation de x_C).

Démonstration.

- Si il existe $y \in C$ vérifiant l'hypothèse de la proposition, pour $z \in C$ on a :

$$||z-x||^2 = ||z-y+y-x||^2 = ||z-y||^2 + ||y-x||^2 - 2\langle z-y, y-x\rangle.$$

Donc l'hypothèse assure que $||z-x||^2 \geq ||y-x||^2$ donc cela étant vrai pour tout $z \in C, ||y-x|| \leq d(x,C)$; mais $y \in C$ donc $||y-x|| \geq d(x,C)$ donc finalement $||x-y|| = d(x,C)$. Ayant unicité du projeté orthogonal d'après le théorème(1.4.1), on a bien $y = x_C$.

- Vérifions maintenant que x_C vérifie bien l'hypothèse de l'énoncé.
On a pour tout $z \in C$: $||x-z|| \geq ||x-x_C||$ et

$$||x-z||^2 = ||x-x_C+x_C-z||^2 = ||x-x_C||^2 - 2\langle x-x_C, x_C-z\rangle + ||x_C-z||^2.$$

Donc

$$2\langle z-x_C, x-x_C\rangle \leq ||z-x_C||^2.$$

Nous avons presque le résultat attendu, mais il nous faut se débarrasser du membre de droite. Pour cela, on paramètre le problème : fixons $z_0 \in C$, ayant C convexe on a pour tout $\lambda \in [0,1]$,

$$z = \lambda z_0 + (1-\lambda)x_C \in C.$$

D'où en appliquant le petit calcul ci-dessus, ayant $z - x_C = \lambda(z_0 - x_C)$, nous avons

$$2\lambda\langle z_0 - x_C, x - x_C\rangle \leq \lambda^2||z_0 - x_C||^2.$$

Pour $\lambda \in]0;1]$ cela donne

$$2\langle z_0 - x_C, x - x_C\rangle \leq \lambda||z_0 - x_C||^2.$$

Avec $\lambda \to 0+$ on a bien le résultat attendu. □

1.4.3 Théorème de représentation de Riesz

Théorème 1.6. *(de représentation de Riesz) Soit H' le* ***dual topologique*** *de H (c'est-à-dire l'ensemble des formes linéaires continues sur H).*

Soit $a \in H$, on note $\varphi_a : x \in H \mapsto \langle a, x\rangle \in \mathbb{K}$, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.

Alors $\varphi : a \in H \mapsto \varphi_a \in H'$ est bien définie, et c'est un isomorphisme de H sur H'.

Démonstration. Tout d'abord, φ est bien définie car l'inégalité de Cauchy-Schwarz assure la continuité de la fonction $\varphi_a(a \in H$ fixé).

φ est injective : si $\varphi_a = 0$ alors en particulier, $\varphi_a(a) = \langle a, a\rangle = ||a||^2 = 0$ donc $a = 0$ et $ker\varphi = \{0\}$.

Montrons que φ est surjective. Soit $L \in H\prime$. Si $L = 0L = \varphi(0)$ et c'est bon ; sinon, notons $E = kerL$, E est un hyperplan de H.

Ayant $E = L^{\langle -1\rangle}(\{0\})$ et L continue, E est un fermé de H. D'après la proposition(1.4.2), $E \oplus E^{\perp} = H$. Ayant L différente de l'application nulle, il existe $u \in E^{\perp}$tel que $L(u) \neq 0$. On a en fait $E = Vecta$: si on prend $x \in E^{\perp}$et si on note $w = x - \frac{L(x)}{L(a)}a$ on a $L(w) = L(x) - \frac{L(x)}{L(a)}L(a)$ par la linéarité, donc $L(w) = L(x) - L(x) = 0 : w \in E$. Or $w \in E^{\perp}$ par construction, donc $w \in E \cap E \perp = \{0\}$ d'où $x = \frac{L(x)}{L(a)}a : x \in Vecta$ et par double inclusion (l'autre inclusion étant évidente) on a bien $E^{\perp} = Vect(a)$.

On a donc pour tout $x \in H : x = \frac{L(x)}{L(a)}a + x_E$, $x_E \in E$. Notons maintenant $b = \frac{L(x)}{\|a\|^2}a$.On pour tout $x \in H$

$$\langle x, b\rangle = \langle \frac{L(x)}{L(a)}a + x_E, \frac{L(x)}{\| a \|^2}a\rangle = \frac{L(x)}{L(a)}\frac{L(x)}{\| a \|^2}\langle a, a\rangle = L(x)$$

car $x_E \in E$ et $a \in E^{\perp}$. D'où l'on a $L = \varphi_b$ et φ est bien surjective.

En conclusion, φ est bien un isomorphisme, il y a donc isomorphisme entre H et son dual topologique. □

1.5 Principe des contractions

De nombreuses questions, liées à l'existence et à l'unicité des solutions de certains types d'équations (par exemple, des équations différentielles),. peuvent être ramenées à la question d'existence et d'unicité d'un point fixe pour une application de l'espace métrique correspondant dans lui-même. Parmi les différents critères d'existence et d'unicité d'un point fixe pour de telles applications, l'un des plus simples et à la fois des plus importants est celui qui porte le nom de principe des contractions. Soit R un espace métrique. Une application A de l'espace R dans lui-même est appelée application contractante ou simplement contraction, s'il existe un nombre $\alpha < 1$ tel que pour tout couple de points $x, y \in R$ on a l'inégalité

$$d(Ax, Ay) \leq \alpha d(x, y). \tag{1.1}$$

Toute application contractante est continue. En effet, si $x_n \to x$, en vertu de (2.1) on a également $Ax_n \to Ax$.
On dit que x est un point fixe pour l'application A, si $Ax = x$. Autrement dit, les points fixes sont les solutions de l'équation $Ax = x$.

Théorème 1.7. ***(Principe des contractions).*** *Toute contraction, définie sur un espace métrique complet R, admet un point fixe et un seul.*

Démonstration. Soit x_0 un point arbitraire de R. Posons $x_1 = Ax_0, x_2 = Ax_1 = A^2x_0, ..., x_n = Ax_{n-1} = A^n x_0$.
Montrons que x_n est une suite de Cauchy. En effet, en posant, pour fixer les idées, $m \geq n$ on a

$$d(x_n, x_m) = d(A^n x_0, A^m x_0) \leq \alpha^n d(x_0, x_{m-n}) \leq \alpha^n \{d(x_0, x_1) + d(x_1, x_2) + ... + d(x_{m-n-1}, x_{m-n})\}$$
$$\leq \alpha^n d(x_0, x_1)\{1 + \alpha + \alpha^2 + ... + \alpha^{m-n+1}\} \leq \alpha^n d(x_0, x_1) \frac{1}{1-\alpha}.$$

Comme $\alpha < 1$ pour n assez grand cette quantité peut devenir aussi petite que l'on veut. L'espace R étant complet, la suite de Cauchy x_n a une limite dans R. Posons

$$x = \lim_{n\to\infty} x_n.$$

Alors, en vertu de la continuité de l'application A, on a

$$Ax = A \lim_{n\to\infty} x_n = \lim_{n\to\infty} Ax_n = \lim_{n\to\infty} x_{n+1} = x.$$

L'existence du point fixe est donc démontrée. Démontrons l'unicité de ce point. Si

$$Ax = x, Ay = y,$$

l'inégalité (2.1) prend la forme

$$d(x, y) \leq \alpha d(x, y);$$

comme $\alpha < 1$, il vient

$$d(x, y) = 0$$

c.-à-d.

$$x = y.$$

$\square$

Les espaces $fonctionnels$

1.6 Espaces L^p

Soit Ω un ouvert de $\mathbb{R}^n$ muni de la mesure de Lebesgue dx.

$$\mathcal{F} = \{f : \Omega \to \mathbb{R}, f \text{ intégrable}\} \longrightarrow \mathbb{R}_+,$$

L'application

$$f \longmapsto \|f\| = \int_\Omega |f(x)|dx.$$

est une semi-norme.
On va définir une relation d'équivalence sur $\mathcal{F}$.

$$\forall f, g \in \mathcal{F} \qquad f\Re g \Leftrightarrow \forall x \in \Omega \qquad f(x) = g(x) \ p.p\ .$$

Définition 1.12. *L'ensemble quotient $\mathcal{F}/\Re$ muni de la norme $\|f\|_1 = \displaystyle\int_\Omega |f(x)|dx$, s'appelle l'espace de Lebesgue et sera noté par L^1.*

1.6.1 Définition et propriétés élémentaire des espaces L^p

Définition 1.13. *Soit $p \in \mathbb{R}$ avec $1 < p < \infty$, on dit que $f \in L^p(\Omega)$ si f est mesurable et $|f|^p \in L^1(\Omega)$.*

Théorème 1.8. *L'application*

$$f \longleftrightarrow \|f\|_p = \left[\int_\Omega |f(x)|^p dx\right]^{1/p}.$$

est une norme sur L^p.

Définition 1.14. *On dit que f est essentiellement bornée sur Ω s'il existe une constante C positive telle que $|f(x)| \leq C$ p.p.*
La plus petite de ces constantes est appelée le sup essentiel de f.
On le note par $\sup .ess|f(x)|$.

Définition 1.15. *On appelle espace de Lebesgue de puissance d'ordre ∞ l'espace, noté $L^\infty(\Omega)$, des classes des fonctions mesurables au sens de Lebesgue, définies presque partout sur Ω à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$ vérifiant :*

$$\sup .ess|f(x)| < +\infty.$$

Théorème 1.9. *L'application de $L^\infty(\Omega)$ dans $\mathbb{R}^+$ définit par*

$$f \longmapsto \|f\|_\infty = \sup_{x\in\Omega} .ess|f(x)|.$$

est une norme.

Notation 1.5. *Soit $1 \leq p \leq \infty$, on désigne par q l'exposant conjugué de p, i.e* $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

Théorème 1.10. *$L^2(\Omega)$ est un espace de Hilbert, le produit scalaire étant donné par :*

$$(u, v) = \int_\Omega u(x)\overline{v(x)}dx,$$

$\left(\textit{qui s'écrit } \displaystyle\int_\Omega u(x)v(x)dx \textit{ pour les fonctions réelles}\right).$

1.7 Les espaces de Sobolev

1.7.1 Définitions et premières propriétés

Définition 1.16. *On pose*

$$H^1(\Omega) := \left\{ u \in L^2(\Omega) : \frac{\partial u}{\partial x_i} \in L^2(\Omega), \forall i = 1, 2, ..., n \right\},$$

la dérivation est à comprendre au sens des distribution. En autres termes, une fonction $u \in L^2(\Omega)$ est dans $H^1(\Omega)$ s'il existe des fonctions $v_1, v_2, ..., v_n$ dans $L^2(\Omega)$ telle que :

$$\int_\Omega u \frac{\partial \varphi}{\partial x_i} \, dx = - \int_\Omega v_i \varphi \, dx \quad , \forall \varphi \in D(\Omega) \ , \quad \forall i = 1, 2, ..., n.$$

L'espace $H^1(\Omega)$ est muni du produit scalaire

$$(u, v)_{H^1(\Omega)} = (u, v) + (\nabla u, \nabla v).$$

La norme associée est

$$\|u\|^2_{H^1(\Omega)} = \|u\|_2^2 + \|\nabla u\|_2^2.$$

Théorème 1.11. *L'espace $H^1(\Omega)$ est un espace de Hilbert séparable.*

De la même façon, on définit les espaces de Sobolev $H^m(\Omega)$, où m est un entier strictement positif par :

$$H^m(\Omega) =: \{u \in L^2(\Omega) : D^\alpha u \in L^2(\Omega), \alpha \in \mathbb{N}^n, |\alpha| \leq m\}.$$

On le munit de la norme naturelle :

$$\|u\|_{H^m(\Omega)} = \left(\sum_{\alpha:|\alpha|\leq m} \int_\Omega | D^\alpha u(x) |^2 \, dx \right)^{1/2},$$

(où $D^\alpha u$ est comprise au sens des distributions).
De façon plus générale, pour tout $1 \leq p \leq \infty$ et pour tout $m \in \mathbb{N}$, $m \geq 1$, on peut définir les espaces de Sobolev. Ces espaces sont construits sur l'espace de Banach L^p.

Définition 1.17. *Pour tout $m \in \mathbb{N}$ et $1 \leq p \leq \infty$, on définit l'espace de Sobolev $W^{m,p}$ comme suit*

$$W^{m,p}(\Omega) = \{u \in L^p(\Omega) : D^\alpha u \in L^p(\Omega), \forall \alpha \in \mathbb{N}^n, |\alpha| \leq m\}.$$

Théorème 1.12. *L'application*

$$u \longleftrightarrow \|u\|_{W^{m,p}(\Omega)} = \begin{cases} \left(\sum\limits_{\alpha:\|\alpha\| \leq m} \|D^\alpha u\|_p^p \right)^{1/p}, & si\ 1 \leq p < \infty; \\ \sup\limits_{1 \leq \|\alpha\| \leq m} \|D^\alpha u\|_\infty, & si\ p = \infty. \end{cases}$$

est une norme sur $W^{m,p}(\Omega)$.

Si $p = 2$ on a : $H^m(\Omega) = W^{m,2}(\Omega)$.

Théorème 1.13. *L'application définie par :*

$$H^m(\Omega) \times H^m(\Omega) \longrightarrow \mathbb{C}$$

$$(u, v) \longrightarrow (u, v)_{H^m(\Omega)} = \sum_{|\alpha| \leq m} (D^\alpha u, D^\alpha v),$$

est un produit scalaire sur $H^m(\Omega)$.

Théorème 1.14. *$H^m(\Omega)$ muni du produit scalaire est un espace de Hilbert.*

1.7.2 L'espace $W_0^{1,p}(\Omega)$

Définition 1.18. *Soit $1 \leq p < \infty$, $W_0^{1,p}(\Omega)$ désigne la fermeture de $C_c^1(\Omega)$:={l'ensemble de tous les fonctions de classe C^1 sur un compacte de Ω} dans $W^{1,p}(\Omega)$.*

On note : $H_0^1(\Omega) = W_0^{1,2}(\Omega)$.

L'espace $W_0^{1,p}(\Omega)$ muni de la norme induite par $W^{1,p}$ est un espace de Banach séparable, il est réflexif si $1 < p < \infty$. $H_0^1(\Omega)$ est un espace de Hilbert pour le produit scalaire de $H^1(\Omega)$.

Théorème 1.15. *$C_c^\infty(\Omega)$ est dense dans $W_0^{1,p}(\Omega)$.*
Autrement dit on peut utiliser indifféremment $C_c^\infty(\Omega)$ au lieu de $C_c^1(\Omega)$ dans la définition de $W_0^{1,p}(\Omega)$.

Proposition. 1.6. *Soit $u \in W^{1,p}(\Omega)$, $u \in W_0^{1,p}(\Omega) \Leftrightarrow u = 0$ sur la frontière de Ω.*

Proposition. 1.7. *On peut définir $W_0^{m,p}$ pour $m > 1$, par :*

$$W_0^{m,p}(\Omega) = \{u \in W^{m,p}(\Omega) : u = Du = ... = D^{m-1}u = 0 \quad , \quad sur\ \partial\Omega\}.$$

1.8 Rappels sur l'optimisation sans contraintes

Le problème que l'on étudie ici est celui de la recherche minimum (ou du maximum) d'une fonction réelle f de n variables réelles $x_1, x_2, ..., x_n$ chacune de ces variables pouvant prendre n'importe quelle valeur de $-\infty$ à $+\infty$. De tels problèmes apparaissent fréquemment dans les applications. Soit $f : \mathbb{R}^n \to \mathbb{R}$ qui à tout $x \in \mathbb{R}^n$, $x = (x_1, x_2, ..., x_n)^T$, associe la valeur réelle :

$$f(x) = f(x_1, x_2, ..., x_n).$$

On cherche à résoudre :

$$\begin{cases} \min f(x), \\ x \in \mathbb{R}^n. \end{cases}$$

Il s'agit donc de déterminer un point x^* de $\mathbb{R}^n$ tel que :

$$\forall x \in \mathbb{R}^n : f(x^*) \leqslant f(x), \tag{1.2}$$

c'est-à-dire un minimum global de f sur $\mathbb{R}^n$.

Une condition suffisante pour l'existence de $x^* \in \mathbb{R}^n$ vérifiant (1.2) est que f soit continue (ou, plus généralement semi-continue inférieurement) et possède la propriété de croissance à l'infini : $f(x) \to +\infty$ lorsque $\|x\| \to +\infty$.

Lorsque l'inégalité stricte : $f(x^*) < f(x)$ est vérifiée $\forall x \in \mathbb{R}^n$, $x \neq x^*$, le minimum global x^* est unique, mais évidemment l'unicité n'est pas toujours garantie.

1.8.1 Conditions nécessaires d'optimalité locale

On suppose ici que f est continue et a des dérivées partielles premières $\partial f \diagup \partial x_i$ et secondes $\partial^2 f \diagup \partial x_i \partial x_j$ continues pour tout $x \in \mathbb{R}^n$. Alors :

Théorème 1.16. *Une condition nécessaire pour que x^* soit un minimum (local ou global) de f est :*

(a) $\nabla f(x^*) = 0$ *(stationnarité).*

(b) le hessien $\nabla^2 f(x^*) = [\partial^2 f \diagup \partial x_i \partial x_j (x^*)]$ *est une matrice semi-définie positive, c'est-à-dire :* $\forall y \in \mathbb{R}^n, \quad y^T.\nabla^2 f(x^*).y \geqslant 0.$

Démonstration. Soit x^* un minimum local (ou global) de f.
Comme f est deux fois continûment différentiable, le développement de Taylor à l'ordre 2 au voisinage de x^* donne :

$$f(x) = f(x^*) + \nabla f^T(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*) + \|x - x^*\|^2 \varepsilon(x - x^*)$$

avec $\varepsilon(x - x^*) \to 0$ quand $x \to x^*$.
Si $\nabla f(x^*) \neq 0$ alors en choisissant $x = x^* - \theta \nabla f(x^*)$ on aurait, pour $\theta > 0$ suffisamment petit : $f(x) < f(x^*)$ ce qui contredirait le fait que x^* est un minimum local. Donc la condition (a) est bien nécessaire, et l'on peut écrire :

$$f(x) = f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*) + \|x - x^*\|^2 \varepsilon(x - x^*).$$

Si la matrice $\nabla^2 f(x^*)$ n'est pas semi-définie positive, c'est qu'il existe un vecteur $d \in \mathbb{R}^n$ $(d \neq 0)$ tel que : $d^T \nabla^2 f(x^*) d < 0$.
En choisissant alors $x = x^* + \theta d$, pour $\theta > 0$ suffisamment petit on aurait $f(x) < f(x^*)$ ce qui contredirait encore l'optimalité locale de x^*.
La condition (b) est donc bien également nécessaire. □

Un point x^* qui vérifie la condition (a) c'est-à-dire : $\partial f / \partial x_i(x^*) = 0$ $(i = 1, ..., n)$ est appelé un *point stationnaire.*

Les conditions (a) et (b) du théorème 1.16 ne sont pas suffisantes pour garantir un minimum local ou global. A fortiori la stationnarité seule, bien que nécessaire, n'est pas une condition suffisante d'optimalité locale.

1.8.2 Conditions suffisantes d'optimalité locale

Théorème 1.17. *Sous les mêmes hypothèses qu'au* §1.8.1, *une condition suffisante pour que x^* soit un optimum local de f sur $\mathbb{R}^n$ est*

(a) $\nabla f(x^*) = 0$ *(stationnarité).*

(b) le hessien $\nabla^2 f(x^*)$ est une matrice définie positive, c'est-à-dire : $\forall y \in \mathbb{R}^n, y \neq 0$: $y^T.\nabla^2 f(x^*).y > 0$.

Démonstration. Considérons un point x^* satisfaisant les deux conditions (a) et (b) du théorème. Le développement de Taylor de f à l'ordre 2 au voisinage de x^*s'écrit alors :

$$f(x) = f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*) + \|x - x^*\|^2 \varepsilon(x - x^*),$$

avec $\varepsilon(x - x^*) \to 0$ quand $x \to x^*$.
Pour toute direction de déplacement $d \in \mathbb{R}^n$ $(\|d\| = 1)$ on a alors :

$$f(x^* + \theta d) = f(x^*) + \frac{\theta^2}{2} d^T \nabla^2 f(x^*) d + \theta^2.\varepsilon(\theta),$$

où $\varepsilon(\theta) \to 0$ quand $\theta \to 0$.
En vertu de la condition (b) on a : $d^T.\nabla^2 f(x^*).d > 0$ et par suite, pour θ suffisamment petit,

on aura : $f(x^* + \theta d) > f(x^*)$.
Ce qui montre que x^* est bien un minimum local de f. $\square$

Remarquons que la condition (b) du théorème 1.17 revient à supposer que f est strictement convexe dans un voisinage de x^*.

1.8.3 Cas des fonctions convexes : condition nécessaire et suffisante d'optimalité globale

Théorème 1.18. *Si f est une fonction convexe continûment différentiable, une condition nécessaire et suffisant pour que x^* soit un optimum global de f sur $\mathbb{R}^n$ est que : $\nabla f(x^*) = 0$.*

Autrement dit, dans le cas convexe, la stationnarité à elle seule constitue une condition nécessaire et suffisante d'optimalité globale.

1.8.4 La méthode du gradient conjugué pour les fonctions quadratiques

On suppose ici que la fonction à minimiser est quadratique strictement convexe de la forme :

$$q(x) = \frac{1}{2}x^T Ax + b^T x + c.$$

L'idée de la méthode est de construire progressivement des directions $d_0, d_1, ..., d_k$ mutuellement conjuguées par rapport à la matrice A de la forme quadratique. À chaque étape k, la direction d_k est obtenue par combinaison linéaire du gradient $-\nabla q(x^k)$ en x^k, et des directions précédentes, $d_0, d_1, ..., d_{k-1}$, les coefficients de la combinaison linéaire étant choisis de telle sorte que d_k soit conjuguées par rapport à toutes les directions précédentes.

En notant $g_k = \nabla q(x^k)$ le gradient de la fonction q en x^k, la méthode prend la forme suivante :

Algorithme du gradient conjugué pour les fonctions quadratiques.

(a) Soit x^0 le point de départ, $g_0 = \nabla q(x^0) = Ax^0 + b$.
Poser $d_0 = -g_0$, $k = 0$.

(b) Pour $(k = 0, 1, ..., n-1)$
À l'itération k on est au point x^k.
Définir $x^{k+1} = x^k + \lambda_k d_k$ avec :

$$\lambda_k = -\frac{g_k^T d_k}{d_k^T A d_k}, \tag{1.3}$$

puis :

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \tag{1.4}$$

avec :

$$\beta_k = \frac{g_{k+1}^T A d_k}{d_k^T A d_k}. \tag{1.5}$$

Fpour.

Pour démontrer la validité de l'algorithme (convergence en au plus n étapes), il suffit donc de vérifier que les directions engendrées par (1.4) et (1.5) sont mutuellement conjuguées.

Théorème 1.19. *À une itération k quelconque de l'algorithme où l'optimum de $q(x)$ n'est pas encore atteint (c'est-à-dire $g_i \neq 0,\ i = 0, 1, ..., k$) on a :*

(a)

$$\lambda_k = \frac{g_k^T g_k}{d_k^T A d_k} \neq 0; \tag{1.6}$$

(b)

$$\beta_k = \frac{g_{k+1}^T [g_{k+1} - g_k]}{g_k^T g_k}; \tag{1.7}$$

$$= \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}; \tag{1.8}$$

(c) Les directions $d_0, d_1, ..., d_{k+1}$ engendrées par l'algorithme sont mutuellement conjuguées.

Démonstration. On raisonne par récurrence sur k en supposant que $d_0, d_1, ..., d_k$ sont mutuellement conjuguées.

(a) Montrons d'abord l'équivalence de (1.3) et de (1.6)
On a : $d_k = -g_k + \beta_{k-1} d_{k-1}$
Donc (1.3) s'écrit :
$$\lambda_k = \frac{g_k^T g_k}{d_k^T A d_k} - \beta_{k-1} \frac{g_k^T d_{k-1}}{d_k^T A d_k}.$$
Comme $(d_0, d_1, ..., d_{k-1})$ sont mutuellement conjuguées, x^k est optimum de $q(x)$ sur la variété V^k passant par x^0 et engendrée par $(d_0, d_1, ..., d_{k-1})$ [14].
Donc $g_k^T d_{k-1} = 0$ d'où l'on déduit (1.6).

(b) Pour démontrer (1.7) remarquons que :

$$g_{k+1} - g_k = A(x^{k+1} - x^k) = \lambda_k A d_k.$$

On a alors :

$$g_{k+1}^T A d_k = \frac{1}{\lambda_k} g_{k+1}^T [g_{k+1} - g_k].$$

et en utilisant (1.6) il vient

$$\beta_k = \frac{g_{k+1}^T [g_{k+1} - g_k]}{g_k^T g_k}.$$

ce qui démontre (1.7).

(1.8) découle alors du fait que : $g_{k+1}^T g_k = 0$ car $g_k = d_k - \beta_{k-1} d_{k-1}$ appartient au sous-espace engendré par $(d_0, d_1, ..., d_k)$ et que g_{k+1} est orthogonal à ce sous-espace.

(c) Montrons enfin que d_{k+1} est conjuguée par rapport à $d_0, d_1, ..., d_k$. On a bien $d_{k+1}^T A d_k = 0$ car, en utilisant (1.4) :

$$\begin{aligned}(-g_{k+1} + \beta_k d_k)^T A d_k &= -g_{k+1}^T A d_k + \beta_k . d_k^T A d_k, \\ &= 0 \text{ (par définition de } \beta_k \text{).}\end{aligned}$$

Vérifions maintenant que :

$$d_{k+1}^T A d_i = 0 \quad \text{pour } i = 0, 1, ..., k-1.$$

On a :

$$d_{k+1}^T A d_i = -g_{k+1}^T A d_i + \beta_k d_k^T A d_i.$$

Le second terme est nul par l'hypothèse de récurrence. Montrons qu'il en est de même du premier terme.

Puisque $x^{i+1} = x^i + \lambda_i d_i$ et que $\lambda_i \neq 0$ on a :

$$A d_i = \frac{1}{\lambda_i}(A x^{i+1} - A x^i) = \frac{1}{\lambda_i}(g_{i+1} - g_i).$$

En écrivons :

$$g_{i+1} = -d_{i+1} + \beta_i d_i.$$

$$g_i = -d_i + \beta_{i-1} d_{i-1}.$$

On voit que Ad_i est combinaison linéaire de d_{i+1}, d_i et d_{i-1} (pour $i = 0$, $g_0 = -d_0$, donc Ad_0 est combinaison linéaire de d_1 et de d_0 seulement).

Mais, puisque $(d_0, d_1, ..., d_k)$ sont mutuellement conjuguées, on sait [14] que le point x^{k+1} est l'optimum de $q(x)$ sur la variété V^{k+1}, engendrée par $(d_0, d_1, ..., d_k)$.

Donc g_{k+1} est orthogonal au sous-espace engendré par $(d_0, d_1, ..., d_k)$ et comme Ad_i appartient à ce sous-espace pour $i = 0, 1, ..., k-1$, on en déduit $g_{k+1}^T A d_i = 0$ ce qui achève la démonstration.

□

Chapitre 2

Formulation variationnelle en dimension un

2.1 Introduction et position du problème

Dans ce qui suit, on désignera par $\mathcal{C}^m(I)$ l'espace vectoriel des fonctions réelles m fois continûment dérivables sur un intervalle $I \subset \mathbb{R}$, m entier ≥ 0. Par ailleurs, on notera f', f'', et $f^{(n)}$ pour $n \geq 3$, les dérivées successives d'une fonction d'une variable réelle. Considérons le problème suivant : étant donné deux fonctions $c, f \in \mathcal{C}^0([0,1])$ et deux constantes α et β, trouver une fonction $u \in \mathcal{C}^2([0,1])$ qui vérifie

$$(P) \begin{cases} -u''(x) + c(x)u(x) = f(x), \quad 0 < x < 1, \\ \quad u(0) = \alpha, \quad u(1) = \beta. \end{cases}$$

Un tel problème est appelé problème aux limites, car la fonction inconnue doit satisfaire les conditions aux limites $u(0) = \alpha$, $u(1) = \beta$ posées à la frontière de l'intervalle ouvert sur lequel l'équation différentielle doit être satisfaite.

Un exemple de situation physique où il est rencontré est celui du fléchissement d'une poutre de longueur 1, étirée selon son axe par une force P, soumise à une charge transversale $f(x)dx$ par unité de longueur dx, et simplement appuyée à ses extrémités 0 et 1. Alors le moment fléchissant $u(x)$ au point d'abscisse x est solution d'un problème aux limites du type ci-dessus, avec $c(x) = \frac{P}{EI(x)}$, où E est le module d'Young du matériau constituant la poutre et $I(x)$ le moment principal d'inertie de la section de la poutre au point d'abscisse x, et avec $\alpha = \beta = 0$.

Si l'on suppose la fonction $c \geqslant 0$ sur l'intervalle $[0, 1]$ (nous ferons d'ailleurs cette hypothèse tout au long de ce paragraphe), on peut montrer que ce problème a une solution et une seule [11], qui sera notée φ.

Remarque 2.1. *La condition $c \geqslant 0$ n'est nullement nécessaire pour l'existence et l'unicité d'une solution, qui sont (par exemple) garantis sous l'hypothèse plus faible*

$$-\pi^2 < \inf\{c(x); 0 \leqslant x \leqslant 1\},$$

ou encore si

$$-(k+l)^2\pi^2 < \inf\{c(x); 0 \leqslant x \leqslant 1\} \leqslant \sup\{c(x); 0 \leqslant x \leqslant 1\} < -k^2\pi^2,$$

pour un entier $k \leqslant 1$. On retiendra simplement que, contrairement aux équations différentielles du second ordre associées à des conditions initiales, c'est-à-dire des conditions du type $u(0) = \alpha$, $u'(0) = \alpha'$, l'existence et l'unicité d'une solution ne sauraient être assurées dans tous les cas. On s'en convainc déjà facilement dans le cas d'équations différentielles à coefficients constants .

2.1.1 Théorème de Lax-Milgram

Avant d'énoncer le théorème on fait les hypothèses suivantes :
Soit H un espace de Hilbert, et soit $a : H \times H \longrightarrow \mathbb{R}$ une application bilinéaire vérifiant de plus,

* a est continue sur $H \times H$, c'est-à-dire : il existe une constante $k > 0$ telle que,

$$\forall (u,v) \in H \times H, \quad |a(u,v)| \leq k\|u\|_H \|v\|_H.$$

** a est coercive, c'est-à-dire : il existe une constante $\alpha > 0$ telle que,

$$\forall u \in H, \quad a(u,u) \geq \alpha \|u\|_H^2;$$

Soit $l : H \longrightarrow \mathbb{R}$ une application linéaire de plus, l est continue, c'est-à-dire : il existe une constante $c > 0$ telle que,

$$\forall v \in H, \quad |l(v)| \leq c\|v\|_H.$$

Théorème 2.1. *(De Lax-Milgram)*

Soit H un espace de Hilbert, et soit $a : H \times H \longrightarrow \mathbb{R}$ une application bilinéaire continue et coercive. Alors pour tout $l \in H^* = \mathscr{L}(H, \mathbb{R})$, il existe $u \in H$ unique tel que,

$$a(u,v) = l(v), \quad \forall v \in H \tag{2.1}$$

Démonstration. Passons par :

a) Pour tout $v \in H$, notons $A(v) \in H$ l'unique élément tel que $a(v,.) = (A(v),.)$ (d'après le théorème de représentation de Riesz). Montrons que A est linéaire et continue et que (2.1) équivaut à $A(u) = f$ où $f \in H$ est tel que $l = (f,.)$. Il est clair que A est linéaire de part la bilinéarité de $a(.,.)$. étant donné $u, v \in H$ on a

$$|(A(u),v)| = |a(u,v)| \leq \|a\|\|u\|\|v\|,$$

ce qui montre que

$$\|A(u)\| = \sup_{\|v\| \leq 1} (A(u),v) \leq \|a\|\|u\|,$$

d'où la continuité de A. On remark alors que (1) équivaut à

$$(A(u),v) = l(v) = (f,v), \quad \forall v \in H,$$

ce que est bien équivalent à $A(u) = f$.

b) pour $\rho > 0$, on pose $T_\rho : H \to H$ par $T_\rho(v) = v - \rho\,(A(v) - f)$. Montrons que l'on peut choisir ρ tel que T_ρ soit une contraction. En déduire (1). En développant à $\|T_\rho(\omega) - T_\rho(v)\|^2$, on obtient que

$$\|T_\rho(\omega) - T_\rho(v)\|^2 \leq (1 - 2\rho\gamma + \rho^2 c^2)\|\omega - v\|^2,$$

où $c = \|a\|$ (la norme de l'application bilinéaire $a(.,.)$). Pour $\rho = \frac{\gamma}{c^2}$, on obtient que T_ρ est une contraction de rapport $\sqrt{1 - \frac{\gamma^2}{c^2}}$, qui admet donc un point fixe unique d'après le Théorème des contractions de Banach (théorème 1.7), ce qui démontre bien (2.1) en notant que $T_\rho(u) = u$ si et seulement si $A(u) = f$.

c) Montrons qu'il existe alors $\gamma > 0$ tel que, pour tout $l \in L(H, \mathbb{R})$, on a $a\|u\| \leq \gamma\|f\|$. Notant que u la solution de (2.1), on a

$$\alpha\|u\|^2 \leq a(u,u) = (f,u) \leq \|f\|\|u\|,$$

d'où le résultat avec $\gamma = \alpha^{-1}$.

□

2.2 Formulation variationnelle de problème aux limites de Dirichlet en dimension un

Étant donné deux fonctions $c, f \in \mathcal{C}^0([0,1])$, trouver une fonction $u \in \mathcal{C}^2([0,1])$ qui vérifie

$$(P_1) \begin{cases} -u''(x) + c(x)u(x) = f(x), & 0 < x < 1, \\ \quad u(0) = u(1) = 0. \end{cases}$$

C'est par commodité que nous supposons ici les conditions aux limites homogènes, et ce n'est pas une restriction. En effet, si elles sont posées sous la forme $u(0) = \alpha$, $u(l) = \beta$, il suffit pour s'y ramener de faire un changement de fonction inconnue en soustrayant la fonction $\alpha(1 - x) + \beta x$.

Introduisons l'espace vectoriel V formé des fonctions continues sur l'intervalle $[0,1]$, nulles aux points 0 et 1, et une fois continûment dérivables par morceaux sur ce même intervalle, c'est-à-dire dérivables en tout point de l'intervalle $[0,1]$ sauf en un nombre fini de points x_i de l'intervalle $]0,1[$, la dérivée coïncidant sur chaque intervalle ouvert entre deux points x_i consécutifs avec la restriction d'une fonction continue sur l'intervalle fermé correspondant (le nombre et la position des points x_i varient avec la fonction considérée). Muni de l'application

$$v \in V \rightarrow \|v\|_V = \left(\int_0^1 (|v'|^2 + |v|^2)dx \right)^{\frac{1}{2}},$$

cet espace vectoriel est normé (c'est même un espace pré hilbertien). Pour la mise en ouvere et l'analyse de la méthode que nous avons en vue, il est essentiel de poser Le problème aux limites sous une forme différente, appelée formulation variationnelle : c'est l'objet du résultat qui suit.

Théorème 2.2.

(1) Si u est solution du problème aux limites (P_1), alors

$$a(u, v) = l(v) \text{ pour tout } v \in V,$$

où la forme bilinéaire $a : V \times V \rightarrow \mathbb{R}$ et la forme linéaire $f : V \rightarrow \mathbb{R}$ ont respectivement pour expressions

$$a(u, v) = \int_0^1 (u'(x)v'(x) + c(x)u(x)v(x))dx, \; l(v) = \int_0^1 f(x)v(x)dx.$$

pour des fonctions quelconques $u, v \in V$.

(2) On suppose $c(x) \geq 0$. Une fonction $u \in V$ est solution des équations $a(u, v) = l(v)$ pour tout $v \in V$ si et seulement si

$$J(u) = \inf_{v \in V} J(v), \quad \text{où} \quad J(v) \overset{déf}{=} \frac{1}{2}a(v, v) - l(v).$$

Démonstration.

(i) Supposons connue une solution $u \in \mathcal{C}^2([0, 1])$ de (P_1) que nous appellerons solution classique de (P_1). Soit v une fonction arbitraire de l'espace V, multiplions par v les deux membre de l'équation vérifiée par u et intégrons entre 0 et 1. On obtient

$$\int_0^1 (u''(x) + c(x)u(x))v(x)dx = \int_0^1 f(x)v(x)dx.$$

Notons x_i, $1 \leq i \leq N$, les points, supposés rangés par ordre croissant, où la dérivée de la fonction v n'est pas définie, et posant $x_0 = 0$, $x_{N+1} = 1$, on peut écrire

$$\begin{aligned}
\int_0^1 u''(x)v(x) &= \sum_{i=0}^{N} \int_{x_i}^{x_{i+1}} u''(x)v(x)dx \\
&= \sum_{i=0}^{N} \left(-\int_{x_i}^{x_{i+1}} u'(x)v'(x)dx + \{u'(x)v(x)\}_{x=x_i}^{x=x_{i+1}} \right) \\
&= -\int_0^1 u'(x)v'(x)dx,
\end{aligned}$$

la dernière égalité ci-dessus provenant de la continuité des fonctions u' et v sur l'intervalle $[0,1]$ et des relations $v(0) = v(1) = 0$. Donc

$$\int_0^1 (u'(x)v'(x) + c(x)u(x)v(x))dx = \int_0^1 f(x)v(x)dx \quad \forall v \in V. \tag{2.2}$$

On a donc établi qu'une solution classique de (P_1) vérifie :

$$u \in V \quad \text{et} \quad a(u,v) = l(v), \quad \forall v \in V,$$

où l'on posé, pour tout $u, v \in V$,

$$a(u,v) = \int_0^1 (u'(x)v'(x) + c(x)u(x)v(x))dx, \quad \text{et} \quad l(v) = \int_0^1 f(x)v(x)dx.$$

Le fondement de la méthode variationnelle est le fait surprenant que (2.2) caractérise la solution de (P_1). L'outil théorique principal est le théorème 2.1 (de Lax-Milgram).

(ii) Démontrons l'inégalité suivante, qui nous servira à plusieurs reprises par la suite : en supposant la fonction $c(x) \geq 0$, il existe une constante $\alpha > 0$ telle que

$$\alpha \|v\|_V^2 \leq a(v,v), \quad \forall v \in V.$$

Pour cela, il suffit d'établir que

$$\alpha \|v\|_V^2 \leq \int_0^1 |v'|^2 dx, \quad \forall v \in V.$$

Or on peut écrire, pour tout $x \in [0,1]$,

$$|v(x)| = \Big| \int_0^x v'(t)dt \Big| \leq \int_0^1 |v'(t)|dt \leq \left(\int_0^1 |v'(t)|^2 dt \right)^{\frac{1}{2}},$$

d'après l'inégalité de Cauchy-Schwarz pour les fonctions.
Donc,

$$\begin{aligned}
|v(x)|^2 \leq \int_0^1 |v'(t)|^2 dt \quad &\Rightarrow \quad \int_0^1 |v(x)|^2 dx \leq \int_0^1 \left(\int_0^1 |v'(t)|^2 dt \right) dx, \\
&\Rightarrow \quad \int_0^1 |v(x)|^2 dx \leq \int_0^1 |v'(x)|^2 dx, \\
&\Rightarrow \quad \int_0^1 \left(|v(x)|^2 + |v'(x)|^2 \right) dx \leq 2 \int_0^1 |v'(x)|^2 dx, \\
&\Rightarrow \quad \frac{1}{2} \|v\|_V^2 \leq \int_0^1 |v'(x)|^2 dx
\end{aligned}$$

L'inégalité annoncée est satisfaite avec $\alpha = 1/2$.

(iii) La caractérisation du point (2) démontre à partir de l'identité (de vérification immédiate) :

$$J(u+v)-J(u)=\{a(u,v)-l(v)\}+\frac{1}{2}a(v,v),\quad \forall u,v\in V.$$

En effet, on en déduit, d'une part

$$J(u+v)-J(u)=\frac{1}{2}a(v,v)\geq\frac{\alpha}{2}\|v\|_V^2\geq 0,\quad \forall v\in V.$$

dès que $a(u,v)-l(v)=0$ pour tout $v\in V$. Donc,

$$a(u,v)=l(v)\Longrightarrow J(u+v)-J(u)\geq 0,$$

et car V est un espace vectoriel, donc $u+v$ soit arbitraire de V. Donc,

$$a(u,v)=l(v)\Longrightarrow J(u)=\inf_{v\in V}J(v)......(*).$$

Par ailleurs, pour $v\in V$ fixé, l'inégalité

$$0\leq J(u+\theta v)-J(u)=\theta\{a(u,v)-l(v)\}+\frac{\theta^2}{2}a(v,v),\quad \forall\theta\in\mathbb{R},$$

car θ est arbitraire de $\mathbb{R}$, il faut et il suffit que $a(u,v)-l(v)=0$, donc car V est un espace vectoriel $u+\theta v$ est arbitraire de V. Donc,

$$J(u)=\inf_{v\in V}J(v)\Longrightarrow a(u,v)=l(v)......(**).$$

De $(*)$ et de $(**)$: le point (2) est bien vérifié.

□

Remarque 2.2. .

(1) Lorsque la fonction c est positive, la démonstration ci-dessus fournit une nouvelle preuve de l'unicité de la solution u [11]. Il en résulte en effet que

$$v\in V\quad et\quad a(v,v)=0\Rightarrow\|v\|_V=0\Rightarrow v=0,$$

de sorte que

$$v\neq 0\Rightarrow J(u+v)-J(u)>0.$$

(2) L'expression $\{a(u,v)-l(v)\}$ n'est autre que la dérivée de la fonction J au point u, appliquée à la fonction v ce qui explique la caractérisation du minimum par l'annulation de la "première variation" de la fonction J, c'est pourquoi les relations "$a(u,v)=l(v)$ pour tout $v\in V$" sont appelées des équations variationnelles.

(3) *Réciproquement, on peut définir le problème aux limites directement par l'une des formulations variationnelles* (1) *ou* (2) *du théorème 2.2. La première difficulté consiste ensuite à démontrer l'existence d'une solution, car celle-ci ne peut être établie en toute généralité que dans un espace complet, à savoir la complétion de l'espace V pour sa norme (qui est ici l'espace de Sobolev $H_0^1(]0,1[)$). Si le résultat "abstrait" d'existence est relativement facile à démontrer (cf.théorème 2.3), c'est l'étude des espaces complétés qui est délicate (surtout en dimension ≥ 2, où toutes ces idées se généralisent). Une seconde difficulté consiste à démontrer que la solution ainsi obtenue est suffisamment régulière pour être également une solution au sens "classique" où nous l'avons entendu jusque là.*

2.2.1 Espace de Sobolev $H_0^1(]0,1[)$

Commençons par préciser quelques notations. Pour $1 \leq p < +\infty$, on notera $L^p(]0,1[)$ l'ensemble des (classes) de fonctions mesurables u sur $]0,1[$ telles que $\int_0^1 |u(x)|^p dx < +\infty$, muni de la norme $\|u\|_{L^p(]0,1[)} = \left(\int_0^1 |u(x)|^p dx\right)^{\frac{1}{p}}$. On notera aussi $L^P_{loc}(]0,1[)$ l'ensemble (des classes) de fonctions mesurables u sur $]0,1[$ telles que $|u|^p$ soit intégrable sur tout compact $K \subset]0,1[$.

Définition 2.1. *On note $H^1(]0,1[)$ l'ensemble des fonctions $u \in L^2(]0,1[)$ pour lesquelles il existe $w \in L^2(]0,1[)$ tel que*

$$u(x_2) - u(x_1) = \int_{x_1}^{x_2} w(s)ds \quad \forall x_1, x_2 \in]0,1[.$$

Il est important que les fonctions de $H^1(]0,1[)$ sont continues sur $]0,1[$ (avec l'abus de notation habituel constante à identifier un élément de $L^2(]0,1[)$ avec un de ses représentants) et même prolongeables continûment à $[0,1]$. Les fonctions de $H^1(]0,1[)$ soit aussi dérivables presque partout et $u'(t) = w(t)$ p.p. on notera dans la suite $u' := w$, étant entendu que cette fonction u' n'est égale que presque partout à la dérivée de u. De façon équivalente, $H^1(]0,1[)$ est l'ensemble des fonctions absolument continues sur $[0,1]$ telles que leur dérivée presque partout appartiennent à $L^2(]0,1[)$. On remarkra aussi que l'on peut prendre $x_1, x_2 \in [0,1]$ dans la définition de $H^1(]0,1[)$ et que l'on peut remplacer "pour tout $x_1, x_2 \in [0,1]$" par "il existe x_1 tel que pour tout $x_2 \in [0,1]$". On pose alors, pour tout $u, v \in H^1(]0,1[)$,

$$\langle u, v \rangle = \int_0^1 (u'(x)v'(x) + u(x)v(x))dx \tag{2.3}$$

et on vérifie aisément que l'on définit ainsi un produit scalaire sur $H^1(]0,1[)$.

Proposition. 2.1. *L'espace $H^1(]0,1[)$ muni de la norme dérivée du produit scalaire 2.3 est un espace de Hilbert.*

Démonstration. Soit $(u_n)_{n\in\mathbb{N}} \subset H^1(]0,1[)$ une suite de Cauchy dans $H^1(]0,1[)$. Il résulte de la définition de la norme dans $H^1(]0,1[)$ que les suites $(u_n)_{n\in\mathbb{N}}$ et $(u'_n)_{n\in\mathbb{N}}$ sont de Cauchy dans $L^2(]0,1[$. Notons u et w leurs limites respectives. Choisissons alors $x_1 \in [0,1]$ tel qu'il existe une sous-suite que l'on note encore $(u_n)_{n\in\mathbb{N}}$ telle que $\lim\limits_{n\to+\infty} u_n(x_1) = u(x_1)$. Pour tout $x \in [0,1]$ et pour tout $n \in \mathbb{N}$, on a

$$u_n(x) = u_n(x1) + \int_{x_1}^{x} u'_n(s)ds.$$

Passons à la limite, il vient

$$u(x) = u(x_1) + \int_{x_1}^{x} w(s)ds,$$

ce qui montre que $u \in H^1(]0,1[)$ et que la suite $(u_n)_{n\in\mathbb{N}}$ converge vers u dans $H^1(]0,1[)$. Enfin observant que, pour tout $x \in [0,1]$ et $m,n \in \mathbb{N}$, on a

$$\begin{aligned} |u_n(x) - u_m(x)| &\leq |u_n(x_1) - u_m(x_1)| + \Big|\int_{x_1}^{x} |u'_n(s) - u'_m(s)|ds\Big|, \\ &\leq |u_n(x_1) - u_m(x_1)| + \sqrt{(1-0)}\ \|u'_n - u'_m\|_{L^2(]0,1[)}, \\ &\leq |u_n(x_1) - u_m(x_1)| + \|u'_n - u'_m\|_{L^2(]0,1[)}. \end{aligned}$$

on obtient donc que toute sous-suite de $(u_n)_{n\in N}$ possède une sous-suite qui converge uniformément vers u sur $[0,1]$, ce qui implique la convergence uniforme de $(u_n)_{n\in N}$ vers u sur $[0,1]$. □

Proposition. 2.2. *Posons*

$$H_0^1(]0,1[) = \{u \in H^1(]0,1[) : u(0) = u(1) = 0\}.$$

$H_0^1(]0,1[)$ *est aussi un espace de Hilbert.*

Démonstration. Montrons que $H^1(]0,1[)$ est un espace de Hilbert quand on le munit du produit scalaire de $H^1(]0,1[)$.
Il découle immédiatement de proposition 2.1 que $H_0^1(]0,1[)$ est fermé dans $H^1(]0,1[)$ on notant le fait, établi dans la démonstration de proposition 2.1 , que la convergence dans $H^1(]0,1[)$ implique la convergence uniforme sur $[0,1]$. □

Étant donné un espace de Hilbert V, une fonction $J : V \to \mathbb{R}$ est appelée fonctionnelle quadratique sur V si elle est de la forme

$$J(v) = \frac{1}{2}a(v,v) - l(v),$$

où $a(.,.) : V \times V \to \mathbb{R}$ est une forme bilinéaire, continue, symétrique ($a(u,v) = a(v,u)$ pour tout $u, v \in V$) et $l : V \to \mathbb{R}$ est une forme linéaire continue. Cette définition généralise de façon naturelle celle d'une fonctionnelle quadratique sur $\mathbb{R}^n$ puisque, grâce au théorème de représentation de Riesz, il existe un opérateur $A \in \mathscr{L}(V)$ et un élément $b \in H$, tous deux définis de façon unique, tels que

$$a(u,v) = (Au, v) = (u, Av), \quad \forall u, v \in V,$$

$$l(v) = (b, v), \quad \forall v \in V,$$

en désignant par $(.,.)$ le produit scalaire de l'espace V.
Le théorème de projection et le théorème de représentation de Riesz permettent alors d'établir simplement un résultat général d'existence pour des problèmes (P_*) posés avec de telles fonctionnelles. On notera que le cas $U = V$ correspond exactement à la formulation variationnelle des problèmes aux limites.

Théorème 2.3. *Soit*

$$J : v \in V \to J(v) = \frac{1}{2}a(v,v) - l(v),$$

une fonctionnelle quadratique sur un espace de Hilbert V. On suppose de plus qu'il existe un nombre α tel que

$$\alpha > 0 \quad et \quad a(v,v) \geq \|v\|_H^2 \quad \forall v \in V.$$

étant donné une partie non vide, convexe, fermée U de V, il existe un et un seul élément u vérifiant

$$(P_*) \qquad\qquad u \in U \quad et \quad J(u) = \inf_{v \in V} J(v).$$

Cet élément u, vérifie également

$$a(u, v - u) \geq l(v - u), \quad \forall v \in U,$$

et, réciproquement, si un élément $u \in V$,vérifie les inéquations ci-dessus, c'est la solution du problème (P_*). Si U est un sous-espace vectoriel, les inéquations précédentes sont remplacées par les équations

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

Démonstration. La forme bilinéaire $a(.,.)$ est également un produit scalaire sur l'espace V, la norme associée étant équivalente à la norme $\|.\|$ associée au produit scalaire $(.,.)$ de l'espace V. En effet, les hypothèses faites entrainent :

$$\sqrt{\alpha}\|v\| \leq \sqrt{a(v,v)} \leq \sqrt{\|a\|}\|v\|,$$

en désignant par $\|a\|$ la norme (dans l'espace $\mathscr{L}_2(V;\mathbb{R})$) de l'application bilinéaire $a(.,.)$. La forme linéaire l étant donc encore continue pour cette nouvelle norme, le théorème de représentation de Riesz montre qu'il existe un élément $c \in V$ et un seul tel que

$$l(v) = a(c, v), \quad \forall v \in V.$$

Par suite, on peut transformer l'expression de la fonctionnelle, en l'écrivant

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

Dans ces conditions, résoudre le problème (P_*) revient à chercher la projection u de l'élément c sur l'ensemble U, au sens du produit scalaire $a(.,.)$. D'après le théorème de projection, il en existe une et une seule, ce qui établit l'existence et l'unicité de la solution u du problème (P_*). D'après le même théorème, cette solution est également caractérisée par les inéquations

$$a(u - c, v - u) \geq 0, \quad \forall v \in U,$$

ou par les équations

$$a(u - c, v) = 0, \quad \forall v \in U.$$

si U est un sous-espace vectoriel, relations qui coïncident avec celles de l'énoncé puisque $a(c, v) = l(v)$ pour tout $v \in V$. □

Remarque 2.3.

(1) Un usage essentiel de la symétrie de la forme bilinéaire a été fait, d'une part, pour conclure que l'expression $a(.,.)$. est un produit scalaire, d'autre part, pour écrire la nouvelle expression de la fonctionnelle.

(2) Les inéquations $a(u, v - u) \geq l(v - u)$ sont des cas particulier des inéquations d'Euler $J'(u)(v - u) \geq 0$ appliquées à la fonctionnelle J, de dérivée donnée par

$$J'(u)v = a(u, v) - l(v), \quad \forall v \in V.$$

(3) On a indiqué aux paragraphes les raisons pour lesquelles les relations $a(u, v) = l(v)$ sont des équations "variationnelles"; c'est dans le même esprit que les relations "$a(u, v - u) \geq l(v - u)$ pour tout $v \in U$" sont appelées des inéquation variationnelles.

Chapitre 3

Éléments finis linéaires unidimensionnels

3.1 Principes généraux de l'approximation

3.1.1 Une famille de problèmes variationnels linéaires

Soit V un espace de Hilbert réel muni du produit scalaire $(.,.)$ et de la norme associée $\|.\|$.
Soit a une forme bilinéaire continue sur V et coercive, c'est à dire telle qu'il existe, pour tout $v \in V$, un nombre $\alpha > 0$ tel que :

$$a(v,v) \geq \alpha \|v\|^2. \tag{3.1}$$

Soit l une forme linéaire continue sur V.
On considère le problème P_0 suivant :

$$\begin{cases} \text{Trouver une fonction } u \text{ appartenant à l'espace de Hilbert } V \text{ telle que :} \\ a(u,v) = l(v), \quad \forall v \in V. \end{cases} \tag{3.2}$$

Sous les hypothèses précédentes : V espace de Hilbert réel, a forme bilinéaire continue sur V et coercive, l une forme linéaire continue sur V , Le théorème de Lax-Milgram conduit aux conclusions suivantes :

1) Le problème P_0 admet une solution unique u dans V
2) Si la forme bilinéaire a est symétrique le problème variationnel P_0 est équivalent au problème de minimisation suivant :

$$\begin{cases} \text{Trouver une fonction } u \in V \text{ qui minimise la forme quadratique,} \\ J(v) = \frac{1}{2}a(v,v) - l(v). \end{cases} \tag{3.3}$$

3.1.2 Approximation interne du problème

On suppose maintenant que l'on connait un sous-espace $V_h \subset V$ de dimension finie, paramétré par h et tel que pour tout $v \in V$, il existe un élément $r_h v \in V_h$ vérifiant :

$$\lim_{h \longrightarrow 0} \|r_h v - v\| = 0. \tag{3.4}$$

On parle d'approximation interne car $V_h \subset V$.
Considérons alors le problème P_h

$$\begin{cases} \text{Trouver une fonction } u_h \text{ appartenant à l'espace de Hilbert } V_h \text{ telle que :} \\ a(u_h, v_h) = l(v_h), \quad \forall v_h \in V_h. \end{cases} \tag{3.5}$$

Ce problème admet également une solution unique car $V_h \subset V$, et donc les hypothèses du théorème de Lax-Milgram sont également vérifiées dans V_h.

Remarque 3.1. *Si la forme bilinéaire a est symétrique le problème variationnel P_h est équivalent au problème de minimisation suivant :*

$$\left\{ \begin{array}{l} \text{Trouver une fonction } u_h \in V_h \text{ qui minimise la forme quadratique,} \\ J(v_h) = \frac{1}{2}a(v_h, v_h) - l(v_h). \end{array} \right. \tag{3.6}$$

et on a donc évidemment

$$J(u_h) \geq J(u).$$

3.1.3 Majoration d'erreur

Théorème 3.1. *Soit M la constante intervenant dans l'hypothèse de continuité de a : $a(u, v) \leq M\|u\|\|v\|$ et α la constante intervenant dans l'hypothèse de coercivité : $a(v, v) \geq \alpha\|v\|^2$, on a la majoration d'erreur suivante :*

$$\|u - u_h\| \leq \frac{M}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|. \tag{3.7}$$

Démonstration. De

$$a(u, v) = l(v), \quad \forall v \in V,$$

et

$$a(u_h, v_h) = l(v_h), \quad \forall v_h \in V_h,$$

on obtient

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

et

$$a(u - u_h, u_h - u + u - v_h) = 0, \quad \forall v_h \in V_h,$$

soit

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

et le résultat. □

Remarque 3.2. *Si a est symétrique :*

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

signifie que u_h est la projection de u dans V_h au sens du produit scalaire a. Dans ce cas on a

$$\|u - u_h\| \leq \sqrt{\frac{M}{\alpha}} \inf_{v_h \in V_h} \|u - v_h\|. \tag{3.8}$$

3.1.4 Un premier exemple d'approximation interne : la méthode de Galerkin

On suppose l'espace de Hilbert V séparable. Il existe donc une base $\{w_j\}_{j=1}^{+\infty}$ dénombrable engendrant un sous-espace dense dans V. On considère alors un sous-ensemble fini $B_m = \{w_j\}_{j=1}^{m}$ et l'espace V_m engendré par B_m.
Soit Π_m l'opérateur de projection orthogonale de V dans V_m on a :

$$\lim_{m \longrightarrow \infty} \|\Pi_m v - v\| = 0, \quad \forall v \in V. \tag{3.9}$$

Le problème P_m suivant :

$$\begin{cases} \text{Trouver une fonction } u_m \text{ appartenant à l'espace } V_m \text{ telle que :} \\ a(u_m, v_m) = l(v_m), \quad \forall v_m \in V_m. \end{cases} \tag{3.10}$$

qui admet une solution unique dans V_m peut s'écrire sous forme d'un système linéaire :

$$\sum_{j=1,m} a_{i,j} u_j = L_i, \quad \forall i = 1, \ldots, m. \tag{3.11}$$

avec

$$a_{i,j} = a(w_j, w_i), \tag{3.12}$$

et

$$L_i = L(w_i). \tag{3.13}$$

3.2 Éléments finis P1 pour le problème aux limites de Dirichlet unidimensionnel

3.2.1 Problème aux limites de Dirichlet homogènes

Reprenons le problème aux limites de Dirichlet homogènes :

$$\begin{cases} \text{Trouver une fonction } u \text{ appartenant à } H_0^1[0,1] \text{ telle que :} \\ \int_0^1 u'(x)v'(x)dx + \int_0^1 c(x)u(x)v(x)dx = \int_0^1 f(x)v(x)dx, \quad \forall v \in H_0^1[0,1]. \end{cases} \tag{3.14}$$

et introduisons une discrétisation de l'intervalle $[0,1]$ en N sous-intervalles ou éléments $T_i = [x_{i-1}, x_i]$

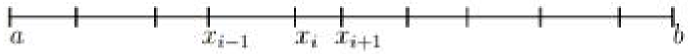

FIGURE 3.1 – Discrétisation.

Les éléments T_i n'ont pas forcément même longueur. $V_{0,h}$ est alors l'espace des fonctions continues affines par morceaux (affines sur les segments T_i) et nulles aux extrémités 0 et 1.

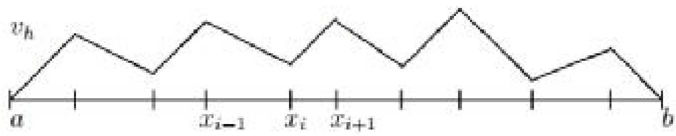

FIGURE 3.2 – Une fonction de $V_{0,h}$.

Rappelons que chaque fonction $v_h \in V_{0,h}$ est déterminée de manière unique par la donnée de ses valeurs aux points x_i pour $i = 1, N-1$. L'espace $V_{0,h}$ est de dimension $N-1$ et est engendré par la base de Lagrange de $V_{0,h}$, formée des $N-1$ fonctions $w_i \in V_{0,h}$ définies par les $N-1$ conditions suivantes :

$$w_i(x_j) = \delta_{ij}, \qquad \forall i = 1, N-1 \quad \text{et} \quad \forall j = 1, N-1. \tag{3.15}$$

Une fonction v_h quelconque sécrit dans cette base :

$$v_h(x) = \sum_{i=1}^{N-1} v_i w_i(x). \tag{3.16}$$

avec $v_i = v_h(x_i)$. Les coefficients v_i sont donc les valeurs de v_h aux points (x_i).

3.2.2 Écriture du problème approché

écrivons le problème approché dans $V_{0,h}$

$$\left\{ \begin{array}{l} \text{Trouver une fonction } u_h \text{ appartenant à } V_{0,h} \text{ telle que } \quad \forall v_h \in V_{0,h}, \\ \displaystyle\int_0^1 u_h'(x)v_h'(x)dx + \int_0^1 c(x)u_h(x)v_h(x)dx = \int_0^1 f(x)v_h(x)dx. \end{array} \right. \tag{3.17}$$

Le problème étant linéaire, l'égalité est vraie pour tout v_h si elle est vraie pour une base de l'espace vectoriel $V_{0,h}$.

D'autre part, écrivons u_h, solution du problème approché dans $V_{0,h}$, dans la base des w_i

$$u_h(x) = \sum_{j=1}^{N-1} u_j w_j(x). \tag{3.18}$$

avec $u_j = u_h(x_j)$ valeur approchée de la solution exacte au point x_j.
On obtient l'écriture suivante du problème approché :

$$\begin{cases} \text{Trouver } u_1, u_2, ..., u_{N-1} \text{ tels que } \quad \forall i = 1, \ldots, N-1 \\ \displaystyle\sum_{j=1}^{N-1} \left(\int_0^1 w_j'(x) w_i'(x) dx + \int_0^1 c(x) w_j(x) w_i(x) dx \right) u_j = \int_0^1 f(x) w_i(x) dx. \end{cases} \tag{3.19}$$

Soit en posant

$$\int_0^1 f(x) w_i(x) dx = F_i, \tag{3.20}$$

et

$$\int_0^1 w_j'(x) w_i'(x) dx + \int_0^1 c(x) w_j(x) w_i(x) dx = a_{ij}, \tag{3.21}$$

$$\sum_{j=1}^{N-1} a_{ij} u_j = F_i, \qquad \forall i = 1, N-1. \tag{3.22}$$

On a ainsi obtenu un système linéaire de $N-1$ équations à $N-1$ inconnues, qui peut s'écrire sous la forme matricielle suivante :

$$A\,U = F. \tag{3.23}$$

Dans la dernière section de ce chapitre, nous allons voir comment calculer les éléments de la matrice A et du second membre F pour le cas $c(x) = 0$.

3.2.3 Problème aux limites de Dirichlet non-homogènes

Le problème aux limites de Dirichlet non-homogènes s'écrit

$$\begin{cases} -u''(x) + c(x)u(x) = f(x), & \forall x \in [0,1], \\ u(0) = \alpha, \quad u(1) = \beta. \end{cases} \tag{3.24}$$

On pose $u = \tilde{u} + u_0$ en choisissant une fonction auxiliaire simple u_0 prenant les valeurs fixées

$$u_0(0) = \alpha, \quad u_0(1) = \beta. \tag{3.25}$$

Le problème se ramène alors à un problème aux limites de Dirichlet homogènes pour la nouvelle inconnue $\tilde{u}$ dont la formulation variationnelle s'écrit :

$$\begin{cases} \text{Trouver une fonction } \tilde{u} \text{ appartenant à } H_0^1[0,1] \text{ telle que :} \\ \int_0^1 \tilde{u}'(x)v'(x)dx + \int_0^1 c(x)\tilde{u}(x)v(x)dx = \int_0^1 f(x)v(x)dx, \\ -\int_0^1 u_0'(x)v'(x)dx - \int_0^1 c(x)u_0(x)v(x)dx, \quad \forall v \in H_0^1[0,1]. \end{cases} \tag{3.26}$$

On obtient donc dans ce cas le problème approché suivant dans $V_{0,h}$:

$$\begin{cases} \text{Trouver une fonction } \tilde{u}_h \text{ appartenant à } V_{0,h} \text{ telle que :} \\ \int_0^1 \tilde{u}_h'(x)v_h'(x)dx + \int_0^1 c(x)\tilde{u}_h(x)v_h(x)dx = \int_0^1 f(x)v_h(x)dx, \\ \int_0^1 u_0'(x)v_h'(x)dx - \int_0^1 c(x)u_0(x)v_h(x)dx, \quad \forall v_h \in V_{0,h}. \end{cases} \tag{3.27}$$

Il reste simplement à préciser le choix pratique de u_0. On prend habituellement pour u_0, la fonction suivante de l'espace V_h des fonctions continues affines par éléments :

$$u_0 = \alpha w_0 + \beta w_N. \tag{3.28}$$

L'existence et l'unicité des solutions des problèmes continus et discrets se démontrent aisément par application du Théorème de Lax-Milgram.
D'un point de vue pratique, la seule modification à apporter au système par rapport au cas Dirichlet homogène concerne les seules composantes 1 et $N-1$ du second membre.

3.3 Application à la résolution de l'equation de Poisson 1D

Considérons le problème au limites suivant :

$$\begin{cases} -u''(x) = f(x), & x \in]0,1[\,; \\ u(0) = u(1) = 0. \end{cases} \tag{3.29}$$

Ce probléme est dit équation de poisson unidimentionnelle, c'est un cas particulier du problème de Dirichlet aux limites homogènes, avec $c(x) = 0$. Résolvons ce problème avec la méthode des éléments finis P1.

La Formulation variationnelle du problème. Soit $v \in H_0^1(\Omega), \Omega =]0,1[$.
En intégrant par partie, on obtient

$$\int_0^1 u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx, \forall v \in V. \tag{3.30}$$

Ce qui est équivalent à écrire

$$a(u, v) = l(v),$$

avec $a(u, v) = \int_0^1 u'(x)v'(x)dx$ et $l(v) = \int_0^1 f(x)v(x)dx$.

Discrétisation : Pour $h > 0$, on cherche $u \in V_{0,h}$ ($V_{0,h}$ est un sev de V), tel que

$$\int_0^1 u_h'(x)v_h'(x)dx = \int_0^1 f(x)v_h(x)dx, \forall v_h \in V_{0,h} \tag{3.31}$$

ou

$$a(u_h, v_h) = l(v_h),\ \forall v \in V_{0,h}.$$

Construisons le maillage suivant :

$$x_0 = 0,\ x_{i+1} = x_i + h, i = 1, \ldots, N,\ \text{avec } N = \frac{1}{h}.$$

Choisissons une base de fonctions de l'espace $V_{0,h}$ vérifiant :

$$w_i(x_j) = \delta_{ij} = \begin{cases} 1, & \text{si } i = j\,; \\ 0, & \text{sinon.} \end{cases}$$

En utilisant la méthode des éléments finis P1, on choisit les fonctions $w_i,\ i = 1, 2, \ldots, N-1$ comme étant les polynôme de degré 1 suivants :

$$w_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, & \text{si } x_{i-1} \leq x \leq x_i\,; \\ \frac{x - x_{i+1}}{x_i - x_{i+1}}, & \text{si } x_i \leq x \leq x_{i+1}\,; \\ 0, & \text{sinon.} \end{cases}$$

On a bien $w_i(x_i) = 1,\ w_i(x_{i+1}) = 0,\ w_i(x_{i-1}) = 0,\ w_i(x_{i-2}) = 0$, etc. Donc

$$w_i(x_j) = \begin{cases} 1, & \text{si } i = j\,; \\ 0, & \text{sinon.} \end{cases}$$

Exprimons u_h dans la base $\{w_i\}_{i=1,\ldots,N-1}$:

$$u_h(x) = \sum_{i=1}^{N-1} u_i w_i(x),$$

avec $u_i = u_h(x_i),\ i = 1, \ldots, N-1$.

En posant $v_h = w_1,\ w_2,\ \ldots,\ w_{N-1}$, la forme (3.31) devient

$$\int_0^1 u_h'(x)w_j'(x)dx = \int_0^1 f(x)w_j(x)dx,\ j = 1, \ldots, N-1$$

ou

$$a(u_h, w_j) = l(w_j),\ j = 1, \dots, N-1.$$

En remplaçant $u_h(x)$ par $\sum_{i=1}^{N-1} u_i w_i(x)$, on obtient

$$\int_0^1 u_h'(x) w_j'(x) dx = \int_0^1 \left(\sum_{i=1}^{N-1} u_i w_i'(x) \right) w_j'(x) dx = \sum_{i=1}^{N-1} u_i \int_0^1 w_i'(x) w_j'(x) dx.$$

D'où

$$a(u_h, w_j) = a\left(\sum_{i=1}^{N-1} u_i w_i, w_j \right) = \sum_{i=1}^{N-1} u_i a(w_i, w_j),$$

Posons

$$a_{ij} = a(w_i, w_j) = \int_0^1 w_i'(x) w_j'(x) dx$$

et

$$F_i = \int_0^1 f(x) w_i(x) dx.$$

On obtient alors le système d'équations linéaires suivant :

$$\sum_{j=1}^{N-1} a_{ij} u_j = F_i,\ i = 1, \dots, N-1 \Leftrightarrow AU = F,$$

avec $A = (a_{ij},\ i = 1, \dots, N-1,\ j = 1, \dots, N-1)$, $F = (F_i,\ i = 1, 2, \dots, N-1)$ et $U = (u_1, u_2, \dots, u_{N-1})$ le vecteur des variables qui représentent les valeurs approchées de u aux noeuds $x_1,\ x_2, \dots, x_{N-1}$.

Calculons les éléments de la matrice A et ceux du second membre F. On a

$$a_{ij} = \int_0^1 w_i'(x) w_j'(x) dx = \int_0^1 w_j'(x) w_i'(x) dx = a_{ji},$$

donc A est symétrique. De plus

$$a_{i,i+2} = \int_{x_{i-1}}^{x_i} w_i'(x) w_{i+2}'(x) dx + \int_{x_i}^{x_{i+1}} w_i'(x) w_{i+2}'(x) dx = 0.$$

De la même manière, on trouve $a_{ij} = 0$, pour $j > i+1$ ou $j - i > 1$.

Le calcul de $w_i'(x)$ nous donne :

$$w_i'(x) = \begin{cases} \frac{1}{x_i - x_{i-1}} = \frac{1}{h}, & \text{si } x_{i-1} \leq x \leq x_i\,; \\ \frac{1}{x_i - x_{i+1}} = -\frac{1}{h}, & \text{si } x_i \leq x \leq x_{i+1}\,; \\ 0, & \text{sinon.} \end{cases}$$

On a aussi

$$a_{ii} = \int_0^1 w_i'(x) w_i'(x) dx$$

$$
\begin{aligned}
&= \int_{x_{i-1}}^{x_i} w_i'(x)w_i'(x)dx + \int_{x_i}^{x_{i+1}} w_i'(x)w_i'(x)dx \\
&= \int_{x_{i-1}}^{x_i} \frac{1}{h^2}dx + \int_{x_i}^{x_{i+1}} \frac{1}{h^2}dx = \frac{2}{h}.
\end{aligned}
$$

Calculons $a_{i,i+1}$ et $a_{i-1,i}$:

$$
\begin{aligned}
a_{i,i+1} &= \int_{x_{i-1}}^{x_i} w_i'(x)w_{i+1}'(x)dx + \int_{x_i}^{x_{i+1}} w_i'(x)w_{i+1}'(x)dx \\
&= \int_{x_i}^{x_{i+1}} -\frac{1}{h} \times \frac{1}{h}dx = -\frac{1}{h^2}(x_{i+1} - x_i) = -\frac{1}{h}. \\
a_{i-1,i} &= \int_{x_{i-1}}^{x_i} w_{i-1}'(x)w_i'(x)dx + \int_{x_i}^{x_{i+1}} w_{i-1}'(x)w_i'(x)dx \\
&= \int_{x_{i-1}}^{x_i} \frac{1}{x_{i-1} - x_i}\frac{1}{h}dx = -\frac{1}{h^2}(x_i - x_{i-1}) = -\frac{1}{h}.
\end{aligned}
$$

Donc on aura :

$$
A = \frac{1}{h}\begin{pmatrix} 2 & -1 & \dots & 0 \\ -1 & 2 & \dots & 0 \\ 0 & -1 & \ddots & -1 \\ 0 & 0 & \dots & 2 \end{pmatrix}.
$$

Calcul des F_i avec la méthode des trapèzes :

Rappelons que la formule des trapèzes nous donne

$$
\int_\alpha^\beta g(x)dx = \frac{\beta - \alpha}{2}\left[g(\alpha) + g(\beta)\right].
$$

On a alors

$$
\begin{aligned}
F_i = \int_0^1 f(x)w_i(x)dx &= \int_{x_{i-1}}^{x_i} f(x)w_i(x)dx + \int_{x_i}^{x_{i+1}} f(x)\varphi_i(x) \\
&= \frac{x_i - x_{i-1}}{2}\left[f(x_i)w_i(x_i) + f(x_{i-1})w_i(x_{i-1})\right] \\
&+ \frac{x_{i+1} - x_i}{2}\left[f(x_{i+1})w_i(x_{i+1}) + f(x_i)w_i(x_i)\right] \\
&= \frac{h}{2}\left[2f(x_i)\right] = hf(x_i).
\end{aligned}
$$

3.3.1 Exemple numérique

Considérons le problème au limites suivant :

$$\begin{cases} -u''(x) = sin(\pi x), & x \in]0,1[\,; \\ u(0) = u(1) = 0. \end{cases} \tag{3.32}$$

On a $f(x) = sin(\pi x)$. Prenons $h = 1/4$ et appliquons la formule des trapèzes pour le calcul des seconds membres, on obtient le système linéaire $AU = F$, avec

$$A = \begin{pmatrix} 8 & -4 & 0 \\ -4 & 8 & -4 \\ 0 & -4 & 8 \end{pmatrix}, F = \begin{pmatrix} hf(x_1) \\ hf(x_2) \\ hf(x_3) \end{pmatrix} = \begin{pmatrix} \frac{1}{4}sin\frac{\pi}{4} \\ \frac{1}{4}sin\frac{\pi}{2} \\ \frac{1}{4}sin\frac{3\pi}{4} \end{pmatrix} = \begin{pmatrix} 0.1768 \\ 0.25 \\ 0.1768 \end{pmatrix}.$$

La solution de ce systéme est :

$$U = (0.0754, 0.1067, 0.0754)^T.$$

En intégrant deux fois et en utilisant les condition aux limites, on trouve la solution exacte suivante :

$$u(x) = \frac{sin\pi x}{\pi^2}.$$

Le tableau suivant nous permet de comparer la solution approchée avec la solution exacte :

x	0.25	0.5	0.75
u_i	0.0754	0.1067	0.0754
$u(x_i)$	0.0716	0.1013	0.0716

Chapitre 4

Résolution de l'équation de Poisson 1D par la programmation quadratique

4.1 Introduction

Dans ce chapitre nous allons montrer comment utiliser les techniques de la programmation quadratique convexe pour résoudre le système obtenu avec la méthodes des éléments finis P1 pour l'équation de Poisson.

4.2 Résolution par la programmation quadratique

Remarquons que le système $AU = F$ est équivalent au système $QU = b$ avec $Q = (q_{ij}, i, j = 1, \ldots, N-1)$, avec

$$q_{ij} = \begin{cases} 2, & \text{si } i = j\,; \\ -1, & \text{si } |j-i| = 1\,; \\ 0, & \text{sinon,} \end{cases}$$

et $b = (b_1, b_2, \ldots, b_{N-1})$, avec $b_i = hF_i,\ i = 1, \ldots, N-1$.
Notons d'abort que Q est tridiagonale, symétrique et définie positive. Il est facile de montrer que la solution du système $QU = b$ est aussi solution du problème de minimisation quadratique strictement convexe sans contrainte suivant :

$$\min_{u \in \mathbb{R}^{N-1}} J(u) = \frac{1}{2} u^T Q u - b^T u. \tag{4.1}$$

En effet, comme le Hessien de J est égal à $Q > 0$, la condition $\nabla J(u) = 0 \Leftrightarrow Qx = b$ est nécessaire et suffisante pour l'optimalité locale et globale du point u. Par conséquent les algorithme itératifs de programmation quadratique strictement convexe peuvent être appliqués pour la résolution de ce problème.

4.2.1 Exemple numérique

Considérons le problème au limites suivant :

$$\begin{cases} -u''(x) = sin(\pi x), & x \in]0,1[\,; \\ u(0) = u(1) = 0. \end{cases} \tag{4.2}$$

La forme quadratique correspondant au système obtenu dans l'exemple numérique du chapitre précédent est donnée par :

$$\min_{u \in \mathbb{R}^3} J(u) = \frac{1}{2} u^T Q u - b^T u, \tag{4.3}$$

avec

$$Q = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \ b = \begin{pmatrix} 0.0442 \\ 0.0625 \\ 0.0442 \end{pmatrix}.$$

4.3 Simulation numérique et résultats expérimentaux

Afin de comparer les approches de résolution des problèmes de programmation quadratique avec celle utilisées pour la résolution des systèmes d'équations, nous avons implémentée la méthode des élénents finis P1 avec Matlab2018 pour la résolution de l'équation de poisson résolue dans l'exemple numérique du chapitre précédent.

Nous avons comparé le temps d'exécution et l'erreur de trois méthodes : la méthode implémentée dan Matlab "quadprog", la méthode de résolution des systèmes d'équations implémentée dans matlab qui consiste à calculer la solution du système $Qu = b$ avec la commande $u = Q\backslash b$ ("syslin") et la méthode du gradient conjugé ("gradientC") en commençant par l'origine et en posant $tol = 10^{-6}$. Nous avons varié le pas de discrétisation $h = 1/4,\ h = 1/10, h = 1/20,\ h = 1/1000$ et $h = 1/10000$. Les résultats obtenus sont présentés dans les tableaux 4.1.

On représente graphiquement la solution exacte et les solutions approchées obtenues pour $N = 20$ (Figure 4.1, 4.2 et 4.3), la précision dessolutions approchée obtenues par les trois méthodes (Figure 4.4) et le temps d'exécution des trois méthodes pour différentes valeur de N (Figure 4.5).

On remarque que pour $N = 20$ la solution approchée trouvée par les trois méthodes est presque identique avec la solution exacte. De plus, nous constatons que la résolution du système d'equation avec la méthode de factorisation directe (opérateur $\backslash$ de Matlab) est plus précise que les autres méthodes pour le problème considéré. Cependant, pour de grande valeur de N, on constate que la méthode du gradient conjugé est plus rapide ce qui montre l'intérêt de la transformation en un programme quadratique équivalent.

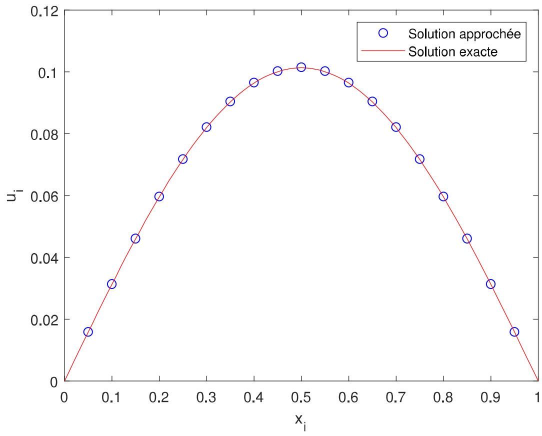

FIGURE 4.1 – Comparaison de la solution approchée obtenue avec syslin et la solution exacte pour $N = 20$.

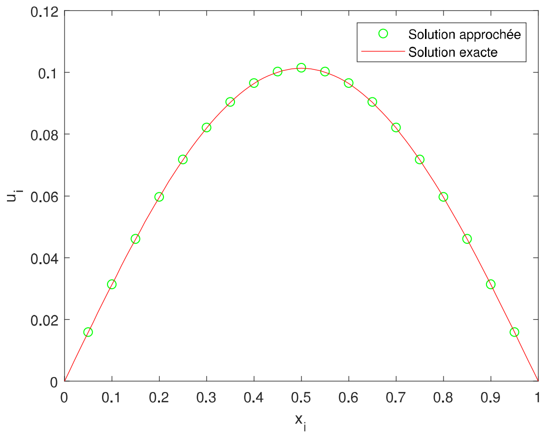

FIGURE 4.2 – Comparaison de la solution approchée obtenue avec quadprog et la solution exacte pour $N = 20$.

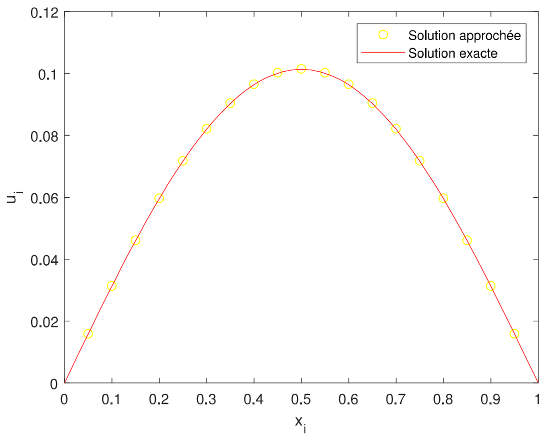

FIGURE 4.3 – Comparaison de la solution approchée obtenue avec gradientC et la solution exacte pour $N = 20$.

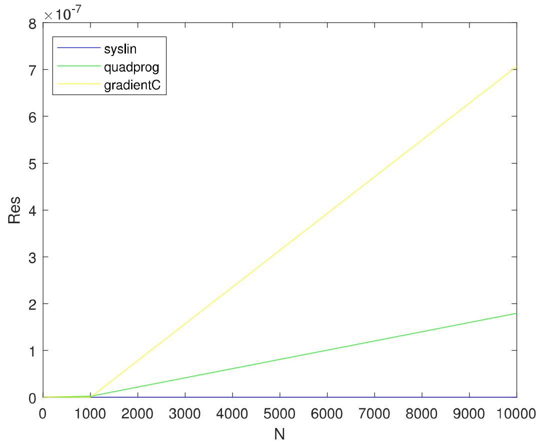

FIGURE 4.4 – Comparaison des solutions approchées obtenues avec les trois méthodes en terme de précision.

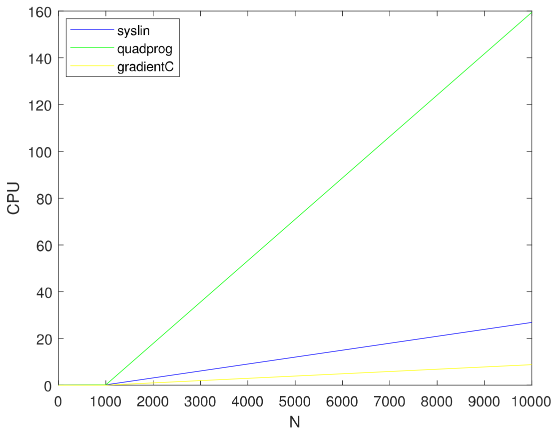

FIGURE 4.5 – Comparaison des temps d'exécution des trois méthodes.

Algorithme	syslin		quadprog		gradientC	
N	Res	CPU	Res	CPU	Res	CPU
4	2.0817e-17	0.0012	7.5470e-12	0.0050	6.9389e-18	7.2949e-04
10	5.5768e-17	1.8863e-04	1.1445e-11	0.0045	4.0367e-17	3.9496e-04
20	7.5253e-17	2.0095e-04	1.6184e-11	0.0047	1.1920e-16	4.0305e-04
1000	4.2220e-16	0.2371	2.3607e-09	0.1395	9.5242e-16	0.0059
10000	1.4039e-15	26.8494	1.7919e-07	159.3953	7.0711e-07	8.8382

TABLE 4.1 – Comparaison des trois méthodes.

Où : $Res = \|Qu - b\|$, avec u la solution approchée trouvée et CPU représente le temps d'exécution.

Conclusion

Nous avons étudié les théorie détaillée des problèmes aux limites unidimentionnels avec des conditions de Dirichlet. On a rappelé les aspects théoriques du problème et les outils de bases nécessaires pour la résolution numérique de ce dernier. Cependant, certains problèmes mathématiques peuvent être difficiles à résoudre analytiquement, c'est pourquoi on fait recours aux méthodes numériques pour trouver une solution rapprochée. Ainsi, nous avons exposé en détail la méthode des éléments finis P1 pour la résolution approchée du problème aux limites de Dirichlet unidimentionnel, puis nous avons illuster la méthode avec l'équation de Poisson 1D. Enfin, en effectuant des simulations numériques avec Matlab pour la résolution de l'équation de Poisson 1D, nous avons montré qu'il est intéressant en terme de temps d'exécution de transformer le système d'équations obtenus par la méthode des éléments finis en un problème de programmation quadratique convexe sans contraintes et de le résoudre ensuite par la méthode du gradient conjugé.

Bibliographie

[1] S. Achour (2017) Formulation variationnelle d'un problème aux limites de Dirichlet sous un problème d'optimisation quadratique. Mémoire de Master en mathématiques, Université de Laghouat.

[2] R. A. Adams (1975) Sobolev spaces, Academic Press, New York.

[3] G. Allaire (2005) Analyse numérique et optimisation. école polytechnique, ISBN : 2-7302-1255-8.

[4] G. Allaire, F. ALOUGES (2015) Analyse variationnelle des équations aux dérivées partielles. école Polytechnique.

[5] D. Azé, J. Baptiste, H. Urruty (2010) Analyse variationnelle et optimisation. Cépadués marss.

[6] A. Bendali (2013) Méthode des éléments finis. Département GMM 4 ème année-Orientation MMN, Toulouse.

[7] E. Blayo (2010) Notes de cours sur la méthode des éléments finis. M1 Mai.

[8] H. Brezis (1983) Analyse fonctionnelle théorie et applications. Masson, Paris.

[9] A. Chikhaoui (2010) Algorithmes d'optimisation des Problèmes de formes quadratiques. Thèse de doctorat en informatique. Université des sciences et de la technologie d'Oran «Mohamed Boudiaf».

[10] P.G. Ciarlet, J.L. Lions (1998) Introduction à l'analyse numérique matricielle et à l'optimisation, Cours et exercices corrigés. Dunod, Paris.

[11] P.G. Ciarlet, B.Miara, J.M. Thomas (1987) Exercices d'analyse numérique matricielle et d'optimisation avec solution. Masson.

[12] J-P. Demailly (2006) Analyse numérique et équations différentielles, EDP Sciences,France.

[13] A. Kolmogorov, S. Fomine (1973) éléments de la théorie des fonctions et de l'analyse fonctionnelle. Mir, Moscou.

[14] M. Minoux (2008) Programmation mathématique, théorie et algorithmes, 2^e édition. Tec et Doc, Paris.

[15] J-H Saiac (2006) Analyse numérique des équations aux dérivées partielles. Polycopié des cours de Calcul Scientifique CSC108 et CSC109.

[16] L. Schwartz (1970) Topologie générale et analyse fonctionnelle. Hermann, Paris.