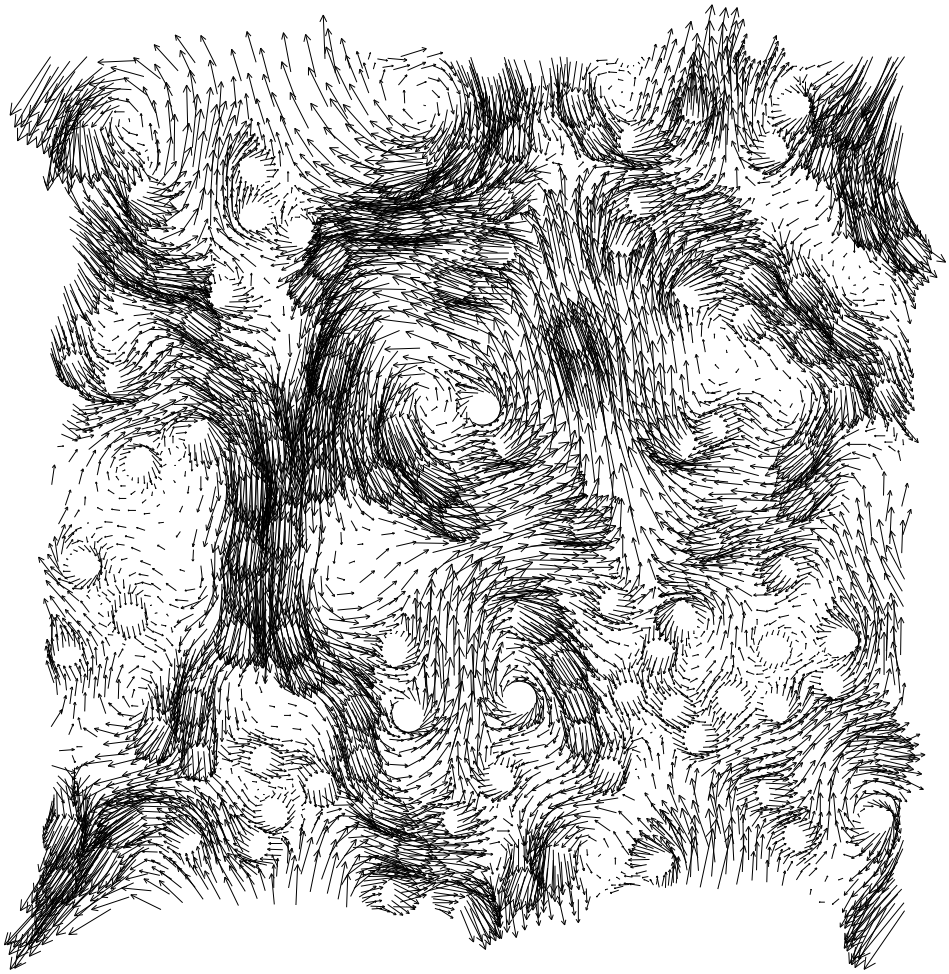
Méthode des éléments finis et optimisation sous contrainte



B.Maury
Laboratoire de Mathématiques d'Orsay

Avant-propos

Ce document a été réalisé en accompagnement d'un cours de M2 donné au Laboratoire de Mathématiques d'Orsay. Son évolution incrémentale au fil des années a pu conduire à des bizarreries de structure, qui je l'espère ne gêneront pas trop la lecture.

Nous avons pris le parti de regrouper les chapitres en deux grandes parties. La première traite des aspects de modélisation au sens large : construction des équations à partir de principes physiques (essentiellement principe de conservation et lois phénoménologique), et démarche permettant de formuler ces problèmes sous une forme adaptée au cadre variationnel sur lequel se base la méthode des éléments finis. La seconde partie regroupe les chapitres plus théoriques : bases d'analyse fonctionnelle, d'analyse numérique, et compléments théoriques sur l'optimisation sous contraintes dans les espaces de Hilbert.

Table des matières

Première partie

Modélisation

Chapitre 1

Éléments de modélisation des milieux continus

1.1. Flux et conservation

On s'intéresse ici à la description de la distribution d'une substance dans l'espace au cours du temps. On notera $\rho(x,t)$ cette densité.

Définition 1.1. (Vecteur flux)
Soit $\mathbf{x}$ un point du domaine, $\mathbf{n}$ un vecteur unitaire, et $D_\varepsilon(\mathbf{n})$ un disque (ou un segment s'il s'agit de la dimension 2) centré en $\mathbf{x}$, d'aire ε (de longueur ε en dimension 2), et normal à $\mathbf{n}$. On note $J(\varepsilon,\mathbf{n})$ la quantité de substance qui traverse D_ε par unité de temps, comptée positivement dans le sens $\mathbf{n}$. S'il existe un vecteur $\mathbf{J}$ tel que, pour tout $\mathbf{n}$, la quantité $J(\varepsilon,\mathbf{n})/\varepsilon$ tende vers une limite quand ε tend vers 0, et que cette limite s'écrive $\mathbf{J}\cdot\mathbf{n}$, on appelle $\mathbf{J}=\mathbf{J}(\mathbf{x})$ le vecteur flux en $\mathbf{x}$.

Équation de conservation. On considère une substance qui se propage selon le vecteur flux $\mathbf{J}$. On écrit que la dérivée en temps de la quantité de substance contenue dans un sous-domaine ω immobile est égal au bilan instantané des flux à travers la frontière.

$$\frac{dn_\omega}{dt}=\frac{d}{dt}\int_\omega \rho(x,t)\,\mathrm{dx}=-\int_{\partial\omega}\mathbf{J}\cdot\mathbf{n}=-\int_\omega \nabla\cdot\mathbf{J}.$$

Cette identité étant vérifiée pour tout ω, on en déduit l'équation

$$\frac{\partial\rho}{\partial t}+\nabla\cdot\mathbf{J}=0. \tag{1.1}$$

Terme source. On peut intégrer à ce modèle des termes-source (ou termes-puits si l'on enlève de la matière), en considérant une quantité f de matière injectée par unité de temps et par unité de volume. Le bilan instantané de matière sur un volume ω s'écrit alors

$$\frac{\mathrm{d}}{\mathrm{dt}}\int_\omega \rho=-\int_{\partial\omega}\mathbf{J}\cdot\mathbf{n}+\int_\omega f,$$

ce qui conduit à l'équation

$$\frac{\partial\rho}{\partial t}+\nabla\cdot\mathbf{J}=f.$$

Modèle 1.1. (Équation d'advection)
On considère une substance décrite par sa densité $\rho(x,t)$, et convectée par un champ de vitesse $\mathbf{u}$. Le vecteur flux s'écrit $\mathbf{J}=\rho\mathbf{u}$, et l'équation correspondante est

$$\frac{\partial\rho}{\partial t}+\nabla\cdot\rho\mathbf{u}=f.$$

1.2. Diffusion

Modèle 1.2. (Loi de Fick)
On dit qu'un phénomène de propagation suit la loi de Fick s'il existe un paramètre positif D tel que

$$\mathbf{J} = -D\nabla\rho.$$

Remarque 1.2. D'un point de vue qualitatif, cette loi exprime le fait que la substance a tendance à aller des zones à forte densité vers les zones à faible densité. On peut donc s'attendre à ce qu'un tel phénomène tende à uniformiser les distributions.

Équation de la chaleur. On considère une substance qui diffuse dans un milieu selon la loi de Fick (modèle 1.2). L'équation de conservation (1.1) s'écrit ici

$$\frac{\partial\rho}{\partial t} - \nabla \cdot D\nabla\rho = 0,$$

ou, dans le cas où D est uniforme,

$$\frac{\partial\rho}{\partial t} - D\triangle\rho = 0. \tag{1.2}$$

Diffusion non isotrope. Dans le cas où le milieu n'est pas isotrope (i.e. la diffusion est plus importante dans certaines direction), on peut introduire une matrice de diffusion définie positive $\mathbf{D}$ qui conduit a une équation formellement analogue.

Conditions aux limites. On suppose que le phénomène de diffusion prend place dans une zone délimitée de l'espace. On note Ω cette zone, et l'on suppose que Ω est un ouvert borné. Il est alors licite de prescrire deux types de condition sur la frontière de Ω.

(i) Conditions de Dirichlet : la valeur de la densité est imposée au bord du domaine.

(ii) Conditions de Neumann : on prescrit le flux $\mathbf{J} \cdot \mathbf{n}$ à travers la frontière du domaine Ω, c'est-à-dire, sous l'hypothèse de flux régi par la loi de Fick, la dérivée normale de la densité, ou plus précisément $-D\partial\rho/\partial n$.

Il est possible de panacher ces deux conditions, c'est-à-dire d'imposer la valeur de ρ sur une partie de la frontière, et la valeur de la dérivée normale sur son complémentaire.

Notons qu'un troisième type de conditions aux limites peut être envisagé, qui implique à la fois la valeur de la fonction et sa dérivée normale, il s'agit des

(iii) Conditions de Robin (ou Fourier) : on prescrit une combinaison linéaire (à coefficient positifs) de la valeur et de la dérivée normale.

Précisons d'où peuvent venir ces dernières conditions en prenant l'exemple de la diffusion de l'oxygène dans le sang au travers de la paroi alvéolaire. On assimile une alvéole à une sphère remplie d'air, au sein duquel l'oxygène diffuse selon la loi de Fick avec un certain paramètre de diffusivité D. La paroi alvéolaire sépare l'alvéole des capillaires dans lesquels circulent le sang, dont les globules rouges vont capter l'oxygène. Au sein de cette paroi, l'oxygène diffuse également et comme elle est très fine, il est licite de négliger au premier ordre la diffusion dans la direction transverse. Si l'on note u_{ext} la concentration en oxygène

dans le sang, on peut écrire que le flux d'oxygène au travers de la paroi est proportionnel à la différence de valeurs de part et d'autre, ce qui conduit à écrire

$$\text{Flux alvéole vers sang} = \beta(u - u_{\text{ext}}),$$

où u est la valeur de la concentration dans l'alvéole au voisinage de la paroi alvéolaire, d'où la condition en tout point de la frontière

$$-D\frac{\partial u}{\partial n} = \beta(u - u_{\text{ext}})\,,\ \text{i.e. } \beta u + D\frac{\partial u}{\partial n} = \beta u_{\text{ext}}.$$

Noter que cette condition présente l'avantage de contenir d'une certaine manière toutes les autres, puisque l'on retrouve des conditions de Neumann en faisant tendre β vers 0, et des conditions de Dirichlet [1] en faisant tendre β vers $+\infty$.

1.3. Écoulements en milieu poreux

On dit qu'un milieu est poreux s'il est constitué d'une phase solide qui ne remplit pas complètement l'espace, de telle sorte qu'un fluide puisse passer au travers.

Modèle 1.3. (Loi de Darcy en milieu isotrope)
On considère l'écoulement d'un fluide visqueux dans un milieu poreux saturé. On dit que cet écoulement suit la Loi de Darcy s'il existe k, appelé conductivité hydraulique [2] du milieu, tel que

$$\mathbf{u} = -k\nabla \mathrm{p},$$

où p la pression au sein du fluide, et $\mathbf{u}$ est la vitesse moyenne locale.

Noter que la vitesse définie ci-dessus doit être interprétée comme un débit par unité de surface (qui est bien homogène à une vitesse), et non pas comme la vitesse effective de particules fluides. En particulier dans le cas d'une porosité petite (fraction de fluide réduite) , on peut avoir une vitesse caractéristique $\tilde{\mathbf{u}}$ du fluide, mais si l'on prend son flux au travers d'une surface Σ, l'écoulement ne se produit qu'au niveau des pores, de telle sorte que le flux effectif est très inférieur à $\int_\Sigma \tilde{\mathbf{u}} \cdot \mathbf{n}$, mais vaudra [3] $\int_\Sigma \mathbf{u} \cdot \mathbf{n}$.

Equation de Darcy. L'écoulement en milieu poreux saturé d'un fluide visqueux incompressible est régi par

$$\nabla \cdot k\nabla \mathrm{p} = 0\,,\ \mathbf{u} = -k\nabla \mathrm{p}$$

où p est la pression au sein du fluide, $\mathbf{u}$ la vitesse, k la conductivité hydraulique.

1. Cette technique est couramment utilisée numériquement pour imposer, dans le cadre des méthodes d'éléments finis, des conditions de Dirichlet sans changer la structure de la matrice : il s'agit de la méthode de pénalisation frontière.

2. Cette conductivité s'exprime K/μ, où K est la *perméabilité intrinsèque* du milieu (qui ne dépend pas du fluide qui s'écoule, et μ la viscosité du fluide. La perméabilité est homogène à une surface, et le lien avec la loi de Poiseuille (13.7) est immédiat, si l'on écrit la vitesse comme un débit par unité de surface :

$$u = \frac{Q}{\pi a^2} = \frac{8a^2}{\mu}\frac{\mathrm{p}_{\text{entrée}} - \mathrm{p}_{\text{sortie}}}{L}.$$

3. Cette définition d'une vitesse de Darcy « locale » n'a donc de sens que comme vitesse moyenne sur des volumes élémentaires représentatifs (petits devant la taille caractéristique du domaine considéré) de taille très supérieure à la taille caractéristique des pores.

Le modèle peut s'écrire sous la forme [4]

$$\begin{cases} \mathbf{u} + k\nabla \mathrm{p} &= \mathbf{f} \\ \nabla \cdot \mathbf{u} &= 0 \end{cases}$$

ou en éliminant la vitesse, ce qui ramène à un problème de Poisson

$$-\nabla \cdot k\nabla \mathrm{p} = \nabla \cdot \mathbf{f}.$$

Si l'on s'intéresse aux solutions d'un tel modèle sur un domaine Ω borné, on peut prescrire sur toute composante Σ de la frontière l'une des deux conditions aux limites suivantes :

(1) Conditions de Dirichlet : on impose la pression $\mathrm{p} = \mathrm{p}_0$ sur Σ.
(2) Conditions de Neumann : on impose $\partial \mathrm{p}/\partial n$, dérivée normale de la pression, ce qui revient à imposer la vitesse normale (ou flux par unité de surface) sur la frontière ($\mathbf{u} \cdot \mathbf{n} = -k\partial_n \mathrm{p}$).

Dissipation d'énergie. L'écoulement régi par le modèle de Darcy est dissipatif, et le taux d'énergie dissipée par unité de volume (en Wm^{-3}) s'écrit [5] $K\left|\nabla \mathrm{p}\right|^2$, de telle sorte que l'énergie dissipée au sein du matériau s'écrit

$$\int_\Omega k\left|\nabla \mathrm{p}\right|^2.$$

.

1.4. Autres modèles

Nous évoquons ici d'autres modèles conduisant à une équation de Poisson.

Modèle de Hele-Shaw. On considère 2 plaques parallèles rigides entre lesquelles se trouve un fluide visqueux. L'écoulement est alors régi par un modèle de type Darcy, basé sur le fait que l'on peut définir une pression (constante dans la direction transverse) et un champ de vitesse bidimensionnel (vitesse moyennée dans la direction transverse), tels que l'on ait

$$\mathbf{u} = -k\nabla \mathrm{p}.$$

Supposons maintenant que l'on injecte un fluide visqueux entre les deux plaques. On note $\Omega(t) \subset \mathbb{R}^2$ le domaine bidimensionnel correspondant à la zone occupée par le fluide, et ν la fonction (ou mesure) terme source [6]. Si l'on suppose la pression nulle dans la zone extérieure au fluide, on peut écrire un modèle d'évolution pour le domaine Ω :

$$\begin{aligned} \nabla \cdot \mathbf{u} = -k\Delta \mathrm{p} &= \nu \quad \text{sur } \Omega(t) \\ \mathrm{p} &= 0 \quad \text{sur } \Gamma(t) = \partial\Omega(t) \end{aligned}$$

4. On parle de problème de type point-selle : formulation duale d'un problème de minimisation sous contrainte.

5. On pourra faire lien avec le $P = RI^2$ qui donne la puissance dissipée au sein d'un fil conducteur de résistance R. Dans le contexte des écoulements en milieu poreux, la pression p joue le rôle du portentiel, et la vitesse u le rôle de l'intensité. La puissance dissipée s'exprime ainsi $k^{-1}\left|\mathbf{u}\right|^2 = k\left|\nabla \mathrm{p}\right|^2$.

6. Le terme source ν est a priori en $m^3 s^{-1}$ par unité de surface, c'est-à-dire en ms^{-1}, mais il est naturel de le diviser par la distance entre les plaques pour obtenir des s^{-1}, qui est bien à la divergence d'une vitesse

et l'on définit la vitesse V du bord du domaine comme la composante normale de la vitesse fluide (qui en en fait la vitesse elle-même car, la pression étant constante sur la frontière, la vitesse tangentielle est nulle)

$$V = \mathbf{u} \cdot \mathbf{n} = -k \frac{\partial \mathrm{p}}{\partial n}.$$

Gravitation. En mécanique classique, une distribution de masse ρ dans l'espace crée un champ gravitationnel v qui vérifie

$$\Delta v = 4\pi G \rho,$$

où G est la constante de gravitation universelle.

Electrostatique. Le champ électrique associé à un potentiel u s'écrit $\mathbf{E} = -\nabla u$, et dans un milieu dans lequel on a une densité de charge ρ, on a $\nabla \cdot \mathbf{E} = \rho$. On obtient donc encore une fois un problème de type point-selle

$$\begin{cases} \mathbf{E} + \nabla u &= 0 \\ \nabla \cdot \mathbf{E} &= \rho \end{cases}$$

ou en éliminant le champ électrique, un problème de Poisson

$$-\Delta u = \rho.$$

1.5. Élasticité linéaire

On considère ici un matériau élastique subissant des petites déformations au voisinage d'un configuration d'équilibre. On note $\mathbf{u}$ le champ de déplacement, et $\mathbf{e}(\mathbf{u})$ le tenseur des taux de déformations défini par

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

Tenseur des contraintes. Soit $\mathbf{x}$ un point du domaine occupé par le matériau, $\mathbf{n}$ un vecteur unitaire, et $D_\varepsilon(\mathbf{n})$ un disque (ou un segment s'il s'agit de la dimension 2) centré en $\mathbf{x}$, d'aire ε (de longueur ε en dimension 2), et perpendiculaire à $\mathbf{n}$. On note $\mathbf{F}(\varepsilon, \mathbf{n})$ la force exercée sur $D_\varepsilon(\mathbf{n})$ par le matériau situé du côté vers lequel pointe $\mathbf{n}$. Si $\mathbf{F}(\varepsilon, \mathbf{n})/\varepsilon$ tend vers un vecteur $\mathbf{S}$ quand ε tend vers 0, et que la correspondance $\mathbf{n} \longmapsto \mathbf{S}(\mathbf{n})$ est linéaire, alors on appelle tenseur des contraintes au point $\mathbf{x}$, et l'on note $\boldsymbol{\sigma}$, le tenseur qui représente cette correspondance :

$$\mathbf{S}(\mathbf{n}) = \boldsymbol{\sigma} \cdot \mathbf{n}.$$

On s'intéresse ici aux matériaux pour lesquels il existe 2 coefficients (appelés coefficients de Lamé) μ et λ tels que

$$\boldsymbol{\sigma} = \mu(\nabla \mathbf{u} + {}^t\nabla \mathbf{u}) + \lambda\left(\nabla \cdot \mathbf{u}\right).$$

On suppose le matériau considéré soumis à l'action d'un champ de force en volume $\mathbf{f}$. L'équilibre statique du système conduit aux équations de l'élasticité linéaire

$$-\nabla \sigma = -\nabla\left(\mu(\nabla \mathbf{u} + {}^t\nabla \mathbf{u}) + \lambda\left(\nabla \cdot \mathbf{u}\right)\right) = \mathbf{f}.$$

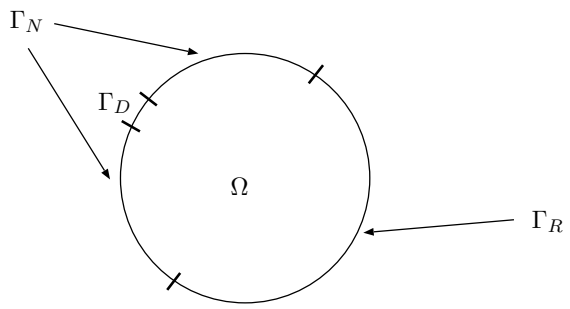


FIGURE 1. Alvéole

Les conditions aux limites peuvent être ici encore de type

(1) Conditions de Dirichlet : on impose le déplacement $\mathbf{u}$

(2) Conditions de Neumann : on impose le tenseur normal des contraintes $\boldsymbol{\sigma} \cdot \mathbf{n}$.

La condition de Neumann correspond en fait à une condition sur le saut de la contrainte normale au travers de l'interface entre le matériau élastique et le milieu extérieur. Ainsi dans le cas d'une frontière libre, c'est à dire dans la réalité en contact avec un milieu considéré comme parfait, c'est à dire au sein duquel le tenseur des contraintes s'écrit $\boldsymbol{\sigma} = -\mathrm{p}_{\mathrm{ext}}\,\mathrm{Id}$, cette continuité s'écrit $\boldsymbol{\sigma} \cdot \mathbf{n} = -\mathrm{p}_{\mathrm{ext}}\mathbf{n}$.

Il est aussi possible de panacher ces conditions. Ainsi pour modéliser une condition de glissement, on écrira $\mathbf{u} \cdot \mathbf{n} = 0$, et $\mathbf{t} \cdot \boldsymbol{\sigma} \cdot \mathbf{n} = 0$ (cisaillement nul).

Energie. L'énergie potentielle élastique stockée dans un matériau de Lamé soumis au champ de déformation $\mathbf{u}$ s'écrit

$$E = \int_\Omega \mu\,|e(\mathbf{u})|^2 + \frac{1}{2}\int_\Omega \lambda(\nabla \cdot \mathbf{u})^2$$

1.6. Exemples de problèmes

1.6.1. Diffusion dans les alvéoles. On considère une alvéole de forme sphérique, dont la frontière se décompose en trois parties (voir figure 1) :

(1) une partie Γ_D, correspondant au raccord avec la bronchiole, où l'on suppose que la concentration est connue (concentration d'oxygène dans l'air que l'on respire) ;

(2) une partie Γ_R correspondant à la zone de la paroi alvéolaire en regard de vaisseaux capillaires, au travers de laquelle l'oxygène va diffuser selon la condition de Robin évoquée ci–dessous ;

(3) une dernière partie Γ_N imperméable à l'oxygène.

On se ramène ainsi à la recherche d'une concentration $u(x,t)$ sur Ω, solution de

$$\left\{\begin{array}{rcll} \dfrac{\partial u}{\partial t} - D\triangle u &=& 0 & \text{dans } \Omega \\ u &=& u_D & \text{sur } \Gamma_D \\ D\dfrac{\partial u}{\partial n} &=& 0 & \text{sur } \Gamma_N \\ \beta u + D\dfrac{\partial u}{\partial n} &=& \beta u_{\text{ext}} & \text{sur } \Gamma_R \end{array}\right. \tag{1.3}$$

Bilan de matière. On peut écrire le bilan d'oxygène en intégrant l'équation sur l'ensemble du domaine Ω, et en intégrant par parties le terme de Laplacien :

$$\frac{d}{dt}\int_\Omega u - \int_\Gamma D\frac{\partial u}{\partial n} = 0.$$

L'intégrale de bord se décompose suivant les 3 composantes de Γ. En utilisant les conditions aux limites, on obtient

$$\frac{d}{dt}\int_\Omega u = \underbrace{\int_{\Gamma_D} D\frac{\partial u}{\partial n}}_{\text{flux à travers } \Gamma_D} + \underbrace{\beta\int_{\Gamma_R}(u_{\text{ext}} - u)}_{\text{flux à travers } \Gamma_R}.$$

On notera que les flux respectifs ne sont pas connus (sauf le flux à travers Γ_N, qui est nul), mais peuvent être calculés à partir de la solution.

1.6.2. Calcul de propriétés effectives. On considère un matériau bidimensionnel occupant un domaine Ω (le carré unité pour fixer les idées). On suppose que la conductivité thermique de ce matériau est non uniforme, donnée par une fonction $k(x)$ de la variable d'espace, de telle sorte que le flux de chaleur s'écrive $J = -k\nabla u$, où u représente le champ de température. On cherche à estimer la conductivité thermique effective de ce matériau. On considère pour cela l'expérience virtuelle suivante : on suppose la température fixée aux bords inférieur et supérieur aux valeurs 0 et 1 respectivement, on suppose les bords latéraux isolés thermiquement, ce qui conduit au problème

$$\left\{\begin{array}{rcll} -\nabla\cdot k\nabla u &=& 0 & \text{in } \Omega \\ u &=& 0 & \text{sur } \Gamma_1 \\ u &=& 1 & \text{sur } \Gamma_3 \\ k\dfrac{\partial u}{\partial n} &=& 0 & \text{sur } \Gamma_2 \bigcup \Gamma_4 \end{array}\right. \tag{1.4}$$

L'objectif est alors d'estimer le facteur de proportionnalité entre le saut de température (ici égal à 1) et le flux de chaleur qui traverse l'échantillon de bas en haut, qui s'exprime

$$\Phi = -\int_{\Gamma_1} k\frac{\partial u}{\partial n}.$$

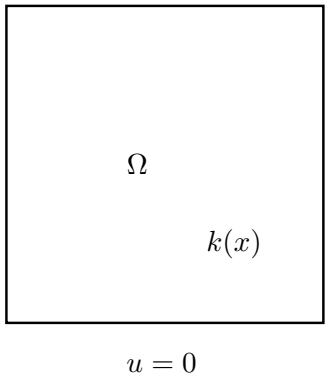


FIGURE 2. Domaine

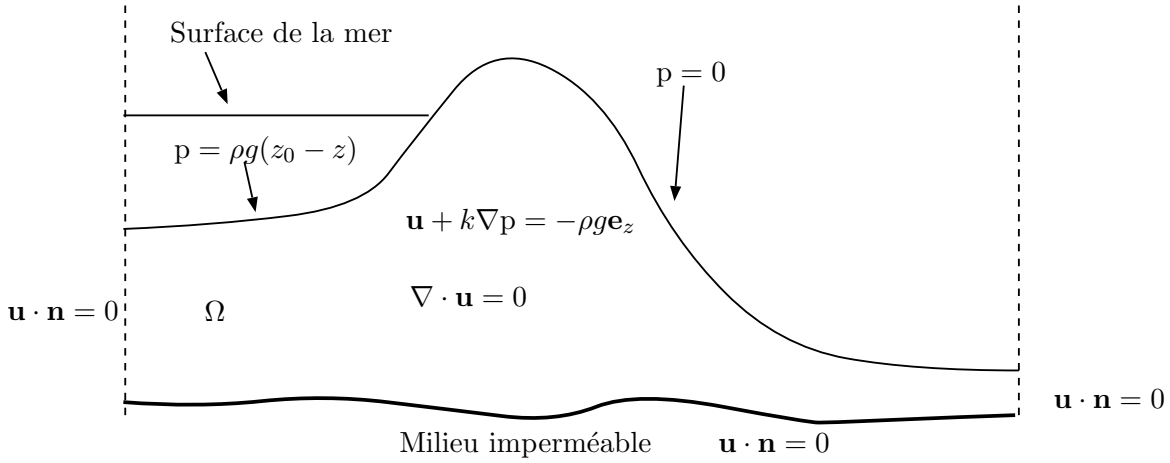


FIGURE 3. Polder

1.6.3. Écoulements en milieu poreux. On considère la situation (représentée sur la figure 3) d'une zone côtière située au dessous du niveau de la mer, protégée par une digue.

On considère la pression atmosphérique constante sur le wdomaine, on la prend égale à 0 (la pression est définie à une constante près, que l'on peut fixer arbitrairement). La pression exercée sur la partie du domaine recouverte par les eaux correspond à la pression hydrostatique. On suppose la partie basse du domaine en contact avec une zone constituée d'un matériau imperméable, et l'on s'intéresse bien sûr à la quantité d'eau susceptible de sourdre vers le polder, zone que l'on cherche à garder sèche.

Chapitre 2

Démarche générale

2.1. Problème de Poisson avec conditions de Dirichlet homogènes

On considère un matériau bidimensionnel conducteur de la chaleur occupant le domaine $\Omega =]0,1[\times]0,1[$. On note $k(x)$ la conductivité thermique au point x, de telle sorte que le flux de chaleur s'écrive $J = -k\nabla u$, où u représente le champ de température. On suppose la conductivité minorée sur le domaine : $k \geq \eta > 0$. On suppose que ce matériau est chauffé, et l'on note f le flux de chaleur injecté par unité de surface.

$$\left\{ \begin{array}{rcll} -\nabla \cdot k\nabla u & = & f & \text{dans } \Omega \\ u & = & 0 & \text{sur } \Gamma \end{array} \right. \tag{2.1}$$

Formulation variationnelle. On obtient [1] la formulation variationnelle de ce problème en multipliant la première équation par une fonction test v régulière qui s'annule sur la partie du bord où la température est imposée. On obtient après intégration par parties

$$\int_\Omega k\nabla u \cdot \nabla v - \int_\Gamma kv\frac{\partial u}{\partial n} = \int fv$$

d'où (les termes de bord s'annulent sur Γ du fait de la nullité de v)

$$\int_\Omega k\nabla u \cdot \nabla v = \int fv.$$

Cette démarche d'élaboration de la formulation variationnelle n'est pas à proprement parler mathématique : ni l'espace dans lequel est censé vivre la solution, ni le sens que l'on peut donner à l'équation de départ, n'ont été précisés. C'est cette formulation variationnelle qui va permettre justement de donner un cadre théorique précis au modèle.

Cadre théorique. Ce problème se met donc sous la forme

$$a(u,v) = \langle \varphi\, , v\rangle \quad \forall v \in V,$$

où $a(\cdot,\cdot)$ est une forme bilinéaire symétrique sur un espace de Hilbert V, et φ une forme linéaire continue sur ce même espace. L'espace V est l'espace de Sobolev $H^1_0(\Omega)$ (voir

1. Cette démarche en elle-même n'est pas mathématique, elle consiste précisément à faire rentrer le problème dans un cadre mathématique. Pour le mathématicien, non seulement le problème (2.1) n'est pas encore bien posé (il n'est pas sous une forme qui permette l'utilisation directe d'un théorème), mais d'une certaine manière il n'est même pas posé (l'espace dans lequel est supposé vivre l'inconnue n'est pas précisé, ni le sens que peuvent avoir les conditions aux limites). Ces remarques peuvent laisser croire que l'obtention de la formulation variationnelle se fait hors de toute règle. Il faut cependant garder à l'esprit qu'un retour (parfaitement mathématisé celui-là) vers l'équation sera nécessaire pour garantir le lien entre le problème initial et la formulation variationnelle.

chapitre 8) des fonction de L^2 dont les dérivées partielles sont aussi dans L^2, et qui sont nulles [2] sur Γ :

Dans le cas où la forme bilinéaire $a(\cdot,\cdot)$ est coercive, c'est à dire (voir définition 6.20) s'il existe $\alpha > 0$ tel que $a(v,v) \geq \alpha |v|^2$ pour tout v dans V, le théorème de Lax Milgram (théorème 6.25) assure l'existence et l'unicité d'une solution dans V.

Cette solution peut être caractérisée comme unique minimiseur de la fonctionnelle

$$J(v) = \frac{1}{2}a(v,v) - \langle \varphi \, , \, v \rangle = \frac{1}{2}\int_\Omega k \, |\nabla v|^2 - \int_\Omega fv.$$

Le point essentiel pour pouvoir utiliser le théorème de Lax-Milgram est la coercivité de la forme bilinéaire, dont nous verrons qu'elle peut être mise à mal pour des matériaux dégénérés (pour le problème de conduction de la chaleur considéré ici, la dégénérescence se produit lorsque la conductivité tend localement vers 0) . Ici, la coercivité de la forme bilinéaire est assurée d'une part par l'hypothèse $k \geq \eta > 0$, et d'autre part par le fait que l'on peut choisir la quantité $(\int |\nabla u|^2)^{1/2}$ comme norme sur l'espace V, grâce à l'un des corollaires de l'inégalité de Poincaré (voir proposition 8.48, page 8.48).

Retour à l'équation de départ. La formulation variationnelle ayant été construite de façon informelle, il est important de préciser en quel sens le problème mis sous forme variationnelle correspond bien au problème initial. Cette étape peut être très délicate dans certains cas (la difficulté dépendant de la régularité de la frontière du domaine, et des conditions aux limites considérées). Le premier pas consiste à établir à partir de la formulation variationnelle que la solution est en fait plus régulière [3] que la régularité naturelle H^1 (qui intervient dans le cadre de l'utilisation du théorème de Lax-Milgram). La solution u est dite solution faible de

$$-\nabla \cdot k \nabla u = f,$$

avec $f \in L^2(\Omega)$. Dans le cas où k est supposé régulier (C^1), la solution appartient en effet à un espace de fonctions plus régulières, l'espace $H^2(\Omega)$ (voir définition 8.21, et la section 8.8.2 pour l'énoncé des théorèmes de régularité), de telle sorte que $\nabla \cdot k \nabla u$ est défini comme fonction de $L^2(\Omega)$, et que l'on peut écrire

$$-\nabla \cdot k \nabla u = f \quad \text{p.p. sur } \Omega.$$

Précisons que l'appartenance à $H^2(\Omega)$ ainsi que l'écriture de l'équation ci-dessus utilisent uniquement la formulation variationnelle pour des fonctions tests à support compact dans Ω (qui sont en particulier nulles au bord).

Les conditions aux limites de Dirichlet sur le bord du domain sont contenues dans l'appartenance de u à l'espace V

Discrétisation en espace. L'approximation de la solution u du problème de départ est basée sur l'introduction d'espaces V_h de fonctions, de dimension finie. Dans le cadre de la méthode des éléments finis dits P^1 (pour polynôme de degré 1), on se donne une suite de triangulations T_h (voir définition 10.9, page 134, pour une définition précise de ce que nous entendons par triangulation), où h est un petit paramètre destiné à tendre vers 0,

2. Le sens que l'on peut donner à l'expression $u_{|\Gamma} = 0$ est précisé dans la section 8.4, page 96.

3. Précisons que ce résultat de régularité interviendra de façon essentielle dans l'analyse d'erreur de la méthode de discrétisation.

qui mesure la finesse de la triangulation. On définit alors V_h comme l'espace des fonctions continues, qui vérifient la condition aux limites, et dont la restriction à chaque triangle de T_h est affine :

$$V_h = \left\{ v_h \in V \,,\ v_{h|K} \text{ est affine sur tout } K \in T_h \right\}.$$

Le problème discret s'écrit

$$\left\{ \begin{array}{l} \text{Trouver } u_h \in V_h \text{ tel que} \\ \displaystyle\int_\Omega k \nabla u_h \cdot \nabla v_h = \int_\Omega f v_h \quad \forall v_h \in V_h. \end{array} \right. \tag{2.2}$$

Formulation matricielle. On numérote $i = 1, 2, \ldots, N_h$ les nœuds de la triangulation qui correspondent à des degrés de liberté (c'est à dire les sommets de T_h qui n'appartiennent pas à Γ). La solution recherchée u_h peut s'écrire

$$u_h = \sum_{j=1}^{N_h} u^j w_j,$$

de telle sorte que (2.2) se ramène au système matriciel (on garde la notation u_h pour désigner le vecteur $(u^1, \ldots, u^{N_h})$

$$A u_h = b_h,$$

où A est une matrice carrée d'ordre N_h , et $b_h \in \mathbb{R}^{N_h}$:

$$A = (a_{ij}) = \left(\int_\Omega k \nabla w_i \cdot \nabla w_j \right) \,,\ b_h = \left(\int_\Omega f w_i \right)_i .$$

Implantation sur `Freefem++` . Le logiciel `Freefem++` permet de calculer u_h en quelques lignes. Précisons que l'assemblage de la matrice et la résolution des systèmes sont gérés par le logiciel sans que l'utilisateur ait à intervenir (si ce n'est pour préciser éventuellement le choix de telle ou telle méthode de résolution). D'autre part, les conditions de Dirichlet non homogènes (conditions $u = 1$ sur Γ_3) ne nécessitent pas l'introduction explicite d'un relèvement de cette condition au bord.

```
int np=50;
mesh Th=square(np,np);

fespace Vh(Th,P1);
Vh u,tu  ;
func k = 1+0.5*sin(y*4*pi) ;
func f = 1 ;
plot(Th,wait=1);

problem Poisson(u,tu)=
    int2d(Th)(k*(dx(u)*dx(tu)+dy(u)*dy(tu)))
   -int2d(Th)(f*v)
   +on(1,2,3,4,u=0);
Poisson  ;  plot(u, wait=1);
```

Estimation d'erreur. ll est important de préciser la confiance que l'on peut avoir dans la précision du calcul. Cette estimation d'erreur se base sur 2 ingrédients.

1) En premier lieu, il s'agit d'établir une inégalité d'approximation du type

$$\inf_{v_h \in V_h} |v_h - u| \leq \varepsilon(h, u),$$

où u est la solution exacte du problème initial, et $\varepsilon(h, u)$ tend vers 0 quand le paramètre de discrétisation h tend lui-même vers 0. Pour le cas des éléments finis d'ordre 1 que nous avons considérés ici, ε est du type $Ch\,\|u\|_{H^2}$, où H^2 désigne l'espace de Sobolev des fonctions de L^2 dont toutes les dérivées secondes sont de carré intégrable. Noter que la régularité de la solution donnée par le théorème d'existence et d'unicité est simplement H^1. Il sera donc nécessaire de montrer que la solution est plus régulière que cela.

2) Le fait que l'estimation d'approximation précédente puisse conduire à une estimation d'erreur sur la solution effectivement calculée (qui a priori n'est pas la meilleure approximation de u par un élément de V_h) se base sur le lemme de Céa (voir chapitre 10), qui utilise encore une fois la coercivité de la forme bilinéaire $a(\cdot, \cdot)$, et s'exprime ici

$$\|u - u_h\| \leq C \inf_{v_h \in V_h} |v_h - u|\,,$$

où C est une nouvelle constante qui dépend des propriétés de la forme bilinéaire. Nous verrons que dans le cas de matériaux inhomogènes cette constante est susceptible d'être très grande, ce qui suggère une dégradation de la précision numérique. La démonstration de ces propriétés fait l'objet du chapitre 10.

Ces propriétés assurent ici que, si l'on considère (T_h) une famille régulière de triangulations de Ω (voir définition 10.11), V_h l'espace d'approximation associé défini précédemment, alors il existe une constante $C > 0$ telle que

$$|u - u_h|_{\Omega,1} \leq Ch\,|f|_{\Omega,0}\,.$$

C'est une application directe de la proposition 8.63, page 107 (ou plus précisément de la proposition 8.65 qui s'applique au cas d'un polyèdre convexe), du théorème d'approximation 10.12, et du lemme de Céa 10.16.

Remarque 2.1. On prendra garde au fait que le lemme de Céa est *non local* (l'estimation de l'erreur par l'erreur d'approximation est globale). En particulier, si la solution a la régularité H^2 sauf au voisinage d'un point (par exemple un coin rentrant), on n'a pas forcément approximation d'ordre 1, même loin du point problématique : la singularité est susceptible de *polluer* l'ensemble de l'approximation.

2.2. Autres conditions aux limites

Conditions de Dirichlet non homogènes. Il est utile de pouvoir prescrire des conditions non nulle au bord du domaine. Reprenons la situation décrite précédemment dans le cas où la condition au bord est $u = g$, où g est une fonction donnée sur Γ. En premier lieu notons qu'il sera impossible d'utiliser l'approche précédente, qui assure l'existence d'une solution dans $H^1(\Omega)$, si g n'est pas la trace d'une fonction de H^1, c'est à dire une fonction qui possède une certaine régularité [4]. Considérons ici le cas d'une fonction g régulière. Une première approche consiste à se ramener au problème précédent en introduisant un

4. Voir remarque 8.76. Il est ainsi illicite de prendre pour g une fonction qui présente des sauts. Plus précisément, s'il est possible de définir la solution d'un tel problème, cela ne peut pas être dans le cadre de l'approche (théorème de Lax Milgram utilisé dans H^1) présentée ci-dessus.

relèvement de g, c'est à dire une fonction sur Ω dont la trace sur Γ est g. Notons $\tilde{g}$ une telle fonction. On cherche alors u sous la forme $\tilde{u} + \tilde{g}$, ce qui conduit au problème de Poisson sur $\tilde{u}$

$$-\Delta\tilde{u} = f + \Delta\tilde{g},$$

avec conditions de Dirichlet homogènes sur $\tilde{u}$.

On peut aussi choisir de rechercher la solution dans l'espace affine V^1 des fonctions de $H^1(\Omega)$ dont la trace s'identifie à g. Le corollaire 6.26, page 73, du Théorème de Lax–Milgram (version affine de ce théorème), permet de gérer directement cette situation.

Conditions de Neuman. On considère la situation (rencontrée dans les exemples du chapitre I) où la dérivée normale est imposée sur une partie de la frontière. Notons Γ_N cette partie, et Γ_D la composante restante, sur laquelle on choisit d'imposer une condition de Dirichlet homogène. Pour fixer les idées, on considère que Ω est le carré unité, et que Γ_N est le bord inférieur. On se donne une donnée $g \in L^2(\Gamma_N)$ sur le bord [5]. Le problème considéré est maintenant (avec $k \equiv 1$)

$$\left\{\begin{array}{rcll} -\Delta u & = & f & \text{in } \Omega \\ u & = & 0 & \text{sur } \Gamma_D \\ \dfrac{\partial u}{\partial n} & = & g & \text{sur } \Gamma_N \end{array}\right. \tag{2.3}$$

On obtient la formulation variationnelle en multipliant par une fonction-test v nulle Γ_D en intégrant par parties, et en remplaçant [6] $\partial u/\partial n$ par g :

$$\int_\Omega \nabla u \cdot \nabla v = \int_\Omega fv + \int_{\Gamma_N} gv.$$

Ce problème se ramène donc à la recherche de $u \in V$ tel que

$$a(u, v) = \langle \varphi \, , \, v \rangle \quad \forall v \in V,$$

avec [7]

$$V = \left\{ u \in H^1(\Omega) \, , \; u_{|\Gamma_D} = 0 \right\}.$$

L'espace V est sous-espace fermé de $H^1(\Omega)$, c'est donc bien un espace de Hilbert, et $a(\cdot, \cdot)$ est une forme bilinéaire continue symétrique. L'intégrale en volume dans le second membre est bien une forme linéaire continue, et

$$\left| \int_{\Gamma_N} gv \right| \; \leq |g|_{L^2(\Gamma_N)} \, |v|_{L^2(\Gamma_N)} \leq C \, |g|_{L^2(\Gamma_N)} \, \|v\|_{H^1(\Omega)} \, ,$$

par continuité de l'application trace, et donc $\varphi \in V'$. Il reste à établir la coercivité de a, ce que permet le corrolaire 8.52, page 103, de l'inégalité de Poincaré généralisée :

$$\int_\Omega |\, \nabla u|^2 \geq \frac{1}{1 + C^2} \left(\int_\Omega u^2 + \int_\Omega |\, \nabla u|^2 \right).$$

5. La question de la régularité de g est un peu délicate. On pourra considérer dans un premier temps $g \in L^2(\Gamma)$, ce qui permet d'obtenir un problème bien posé. En revanche si l'on souhaite démontrer la régularité H^2 de la solution, il est nécessaire de prendre une donnée plus régulière, en l'occurrence $H^{1/2}(\Gamma)$.

6. Il est essentiel de faire disparaître toute trace de $\partial u/\partial n$, car cette quantité n'est pas définie pour des fonctions de H^1. Or la forme bilinéaire impose que l'on se place dans H^1 pour utiliser le théorème de Lax-Milgram.

7. En toute rigueur la condition de Dirichlet sur Γ_D devrait s'écrire en utilisant l'opérateur de trace γ_0. Nous utiserons pourtant dans la suite la notation $u_{|\Gamma_D}$ pour désigner la trace de u sur Γ_D.

Le problème admet donc une unique solution $u \in V$.

Remarque 2.2. On peut choisir de munir V d'une autre norme. Ici, l'inégalité de Poincaré généralisée assure que la semi-norme $|u|_1$ est en fait une norme équivalent à la norme H^1 (avec la partie L^2). On peut donc choisir de munir V de cette norme, et par suite la forme est bien sûr coercive, avec une constante de coercivité égale à 1. Dans ce cas l'existence et l'unicité sont directement données par le théorème de Riez-Fréchet.

Retour à l'équation de départ. Il s'agit de montrer en premier lieu que la solution est H^2, de façon à donner un sens à Δu comme fonction [8]. Cette régularité est assurée sous certaines hypothèses, en particulier ici dans le cas de conditions mixtes dans le cas où le raccord entre les différentes composantes se fait à angle droit (voir remarque 8.66, page 108). Nous supposerons ici que la donnée g a été choisie de telle sorte que cette régularité H^2 soit vérifiée.

Cet exemple va nous permettre de faire la distinction entre condition *essentielle* (conditions de Dirichlet), et condition *naturelle* (de Neuman en l'occurence, mais il pourrait s'agir des conditions de Robin). Dans le premier cas, la condition au bord est dans la définition de l'espace sur lequel on travaille : on a $u(x) = 0$ presque partout sur Γ_D par appartenance de u à V. Les conditions de Neuman ont en revanche disparu en tant que telles du problème sous sa forme variationnelle, il est important de vérifier qu'elles sont bien vérifiées dans un certain sens par la solution. On utilise pour cela la régularité H^2 de la solution. On considère alors la formulation variationnelle pour des fonctions-test régulières qui s'annulent sur Γ_D, mais pas forcément sur Γ_N. On utilise alors la formule de Green (voir proposition 8.38), ce qu'autorise la régularité H^2 de la solution u, pour obtenir

$$\int_\Omega (-\Delta u)\, v + \int_{\Gamma_N} \frac{\partial u}{\partial n} v - \int_{\Gamma_N} gv = \int_\Omega fv.$$

Comme l'équation de Poisson est vérifiée presque partout, il reste

$$\int_{\Gamma_N} \left(\frac{\partial u}{\partial n} - g \right) v = 0.$$

La fonction v pouvant être choisie arbitrairement, on en déduit $\partial_n u = g$ presque partout sur Γ_N.

Discrétisation en espace. La discrétisation en espace ne change pas significativement du cas Dirichlet homogène, si ce n'est que les points du maillage situés sur Γ_N correspondent maintenant à des degrés de liberté, et que le second membre contient des termes provenant d'intégrales surfaciques impliquant les fonctions-test associées à ces nouveaux degrés de liberté :

$$b_h = \left(\int_\Omega f v_h + \int_{\Gamma_N} g w_i \right)_i .$$

8. Il existe une autre manière (que nous ne privilégierons pas ici) de donner un sens à l'équation de Poisson sans l'aide d'aucun théorème de régularité (voir section 8.9.2, page 111). La formulation variationnelle assure que ∇u admet une divergence faible L^2. On peut donc donner un sens à Δu comme la divergence faible de ∇u, en gardant à l'esprit qu'il s'agit d'une notation globale, et qu'en particulier les dérivées secondes ne sont pas nécessairement définies comme des fonctions de L^2. On peut pousser la démarche jusqu'à donner un sens à $\partial u/\partial n$ comme la trace normale du champ de vecteur $\nabla u \in H_{\mathrm{div}}$ (voir remarque 8.81), page 111). Cette trace est alors définie dans un sens faible, ce qui interdit par exemple l'écriture $\partial_n u = g$ p.p.

Chapitre 3

Minimisation sous contrainte

3.1. Préliminaires, introduction

Considérons une fonctionnelle J de $\mathbb{R}^n$ dans $\mathbb{R}$, lisse (au moins continûment différentiable), que l'on cherche à minimiser sur un espace affine K de $\mathbb{R}^n$, dont l'espace vectoriel sous-jacent K_0 est défini comme le noyau d'une matrice $B \in \mathcal{M}_{mn}(\mathbb{R})$:

$$K = U + \ker B.$$

Notons u un extremum de J sur K. Pour tout $h \in K_0$ (donc tel que $u + \mathbb{R}h \subset K$), tout $t \in \mathbb{R}$, on a

$$J(u+th) \geq J(u)\,, \quad \text{d'où} \quad t\nabla J \cdot h + o(t) \geq 0 \quad \forall t \in \mathbb{R}.$$

On a donc $\nabla J \cdot h = 0$ pour tout $h \in \ker B$, c'est à dire

$$\nabla J \subset (\ker B)^{\perp} = \mathrm{Im}(B^{\star}).$$

En conséquence il existe $\lambda \in \mathbb{R}^m$ tel que le couple (u, λ) vérifie

$$\begin{aligned} \nabla J(u) + B^{\star}\lambda &= 0 \\ Bu &= BU. \end{aligned}$$

Dans le cas où J est une fonctionnelle quadratique, et où la contrainte est linéaire ($U = 0$) :

$$v \in \mathbb{R}^N \longmapsto J(v) = \frac{1}{2}\,(Au, v) - (b, v),$$

avec A matrice symétrique, on obtient le système matriciel

$$A + B^{\star}\lambda = b \tag{3.1}$$

$$Bu = 0. \tag{3.2}$$

Remarque 3.1. On parle d'un problème sous forme point-selle. En effet, on peut vérifier (voir proposition 9.7 pour le cas de la dimension infinie) que si l'on introduit le Lagrangien du problème

$$L(v, \mu) = J(v) + Bu \cdot \mu,$$

alors (u, λ) est solution du système précédent si et seulement si le couple est point-selle pour L, c'est-à-dire si

$$L(u, \mu) \leq L(u, \lambda) \leq L(v, \lambda) \quad \forall v \in \mathbb{R}^n\,,\ \mu \in \mathbb{R}^m.$$

Interprétation des multiplicateurs de Lagrange.

Considérons une chaîne horizontale de $n+1$ masses 0, 1, 2, ..., n, reliées entre elles (0 reliée à 1, 1 à 2, etc...) par des ressorts de longueur au repos nulle et de raideur k. Les

positions de ces masses sont représentées par le vecteur position $(x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1}$. L'énergie potentielle du système s'écrit

$$J(\mathbf{x}) = \frac{1}{2}k\sum_{i=1}^{n} |x_i - x_{i-1}|^2 = \frac{1}{2}k(A\mathbf{x}, \mathbf{x}),$$

où A est (à une constante multiplicative près) la matrice du Laplacien discret avec conditions de Neuman. Tout point diagonal $(x, x, \ldots, x)$ de $\mathbb{R}^{n+1}$ minimise cette énergie. On s'intéresse maintenant à la situation où la masse 0 est fixée au point $x_0 = 0$, et la masse n au point $x_n = L > 0$. Il s'agit donc maintenant de minimiser J sur l'espace affine

$$E = \{\mathbf{x}\,,\ x_0 = 0\,,\ x_n = L\} = X + \ker B\,, \quad \text{avec } B\ :\ \mathbf{x} \in \mathbb{R}^{n+1} \longmapsto (x_0, x_n) \in \mathbb{R}^2.$$

La matrice B s'écrit

$$B = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

D'après ce qui précède, il existe donc $\lambda = (\lambda_0, \lambda_1) \in \mathbb{R}^2$ tel que (3.1) soit satisfaite. Écrivons les première et dernière lignes de ce système :

$$\begin{aligned} k(x_0 - x_1) + \lambda_0 &= 0 \\ k(-x_{n-1} + x_n) + \lambda_1 &= 0. \end{aligned}$$

Ces relations expriment l'équilibre des masses extrêmales, et permettent d'interpréter $-\lambda_0$ (resp. $-\lambda_1$) comme la force exercée par le support en 0 sur la masse 0 (resp. par le support en 1 sur la masse n). On peut préciser la configuration minimisante en notant que, pour $i = 1, \ldots, n-1$, on a

$$x_{i+1} - x_i = x_i - x_{i-1},$$

de telle sorte que les longueurs des ressorts sont toutes identiques, égales L/n, et ainsi

$$\lambda_0 = -\lambda_1 = kL/n.$$

Cet exemple permet aussi d'illustrer et d'interpréter mécaniquement une méthode très utilisée en pratique, la méthode de pénalisation. Elle consiste à relaxer la contrainte, et à ajouter à la fonctionnelle à minimiser un terme supplémentaire qui pénalise la non vérification des contraintes. Dans l'exemple considéré, elle consiste à considérer la fonctionnelle

$$J_\varepsilon(\mathbf{x}) = \frac{1}{2}k\sum_{i=1}^{n} |x_i - x_{i-1}|^2 + \frac{1}{2\varepsilon}\left(|x_0|^2 + |x_n - L|^2\right).$$

Noter que cela revient à supposer les masses 0 et n attachées à des supports respectivement en 0 et L par des ressorts dont la raideur $1/\varepsilon$ tend vers l'infini. Ce problème rentre dans le cadre de la section 9.3, page 124. On peut ainsi montrer la convergence des minimiseurs $\mathbf{x}_\varepsilon$ vers la solution du problème contraint $\mathbf{x}$.

Noter que la remarque 9.23 permet également d'affirmer que la force exercée par exemple par le ressort en 0 de raideur $1/\varepsilon$, force qui vaut $-x_0^\varepsilon/\varepsilon$, tend vers le multiplicateur de Lagrange $-\lambda_0$ introduit ci-dessus.

Commentaires.

Précisons les difficultés liées à cette approche.

En premier lieu, noter que la manière d'écrire les contraintes n'est pas unique. On peut rajouter par exemple $x_n - x_0 = L$. On aura alors un troisième multiplicateur de Lagrange, qui correspondrait à la tension (positive ou négative) au sein d'une barre rigide qui relierait les points extrêmaux. La non unicité met en évidence le fait concret qu'il est a priori impossible de prévoir la tension effective au sein de ce raidisseur, ainsi que l'effort au niveau des supports. Dans la réalité, il peut se produire par exemple que seuls les supports fixes soient actifs, jusqu'à ce que l'un d'entre eux se détériore et finisse par lâcher, pour être relayé par le raidisseur, sans que rien ne transparaisse au niveau de ce que nous appelerons par la suite les variables primales (i.e. les positions des ressorts). On parlera dans un contexte mécanique de situation *hyperstatique* (il y a trop de contrainte), par opposition aux situations *isostatiques* (jeu minimal de contraintes assurant l'unicité des multiplicateurs de Lagrange). On notera qu'il y a un lien fort entre l'expression mathématique d'un ensemble de contraintes et les moyens que l'on pourrait se donner pour les réaliser en pratique. D'autre part l'approche de pénalisation est basée sur une réalisation approchée des contraintes, ce qui est en général assez réaliste dans un contexte mécanique.

L'exemple du pont rigide entre les points extrémaux évoqué plus haut est un peu caricatural car la troisième contrainte est manifestement redondante. Dans des situations plus compliquées pourtant, il peut ne pas être aisé de supprimer des contraintes pour parvenir à un jeu minimal équivalent qui assurera l'unicité des multiplicateurs de Lagrange. D'autre part certains systèmes réels très courants conduisent à une non unicité. Ainsi, pour la chaise à 4 pieds posés sur un sol horizontal, on aura un multiplicateur de Lagrange associé à chacun des 4 contacts avec le sol. Or 3 contacts suffisent pour que la chaise ne rentre pas dans le sol (nous ne considérons pas ici les questions de stabilité). Il est ainsi impossible de prévoir, même si l'on dispose de toutes les informations, quel est l'effort au niveau de chacun des pieds d'une chaise parfaitement équilibrée. Noter également que ces efforts sont susceptibles de changer au cours du temps de façon très irrégulière.

Ce premier problème apparaît pour des situations extrêmement simples, pour un nombre fini de degrés de liberté (et de contraintes). Le second problème, qui affecte l'existence-même des multiplicateurs de Lagrange, est lié au fait que, en dimension infinie, on a seulement $\mathrm{Im}(B^\star) \subset (\ker B)^\perp$ avec inclusion stricte si l'image de $B^\star$ n'est pas fermée. Dans ce dernier cas, on peut ne pas avoir existence du multiplicateur de Lagrange. On verra pourtant que de telles formulations, mal posées dans un certain sens, peuvent être utilisées numériquement de façon tout à fait rigoureuse pour approcher la solution du problème de minimisation (qui lui, dans les situations que nous rencontrerons, est toujours bien posé).

3.2. Cadre théorique

Avant d'aborder dans la section suivante un exemple de problème de minimisation sous contrainte, nous donnons ici une vue d'ensemble de la démarche de formalisation mathématique. Nous allons considérer des problèmes consistant à minimiser une fonctionnelle quadratique

$$J(v) = \frac{1}{2}a(v,v) - \langle \varphi\,,\, v\rangle,$$

sur un sous-espace vectoriel K d'un espace de Hilbert V. La forme bilinéaire $a(\cdot,\cdot)$ est supposée symétrique, continue et coercive. Le théorème de Lax-Milgram 6.25 assure l'existence

et l'unicité d'une solution à ce problème, dont la formulation variationnelle est

$$a(u,v) = \langle \varphi \, , \, v \rangle \quad \forall v \in K,$$

dont on déduit immédiatement que la forme linéaire ξ définie par

$$\langle \xi \, , \, v \rangle = \langle \varphi \, , \, v \rangle - a(u,v),$$

est dans $K^\perp$. Si l'on suppose K de la forme $\ker B$, où B est une application linéaire continue de V dans un espace de hilbert Λ, on peut se demander si la démarche menée en dimension finie au début de ce chapitre est valide. La proposition 7.18, page 83, assure l'identité

$$(\ker B)^\perp = \overline{\mathrm{Im}(B^\star)}.$$

Dans le cas où l'image de B est fermée (et donc celle de $B^\star$, voir ptoposition 7.19), on aura donc existence de $\lambda \in \Lambda$ tel que ξ s'écrive $B^\star\lambda$, ce que l'on écrira en pratique sous la forme suivante : il existe $\lambda \in \Lambda$ tel que

$$\left\{ \begin{array}{lcll} a(u,v) + (\lambda, Bv) & = & \langle \varphi \, , \, v \rangle & \forall v \in V \\ (\mu, Bu) & = & 0 & \forall \mu \in \Lambda. \end{array} \right.$$

Si B est surjectif, alors l'injectivité de B assure l'unicité de λ.

Remarque 3.2. (Intérêt des problèmes mal posés)
Dans la suite, nous serons parfois amenés à considérer des problèmes mal posés, c'est à dire pour lesquels on n'a pas existence du multiplicateur de Lagrange, et même à bâtir des schémas numériques sur ces formulations a priori mal posées. Précisons, bien que cela dépasse le cadre de ce cours, que de telle hérésies apparentes peuvent être justifiées tout à fait rigoureusement : en effet, si λ n'est pas défini, la forme linéaire ξ (qui exprime l'action du multiplicateur de Lagrange) est, elle, parfaitement définie, et l'on peut espérer [1] construire une suite de multiplicateurs approchés λ_h tels que la forme linéaire associé

$$v \longmapsto \langle \lambda_h \, , \, Bv \rangle$$

tende vers ξ.

Nous présentons ici un certain nombre de situations, dans le cadre des espaces de Sobolev, qui peuvent être mises en forme dans le cadre du formalisme introduit dans les chapitres précédent. Nous commençons par quelques problèmes pour lesquels la contrainte est de nature géométrique (prise en compte d'obstacles dans le domaine, dans un sens qui sera précisé par la suite), qui présentent l'avantage de pouvoir être traités (jusqu'à l'implémentation numérique effective) par toutes les méthodes que nous avons présentées. Nous poursuivrons par l'étude du problème de Darcy, puis du problème de Stokes incompressible, pour lequel nous nous concentrerons sur l'approche par dualité.

1. On peut établir rigoureusement des résultats de convergence de la méthode numérique associée, pour la partie primale, voir proposition 11.7, page 146.

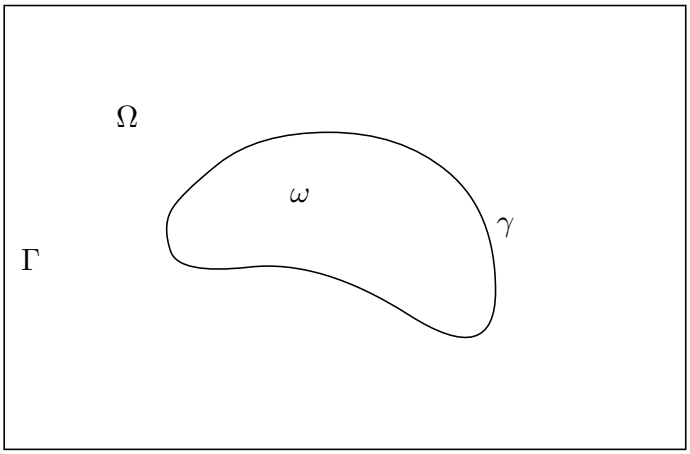


FIGURE 1. Géométrie.

3.3. Problème de Poisson sur un domaine perforé

On considère un domaine borné Ω du plan $\mathbb{R}^2$, et ω un sous-domaine fortement inclus dans Ω, c'est-à-dire que $\overline{\omega} \subset \Omega$. On notera Γ la frontière de Ω, et γ la frontière de ω (voir figure 1). On fait les hypothèses de régularité suivantes [2] :

(1) γ est de classe C^2
(2) Ω est un polyèdre convexe, ou bien sa frontière Γ est C^2.

Étant donnée f une fonction de $L^2(\Omega \setminus \overline{\omega})$, on s'intéresse au problème suivant :

$$\left\{ \begin{array}{rcll} -\triangle u & = & f & \text{dans} \quad \Omega \setminus \overline{\omega} \\ u & = & 0 & \text{sur} \quad \partial\Omega \cup \partial\omega. \end{array} \right. \tag{3.3}$$

Nous allons considérer dans un premier temps l'approche directe, qui consiste simplement à formuler le problème dans un espace fonctionnel qui vérifie les conditions aux limites à la fois sur $\partial\Omega$ et $\partial\omega$. D'un point de vue théorique, il est superflu de considérer d'autres formulations que celle-ci. En revanche sur le plan pratique, en vue d'une implémentation effective sur machine, cette première approche nécessite un maillage du domaine dans lequel vit effectivement la solution, ce qui exclut par exemple l'utilisation d'un maillage cartésien régulier. Il peut donc être intéressant d'introduire des formulations qui font intervenir un champ défini sur le domaine Ω en entier.

3.3.1. Approche directe. L'approche variationnelle directe est basée sur la fonctionnelle

$$\begin{array}{rcl} H_0^1(\Omega \setminus \overline{\omega}) & \longrightarrow & \mathbb{R} \\ v & \longmapsto & J(v) = \dfrac{1}{2} \displaystyle\int_{\Omega\setminus\overline{\omega}} |\nabla v|^2 - \int_{\Omega\setminus\overline{\omega}} fv, \end{array}$$

2. On notera que l'on exclut le cas où ω serait un polyèdre convexe. En effet, il est important de contrôler la régularité de la solution au problème de Poisson sur $\Omega \setminus \overline{\omega}$. Pour ce domaine, les sommets (ou arêtes) de ω sont vus comme des coins rentrant, de telle sorte que la solution ne sera pas en général dans $H^2(\Omega \setminus \overline{\omega})$.

que l'on cherche à minimiser sur l'espace $H_0^1(\Omega\backslash\overline{\omega})$ lui-même. Le théorème de Lax-Milgram assure immédiatement l'équivalence entre ce problème de minimisation et la formulation variationnelle suivante

$$\int_{\Omega\backslash\overline{\omega}} \nabla u \cdot \nabla v = \int fv \quad \forall v \in H_0^1(\Omega \setminus \overline{\omega}).$$

3.3.2. Pénalisation. On se place maintenant sur l'espace $V = H_0^1(\Omega)$, et l'on prolonge f par 0 à l'intérieur de ω, en conservant la notation f pour cette fonction de $L^2(\Omega)$. On réécrit le problème initial comme un problème de minimisation sous contrainte dans le nouvel espace

$$(\mathcal{P}) \quad \begin{cases} u \in K = \left\{ v \in V\,,\ v_{|\omega} = 0 \right\}, \\ J(u) = \inf\limits_{v\in K} J(v), \end{cases}$$

où J est la fonctionnelle définie par

$$J(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv.$$

Nous allons considérer deux manières de formuler ce problème par pénalisation, une « bonne » (rarement utilisée en pratique), et une « mauvaise » (très utilisée). Nous commencerons par la méthode qui permet d'utiliser au mieux les outils théoriques, puis nous décrirons la méthode la plus courante (qui est aussi la plus naturelle et la plus facile à mettre en œuvre).

Pénalisation fermée. La démarche proposée dans la section 9.3 consiste par exemple à écrire l'espace des u admissibles comme

$$K = \left\{ v \in V\,,\ \int_\omega (uv + \nabla u \cdot \nabla v) = 0\,,\ \forall v \in V \right\},$$

et à introduire, pour $\varepsilon > 0$, la fonctionnelle

$$J_\varepsilon^1(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv + \frac{1}{2\varepsilon}\int_\omega \left(v^2 + |\nabla v|^2\right).$$

Le problème pénalisé s'écrit

$$(\mathcal{P}_\varepsilon) \quad \begin{cases} u \in V, \\ J_\varepsilon^1(u) = \inf\limits_{v\in V} J_\varepsilon(v), \end{cases}$$

Notons u_1^ε le minimiseur de J_ε^1 sur V. On sait que u_1^ε tend vers u dans $H_0^1(\Omega)$, et l'on peut préciser la vitesse de convergence.

Proposition 3.3. La distance de u_1^ε à u vérifie

$$\|u_1^\varepsilon - u\|_{H^1} \leq C\varepsilon.$$

Démonstration : Il suffit pour cela de remarquer que la fonctionnelle pénalisée s'écrit

$$b(u, v) = (\Psi u, \Psi v)\,,$$

où Ψ est l'opérateur de restriction d'une fonction de $H_0^1(\Omega)$ à ω. Cet opérateur étant surjectif (car ω est régulier), la proposition 9.31, page 129, assure l'estimation annoncée. □

Pénalisation non fermée. Nous considérons ici l'écriture de K sous la forme

$$K = \left\{ v \in V\,,\ \int_\omega vw = 0\,,\ \forall w \in V \right\}.$$

La fonctionnelle pénalisée s'écrit

$$J^0_\varepsilon(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv + \frac{1}{2\varepsilon}\int_\omega v^2.$$

Notons u^ε_0 la solution du problème pénalisé. On sait que u^ε_0 tend vers u fortement dans V, mais les arguments utilisés précédemment pour quantifier la vitesse de convergence ne sont plus utilisables car la restriction d'une fonction de V à ω en tant que fonction de $L^2(\omega)$ n'est pas à image fermée (son image est dense et l'application n'est pas surjective). L'exercice 3.4, page 41, permet de vérifier sur un cas particulier que la convergence peut effectivement être plus lente que ε.

Pénalisation frontière. Une alternative consiste à travailler sur un nouvel espace contraint qui ne fait intervenir que les valeurs de la fonction sur la frontière de ω. On introduit ainsi et l'on considère la fonctionnelle

$$J_\varepsilon(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv - \frac{1}{2\varepsilon}\int_\gamma v^2. \tag{3.4}$$

Bien que l'on ne soit pas dans le cas d'une pénalisation fermée, au sens indiqué précédemment, la méthode converge à l'ordre 1. En effet, on peut expliciter le ξ de la proposition 9.2.

Proposition 3.4. On note $u \in V = H^1_0(\Omega)$ le prolongement par 0 dans ω de la solution du problème (3.3). On note toujours f le prolongement par 0 du second membre dans ω. On a alors

$$\int_\Omega \nabla u \cdot \nabla v + \langle \xi\,, v\rangle = \int_\Omega fv \quad \forall v \in V, \tag{3.5}$$

avec

$$\langle \xi\,, v\rangle = -\int_\gamma \frac{\partial u}{\partial n} v,$$

où la dérivée normale est prise du côté extérieur à ω.

Démonstration : On définit la forme ξ par

$$\langle \xi\,, v\rangle = -\int_\Omega \nabla u \cdot \nabla v + \int_\Omega fv \quad \forall v \in V.$$

Pour toute fonction test v, on intègre maintenant par parties la première intégrale de (3.5) sur chacun des sous-domaines ω et $\Omega \setminus \overline{\omega}$, où u est H^2 (on ne peut pas intégrer par parties globalement, car u n'est pas H^2 sur Ω). Il vient

$$-\int_{\Omega\setminus\overline{\omega}} v\triangle u + \int_\gamma \frac{\partial u}{\partial n} v + \langle \xi\,, v\rangle = \int_\Omega fv \quad \forall v \in V,$$

d'où, comme u est solution de $-\triangle u = f$ sur $\Omega \setminus \overline{\omega}$,

$$\langle \xi\,, v\rangle = -\int_\gamma \frac{\partial u}{\partial n} v,$$

qui est l'expression annoncée. □

Proposition 3.5. On note u^ε l'unique fonction de $V = H^1_0(\Omega)$ qui réalise le minimum de la fonctionnelle J_ε définie par (3.4), et $u \in V = H^1_0(\Omega)$ le prolongement par 0 dans ω de la solution du problème (3.3). On a convergence à l'ordre 1 de u^ε vers u.

DÉMONSTRATION : La proposition 3.4 donne l'expression explicite de ξ. Comme u est dans $H^2(\Omega \setminus \overline{\omega})$, sa dérivée normale est dans $L^2(\gamma)$. On peut donc utiliser la proposition 9.24, page 126, qui assure la convergence à l'ordre 1. □

3.3.3. Formulation point-selle. On pourrait déduire un peu hâtivement de la proposition (3.5) qu'une formulation point-selle pertinente du problème est nécessairement basée sur une expression de la contrainte exprimée sur la frontière de ω. Ceci n'est vrai que si l'on exprime les dualités par des produits de type L^2. Ainsi pour la première formulation que nous allons considérer, le multiplicateur de Lagrange (défini de façon unique) est identifié à une fonction qui vit dans ω, et non pas sur sa frontière. Ceci n'est pas incompatible avec ce qui précède car la dualité considérée fait intervenir des dérivées (le gradient) qui délocalisent l'action du multiplicateur.

Introduisons $\Lambda = H^1(\omega)$, et l'opérateur B de V dans Λ qui à une fonction de $H^1_0(\Omega)$ associe sa restriction à ω en tant que fonction de $H^1(\omega)$. On écrit (dans un premier temps sous une forme semi-abstraite, c'est-à-dire sans identifier les éléments du dual de Λ à des fonctions)

$$K = \{v \in V\,,\ \langle \mu\,, Bv\rangle = 0 \quad \forall \mu \in \Lambda\}\,.$$

Le Lagrangien s'écrit

$$L(v,\mu) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv + \langle \mu\,, Bv\rangle,$$

et le problème s'écrit sous la forme variationnelle suivante

$$\left\{\begin{array}{lcll} \displaystyle\int_\Omega \nabla u\cdot\nabla v \ + \ \langle \lambda\,, Bv\rangle & = & \displaystyle\int_\Omega fv & \forall v \in V \\ \langle \mu\,, Bu\rangle & = & 0 & \forall \mu \in \Lambda'. \end{array}\right. \tag{3.6}$$

L'opérateur B étant surjectif (car la frontière de ω est régulière), l'approche menée dans la section **??** conduit à un problème bien posé, comme l'exprime la proposition suivante.

Proposition 3.6. Il existe un unique couple $(u,\lambda) \in V \times \Lambda'$ point-selle de L, ou de façon équivalente, solution de (3.6).

Le multiplicateur de Lagrange λ est pour l'instant défini de façon abstraite, c'est-à-dire comme forme linéaire sur $\Lambda = H^1(\omega)$, et non pas comme une fonction. Pour écrire λ sous forme de fonction (et exhiber une formulation variationnelle directement exploitable numériquement), l'approche la plus naturelle consiste à utiliser l'identification de Riez-Fréchet. On garde la notation λ pour désigner une fonction de $H^1(\omega)$. La proposition précédente assure donc l'existence et l'unicité d'un multiplicateur de Lagrange (en tant que fonction) $\lambda \in \Lambda$ tel que (u,λ) est solution de

$$\left\{\begin{array}{lcll} \displaystyle\int_\Omega \nabla u\cdot\nabla v \ + \ \int_\omega \lambda v + \int_\omega \nabla\lambda\cdot\nabla v & = & \int_\Omega fv & \forall v \in V \\ \displaystyle\int_\omega \mu u + \int_\omega \nabla\mu\cdot\nabla u & = & 0 & \forall \mu \in \Lambda. \end{array}\right. \tag{3.7}$$

Mais, comme dans le cas de la pénalisation, l'approche la plus couramment menée est basée sur la formulation point-selle mal posée que nous allons décrire à présent. On écrit l'ensemble admissible de la façon la plus simple

$$K = \left\{ v \in V \,,\ \int_B v\mu = 0 \quad \forall \mu \in L^2(\omega) \right\},$$

de telle sorte que l'espace des multiplicateurs de Lagrange est $\Lambda = L^2(\omega)$. Le Lagrangien s'écrit

$$L(v,\mu) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv + \int_\omega v\mu,$$

et le problème s'écrit sous la forme variationnelle suivante

$$\left\{ \begin{array}{lcll} \displaystyle\int_\Omega \nabla u \cdot \nabla v \ + \ \int_\omega v\lambda & = & \displaystyle\int_\Omega fv & \forall v \in V \\ \displaystyle\int_\omega u\mu & = & 0 & \forall \mu \in \Lambda. \end{array} \right. \tag{3.8}$$

Cette manière en quelque sorte canonique d'écrire la contrainte de façon duale conduit à un problème mal posé. En effet, l'opérateur B correspondant est l'application qui à une fonction de $V = H^1_0(\Omega)$ associe sa restriction à ω, vue comme une fonction de $L^2(B)$. Or cette application n'est pas à image fermée : son image est dense sans que l'application soit surjective. L'existence d'un multiplicateur de Lagrange (et donc d'une solution à (3.8)) n'est donc pas assurée [3].

En revanche, si l'on discrétise en espace, comme on le verra plus loin, l'existence d'un multiplicateur de Lagrange discret λ_h est automatiquement assurée du fait de la dimension finie (toutes les applications linéaires sont alors à image fermée). Il est possible que la suite des λ_h ne converge dans aucun espace de fonction, mais comme seule la composante primale du point-selle nous intéresse, cette méthode peut être convergente pour ce qui est de cette composante primale.

Exercice 3.1. On prend maintenant $\Lambda = H^1(\omega)$, B est toujours l'opérateur de restriction à ω, et l'on considère la formulation variationnelle (3.8) associée à ce nouvel espace. Pourquoi ce nouveau problème est-il lui-même mal posé, alors que l'application de restriction est maintenant surjective ?

3.4. Obstacle de conductivité infinie

On considère comme précédemment un domaine Ω du plan, et ω un sous-domaine fortement inclus dans Ω, c'est-à-dire que $\overline{\omega} \subset \Omega$. Le problème que nous allons considérer maintenant est issu du modèle physique suivant. On considère une plaque conductrice de la chaleur, dont on suppose que les bords sont à température nulle, et l'on suppose qu'une partie de cette plaque (qui correspondra au sous-domaine ω) a une conductivité infinie, de telle sorte que la température y est uniforme. On suppose qu'on chauffe la plaque sur la partie où la température est finie. On cherche ainsi un champ de température solution de

3. On peut même affirmer qu'en général on n'a pas existence. En effet s'il existe $\lambda \in L^2(\omega)$ tel que (3.8) soit vérifiée, alors u est dans $H^2(\Omega)$. Or, si l'on peut espérer que u soit régulière sur $\Omega \setminus \overline{\omega}$ ainsi que sur ω, on n'a pas raccord des dérivées normales le long de la frontière de ω, ce qui serait pourtant le cas u était dans $H^2(\Omega)$.

l'équation de la chaleur, dans $\overline{B} \subset \Omega$, tel que la température est constante sur la frontière de B, et tel que le flux de chaleur à travers cette frontière est nul [4].

On se donne donc f une fonction de $L^2(\Omega \setminus \overline{\omega})$, et l'on s'intéresse au problème suivant :

$$\left\{ \begin{array}{rcll} -\triangle u & = & f & \text{dans } \Omega \setminus \overline{\omega} \\ u & = & 0 & \text{sur } \partial\Omega \\ u & = & U & \text{sur } \partial\omega \\ \displaystyle\int_{\partial\omega} \frac{\partial u}{\partial n} & = & 0, & \end{array} \right. \tag{3.9}$$

où U est une constante réelle dont la valeur est inconnue.

3.4.1. Approche directe, minimisation sous contrainte. On introduit l'espace

$$H^1_C(\Omega \setminus \overline{\omega}) = \left\{ u \in H^1(\Omega \setminus \overline{\omega}),\ u = 0 \text{ sur } \partial\Omega,\ u = \text{cste sur } \partial\omega \right\}.$$

L'approche variationnelle directe est basée sur la fonctionnelle

$$\begin{array}{rcl} H^1_C(\Omega \setminus \overline{\omega}) & \longrightarrow & \mathbb{R} \\ v & \longmapsto & J(v) = \dfrac{1}{2} \displaystyle\int_{\Omega\setminus\overline{\omega}} |\nabla v|^2 - \int_{\Omega\setminus\overline{\omega}} fv, \end{array}$$

Le problème $\mathcal{P}$ consiste donc à minimiser J sur $H^1_C(\Omega \setminus \overline{\omega})$. On notera que la condition de flux nul a disparu. Il s'agit en fait d'une condition dite « naturelle », qui dérive du problème de minimisation, comme le précise la proposition suivante.

Proposition 3.7. Soit $u \in H^1_C(\Omega \setminus \overline{\omega})$ la fonction qui minimise la fonctionnelle J sur $H^1_C(\Omega \setminus \overline{\omega})$. Alors u est solution du problème (3.9).

Démonstration : On note U la valeur de u sur la frontière de ω, et l'on construit un relèvement $\tilde{U}$ de U, de régularité C^2, à support compact dans Ω. La fonction $u - \tilde{U}$ est dans $H^1_0(\Omega \setminus \overline{\omega})$, et c'est la solution faible de l'équation

$$-\triangle w = f + \triangle \tilde{U},$$

avec conditions de Dirichlet homogènes. C'est donc un élément de $H^2(\Omega\setminus\overline{\omega})$, et par suite u lui-même a une régularité H^2. On considère maintenant des fonctions-test dans $H^1_0(\Omega\setminus\overline{\omega})$. Par intégration par parties, on obtient $-\triangle u = f$ dans $\Omega \setminus \overline{\omega}$. Pour retrouver la condition de flux nul à travers l'interface, on prend maintenant une fonction test non nulle sur γ, qui prend par exemple la valeur 1. On utilise de nouveau la formule de Green pour obtenir

$$-\int_{\Omega\setminus\overline{\omega}} v\triangle u + \int_\gamma \frac{\partial u}{\partial n} v = \int fv,$$

d'où

$$\int_\gamma \frac{\partial u}{\partial n} = 0.$$

□

4. Ce modèle présente peu d'intérêt en lui-même. En revanche il contient, sous une forme plus abordable, l'essentiel des difficultés que nous rencontrerons lors de la modélisation de particules rigides dans un fluide.

Le problème de minimisation sous contrainte associé s'écrit de la façon suivante : on introduit

$$K = \left\{ u \in H^1(\Omega)\,,\ u = 0 \text{ sur } \partial\Omega\,,\ u = \text{cste dans } \omega \right\},$$

et l'on considère le problème de minimisation sur K de la fonctionnelle

$$\begin{array}{rcl} V & \longrightarrow & \mathbb{R} \\ v & \longmapsto & J(v) = \dfrac{1}{2}\displaystyle\int_{\Omega\setminus\overline{\omega}} |\nabla u|^2 - \int_{\Omega\setminus\overline{\omega}} fv, \end{array}$$

3.4.2. Pénalisation. La formulation pénalisée est ici particulièrement naturelle puisqu'elle est directement liée au modèle physique proposé. Nous allons simplement supposer que la zone recouverte par ω est caractérisée par une conductivité qui tend vers l'infini (égale à $1+1/\varepsilon$). On se place sur l'espace $V = H^1_0(\Omega)$, et l'on prolonge f par 0 à l'intérieur de ω, en conservant la notation f pour cette fonction de $L^2(\Omega)$. On réécrit le problème initial comme un problème de minimisation sous contrainte dans le nouvel espace

$$(\mathcal{P}) \quad \left\{ \begin{array}{l} u \in K = \left\{ v \in V\,,\ v_{|\omega} = \text{cste} \right\}, \\ J(u) = \displaystyle\inf_{v\in K} J(v), \end{array} \right.$$

où J est la fonctionnelle définie par

$$J(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv.$$

La démarche proposée dans la section 9.3 consiste ici à écrire l'espace des u admissibles comme

$$K = \left\{ v \in V\,,\ \int_\omega \nabla u \cdot \nabla v = 0\,,\ \forall v \in V \right\},$$

et à introduire, pour $\varepsilon > 0$, la fonctionnelle

$$J_\varepsilon(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv + \frac{1}{2\varepsilon}\int_\omega |\nabla u|^2\,.$$

On peut vérifier qu'il s'agit de ce que nous avons appelé une pénalisation fermée. On peut en effet vérifier que l'opérateur qui à une fonction de $H^1(\Omega)$ associe la restriction de son gradient à ω comme fonction de $L^2(\omega)^N$ est à image fermée (cette propriété, qui fonde le caractère bien posé de la formulation point-selle, est établie dans la section suivante). On dispose donc ici d'une estimation optimale de l'erreur en $\mathcal{O}(\varepsilon)$.

3.4.3. Dualité. On introduit $\Lambda = L^2(\omega)^N$, et l'opérateur B de V dans Λ qui à une fonction de $H^1_0(\Omega)$ associe la restriction de son gradient à ω. On écrit

$$K = \left\{ v \in V\,,\ \int_\omega \nabla v \cdot \mathbf{z} = 0 \quad \forall \mathbf{z} \in \Lambda \right\}.$$

Le Lagrangien s'écrit

$$L(v, \mathbf{z}) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv + \int_\omega \nabla v \cdot \mathbf{z},$$

et le problème s'écrit sous la forme variationnelle suivante

$$\left\{\begin{array}{lcll} \displaystyle\int_\Omega \nabla u\cdot\nabla v \;+\; \int_\omega \nabla v\cdot\mathbf{y} & = & \displaystyle\int_\Omega fv & \forall v\in V \\ \displaystyle\int_\omega \nabla u\cdot\mathbf{z} & = & 0 & \forall \mathbf{z}\in\Lambda. \end{array}\right. \tag{3.10}$$

La fonction B utilisée pour la formulation est naturellement définie de $V = H^1_0(\Omega)$ dans $L^2(\omega)^N$, qui à une fonction de V associe la restriction de son gradient à ω. Cette application n'est, de façon évidente, pas surjective, car les champs de $L^2(\omega)^N$ ne sont pas tous des champs de gradient. On peut néammoins vérifier que le problème de point-selle est bien posé pour ce qui est de l'existence.

Proposition 3.8. L'opérateur B défini ci-dessus est à image fermée.

Démonstration : Soit $\mathbf{w} = Bu$. On définit $\tilde{u}\in H^1(\omega)$ par $\tilde{u} = u_{|\omega} - \overline{u}_{|\omega}$, où $\overline{u}_{|\omega}$ est la valeur moyenne de u sur ω, de telle sorte que $\nabla u_{|\omega} = \nabla\tilde{u}$. D'après l'inégalité de Poincaré-Wirtinger, il existe une constante telle que

$$\|\tilde{u}\|_{H^1(\omega)} \le C\,\|\mathbf{w}\|_{L^2(\omega)^N}.$$

Comme l'opérateur de restriction de $H^1_0(\Omega)$ vers $H^1(\omega)$ est surjectif, il existe d'autre part une constante C' telle que $\tilde{\mathbf{u}}$ admette un antécédent (pour l'opérateur de restriction) dans $H^1_0(\Omega)$ tel que

$$\|\mathbf{u}\|_{H^1_0(\Omega)} \le C'\,\|\tilde{\mathbf{u}}\|_{H^1(\omega)}.$$

On construit ainsi un antécédent (pour l'opérateur B) à $\mathbf{w}$, dont la norme est constrôlée par celle de $\mathbf{w}$. L'opérateur B est donc à image fermée. □

Remarque 3.9. On peut modifier la formulation du problème de façon à avoir unicité du multiplicateur de Lagrange en prenant pour Λ l'espace de champs de gradient de fonctions de $H^1(\omega)$. Cette extension est laissée en exercice.

Exercice 3.2. Proposer un modèle correspondant à la situation où l'on chauffe maintenant l'ensemble de la plaque (y compris la partie ω), et préciser les modifications qu'il faut alors apporter à la démarche ci-dessus.

3.5. Problème de Darcy

Soit Ω un domaine borné régulier. Les équations de Darcy que nous considérons ici modélisent l'écoulement d'un fluide dans un milieu poreux homogène isotrope. Elles expriment le fait que que le champs de vitesse est proportionnel au gradient local de la pression. Le modèle écrit par les physicien est donné directement sous forme de point-selle. C'est ainsi sous cette forme que nous l'introduirons, puis nous remonterons au problème de minimisation sous contrainte sous-jacent. Noter que le problème $\mathcal{P}''$ ci-dessous n'est pas un problème bien posé (ni même simplement « posé » à strictement parler) mathématiquement, tant que les espaces dans lesquels on cherche les inconnues n'ont pas été précisé. La démarche ici consiste justement à construire un cadre mathématique à ce problème, ce qui permettra de préciser dans quel sens les identités ont lieu (proposition 3.10).

Nous considérons dans un premier temps une situation où les bords sont libres (le fluide peut sortir du domaine ou y rentrer), et la pression au niveau du bord est imposée. On cherche un champ de vitesse $\mathbf{u} = (u_1, u_2)$ et un champ de pression p définis sur Ω (les régularités de ces champs seront précisées par la suite) tels que

$$(\mathcal{P}'') \quad \begin{cases} \mathbf{u} + \nabla \mathrm{p} &= \mathbf{f} \quad \text{dans } \Omega, \\ \nabla \cdot \mathbf{u} &= 0 \quad \text{dans } \Omega, \\ \mathrm{p} &= 0 \quad \text{sur } \Gamma, \end{cases} \tag{3.11}$$

où $\mathbf{f}$ est un champ de force donné. On se place sur l'espace en vitesses $V = L^2(\Omega)^2$. On pose $\Lambda = H^1_0(\Omega)$, et l'on introduit l'application B de V dans $\Lambda' = H^{-1}$ qui à $\mathbf{v} \in V$ associe la forme linéaire $B\mathbf{v}$ définie par

$$\langle B\mathbf{v}\,,\, \mathrm{q}\rangle = \int_\Omega \mathbf{v} \cdot \nabla \mathrm{q}.$$

On définit alors $K = \ker B$, et le problème de minimisation sous contrainte s'écrit

$$(\mathcal{P}) \quad \begin{cases} \mathbf{u} \in K = \left\{ \mathbf{v} \in L^2(\Omega)^2\,,\ \int_\Omega \mathbf{v} \cdot \nabla \mathrm{q} = 0 \quad \forall \mathrm{q} \in H^1_0(\Omega) \right\}, \\ J(\mathbf{u}) = \inf\limits_{\mathbf{v} \in K} J(\mathbf{v}), \quad \text{avec } J(\mathbf{v}) = \dfrac{1}{2}\displaystyle\int_\Omega |\mathbf{v}|^2 - \int_\Omega \mathbf{v} \cdot \mathbf{f}. \end{cases} \tag{3.12}$$

Proposition 3.10. Soit Ω un domaine borné de frontière Lipschitz, et $\mathbf{f} \in L^2(\Omega)$. Le problème de minimisation (3.12) ci-dessus admet une solution unique $\mathbf{u} \in K$. Cette solution admet une divergence faible L^2, nulle presque partout sur Ω, et il existe un unique $\mathrm{p} \in \Lambda = H^1_0(\Omega)$ tel que

$$\mathbf{u} + \nabla \mathrm{p} = \mathbf{f} \quad \text{p.p.}$$

Démonstration : Le problème (3.12) consiste à minimiser une fonctionnelle quadratique sur un sous-espace K fermé (K s'exprime comme le noyau d'une application linéaire continue). Il admet donc une solution unique $\mathbf{u} \in K$. L'appartenance à K entraîne l'existence d'une divergence faible pour $\mathbf{u}$, d'après la proposition 8.78, page 111, nulle presque partout sur Ω.

Il reste à vérifier que le problème de point-selle associé est bien posé. En effet, l'application B est surjective, car son adjoint $B^\star : \mathrm{q} \longmapsto \nabla \mathrm{q}$ est tel que

$$|B^\star \mathrm{q}| = |\nabla \mathrm{q}|_{L^2(\Omega)} \geq \alpha\, |\mathrm{q}|_{H^1_0(\Omega)}\,,$$

d'après l'inégalité de Poincaré 8.48, page 102, ce qui assure bien la surjectivité de B selon la proposition 7.20, page 84. On se trouve donc dans le cadre des hypothèses du théorème 9.9, page 119, qui assure l'existence et l'unicité d'un couple $(\mathbf{u}, \mathrm{p})$ tel que $\mathbf{u} + \nabla \mathrm{p} = \mathbf{f}$, au sens de l'identité entre fonctions de $L^2(\Omega)$. □

3.6. Problème de Stokes

On considère un domaine Ω du plan, dans lequel on cherche à résoudre le problème de Stokes. Contrairement aux problèmes envisagés précédemment, ce problème est décrit en général sous une forme de point-selle (c'est sous cette forme qu'il intervient en physique). C'est donc sous cette forme (qui correspond à la forme abstraite (9.5), page 119) que nous le présenterons.

On cherche un champ de vitesse $\mathbf{u} = (u_1, u_2)$ et un champ de pression p définis sur Ω (les régularités de ces champs seront précisées par la suite) tels que

$$\mathcal{P}'' \quad \left\{ \begin{array}{rcl} -\triangle \mathbf{u} + \nabla \mathrm{p} & = & \mathbf{f}, \\ \nabla \cdot \mathbf{u} & = & 0, \end{array} \right. \tag{3.13}$$

où $\mathbf{f}$ est un champ de force donné. La première des deux équations ci-dessus exprime l'équilibre des forces en chaque point du fluide, et la seconde exprime l'incompressibilité du fluide.

Nous allons maintenant préciser comment ce problème rentre le cadre de ce qui a été vu précédemment, en repartant du point de départ usuel qui est le problème de minimisation sous contrainte, puis en reconstruisant le problème de Stokes tel qu'énoncé ci-dessus à partir de la formulation point-selle.

On introduit les espaces

$$V = H_0^1(\Omega)^2 = \left\{ \mathbf{u} = (u_1, u_2)\,,\ u_1, u_2 \in H_0^1(\Omega) \right\}\ ,\ K = \{ \mathbf{u} \in V\,,\ \nabla \cdot \mathbf{u} = 0 \}\ ,$$

On considère le problème de minimisation sous contrainte

$$(\mathcal{P}) \quad \left\{ \begin{array}{l} \mathbf{u} \in K, \\ J(\mathbf{u}) = \inf\limits_{\mathbf{v} \in K} J(\mathbf{v}), \end{array} \right. \tag{3.14}$$

où J est la fonctionnelle

$$J(\mathbf{v}) = \frac{1}{2} \int_\Omega |\nabla \mathbf{v}|^2 - \int_\Omega \mathbf{f} \cdot \mathbf{v}$$

Proposition 3.11. La fonctionnelle J admet un unique minimiseur sur K.

Démonstration : L'application $\mathbf{v} \mapsto \nabla \cdot \mathbf{v}$ étant linéaire continue (de V dans $L^2(\Omega)$), l'ensemble K est un sous-espace vectoriel fermé de V. De plus la fonctionnelle J est du type

$$J(\mathbf{v}) = \frac{1}{2} a(\mathbf{v}, \mathbf{v}) - \langle \varphi\,, \mathbf{v} \rangle,$$

où $a(\cdot, \cdot)$ est une forme bilinéaire symétrique continue et coercive sur V, et $\varphi \in V'$. On se trouve donc bien dans les hypothèses du début du chapitre **??** (voir page **??**). Le théorème de Lax-Milgram assure l'existence et l'unicité d'un minimiseur. □

En vue d'écrire ce problème sous la forme d'une recherche de point-selle, nous introduisons maintenant l'espace

$$\Lambda = L_0^2(\Omega) = \left\{ \mathrm{p} \in L^2(\Omega)\,,\ \int_\Omega \mathrm{p} = 0 \right\},$$

et l'opérateur

$$B\ :\ \mathbf{v} \in V \longmapsto B\mathbf{v} = -\nabla \cdot \mathbf{v}.$$

L'espace K peut s'écrire

$$K = \left\{ \mathbf{v} \in V\,,\ -\int_\Omega \mathrm{q} \nabla \cdot \mathbf{v} = 0 \quad \forall \mathrm{q} \in \Lambda \right\},$$

ce qui conduit au Lagrangien

$$L(\mathbf{v}, \mathrm{q}) = \frac{1}{2} \int_\Omega |\nabla \mathbf{v}|^2 - \int_\Omega \mathbf{f} \cdot \mathbf{v} - \int_\Omega \mathrm{q} \nabla \cdot \mathbf{v}.$$

Le caractère bien posé de la formulation point-selle est assuré par la

Proposition 3.12. Soit Ω un domaine borné de frontière Γ Lipschitz, et $\mathbf{f} \in L^2(\Omega)^N$. Le Lagrangien L défini ci-dessus admet un unique point-selle $(\mathbf{u}, \mathrm{p}) \in V \times \Lambda$, où $\mathbf{u}$ est la solution du problème de minimisation sous contrainte (3.14). De façon équivalente (voir proposition 9.7, page 119), il existe un unique couple $(\mathbf{u}, \mathrm{p}) \in H_0^1(\Omega)^N \times L_0^2(\Omega)$ tel que

$$\int_\Omega \nabla\mathbf{u} : \nabla\mathbf{v} - \int_\Omega \mathrm{p}\nabla\cdot\mathbf{v} = \int_\Omega \mathbf{f}\cdot\mathbf{v} \quad \forall \mathbf{v} \in H_0^1(\Omega)^N \tag{3.15}$$

$$\int_\Omega \mathrm{q}\nabla\cdot\mathbf{u} = 0 \quad \forall \mathrm{q} \in L_0^2(\Omega). \tag{3.16}$$

Démonstration : Malgré l'analogie formelle avec le problème de Darcy (l'opérateur B est l'opérateur de divergence dans les deux cas), la démonstration est plus délicate, et nous référons à des ouvrages de références pour les aspects techniques. L'existence et l'unicité d'un point-selle est une conséquence de la surjectivité de l'opérateur de divergence B, qui est assurée par le lemme 3.13 ci-après.

Signalons que certains auteurs choisissent de ne pas utiliser explicitement la formulation point-selle, mais déduisent directement l'existence et l'unicité du couple vitesse-pression du théorème de de Rham 3.14 ci-dessous. En effet, la solution $\mathbf{u}$ vérifie (voir proposition 9.2, page 116)

$$\int_\Omega \nabla\mathbf{u}\cdot\nabla\mathbf{v} + \langle \xi\,,\mathbf{v}\rangle = \int_\Omega \mathbf{f}\cdot\mathbf{v} \quad \forall \mathbf{v} \in H_0^1(\Omega)^N$$

avec $\langle \xi\,,\mathbf{v}\rangle = 0$ pour tout $\mathbf{u}$ dans K, qui est constitué des champs à divergence nulle. Le théorème de de Rham assure donc l'existence et l'unicité d'une pression $\mathrm{p} \in L_0^2(\Omega)$ telle que la formulation variationnelle mixte (3.15)(3.16) est vérifiée. □

Lemme 3.13. Soit Ω un domaine connexe, borné, de frontière Γ Lipschitzienne, et soit q dans $L_0^2(\Omega)$. Il existe $\mathbf{v} \in H_0^1(\Omega)$ tel que $\nabla\cdot\mathbf{v} = \mathrm{q}$.

Démonstration : On se reportera à [**9**, lemme 3.2] pour la démonstration de ce résultat. Noter que le théorème de l'application ouverte assure l'existence d'une constante C telle que l'antécédent $\mathbf{v}$ peut être choisi tel que $\|\mathbf{v}\|_{H^1} \leq C\,\|\mathrm{q}\|_{L^2}$. □

Théorème 3.14. (De Rham)
Soit Ω un domaine connexe, borné, de frontière Γ Lipschitz. Soit ξ une forme linéaire sur $H_0^1(\Omega)^N$ qui s'annule contre tout champ dont la divergence est nulle :

$$\nabla\cdot\mathbf{v} = 0 \quad \text{p.p.} \Longrightarrow \langle \xi\,,\mathbf{v}\rangle = 0.$$

Alors il existe un unique $\mathrm{p} \in L_0^2(\Omega)$ tel que

$$\langle \xi\,,\mathbf{v}\rangle = \int_\Omega \mathrm{p}\nabla\cdot\mathbf{v} \quad \forall \mathbf{v} \in H_0^1(\Omega)^N.$$

Démonstration : Ce théorème est une conséquence directe du lemme 3.13 ci-dessus ainsi que de la proposition 7.20, page **??**. En effet, comme l'image de B est fermée, on a notamment

$$(\ker B)^\perp \subset B^\star(\Lambda'),$$

qui assure l'exsitence de p, qui est unique car $B^\star$ est injective par surjectivité de B. □

Remarque 3.15. Comme il a été précisé, établir l'existence et l'unicité d'une solution pour le problème de Stokes en formulation vitesse-pression est plus délicat que pour le problème de Darcy. Cette différence peut se préciser ainsi : dans le cas de Darcy, la démonstration repose sur une inégalité qui assure l'injectivité de $B^\star$ et le caractère fermé de son image. L'opérateur $B^\star$ va de $H_0^1(\Omega)$ dans $L^2(\Omega)^2$, et l'inégalité est conséquence directe de l'inégalité de Poincaré

$$\|\mathrm{q}\|_{L^2(\Omega)} \leq C \, \|\nabla \mathrm{q}\|_{L^2(\Omega)^N} \quad \forall \mathrm{q} \in H_0^1(\Omega),$$

qui est vérifiée dès que Ω est borné dans une direction (voir proposition 8.48, page 102). Dans le cas de Stokes, la surjectivité de l'opérateur B peut être établie comme conséquence directe d'une inégalité à première vue très similaire, l'opérateur $B^\star$ étant toujours dans un certain sens l'opérateur de gradient, mais vu cette fois comme un opérateur de $L^2(\Omega)$ dans $H^{-1}(\Omega) = (H_0^1(\Omega)^N)'$. Cette inégalité peut s'écrire

$$\|\mathrm{q}\|_{L^2(\Omega)} \leq C \, \|\nabla \mathrm{q}\|_{H^{-1}(\Omega)} \quad \forall \mathrm{q} \in L_0^2(\Omega),$$

où $\nabla \mathrm{q}$ représente la forme linéaire sur $H_0^1(\Omega)^N$ définie par

$$\mathbf{v} \longmapsto \int_\Omega \mathrm{q} \nabla \cdot \mathbf{v} \,, \ \|\nabla \mathrm{q}\|_{H^{-1}(\Omega)} = \sup_{\mathbf{v} \in H_0^1(\Omega)} \frac{\displaystyle\int_\Omega \mathrm{q} \nabla \cdot \mathbf{v}}{\|\mathbf{v}\|_{H_0^1(\Omega)^N}}.$$

Proposition 3.16. Soit Ω un domaine borné de classe C^2 et $\mathbf{f} \in L^2(\Omega)^N$. Soit $(\mathbf{u}, \mathrm{p})$ le point-selle de L dont l'existence et l'unicité est assurée par la proposition 3.12. On a alors $\mathbf{u} \in H^2(\Omega)$, $\mathrm{p} \in H^1(\Omega)$, et l'on a presque partout sur Ω

$$-\triangle \mathbf{u} + \nabla \mathrm{p} = \mathbf{f}.$$

3.7. Exercices

Exercice 3.3. On se place sur l'intervalle $I =]0,1[$, on note V l'espace de Sobolev $H_0^1(I)$, et l'on se donne deux réels x_1 et x_2, avec $0 < x_1 < x_2 < 1$. On considère l'ensemble

$$K = \{v \in V \,, \ v(x_1) = v(x_2)\},$$

et, pour $f \in L^2(I)$ donné, la fonctionnelle

$$v \longmapsto J(v) = \frac{1}{2} \int_I |v'|^2 - \int_I f v.$$

1) Montrer qu'il existe un unique $u \in K$ tel que

$$J(u) = \inf_{v \in K} J(v).$$

2) Montrer qu'il existe $\lambda \in \mathbb{R}$ tel que

$$\int_I u'v' = \int_I fv - \lambda \left(v(x_2) - v(x_1)\right) \quad \forall v \in V. \quad (\star)$$

3) Montrer que la restriction de u à chacun des sous-intervalles $]0, x_1[$, $]x_1, x_2[$, et $]x_2, 1[$, est de régularité H^2, et écrire l'équation dont u est solution sur chacun de ces sous-intervalles.

4) a) Montrer que u est continûment dérivable sur chacun des trois sous-intervalles.

b) Montrer que le saut des dérivées en x_1 est égal à l'opposé du saut des dérivées en x_2.

5) a) Pour tout $\lambda \in \mathbb{R}$, montrer que le problème $(\star)$ admet une solution $u_\lambda \in V$ unique.

b) Exprimer u en fonction de u_0 et u_1.

EXERCICE 3.4. On se place sur l'intervalle $I =]0,1[$, et l'on cherche à minimiser

$$J(v) = \frac{1}{2}\int_I \left|u'\right|^2$$

sur l'ensemble V des $v \in H^1(I)$ qui prennent la valeur 1 en 1, et qui s'annulent sur l'intervalle $\omega =]0,1/2[$. On introduit pour cela la fonctionnelle pénalisée

$$J_\varepsilon = \frac{1}{2}\int_I \left|u'\right|^2 + \frac{1}{2\varepsilon}\int_\omega |u|^2 ,$$

et l'on note u^ε la fonction qui réalise le minimum de J_ε sur V, l'ensemble des fonctions de $H^1(I)$, nulles en 0, et qui prennent la valeur 1 en 1.

1) Écrire les équations différentielles dont u_ε est solution sur chacun des intervalles $]0,1/2[$ et $]1/2,1[$, et préciser les conditions de raccord en $1/2$.

2) Expliciter u_ε.

3) Montrer que $|u_\varepsilon - u|_1$ est de l'ordre de $\varepsilon^{1/4}$

4) Que peut on dire des normes L^2 des restrictions à $]1/2,1[$ de $u_\varepsilon - u$ et de sa dérivée ?

3.8. Inclusions rigides dans un fluide de Stokes

On s'intéresse ici à un problème faisant intervenir deux type de contraintes. D'une part on considère un écoulement incompressible, et d'autre part on suppose qu'une partie du domaine est indéformable. On se conformera à l'approche menée dans la section précédente pour ce qui est de la contrainte d'incompressibilité. On se propose en revanche d'explorer les trois manières de prendre en compte la contrainte de mouvement rigide sur une partie du domaine.

On considère comme précédemment un domaine du plan, dans lequel on cherche à résoudre le problème de Stokes présenté précédemment, mais l'on suppose maintenant que seule une partie $\Omega \setminus \overline{\omega}$ est occupée par un fluide, le domaine ω correspondant à une zone occupée par un milieu rigide. Pour fixer les idées on supposera que ω est un disque ouvert fortement inclus dans Ω, mais on pourra généraliser l'approche sans difficulté au cas où ω a plusieurs composantes connexes, non nécessairement circulaires. On se place en régime non inertiel, ce qui conduit à écrire d'une part, comme précédemment, l'équilibre des forces en chaque point du fluide (équation de Stokes), ainsi que l'équilibre des forces exercées par le fluide sur la particule. De plus le caractère visqueux du fluide impose une condition de non-glissement à la surface de la particule : en tout point de γ, la vitesse du fluide s'identifie à la vitesse du milieu rigide.

On note $\mathbf{U}$ et η les vitesses de translation et de rotation du disque ω, qui sont des inconnues du problème. On cherche ainsi un champ de vitesse $\mathbf{u} = (u_1, u_2)$ défini sur $\Omega \setminus \overline{\omega}$, $(\mathbf{U}, \eta) \in$

$\mathbb{R}^2 \times \mathbb{R}$, et un champ de pression p défini sur $\Omega \setminus \overline{\omega}$ tels que

$$\left\{\begin{array}{rcll} -\triangle \mathbf{u} + \nabla \mathrm{p} &=& \mathbf{f} & \text{dans } \Omega \setminus \overline{\omega}, \\ \nabla \cdot \mathbf{u} &=& 0 & \text{dans } \Omega \setminus \overline{\omega}, \\ \mathbf{u} &=& 0 & \text{sur } \partial\Omega, \\ \mathbf{u} &=& \mathbf{U} + \eta \wedge \mathbf{r} & \text{sur } \partial\omega, \\ \displaystyle\int_\gamma \sigma \cdot \mathbf{n} &=& 0 & \\ \displaystyle\int_\gamma \mathbf{r} \wedge \sigma \cdot \mathbf{n} &=& 0 & \end{array}\right. \tag{3.17}$$

où $\mathbf{f}$ est un champ de force donné, $\mathbf{r}$ est le rayon vecteur relativement au centre du disque ω, et $\sigma \cdot \mathbf{n}$ est la contrainte normale

$$\sigma \cdot \mathbf{n} = (\nabla \mathbf{u} + {}^t\nabla \mathbf{u}) \cdot \mathbf{n} - \mathrm{p}\mathbf{n}.$$

Les méthodes de résolutions de ce problème que nous proposons ici sont basées sur la formulation sous forme de problème de minimisation sous contrainte. Nous introduisons les espaces

$$V = H_0^1(\Omega)^2 = \left\{\mathbf{u} = (u_1, u_2)\,,\ u_1, u_2 \in H_0^1(\Omega)\right\}\,,\ K_\nabla = \left\{\mathbf{u} \in V\,,\ \nabla \cdot \mathbf{u} = 0 \quad \text{p.p. } \right\}\,,$$

$$K_\omega = \left\{\mathbf{u} \in V\,,\ \exists\, (\mathbf{U}, \eta) \in \mathbb{R}^2 \times \mathbb{R} \ \text{ t.q. } \ \mathbf{u} = \mathbf{U} + \eta \wedge \mathbf{r} \quad \text{p.p. dans } \omega\right\},$$

et la fonctionnelle

$$J(\mathbf{v}) = \frac{1}{4}\int_\Omega \left|\nabla \mathbf{v} + {}^t\nabla \mathbf{v}\right|^2 - \int_\Omega \mathbf{f} \cdot \mathbf{v}.$$

Le problème de minimisation sous contrainte associé au système (3.17) s'écrit

$$\left\{\begin{array}{l} u \in K_\nabla \cap K_\omega, \\ J(u) = \displaystyle\inf_{v \in K_\nabla \cap K_\omega} J(v), \end{array}\right. \tag{3.18}$$

Remarque 3.17. On pourra être amené par la suite à désigner un élément de K_ω en explicitant les degrés de libertés associés au mouvement rigide de la particule. $(\mathbf{u}, \mathbf{V}, \eta) \in K_\omega$.

3.8.1. Approche directe / duale. Nous allons traiter la contrainte d'appartenance à K_∇ par dualité (comme dans la section précédente), et la contrainte d'appartenance à K_B de façon directe. Comme les champs de K_ω sont à divergence nulle dans B, la contrainte d'incompressibilité dans l'obstacle est redondante. On peut donc choisir de l'écrire de façon duale contre des multiplicateurs de Lagrange supportés à l'extérieur de ω. Néammoins, comme nous prévoyons d'imposer la contrainte de mouvement rigide de façon approchée (par pénalisation, voir section suivante), il est préférable d'imposer la contrainte d'incompressibilité sur l'ensemble du domaine Ω. On note ainsi comme précédemment

$$\Lambda = L_0^2(\Omega) = \left\{\mathrm{p} \in L^2(\Omega)\,,\ \int_\Omega \mathrm{p} = 0\right\}.$$

On définit le Lagrangien de $K_\omega \times L^2(\Omega)$ dans $\mathbb{R}$ par

$$L(\mathbf{v}, \mathrm{q}) = \int_\Omega |D(\mathbf{v})|^2 - \int_\Omega \mathbf{f} \cdot \mathbf{v} - \int_\Omega \mathrm{q} \cdot \nabla \mathbf{v},$$

avec

$$D(\mathbf{v}) = \frac{1}{2}\left(\nabla\mathbf{v} + {}^t\nabla\mathbf{v}\right).$$

La formulation que nous appellerons formulation directe (alors qu'elle n'est en fait que semi-directe : l'incompressibilité est traitée de façon duale) est la suivante

$$\begin{cases} \text{Trouver } (\mathbf{u}, \mathbf{U}, \eta) \in K_\omega \text{ et } \mathrm{p} \in \Lambda \text{ tel que :} \\ \dfrac{1}{2}\displaystyle\int_\Omega (\nabla\mathbf{u} + {}^t\nabla\mathbf{u}) : (\nabla\tilde{\mathbf{u}} + {}^t\nabla\tilde{\mathbf{u}}) - \int_\Omega \mathrm{p}\nabla\cdot\tilde{\mathbf{u}} = \int_\Omega \mathbf{f}\cdot\tilde{\mathbf{u}}, \quad \forall(\tilde{\mathbf{u}}, \tilde{\mathbf{V}}, \tilde{\eta}) \in K_\omega, \\ \displaystyle\int_\Omega \mathrm{q}\nabla\cdot\mathbf{u} = 0 \quad \forall \mathrm{q} \in \Lambda. \end{cases} \tag{3.19}$$

Le lien entre le système d'équations aux dérivées partielles (3.17) et le problème variationnel précédent est assuré par la proposition suivante.

Proposition 3.18. Soit $(\mathbf{u}, \mathrm{p})$ un couple vitesse-pression de $H_0^1(\Omega)^2 \times L^2(\Omega)$. On suppose que les restrictions de $\mathbf{u}$ et p à Ω sont respectivement dans $H^2(\Omega\setminus\overline{\omega})^2$ et $H^1(\Omega\setminus\overline{\omega})$. Alors $(\mathbf{u}, \mathbf{U}, \eta, \mathrm{p})$ est solution de (3.17) si et seulement si $(\mathbf{u}, \mathbf{U}, \eta, \mathrm{p})$ est solution de (3.19).

Démonstration : La partie la plus délicate consiste à passer de la formulation variationnelle au système déquations aux dérivées partielles. On considère une solution $(\mathbf{u}, \mathrm{p})$ de $\mathcal{P}$. On considère dans un premier temps des fonctions-test à support dans $\Omega\setminus\overline{\omega}$. D'après la proposition 13.17, page 167, il vient

$$-\triangle\mathbf{u} + \nabla\mathrm{p} = \mathbf{f} \quad \text{dans } \Omega\setminus\overline{\omega}.$$

On considère maintenant des fonctions-test $\tilde{\mathbf{U}}$ de K_ω qui ne s'annulent pas nécessairement sur ω. Ces fonctions $\tilde{\mathbf{U}} = (\tilde{\mathbf{u}}, \tilde{\mathbf{U}}, \tilde{\eta})$ vérifient la condition de mouvement rigide dans ω. L'intégration par parties (en remarquant que les intégrales du membre de gauche peuvent en fait se restreindre à $\Omega\setminus\overline{\omega}$) donne donc

$$-\int_{\Omega\setminus\overline{\omega}} \triangle\mathbf{u}\cdot\tilde{\mathbf{u}} + \int_{\Omega\setminus\overline{\omega}} \tilde{\mathbf{u}}\cdot\nabla\mathrm{p} + \int_\gamma \left((\nabla\mathbf{u} + {}^t\nabla\mathbf{u})\cdot\mathbf{n} - \mathrm{p}\mathbf{n}\right)\cdot\tilde{\mathbf{u}} = \int_\Omega \mathbf{f}\cdot\tilde{\mathbf{u}}.$$

On a donc, d'après ce qui précède,

$$\int_\gamma \left((\nabla\mathbf{u} + {}^t\nabla\mathbf{u})\cdot\mathbf{n} - \mathrm{p}\mathbf{n}\right)\cdot\left(\tilde{\mathbf{U}} + \tilde{\eta}\wedge\mathbf{r}\right) = 0 \quad \forall(\tilde{\mathbf{U}}, \tilde{\eta}) \in \mathbb{R}^2\times\mathbb{R}.$$

On a donc en particulier (prendre $\tilde{\eta} = 0$, $\tilde{\mathbf{U}} = \mathbf{e}_x$ puis $\tilde{\mathbf{U}} = \mathbf{e}_y$)

$$\int_\gamma \sigma\cdot\mathbf{n} = 0,$$

et (on prend $\tilde{\mathbf{U}} = 0$, $\tilde{\eta} = 1$, et on utilise $(\sigma\cdot\mathbf{n})\cdot(\tilde{\eta}\wedge\mathbf{r}) = \tilde{\eta}\cdot(\mathbf{r}\wedge\sigma\cdot\mathbf{n})$)

$$\int_\gamma \mathbf{r}\wedge\sigma\cdot\mathbf{n} = 0.$$

On a donc bien $(\mathbf{u}, \mathrm{p})$ solution de (3.17).

La réciproque se démontre aisément en intégrant par parties l'équation de Stokes multipliée par une fonction test de K_ω et intégrée sur $\Omega\setminus\overline{\omega}$. □

3.8.2. Pénalisation. En premier, nous introduisons une nouvelle expression de K_ω, donnée par la proposition suivante.

Proposition 3.19. L'ensemble K_ω peut s'écrire

$$K_\omega = \left\{ \mathbf{u} \in V\,,\ \nabla\mathbf{u} + {}^t\nabla\mathbf{u} = 0 \quad \text{p.p. dans } \omega \right\},$$

Démonstration : Voir Allaire [**1**]. □

La méthode de pénalisation appliquée à la contrainte de mouvement rigide conduit ainsi à la formulation variationnelle suivante

$$\mathcal{P}_\varepsilon \begin{cases} \text{Trouver } \mathbf{u} \in V \text{ et } \mathrm{p} \in \Lambda \text{ tel que :} \\ \dfrac{1}{2}\displaystyle\int_\Omega (\nabla\mathbf{u} + {}^t\nabla\mathbf{u}) : (\nabla\tilde{\mathbf{u}} + {}^t\nabla\tilde{\mathbf{u}}) - \int_\Omega \mathrm{p}\nabla\cdot\tilde{\mathbf{u}} \\ +\dfrac{1}{2\varepsilon}\displaystyle\int_\omega (\nabla\mathbf{u} + {}^t\nabla\mathbf{u}) : (\nabla\tilde{\mathbf{u}} + {}^t\nabla\tilde{\mathbf{u}}) = \int_\Omega \mathbf{f}\cdot\tilde{\mathbf{u}}, \quad \forall\tilde{\mathbf{u}} \in V, \\ \displaystyle\int_{\eta_F} \mathrm{q}\nabla\cdot\mathbf{u} = 0 \quad \forall \mathrm{q} \in \Lambda, \end{cases}$$

Proposition 3.20. Pour $\varepsilon > 0$, on note u^ε la partie primale de la solution de $\mathcal{P}_\varepsilon$. Alors

$$\|u - u^\varepsilon\|_{H^1(\Omega)} = \mathcal{O}(\varepsilon).$$

Démonstration : C'est une application directe des propriétés établies pour la méthode de pénalisation appliquée à la minimisation de la fonctionnelle J sur $K_\omega \cap K_\nabla$. On prendra garde au fait que l'espace de référence (qui correspond à l'espace V du cadre abstrait) est ici l'espace K_∇ des champ H^1 à divergence nulle. La convergence forte de $\mathbf{u}^\varepsilon$ vers $\mathbf{u}$ dans K_∇ (et donc dans V) est une conséquence du théorème 9.22, page 124. L'ordre 1 de convergence est assuré par la proposition 9.31, page 129, que l'on peut utiliser ici car l'application

$$B \ :\ u \in H^1_0(\Omega)^2 \longmapsto (\nabla\mathbf{u} + {}^t\nabla\mathbf{u})_{|\omega} \in L^2(\omega)^4$$

est à image fermée. En effet, pour tout $\mathbf{w} = (\nabla\mathbf{u} + {}^t\nabla\mathbf{u})_{|\omega} \in \mathrm{Im}(B)$, on peut introduire la projection $\tilde{\mathbf{u}}$, dans $H^1(\omega)$, de $\mathbf{u}_{|\omega}$ sur l'orthogonal des champs correspondant à un mouvement rigide dans Ω. On a $D(\tilde{\mathbf{u}}) = D(\mathbf{u})$ sur ω. Du fait de cette orthogonalité et d'après le corollaire 8.55, page 104, il existe une constance $C > 0$ telle que

$$\|\tilde{\mathbf{u}}\|_{H^1(\omega)} \leq C\,|D(\tilde{\mathbf{u}})|_{0,\omega} = C\,|D(\mathbf{u})|_{0,\omega}\,.$$

On applique maintenant à $\tilde{\mathbf{u}}$ l'opérateur de prolongement de la proposition 8.28, page 97 (que l'on peut choisir tel que son image est dans $H^1_0(\Omega)^2$). La fonction $P(\tilde{\mathbf{u}})$ est bien un antécédent de $\mathbf{w}$, dont la norme est contrôlée par celle de $\mathbf{w}$, ce qui assure le caractère fermé de B (voir proposition 7.15, page 82).

Chapitre 4

Estimation d'erreur pour les problèmes sous contraintes

Nous présentons dans ce chapitre comment l'analyse théorique menée précédemment peut-être appliquée à l'analyse d'erreur effective pour un certain nombre de problèmes. Nous distinguerons les situations de contraintes "distribuée" (typiquement contrainte de Divergence nulle pour Stokes) pour lesquelles on s'attachera à vérifier la condition inf-sup, et les situations de contraintes géométriques (typiquement une inclusion traité comme une contrainte pour permettre l'utilisation de maillages cartésiens.

4.1. Contrainte distribuée

4.1.1. Approximation numérique du problème de Stokes. On s'intéresse ici à la discrétisation du problème de Stokes basée sur l'utilisation de l'élément P^1 bulle - P^1. Étant donnée une triangulation T_h, on introduit donc l'espace V_h donc chaque composante est somme d'une fonction continue P^1 et d'une combinaison linéaire de fonctions dites "bulles", produit des coordonnées barycentriques dans chaque élément. L'espace Λ_h est l'espace P^1 usuel.

Proposition 4.1. Le couple (V_h, Λ_h) vérifie la condition inf-sup discrète.

Démonstration : On cherche à utiliser le lemme 11.8. Pour un élément quelconque de $H_0^1(\Omega)^N$, on considère tout d'abord $R_h v$, où R_h est l'opérateur dit de Clément, dont nous admettrons ici l'existence, qui est tel que

$$|\mathbf{v} - R_h\mathbf{v}|_{0,K} \leq Ch_K \, |\mathbf{v}|_{1,\Delta_K} \, ,$$

où Δ_K est l'ensemble des éléments en contact avec K, et qui est uniformément stable en semi-norme H^1, i.e.

$$|R_h\mathbf{v}|_{1,\Omega} \leq C \, |\mathbf{v}|_{1,\Omega} \, .$$

Cet opérateur joue le rôle d'un opérateur d'interpolation locale (il ne fait intervenir que des valeurs de la fonction au voisinage de l'élément considéré), bien défini pour des fonctions peu régulières (notamment non continues). On corrige ensuite $R_h\mathbf{v}$ au niveau de chaque élément de façon à conserver la valeur moyenne. Plus précisément on écrira sur K

$$\Pi_h\mathbf{v} = R_h\mathbf{v} + \mathbf{w}_h,$$

où $\mathbf{w}_h$ est une fonction de type bulle (plus précisément un vecteur dont chaque composante est de type bulle), choisie de telle sorte que

$$\int_K \Pi_h\mathbf{v} = \int_K R_h\mathbf{v} + \int_K \mathbf{w}_h = \int_K \mathbf{v},$$

sur chaque éléments K de la triangulation. Cette condition définit de façon unique la fonction bulle, le poids de la bulle dans chaque élément pouvant être choisi indépendamment des autres.

Notons $\overline{\mathbf{w}}_h$ la valeur moyenne de $\mathbf{w}_h$ sur K. On a

$$\overline{\mathbf{w}}_h = \frac{\int_K \mathbf{w}_h}{|K|} = |K|^{-1} \left| \int_K \mathbf{v} - \int_K R_h \mathbf{v} \right| \leq |K|^{-1/2} \, |\mathbf{v} - R_h \mathbf{v}|_{0,K} \leq C \mid K|^{-1/2} \, h_K \, |\mathbf{v}|_{1,\Delta_K}$$

Par ailleurs, sur l'élément de référence, la semi norme H^1 d'une fonction bulle est contrôlée par la valeur moyenne (on est sur un espace vectoriel de dimension 1 (!), sur lequel toutes les normes sont équivalentes). En effectuant le changement de variable vers l'élément K considéré, on obtient

$$|\mathbf{w}_h|_{1,K} \leq C \rho_K^{-1} \overline{\mathbf{w}}_h \, |K|^{1/2} \, ,$$

et l'on peut, en supposant que le rapport d'aspect h_K/ρ_K est majoré, obtenir une inégalité analogue en remplaçant ρ_K par h_K.

On a donc au final

$$|\mathbf{w}_h|_{1,K} \leq C \, |\mathbf{v}|_{1,\Delta_K} \, .$$

Comme la semi-norme dans le membre de droite fait intervenir un voisinage de K, on a (l'intégrale globale fait apparaître chaque intégrale sur les éléments un nombre de fois contrôlé)

$$|\mathbf{w}_h|_{1,\Omega} \leq C \, |\mathbf{v}|_{1,\Omega} \, .$$

On a donc bien $|\Pi_h v|_{1,\Omega} \leq C \, |v|_{1,\Omega}$, et on a par construction

$$\int_\Omega q_h \nabla \cdot (\mathbf{v} - \Pi_h \mathbf{v}) = - \int_\Omega (\mathbf{v} - \Pi_h \mathbf{v}) \cdot \nabla q_h + \int_{\partial\Omega} q_h (\mathbf{v} - \Pi_h \mathbf{v}) \cdot \mathbf{n}.$$

Le second terme est nul car les vitesses v et v_h sont nulles au bord, et le premier est nul également car ∇q_h est constant sur chaque élément. On en déduit la condition inf-sup discrète grâce au lemme de Fortin 11.8. □

Remarque : Nous avons considéré ici le cas de conditions de Dirichlet homogènes en vitesse, ce qui impose de prendre des pressions à moyenne nulle. Dans la proposition précédente, nous avons donc implicitement supposé que les pressions discrètes étaient elles aussi à moyenne nulle. En pratique, on évite la construction d'un espace de pressions discrètes à moyenne nulle, et l'on travaille simplement avec des pressions affines par morceaux. En conséquence, la pression discrète est elle-même définie à une constante près, et le système matriciel sous forme point-selle résultant de la discrétisation proposée n'est pas inversible. Si l'on utilise une méthode de résolution itérative, ce caractère singulier peut ne pas être un problème, en revanche une méthode de résolution directe (de type pivot de Gauss) ne va pas marcher [1]. En pratique, on peut être amener par exemple à stabiliser le calcul en rajoutant un (petit) terme diagonal en pression dans la matrice point-selle.

1. Ou, pire, elle va donner l'impression de marcher car le pivot qui devrait être nul en arithmétique exacte ne le sera pas du fait des erreurs d'arrondi, et le logiciel risque de produire un résultat aberrant.

4.2. Contraintes géométriques

4.2.1. Exemples en dimension 1. Nous montrons dans cette section comment les propriétés d'approximation et le cadre abstrait introduits au chapitre précédent vont nous permettre de montrer que la discrétisation en espace de problèmes éventuellement « mal posés » (c'est-à-dire pour lesquels la contrainte s'écrit par l'intermédiaire d'une application qui n'est pas à image fermée) est convergente dans un sens que nous préciserons.

EXEMPLE 4.2. Nous étudions ici un cas d'école qui contient une bonne part des difficultés propres au caractère mal posé du problème, dans un cadre monodimensionnel qui limite les complications techniques.

Soit $I =]0,1[$, $\omega =]a,b[$, avec $0 < a < b < 1$, et $f \in L^2(I)$. On note $V = H_0^1(I)$, et l'on s'intéresse au problème consistant à trouver

$$u \in K\,,\ J(u) = \inf_K J\,,\quad \text{avec } J(v) = \frac{1}{2}\int_I u'v' - \int_I fv\,,\ K = \left\{v \in V\,,\ v_{|\omega} = 0\right\}. \tag{4.1}$$

On considère la formulation point-selle la plus naturelle de ce problème (dont on sait qu'elle est mal posée au niveau continu, voir section 3.3.3, page 32)

$$\left\{\begin{array}{rcll} \displaystyle\int_I u'v' & + & \displaystyle\int_\omega \lambda v = \displaystyle\int_I fv & \forall v \in V \\ \displaystyle\int_\omega \mu u & & = 0 & \forall \mu \in \Lambda. \end{array}\right. \tag{4.2}$$

Nous allons montrer que ce problème, malgré son caractère mal posé au niveau continu, peut être discrétisé en espace, et conduire à une méthode qui converge dans un sens que nous précisons plus loin. On introduit deux paramètres de discrétisation h et H, auxquels on associe respectivement des subdivisions uniformes de I et ω. On note V_h l'espace des fonctions continues sur I, nulles aux extrémités, et affines sur chaque intervalle du type $]nh,(n+1)h[$. On définit Λ_H comme l'ensemble des application de ω dans $\mathbb{R}$, constantes sur chaque intervalle du type $]a+kH, a+(k+1)H[$. On considère maintenant le problème discret qui consiste à chercher $(u_h, \lambda_h) \in V_h \times \Lambda_H$ tel que

$$\left\{\begin{array}{rcll} \displaystyle\int_I u_h'v_h' & + & \displaystyle\int_\omega \lambda v_h = \displaystyle\int_I fv_h & \forall v_h \in V_h \\ \displaystyle\int_\omega \mu_H u_h & & = 0 & \forall \mu_H \in \Lambda_H. \end{array}\right. \tag{4.3}$$

Proposition 4.3. On se donne $f \in L^2(I)$, nulle presque partout sur ω. Le problème (4.3) admet une solution (u_h, λ_H), unique pour ce qui est de sa composante primale u_h, et il existe une constante C telle que

$$\|u_h - u\|_{H^1} \leq C\left(\sqrt{h} + \sqrt{H}\right).$$

DÉMONSTRATION : On rappelle que B_H est défini comme opérateur de V dans Λ_H (ou de V_h dans Λ_H selon la situation) par

$$(B_H v, \mu_H) = (Bv, \mu_H) \quad \forall \mu_H \in \Lambda_H.$$

Noter tout d'abord que l'on n'a pas toujours unicité du multiplicateur de Lagrange pour ce problème (si H est très petit, B_H ne peut être surjectif, et par suite $B_H^\star$ est non injectif). En revanche, l'espace Λ_H étant de dimension finie, l'application B_H est à image fermée, et

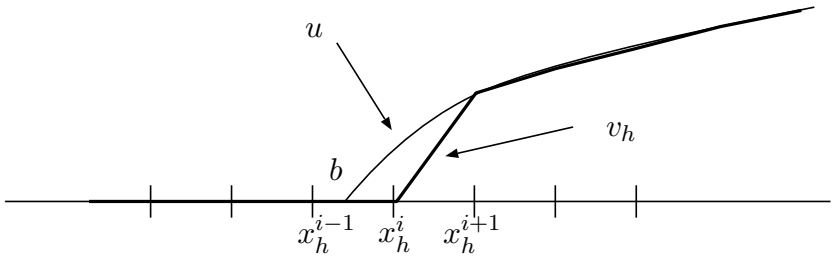


FIGURE 1. Construction de v_h.

le problème admet donc une solution (u_h, λ_H), unique pour ce qui est de sa composante primale u_h, définie comme le minimiseur de la fonctionnelle J sur

$$K_h^H = \{v_h \in V_h\,,\ (\mu_H, Bu_h) = 0 \quad \forall \mu_H \in \Lambda_H\}\,.$$

Nous allons utiliser maintenant l'estimation établie par la proposition 11.7, page 146, qui fait intervenir le ξ défini abstraitement par la proposition 9.2, page 116, et qui s'exprime ici

$$\langle \xi\,, v\rangle = \int_I fv - \int_I u'v' \quad \forall v \in V,$$

où u est la solution exacte du problème de minimisation sous contrainte (4.4). On peut ainsi exprimer [2]

$$\langle \xi\,, v\rangle = \int_\omega fv - u'(a^-)v(a) + u'(b^+)v(b) - u'(a^-)v(a) + u'(b^+)v(b),$$

car on a supposé f nulle sur ω. Il s'agit donc de construite dans un premier temps μ_H dans Λ_H, tel que $B_H^\star \mu_H$ approche ξ dans V'. On introduit pour cela la fonction

$$\mu_H^b = \frac{1}{H}\mathbb{1}_{]b-H,b[}.$$

On a, pour tout $v \in V$,

$$\begin{aligned}\left|\left\langle \mu_H^b\,, v\right\rangle - v(b)\right| &= \frac{1}{H}\left|\int_{b-H}^b (v(x) - v(b))\,dx\right| = \frac{1}{H}\left|\int_{b-H}^b \int_b^x v'(t)\,dt\,dx\right| \\ &\le \frac{1}{H}\int_{b-H}^b |b-x|^{1/2}\left(\int_x^b |v'|^2\right)^{1/2} \le \sqrt{H}\,\|v\|_{H^1}\,.\end{aligned}$$

Si l'on définit maintenant

$$\mu_H = -u'(a^-)\frac{1}{H}\mathbb{1}_{]a,a+H[} + u'(b^+)\frac{1}{H}\mathbb{1}_{]b-H,b[},$$

on a ainsi

$$\|B_H^\star \mu_H - \xi\|_{V'} \le C\sqrt{H}.$$

Il s'agit maintenant de traiter le premier terme de l'estimation abstraite (proposition 11.7). Pour cela, on introduit $I_h u$, l'interpolée affine par morceaux de la solution exacte u sur la subdivision de pas h, et l'on définit v_h comme la fonction continue affine par morceaux qui est égale à $I_h u$ sur tous les points, sauf sur ceux qui appartiennent à au moins un segment dont l'intersection avec $]a, b[$ est non vide (voir Fig. 1 la construction de v_h au voisinage de b).

2. On pourra écrire $\xi = f\mathbb{1}_\omega - u'(a^-)\delta_a + u'(b^+)\delta_b$.

La quantité $|u - v_h|^2_{I,1}$ peut se décomposer en la somme de deux termes. Le premier terme correspond à une intégrale sur un domaine où v_h s'identifie à $I_h u$ est sur lequel u est H^2. Ce premier terme est donc d'ordre 2 en h d'après les estimations usuelles. Le second terme correspond à une intégrale sur la réunion de deux intervalles (l'un contient a, l'autre b, composés chacun de deux segments. On se propose d'estimer le terme correspondant au voisinage de b. Notons $\eta = u(x_h^{i+1})$. Comme u est de régularité H^2 de u sur $]b, 1[$, elle est C^1, nulle en b, et ainsi on a $|\eta| \leq Ch$. On a donc

$$\int_{x_h^{i-1}}^{x_h^{i+1}} \left|v_h' - u'\right|^2 \leq 2\left(\int_{x_h^{i-1}}^{x_h^{i+1}} \left|v_h'\right|^2 + \int_{x_h^{i-1}}^{x_h^{i+1}} \left|u'\right|^2\right) \leq Ch,$$

d'où l'on déduit finalement que $|u - v_h|_{I,1}$ est d'ordre $\sqrt{h}$. □

Exemple 4.4. The following example illustrates the situation where the constraint for discretized problem is not the continuous constraint (see Section 11.2.2). We consider again $I =]0, 1[$, $V = H_0^1(I)$, $\omega =]a, b[$, with $0 < a < b < 1$, and $f \in L^2(I)$. Our purpose is to approximate u, defined as

$$J(u) = \inf_K J\,, \quad \text{with } J(v) = \frac{1}{2}\int_I u'v' - \int_I fv\,,\ K = \left\{v \in V\,,\ \int_\omega v = 0\right\}. \tag{4.4}$$

We introduce a uniform discretization of I,

$$0 = x_0^h < x_1^h = h < \cdots < x_N^h = 1,$$

and we suppose that the constraint is prescribed in a way which is consistent with the global discretization :

$$B_h v = \int_{a^h}^{b^h} v\,,\ \Lambda = \mathbb{R},$$

where a^h (resp. b^h) is the larger discretization point smaller than a (resp. the smaller discretization point larger than b), so that $]a, b[\subset]a^h, b^h[$, and $\left|]a^h, b^h[\backslash]a, b[\right| \leq 2h$.

4.2.2. Exercices.

Exercice 4.1. Consider the situation which combines the difficulties of examples 4.2 and 4.4, i.e. consider the minimization problem of example 4.2, with a constraint which is imposed in the spirit of example 4.4 : B_h (the subscript h is no longer justified) is defined by

$$(B_h v, \mu_h) = \int_{a^h}^{b^h} v\mu_h \in \Lambda_h,$$

where Λ_h is defined as the set of all those functions in $L^2(]a^h, b^h[)$ which are piecewise constant with respect to the h-discretization.

4.2.3. Problème de Poisson sur un domaine troué. Nous abordons maintenant l'analyse numérique des méthodes de pénalisation et de point-selle appliquées au problème de Poisson dans un domaine bidimensionnel perforé. La première partie est consacrée aux propriétés d'approximation de la solution exacte par une fonction discrète rattachée à un maillage qui ne respecte pas la géométrie de l'obstacle

Afin de simplifier les notations, nous allons nous placer dans un cadre géométrique particulier.

On considère $\Omega \subset \mathbb{R}^2$ un carré du plan, et ω un disque centré en l'origine que l'on suppose fortement inclus dans Ω. On considère une suite de triangulations régulières (T_h) de Ω. On note V_h l'espace des fonctions continues sur Ω dont la restriction à chaque triangle de T_h est affine.

Dans toute la suite on notera u la solution du problème

$$\begin{aligned} -\triangle u &= f \text{ dans } \partial\Omega \setminus \overline{\omega} \\ u &= 0 \text{ sur } \partial\Omega \\ u &= 0 \text{ sur } \partial\omega. \end{aligned}$$

On note toujours u le prolongement par 0 de cette fonction dans ω, de telle sorte que $u \in H_0^1(\Omega)$.

4.2.3.1. *Approximation de u sur l'espace discrétisé contraint.* On note $I_h u$ l'interpolée de u sur T_h. On définit $\tilde{u}_h$ comme la fonction de V_h qui vaut 0 sur tous les sommets des triangles de T_h qui ont une intersection non vide avec $\overline{\omega}$, et qui s'identifie à $I_h u$ sur tous les autres sommets. On introduit

$$\omega_{2h} = \{x \in \Omega\,,\ x \notin \overline{\omega}\,,\ d(x, \overline{\omega}) < 2h\}\,.$$

Les trois lemmes suivant vont nous permettre d'établir le résultat principal d'approximation de la solution exacte par une fonction de V_h.

Lemme 4.5. On a

$$\begin{aligned} |u - \tilde{u}_h|_{\Omega\setminus(\omega\cup\overline{\omega_{2h}}),0} &\leq Ch^2\,|u|_{\Omega,2}\,, \\ |u - \tilde{u}_h|_{\Omega\setminus(\omega\cup\overline{\omega_{2h}}),1} &\leq Ch\,|u|_{\Omega,2}\,, \end{aligned}$$

Démonstration : C'est une application directe de la propriété d'estimation standard. En effet, tous les triangles de T_h dont l'intersection avec $\Omega \setminus (\omega \cup \overline{\omega_{2h}})$ est non vide ne rencontrent pas ω, de telle sorte que $\tilde{u}_u = I_h u$ sur ces triangles. □

Lemme 4.6.

$$\begin{aligned} |u|_{\omega_{2h},0} &\leq Ch^{3/2}\,|u|_{\Omega,2}\,, \\ |u|_{\omega_{2h},1} &\leq Ch^{1/2}\,|u|_{\Omega,2}\,. \end{aligned}$$

Démonstration : On a

$$|u|^2_{\omega_{2h},1} = \int_0^{2\pi}\int_R^{R+2h} |\nabla u|^2\, r\, dr\, d\theta.$$

Or pour toute dérivée partielle de u, on a

$$\partial_i(r,\theta) = \partial_i(R,\theta) + \int_R^r \partial_r\partial_i u\, dr.$$

d'où

$$\begin{aligned} |u|^2_{\omega_{2h},1} &\leq 2\int_0^{2\pi}\int_R^{R+2h} |\partial_i(R,\theta)|^2\, r\, dr\, d\theta + 2\int_0^{2\pi}\int_R^{R+2h} \left|\int_R^r \partial_r\partial_i u\, ds\right|^2 r\, dr\, d\theta \\ &\leq Ch\int_{\partial B}\left|\frac{\partial u}{\partial n}\right|^2 + 2h\int_0^{2\pi}\int_R^{R+2h}\left(\int_R^{R+2h} |\partial_r\partial_i u|^2\, ds\right) r\, dr\, d\theta \\ &\leq Ch\,|u|^2_{\Omega,2} + C'h^2\,|u|^2_{\Omega,2}\,, \end{aligned}$$

d'où la seconde inégalité. La première se déduit de la seconde par une démonstration parfaitement analogue à celle de l'inégalité de Poincaré sur la bande (curviligne) ω_{2h}, dont l'épaisseur est en $\mathcal{O}(h)$. □

Lemme 4.7. On a

$$|\tilde{u}_h|_{\omega_{2h},0} \leq Ch^{3/2} |u|_{\Omega,2} ,$$
$$|\tilde{u}_h|_{\omega_{2h},1} \leq Ch^{1/2} |u|_{\Omega,2} .$$

La démonstration de ce dernier lemme est basée sur l'équivalence des normes L^∞ et L^2 sur l'espace des fonctions affines sur un triangle, et sur une inégalité inverse qui relie la norme L^2 du gradient à la norme L^∞. On notera $|w|_{K,\infty}$ la norme L^∞ d'une fonction w sur K. Nous allons énoncer (et démontrer) sous forme de lemmes ces deux propriétés avant d'aborder la démonstration du lemme 4.7.

Lemme 4.8. (Équivalence entre les normes L^2 et L^∞ dans un triangle)
Il existe des constantes C_1 et C_2 (universelles) telle que, pour tout triangle K non dégénéré (*i.e.* d'aire $|K|$ non nulle), toute fonction w_h affine sur K, on ait

$$C_1 |K| |w_h|^2_{K,\infty} \leq |w_h|^2_{K,0} \leq C_2 |K| |w_h|^2_{K,\infty} .$$

Lemme 4.9. (Inégalité inverse dans un triangle)
Il existe une constante universelle telle que, pour tout triangle K non dégénéré, toute fonction w_h affine sur K, on ait

$$|w_h|^2_{K,1} \leq C \frac{|K|}{\rho_K^2} |w_h|^2_{K,L^\infty} .$$

Démonstration : Soit w_h une fonction affine dans K, nulle en deux des sommets de K. Son gradient a pour module la norme L^∞ de cette fonction divisée par une des hauteurs du triangle K, qui est supérieure à ρ_K. Comme toute fonction affine w_h sur K s'écrit comme somme de telles fonctions, on a

$$|\nabla w_h|^2 \leq 3 \frac{|w_h|^2_{K,\infty}}{\rho_K^2},$$

d'où l'on déduit l'estimation proposée. □

Démonstration du lemme 4.7. : Les triangles de T_h qui « posent problème » sont ceux sur lesquels $\tilde{u}_h$ n'est ni identiquement nul ni identiquement égal à $I_h u$. Sur chacun de ces triangles, $\tilde{u}_h$ prend la valeur $I_h u$ sur 1 ou 2 sommets, et symétriquement la valeur 0 sur 2 ou 1 sommets. Notons $\tilde{\omega}_{2h}$ le sous-domaine de ω_{2h} défini comme l'union de tels triangles. On décompose le carré de l'erreur comme somme d'une intégrale sur $\tilde{\omega}_{2h}$ et d'une intégrale sur son complémentaire dans ω_{2h}. Les contributions associées à ce complémentaire correspondent à des triangles dans lesquels $\tilde{u}_h$ est soit nul, soit égal à $I_h u$, ce qui assure que l'on peut majorer la contribution par un $\mathcal{O}(h^3)$.

Pour ce qui est de l'intégrale sur $\tilde{\omega}_{2h}$, précisons dans un premier temps ce qui se passe au niveau d'un triangle de $\tilde{\omega}_{2h}$. Soit K l'un de ces triangles. On a

$$|\tilde{u}_h|^2_{K,0} \leq C_2 |K| |\tilde{u}_h|^2_{K,\infty} \leq C_2 |K| |I_h u|^2_{K,\infty} \leq \frac{C_2}{C_1} |I_h u|^2_{K,0} \leq C \left(|u|^2_{K,0} + h^4 |u|^2_{K,2} \right),$$

d'où

$$|\tilde{u}_h|^2_{\tilde{\omega}_{2h},0} = \sum_{K\in\tilde{\omega}_{2h}} |\tilde{u}_h|^2_{K,0} \leq C\left(|u|^2_{\tilde{\omega}_{2h},0} + h^4\,|u|^2_{K,2}\right) \leq C'h^3\,|u|^2_{\Omega,2}\,,$$

d'après le lemme 4.6.

Pour l'estimation sur la norme L^2 du gradient, on utilise ce qui précède et l'inégalité inverse du lemme 4.9.

$$|\tilde{u}_h|^2_{K,1} \leq C\frac{|K|}{\rho_K^2}\,|\tilde{u}_h|^2_{K,\infty} \leq \frac{C'}{\rho_K^2}\,|\tilde{u}_h|^2_{K,0}\,,$$

d'où l'on déduit, d'après l'estimation sur la norme L^2,

$$|\tilde{u}_h|_{\tilde{\omega}_{2h},1} \leq Ch^{1/2}\,|u|_{\Omega,2}$$

car on a supposé que la suite des triangulations est régulière (de telle sorte que h_K/ρ_K est majoré). □

Proposition 4.10. Il existe une constante C telle que

$$\inf_{w_h\in V_h\cap K} |u - w_h|_{\Omega,0} \leq Ch^{3/2}\,|u|_{\Omega,2}\,,$$

$$\inf_{w_h\in V_h\cap K} |u - w_h|_{\Omega,1} \leq Ch^{1/2}\,|u|_{\Omega,2}\,.$$

Démonstration : C'est une conséquence immédiate des lemmes précédents. On prend $w_h = \tilde{u}_h$ tel qu'il a été construit précédemment, et on décompose le carré de chacune de ces erreurs sur les domaines ω, ω_{2h}, et $\Omega\setminus(\omega\cup\overline{\omega_{2h}})$. La première contribution est nulle. La troisième correspond à une approximation optimale (voir lemme 4.5). Enfin, pour la seconde contribution, on écrit (par exemple pour la norme L^2)

$$\int_{\omega_{2h}} |\tilde{u}_h - u|^2 \leq 2\int_{\omega_{2h}} |\tilde{u}_h|^2 + 2\int_{\omega_{2h}} |u|^2\,.$$

La seconde de ces intégrales est en $\mathcal{O}(h^3)$ d'après le lemme 4.6, et le lemme 4.7 donne la même majoration pour la première intégrale. Le raisonnement est en tout point analogue pour la majoration de l'erreur en norme H^1. □

4.2.3.2. *Méthode de pénalisation.* On note $u_h^\varepsilon \in V_h$ la solution du problème pénalisé dicrétisé en espace, sans hypothèse pour l'instant sur la manière dont la pénalisation est effectuée (i.e. la forme bilinéaire $b(\cdot,\cdot)$ est simplement telle que $b(u,u) = 0$ si et seulement si $u \in K$. On a alors convergence de la méthode au sens suivant

Proposition 4.11. La suite $(u^\varepsilon_{h_\varepsilon})$ tend vers u dans $H^1_0(\Omega)$ quand h et ε tendent vers 0.

Démonstration : C'est une application directe de la proposition **??**, qui assure

$$|u_h^\varepsilon - u| \leq C\left(\min_{v_h\in V_h\cap K} |v_h - u| + \sqrt{|u^\varepsilon - u|}\right).$$

La convergence du premier terme vers 0 est assurée par la proposition 4.10, et la convergence du second par le théorème 9.22. □

La proposition précédente ne donne aucune indication sur la vitesse de convergence. On peut préciser cette vitesse (et prescrire un choix optimal de h relativement à ε) dans le cas de la pénalisation fermée, comme l'établit la proposition suivante.

Proposition 4.12. On suppose que $b(\cdot,\cdot)$ est de la forme

$$b(u,v) = (\Psi u, \Psi v)\,,$$

où Ψ est linéaire continue de V dans un espace de Hilbert Λ, à image fermée. On a alors

$$|u_h^\varepsilon - u| = \mathcal{O}(\sqrt{h}) + \mathcal{O}(\sqrt{\varepsilon}).$$

En particulier, si $\varepsilon = h$, on a convergence en $\sqrt{h}$.

Démonstration : C'est une conséquence directe de la proposition 9.31, page 129, et de la proposition 4.10. □

4.2.3.3. *Méthode de point-selle.* On se propose ici de faire l'analyse numérique de méthodes de point-selle associées les plus utilisées en pratique, qui sont associées à des formulations mal posées, c'est à dire pour lesquelles le problème continu n'admet pas de solution. On se propose d'utiliser l'estimation de la proposition 11.7, page 146, qui assure

$$|u - u_h| \leq C \left(\inf_{w_h \in K_h^H} |u - w_h| + \inf_{\mu_H \in \Lambda_H} |\xi - B_H^\star \mu_H| \right).$$

La démarche ci-dessus permet de majorer le premier terme par $C\sqrt{h}$. Le second terme est plus délicat, et son estimation dépend bien entendu de la manière dont sont gérées les contraintes. On se propose ici d'analyser deux approches usuelles, l'une basée sur une écriture de la contrainte sur la frontière de ω, l'autre basée sur des multiplicateurs de Lagrange qui vivent dans l'intérieur de ω (on parle de multiplicateurs *distribués*). Rappellons tout d'abord que ξ est la forme linéaire

$$v \in H_0^1(\Omega) \longmapsto < \xi, v > = -\int_\gamma \frac{\partial u}{\partial n} v.$$

Nous noterons $g = -\partial u/\partial n$ dans la suite, et nous supposerons que g est dans $H^1(\gamma)$. Il s'agit donc de construire $\mu_H \in \Lambda_H$ tel que

$$v \longmapsto < B^\star \mu_H, v >$$

tende (fortement) vers ξ dans le dual de $H_0^1(\Omega)$.

Contrainte sur la frontière.

Le lemme suivant précise comment il est possible d'approcher une forme linéaire du type $v \longmapsto \int_\gamma gv$ par une intégrale sur une ligne brisée contre une fonction constante par morceaux. Nous supposerons un peu plus de régularité sur g (qui représente $\partial u/\partial n$) que ce que le problème de départ autorise *a priori*[3]. Nous supposerons ainsi que g a une régularité H^1.

3. Cette propriété sur $g = \partial u/\partial n$ est assurée si l'on prend un second membre plus régulier $f \in H^{1/2}(\Omega \setminus \overline{\omega})$, ce qui est le cas dans toutes les applications que nous considérerons. Noter que si f n'est pas plus que L^2 au voisinage du bord, on pourrait établir une propriété d'approximation en $\sqrt{H}$.

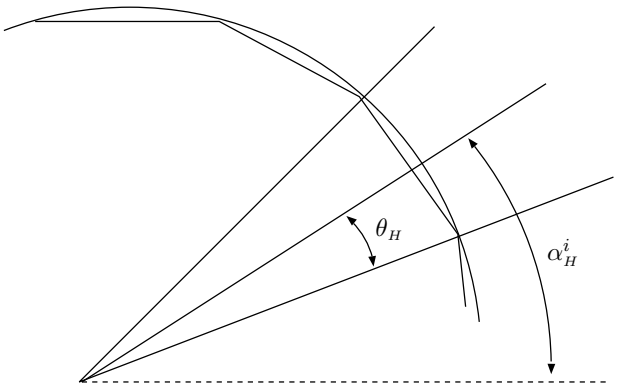


FIGURE 2. Définition de γ_H.

Lemme 4.13. Soit ω le disque de centre 0 et de rayon R, et Ω un domaine borné régulier qui contient fortement $\overline{\omega}$. Soit $H > 0$ tel que $2\pi R/H = N_H$ soit un entier. On introduit une subdivision uniforme de $]0, 2\pi[$

$$\alpha_H^1 < \cdots < \alpha_H^{N_H}\ ,\ \alpha_H^{i+1} - \alpha_H^i = 2\theta_H\ ,\ \theta_H = \pi/N_H.$$

et l'on définit γ_H comme la frontière du polygone de sommets les points du cercle associés aux angles $\alpha_H^i - \theta_H$ (voir figure 2). On appelle Λ_H l'espace des fonctions définies sur γ_H, constantes sur chaque arête de γ_H, muni de la norme L^2. On introduit l'opérateur B_H de V dans Λ_H défini par

$$(B_H v, \mu_H) = \int_{\gamma_H} \mu_H v \quad \forall \mu_H \in \Lambda_H.$$

Pour tout $g \in H^1(\gamma)$ donné, il existe une constante C telle que

$$\inf_{\mu_H \in \Lambda_H} \|B_H^\star \mu_H - \xi\|_{H^{-1}(\Omega)} \leq CH,$$

où ξ est la forme linéaire définie par $\langle \xi\ , v\rangle = \int gv$.

DÉMONSTRATION : On définit $g_H = g_H(\theta)$ comme la projection L^2 de g sur les fonctions constantes sur chaque sous-intervalle $]\alpha_H^i - \theta_H, \alpha_H^i + \theta_H[$, et l'on définit $\mu_H \in \Lambda_H$ comme la fonction qui prend la valeur de g_H sur le segment correspondant. Soit v une fonction régulière sur Ω. On peut écrire l'intégrale sur la ligne brisée comme une somme d'intégrales angulaires :

$$\int_{\gamma_H} \mu_H v - \int_\gamma gv \ = \ \sum_{i=1,N_H} \int_{\alpha_H^i - \theta_H}^{\alpha_H^i + \theta_H} v(r,\theta) g_H(\theta) \frac{R_H}{\cos^2\theta} d\theta - \int_\gamma gv.$$

On écrit

$$v(r,\theta) = v(R,\theta) - \int_r^R \partial r v\, ds.$$

On a ainsi

$$\begin{aligned}\int_{\gamma_H} \mu_H v - \int_\gamma g v &= \sum_{i=1,N_H} \int_{\alpha_H^i-\theta_H}^{\alpha_H^i+\theta_H} v(R,\theta)\left(g_H(\theta)\frac{R_H}{\cos^2\theta} - g(\theta)R\right) d\theta \\ &\quad - \sum_{i=1,N_H} \int_{\alpha_H^i-\theta_H}^{\alpha_H^i+\theta_H} v(R,\theta)\left(\int_r^R \partial_r v\, ds\right)\frac{R_H}{\cos^2\theta} d\theta\end{aligned}$$

et l'on majore chacune de ces intégrales par un $\mathcal{O}(H)$. (Démonstration à terminer).

□

Multiplicateurs distribués. On se limite ici à une version quelque peu idéalisée de ce qui est réalisé en pratique, mais dont l'analyse illustre bien le cas général. On note H le pas de discrétisation associé au multiplicateur. On discrétise la couronne des points de ω. On définit la projection dans $L^2(]0,2\pi[)$ de g sur l'espace des fonctions constantes sur chaque intervalle $]\theta_1^i,\theta_2^i[$. On peut exprimer explicitement

$$g_H|_{K_H^i} = \frac{\displaystyle\int_{\theta_1^i}^{\theta_2^i} dR\, d\theta}{R(\theta_2^i-\theta_1^i)}.$$

On définit μ_h comme la fonction constante sur chaque K_i^H, qui prend la valeur

$$\mu_H^i = \mu_H|_{K_i^H} = \frac{Rg_H}{\displaystyle\int_{R-H}^R r\, dr}.$$

et on note ξ_H la forme linéaire associée à μ_H par dualité L^2 :

$$\langle \xi_H\, , v\rangle = \langle B^\star \mu_H\, , v\rangle = \int_\omega \mu_H v \quad \forall v \in H_0^1(\Omega).$$

Lemme 4.14. On suppose que Λ_H, sous-espace de $L^2(\omega)$, contient les fonctions qui valent 1 sur l'un des K_H^i et qui sont nulles partout ailleurs. On a alors

$$\inf_{\mu_H\in\Lambda_H} |\xi - B^\star\mu_H| \le CH^{1/2}.$$

et par suite la méthode est convergente en $\mathcal{O}(h^{1/2}) + \mathcal{O}(H^{1/2})$.

Soit maintenant $v \in V = H_0^1(\Omega)$. On suppose dans un premier temps v régulière (au moins C^1). On utilisera l'expression de v en coordonnées polaires au voisinage de γ

$$v(r,\theta) = v(R,\theta) - \int_r^R \partial_r v(s,\theta)\, ds.$$

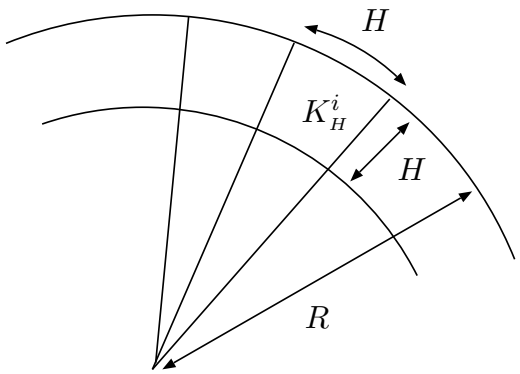


FIGURE 3. Définition de K_H^i.

On a

$$\begin{aligned} < B^\star \mu_H, v > - < \xi, v > &= \int_\Omega \mu_H v - \int_\gamma g v \\ &= \int_0^{2\pi} \int_{R-H}^{R} \mu_H v(R,\theta) r \, dr \, d\theta - \int_0^{2\pi} g v R d\theta \\ &\quad - \int_0^{2\pi} \int_{R-H}^{R} \mu_H \left(\int_r^R \partial_r v(s,\theta) \, ds \right) dr \, d\theta. \end{aligned}$$

Le premier terme s'écrit

$$\begin{aligned} \sum_{i=1}^{N_H} \int_{\theta_1^i}^{\theta_2^i} \int_{R-H}^{R} \mu_H v(R,\theta) &= \sum_{i=1}^{N_H} \int_{R-H}^{R} \mu_H^i r \, dr \int_{\theta_1^i}^{\theta_2^i} v(R,\theta) d\theta \\ &= \sum_{i=1}^{N_H} \int_{\theta_1^i}^{\theta_2^i} g_H(\theta) v(R,\theta) R\theta, \end{aligned}$$

de telle sorte que

$$\begin{aligned} \left| \int_0^{2\pi} \int_{R-H}^{R} \mu_H v(R,\theta) r \, dr \, d\theta - \int_0^{2\pi} g v R d\theta \right| &\leq C \left| g_H - g \right|_{]0,2\pi[,0} \left| v \right|_{]0,2\pi[,0} \\ &\leq C \left| g_H - g \right|_{]0,2\pi[,0} \left\| v \right\|_{H_0^1(\Omega)}. \end{aligned}$$

Comme g est supposé de régularité H^1, la fonction g_H approche g à un $\mathcal{O}(H)$ près.

Le second terme se majore

$$\begin{aligned} \int_0^{2\pi} \int_{R-H}^{R} \mu_H \left(\int_r^R \partial_r v(s,\theta) \, ds \right) dr \, d\theta &\leq \sqrt{H} \int_0^{2\pi} \int_{R-H}^{R} \mu_H \left(\int_{R-H}^{R} |\nabla v|^2 \, ds \right)^{1/2} dr \, d\theta \\ &= \sqrt{H} \int_0^{2\pi} \sqrt{R} g_H \left(\int_r^R |\nabla v|^2 \, R \, ds \right)^{1/2} d\theta \\ &\leq C \sqrt{H} \left| g_H \right|_{\gamma,0} \left| v \right|_{\Omega,1}. \end{aligned}$$

On a donc

$$\left\| \xi_H - \xi \right\|_{V'} = \mathcal{O}(\sqrt{H}).$$

Chapitre 5

Résolution effective

Ce chapitre regroupe un certain nombre de considérations sur les différentes méthodes de résolution d'un système linéaire (présentées en détail dans le chapitre **??**), leurs conditions d'utilisation, leurs avantages respectifs, leur vitesse de convergence le cas échéant.

Les méthodes de résolution des systèmes linéaires se classent en trois grandes catégories

1) Méthodes directes : ces méthodes conduisent à une résolution exacte du système (sous l'hypothèse idéalisée que les opérations élémentaires soient effectuées à précision infinie).

2) Méthodes itératives : basées sur un procédé itératif qui construit une suite d'approximations de la solution, qui converge vers cette solution.

3) Méthodes rapides. Il s'agit en fait de méthodes directes, mais qui méritent de figurer à part dans cette classification. La plupart ne sont applicables que dans des situations très particulières : opérateur de type Laplacien non perturbé (les inhomogénéités sont exclues), sur un maillage cartésien.

Précisons tout de suite que cette classification n'est pas étanche, car certaines méthodes directes au sens précédent (comme la méthode du gradient conjugué) produisent une suite d'approximations de la solution, et peuvent être utilisées (et donc considérées) comme des méthodes itératives. Nous réserverons donc l'adjectif *directes* aux méthodes qui ne donnent aucune approximation de la solution avant d'avoir été menées à terme.

Remarque 5.1. Une distinction essentielle entre les deux types d'approches est liée à la manière d'appréhender la matrice dans l'implantation effective de ces méthodes sur ordinateur. Les méthodes directes nécessitent un stockage de la matrice, alors que les méthodes itératives utilisent typiquement des produits matrice-vecteur successifs (la matrice elle-même n'est pas touchée), de telle sorte que l'on peut effectuer l'assemblage en temps réel et économiser la place mémoire de stockage de la matrice. Cet avantage est particulièrement sensible dans le cas d'un maillage structuré, qui conduit à une matrice qui ne contient qu'un petit nombre d'entrées différentes.

Nous supposerons ici que les opérations élémentaires sont réalisées avec une précision infinie : les erreurs d'arrondi ne sont pas prises en compte.

Nous commençons par aborder la notion de conditionnement d'une matrice, ingrédient essentiel pour mesurer la difficulté de résolution d'un système linéaire.

5.1. Conditionnement

Étant donnée une norme matricelle subordonnée[1], le conditionnement d'une matrice A inversible est défini comme $\kappa(A) = \|A\| \left\|A^{-1}\right\|$. Nous ne considérerons ici que la norme subordonnée à la norme euclidienne sur $\mathbb{R}^n$. Ce conditionnement peut être introduit comme majorant du facteur d'amplification entre une perturbation sur les données (second membre du système) ou sur la matrice elle-même, et la solution du problème (voir section 12.1, page 149). Au-delà de ce rôle d'estimateur de l'instabilité d'un problème (très important pour des problèmes « réels », pour lesquels les données sont susceptibles d'être entâchées d'erreurs), il intervient de façon essentielle dans les estimations de vitesses de convergence des méthodes itératives (notamment des méthodes de gradient décrites ci-après). Plus précisément, on peut estimer (voir remarque 12.1, page 149) à $C\kappa$ (resp. $C\sqrt{\kappa}$) le nombre d'itérations nécessaires (donc le temps de calcul pour une taille de matrice donnée) pour obtenir une précision donnée par une méthode de gradient (resp. de gradient conjugué).

Nous nous intéresserons ici principalement à la résolution de systèmes linéaires pour lesquels la matrice A provient de la discrétisation par éléments finis d'un problème elliptique, auquel cas A est symétrique définie positive. Le conditionnement est alors égal au rapport des valeurs propres extrêmes λ_n/λ_1. Avant d'aborder la question de l'estimation effective de ce conditionnement, il est indispensable de bien comprendre les deux remarques suivantes :

Remarque 5.2. Bien que les deux notions puissent interagir comme nous le verrons, il faut distinguer le conditionnement d'une matrice A issue de la discrétisation d'une forme bilinéaire $a(\cdot,\cdot)$, de la quantité $\|a\|/\alpha$ (où α est la constante de coercivité de $a(\cdot,\cdot)$). Prenons l'exemple du Laplacien en dimension 1 avec conditions de Dirichlet homogènes. La forme bilinéaire s'écrit au niveau continu

$$a(u,v) = \int_I u'v',$$

de telle sorte que, si l'on munit H_0^1 de la norme $|u|_1 = \int (u')^2$, le problème au niveau continu est parfaitement « conditionné » puisque $\|a\|/\alpha = 1$. Le problème consistant à trouver u tel que

$$a(u,v) = \langle \varphi\,,\, v\rangle \quad \forall v \in H_0^1(\Omega),$$

consiste exactement à réaliser l'identification de Riez-Fréchet ; l'inversion est donc une isométrie entre V' et V, et, si l'on perturbe le second membre abstrait φ par $d\varphi$, on a, avec des notations évidentes,

$$\frac{|du|}{|u|} \leq \frac{\|d\varphi\|}{\|\varphi\|}.$$

Le Laplacien est vu ici comme un opérateur de H_0^1 dans son dual H^{-1}, cette application est alors une isométrie, et comme telle parfaitement conditionnée.

Pour le problème discret sur le domaine $I =]0,1[$, le second membre est donné sous la forme d'une fonction de $L^2(I)$. Le système linéaire dans $\mathbb{R}^n$ obtenu peut être vu comme la version discrète du Laplacien, mais pour la norme euclidienne sur $\mathbb{R}^n$. Cette norme euclidienne correspond (à un facteur multiplicatif près) à la norme L^2 sur I. L'opérateur discret A correspond ainsi (toujours à un facteur multiplicatif près, qui n'affectera pas le

1. Norme du type $\|A\| = \sup \|Au\|/\|u\|$, où $\|\cdot\|$ est une norme de $\mathbb{R}^n$.

conditionnement) au Laplacien *comme opérateur de L^2 dans L^2*. Cet opérateur au niveau continu est dit non-borné, il n'est pas défini pour toutes les fonctions, et son spectre est une suite de réels positifs [2] qui tend vers $+\infty$: il est infiniment mal conditionné. Le fait de discrétiser tronque les hautes fréquences du spectre : seules les oscillations de fréquence au maximum de l'ordre de h ($1/h^2$ pour les valeurs propres correspondantes, puisqu'on dérive deux fois) peuvent exister. En conséquence, on obtient un conditionnement de l'ordre de $1/h^2$, ce que l'on peut vérifier sur le Laplacien discret en différence finie (voir exemple 12.3). Cette explosion du conditionnement reproduit simplement au niveau discret le caractère non borné du Laplacien comme opérateur de L^2 dans L^2.

Remarque 5.3. On prendra garde au fait que, en éléments finis (dans le cas d'un maillage uniforme) tout se passe à une échelle de l'ordre du volume d'un élément (h^d en dimension d), mais cela n'invalide pas ces considérations sur le conditionnement, qui est homogène d'ordre 0. Plus précisément, la discrétisation du problème de valeur propre continu, par exemple

$$-\triangle u = \lambda u,$$

dans le cas du Laplacien, conduit à résoudre le problème aux valeurs propres généralisé

$$Au_h = \lambda^h M u_h,$$

où M est la matrice dite de masse, telle que $(Mu_h, v_h) = \int u_h v_h$. Dans le cas d'un maillage uniforme, cette matrice est spectralement proche de $h^d\,\mathrm{Id}$ (voir exercice **??**). Or on peut montrer que les premières valeurs propres de ce problème discrétisé tendent vers les premières valeurs propres de l'opérateur continu. On appellera dans la suite λ_1^h/h^d la première valeur propre renormalisée par la matrice de masse, ou simplement renormalisée, qui se comporte comme la plus petite valeur propre λ_1 du problème continu.

Il est possible d'évaluer le conditionnement d'une matrice résultant de la discrétisation d'un problème elliptique en se fondant sur les considérations suivantes :

(i) la plus petite valeur propre est associée aux grandes longueur d'onde en espace, qui sont bien capturées par la méthode numérique. En conséquence, λ_1 ne dépend pas au premier ordre du pas de discrétisation (sous réserve de renormaliser par la matrice de masse conformément à la remarque 5.3 ci-dessus), mais seulement de la géométrie. C'est la première valeur propre de l'opérateur avec les conditions aux limites considérées. On se reportera à la section 10.4, page 138 pour une preuve de ces considérations.

(ii) la plus grande valeur propre est un reflet de la discrétisation. Elle ne correspond à rien de physique, mais aux fonctions de hautes fréquences spatiales qui vivent dans l'espace discret. Conformément à la remarque 5.2, qui sera développée ci-dessous, elle se comporte en $1/h^2$, où h est la taille du *plus petit* élément [3].

Estimation effective du conditionnement. Soit Ω un domaine de $\mathbb{R}^2$ (l'approche se généralise sans problème aux dimensions supérieures). La plus petite valeur propre du

2. Sur I, les vecteurs propres sont les fonctions $\sin(k\pi x)$, de valeurs propres associées $\pi^2 k^2$.

3. On prendra garde que dans le cas de maillages non uniformes (tailles des éléments très différentes) et non isotropes (possiblité d'avoir des éléments très allongés), le h qui intervient dans l'estimation d'erreur est ce qu'on appelle le diamètre de la triangulation, à savoir le diamètre du *plus grand* éléments, alors qu'ici h désigne la plus haute fréquence susceptible d'être capturée par le maillage, c'est à dire en fait le *plus petit* ρ_K (diamètre du cercle inscrit dans un élement K de la triangulation).

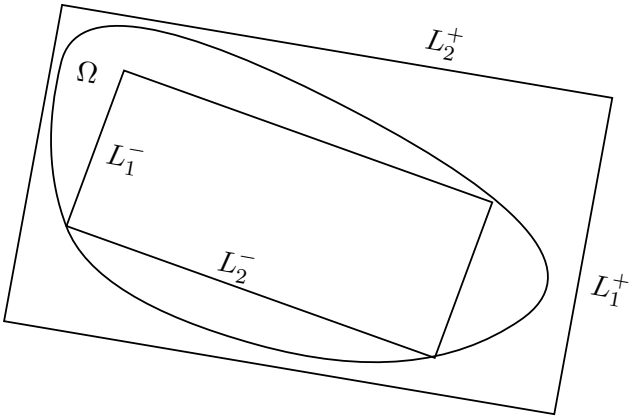


FIGURE 1. Estimation de λ_1

Laplacien avec conditions de Dirichlet homogènes s'écrit, d'après le théorème de Courant-Fisher,

$$\lambda_1 = \inf_{v \in H_0^1(\Omega)} \frac{(Av, v)}{|v|^2},$$

de telle sorte que pour tout Ω^+ contenant Ω, tout Ω^- contenu dans Ω, on a

$$\lambda_1(\Omega^+) \leq \lambda_1(\Omega) \leq \lambda_1(\Omega^-).$$

Or la proposition 13.15, page 165, donne l'expression de la plus petite valeur propre dans un rectangle $L_1 \times L_2$:

$$\lambda_1 = \pi^2 \left(\frac{1}{L_1^2} + \frac{1}{L_2^2} \right).$$

Ces propriétés permettent d'encadrer assez précisément Λ_1 pour des domaines pas trop éloignés d'un rectangle. Ainsi, pour le domaine représenté sur la figure 1, on a

$$\pi^2 \left(\frac{1}{(L_1^+)^2} + \frac{1}{(L_2^+)^2} \right) \leq \lambda_1 \leq \pi^2 \left(\frac{1}{(L_1^-)^2} + \frac{1}{(L_2^-)^2} \right).$$

La plus grande valeur propre correspond aux plus grandes fréquences rendues possibles par le maillage. On peut l'estimer en suivant la démarche suivante. Notons tout d'abord que le théorème de Gerchgorin (voir théorème 13.14, page 164) permet de majorer cette plus grande valeur propre par la somme des valeurs absolues des éléments de n'importe quelle ligne de la matrice. Pour la matrice du Laplacien dicrétisé par éléments finis sur un maillage uniforme (diamètre h), les éléments de la matrice sont tous de même ordre

$$\int_\Omega \nabla w_i \cdot \nabla w_j \approx (\text{Volume de l'élément}) \times (\text{valeur du gradient})^2 \approx h^d \times h^{-2} = h^{d-2},$$

ce qui donne (C est une constante universelle pour des maillages uniformes réguliers) $\lambda_N \leq C h^{d-2}$. Le théorème 13.13, page 164, permet de minorer cette plus grande valeur propre :

$$\lambda_N \geq \frac{(Av, v)}{|v^2|},$$

pour tout vecteur v. Si l'on prend, dans le cas du Laplacien discret, le vecteur associé à une fonction de base (dont la norme euclidienne est 1), on obtient

$$\lambda_N \geq \int_\Omega |\nabla w_i|^2 \approx h^d \times h^{-1} = h^{d-2}.$$

La plus grande valeur propre se comporte donc comme h^{d-2}. Si l'on normalise par la matrice de masse (ce qui revient essentiellement à diviser par le volume caractéristique d'un élément h^d), on obtient h^{-2}.

En conclusion, la plus petite valeur propre (normalisée) étant essentiellement une constante du problème (elle ne dépend pas du paramètre de discrétisation), on obtient un conditionnement de l'ordre de $1/h^2$.

Remarque 5.4. On notera que le conditionnement du Laplacien (ou d'un opérateur du même type, comme celui de l'élasticité) à nombre de points fixé (donc taille de matrice fixée) décroit avec la dimension. En effet, pour N points de maillage en dimension d, la taille des mailles varie comme $N^{-1/d}$, de telle sorte que le conditionnement se comporte en $N^{2/d}$. Le nombre de points total n'intervient pas en effet dans le conditionnement, mais bien le nombre de points *dans une direction d'espace*, qui conditionne la capacité à capturer des hautes fréquences.

5.2. Méthodes directes

Les méthodes directes présentées dans le chapitre **??** (voir section 12.2) sont en général inutilisables pour des problèmes de taille industrielle. Elles sont néammoins importantes pour les raisons suivantes :

(1) Elle peuvent être utilisées sur des cas test de taille raisonnable (typiquement des problèmes elliptiques bidimensionnels sur des maillages de l'ordre de 100×100), et permettent de tester des méthodes numériques en s'affranchissant des erreurs liées à la résolution effective des systèmes. La méthode utilisée par défaut dans `Freefem++` est ainsi une méthode directe, de type décomposition LU.

(2) Elles sont à la base d'un certain nombre de préconditionneurs, utilisés pour accélérer la convergence de méthodes itératives (par exemple préconditionnement LU incomplet).

(3) Une approche de décomposition de domaine permet parfois de se ramener à des sous-systèmes à résoudre de taille raisonnable, même si le problème global excluait a priori l'utilisation de méthodes directes. On peut alors tirer parti du fait que l'on doive résoudre un grand nombre de fois (dans le cadre d'un algorithme itératif lié à l'approche de décomposition de domaine) le même système pour des données différentes. On peut ainsi, par exemple, construire la décomposition LU d'une matrice associée à un sous-domaine, la conserver en mémoire, et utiliser cette même décomposition pour résoudre le système local (par la résolution de deux systèmes triangulaires) à chaque itération.

5.3. Méthodes itératives

Nous décrivons ici les méthodes de type gradient, basées sur des principes très généraux qui s'étendent à la minimisation de fonctionnelles non quadratiques.

La plus simple de ces méthodes est la

Méthode de gradient à pas fixe. Il s'agit d'une stratégie très générale de recherche de minimum d'une fonctionnelle, que l'on peut voir comme une discrétisation en temps (il s'agit d'un temps virtuel sans signification physique) d'une équation d'évolution appelée flot-gradient. Considérons une fonctionnelle J dérivable d'un espace de Hilbert H dans $\mathbb{R}$, et l'équation différentielle

$$\frac{du}{dt} = -\nabla J(u)\,,\ u(0) = u_0.$$

Si la trajectoire tend, quand t tend vers $+\infty$, vers un point u_∞ de H, alors u_∞ est point fixe de l'équation, et donc $\nabla J(u_\infty) = 0$. Il s'ensuit que u_∞ est un point critique de J. Dans le cas général, il peut s'agir d'un point-selle (la fonctionnelle augmente si l'on perturbe u_∞ dans certaines directions, diminue pour d'autres direction), ou d'un extremum local. Dans le cas où la fonctionnelle est convexe, le point limite, s'il existe, minimise la fonctionnelle J sur H. La méthode de gradient consiste à construire la suite associée à la discrétisation de cette équation différentielle par une méthode d'Euler explicite : pour un pas de temps $\rho > 0$ fixé, on construit

$$u^0 = u_0\,,\ \frac{u^{k+1} - u^k}{\rho} = -\nabla J(u^k).$$

Cette approche permet de construire une approximation de la solution du système $Au = b$, où A est symétrique définie positive, et $b \in \mathbb{R}^n$ un second membre. On introduit la fonctionnelle

$$J(v) = \frac{1}{2}(Av\,,v) - (b,v)\,,\ \text{de gradient } \nabla J(v) = Av - b,$$

et l'on introduit un pas de temps $\rho > 0$. La discrétisation du flot gradient par une méthode d'Euler explicite s'écrit donc

$$u^{k+1} = u^k - \rho\nabla J(u^k) = u^k - \rho(Au^k - b),$$

qui est exactement l'algorithme 12.9. Pour ρ assez petit, on a convergence de l'algorithme vers la solution u du système linéaire, ce qu'on peut voir comme un résultat sur le comportement asymptotique de l'algorithme de discrétisation en temps.

Remarque 5.5. Cet algorithme peut être utilisé dans le cas où la matrice est simplement symétrique positive. Dans le cas où b est dans l'image de A (cas pour lequel des solutions existent), on vérifie immédiatement que l'on a convergence vers une solution dont la projection (orthogonale) sur $\ker A$ s'identifie à celle de u^0. Cette propriété peut être utile dans le cas où l'on résout un problème dégénéré, comme un problème de déformation d'un solide élastique qui glisse sans frottement sur un support plan (le mouvement de translation parallèlement au support, qui n'introduit aucune déformation, donc laisse inchangée l'énergie élastique, est dans le noyau de la matrice A). Noter que dans le cas $u^0 = 0$ (qui correspond à ce que l'on fait en pratique, à moins de disposer d'une première approximation de la solution), la solution trouvée est dans l'orthogonal du noyau de A.

Les sections 12.3.2 et 12.3.3 décrivent les méthodes de gradient à pas optimal et de gradient conjugués, extensions de la méthode précédente, qui permettent d'approcher la solution beaucoup plus rapidement.

Remarque 5.6. La méthode de gradient conjugué apparaît (à la fois selon l'analyse de convergence et dans les faits) comme la meilleure méthode de gradient parmi celles présentées. On prendra néammoins garde au fait que ses qualités sont liées à la préservation de certaines orthogonalités ou relation de conjugaison entre les résidus et directions de descente (voir proposition 12.15), propriétés qui sont établies par récurrence. Or il arrive souvent que l'on utilise cet algorithme dans le cas où le produit matrice vecteur fasse intervenir un système à résoudre. Si ce système, que l'on résout à chaque itération, est résolu lui-même de façon imparfaite (par exemple par utilisation d'une méthode itérative, ou du simple fait des erreurs d'arrondi machine), les propriétés ne seront pas préservées, et la vitesse de convergence, ni même la convergence elle-même, ne sont garanties. La méthode de gradient à pas fixe est plus robuste en ce sens que chaque étape est semblable à la première [4] : l'algorithme refait le point à chaque étape, en quelque sorte. On peut d'ailleurs montrer (voir [**8**]) la convergence de ce type d'algorithme dans le cas où l'on commet à chaque itération une certaine erreur (si cette erreur n'est pas trop importante). L'algorithme du gradient conjugué se comporte pourtant bien en pratique, même si l'étape intermédiaire est résolue imparfaitement, auquel cas le résidu peut ne pas être strictement décroissant.

4. On pourra penser à un randonneur perdu la nuit dans la montagne. S'il s'arrange pour faire régulièrement un pas dans une direction qui abaisse sensiblement son altitude, même si ça n'est pas la direction de plus grande pente, il finira par atteindre la vallée. S'il s'avise de se suivre l'idée du chemin optimal qu'il peut avoir en tête, alors les erreurs accumulées peuvent le dévier de son objectif.

Deuxième partie

Aspects théoriques

Chapitre 6

Éléments d'analyse Hilbertienne

6.1. Généralités sur les espaces de Hilbert

Définition 6.1. (Produit scalaire)
Soit H un espace vectoriel sur $\mathbb{R}$. On appelle produit scalaire une forme bilinéaire (u, v) de $H \times H$ dans $\mathbb{R}$, symétrique, définie et positive : $(u, v) = (v, u)$, $(u, u) \geq 0 \quad \forall u \in H$ et $(u, u) = 0 \iff u = 0$.

Un produit scalaire définit sur H une structure d'espace vectoriel normé pour $u \longmapsto |u| = (u, u)^{1/2}$.

Définition 6.2. (Espace de Hilbert)
On appelle espace de Hilbert un espace vectoriel muni d'un produit scalaire, et qui est complet pour la norme associée.

Exemple 6.3. L'exemple le plus simple d'espace de Hilbert de dimension infinie est l'espace ℓ^2 des suites de carré intégrable. On peut définir par extension une infinité de nouveaux espaces dits « à poids » en introduisant, pour $\gamma = (\gamma_n)$ une suite quelconque de réels strictement positifs,

$$\ell^2_\gamma = \left\{ (u_n) \in \mathbb{R}^{\mathbb{N}},\ \sum \gamma_n |u_n|^2 < +\infty \right\}.$$

Proposition 6.4. (Inégalité de Cauchy-Schwarz)
Tout produit scalaire vérifie l'inégalité de Cauchy-Schwarz

$$|(u, v)| \leq (u, u)^{1/2}(v, v)^{1/2} \quad \forall u, v \in H.$$

Démonstration : On écrit que $(u + tv, u + tv)$ est positif, pour tout $t \in \mathbb{R}$, notamment pour $t = -(u, v)/|v|^2$ qui réalise le minimum. □

Proposition 6.5. (Identité du parallélogramme)
Toute norme issue d'un produit scalaire vérifie l'identité du parallélogramme

$$\left|\frac{u+v}{2}\right|^2 + \left|\frac{u-v}{2}\right|^2 = \frac{1}{2}(|u|^2 + |v|^2).$$

Proposition 6.6. Tout sous-espace vectoriel fermé d'un espace de Hilbert est un espace de Hilbert (pour le même produit scalaire).

Démonstration : La propriété découle simplement du fait que la restriction d'un produit scalaire à un sous-espace est un produit scalaire, et qu'un sous-espace fermé d'un espace complet est complet. □

Définition 6.7. (Séparabilité)
On dit qu'un espace de Hilbert H est séparable s'il existe un sous-ensemble de H dénombrable et dense dans H.

Théorème 6.8. (Projection sur un convexe fermé)
Soit H un espace de Hilbert et K un convexe fermé non vide de H. Pour tout $z \in H$, il existe un unique $u \in K$ (appelée projection de z sur K) tel que

$$|z-u| = \min_{v\in K} |z-v| = \mathrm{dist}(z,K).$$

La projection u est caractérisée par la propriété

$$\begin{cases} u \in K \\ (z-u, v-u) \leq 0 \quad \forall v \in K. \end{cases} \tag{6.1}$$

On notera $u = P_K z$.

Démonstration : On considère une suite minimisante (u_n)

$$u_n \in K\,,\ |z-u_n| \longrightarrow d = \mathrm{dist}(z,K).$$

Pour p, $q \in \mathbb{N}$, on applique l'identité du parallélogramme à $u_p - z$ et $u_q - z$:

$$\left|\frac{u_p+u_q}{2} - z\right|^2 + \left|\frac{u_p-u_q}{2}\right|^2 = \frac{1}{2}(|u_p-z|^2 + |u_q-z|^2).$$

Comme K est convexe $(u_p+u_q)/2 \in K$,

$$\left|\frac{u_p+u_q}{2} - z\right|^2 \geq d^2.$$

On a donc

$$\left|\frac{u_p-u_q}{2}\right|^2 \leq d^2 - d^2 + \varepsilon_p + \varepsilon_q = \varepsilon_p + \varepsilon_q,$$

avec $\varepsilon_n = |u_n - z|^2 - d^2 \longrightarrow 0$. La suite u_n est donc de Cauchy dans H complet, donc converge vers $u \in H$. Comme K est fermé, $u \in K$, et par continuité de la norme, $|u-z| = \mathrm{dist}(z,K)$.

On écrit ensuite simplement que pour tout $v \in K$, l'inégalité $|z-w|^2 \geq |z-u|^2$ est vérifiée pour tout w du segment $[u,v]$ (qu'on écrit $w = u + t(v-u)$, $t \in [0,1]$). □

La démonstration du théorème précédent suggère que toute suite minimisante (u_n) tend nécessairement vers le minimiseur. L'exercice suivant précise cette propriété, en explicitant la vitesse de convergence de la suite des minimiseurs en fonction de la vitesse de convergence de $|u_n - z|$ vers $|u-z|$.

Exercice 6.1. Soit H un espace de Hilbert, K un convexe fermé non vide de H, $z \in H$. On note u la projection de z sur K. Montrer que

$$|v-u| \leq |v-z| \quad \forall v \in K.$$

Exercice 6.2. Soit H un espace de Hilbert, K un convexe fermé non vide de H, $z \in H$. On note u la projection de z sur K. Pour tout $v \in K$, note $d_v = |v-z|$, et $\varepsilon = d_v - d$. Estimer $|v-u|$ en fonction de d_v et ε.

Exercice 6.3. Soit $H = \ell^2$ et K l'ensemble des suites à termes positifs ou nuls. Exprimer la projection d'un élément $z = (z_n)$ sur K.

Remarque 6.9. Si K est un sous-espace affine fermé de H, alors la caractérisation (6.1) prend la forme

$$\begin{cases} u \in K \\ (z-u, v-u) = 0 \quad \forall v \in K, \end{cases} \tag{6.2}$$

et si K est un sous-espace vectoriel de H, on a

$$\begin{cases} u \in K \\ (z-u, v) = 0 \quad \forall v \in K. \end{cases} \tag{6.3}$$

On peut vérifier que l'application de projection P_K définie par le théorème précédent est 1-lipschitzienne

Proposition 6.10. Sous les hypothèses du théorème précédent, on a, pour tous $f, g \in H$,

$$|P_K f - P_K g| \leq |f - g|$$

Démonstration : On utilise la caractérisation de la projection (6.1) :

$$\begin{aligned} (f - P_K f, P_K g - P_K f) &\leq 0, \\ (g - P_K g, P_K f - P_K g) &\leq 0. \end{aligned}$$

En additionnant, il vient,

$$|P_K f - P_K g|^2 \leq (f - g, P_K f - P_K g) \leq |f - g|\,|P_K f - P_K g|\,,$$

d'où l'inégalité annoncée. □

Remarque 6.11. Ne pas confondre le résultat précédent avec le caractère 1-lipschitzien de la fonction distance à un ensemble quelconque, dans tout espace vectoriel normé.

La proposition ci-dessus exprime la stabilité de la projection par rapport à l'élément projeté. On peut se demander si cette projection est stable par rapport à l'ensemble sur lequel on projette. C'est l'objet de l'exercice suivant :

Exercice 6.4. Soit H un espace de Hilbert, et z un élément de H fixé. Pour tout couple (K, K') de convexes fermés bornés, on définit leur distance de Hausdorff par

$$d_H(K, K') = \max\left(\sup_{v \in K} d(v, K'),\ \sup_{v' \in K'} d(v', K)\right).$$

On note $u = P_K z$, $u' = P_{K'} z$. Majorer $|u - u'|$ en fonction de $d_H(K, K')$.

Proposition 6.12. Soit H un espace de Hilbert et K un sous-espace vectoriel fermé de H. Tout u de H s'écrit

$$u = P_K u + P_{K^\perp} u.$$

Démonstration : On vérifie immédiatement que $u - P_K u$ vérifie les identités qui caractérisent la projection de u sur $K^\perp$. □

Proposition 6.13. (Caractérisation de la densité)
Soit H un espace de Hilbert et K un sous-espace de H tel que l'implication suivante soit vérifiée :

$$(h, w) = 0 \quad \forall w \in K \Longrightarrow h = 0.$$

Alors K est dense dans H

DÉMONSTRATION : Si K n'est pas dense dans H, alors il existe $u \in H$, $u \notin \overline{K}$. On pose $h = u - P_{\overline{K}} u$. On a $(h, w) = 0$ pour tout $w \in K$, et $h \neq 0$ car $u \notin \overline{K}$. □

Théorème 6.14. (Hahn-Banach)
Soit H un espace de Hilbert, $K \subset H$ un convexe fermé, et z un point de H qui n'appartient pas à K. Alors il existe un hyperplan fermé qui sépare K et z au sens strict, c'est-à-dire qu'il existe h et x_0 dans H tels que

$$(x - x_0, h) \leq 0 < (z - x_0, h) \quad \forall x \in K.$$

DÉMONSTRATION : On introduit la projection $u = P_K z$ de z sur K, on définit x_0 comme $(z + u)/2$, et $h = z - u$. Pour tout $x \in K$, on a

$$(x - x_0, h) \;=\; \underbrace{(x - u, z - u)}_{\leq 0} + \underbrace{(u - x_0, h)}_{= -|h|^2/2 \leq 0}$$

et on a par ailleurs $(z - x_0, h) = |h|^2 / 2 > 0$. □

EXERCICE 6.5. (Lemme des noyaux)
Soient u, u_1, …, u_n, des éléments d'un espace de Hilbert H. Montrer l'équivalence suivante

$$\left(\bigcap u_i^{\perp} \right) \subset u^{\perp} \quad \Longleftrightarrow \quad \exists \lambda_1, \ldots, \lambda_n \,,\ u = \sum \lambda_i u_i.$$

Définition 6.15. (Orthogonal d'un ensemble)
Soit H un espace de Hilbert et K un sous-ensemble de H. On appelle orthogonal de K l'ensemble

$$K^{\perp} = \{ v \in V \,,\ (v, u) = 0 \quad \forall u \in K \} \,.$$

On vérifie immédiatement que c'est un sous-espace vectoriel fermé.

Proposition 6.16. Soit H un espace de Hilbert et K un sous-espace vectoriel fermé de H. On a

$$K^{\perp\perp} = K.$$

Tout espace de Hilbert peut s'identifier à son dual, comme l'exprime le théorème suivant.

Théorème 6.17. (Riesz-Fréchet)
Soit $\varphi \in H'$ (dual topologique de H). Il existe $f \in H$ unique tel que

$$\langle \varphi \,, u \rangle = (f, u) \quad \forall u \in H. \tag{6.4}$$

De plus, on a $|f| = \|\varphi\|_{H'}$.

DÉMONSTRATION : Si φ est la forme nulle, le résultat est immédiat. Dans le cas contraire, on introduit K le noyau de φ. C'est un hyperplan fermé de H. On construit ensuite un $h \in S_H \cap K^{\perp}$. Pour celà on considère $z \notin K$. D'après la caractérisation (6.3), on a $(z - P_K z, v) = 0$ pour tout $v \in K$. Le vecteur

$$h = \frac{z - P_K z}{|z - P_K z|}$$

convient donc. Pour finir on remarque que tout $v \in H$ peut s'écrire

$$v = \frac{\langle \varphi \,, v \rangle}{\langle \varphi \,, h \rangle} h + \left(v - \frac{\langle \varphi \,, v \rangle}{\langle \varphi \,, h \rangle} h \right) = \lambda h + w,$$

avec $w \in K$. On a donc, pour tout $v \in H$ (on prend le produit scalaire de l'identité précédente avec h),

$$\langle \varphi \, , v \rangle = \langle \varphi \, , h \rangle \, (v, h)$$

d'où l'identité (6.4) avec $f = \langle \varphi \, , h \rangle \, h$. L'unicité d'un tel f est immédiate. □

On prendra garde au fait que cette identification dépend du produit scalaire choisi.

L'identification entre H et son espace dual permet d'étendre immédiatement la caractérisation de la densité 6.13 à un sous-espace du dual :

Proposition 6.18. (Caractérisation de la densité dans le dual)
Soit H un espace de Hilbert et K un sous-espace de H' tel que l'implication suivante soit vérifiée :

$$\langle \varphi \, , h \rangle = 0 \quad \forall \varphi \in K \Longrightarrow h = 0.$$

Alors K est dense dans H'.

Proposition 6.19. (Continuité d'une forme bilinéaire)
Soit $a : H \times H \longrightarrow \mathbb{R}$ une forme bilinéaire. Alors a est continue si et seulement s'il existe une constante $\|a\|$ telle que

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

Démonstration : On suppose a continue. La continuité en 0 assure l'existence d'un r tel que $|a(u, v)| \leq 1$ sur $\overline{B(0, r)} \times \overline{B(0, r)}$. On a donc, pour tous u, v, non nuls

$$\left| a \left(r \frac{u}{|u|}, r \frac{v}{|v|} \right) \right| \leq 1 \Longrightarrow |a(u, v)| \leq \frac{1}{r^2} \, |u| \, |v| \, .$$

Réciproquement, le développement

$$a(u + h, v + k) = a(u, v) + a(h, v) + a(u, k) + a(h, k)$$

assure la continuité en tout $(u, v) \in H \times H$. □

Définition 6.20. (Coercivité d'une forme bilinéaire)
Soit $a : H \times H \longrightarrow \mathbb{R}$ une forme bilinéaire. On dit que a est coercive s'il existe $\alpha > 0$ tel que

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

Remarque 6.21. En dimension finie, et dans le cas où la forme est symétrique ($a(u, v) = a(v, u)$), on retrouve la notion de forme symétrique définie positive. Le plus grand coefficient α est alors la plus petite valeur propre de la matrice associée, et la plus petite constante $\|a\|$ de la continuité sa plus grande valeur propre.

Exercice 6.6. Soit $\alpha = (\alpha_n)$ une suite bornée de réels, et

$$a \ : \ (u, v) \in \ell^2 \times \ell^2 \longmapsto \sum_{n=0}^{+\infty} \alpha_n u_n v_n.$$

A quelle condition sur α la forme bilinéaire $a(\cdot, \cdot)$ est-elle coercive ?

Remarque 6.22. On verra qu'il existe une définition plus générale de la coercivité (pour des fonctionnelles quelconques, voir théorème 6.40), équivalente à la définition ci-dessus dans le cas particulier des formes bilinéaires.

Proposition 6.23. Soit H un espace de Hilbert, et a une forme bilinéaire et continue sur l'espace produit $H \times H$. Pour tout $u \in H$, on note Au l'élément de H qui s'identifie à la forme linéaire $a(u, \cdot)$:

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

L'application $u \longmapsto Au$ est linéaire et continue. De plus si $a(\cdot, \cdot)$ est coercive, alors l'application A est une bijection.

Démonstration : L'application A est évidemment linéaire, et

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

où $\|a\|$ est la constante de continuité de a.

Si a est coercive, on a $(Au, u) = a(u, u) \geq \alpha\,|u|^2$, et donc $|Au| \geq \alpha\,|u|$ pour tout u dans H. On vérifie que l'image est fermée en considérant une suite (Au_n) qui converge vers un élément de l'image w. Comme (Au_n) converge, elle est de Cauchy, donc (u_n) est également de Cauchy d'après l'inégalité précédemment démontrée. Elle converge donc vers $u \in H$ qui vérifie $Au = w$ par continuité de A. On a de plus, pour tout $g \in H$,

$$(g, Au) = 0 \quad \forall u \in H \Longrightarrow (g, Ag) = a(g, g) = 0$$

qui entraîne $g = 0$ par coercivité de a. L'image de A est donc fermée et dense dans H : c'est l'espace H lui-même. L'injectivité est une conséquence immédiate de la coercivité. □

Remarque 6.24. On peut choisir de définir A comme un opérateur de H dans H', en écrivant alors $\langle Au\,, v\rangle = a(u, v)$ pour tout $v \in H$. Les résultats précédents s'étendent bien entendu à cette situation.

On verra que l'opérateur A est bicontinu (*i.e.* son inverse est lui-même continu), mais cette propriété n'est pas utile pour démontrer le point essentiel de cette section, conséquence directe de la proposition qui précède :

Théorème 6.25. (Lax-Milgram)
Soit H un espace de Hilbert, et a une forme bilinéaire continue et coercive sur $H \times H$. Pour tout $\varphi \in H'$, il existe un $u \in H$ unique tel que

$$a(u, v) = \langle \varphi\,, v\rangle \quad \forall v \in H. \tag{6.5}$$

Si a est symétrique, u est l'unique élément de H qui réalise le minimum de la fonctionnelle

$$v \longmapsto J(v) = \frac{1}{2} a(v, v) - \langle \varphi\,, v\rangle.$$

Démonstration : D'après le théorème de représentation de Riesz-Fréchet, il existe un unique $f \in H$ tel que

$$(f, v) = \langle \varphi\,, v\rangle \quad \forall v \in H.$$

On introduit l'opérateur A associé à $a(\cdot, \cdot)$, qui est bijectif (voir proposition 6.23). Il existe donc une unique solution u à l'équation $Au = f$.

On suppose maintenant $a(\cdot, \cdot)$ symétrique. On note toujours u la solution du problème variationnel (6.6). Pour tout $h \in H$, l'application

$$t \longmapsto \psi(t) = J(u + th) - J(u)$$

est convexe, nulle en 0, de dérivée nulle en 0. Elle est donc positive, et ainsi $J(u+h) \geq J(u)$ pour tout $h \in H$.

De la même manière, si w minimise J, on écrit que la dérivée de la fonction $J(w+th)-J(w)$ est nulle en 0, ce qui est exactement la formulation variationnelle (6.6). □

Corollaire 6.26. Soit H un espace de Hilbert, $K \subset H$ un sous-espace affine fermé, K^0 l'espace vectoriel sous-jacent. et a une forme bilinéaire continue sur $H \times H$, coercive sur K^0. Pour tout $\varphi \in H'$, il existe un $u \in K$ unique tel que

$$a(u, v) = \langle \varphi \, , v \rangle \quad \forall v \in K^0. \tag{6.6}$$

Si a est symétrique, u est l'unique élément de K qui réalise le minimum de la fonctionnelle

$$v \longmapsto J(v) = \frac{1}{2} a(v, v) - \langle \varphi \, , v \rangle.$$

Démonstration : On écrit simplement $K = U + K^0$, et l'on cherche la solution sous la forme $u = U + \tilde{u}$, pour se ramener au problème

$$a(\tilde{u}, v) = \langle \varphi \, , v \rangle - a(U, v) \quad \forall v \in K^0,$$

qui rentre dans le cadre du théorème de Lax-Milgram. Le principe de minimisation s'en déduit, du fait que

$$\begin{aligned} J(U + \tilde{v}, U + \tilde{v}) &= J(U, U) + \frac{1}{2} a(\tilde{v}, \tilde{v}) + a(U, \tilde{v}) - \langle \varphi \, , U \rangle - \langle \varphi \, , \tilde{v} \rangle \\ &= \frac{1}{2} a(\tilde{v}, \tilde{v}) - (\langle \varphi \, , \tilde{v} \rangle - a(U, \tilde{v})) + \text{ constante} \end{aligned}$$

□

L'identification établie ci-dessus permet de donner un sens à la notion de différentielle d'une application à valeurs dans $\mathbb{R}$ en tant qu'élément de l'espace de Hilbert :

Définition 6.27. (Différentiabilité)
Soit J une application de H dans $\mathbb{R}$, et $u \in H$. On dit que J est différentiable en u s'il existe $\varphi \in H'$ tel que l'on ait, pour h au voisinage de 0,

$$J(u + h) = J(u) + \langle \varphi \, , h \rangle + |h| \, \varepsilon(h),$$

où $\varepsilon : H \longrightarrow H$ est telle que $\varepsilon(h) \longrightarrow 0$ quand $h \longrightarrow 0$. Si un tel φ existe, on peut l'identifier à un élément de H que l'on note $J'(u)$. On dira que J est différentiable si elle admet une différentielle en tout point, et que J est C^1 si l'application $u \longmapsto J'(u)$ est continue.

6.2. Convergence faible

Comme précédemment H désigne un espace de Hilbert réel muni du produit scalaire $(.,.)$ et de la norme $|\ |$.

Définition 6.28. (Convergence faible)
Soit (u_n) une suite d'éléments de H. On dit que (u_n) converge faiblement vers u dans H, et on note $u_n \rightharpoonup u$, si

$$(u_n, v) \to (u, v) \quad \forall v \in H,$$

ou de façon équivalente, si

$$< \varphi, u_n > \longrightarrow < \varphi, u > \quad \forall \varphi \in H'.$$

Proposition 6.29. Soit (u_n) une suite d'un espace de Hilbert H. Si $u_n \rightharpoonup u$, alors (u_n) est bornée et $|u| \leq \liminf |u_n|$.

Démonstration : C'est une conséquence directe du corollaire 7.8 au théorème de Banach-Steinhaus. □

Proposition 6.30. Si $u_n \rightharpoonup u$ et $|u_n| \to |u|$, alors la suite u_n converge fortement vers u.

Démonstration : On écrit

$$|u_n - u|^2 = |u_n|^2 - 2(u_n, u) + |u|^2 .$$

On a $(u_n, u) \to |u|^2$ d'où $|u_n - u|^2 \to 0$. □

Proposition 6.31. Soient E et F deux espaces de Hilbert, et $T \in \mathcal{L}(E, F)$. Alors

$$u_n \rightharpoonup u \Longrightarrow Tu_n \rightharpoonup Tu.$$

Démonstration : On écrit simplement que, pour tout $z \in F$,

$$(Tu_n, z) = (u_n, T^\star z) \longrightarrow (u, T^\star z) = (Tu, z),$$

qui exprime la convergence faible de Tu_n vers Tu. □

Le résultat fondamental de cette section est le suivant.

Théorème 6.32. Soit (u_n) une suite **bornée** dans un espace de Hilbert H. Alors on peut extraire une sous-suite convergeant faiblement vers u dans H.

Démonstration : On raisonne d'abord dans le cas où H est séparable. Il existe donc une famille dénombrable $\{x_k\}_{k\in\mathbb{N}}$ dense dans H. On se propose de suivre le procédé d'extraction diagonale de Cantor.

(1) Comme (u_n, x_1) est bornée dans $\mathbb{R}$ on peut extraire une suite $u_{j_1(n)}$ telle que $(u_{j_1(n)}, x_1)$ converge.
(2) Comme $(u_{j_1(n)}, x_2)$ est bornée dans $\mathbb{R}$ on peut extraire de $u_{j_1(n)}$ une suite $u_{j_1 \circ j_2(n)}$ telle que $(u_{j_1\circ j_2(n)}, x_2)$ converge.
(3) Par récurrence, on construit une suite de sous-suites emboitées $u_{j_1\circ j_2 \circ \cdots \circ j_k(n)}$ telle que $(u_{j_1\circ j_2\circ\cdots\circ j_k(n)}, x_k)$ converge, pour tout k.

(4) On utilise à présent le procédé d'extraction diagonale : on pose $\varphi(k) = j_1 \circ j_2 \circ \cdots \circ j_k(k)$ (de telle sorte que φ est strictement croissante), et on considère $u_{\varphi(n)}$. Pour tout k, on remarque que $u_{\varphi(n)}$, à partir du rang k, est aussi une suite extraite de $(u_{j_1 \circ j_2 \circ \cdots \circ j_k(n)})$, de telle sorte que $(u_{\varphi(n)}, x_k)$ converge lorsque $n \to +\infty$.

(5) On utilise ensuite la densité des x_k. Pour tout $x \in H$, on montre que $(u_{\varphi(n)}, x)$ est une suite de Cauchy : soit $\varepsilon > 0$, il existe (x_k) tel que $|x - x_k| < \varepsilon$. Comme $(u_{\varphi(n)}, x_k)$ est de Cauchy, il existe un N au-delà duquel $\left|(u_{\varphi(p)}, x_k) - (u_{\varphi(q)}, x_k)\right| < \varepsilon$. Pour tous p, q supérieurs à N, on a donc

$$\begin{aligned}\left|(u_{\varphi(p)}, x) - (u_{\varphi(q)}, x)\right| &\leq \left|(u_{\varphi(p)}, x) - (u_{\varphi(p)}, x_k)\right| + \left|(u_{\varphi(p)}, x_k) - (u_{\varphi(q)}, x_k)\right| \\ &\quad + \left|(u_{\varphi(q)}, x_k) - (u_{\varphi(q)}, x)\right| \\ &\leq M\varepsilon + \varepsilon + M\varepsilon = (1 + 2M)\varepsilon,\end{aligned}$$

où M est un majorant de $|u_n|$.

On a donc démontré que, pour tout $x \in H$, $(u_{\varphi(n)}, x)$ converge vers un élément de H que l'on note $h(x)$. L'application $x \mapsto h(x) \in \mathbb{R}$ est linéaire, et on a pour tout $x \in H$

$$|h(x)| = \lim_{n \to \infty} \left|(u_{\varphi(n)}, x)\right| \leq M\,|x|\,,$$

d'où h continue [1] sur H. D'après le théorème de Riesz-Fréchet, cette forme s'identifie à un élément u de H. On a donc convergence faible de la suite extraite vers u.

Dans le cas où le Hilbert n'est pas séparable, on se place dans l'adhérence de l'espace vectoriel engendré par les termes de la suite, qui est un espace de Hilbert séparable (pour le même produit scalaire) par construction. La convergence faible vers un u de ce sous-espace entraîne la convergence faible dans H.

6.3. Minimisation de fonctionnelles convexes

Commençons par définir un certain nombre de notions générales afférentes aux applications à valeurs dans $\mathbb{R} \cup \{+\infty\}$.

Définition 6.33. (Domaine)
Soit E un ensemble et J une application de E dans $\mathbb{R} \cup \{+\infty\}$. On appelle domaine de J l'ensemble

$$D(J) = \{x \in E\,,\ J(x) < +\infty\}\,.$$

Définition 6.34. (Semi-continuité inférieure)
Soit E un espace topologique, et J une application de E dans $\mathbb{R} \cup \{+\infty\}$. On dit que J est semi-continue inférieurement (s.c.i. en abrégé) si, pour tout $\lambda \in \mathbb{R}$, l'ensemble

$$E_\lambda = \{x \in E\,,\ J(x) \leq \lambda\}$$

est fermé.

Définition 6.35. (Convexité)
Soit E un espace vectoriel, et J une application de E dans $\mathbb{R} \cup \{+\infty\}$. On dit que J est convexe si

$$J(\theta x + (1-\theta)y) \leq \theta J(x) + (1-\theta)J(y) \quad \forall x, y \in E \quad \forall \theta \in]0,1[,$$

1. Remarquer qu'il n'est pas nécessaire ici d'utiliser le théorème de Banach–Steinhaus, du fait de l'hypothèse (u_n) bornée.

ou, de façon équivalente, si l'ensemble (appelé épigraphe de J)

$$\operatorname{epi} J = \{(x,\lambda) \in E \times \mathbb{R}\,,\ J(x) \leq \lambda\}\,,$$

est convexe.

On dit que J est strictement convexe si

$$J(\theta x + (1-\theta)y) < \theta J(x) + (1-\theta)J(y) \quad \forall x, y \in E \quad \forall \theta \in]0,1[.$$

Définition 6.36. (Coercivité)
Soit E un vectoriel normé, et J une application de E dans $\mathbb{R} \cup \{+\infty\}$. On dit que J est coercive si

$$\lim_{\|x\| \to +\infty} J(x) = +\infty.$$

Théorème 6.37. (Banach-Saks)
Soit $(x_n)_{n\in\mathbb{N}}$ une suite de H faiblement convergente vers un élément x de H. Alors il existe une suite extraite $y_n = x_{\varphi(n)}$ telle que la suite des moyennes de Césaro

$$\sigma_n = \frac{1}{n}\sum_{k=1}^{n} y_k$$

converge fortement vers x.

Démonstration : Quitte à remplacer la suite x_n par $x_n - x$, on peut supposer sans perte de généralité que $x_n \rightharpoonup 0$. On construit maintenant la suite y_n de la façon suivante :

(1) On prend $y_1 = x_1$.
(2) Comme x_n converge faiblement vers 0, il existe un indice $\varphi(2)$ tel que
$$\left|(y_1, x_{\varphi(2)})\right| = |(y_1, y_2)| \leq \frac{1}{2}.$$
(3) Par récurrence, on construit à partir des termes déjà construits y_1, y_2, …, y_{n-1}, le $n-$ième terme y_n tel que
$$|(y_i, y_n)| \leq \frac{1}{n} \quad \forall i = 1, 2, \ldots, n-1.$$

On pose

$$\sigma_n = \frac{1}{n}\sum_{k=1}^{n} y_k.$$

Montrons que σ_n tend (fortement) vers 0. On développe

$$|\sigma_n|^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(y_i, y_j),$$

ce qui donne

$$\begin{aligned} |\sigma_n|^2 &\leq \frac{1}{n^2}\left(\sum_{i=1}^{n}|y_i|^2 + 2\sum_{k=1}^{n}\sum_{\ell=1}^{k-1}|(y_\ell, y_k)|\right) \leq \frac{1}{n^2}\left(nM^2 + 2\sum_{k=1}^{n}\frac{k-1}{k}\right) \\ &\leq \frac{1}{n^2}\left(nM^2 + 2n\right) = \frac{M^2+2}{n}, \end{aligned}$$

et donc $\sigma_n \to 0$. □

Ce théorème a plusieurs conséquences importantes, dont la première est le

Théorème 6.38. Soit $K \subset H$ un ensemble convexe fermé de H. Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de K qui converge faiblement vers x. Alors $x \in K$. On dit que K est faiblement séquentiellement fermé.

Démonstration : Le résultat est une conséquence directe du théorème 6.37. □

Exercice 6.7. Montrer que le résultat est faux en général si l'on supprime l'hypothèse de convexité (donner par exemple une suite dans la sphère unité de ℓ^2 qui converge faiblement vers 0).

Une autre conséquence importante du théorème 6.37 est le

Théorème 6.39. Soit $J : H \longrightarrow \mathbb{R}$ une fonction convexe continue s.c.i, $J \not\equiv +\infty$. Pour toute suite $(x_n)_{n\in\mathbb{N}}$ de H telle que $x_n \rightharpoonup x$, on a

$$J(x) \leq \liminf J(x_n).$$

(On dit que J est faiblement séquentiellement s.c.i.)

Démonstration : Soit $L := \liminf J(x_n)$ (a priori, $-\infty \leq L \leq +\infty$). Soit y_n une suite extraite telle que l'on ait

$$J(y_n) \longrightarrow L,$$

et telle que

$$\sigma_n = \frac{1}{n}\sum_{i=1}^{n} y_n \longrightarrow x.$$

par semi–continuité inférieure de J, on a $J(x) \leq \liminf J(\sigma_n)$. D'autre part, J étant convexe

$$J(\sigma_n) \leq \frac{1}{n}\sum_{i=1}^{n} J(y_n) \to L.$$

On a donc bien $J(x) \leq L$. □

Ce théorème va nous permettre d'établir le résultat principal de minimisation :

Théorème 6.40. Soit $J : H \longrightarrow \mathbb{R}$ une fonction convexe s.c.i., $J \not\equiv +\infty$. On suppose que J est coercive, c'est-à-dire que

$$\lim_{|x|\to+\infty} J(x) = +\infty.$$

Alors il existe $u \in H$ tel que

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

Plus généralement, si $K \subset H$ est un convexe fermé, il existe $u \in K$ tel que

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

Enfin, si J est strictement convexe, alors ces minima sont uniques.

DÉMONSTRATION : Soit $(x_n)_{n\in\mathbb{N}}$ une suite minimisante : $x_n \in K$ et

$$J(x_n) \to M := \inf_K J.$$

Comme J est coercive, x_n est bornée. Il existe donc une suite extraite y_n telle que $y_n \rightharpoonup x$. Comme K est un convexe fermé, $x \in K$, et

$$J(x) \leq \liminf J(x_n) = M.$$

Mais comme $J(x)$ M par définition de M, on a $J(x) = M$. □

On remarquera que, pour le résultat concernant K, il suffit que J soit définie sur K. La coercivité signifie que, ou bien K est borné, ou bien

$$\lim_{|x|\to+\infty, x\in K} J(x) = +\infty.$$

Chapitre 7

Autour du théorème de Banach-Steinhaus

Définition 7.1. On appelle espace de Banach tout espace vectoriel normé complet.

Définition 7.2. Soient E et F deux espaces vectoriels normés. On note $\mathcal{L}(E,F)$ l'espace des applications linéaires continues de E dans F. C'est un espace vectoriel normé pour la norme

$$\|T\|_{\mathcal{L}(E,F)} = \sup_{u\neq 0} \frac{\|Tu\|_F}{\|u\|_E} = \sup_{u\in B_E} \|Tu\|_F.$$

Cet espace est complet dès que F est complet. Lorsque $F = E$, on notera simplement $\mathcal{L}(E)$.

Définition 7.3. (Adjoint)
Soient V et Λ deux espaces vectoriels normés, et $T \in \mathcal{L}(V,\Lambda)$. On définit l'adjoint de T comme l'opérateur $T^\star$ de Λ' dans V' qui à $\varphi \in \Lambda'$ associe

$$T^\star\varphi \ : \ u \longmapsto \langle T^\star\varphi\,,u\rangle = \langle \varphi\,, Tu\rangle.$$

On vérifie immédiatement que $T^\star \in \mathcal{L}(\Lambda', V')$, avec $\|T^\star\| = \|T\|$.

Proposition 7.4. Soit E un espace vectoriel normé. Alors E est un espace de Banach si et seulement si toute série absolument convergente est convergente.

Proposition 7.5. Soit E un espace de Banach, et K un sous-espace vectoriel fermé de E. Pour tout $\tilde{x} \in E/K$, on définit

$$\|\tilde{x}\|_{E/K} = \inf_{y\in\tilde{x}} \|y\| = \inf_{h\in K} \|x-h\|.$$

L'espace E/K est complet pour la norme $\|\cdot\|_{E/K}$.

Démonstration : On note dans la suite $\|\cdot\|$ la norme quotient. On vérifie immédiatement que $\|\cdot\|$ est une semi-norme. Soit $\tilde{x}$ tel que$\|\tilde{x}\| = 0$ et soit x un représentant de la classe $\|x\|$. La borne inférieure de $\|x-z\|$ pour z décrivant K est 0, ce qui signifie que la distance de x à M est nulle, d'où $x \in \overline{K} = K$, d'où $\tilde{x} = 0$. On montre que E est complet en utilisant a proposition 7.4. Soit donc $(\tilde{x}_n)$ telle que $\sum |\tilde{x}_n| < +\infty$. Par définition de la norme quotient, pour tout $n \in \mathbb{N}$, il existe $z_n \in M$ tel que

$$|\tilde{x}_n| \leq \|x_n + z_n\| \leq |\tilde{x}_n| + \frac{1}{2^n}.$$

La série $\sum(x_n + z_n)$ est donc absolument convergente dans E. Comme E est complet, cette série converge vers $y \in E$. On vérifie immédiatement que la série $\sum \tilde{x_n}$ converge vers $\tilde{y}$. □

Lemme 7.6. (Baire)
Soit X un espace métrique complet, et $(X_n)_{n\in\mathbb{N}}$ une suite de fermés de X. On suppose que

$$\operatorname{int}(X_n) = \emptyset \quad \forall n \in \mathbb{N}.$$

On a alors

$$\operatorname{int}\left(\bigcup_{n=0}^{+\infty} X_n\right) = \emptyset.$$

On utilise pour la démonstration une formulation équivalente du théorème : soit X un espace métrique complet, et $(U_n)_{n\in\mathbb{N}}$ une suite d'ouverts de X denses dans X. Alors l'intersection des U_n est dense dans X.

Démonstration : On introduit

$$U = \bigcap_{n\in\mathbb{N}} U_n,$$

et on se donne $x \in X$. Pour toute boule $B(x,r)$, on va construire une suite u_n qui converge vers une limite u dans $B(x,r)\cap U$, ce qui établira la densité de U. Comme U_0 est un ouvert dense, il existe $u_0 \in U_0$ et $r_0 > 0$, avec $r_0 \leq r/2$, tel que

$$\overline{B}(u_0, r_0) \subset B(x,r) \cap U_0.$$

On construit par récurrence la suite (u_n) de la façon suivante : supposons u_k et r_k construits pour $k \leq n$, la densité de l'ouvert U_{n+1} assure l'existence d'une boule fermée $\overline{B}(u_{n+1}, r_{n+1})$ incluse dans $U_{n+1} \cap B(u_n, r_n)$, et telle que $r_{n+1} \leq r_n/2$. Les suites (u_n) et $(r_n)_{n\in\mathbb{N}}$ ainsi construites, on vérifie immédiatement que

$$d(u_n, u_{n+1}) < r_n \leq \frac{r}{2^n},$$

d'où l'on déduit que (u_n) est de Cauchy, donc qu'elle converge vers une limite $u \in X$. Par construction, u est dans $B(x,r)$ et dans chacune des boules fermées $\overline{B}(u_n, r_n)$, donc dans U, ce qui termine la démonstration. □

Théorème 7.7. (Banach-Steinhaus)
Soient E et F deux espaces vectoriels normés et $(T_a)_{a\in A}$ une famille d'opérateurs de $\mathcal{L}(E,F)$. On suppose

$$\sup_{a\in A} \|T_a x\|_F < +\infty \quad \forall x \in E. \tag{7.1}$$

On a alors

$$\sup_{a\in A} \|T_a\|_{\mathcal{L}(E,F)} < +\infty.$$

Démonstration : On introduit les ensembles

$$E_n = \left\{ x \in E\,,\ \sup_{a\in A} \|T_a x\|_F \leq n \right\}.$$

Les E_n sont des fermés de E comme intersection de fermés. D'autre part leur réunion est E tout entier, d'après l'hypothèse. L'un des E_n est donc d'intérieur non vide. Soit n_o tel que $\operatorname{int}(E_{n_o}) \neq \emptyset$. Il existe $x_0 \in E$ et $\rho > 0$ tel que $\overline{B}(x_0, \rho) \subset E_{n_o}$, d'où

$$\|T_a(x_0 + \rho u)\| \leq n_o \quad \forall u \in B_E.$$

On a donc, pour tout $a \in A$,

$$\|T_a\|_{\mathcal{L}(E,F)} \leq \frac{1}{\rho}\left(n_o + \sup_{a\in A}\|T_a(x_0)\|_F\right),$$

ce qui conclut la démonstration. □

EXERCICE 7.1. Montrer qu'un espace de Banach est de dimension soit finie soit non dénombrable.

Corollaire 7.8. Soient E et F deux espaces de Banach et $(T_n)_{n\in\mathbb{N}}$ une suite d'opérateurs de $\mathcal{L}(E,F)$ telle que, pour tout $x \in E$, $T_n x$ converge vers un élément de F, que l'on note Tx. La suite (T_n) est alors nécessairement bornée dans $\mathcal{L}(E,F)$. De plus, l'opérateur limite T est dans $\mathcal{L}(E,F)$, et sa norme vérifie

$$\|T\|_{\mathcal{L}(E,F)} \leq \liminf_{n\to+\infty}\|T_n\|_{\mathcal{L}(E,F)}.$$

DÉMONSTRATION : On applique le théorème de Banach-Steinhaus à la suite d'opérateurs (T_n). On en déduit que (T_n) est bornée, et que l'on a

$$\|T\| = \sup_{x\in B_E}\|Tx\|_{\mathcal{L}(E,F)} = \sup_{x\in B_E}\lim_{n\to+\infty}\|T_n x\|_{\mathcal{L}(E,F)} \leq \liminf_{n\to+\infty}\|T_n\|_{\mathcal{L}(E,F)},$$

d'où la majoration de $\|T\|_{\mathcal{L}(E,F)}$.

Remarque 7.9. La dernière inégalité du corollaire précédent peut être stricte. Considérer par exemple $E = \ell^2$ et la suite des formes linéaires

$$T_k \ : \ x = (x_n)_{n\in\mathbb{N}} \longmapsto x_k \in \mathbb{R}.$$

Cette suite converge ponctuellement vers la forme linéaire nulle. Cet exemple permet d'autre part de vérifier que l'on n'a pas en général convergence de T_k vers T pour la norme d'opérateur.

Remarque 7.10. On prendra garde au fait que l'hypothèse (7.1) du théorème de Banach-Steinaus, (tout comme l'hypothèse de convergence de $T_n x$ du corollaire ci-dessus), doit être vérifiée pour tout x de E, et non pas seulement sur un sous-ensemble dense.

Théorème 7.11. (Application ouverte)
Soient E et F deux espaces vectoriels normés et soit $T \in \mathcal{L}(E,F)$ surjectif. Alors il existe une constante c telle

$$B_F(0,c) \subset T(B_E).$$

DÉMONSTRATION : La démonstration s'effectue en deux étapes. On montre dans un premier temps que l'adhérence de $T(B_E)$ contient une boule ouverte centrée en 0. On note $K = \overline{T(B_E)}$ cette adhérence. Les $K_n = nK$ sont des fermés de F par construction, et leur union est F tout entier (car T est surjective). L'espace d'arrivée F étant complet, il en existe donc un d'intérieur non vide (d'après le lemme de Baire), donc K lui–même est d'intérieur non vide (les K_n sont homothétiques à K) : K contient une boule ouverte $B = B(a,2c)$. Par symétrie de K, $-K$ est également dans K, et par convexité (K est l'adhérence de l'image d'un convexe par une application linéaire),

$$\begin{aligned}\frac{1}{2}B + \frac{1}{2}(-B) &= \left\{\frac{1}{2}a - \frac{1}{2}a + \frac{1}{2}h_1 - \frac{1}{2}h_2\,,\ h_1\,,\ h_2 \in B(0,2c)\right\}\\ &= \{0+h\,,\ h\in B(0,2c)\}\\ &= B(0,2c) \subset K = \overline{T(B_E)}.\end{aligned}$$

On va maintenant montrer, en utilisant cette fois la complétude de l'espace de départ E, que $T(B_E)$ contient $B(0,c)$. On se donne $y \in B(0,c)$, et on cherche à construire un antécédent x dans B_E. D'après ce que l'on vient détablir, pour tout $\varepsilon > 0$, il existe z dans $1/2B_E$ tel que $\|y - Tz\| < \varepsilon$. On construit ainsi z_1 tel que

$$\|y - Tz_1\| < \frac{c}{2}\,,\ \|z_1\| \leq \frac{1}{2}.$$

On construit de la même manière z_2

$$\|(y - Tz_1) - Tz_2\| < \frac{c}{4}\,,\ \|z_1\| \leq \frac{1}{4},$$

puis par récurrence les termes de la suite (z_n) tels que

$$\|z_n\| \leq \frac{1}{2^n}.$$

Par construction la suite $x_n = z_1 + \cdots + z_n$ est de Cauchy, donc converge vers un certain x de norme inférieure ou égale à 1, et on a bien $y = Tx$ par continuité de T. □

On en déduit le

Corollaire 7.12. Soient E et F deux espaces de Banach. et soit $T \in \mathcal{L}(E,F)$ bijectif . Alors T^{-1} est continu de F dans E.

Démonstration : D'après le théorème de l'application ouverte, pour tout élément y de F, $cy/(2\|y\|)$ admet un antécédent de norme 1. On a donc

$$\left\|T^{-1}y\right\| = \frac{2}{c}\|y\| \left\|T^{-1}\left(\frac{c}{2\|y\|}y\right)\right\| \leq \frac{2}{c}\|y\|\,,$$

d'où la continuité de l'opérateur réciproque. □

Dans le cas où T n'est pas surjectif, on peut appliquer ce qui précède à l'application $\tilde{T}$, bijection canoniquement associée à T comme le précise le corollaire ci -dessous.

Corollaire 7.13. Soient V et Λ deux espaces de Banach, et $T \in \mathcal{L}(V,\Lambda)$. On suppose que l'image de T est fermée. L'application $\tilde{T}$ définie de $V/\ker T$ dans $T(V)$ par $\tilde{T}\tilde{x} = Tx$ est une bijection bicontinue. En particulier, il existe une constante α telle que

$$\|\tilde{u}\|_{V/\ker T} = \inf_{h \in \ker T} \|u - h\| \leq \alpha \|Tu\|\,.$$

Remarque 7.14. Dans le cas où V est un espace de Hilbert, l'infimum est atteint pour h égal à la projection de u sur $\ker T$, l'inégalité ci-dessus devient

$$\left|P_{(\ker T)^\circ} u\right| \leq \alpha \|Tu\|\,.$$

Proposition 7.15. Soient V et Λ deux espaces de Banach, et $T \in \mathcal{L}(V,\Lambda)$. L'image de T est fermée si et seulement s'il existe $\alpha > 0$ tel que

$$\forall y \in T(V)\,,\ \exists x \in V\,,\ \|x\| \leq \alpha\|y\|\ ,\ y = Tx. \tag{7.2}$$

Démonstration : La condition nécessaire est une conséquence directe du corollaire précédent. En effet, si l'on note α la constante de continuité de l'aplication $\tilde{T}^{-1}$, on a

$$\forall y \in T(V)\,,\ \left\|\tilde{T}^{-1}y\right\|_{V/\ker T} \leq \alpha\|y\|\,.$$

Soit z un élément de la classe $\tilde{T}^{-1}y$, on a

$$\left\|\tilde{T}^{-1}y\right\|_{V/\ker T} = \|z - P_{\ker T}z\|,$$

d'où la propriété avec $x = z - P_{\ker T}z$.

Réciproquement, si un tel α existe, alors pour tout suite (x_n) telle que $Tx_n \to y$, on peut construite une suite bornée x'_n avec $Tx_n = Tx'_n$, dont on peut extraire une sous-suite faiblement convergence (toujours notée (x'_n)) vers $x \in V$. La proposition 6.31 assure alors la convergence faible de Tx'_n vers Tx, d'où $y = Tx \in T(V)$. □

Remarque 7.16. On déduit immédiatement de ce qui précède que l'image d'un sous-espace fermé par une application linéaire injective à image fermée est fermée (comme image réciproque d'un fermé par l'application réciproque, qui est continue).

Définition 7.17. (Polaire d'un ensemble)
Soit V un espace de Banach et K un sous-espace vectoriel de V. On appelle polaire de K l'ensemble

$$K^\circ = \left\{\varphi \in V',\ \langle \varphi\,,\, u\rangle = 0 \quad \forall u \in K\right\}.$$

Les propriétés qui suivent sont essentielles pour établir les résultats afférents à l'existence et l'unicité de point-selle On se reportera à Brezis [**3**] pour un exposé plus complet des propriétés de l'opérateur adjoint.

Proposition 7.18. Soient V et Λ deux espaces de Banach, et $T \in \mathcal{L}(V, \Lambda)$. On a

$$\overline{\mathrm{Im}(T^\star)} \subset (\ker T)^\circ.$$

Dans le cas où V est un espace de Hilbert (et plus généralement dans le cas où V est réflexif), on a l'identité

$$\overline{\mathrm{Im}(T^\star)} = (\ker T)^\circ.$$

Démonstration : Soit $\varphi \in T^\star(\Lambda')$, donc de la forme $T^\star\lambda$. On a, pour tout $u \in \ker T$,

$$\langle \varphi\,,\, u\rangle = \langle T^\star\lambda\,,\, u\rangle = \langle \lambda\,,\, Tu\rangle = 0,$$

d'où $T^\star(\Lambda') \subset (\ker T)^\circ$. Comme $(\ker T)^\circ$ est fermé, cela entraîne $\overline{T^\star(\Lambda')} \subset (\ker T)^\circ$.

Montrons que cette inclusion ne peut être stricte dans le cas Hilbertien. Supposons qu'elle le soit. Il existe alors $\varphi_0 \in (\ker T)^\circ$ non élément de l'adhérence de $T^\star(\Lambda')$. Le théorème de Hahn-Banach permet de séparer strictement φ_0 du convexe fermé $\overline{T^\star(\Lambda')}$: il existe [1] $h \in V$ et $\alpha \in \mathbb{R}$ tels que

$$(T^\star\lambda, h) \leq \alpha < \langle \varphi_0\,,\, h\rangle \quad \forall \lambda \in \Lambda'.$$

Comme Λ' est un espace vectoriel, l'ensemble des valeurs prises par $(T^\star\lambda, h)$ est soit $\{0\}$ soit $\mathbb{R}$ tout entier. D'après l'inégalité précédente, c'est nécessairement $\{0\}$. On a donc $\langle \lambda\,, Th\rangle = 0$ pour tout $\lambda \in \Lambda'$ d'où $h \in \ker T$, mais alors $\langle \varphi_0\,,\, h\rangle = 0$, ce qui est en contradiction avec l'inégalité ci-dessus. On a donc bien identité entre les deux ensembles. □

1. C'est ici qu'intervient l'hypothèse de réflexivité de V, dans le fait que la forme linéaire sur V' est de la forme $\varphi \mapsto \langle \varphi\,,\, h\rangle$

Proposition 7.19. Soient V et Λ deux espaces de Banach, et $T \in \mathcal{L}(V, \Lambda)$. Les assertions suivantes sont équivalentes :

(i) $\mathrm{Im}(T)$ est fermée.

(ii) $\mathrm{Im}(T^\star)$ est fermée.

(iii) Il existe $C > 0$ tel que

$$\forall z \in \mathrm{Im}(T)\,,\ \exists u \in V\,,\ z = Tu\,,\ \|u\| \leq C\,\|z\|\,,$$

ou, de façon équivalente

$$\|\tilde{u}\|_{V/\ker T} \leq C\,\|Tu\|\,.$$

(iv) Il existe $\beta > 0$ tel que

$$\sup_{u \in V} \frac{\langle \lambda\,, Tu\rangle}{\|u\|} \geq \beta\,\|\lambda\|_{\Lambda'/\ker T^\star}\,.$$

Démonstration : $(i) \Longrightarrow (ii)$ On suppose que l'image de T est fermée. On introduit l'opérateur $\tilde{T}$ qui est l'opérateur T considéré comme allant de V dans $\tilde{\Lambda} = T(V)$, qui est un espace de Hilbert par hypothèse. L'opérateur $\tilde{T}$ est surjectif par construction. Montrons que l'image de $\tilde{T}^\star$ s'identifie à celle de $T^\star$. Pour tout $\tilde{\lambda} \in \tilde{\Lambda}'$, on construit $\lambda \in \Lambda'$ de la façon suivante

$$\lambda : g \longmapsto < \tilde{\lambda}, P_{\tilde{\Lambda}} g >,$$

où $P_{\tilde{\Lambda}}$ est la projection de g sur $\tilde{\Lambda}$. Pour tout $\varphi \in \tilde{T}^\star\tilde{\Lambda}'$, d'antécédent $\tilde{\lambda}$, on a $\varphi = T^\star\lambda$, où λ est construit selon le procédé ci-dessus. On a donc $\tilde{T}^\star\tilde{\Lambda}' \subset T^\star\Lambda'$. L'inclusion réciproque est immédiate : $v = T^\star\lambda$ implique $v = \tilde{T}^\star\tilde{\lambda}$, où $\tilde{\lambda}$ est la restriction de λ à $\tilde{\Lambda}$. Il s'agit donc de vérifier que l'image de $\tilde{T}^\star$ est fermée. Pour alléger les notations, on suppose que T lui-même est surjectif. La proposition 7.20 nous assure l'existence d'une constante $\alpha > 0$ telle que

$$\|\mu\| \leq \alpha\,\|T^\star\mu\| \quad \forall \mu \in \Lambda'.$$

On considère pour finir une suite de $T^\star(\Lambda)$, donc du type $(T^\star\lambda_n)$, qui converge vers $\varphi \in V'$. D'après l'inégalité ci-dessus, le critère de Cauchy de $(T^\star\lambda_n)$ implique que la suite (λ_n) est elle-même de Cauchy. Donc elle converge dans Λ' vers λ, et l'on a par continuité $T^\star\lambda = \varphi$, donc $\varphi \in T^\star(\Lambda)$.

Proposition 7.20. Soient V et Λ deux espaces de Banach, et $T \in \mathcal{L}(V, \Lambda)$. Les assertions suivantes sont équivalentes.

(i) T est surjectif.

(ii) Il existe $\alpha > 0$ tel que

$$\|\mu\| \leq \alpha\,\|T^\star\mu\| \quad \forall \mu \in \Lambda'.$$

(iii) Il existe $\beta > 0$ tel que

$$\sup_{u \in V} \frac{|\langle \lambda\,, Tu\rangle|}{|u|\,\|\lambda\|} \geq \beta \quad \forall \lambda \in \Lambda'.$$

Démonstration : $(i) \Longrightarrow (ii)$ On suppose que T est surjectif. On note α la constante de la proposition 7.15. Pour tout $z \in \Lambda$, il existe $u \in V$ de norme inférieure ou égale à $\alpha\,\|z\|$ tel que $Tu = z$. Soit maintenant $\mu \in \Lambda'$. On a

$$\langle \mu\,, z\rangle = \langle \mu\,, Tu\rangle = \langle T^\star\mu\,, u\rangle \leq \alpha\,\|z\|\,\|T^\star\mu\|\,.$$

On en déduit l'inégalité

$$\|\mu\| = \sup_{\|z\|\leq 1} \langle \mu \, , z\rangle \leq \alpha \, \|T^\star \mu\| \, .$$

$(ii) \Longrightarrow (i)$ On suppose maintenant qu'il existe $\alpha > 0$ tel que l'inégalité ci-dessus soit vérifiée pour tout $\mu \in \Lambda'$. Montrons dans un premier temps que l'adhérence de $T(B_V)$ contient une boule ouverte centrée en 0. Si ça n'est pas le cas, alors pour tout n il existe z_n de norme inférieure à $1/n$ à l'extérieur de $\overline{T(B_V)}$. Pour tout n, il existe donc d'après le théorème de Hahn-Banach un hyperplan associé à la forme linéaire $\lambda_n \in \Lambda'$ qui sépare strictement z_n du convexe fermé $\overline{T(B_V)}$:

$$\langle \lambda_n \, , z_n\rangle > \sup_{u\in B_V} \langle \lambda_n \, , Tu\rangle = \sup_{u\in B_V} \langle T^\star \lambda_n \, , u\rangle = \|T^\star \lambda_n\| \geq \|\lambda_n\| \, /\alpha.$$

On a donc $\|z_n\| \geq 1/\alpha$, ce qui est absurde. Reprenant le raisonnement effectué dans la seconde partie de la démonstration du théorème de l'application ouverte, on montre que $T(B_V)$ (et non pas seulement son adhérence) contient une boule ouverte. L'image de T contient donc l'espace vectoriel engendré par cette boule ouverte, qui est Λ tout entier.

$(ii) \Longleftrightarrow (iii)$ On a, pour tout $\lambda \in \Lambda'$,

$$\sup_{u\in V} \frac{\|\langle \lambda \, , Tu\rangle\|}{\|u\|} = \sup_{u\in V} \frac{\|\langle T^\star \lambda \, , u\rangle\|}{\|u\|} = \|T^\star \lambda\| \, .$$

On a donc l'équivalence

$$\exists \alpha > 0 \, , \ \|\mu\| \leq \alpha \, \|T^\star \mu\| \quad \forall \mu \in \Lambda' \Longleftrightarrow \inf_{\lambda\in\Lambda'} \sup_{u\in V} \frac{\|\langle \lambda \, , Tu\rangle\|}{\|u\| \, \|\lambda\|} = \beta = 1/\alpha > 0.$$

Chapitre 8

Espaces de Sobolev

Soit N un entier ≥ 1. Dans tout ce chapitre, on désignera par Ω un ouvert de $\mathbb{R}^N$ muni de la mesure de Lebesgue dx.

8.1. Vue d'ensemble

Ce chapitre présente les définitions et propriétés principales des espaces de Sobolev, cadre fonctionnel naturel pour les problèmes étudiés dans ce cours. Cette section décrit ce qu'il est essentiel de connaître, en particulier sur la notion de trace, les inégalités de type Poincaré, et la régularité des solutions à des problèmes elliptiques.

Définitions. L'espace de Sobolev $H^1(\Omega)$ est l'espace des fonctions de L^2 dont le gradient est dans $L^2(\Omega)^N$. On peut en donner une définition rigoureuse dans l'esprit des distributions (voir définition 8.10)) ou, pour $\Omega = \mathbb{R}^N$, à l'aide de la transformée de Fourier (voir théorème 8.72). L'espace $H^2(\Omega)$ (et, de la même manière, les espaces H^3, etc...) est défini comme l'espace des fonctions possédant un gradient dans L^2 dont chaque composante est dans H^1.

On définit le sous-espace H^1_0 des fonctions nulles au bord en considérant l'adhérence des fonctions régulières à support compact dans Ω (i.e. fonctions nulles sur un voisinage de la frontière).

Traces. Pour tout domaine dont la frontière $\Gamma = \partial\Omega$ est raisonnablement régulière, on peut définir une application

$$\gamma_0 \ : \ u \in H^1(\Omega) \longmapsto \gamma_0(u) \in L^2(\Gamma).$$

La définition de cette application est délicate, car les fonctions de H^1 sont définies en fait comme des classes de fonctions (dont la valeur prise sur un ensemble de mesure nulle n'a a priori pas de signification). Le bord étant de mesure nulle, la notion de restriction n'a en particulier aucun sens a priori. Elle a en revanche un sens pour les fonctions régulières, et cette application est définie justement par densité des fonctions régulières. Cette construction repose donc sur deux arguments :

(i) densité de l'espace des fonctions régulières $\mathcal{D}(\overline{\Omega})$ (restriction à Ω des fonctions régulières sur $\mathbb{R}^N$) ;

(ii) continuité de l'application restriction pour les normes indiquées ci dessus.

Le point (i) est vérifié pour des domaines dont le bord est Lipschitz (on montre que l'on peut écrire toute fonction de $H^1(\Omega)$ comme restriction à Ω d'une fonction de $H^1(\mathbb{R}^N)$, et

l'on régularise par convolution), et le second point peut se démontrer par exemple à l'aide de la transformée de Fourier, ou de façon directe (voir proposition 8.30), dans le cas d'un demi-espace. Le cas général est obtenu par utilisation de cartes locales.

L'application trace définie ci-dessus n'est **pas surjective**. L'image de γ_0 est un espace strictement inclus dans $L^2(\Gamma)$, noté $H^{1/2}(\Gamma)$. Une conséquence très importante est que γ_0, vue comme application à image dans $L^2(\Gamma)$, n'est pas à image fermée [1].

On définit de la même manière, pour toute fonction de $H^2(\Omega)$, la trace de sa dérivée normale

$$\gamma_1 \ : \ u \in H^2(\Omega) \longmapsto \gamma_1(u) \in L^2(\Gamma),$$

qui s'identifie à $\partial u/\partial n$ pour des fonctions régulières, où $\partial u/\partial n$ est la dérivée dans la direction de la normale sortante.

On considère une partition de Ω en sous-domaines Ω_i qui ne se recouvrent pas, et une fonction $u \in L^2(\Omega)$ dont la restriction à chaque Ω_i est dans H^1. Si les traces sur les interfaces coïncident, alors u est dans $H^1(\Omega)$ (proposition 8.39, page 99).

Si maintenant u est H^2 sur chaque sous-domaine, si les traces coïncident, et si les dérivées normales coïncident, alors la fonction u est dans H^2. Ces conditions sont nécessaires : en particulier une fonction constante par morceaux n'est dans H^1 globalement que si les constantes sont les mêmes [2].

Pour u dans H^2 et v dans H^1, on a la formule de Green (proposition 8.38)

$$-\int_\Omega v\triangle u = \int_\Omega \nabla u\cdot\nabla v - \int_\Gamma \frac{\partial u}{\partial n} v.$$

Injections. Sur un domaine borné, une application continûment différentiable est dans $H^1(\Omega)$, et son gradient au sens classique s'identifie au gradient au sens de Sobolev. Inversement, les fonctions des espaces de Sobolev peuvent vérifier des propriétés de régularité au sens classique. La dimension d'espace intervient de façon essentielle dans ces propriétés [3]. En particulier, les fonctions de H^1 sont continues en dimension 1 (elles sont même 1/2 Höldériennes), mais ne le sont pas nécessairement en dimension supérieure.

1. Ce fait a des conséquences importantes sur les formulations par dualité de problèmes sous contraintes au bord.

2. On peut donner une formulation énergétique de cette non appartenance. Dans le cas d'un écoulement régi par la loi de Darcy par exemple, pour un domaine de perméabilité uniforme, un champ de pression qui subit un saut à travers une interface n'est pas dans H^1. Cela signifie simplement que la puissance dissipée associée à l'écoulement induit par un tel champ de pression serait infinie. De la même manière en élasticité, la norme H^1 fait intervenir la norme L^2 du gradient du déplacement, cela signifie que l'on ne peut avoir de saut du déplacement à travers une interface. Noter que si cela n'était pas le cas, il serait nécessaire de revoir les modèles d'élasticité linéaire, puisque cela signifierait que l'on peut "casser" un matériau en deux avec une énergie finie.

3. Les espaces de Sobolev sont basés sur une caractérisation de la régularité en moyenne quadratique. Prenons l'exemple de singularités radiales au voisinage d'un point susceptibles d'apparaître pour de telles fonctions : le critère de régularité fait intervenir l'élément d'intégration en r^{N-1}. Plus N est grand, plus le comportement local est susceptible d'être singulier pour une même régularité au sens de Sobolev. Ainsi en dimension 1 les fonctions de H^1 sont continues, en dimension 2 la fonction radiale $\varphi(r) = |\ln r|^\alpha$ est dans H^1 pour $\alpha < 1/2$, et en dimension $N \geq 3$ $\varphi(r) = 1/|r|^\alpha$ est dans H^1 dès que $\alpha < N/2 - 1$.

On a en revanche continuité des fonctions de H^2 pour les dimensions physiques $N = 2$ et 3 (voir théorème 8.42, page 100), ce qui permet de parler de valeur ponctuelle d'une fonction dès que l'on a régularité H^2 au voisinage du point considéré.

Les propriétés de compacité des injections entre espaces de Sobolev jouent un rôle très important, pour la démonstration des inégalités de type Poincaré évoquées ci-après, et pour le théorème de stabilité sur lequel se fonde toute l'analyse d'erreur pour les méthodes d'éléments finis. On a en particulier, dans le cas d'un domaine borné régulier, injection compacte de $H^1(\Omega)$ dans $L^2(\Omega)$ (et de la même manière de $H^2(\Omega)$ dans $H^1(\Omega)$ (voir théorème de Rellich 8.43, page 100).

Inégalités de type Poincaré. Les théorèmes d'existence d'une solution aux problèmes abordés ici sont basés sur l'utilisation du théorème de Lax-Milgram, qui repose essentiellement sur la coercivité de la forme bilinéaire intervenant dans la formulation variationnelle, coercivité exprimée vis à vis de la norme H^1 :

$$a(u,u) \geq \alpha \left(\int_\Omega |u|^2 + \int_\Omega |\nabla u|^2 \right).$$

Or dans la plupart des situations physiques rencontrées, la forme bilinéaire ne fait intervenir que le gradient de la fonction [4]. Il est alors nécessaire de disposer d'un contrôle de la norme L^2 par la semi-norme H^1, du type :

$$|u|_0 = \left(\int_\Omega |u|^2 \right)^{1/2} \leq C\,|u|_1 = C \left(\int_\Omega |\nabla u|^2 \right)^{1/2}.$$

Cette inégalité s'appelle l'inégalité de Poincaré, et elle est notamment vérifiée pour toutes les fonctions de H^1_0 dans le cas d'un domaine borné (ou simplement borné dans une direction, voir proposition 8.48). Sur $H^1(\Omega)$ cette inégalité est évidemment fausse, puisqu'elle est invalidée par les fonctions constantes, mais on peut montrer qu'elle est vérifiée dès que le problème des fonctions constantes est géré d'une manière ou d'une autre, en particulier lorsque l'on travaille avec des fonctions nulles sur une partie du bord. Plus généralement, on aura recours à une inégalité de type Poincaré qui permet de minorer la forme quadratique par la norme L^2, dès que la forme quadratique s'écrit comme somme du terme de semi norme H^1 et d'un terme qui « voit les constantes ». La forme générale de cette propriété est donnée par la proposition 8.51, page 103 :

$$|u|_0 \leq C\,(|Tu|_M + |u|_1) \quad \forall u \in H^1(\Omega),$$

où T est un opérateur linéaire non nul sur la fonction identiquement égale à 1.

Régularité des solutions de problèmes elliptiques. On s'intéresse aux solutions de problèmes du type

$$\alpha u - \nabla \cdot k \nabla u = f \quad \text{dans } \Omega$$

avec conditions au bord du domaine Ω, dans le cas où $k = k(x)$ est un champ scalaire borné et minoré par une constante positives, et $f \in L^2$. On supposera k régulier (C^1) sur Ω. Dès que l'on peut vérifier la coercivité de la forme bilinéaire associée, le théorème de Lax-Milgram assure l'existence d'une solution dans $H^1(\Omega)$, qui correspond à la régularité « native » de la solution.

4. Ainsi pour l'équation de Darcy, l'énergie dissipée s'écrit $\int k\,|\nabla \mathrm{p}|^2$.

Mais pour un second membre dans L^2, du fait que l'opérateur Laplacien est d'ordre 2, on peut espérer avoir une meilleure régularité [5] (en particulier H^2).

Cette propriété est triviale en dimension 1, puisque le Laplacien est l'opérateur de dérivée seconde. C'est moins évident en dimension supérieure ou égale à 2, puisque le Laplacien ne fait intervenir qu'une seule combinaison linéaire des coefficients de la matrice Hessienne ($N(N+1)/2$ coefficients distincts, qui doivent tous appartenir à L^2). Cette propriété est pourtant vraie en général pour k continûment différentiable (voir proposition 8.61 dans le cas d'un k uniforme) si l'on exclut le voisinage de la frontière. Le problème de la régularité jusqu'au bord est très délicat. Pour des conditions d'un même type (tout Dirichlet ou tout Neuman sur l'ensemble de la frontière), on a régularité H^2 dans le cas d'un domaine de classe C^2, et dans le cas d'un domaine polygonal convexe. Le panachage de conditions aux limites est lui-même assez problématique, et invalide en général la régularité H^2, même pour des domaines réguliers. On retiendra simplement ici que l'on conserve cette régularité dans le cas d'un passage de conditions de Dirichlet à conditions de Neuman (constante pour Dirichlet, homogènes pour Neumann) au travers d'une jonction qui se fait à angle droit (comme dans la situation très courante en pratique du problème (1.4), page 17).

Il est important [6] d'identifier les situations pour lesquelles on n'a pas régularité :

(*i*) Le coefficient k de conductivité (ou les coefficients de Lamé pour l'élasticité) est constant par morceaux. On a alors régularité H^2 sur les sous-domaines correspondant aux divers matériaux, mais le fait que l'on n'ait pas raccord des dérivées normales à travers les interfaces entre matériaux exclut l'appartenance à H^2 de la solution [7].

(*ii*) Le domaine présente des singularités. Noter que si le domaine est convexe, la solution reste H^2 pour des singularités Lipschitziennes. Pour des coins rentrants en revanche, la solution n'est pas H^2. On se reportera à l'exercice **??**, page **??**, pour un exemple de fonction harmonique non régulière au voisinage d'un coin rentrant [8].

5. Il est souvent essentiel de montrer que la régularité est supérieure, pour au moins deux raisons. D'une part, si l'on cherche à montrer que la solution de la formulation variationnelle vérifie l'équation de départ avec conditions aux limites, dans le cas où des conditions aux limites de type Neuman interviennent , la définition de la dérivée normale comme fonction au bord nécéssite que l'on ait une régularité H^2 sur la solution. En second lieu, l'ordre de convergence des méthodes d'éléments finis dépend de la régularité de la solution. En particuliers les méthodes basées sur des fonctions affines par triangles ne sont d'ordre 1 vis-à-vis du paramètre de discrétisation que pour des fonctions de H^2.

6. Notamment pour des motivations numériques : on pourra par exemple être amené à raffiner au voisinage d'un point ou d'une zone en laquelle on s'attend à avoir une solution moins régulère. De la même manière, dans le cas d'un domaine à propriétés lisse par morceaux, il peut être avantageux d'utiliser un maillage qui respecte la décomposition donnée par le modèle.

7. La solution est en fait dans $H^{3/2-\varepsilon}$ pour tout $\varepsilon > 0$.

8. Un cas extrême de cette situation se rencontre dans le cas de fissures au sein d'un matériau : le domaine est alors le domaine initial moins la fissure (noter que ce cas ne rentre pas dans le cadre de la définition 8.26, page 97, car le domaine n'est plus d'un seul côté de sa frontière). On a donc un coin rentrant d'angle 2π, ce qui favorise l'apparition d'une singularité au voisinage de l'extrémité de la fissure, de nature (par apparition de très fortes contraintes susceptible d'endommager localement le matériau) à propager cette fissure.

(*iii*) Passage d'une condition aux limites de Dirichlet (valeur/déplacement imposée) à une condition de Neuman (flux/force imposé). Le phénomène est particulièrement net lorsque la transition se fait au niveau d'un point non singulier de la frontière [9].

8.2. Rappels sur l'espace $L^2(\Omega)$

On désigne par Ω un ouvert de $\mathbb{R}^N$ muni de la mesure de Lebesgue dx.

Définition 8.1. On définit l'espace $L^2(\Omega)$ comme

$$L^2(\Omega) = \left\{ f : \Omega \to \mathbb{R}\,,\ f \quad \text{mesurable}, \quad \int_\Omega |f(x)|^2\, dx < +\infty \right\}.$$

On le munit de la norme $\|f\|_2 = \left(\int_\Omega |f|^2\right)^{1/2}$. On notera $L^2(\Omega)^N$ l'espace des champs de vecteurs dont chaque composante appartient à $L^2(\Omega)$.

Proposition 8.2. L'espace $L^2(\Omega)$ est un espace de Hilbert pour le produit scalaire

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

comme pour tout produit du type

$$(u, v)_k = \int_\Omega k(x)u(x)v(x)\, dx,$$

où k est une fonction mesurable telle que $0 < m \leq k(x) \leq M$ presque partout.

Démonstration : Le fait que cette forme bilinéaire soit bien définie sur $L^2 \times L^2$ est conséquence directe de l'inégalité de Cauchy-Schwarz. Il s'agit alors de montrer que L^2 est bien complet pour la norme associée. Pour cela on considère une suite de Cauchy, on montre par un argument de convergence monotone que la suite converge presque partout vers une limite, que la limite appartient bien à L^2, et que l'on a bien convergence pour la norme L^2 vers cette limite. On trouvera une démonstration détaillée dans [**3**], page 57.

Définition 8.3. (Suite régularisante)
On appelle suite régularisante une suite (ρ_n) de fontions C^∞ de $\mathbb{R}^N$ dans $\mathbb{R}$ telle que, pour tout $n \in \mathbb{N}$,

$$\mathrm{supp}(\rho_n) \subset B(0, 1/n)\,,\ \int_{\mathbb{R}} \rho_n = 1\,,\ \rho_n(x) \geq 0 \quad \forall x \in \mathbb{R}.$$

Proposition 8.4. Soit $f \in L^2(\mathbb{R}^N)$. On définit la fonction $\rho_n \star f$ par

$$(\rho_n \star f)(x) = \int_{\mathbb{R}^N} \rho_n(x - y)f(y)\, dy.$$

Alors la fonction $\rho_n \star f$ est dans $C^\infty(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$. On a

$$\rho_n \star f \longrightarrow f \quad \text{dans} \quad L^2(\mathbb{R}^N).$$

9. Noter le caractère très banal de cette situation dans la réalité physique : elle correspond par exemple à un matériau parallélépipédique fixé à un mur vertical sur l'un de ses côté, dans le cas où l'adhésion n'est pas totale (par exemple si le matériau a été collé et que la colle n'a pas été répartie sur l'ensemble de la zone à fixer).

Remarque 8.5. Toute fonction f de $L^2(\Omega)$ peut être prolongée par 0 à $\mathbb{R}^N$ tout entier. On peut donc appliquer ce qui précède. Les propriétés de convergence énoncées ci-dessus s'appliquent ainsi à la restriction de $\rho_n \star f$ à Ω.

Définition 8.6. On note $\mathcal{D}(\Omega)$ l'espace des fonctions $\mathcal{C}^\infty$ à support compact dans Ω. On vérifie que cet espace est non vide en considérant une boule ouverte $B(a,r)$ dont l'adhérence est dans Ω, et la fonction

$$\varphi(x) = \exp\left(\frac{1}{|x-a|^2 - r^2}\right) \text{ si } x \in B(a,r)\,,\ \varphi(x) = 0 \text{ si } x \notin B(a,r).$$

Proposition 8.7. L'espace $\mathcal{D}(\Omega)$ est dense dans $L^2(\Omega)$.

Remarque 8.8. L'appartenance à L^2 n'exige aucune régularité en espace (aucune « corrélation spatiale » n'est exigée). En particulier, si l'on considère une partition de Ω sous la forme $\overline{\Omega} = \overline{\Omega}_1 \cup \overline{\Omega}_2$, $\Omega_1 \cap \Omega_2 = \emptyset$, où les Ω_i sont des ouverts tels que $\partial\Omega_1 \cap \partial\Omega_2$ est de mesure nulle, pour toutes fonctions $f_i \in L^2(\Omega_i)$, la fonction f dont la restriction à Ω_i est f_i est dans $L^2(\Omega)$. Nous verrons qu'une telle construction par morceaux d'une fonction est en général impossible pour les espaces de Sobolev.

8.3. Définitions, propriétés générales

Définition 8.9. (Gradient)
Soit φ une fonction C^1 de Ω dans $\mathbb{R}$. On appelle gradient de φ la fonction de Ω dans $\mathbb{R}^N$ définie par

$$\nabla\varphi = \begin{pmatrix} \dfrac{\partial\varphi}{\partial x_1} \\ \vdots \\ \dfrac{\partial\varphi}{\partial x_N} \end{pmatrix}.$$

Définition 8.10. On définit l'espace de Sobolev $H^1(\Omega)$ comme l'ensemble des fonctions u dans $L^2(\Omega)$ telles qu'il existe $\mathbf{v} = (v_1, \ldots, v_N) \in (L^2(\Omega))^N$ vérifiant

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

On notera alors $\mathbf{v} = \nabla u$.

La fonction ∇u de $\mathbb{R}$ dans $\mathbb{R}^N$ est ainsi définie comme l'unique fonction vectorielle à composantes dans $L^2(\Omega)$ telle que l'identité entre vecteurs de $\mathbb{R}^N$

$$\int_\Omega u\nabla\varphi = -\int_\Omega \varphi\nabla u$$

soit vérifiée pour tout $\varphi \in \mathcal{D}(\Omega)$.

On notera $H^1(\Omega)^N$ l'espace des champs de vecteurs dont chaque composante appartient à $H^1(\Omega)$. Le gradient $\nabla\mathbf{u}$ est alors une matrice dont la ligne i est le gradient de la i-ème composante de $\mathbf{u}$.

Proposition 8.11. L'espace $H^1(\Omega)$ muni de la norme $\|\cdot\|$ définie par

$$\|v\|^2 = \int_\Omega u^2 + \int_\Omega |\nabla u|^2$$

est un espace de Hilbert séparable [10]. .

DÉMONSTRATION : On construit pour cela une isométrie entre $H^1(\Omega)$ et un sous-espace fermé de $L^2(\Omega) \times L^2(\Omega)^N$. Voir [**3**, Prop. IX.1]. □

NOTATION : On désignera par $|u|_{0,\Omega}$ la norme L^2 de u sur Ω (nous omettrons Ω quand il n'y a pas d'ambigüité), et par $|u|_{1,\Omega}$ la semi norme H^1 :

$$|u|^2_{1,\Omega} = \int_\Omega |\nabla u|^2 ,$$

de telle sorte que

$$\|u\|^2_{H^1} = |u|^2_{0,\Omega} + |u|^2_{1,\Omega} .$$

Proposition 8.12. Si $u \in C^1(\Omega) \cap L^2(\Omega)$ et $\nabla u \in (L^2(\Omega))^N$, alors $u \in H^1(\Omega)$, et le gradient de u au sens classique (définition 8.9) s'identifie au gradient au sens de Sobolev (définition 8.10).

Remarque 8.13. Nous verrons que les fonctions de $H^1(\Omega)$ ne sont pas nécessairement continues en dimension supérieure ou égale à deux (voir exercice **??**, page **??**), mais elles possèdent une certaine régularité en espace. La régularité H^1 exclut notamment les discontinuités franches au travers d'une hypersurface [11] (voir exercice **??**, page **??**).

Proposition 8.14. Soit $u \in H^1(\Omega)$ telle que $\nabla u = 0$ presque partout sur Ω. Alors u est constante sur chaque composante connexe de Ω.

DÉMONSTRATION : Voir Allaire [**1**, Prop. 4.2.5]. □

En dimension 1, une fonction peut s'écrire comme intégrale de sa dérivée, comme le précise la proposition suivante.

Proposition 8.15. Soit I un intervalle de $\mathbb{R}$. Toute fonction $u \in H^1(I)$ admet un représentant continu $\tilde{u}$, qui vérifie

$$\tilde{u}(x) = u(x) \qquad \text{p.p. sur I} \quad , \ \tilde{u}(y) - \tilde{u}(x) = \int_x^y u'(t)\, dt.$$

Cette fonction continue sur I est prolongeable par continuité aux extrémités de I.

DÉMONSTRATION : Voir Brezis [**3**, Th. VIII.2]. □

10. Il contient un sous-ensemble dénombrable et dense

11. Ce fait est fondamental en mécanique des milieux continus. Si l'on identifie **u** à un champ de déplacements élémentaires, cela implique qu'un champ de déplacement qui tendrait à décomposer un matériau en deux sous-parties disjointes est exclu du point de vue énergétique.

Proposition 8.16. Soit u une fonction de $L^2(\Omega)$. Les assertions suivantes sont équivalentes :

(i) $u \in H^1(\Omega)$.

(ii) Il existe une constante C telle que

$$\left| \int_\Omega u \nabla \varphi \right| \leq C \|\varphi\|_{L^2} \quad \forall \varphi \in \mathcal{D}(\Omega).$$

(iii) Il existe une constante C telle que, pour tout $\omega \subset\subset \Omega$, pour tout h tel que $|h| < \operatorname{dist}(\omega, \Omega^c)$,

$$\|\tau_h u - u\|_{L^2(\omega)} \leq C |h|.$$

Démonstration : $(i) \Longrightarrow (ii)$ est une conséquence immédiate de la définition.

$(ii) \Longrightarrow (i)$ Pour i entre 1 et N, on considère la forme linéaire définie sur $C_c^\infty \subset L^2(\Omega)$

$$\varphi \longmapsto \int_\Omega v \partial_i \varphi.$$

Cette forme linéaire est continue pour la norme L^2 par hypothèse. Elle se prolonge donc par densité de $C_c^\infty(\Omega)$ en une forme linéaire continue sur $L^2(\Omega)$. Le théorème de représentation de Riesz-Fréchet assure donc l'existence de $w_i \in L^2(\Omega)$ tel que

$$\int_\Omega v \partial_i \varphi = -\int_\Omega w_i \varphi,$$

d'où $u \in H^1$ avec $\nabla u = (w_1, \ldots, w_N)$. □

$(i) \Longrightarrow (iii)$ Soit $\omega \subset\subset \Omega$, et $h < \operatorname{dist}(\omega, \Omega^c)$. On considère dans un premier temps une fonction u régulière ($u \in \mathcal{D}(\Omega)$). On a

$$u(x+h) = u(x) + \int_0^1 \nabla u(x+th) \cdot h \, dt,$$

d'où

$$|u(x+h) - u(x)|^2 \leq |h|^2 \int_0^1 |\nabla u(x+th)|^2,$$

et donc

$$\int_\omega |\tau_h u - u(x)|^2 \leq |h|^2 \int_\omega \int_0^1 |\nabla u(x+th)|^2 \leq |h|^2 \int_0^1 \int_\omega |\nabla u(x+th)|^2.$$

On choisit maintenant ω' fortement inclus dans Ω, qui contient tous les translatés de ω par th, pour $t \in [0,1]$. On a

$$\|\tau_h u - u\|_{L^2} \leq |h| \int_{\omega'} |\nabla u|^2.$$

On conclut en utilisant la propriété de densité 8.18.

$(iii) \Longrightarrow (ii)$ Soit $\varphi \in C_c^\infty(\Omega)$, et $\omega \subset\subset \Omega$ qui contient le support de φ. Pour tout h tel que $h < \operatorname{dist}(\omega, \Omega^c)$, on a

$$\left| \int_\omega (\tau_h u - u)\varphi \right| \leq C \|\varphi\|_{L^2(\omega)} |h| \leq C \|\varphi\|_{L^2(\Omega)} |h|.$$

D'autre part,

$$\int_\omega (u(x+h) - u(x))\varphi(x) = \int_\Omega (u(x+h) - u(x))\varphi(x) = \int_\Omega u(y)(\varphi(y-h) - \varphi(y)).$$

La majoration (*iii*) implique donc

$$\left| \int_\Omega u(y) \frac{\varphi(y-h) - \varphi(y)}{|h|} \right| \leq C \, \|\varphi\|_{L^2} \, .$$

On conclut en prenant h de la forme $t\mathbf{e}_i$ et en faisant tendre t vers 0. □

Proposition 8.17. L'espace $\mathcal{D}(\mathbb{R}^N)$ est dense dans $H^1(\mathbb{R}^N)$.

NOTATION : On dit que ω est fortement inclus dans Ω si $\overline{\omega}$ est compact et inclus dans Ω. On note $\omega \subset\subset \Omega$.

Proposition 8.18. Pour tout $\omega \subset\subset \Omega$, tout $u \in H^1(\Omega)$, il existe une suite (u_n) dans $\mathcal{D}(\Omega)$ telle que

$$u_n \longrightarrow u \text{ dans } L^2(\Omega) \, , \ \nabla u_n \longrightarrow \nabla u \text{ dans } L^2(\omega)^N .$$

Corollaire 8.19. Soit (ω_n) une suite de domaines fortements inclus dans Ω, et $u \in H^1(\Omega)$. Il existe une suite (u_n) dans $\mathcal{D}(\Omega)$ telle que

$$\|u_n - u\|_{L^2(\Omega)} \longrightarrow 0 \, , \ \|\nabla u_n - \nabla u\|_{L^2(\omega_n)^N} \longrightarrow 0.$$

Définition 8.20. On définit $H_0^1(\Omega)$ comme l'adhérence de $\mathcal{D}(\Omega)$ dans $H^1(\Omega)$.

Noter que, d'après la proposition 8.18, on a $H_0^1(\mathbb{R}^N) = H^1(\mathbb{R}^N)$

Par rapport à H_0^1, l'espace H^1 peut se décrire comme l'ensemble des fonctions L^2 de gradient L^2 qui peuvent « prendre des valeurs non nulles sur le bord » . Cette expression ne pourra se voir donner un cadre mathématique précis qu'après que l'on aura défini la notion de régularité du bord (voir, section 8.4, la définition de l'opérateur trace sur le bord γ_0). On peut néammoins dès maintenant donner un sens abstrait à la notion de valeur au bord, sans faire aucune hypothèse sur la géométrie de Ω. Par analogie avec l'espace des traces des fonctions de H^1 dans le cas d'un bord régulier (voir définition 8.32), nous noterons $\tilde{H}^{1/2}$ l'espace abstrait correspondant.

Définition 8.21. On définit l'espace $H^2(\Omega)$ comme l'ensemble des fonctions de $H^1(\Omega)$ dont toutes les dérivées partielles par rapport à l'une des composantes sont elles-mêmes dans $H^1(\Omega)$. C'est un espace de Hilbert muni de la norme

$$\|u\|^2_{H^2(\Omega)} = |u|_0^2 + \sum_i \left| \frac{\partial u}{\partial x_i} \right|_0^2 + \sum_{i,j} \left| \frac{\partial^2 u}{\partial x_i \partial x_j} \right|_0^2 = |u|^2_{0,\Omega} + |u|^2_{1,\Omega} + |u|^2_{2,\Omega} \, .$$

On peut définir de façon analogue les espaces $H^m(\Omega)$ pour $m = 3, 4, \ldots$, mais nous n'utiliserons ici que $m \leq 2$.

Définition 8.22. (Espace H^m_{loc})
Soit m un entier positif (on utilisera le cas $m = 2$ dans la suite). On définit l'espace $H^m_{\text{loc}}(\Omega)$ comme l'espace vectoriel des (classes de) fonctions de Ω dans $\mathbb{R}$ dont la restriction à ω est dans $H^m(\omega)$, pour tout ω fortement inclus dans Ω. De façon équivalente, c'est l'ensemble des fonctions u de Ω dans $\mathbb{R}$ telles que θu est dans $H^m(\Omega)$ pour tout θ dans $\mathcal{D}(\Omega)$.

Noter que l'appartenance d'une fonction à H^m_{loc} permet de parler de ses dérivées m-ièmes comme de fonctions (mesurables) définies sur Ω. On donne ainsi un sens à des expressions du type $\partial^m u / \partial x_i^m = g$ presque partout dans Ω, où g est une fonction de L^2_{loc}.

8.4. Traces

En élasticité, le problème le plus couramment rencontré consiste à trouver le champ de déplacement d'un solide déformable soumis à certaines sollicitations sur son bord (déplacement imposé). Ces sollicitations au bord ne peuvent avoir un sens que si l'on est capable de parler d'un champ de déplacement sur le bord du domaine. Lorsque l'on considère des fonctions régulières (au moins continues sur $\overline{\Omega}$), on peut parler simplement de la restriction de la fonction à $\partial\Omega$. Dans le contexte présent, nous avons vu que les fonctions de $H^1(\Omega)$ ne sont pas nécessairement continues, et ne sont définies a priori que comme des classes de fonctions (à un ensemble de mesure nulle près). La frontière d'un ouvert régulier étant de mesure nulle, la notion de restriction n'a pas de sens. Nous allons montrer ici qu'il est possible de donner un sens précis à cette notion de trace, dès que les fonctions que l'on considère ont une régularité suffisante en espace.

Définition 8.23. (Espace des traces abstrait)
On définit l'espace $\tilde{H}^{1/2}$ comme l'espace quotient $H^1(\Omega)/H_0^1(\Omega)$. C'est un espace vectoriel normé pour la norme quotient

$$\|\tilde{u}\|_{H^1/H_0^1} = \inf_{v\in\tilde{u}} \|v\|_{H^1} = \inf_{h\in H_0^1} \|u-h\|_{H^1}\,.$$

Noter que, d'après la définition de H_0^1, on a aussi $\|\tilde{u}\|_{H^1/H_0^1} = \inf_{h\in\mathcal{D}(\Omega)} \|u-h\|_{H^1}$.

Remarque 8.24. On a $H_0^1(\mathbb{R}^N) = H^1(\mathbb{R}^N)$ (d'après la proposition 8.17), et l'on peut avoir $H_0^1(\Omega) = H^1(\Omega)$ même si Ω est strictement inclus dans $\mathbb{R}^N$ (de telle sorte que $\mathcal{D}(\Omega)$ soit strictement inclus dans $\mathcal{D}(\mathbb{R}^N)$). L'espace quotient défini précédemment est alors l'espace trivial $\{0\}$. C'est le cas par exemple de $\mathbb{R}^2$ privé d'un point, ou de $\mathbb{R}^3$ privé d'un point ou d'une droite (voir l'exercice **??** sur la notion de *capacité*).

Proposition 8.25. Soit $u \in H_0^1(\Omega)$. On définit $\tilde{u}$ comme la fonction qui vaut $u(x)$ pour tout $x \in \Omega$, et qui prend la valeur 0 à l'extérieur de Ω. Alors $\tilde{u} \in H^1(\mathbb{R}^N)$.

Démonstration : Tout d'abord remarquons que $\tilde{u}$ est dans $L^2(\mathbb{R}^N)$. Par définition de H_0^1, u est limite d'une suite (u_n) de fonctions C^∞ à support compact dans Ω. Pour tout $\varphi \in \mathcal{D}(\mathbb{R}^N)$, on a

$$\begin{aligned}\int_{\mathbb{R}^N} \tilde{u}\nabla\varphi &= \int_\Omega u\nabla\varphi = \lim_{n\to+\infty}\int_\Omega u_n\nabla\varphi \\ &= -\lim_{n\to+\infty}\int_\Omega \varphi\nabla u_n = -\int_\Omega \varphi\nabla u = -\int_{\mathbb{R}^N}\varphi\mathbf{v}.\end{aligned}$$

où $\mathbf{v}$ est le champ de vecteurs qui vaut ∇u dans Ω, et 0 à l'extérieur de Ω. □

Dans cette section nous précisons les propriétés qui vont nous permettre de définir des valeurs au bord pour des fonctions appartenant aux espaces de Sobolev introduits précédemment. On se reportera à [**13**], [**3**], ou [**11**] pour les démonstrations détaillées.

On définit le cylindre $Q_{\rho h}$ de $\mathbb{R}^N$ par

$$Q_{\rho h} = \left\{x\in\mathbb{R}^N\,,\ x=(x',x_N)=(x_1,\dots,x_N)\,,\ \left|x'\right|<\rho\,,\ -h<x_N<h\right\}.$$

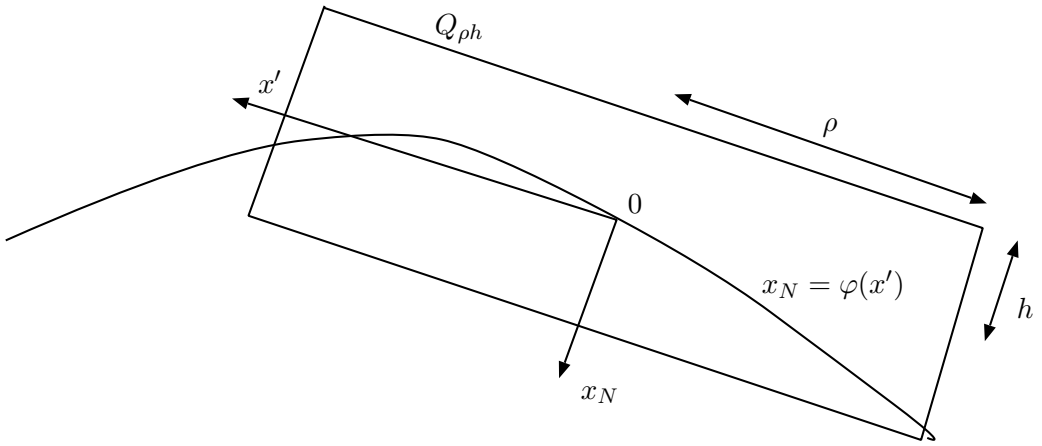


FIGURE 1. Régularité de la frontière

Dans la définition qui suit, "X" représente une régularité fonctionnelle du type C^o, Lipschitz, C^k, etc...

Définition 8.26. Soit Ω un ouvert de $\mathbb{R}^N$. On dit que la frontière de Ω est de classe X si en tout point $a \in \partial\Omega$, il existe un système de coordonnées et ρ, $h > 0$, tels qu'il existe une application

$$\varphi \,:\, \left\{x' \in \mathbb{R}^{N-1}\,,\ \left|x'\right| < \rho\right\} \longrightarrow \mathbb{R}$$

de classe X telle que

(i) $\forall x'$, $|x'| < \rho \Rightarrow |\varphi(x')| < h$,

(ii) $\varphi(0) = 0$,

(iii) $Q_{\rho h} \cap \partial\Omega$ coïncide avec le graphe de φ,

(iv) $U \cap \Omega = \{(x', x_N)\,,\ |x'| \leq \rho\,,\ \varphi(x') < x_N < h\}$.

Définition 8.27. (vecteur normal)
Soit Ω un ouvert de classe $\mathcal{C}^1$, a un point de $\Gamma = \partial\Omega$. On note φ l'application définie ci-dessus. On appelle vecteur normal à Γ au point a le vecteur

$$\mathbf{n} = \frac{(\nabla\varphi, -1)}{|(\nabla\varphi, -1)|}.$$

Noter que l'on peut définir presque partout un tel vecteur sur une frontière supposée seulement Lipschitzienne.

On note $\mathcal{D}(\overline{\Omega})$ l'ensemble des restrictions des fonctions de $\mathcal{D}(\mathbb{R}^N)$ à $\overline{\Omega}$.

Proposition 8.28. Soit Ω un ouvert de frontière Γ Lipschitzienne et bornée. Il existe un opérateur de prolongement

$$P \,:\, H^1(\Omega) \longrightarrow H^1(\mathbb{R}^N),$$

linéaire continu, tel que, pour tout $u \in H^1(\Omega)$, la restriction de Pu à Ω s'identifie à u.

DÉMONSTRATION : Voir Brezis [**3**, Th. IX.7] dans le cas d'un ouvert C^1. L'ingrédient principal de la démonstration est le prolongement par réflexion dont nous indiquons ici le principe dans le cas $N = 1$. On considère $u \in H^1(]0,1[)$, et l'on construit $\tilde{u}$ comme la fonction qui s'identifie à u sur $]0,1[$, et telle que $\tilde{u}(x) = u(-x)$ sur $]-1,0[$. La fonction $\tilde{u}$ est dans $L^2(]-1,1[)$, et sa dérivée $\tilde{u}'$ est définie presque partout sur $]-1,1[$ (avec $\tilde{u}'(-x) = -u'(x)$ pour $x > 0$. Nous allons montrer que cette fonction $\tilde{u}'$ est bien la dérivée de u au sens de Sobolev sur $]-1,1[$. Pour toute fonction-test $\varphi \in \mathcal{D}(]-1,1[)$, si l'on note $\tilde{\varphi}(x) = \varphi(-x)$, on a

$$\int_{-1}^{1} u\varphi' = \int_{-1}^{0} u\varphi' + \int_{0}^{1} u\varphi' = -\int_{0}^{1} u\tilde{\varphi}' + \int_{0}^{1} u\varphi' = \int_{0}^{1} u(\varphi - \tilde{\varphi})'.$$

Notons $\psi = \varphi - \tilde{\varphi}$. On ne peut pas utiliser l'appartenance de u à $H^1(]0,1[)$ car ψ n'est pas à support compact dans $]0,1[$. On se ramène à une fonction à support compact en introduisant, pour $\varepsilon > 0$, la fonction $x \longmapsto \eta_\varepsilon(x) = \eta(x/\varepsilon)$, où η est une fonction C^∞ sur $\mathbb{R}^+$, nulle sur $[0,1/2]$ et sur $[1,+\infty[$. La fonction $\psi_\varepsilon = \eta_\varepsilon\psi$ est dans $\mathcal{D}(]0,1[)$. On a d'une part

$$\int_{0}^{1} u\psi_\varepsilon' = -\int_{0}^{1} \psi_\varepsilon u' \longrightarrow -\int_{0}^{1} \psi u' = -\int_{-1}^{1} \varphi\tilde{u}',$$

et d'autre part

$$\int_{0}^{1} u\psi_\varepsilon' = \int_{0}^{1} \eta_\varepsilon\psi' u + \int_{0}^{1} \eta_\varepsilon'\psi u.$$

Le second terme se majore (en utilisant $\psi(x) = \mathcal{O}(x)$ et $|\eta_\varepsilon'| \leq C/\varepsilon$),

$$\left|\int_{0}^{1} \eta_\varepsilon'\psi u\right| = \left|\int_{0}^{\varepsilon} \eta_\varepsilon'\psi u\right| \leq C\varepsilon\frac{1}{\varepsilon}\int_{0}^{\varepsilon} |u| \leq C\sqrt{\varepsilon}.$$

d'où $\int_0^1 u\psi_\varepsilon' \longrightarrow \int_0^1 \psi' u$,

On a donc $\tilde{u} \in H^1(]-1,1[)$. □

Proposition 8.29. Soit Ω un ouvert de frontière Γ Lipschitzienne. Alors $\mathcal{D}(\overline{\Omega})$ est dense dans $H^1(\Omega)$.

Proposition 8.30. Soit Ω un ouvert de frontière Γ Lipschitzienne et bornée. L'application

$$\gamma_0 : \varphi \in \mathcal{D}(\overline{\Omega}) \longmapsto \varphi_{|\Gamma},$$

se prolonge par continuité en une application linéaire de $H^1(\Omega)$ dans $L^2(\Gamma)$.

DÉMONSTRATION : On se limite ici à une démonstration dans le cas du demi espace $\mathbb{R}^{N-1} \times \mathbb{R}^+$ (pour lequel le résultat est vrai malgré le caractère non borné), et l'on se reportera à [**3**] pour une démonstration plus complète. On peut se limiter à des fonctions régulières nulles pour $x_N \geq 1$. Pour une telle fonction, on a

$$\varphi(x', 0) = \int_{1}^{0} \partial_N\varphi,$$

d'où

$$\int_{\mathbb{R}^{N-1}} \varphi(x', 0)^2 = \int_{\mathbb{R}^N} \left(\int_{1}^{0} \partial_N\varphi\right)^2 \leq \int_{\mathbb{R}^N} |\partial_N\varphi|^2 \leq \int_{\mathbb{R}^N} |\nabla u|^2 .$$

□

Remarque 8.31. On notera que seul le contrôle sur la dérivée dans la direction verticale (normale à la frontière) a été utilisé dans la démonstration précédente. La rigidité transverse (selon $\mathbb{R}^{N-1}$ dans le cas précédent) va conditionner la régularité de la trace (dont on peut montrer qu'elle est strictement plus régulière que L^2).

Définition 8.32. (Espace $H^{1/2}(\Gamma)$)
On note $H^{1/2}(\Gamma) \subset L^2(\Gamma)$ l'image de l'application $\gamma_0 : H^1(\Omega) \longmapsto L^2(\Gamma)$ définie ci-dessus. C'est un espace de Banach pour la norme

$$\|g\|_{H^{1/2}(\Gamma)} = \inf_{\gamma_0 v = g} \|v\|_{H^1(\Omega)}.$$

Remarque 8.33. L'espace $H^{1/2}$ peut se définir sur l'espace entier par la transformée de Fourier (voir définition 8.73), puis par cartes locales sur une variété régulière. Il est essentiel de garder à l'esprit que l'inclusion de $H^{1/2}$ est stricte. En particulier, l'appartenance à $H^{1/2}$ exclut les discontinuités franches (voir remarque 8.76, page 110).

Proposition 8.34. L'espace $H_0^1(\Omega)$ est constitué des fonctions de $H^1(\Omega)$ dont la trace sur $\partial\Omega$ est nulle.

Démonstration : Voir Raviart [**13**]. □

Définition 8.35. (Dérivée normale)
Soit Ω un domaine de frontière Lipschitzienne. On note $\mathbf{n}$ le vecteur normal à Γ dirigé vers l'extérieur de Ω. Ce vecteur est défini presque partout. Pour toute fonction $\varphi \in \mathcal{D}(\overline{\Omega})$, on appelle dérivée normale de φ en un point de Γ la quantité

$$\frac{\partial \varphi}{\partial n} = \nabla\varphi \cdot \mathbf{n}.$$

Définition 8.36. Soit Ω un ouvert borné de frontière Γ lipschitzienne. On définit γ_1 comme l'application de $H^2(\Omega)$ dans $L^2(\Gamma)$ qui à $u \in H^2(\Omega)$ associe $\nabla u \cdot \mathbf{n}$, où la trace de chaque composante de ∇u est définie comme précédemment. On notera

$$\gamma_1 u = \frac{\partial u}{\partial n}.$$

Noter que l'on n'utilise pas ici la densité de $\mathcal{D}(\overline{\Omega})$ dans $H^2(\Omega)$ (qui, de fait, n'est pas exigée).

Proposition 8.37. (Première formule de Green)
Soit Ω un ouvert borné de frontière Γ Lipschitzienne. Pour tous u et v dans $H^1(\Omega)$, on a

$$\int_\Omega v\nabla u = -\int_\Omega u\nabla v + \int_\Gamma uv\mathbf{n}.$$

Proposition 8.38. (Deuxième formule de Green)
Soit Ω un ouvert borné de frontière Γ lipschitzienne. Pour tout u dans $H^2(\Omega)$ et tout v dans $H^1(\Omega)$, on a

$$-\int_\Omega v\triangle u = \int_\Omega \nabla u \cdot \nabla v - \int_\Gamma \frac{\partial u}{\partial n} v.$$

Proposition 8.39. Soit Ω un ouvert borné de frontière Γ lipschitzienne. On suppose que Ω se décompose de la façon suivante

$$\overline{\Omega} = \bigcup_{i=1,\dots,p} \overline{\Omega}_i,$$

où les Ω_i sont des ouverts de frontière lipschitzienne, inclus dans Ω, deux à deux disjoints. On note $\Gamma_{ij} = \overline{\Omega}_i \cap \overline{\Omega}_j$. Soit u une fonction définie sur Ω, dont la restriction u_i à Ω_i est dans $H^1(\Omega_i)$ pour tout $i = 1, \ldots, p$. On suppose que pour tous i, j tels que $\Gamma_{ij} \neq \emptyset$ les traces de u_i et u_j sur Γ_{ij} s'identifient. Alors u est dans $H^1(\Omega)$.

DÉMONSTRATION : On note $\mathbf{v}$ la fonction de $L^2(\Omega)$ qui s'identifie à ∇u sur chacun des Ω_r. Pour tout $\varphi \in \mathcal{D}(\mathbb{R}^N)$, on a (en utilisant la proposition 8.37 sur chacun des Ω_r),

$$\begin{aligned}\int_\Omega \mathbf{v}\varphi &= \sum_{i=1}^{p} \int_{\Omega_i} \mathbf{v}\varphi \\ &= -\sum_{i=1}^{p} \int_{\Omega_i} u\nabla\varphi + \sum_{i,j} \int_{\Gamma_{ij}} u\varphi(\mathbf{n}_i + \mathbf{n}_j),\end{aligned}$$

où $\mathbf{n}_i$ (resp. $\mathbf{n}_j$) est la normale à Γ_{ij} sortante au domaine Ω_i (resp. Ω_j), de telle sorte que $\mathbf{n}_i + \mathbf{n}_j = 0$. On a donc bien $u \in H^1(\Omega)$ avec $\nabla u = \mathbf{v}$. □

Remarque 8.40. On prendra garde au fait que (on reprend les notation du théorème précédent), même si u est dans $H^2(\Omega_i)$ pour tout i, le raccord des traces sur les interfaces ne suffit pas pour assurer l'appartenance de u à $H^2(\Omega)$. Cette remarque est à la base des difficultés que l'on peut avoir à approcher une fonction sur un maillage qui ne respecte pas la géométrie.

Proposition 8.41. On se replace dans le cadre des notations de la proposition précédente. Soit u une fonction définie sur Ω, dont la restriction u_i à Ω_i est dans $H^2(\Omega_i)$ pour tout $i = 1, \ldots, R$. On suppose que pour tous i, j tels que $\Gamma_{ij} \neq \emptyset$ les traces de u_i et u_j sur Γ_{ij} s'identifient. On suppose d'autre part le raccord des dérivées normales : $\partial u_i/\partial n = \partial u_j/\partial n$ sur Γ_{ij}. Alors u est dans $H^2(\Omega)$.

8.5. Injections

Théorème 8.42. Soit Ω un domaine borné de frontière Lipschitzienne. Alors, pour tout entier $m > N/2$, $H^m(\Omega)$ s'injecte de façon continue dans $C^0(\overline{\Omega})$. En particulier les fonctions de $H^2(\Omega)$ sont continues pour les dimensions physiques $N = 1$, 2, ou 3.

On retrouve notamment le fait déjà énoncé que les fonctions de $H^1(I)$, où I est un intervalle réel, sont continues. En revanche, le théorème ne s'applique pas à $H^1(\Omega)$ en dimension 2. Il existe effectivement des fonctions de $H^1(\mathbb{R}^2)$ qui ne sont pas continues.

On notera également qu'une fonction de $H^2(\Omega)$ est continue sur Ω, sans hypothèse de régularité, car tout $x \in \Omega$ est dans une boule incluse dans Ω. En l'absence de régularité du bord, il est en revanche possible que l'on n'ait pas $\|u\|_\infty \leq C \|u\|_{H^2}$.

Théorème 8.43. (Rellich)
Soit Ω un domaine borné de frontière Lipschitzienne. Alors l'injection de $H^1(\Omega)$ dans $L^2(\Omega)$ est compacte. L'injection de $H^1_0(\Omega)$ dans $L^2(\Omega)$ est compacte pour tout Ω borné (sans hypothèse de régularité). De même, l'injection de $H^{m+1}(\Omega)$ dans $H^m(\Omega)$ est compacte.

DÉMONSTRATION : On se reportera à la section consacrée à la transformée de Fourier (voir théorème 8.74) pour une démonstration de ce théorème. On peut également démontrer la compacité de l'injection en utilisant le point (*iii*) de la caractérisation 8.16 de $H^1(\Omega)$, et le théorème de Riesz-Fréchet-Kolmogorov qui donne un critère suffisant de relative compacité pour des familles de fonctions de $L^2(\Omega)$ (voir Brezis [**3**, Th. IV.25 & Cor. IV.26]). □

EXERCICE 8.1. Montrer que l'injection de $H^1(\Omega)$ dans $L^2(\Omega)$ n'est jamais compacte quand Ω n'est pas borné.

8.6. Champs de vecteurs

Définition 8.44. (Gradient, divergence, taux de déformation d'un champ de vecteurs)
Soit $\mathbf{u} = (u_1, u_2, \dots, u_N)$ un champ de vecteurs régulier défini sur un domaine Ω. On appelle gradient de $\mathbf{u}$ le champ de matrices

$$\nabla \mathbf{u} = \begin{pmatrix} \dfrac{\partial u_1}{\partial x_1} & \cdots & \dfrac{\partial u_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial u_N}{\partial x_1} & \cdots & \dfrac{\partial u_N}{\partial x_N} \end{pmatrix}.$$

On appelle divergence de $\mathbf{u}$ le champ scalaire

$$\nabla \cdot \mathbf{u} = \sum_{1 \le i \le N} \frac{\partial u_i}{\partial x_i}.$$

On notera $e(\mathbf{u})$ la partie symétrique du gradient de $\mathbf{u}$ (tenseur des taux de déformation) :

$$e(\mathbf{u}) = \frac{1}{2}\left(\nabla \mathbf{u} + {}^t\nabla \mathbf{u}\right) = \frac{1}{2} \begin{pmatrix} 2\dfrac{\partial u_1}{\partial x_1} & \cdots & \dfrac{\partial u_1}{\partial x_N} + \dfrac{\partial u_N}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial u_N}{\partial x_1} + \dfrac{\partial u_1}{\partial x_N} & \cdots & 2\dfrac{\partial u_N}{\partial x_N} \end{pmatrix}.$$

Les notions ci-dessus s'étendent immédiatement aux champs dont les composantes appartiennent à l'espace de Sobolev H^1. Les identités ci-dessus se lisent alors comme des identités entre fonctions de $L^2(\Omega)$.

Proposition 8.45. Soit Ω un domaine borné de frontière lipschitzienne, et $\mathbf{u}$ un champ de vecteurs de $H^1(\Omega)^N$. On a

$$\int_\Omega \nabla \cdot \mathbf{u} = \int_\Gamma \mathbf{u} \cdot \mathbf{n}.$$

NOTATION : Soient $\mathbf{u}$ et $\mathbf{v}$ deux champs de vecteurs. On note $\nabla \mathbf{u} : \nabla \mathbf{v}$ le champ scalaire

$$\nabla \mathbf{u} : \nabla \mathbf{v} = \sum_{1 \le i,j \le N} \frac{\partial u_i}{\partial x_j} \frac{\partial v_i}{\partial x_j}.$$

Proposition 8.46. (Formule de Green pour les champs de vecteurs)
Soit Ω un ouvert borné de frontière Γ de frontière lipschitzienne. Pour tout $\mathbf{u}$ dans $H^2(\Omega)^N$ et tout $\mathbf{v}$ dans $H^1(\Omega)^N$, on a

$$-\int_\Omega \mathbf{v}\cdot\triangle\mathbf{u} = \int_\Omega \nabla\mathbf{u} : \nabla\mathbf{v} - \int_\Gamma \mathbf{v}\cdot(\nabla\mathbf{u}\cdot\mathbf{n})$$

Proposition 8.47. Soit Ω un ouvert borné de frontière Γ de frontière lipschitzienne. Pour tout $\mathbf{u}$ dans $H^2(\Omega)^N$ et tout $\mathbf{v}$ dans $H^1(\Omega)^N$, on a

$$-\int_\Omega \mathbf{v}\cdot\nabla\cdot\left(\nabla\mathbf{u}+{}^t\nabla\mathbf{u}\right) = \frac{1}{2}\int_\Omega \left(\nabla\mathbf{u}+{}^t\nabla\mathbf{u}\right) : \left(\nabla\mathbf{v}+{}^t\nabla\mathbf{v}\right) - \int_\Gamma \mathbf{v}\cdot\left(\nabla\mathbf{u}+{}^t\nabla\mathbf{u}\right)\cdot\mathbf{n}$$

8.7. Inégalités de Poincaré, de Korn

Proposition 8.48. (Inégalité de Poincaré)
Soit Ω un domaine de $\mathbb{R}^N$ borné dans une direction, c'est-à-dire tel que

$$\Omega \subset \left\{x\in\mathbb{R}^N\,,\ \xi\cdot x\in]a,b[\right\}.$$

Alors il existe une constante $C>0$ telle que

$$\left(\int_\Omega |u|^2\right)^{1/2} \le C\left(\int_\Omega |\nabla u|^2\right)^{1/2} \qquad \forall u\in H_0^1(\Omega).$$

Démonstration : On note toujours u le prolongement par 0 de u sur $\mathbb{R}^N$ tout entier. Quitte à effectuer une translation et une rotation du système de coordonnées, on suppose que la bande qui contient Ω se met sous la forme

$$\left\{x=(x_1,\dots,x_N)=(x',x_N)\in\mathbb{R}^N\,,\ x_N\in]0,L[\right\}.$$

On suppose dans un premier temps u régulière. Pour tout $x=(x',x_N)\in\Omega$, on a

$$u(x',x_N) = u(x',0) + \int_0^{x_N} \partial_N u = \int_0^{x_N} \partial_N u,$$

d'où, d'aprés l'inégalité de Cauchy-Schwarz,

$$u(x',x_N)^2 \le L\int_0^L |\nabla u|^2.$$

On a donc

$$\begin{aligned}\int_\Omega u^2 &\le L\int_{\mathbb{R}^{N-1}}\int_0^L\int_0^L |\nabla u|^2\\ &\le L^2\int_{\mathbb{R}^{N-1}}\int_0^L |\nabla u|^2 = \int_\Omega |\nabla u|^2.\end{aligned}$$

On conclut en utilisant la densité des fonctions régulières. □

Remarque 8.49. On appelle constante de Poincaré du domaine Ω le plus petit réel C_Ω tel que l'inégalité ci-dessus est vérifiée. On a

$$\frac{1}{C_\Omega^2} = \inf_{u\neq 0} \frac{\displaystyle\int_\Omega |\nabla u|^2}{\displaystyle\int_\Omega |u|^2}.$$

On peut ainsi montrer $1/C_\Omega^2 = \lambda_1$, où λ_1 est la plus petite valeur propre du Laplacien avec conditions de Dirichlet, c'est-à-dire le plus petit réel tel qu'il existe $u \in H_0^1(\Omega)$ non nul vérifiant [12]

$$-\triangle u = \lambda u.$$

La proposition précédente assure $\lambda_1 \geq 1/L^2$, pour tout domaine Ω inclus dans une bande d'épaisseur L.

Corollaire 8.50. Soit Ω un domaine de $\mathbb{R}^N$ borné dans une direction. Alors la forme bilinéaire

$$(u, v) \longmapsto \int_\Omega \nabla u \cdot \nabla v$$

est un produit scalaire sur $H_0^1(\Omega)$, qui induit une norme équivalente à la norme de départ.

L'inégalité de Poincaré énoncée ci-dessus est un cas particulier d'une inégalité plus générale :

Proposition 8.51. (Inégalité de Poincaré généralisée)
Soit Ω un domaine régulier, borné, et connexe, et T une application linéaire continue de $H^1(\Omega)$ dans un espace de Hilbert M. On suppose que l'image par T d'une fonction constante non nulle est elle-même non nulle. Alors il existe une constante C telle que

$$|u|_0 \leq C\,(|Tu|_M + |\nabla u|_0) \quad \forall u \in H^1(\Omega).$$

Démonstration : On raisonne par l'absurde. Si la propriété est fausse, alors pour tout n on peut construire $u_n \in H^1(\Omega)$ tel que

$$\|u_n\|_{L^2} > n\,(|Tu_n|_M + |\nabla u_n|_0) \quad \forall u \in H^1(\Omega).$$

On peut choisir u_n tel que $\|u_n\| = 1$. La suite u_n étant bornée dans H^1, on peut en extraire une sous-suite (que nous noterons toujours (u_n)) qui converge fortement dans $L^2(\Omega)$ (l'injection de $H^1(\Omega)$ dans $L^2(\Omega)$ étant compacte), vers $u \in L^2(\Omega)$). Comme la suite (∇u_n) tend vers 0 dans L^2, elle est de Cauchy, et par suite (u_n) est de Cauchy dans H^1. Elle converge donc dans H^1 vers une limite, qui est nécessairement la limite u dans L^2. Comme Tu_n tend vers 0, on a nécessairement $Tu = 0$. D'autre part, comme $(\nabla u_n) \to 0$, on a $\nabla u = 0$, et ainsi u est constante sur Ω (voir proposition 8.14, page 93). Comme $Tu = 0$, cette constante est nulle, ce qui est absurde car $\|u\| = \lim \|u_n\| = 1$ □

La démonstration ci-dessus permet d'établir directement la propriété suivante :

Corollaire 8.52. Soit Ω un domaine régulier, borné, et connexe, et V un sous-espace fermé de $H^1(\Omega)$ qui ne contient aucune fonction constante autre que 0. Alors il existe $C > 0$ tel que

$$|u|_0 \leq C\,|\nabla u|_0 \quad \forall u \in V.$$

Remarque 8.53. Ce corollaire s'appliquera notamment au cas où V est un espace de fonctions qui s'annulent sur une partie de la frontière de mesure non nulle. Sur un tel espace, $|u|_1$ est une norme équivalent à la norme H^1.

12. L'opérateur de Laplace $-\triangle$, qui fait intervenir des dérivées secondes, n'est *a priori* défini pour des fonctions de H^1 qu'au sens des distributions. On verra par la suite que ces dérivées secondes du minimiseur u peuvent en fait être définies dans le cadre de ce chapitre, c'est-à-dire en tant que fonctions de $L^2(\Omega)$ (ou tout du moins L^2_{loc} sans hypothèse sur le domaine), de telle sorte que l'on pourra écrire $-\triangle u = \lambda u$ presque partout.

Théorème 8.54. (Inégalité de Korn)
Soit Ω un domaine de $\mathbb{R}^N$ borné régulier. Alors il existe une constante $C > 0$ telle que

$$\|\mathbf{u}\|_{H^1(\Omega)} \leq C \left(\int_\Omega |\mathbf{u}|^2 + \int_\Omega \left| \nabla \mathbf{u} + {}^t\nabla \mathbf{u} \right|^2 \right)^{1/2} \quad \forall \mathbf{u} \in H^1(\Omega)^N.$$

DÉMONSTRATION : On trouvera une démonstration de cette propriété dans [**6**]. La première étape consiste à vérifier que l'on peut reconstruire toutes les dérivées secondes d'une fonction à partir de la connaissance des dérivées partielles des quantités qui apparaissent dans le tenseur $\nabla \mathbf{u} + {}^t\nabla \mathbf{u}$. La suite est délicate, elle se base sur une caractérisation des fonctions de L^2 comme distributions dans $H^{-1}(\Omega)$, dual topologique de $H_0^1(\Omega)$, donc les dérivées partielles par rapport à chacune des variables d'espace sont aussi dans $H^{-1}(\Omega)$. Pour la démonstration complète de cette propriété, voir [**7**, Th. 3.2].

Nous présentons ici une démonstration de cette propriété dans le cas de H_0^1. Pour tout champ de déplacement régulier $\mathbf{u}$, on a

$$\frac{1}{2} \int_\Omega \left| \nabla \mathbf{u} + {}^t\nabla \mathbf{u} \right|^2 = \int_\Omega (\nabla \mathbf{u} + {}^t\nabla \mathbf{u}) : {}^t\nabla \mathbf{u} = \int_\Omega |\nabla \mathbf{u}|^2 + \int_\Omega \nabla \mathbf{u} : {}^t\nabla \mathbf{u} = \int_\Omega |\nabla \mathbf{u}|^2 + \int_\Omega (\nabla \cdot \mathbf{u})^2$$

d'après la proposition 13.17, page 167. □

A l'aide de cette inégalité on démontre (de façon analogue à la proposition 8.51) une inégalité généralisée qui prend la forme suivante :

Corollaire 8.55. (Inégalité de Korn généralisée)
Soit Ω un domaine de $\mathbb{R}^N$ borné régulier, de frontière Γ. Soit V un sous-espace de $H^1(\Omega)^N$ qui ne contient aucun déplacement élémentaire rigide non trivial [13]. Il existe alors une constante $C > 0$ telle que

$$\|\mathbf{u}\|_{H^1(\Omega)} \leq C \, |e(\mathbf{u})|_0 \, .$$

Remarque 8.56. Ce corollaire s'applique en particulier au cas où le matériau est fixé sur une partie de la frontière, où V est alors le sous-espace des champ de $H^1(\Omega)^N$ qui s'annulent sur $\Gamma_0 \subset \Gamma$, partie de la frontière de mesure non nulle.

8.8. Problèmes aux limites elliptiques

Nous précisons ici la démarche qui permet de donner un cadre mathématique à un certain type de problèmes elliptiques.

8.8.1. Existence et unicité de solutions. Nous présentons dans cette section des résultats classiques d'existence et d'unicité de solutions pour le problème de Poisson.

Conditions aux limites de Dirichlet.

13. C'est-à-dire que $\mathbf{a} + \boldsymbol{\omega} \wedge \mathbf{x} \in V$ implique $\mathbf{a} = \boldsymbol{\omega} = 0$.

On s'intéresse ici à la résolution [14] de problèmes du type

$$\left\{ \begin{array}{rcl} -\triangle u & = & f \quad \text{dans} \quad \Omega \\ u & = & 0 \quad \text{sur} \quad \partial\Omega, \end{array} \right. \tag{8.1}$$

où f est une fonction de $L^2(\Omega)$ donnée. On parlera du problème de Poisson dans le domaine Ω.

Définition 8.57. (Solution faible)
On appellera solution faible de (8.1) une fonction de $H_0^1(\Omega)$ telle que

$$\int_\Omega \nabla u \cdot \nabla v = \int_\Omega fv \quad \forall v \ H_0^1(\Omega). \tag{8.2}$$

Proposition 8.58. (Principe de Dirichlet)
On suppose Ω borné dans une direction. Soit $f \in L^2(\Omega)$. Alors le problème 8.1 admet une unique solution faible : il existe un unique $u \in H_0^1(\Omega)$ solution de la formulation variationnelle (8.2). C'est l'unique élément de $H_0^1(\Omega)$ qui minimise la fonctionnelle

$$v \longmapsto \frac{1}{2}\int_\Omega |\nabla v|^2 - \int fv.$$

Démonstration : C'est une application directe du théorème de Lax-Milgram, avec

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v\,, \ \langle \varphi\,, v\rangle = \int_\Omega fv.$$

Noter que la forme bilinéaire $a(\cdot,\cdot)$ est bien coercive grâce à l'inégalité de Poincaré (proposition 8.48, page 102). □

Conditions aux limites de Neumann.

On considère maintenant des conditions au bord de type Neumann. Comme ces conditions ne font intervenir que les dérivées, comme l'opérateur de Laplacien lui-même, le problème de Poisson avec de telles conditions est évidemment mal posé (si l'on ajoute une fonction constante, qui est bien dans $H^1(\Omega)$ dès que Ω est borné, à n'importe quelle solution, on obtient bien une autre solution). On verra à la fin de cette section que ce problème est pourtant bien posé dans un certain espace, sous réserve que f vérifie une certaine condition. Dans un premier temps, nous utilisons un moyen élémentaire de contourner ce problème, qui consiste à rajouter au Laplacien un terme d'ordre 0. On s'intéressera donc au problème suivant

$$\left\{ \begin{array}{rcl} u - \triangle u & = & f \quad \text{dans} \quad \Omega \\ \dfrac{\partial u}{\partial n} & = & 0 \quad \text{sur} \quad \partial\Omega, \end{array} \right. \tag{8.3}$$

où f est donnée.

Définition 8.59. On appellera solution classique (dans le cas où f est au moins continue) une fonction de $C^2(\overline{\Omega})$ qui vérifie le système ci-dessus, et solution faible une fonction de $H^1(\Omega)$ telle que

$$\int_\Omega uv + \int_\Omega \nabla u \cdot \nabla v = \int_\Omega fv \quad \forall v \ H^1(\Omega). \tag{8.4}$$

14. Résolution au sens théorique du terme : on se préoccupe ici du caractère bien posé des problèmes (existence et unicité de solution, dépendance continue par rapport aux données), et à la régularité des solutions.

L'existence et l'unicité d'une solution faible est immédiate sans qu'il soit nécessaire de faire des hypothèses sur le domaine, comme le précise la proposition ci-dessous. Il est en revanche délicat de préciser en quel sens une solution faible est solution de (8.3), car la dérivée normale n'est en général pas définie sur le bord.

Proposition 8.60. Soit $f \in L^2(\Omega)$. Alors le problème 8.3 admet une unique solution faible. Cette solution faible est l'élément de $H_0^1(\Omega)$ qui minimise la fonctionnelle

$$v \longmapsto \frac{1}{2}\int_\Omega |v|^2 + \frac{1}{2}\int_\Omega |\nabla v|^2 - \int fv.$$

Démonstration : C'est de nouveau une application directe du théorème de Lax-Milgram dans $H = H^1(\Omega)$. □

8.8.2. Régularité des solutions faibles. Nous abordons maintenant le problème de régularité des solutions faibles construites précédemment. Il s'agit notamment de déterminer si l'équation de départ est vérifiée comme identité entre fonctions mesurables (auquel cas il est licite de préciser *presque partout*), ou dans un sens plus faible. On considère ainsi des équations aux dérivées partielles du type

$$-\triangle u = f\,,\ u - \triangle u = f \text{ ou } -\nabla k \cdot \nabla u = f,$$

où $\triangle$ est le Laplacien $\triangle = \sum \partial^2/\partial x_i^2$, k est un champ scalaire régulier tel que $0 < m \leq k(x) \leq M < +\infty$.

Proposition 8.61. Soit Ω un domaine de $\mathbb{R}^N$ et $u \in H^1(\Omega)$. On suppose qu'il existe $f \in L^2(\Omega)$ tel que

$$\int_\Omega \nabla u \cdot \nabla \varphi = \int_\Omega f\varphi \quad \forall \varphi \in \mathcal{D}(\Omega).$$

Alors u est dans $H^2_{\text{loc}}(\Omega)$ et vérifie

$$-\triangle u = f \quad \text{p.p.}$$

Démonstration : On suppose dans un premier temps que Ω est l'espace $\mathbb{R}^N$ tout entier. Comme $\mathcal{D}(\Omega)$ est alors dense dans $H^1(\Omega)$, la formulation variationnelle est vérifiée pour toute fonction test de $H^1(\Omega)$, en particulier les fonctions-test particulières que nous allons construire à partir de u. Pour $h \in \mathbb{R}^N$, on introduit

$$D_h u = \frac{1}{|h|}\left(\tau_h u - u\right),$$

et l'on écrit la formulation variationnelle avec $v = D_{-h}D_h u$. Il vient

$$\int_{\mathbb{R}^N} \nabla u \cdot \nabla v = \frac{1}{|h|^2}\int_{\mathbb{R}^N} \nabla u \cdot \left(\tau_h \nabla u - 2\nabla u + \tau_{-h}\nabla u\right).$$

On peut écrire

$$\int_{\mathbb{R}^N} \nabla u \cdot \left(-\nabla u + \tau_{-h}\nabla u\right) = \int_{\mathbb{R}^N} \tau_h \nabla u \cdot \left(-\tau_h \nabla u + \nabla u\right),$$

d'où finalement

$$\int_{\mathbb{R}^N} |D_h \nabla u|^2 \leq \|f\|_{L^2}\,\|D_{-h}D_h u\|_{L^2} \leq \|f\|_{L^2}\,\|\nabla D_h u\|_{L^2} = \|f\|_{L^2}\,\|D_h \nabla u\|_{L^2},$$

d'après la proposition 8.16 ($(i) \Rightarrow (iii)$). On a donc

$$\|D_h \nabla u\|_{L^2} \leq \|f\|_{L^2}$$

pour tout $h \in \mathbb{R}^N$. On a donc $\|D_h \partial_i u\|_{L^2}$ uniformément borné, et donc, toujours d'après la proposition 8.16, $\partial_i u \in H^1(\mathbb{R}^N)$ pour tout $i = 1, \ldots, N$.

Dans le cas général on considère une fonction $\theta \in \mathcal{D}(\Omega)$. On a

$$\nabla(\theta u) \cdot \nabla \varphi = \nabla u \cdot \nabla(\theta\varphi) + \nabla\theta \cdot \nabla(u\varphi) - 2\varphi \nabla u \cdot \nabla\varphi,$$

et ainsi la fonction $\theta u \in H^1(\mathbb{R}^N)$ vérifie

$$\int_{\mathbb{R}^N} \nabla(\theta u) \cdot \nabla\varphi = \int_{\mathbb{R}^N} \theta f \varphi - 2\int_{\mathbb{R}^N} \varphi \nabla u \cdot \nabla\theta - \int_{\mathbb{R}^N} \varphi u \triangle\theta = \int_{\mathbb{R}^N} g\varphi \quad \forall \varphi \in \mathcal{D}(\Omega).$$

avec $g \in L^2(\mathbb{R}^N)$. La fonction θu est donc dans $H^2(\mathbb{R}^N)$ d'après ce qui précède. On a donc bien $u \in H^2_{\text{loc}}(\Omega)$. □

Proposition 8.62. On suppose Ω borné dans une direction. Soit f un élément de $L^2(\Omega)$. La solution faible $u \in H^1_0(\Omega)$ de (8.2)) avec conditions de Dirichlet homogènes est dans $H^2_{\text{loc}}(\Omega)$ et vérifie

$$-\triangle u = f \quad \text{p.p.}$$

Démonstration : C'est une application directe de la proposition 8.61. □

Le passage de la régularité H^2_{loc} à l'appartenance à $H^2(\Omega)$ est loin d'être immédiat. Nous nous bornerons ici à énoncer des résultats de régularité dans un certain nombre de situations.

Proposition 8.63. Soit Ω un domaine de classe C^2, borné dans une direction, et de frontière Γ bornée. Pour tout f dans $L^2(\Omega)$, la solution faible de $-\triangle u = f$ avec conditions aux limites de Dirichlet homogènes appartient à H^2, et il existe une constante C (qui dépend du domaine Ω) telle que

$$\|u\|_{H^2} \leq C \|f\|_{L^2}.$$

Démonstration : L'appartenance à $H^2_{\text{loc}}(\Omega)$ est assurée par la proposition 8.61. On se reportera à Brezis [**3**, Th. IX.25] pour une étude détaillée de la régularité près du bord. La démonstration, très technique, utilise des changements de variables permettant de se ramener au cas d'une frontière hyperplane. Pour ce dernier cas, la régularité jusqu'au bord est démontrée selon une méthode de translation analogue à celle utilisée dans la proposition 8.61, les translations étant effectuées parallèlement au bord considéré. □

Proposition 8.64. Les conclusions du théorème ci-dessus sont valides si l'on suppose le domaine polyédrique et convexe.

Proposition 8.65. Les conclusions du théorème ci-dessus s'appliquent à l'équation

$$-\nabla \cdot k \nabla u = f,$$

où k est une fonction C^1 de la variable d'espace sur $\overline{\Omega}$, minorée par une constante

Remarque 8.66. Le cas de conditions aux limites panachées (Dirichlet sur une partie du bord, Neumann sur une autre) et très délicat. Nous admettrons que le passage d'un type de condition à l'autre ne pose pas de problème lorsque les deux composantes de la frontière se rencontrent à angle droit. On trouvera dans [**5**] une analyse détaillée de la régularité dans ce type de situation, en fonction de l'angle du raccord entre les composantes.

Remarque 8.67. Le résultat ne s'étend pas au cas d'un polyèdre non convexe, comme l'illustre l'exercice **??**, page **??**.

Remarque 8.68. Si l'on considère le problème

$$u - \triangle u = f,$$

avec conditions aux limites de Dirichlet, tout ce qui a été dit précédemment reste valable, sans que l'on ait besoin de l'hypothèse que Ω soit borné dans une direction pour assurer l'existence et l'unicité d'une solution faible.

Proposition 8.69. Soit Ω un domaine de frontière C^2 et bornée, et f un élément de $L^2(\Omega)$. La solution de (8.4) appartient à H^2, et sa dérivée normale est nulle sur $\Gamma = \partial\Omega$.

8.9. Compléments

8.9.1. Espaces de Sobolev et transformation de Fourier. On peut définir les espaces de Sobolev l'aide de la transformée de Fourier. Cette approche est particulièrement adaptée aux problèmes posés sur l'espace tout entier, ou en géométrie périodique, ce qui la place un peu en marge de cet ouvrage dont l'un des objectifs est précisément la prise en compte de géométries complexes en domaines bornés. Nous indiquons néammoins ici certains éléments de cette approche, qui permet notamment de bien comprendre le théorème de Rellich, qui est à la base de l'analyse de la méthode des éléments finis.

Définition 8.70. Soit $u \in L^2(\mathbb{R}^N)$. On définit sa transformée de Fourier comme la fonction définie par

$$\tilde{u}(\xi) = \frac{1}{(2\pi)^{-n/2}} \int_{\mathbb{R}^N} e^{-i\xi\cdot x} u(x)\, dx.$$

Théorème 8.71. L'application $u \longmapsto \tilde{u}$ est une isométrie de $L^2(\mathbb{R}^N)$ sur lui-même.

L'espace $H^1(\mathbb{R}^N)$ peut se définir à l'aide de la transformée de Fourier, ce que nous présentons ici comme un thérorème si l'on prend la définition 8.10, page 92 comme référence.

Théorème 8.72. L'espace $H^1(\mathbb{R}^N)$ est l'ensemble des fonctions u de $L^2(\mathbb{R}^N)$ telles que

$$\left(1 + |\xi|^2\right)^{1/2} \tilde{u} \in L^2(\mathbb{R}^N).$$

On définit de la même manière les espaces $H^s(\mathbb{R}^N)$ pour tout $s \geq 0$:

Définition 8.73. L'espace $H^s(\mathbb{R}^N)$ est l'ensemble des fonctions u de $L^2(\mathbb{R}^N)$ telles que

$$\int_{\mathbb{R}^N} \left(1 + |\xi|^2\right)^s |\tilde{u}|^2 < +\infty.$$

Nous démontrons à présent le théorème de Rellich 8.43 déjà énoncé à la page 100.

Théorème 8.74. Soit Ω un domaine borné de frontière lipschitzienne. L'injection de $H^1(\Omega)$ dans $L^2(\Omega)$ est compacte.

DÉMONSTRATION : On considère une suite (u_n) bornée dans $H^1(\Omega)$. On note P l'opérateur de prolongement de la proposition 8.28, page 97. On choisit P de telle sorte que Pv soit nul à l'extérieur d'un borné K, pour tout $v \in H^1(\Omega)$. On conserve la notation (u_n) pour désigner l'image par P de la suite initiale. D'après le théorème 6.32, page 74, on peut en extraire une sous-suite qui converge faiblement dans $H^1(\mathbb{R}^N)$. On notera toujours (u_n) cette sous-suite. Quitte à translater la suite, on suppose que la limite faible est 0. On écrit à présent, pour tout $M \geq 0$

$$\|u_n\|_{L^2}^2 = \|\tilde{u}_n\|_{L^2}^2 = \int_{|\xi|<M} |\tilde{u}_n|^2 + \int_{|\xi|>M} |\tilde{u}_n|^2 \leq \int_{|\xi|<M} |\tilde{u}_n|^2 + \frac{1}{1+M^2} \int_{|\xi|>M} \left(1+|\xi|^2\right) |\tilde{u}_n|^2 .$$

Le second terme tend vers 0 quand M tend vers $+\infty$. Il suffit donc de montrer que, pour M fixé, le premier terme tend vers 0. On a, pour tout ξ,

$$\tilde{u}_n(\xi) = \frac{1}{(2\pi)^{-n/2}} \int_{\mathbb{R}^N} e^{-i\xi\cdot x} u_n(x)\, dx = \frac{1}{(2\pi)^{-n/2}} \int_{\mathbb{R}^N} \chi_K e^{-i\xi\cdot x} u_n(x)\, dx,$$

où χ_K est la fonction caractéristique de K (de telle sorte que $\chi_K e^{-i\xi\cdot x}$ est dans $L^2(\mathbb{R})$), Cette quantité tend donc vers 0 quand n tend vers $+\infty$ d'après la convergence faible de u_n vers 0 dans L^2. Comme par ailleurs $|\tilde{u}_n(\xi)|^2$ est majoré par une constante, le théorème de convergence dominée assure donc la convergence de $|\tilde{u}_n(\xi)|^2$ vers 0 dans $L^1(B(0,M))$. On a donc bien convergence vers 0 de $\|u_n\|_{L^2}$. □

Théorème 8.75. Soit $s > 1/2$. L'application γ_0 qui à une fonction régulière sur $\mathbb{R}^N$ associe sa restriction à $\left\{(0,x'),\ x' \in \mathbb{R}^{N-1}\right\}$ est continue de $H^s(\mathbb{R}^N)$ dans $H^{s-1/2}(\mathbb{R}^{N-1})$. Elle s'étend donc par densité en un opérateur de $\mathcal{L}\left(H^s(\mathbb{R}^N), H^{s-1/2}(\mathbb{R}^{N-1})\right)$.

DÉMONSTRATION : Soit $u \in \mathcal{D}(\mathbb{R}^N)$. Il s'agit de contrôler la norme $H^{s-1/2}(\mathbb{R}^{N-1})$ de $u(\cdot,x')$ par la norme $H^s(\mathbb{R}^N)$ de u. On note $\tilde{u}(0,\xi')$ la transformée de Fourier de $u(0,x')$ par rapport à la seconde variable, et $\hat{u}(\xi_1,\xi')$ la tranformée de Fourier par rapport à l'ensemble des variables. On a

$$\begin{aligned}
\|u(0,\cdot)\|_{H^{s-1/2}(\mathbb{R}^{N-1})}^2 &= \int_{\mathbb{R}^{N-1}} \left(1+\left|\xi'\right|^2\right)^{s-1/2} \left|\tilde{u}(0,\xi')\right|^2 d\xi' \\
&= \int_{\mathbb{R}^{N-1}} \left(1+\left|\xi'\right|^2\right)^{s-1/2} \left|\int_{\mathbb{R}} e^{i0\cdot\xi_1} \hat{u}(\xi_1,\xi') d\xi_1\right|^2 d\xi' \\
&\leq \int_{\mathbb{R}^{N-1}} \left(1+\left|\xi'\right|^2\right)^{s-1/2} d\xi' \int_{\mathbb{R}} \left|\ \hat{u}(\xi_1,\xi')\right|^2 \left(1+|\xi_1|^2+\left|\xi'\right|^2\right)^s d\xi_1 \\
&\quad \int_{\mathbb{R}} \left(1+|\xi_1|^2+\left|\xi'\right|^2\right)^{-s} d\xi_1
\end{aligned}$$

Le changement de variable $\zeta = \xi_1/(1+|\xi'|^2)^{1/2}$ dans l'intégrale ci-dessus conduit à

$$\int_{\mathbb{R}} \left(1+|\xi_1|^2+\left|\xi'\right|^2\right)^{-s} d\xi_1 = \frac{1}{(1+\left|\xi'\right|^2)^{s-1/2}} \int_{\mathbb{R}} \frac{d\zeta}{\left(1+|\zeta|^2\right)^s}.$$

En notant C l'intégrale (convergente dès que $s > 1/2$) ci-dessus, on obtient finalement

$$\|u(0,\cdot)\|_{H^{s-1/2}(\mathbb{R}^{N-1})}^2 \leq C \int_{\mathbb{R}^N} \left(1+|\xi_1|^2+\left|\xi'\right|^2\right)^s \left|\hat{u}(\xi_1,\xi')\right|^2 d\xi_1 d\xi' = C\,\|u\|_{H^1(\mathbb{R}^N)}^2 ,$$

d'où la continuité de γ_0 pour les normes voulues. □

Remarque 8.76. L'inclusion de $H^{1/2}(\Gamma)$ dans $L^2(\Omega)$ est stricte. On peut notamment vérifier que l'appartenance à $H^{1/2}(\Gamma)$ exclut les discontinuités franches. Prenons par exemple $\Gamma =]0,1[$, et considérons la base hilbertienne de $L^2(\Gamma)$ des $\varphi_k(x) = \sqrt{2}\sin(k\pi x)$. On peut caractériser l'appartenance à $H^{1/2}(\Gamma)$ d'une fonction u à l'aide des coefficient de Fourier de $u \in L^2(\Gamma)$:

$$u_k = \int_0^1 \varphi_k u = \sqrt{2}\int_0^1 u(x)\sin(k\pi x),$$

de la façon suivante :

$$u \in H^{1/2}(\Gamma) \Longleftrightarrow \sum_k k\,|u_k| < +\infty.$$

Pour la fonction caractéristique de l'intervalle $]0,1/2[$, les coefficients impairs valent

$$u_{2k+1} = \sqrt{2}\int_0^1 u(x)\sin(k\pi x) = \sqrt{2}\int_{1/2-1/2(2k+1)}^{1/2} \sin(k\pi x) = (-1)^k \frac{\sqrt{2}}{(2k+1)\pi}.$$

Les coefficients pour les modes multiples de 4 sont nuls, et

$$u_{4k+2} = \sqrt{2}\int_0^1 u(x)\sin(k\pi x) = \sqrt{2}\int_{1/2-(4k+2)}^{1/2} \sin((4k+2)\pi x) = \frac{2\sqrt{2}}{(4k+2)\pi}.$$

La série qui intervient dans la caractérisation de $H^{1/2}$ est donc divergente : u n'appartient pas à cet espace [15].

Théorème 8.77. Les applications de $H^s(\Omega)$ sont continues sur Ω dès que $s > N/2$.

Démonstration : Il suffit de montrer que toute application de $H^s(\Omega)$ est continue en 0. On écrit (à une constante multiplicative près)

$$|u(x) - u(0)| = \left| \int_{\mathbb{R}^N} (e^{ix\cdot\xi} - 1)\hat{u}(\xi)d\xi \right|$$

et on décompose cette intégrale sur une boule de rayon $M > 0$ et son complémentaire. La partie infinie s'écrit

$$\left| \int_{|x|>M} (e^{ix\cdot\xi} - 1)\hat{u}(\xi)d\xi \right| \leq 2\int_{|x|>M} |\ \hat{u}(\xi)|\ d\xi$$

$$\leq \left(\int_{|x|>M} (1+|\xi|^2)^s\ |\ \hat{u}(\xi)|^2\ d\xi \right)^{1/2} \left(\int_{|x|>M} (1+|\xi|^2)^{-s}\ d\xi \right)^{1/2}.$$

Le second facteur tend vers 0 avec M dès que $s > N/2$. A M fixé, la quantité

$$\left| \int_{|x|<M} (e^{ix\cdot\xi} - 1)\hat{u}(\xi)d\xi \right|$$

tend vers 0 par convergence dominée. □

15. Dans un contexte d'élasticité linéaire, la conséquence de ce fait est qu'une condition de déplacement imposé qui modéliserait un déchirement de la frontière ne peut se faire qu'au prix d'une dépense énergétique infinie.

8.9.2. Approche $H_{\mathbf{div}}$. Nous décrivons ici une approche qui permet de donner un sens aux équations de type problème de Poisson comme identité entre fonctions de L^2 sans passer par la régularité H^2.

Proposition 8.78. Soit Ω un domaine quelconque, et $\mathbf{v} \in L^2(\Omega)^N$. On a l'équivalence suivante :

$$\exists C\,,\ \left|\int_\Omega \mathbf{v}\cdot\nabla\varphi\right| \le C\,\|\varphi\|_{L^2(\Omega)} \quad \forall\varphi\in\mathcal{D}(\Omega) \Longleftrightarrow \exists \mathrm{q}\in L^2(\Omega) \text{ tel que } \int_\Omega \mathbf{v}\cdot\nabla\varphi = -\int_\Omega \mathrm{q}\varphi.$$

On dit alors que $\mathbf{v}$ admet une divergence faible dans $L^2(\Omega)$, et l'on écrit $\nabla\cdot\mathbf{v} = \mathrm{q}$.

Démonstration : La condition suffisante est conséquence immédiate de l'inégalité de Cauchy-Schwarz. Pour la condition nécessaire, on considère la forme linéaire

$$\varphi \longmapsto \int_\Omega \mathbf{v}\cdot\nabla\varphi$$

définie sur $\mathcal{D}(\Omega)$. Comme elle est continue pour la norme $L^2(\Omega)$ d'après l'hypothèse, cette forme se prolonge par densité en une forme linéaire continue sur $L^2(\Omega)$. Comme il s'agit d'un espace de Hilbert, cette forme admet un représentant $\mathrm{q}\in L^2(\Omega)$. □

Définition 8.79. (Espace H_{div})
On notera H_{div} l'ensemble des champs de vecteurs $\mathbf{u}\in L^2(\Omega)^N$ qui admettent une divergence faible L^2 au sens de la proposition précédente.

Proposition 8.80. L'espace H_{div} est un espace de Hilbert pour le produit scalaire

$$(\mathbf{u},\mathbf{v})_{H_{\mathrm{div}}} = \int_\Omega \mathbf{u}\cdot\mathbf{v} + \int_\Omega (\nabla\cdot\mathbf{u})\,(\nabla\cdot\mathbf{v})\,.$$

Démonstration : On considère une suite de Cauchy $(\mathbf{u}_n)$ dans H_{div}. On a $\mathbf{u}_n\to\mathbf{u}\in L^2$, et $\nabla\cdot\mathbf{u}_n\to\mathrm{q}\in L^2$. On a

$$\int_\Omega \mathbf{u}\cdot\nabla\varphi = \lim\int_\Omega \mathbf{u}_n\cdot\nabla\varphi = -\lim\int_\Omega \varphi\,\nabla\cdot\mathbf{u}_n = -\int_\Omega \varphi\,\mathrm{q},$$

d'où l'on déduit que $\mathbf{u}$ est dans H_{div}, avec $\nabla\cdot\mathbf{u}=\mathrm{q}$. On vérifie immédiatement la convergence de $\mathbf{u}_n$ vers $\mathbf{u}$ pour la norme de H_{div}. □

Remarque 8.81. On peut identifier la trace normale d'un champ de H_{div} à un élément du dual topologique de $H^{1/2}$. On considère Ω un ouvert de frontière Γ Lipschitzienne et bornée. L'application qui à $u\in\mathcal{D}(\overline{\Omega})$ associe la restriction à Γ de la quantité $\nabla\mathbf{u}\cdot\mathbf{n}$ peut être identifiée à un élément du dual de $H^1(\Omega)$ grâce au fait que, pour toute fonction $\varphi\in\mathcal{D}(\overline{\Omega})$,

$$\int_\Gamma \varphi\mathbf{u}\cdot\mathbf{n} = \int_\Omega \varphi\nabla\cdot\mathbf{u} + \int_\Omega \mathbf{u}\cdot\nabla\varphi.$$

L'application $\varphi\mapsto\int_\Gamma\varphi\mathbf{u}\cdot\mathbf{n}$ se prolonge donc par continuité en une forme linéaire continue sur $H^1(\Omega)$, que nous noterons $\psi_{\mathbf{u}}$. Vérifions que $<\psi_{\mathbf{u}},v>$ ne dépend que de la valeur de v sur le bord. Il suffit pour cela de vérifier que H^1_0 est dans le noyau de $\Psi_{\mathbf{u}}$. Considérons donc $v\in H^1_0(\Omega)$. D'après la proposition 8.34, v s'écrit comme limite de fonctions v_n dans $\mathcal{D}(\Omega)$. On note ω_n le support de v_n. En admettant que la propriété de densité 8.19, page 95, s'étend à H_{div} c'est-à-dire qu'il existe $\mathbf{u}_n\in\mathcal{D}(\Omega)^N$ tel que

$$\|\mathbf{u}_n-\mathbf{u}\|_{L^2(\Omega)}\to 0\,,\ \|\nabla\cdot\mathbf{u}_n-\nabla\cdot\mathbf{u}\|_{L^2(\omega_n)}\to 0,$$

on obtient $\langle \Psi_{\mathbf{u}}\,, v\rangle = 0$. La forme linéaire s'annule donc sur H^1_0, et par suite elle peut être vue comme une forme linéaire sur l'espace quotient H^1/H^1_0 que nous avons défini comme $\tilde{H}^{1/2}$. Comme $\tilde{H}^{1/2}$ s'identifie à $H^{1/2}$ dans le cas d'une frontière Lipschitz (par l'isométrie $\tilde{v} \in \tilde{H}^{1/2} \mapsto \gamma_0 v$), on a bien donné un sens à $\mathbf{u}\cdot\mathbf{n}$ sur Γ en tant qu'élement du dual de $H^{1/2}(\Gamma)$. On écrira ainsi

$$\mathbf{u}\cdot\mathbf{n}|_\Gamma \in H^{-1/2}(\Gamma),$$

en prenant bien garde au fait qu'il s'agit d'une identification faite selon le procédé ci-dessus. Il est en particulier illicite d'écrire « presque partout » à côté d'une égalité identifiant deux élements de cet espace.

Considérons maintenant la formulation variationnelle

$$\int_\Omega \nabla u\cdot\nabla v = \int_\Omega fv \quad \forall v \in \mathcal{D}(\Omega).$$

Cela implique que ∇v possède une divergence faible L^2. Si l'on décide de désigner par $\triangle$ l'opérateur $\nabla\cdot\nabla$, à valeurs dans $L^2(\Omega)$, défini sur l'ensemble des champs de $H^1(\Omega)$ dont le gradient admet une divergence L^2, alors on peut écrire

$$-\triangle u = f \quad \text{p.p.}$$

D'après la remarque qui précède, on peut aussi donner un sens à la trace normale du gradient $\partial u/\partial n$, non pas en tant que fonction, mais en temps que forme linéaire sur l'espace $H^{1/2}(\Gamma)$ des traces des fonctions de H^1.

8.10. Inégalité de Poincaré sur domaines étroits

Proposition 8.82. Soit Ω un domaine borné du plan de frontière Lipschitz, et ω un domaine fortement inclus dans Ω. On suppose que $\gamma = \partial\omega$ est une courbe fermée de régularité C^2. Pour $\varepsilon > 0$, on note

$$\omega_\varepsilon = \{x \in \Omega\,,\ \mathrm{dist}(x,\gamma) < \varepsilon\}\,.$$

Alors il existe une constante $C > 0$ telle que

$$|u|_{L^2(\omega_\varepsilon)} \le C\sqrt{\varepsilon}\,\|u\|_{H^1(\Omega)} \quad \forall u \in H^1(\Omega).$$

Démonstration : Pour ε assez petit, tout point x de ω_ε peut se représenter de façon unique à l'aide du couple (s,η), où s est l'abscisse curviligne le long de γ de la projection de x sur γ, et η la distance signée à γ. On a

$$x = \xi(s) + \eta\mathbf{n}\,,\ dx = \mathbf{t}ds + \mathbf{n}d\eta + \eta\frac{\mathbf{n}}{R},$$

où R est le rayon de courbure de γ au point d'abscisse curviligne s. Le jacobien de la transformation $(s,\eta) \mapsto x$ s'écrit

$$J(s,\eta) = \begin{vmatrix} 1+\dfrac{\eta}{R} & 0 \\ 0 & 1 \end{vmatrix}.$$

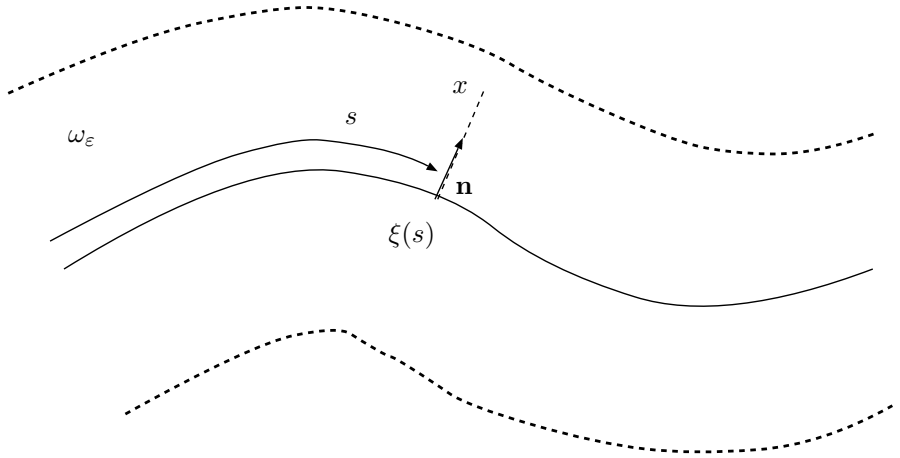


FIGURE 2. Paramétrisation de ω_ε

Ce jacobien est compris entre 1/2 et 3/2 sur ω_ε dès que ε est plus petit que la moitié du minimum du rayon de courbure sur γ, ce que nous supposerons désormais. Soit maintenant une fonction v régulière (au moins C^1 sur Ω). On écrit

$$\begin{aligned} v(s,\eta)^2 &= \left(v(s,0) + \int_0^\eta \nabla v \cdot \mathbf{n}(s)\right)^2 \\ &\leq 2v(s,0)^2 + 8\varepsilon \int_{-\varepsilon}^{\varepsilon} |\nabla u|^2 \end{aligned}$$

et ainsi

$$\begin{aligned} \int_{\omega_\varepsilon} v^2 &= \int_0^\ell \int_{-\varepsilon}^{\varepsilon} v^2 J(s,\eta) d\eta\, ds \leq \frac{3}{2} \int_0^\ell \int_{-\varepsilon}^{\varepsilon} v^2\, d\eta\, ds \\ &\leq 6\varepsilon \int_0^\ell v^2\, ds + 12\varepsilon \int_0^\ell \int_{-\varepsilon}^{\varepsilon} \left(\int_{-\varepsilon}^{\varepsilon} |\nabla u|^2\right) d\eta\, ds \\ &\leq 6\varepsilon \int_\gamma v^2 + 24\varepsilon^2 \int_0^\ell \int_{-\varepsilon}^{\varepsilon} |\nabla u|^2\, d\eta\, ds \\ &\leq 6\varepsilon \int_\gamma v^2 + 48\varepsilon^2 \underbrace{\int_0^\ell \int_{-\varepsilon}^{\varepsilon} |\nabla u|^2 J(s,\eta)\, d\eta\, ds}_{=\int_{\omega_\varepsilon} |\nabla u|^2} \\ &\leq C\varepsilon \int_\Omega |\nabla u|^2 , \end{aligned}$$

d'où l'on déduit la majoration annoncée, en raisonnant par densité. □

Proposition 8.83. Sous les mêmes hypothèses que précédemment, on a

$$|u|_{L^2(\omega_\varepsilon)} \leq C\varepsilon \, \|\nabla u\|_{L^2(\Omega)} \quad \forall u \in H^1(\Omega)\,,\ u_{|\gamma} = 0.$$

DÉMONSTRATION : La démonstration est parfaitement analogue à ce qui précède, si ce n'est que le terme d'intégrale sur γ s'annule ici, ce qui conduit à une majoration en $\mathcal{O}(\varepsilon)$. □

Chapitre 9

Minimisation quadratique sous contrainte affine

9.1. Cadre abstrait

On désigne par V un espace de Hilbert muni du produit scalaire $(\cdot,\cdot)$ et de la norme $|\cdot|$ associée. On se donne $a(\cdot,\cdot)$ une forme bilinéaire symétrique continue et coercive, de constante de coercivité α (voir définition 6.20, page 71), φ un élément de V', et l'on définit la fonctionnelle

$$J : v \in V \longmapsto J(v) = \frac{1}{2}a(v,v) - \langle \varphi\,,\, v\rangle.$$

On s'intéresse au problème de minimisation sous contrainte

$$(\mathcal{P}) \quad \begin{cases} u \in K, \\ J(u) = \inf\limits_{v\in K} J(v), \end{cases} \tag{9.1}$$

où K est un ensemble convexe fermé de V. Dans la suite on notera A l'opérateur associé à la forme bilinéaire a, vu comme un opérateur de V dans V' (voir remarque 6.24, page 72), défini par

$$\langle Au\,,\, v\rangle = a(u,v) \quad \forall v \in V.$$

Nous nous limiterons ici au cas où K est un espace affine fermé :

$$K = u_0 + K_0,$$

où u_0 est un élément de V et K_0 est un sous-espace vectoriel fermé de V. En premier lieu, nous énonçons une extension immédiate du théorème de Lax-Milgram qui assure l'existence et l'unicité d'une solution au problème ci-dessus.

Proposition 9.1. Le problème $(\mathcal{P})$ admet une solution unique $u \in K$, caractérisée par

$$a(u,v) = \langle \varphi\,,\, v\rangle \quad \forall v \in K_0. \tag{9.2}$$

On a de plus la majoration

$$|u| \leq |u_0| + \frac{1}{\alpha}\left(\|a\|\,|u_0| + \|\varphi\|\right).$$

Démonstration : On peut se ramener au problème de minimisation de la fonctionnelle

$$\tilde{u} \longmapsto \tilde{J}(\tilde{u}) = J(u_0 + \tilde{u}),$$

sur le sous-espace vectoriel fermé K_0, pour lequel le théorème de Lax-Milgram assure l'existence et l'unicité d'une solution, ainsi que l'équivalence avec la formulation variationnelle

$$a(\tilde{u},v) = \langle \varphi\,,\, v\rangle - a(u_0,v) \quad \forall v \in K_0.$$

Prenant $v = \tilde{u}$ dans ce qui précède, on en déduit

$$\alpha\,|\tilde{u}|^2 \leq \|\varphi\|\,|\tilde{u}| + \|a\|\,|u_0|\,|\tilde{u}|$$

d'où l'on déduit la majoration sur $|u|$. □

Dans toute la suite u désigne la solution unique du problème $(\mathcal{P})$.

Nous terminons cette introduction par une propriété à la fois élémentaire et essentielle à la compréhension des différentes formulations qui vont être présentées ici.

Proposition 9.2. On a

$$a(u,v) + \langle \xi \, , v \rangle = \langle \varphi \, , v \rangle \quad \forall v \in V, \tag{9.3}$$

avec $\xi \in K_0^{\perp}$.

Démonstration : C'est une simple réécriture de la formulation variationnelle « interne » (9.2) : la forme linéaire ξ définie par

$$v \longmapsto \langle \varphi \, , v \rangle - a(u,v)$$

s'annule contre tout élément de K_0. □

Remarque 9.3. Cette forme linéaire ξ aura en général une signification physique claire dans les applications. Ainsi, dans le cas où le problème considéré traduit un équilibre des forces au sein d'un milieu (la formulation variationnelle (9.2) exprime alors ce qui s'appelle le principe des puissances virtuelles), ξ pourra être vu comme un champ de force permettant d'assurer la contrainte, ou plus précisément, si v désigne un champ de déformation élémentaire du milieu, $\langle \xi \, , v \rangle$ désigne le travail élémentaire des forces qui assurent la contrainte. Noter que pour tout déplacement élémentaire respectant la contrainte (c'est-à-dire $v \in K_0$), alors ce travail est nul. Dans le cas de l'équation de la chaleur, alors ξ représente un flux de chaleur par unité de volume, qu'il faut fournir au système pour assurer que la contrainte soit réalisée (voir section 3.3, page 29).

Remarque 9.4. La forme linéaire ξ de la proposition précédente est unique. On distinguera bien ce fait de l'éventuelle unicité de ce que nous définissons dans la section **??** comme un multiplicateur de Lagrange. La non unicité de ce multiplicateur de Lagrange tient au fait que nous chercherons à écrire ξ sous la forme $B^{\star}\lambda$, où B est un opérateur pas toujours surjectif ni même à image dense (de telle sorte que $B^{\star}$ peut ne pas être injectif).

Précisons maintenant la manière la plus directe de formuler le problème, sur laquelle se baseront les approches numériques dites directes, basées sur la discrétisation de l'espace des degrés de libertés réels [1].

Approche directe. On suppose pour alléger l'écriture que $u_0 = 0$, de telle sorte que K est un sous-espace vectoriel de V. Cette première approche consiste à reformuler le problème dans l'espace K. Dans certains cas , pour la résolution d'un problème elliptique avec conditions de Dirichlet homogènes à la frontière d'un obstacle par exemple, cette reformulation est parfaitement naturelle. Dans l'exemple proposé, travailler dans l'espace K consistera simplement à écrire le problème de minimisation sur l'espace des fonctions qui vérifient la condition au bord de l'obstacle, ce qui nécessitera simplement de discrétiser en espace sur des maillages dont la frontière approche le bord de l'obstacle.

1. Pour des problèmes avec contrainte localisée sur un sous-domaine, il s'agit de l'approche consistant à utiliser un maillage qui suit la frontière du sous-domaine.

Nous proposons ici une formulation abstraite qui s'appliquera à des cas plus généraux. On suppose que K se met sous la forme [2]

$$K = \tilde{K} + \tilde{T}Z = \left\{ u + \tilde{T}z\,,\ u \in \tilde{K}\,,\ z \in Z \right\},$$

où $\tilde{K}$ est un sous-espace vectoriel fermé de V, Z un espace de Hilbert, et $\tilde{T} \in \mathcal{L}(Z, V)$. On appelle T l'application de $\tilde{K} \times Z$ dans K qui à (u, z) associe $u + \tilde{T}z$, et l'on suppose que T est bijective. Nous écrivons le terme linéaire $\langle \varphi\,, v \rangle$ sous la forme d'un produit scalaire (f, v). On peut alors reformuler le problème dans $\tilde{K} \times Z$. On cherche ainsi $U = (u, z) \in \tilde{K} \times Z$ qui minimise la fonctionnelle

$$\frac{1}{2} a(TU, TU) - (f, TU).$$

La formulation variationnelle de ce problème est

$$a(TU, TV) = (ATU, TV) = (f, TV) \quad \forall V \in \tilde{K} \times Z,$$

d'où

$$(T^\star ATU, V) = (T^\star f, V) \quad \forall V \in \tilde{K} \times Z,$$

qui peut donc s'écrire

$$T^\star ATU = T^\star \varphi.$$

Comme il apparaîtra dans les applications, on retrouve au niveau matricel un système du type de celui qui précède. L'avantage est qu'il ne sera pas nécessaire d'assembler la matrice associée au problème contraint, mais une matrice associée à un problème sans contrainte (correspondant à l'opérateur A).

9.2. Formulation point-selle

On suppose ici que l'ensemble admissible s'écrit $K = B^{-1}(\{z\})$, où B est une application linéaire continue de H dans un espace de Hilbert Λ, et $z = Bu_0$ un élément de son image. On notera $K_0 = \ker B$ l'espace vectoriel associé.

L'approche que nous allons mener consiste à réécrire la contrainte de façon duale :

$$u \in K \Longleftrightarrow (\lambda, Bu - z) = 0 \quad \forall \lambda \in \Lambda,$$

et à remplacer le problème de minimisation initial par un problème de recherche de point-selle pour une nouvelle fonctionnelle $L(\cdot, \cdot)$, appelée Lagrangien, définie sur l'espace produit $V \times \Lambda'$. Il peut sembler étonnant de remplacer un problème de minimisation par ce nouveau problème, *a priori* plus compliqué ; le point essentiel est que le nouveau problème n'est plus contraint.

2. L'intérêt (pour le moins peu manifeste) d'une telle approche apparaîtra clairement dans les applications. Pour anticiper sur la résolution effective de tels problèmes, disons simplement ici que $\tilde{K}$ est un sous-espace de K que l'on peut approcher simplement par des sous-espaces de dimensions finies. Cette approche est justifiée quand K ne peut pas lui-même être aisément approché selon les méthodes usuelles.

Nous considérerons ainsi dans cette section le jeu d'hypothèses suivant

$$\left.\begin{array}{l} V \text{ et } \Lambda \text{ espaces de Hilbert,} \\ a(\cdot,\cdot) \text{ bilinéaire symétrique coercive continue sur } V\times V\,,\ \varphi\in V' \\ B\in\mathcal{L}\,(V,\Lambda), \\ u_0\in V\,,\ K=\{u\in V\,,\ Bu=z=Bu_0\} \\ J(v)=\dfrac{1}{2}a(v,v)-\langle\varphi\,,v\rangle,\quad u=\arg\min\limits_K J, \\ L\ :\ (v,\mu)\in V\times\Lambda\longmapsto L(v,\mu)=J(v)+(\mu,Bv-z). \end{array}\right\} \tag{9.4}$$

On appelera comme précédemment $\mathcal{P}$ le problème de minimisation sur K dont u est solution, et $\mathcal{P}'$ le problème suivant :

De façon à disposer d'un cadre théorique directement applicable à la plupart des situations que nous allons rencontrer, on fera systématiquement dans ce qui suit l'identification entre Λ et son dual, mais l'on n'identifiera par V et son dual V' de telle sorte que $B^\star$ est un opérateur de Λ dans V'.

Définition 9.5. (Point-selle)
On appelle point-selle de L tout couple (u,λ) solution du problème $\mathcal{P}'$ suivant

$$(\mathcal{P}')\quad\left\{\begin{array}{l}(u,\lambda)\in V\times\Lambda,\\ L(u,\mu)\le L(u,\lambda)\le L(v,\lambda)\quad\forall v\in V\,,\ \forall\mu\in\Lambda.\end{array}\right.$$

On appellera u la composante primale du point-selle (u,λ), et λ sa composante duale.

On verra plus loin que l'existence d'un point-selle n'est pas garantie en général. En revanche si un point-selle existe, alors sa composante primale est solution du problème de minimisation sous contrainte initial, comme l'exprime la

Proposition 9.6. Si $(u,\lambda)\in V\times\Lambda$ est solution de $(\mathcal{P}')$, alors u est solution de $(\mathcal{P})$.

Démonstration : Soit $(u,\lambda)\in V\times\Lambda$ une solution de $(\mathcal{P}')$. S'il existe $\mu_0\in\Lambda$ tel que

$$(\mu_0,Bu-z)\neq 0,$$

alors (la droite $\mathbb{R}\mu_0$ est dans Λ)

$$\sup_{\mu\in\Lambda'}L(u,\mu)=+\infty,$$

ce qui est contraire aux hypothèses. On a donc $(\mu_0,Bu-z)=0$ pour tout $\mu_0\in\Lambda$, d'où $Bu=z$ et donc $u\in K$. D'autre part pour tout v dans K, on a $Bv=z$, d'où

$$J(v)=L(v,\lambda)\ge L(u,\lambda)=J(u),$$

d'où l'on déduit que $u\in K$ est bien la solution de $\mathcal{P}$. □

Nous utiliserons dans la suite une autre formulation du problème $\mathcal{P}'$, très proche formellement du système matriciel qui résultera de la discrétisation en espace de tels problèmes, qui fait intervenir l'adjoint de B défini par

$$\langle B^\star\mu\,,v\rangle=(\mu,Bv)\quad\forall v\in V\,,\ \mu\in\Lambda.$$

Proposition 9.7. On se place dans le cadre des hypothèses (9.4). Le couple (u, λ) est point-selle de L si et seulement s'il est solution du problème $(\mathcal{P}'')$ suivant

$$(\mathcal{P}'') \qquad \left\{ \begin{array}{lcl} Au + B^\star\lambda & = & \varphi \\ Bu & = & z. \end{array} \right. \tag{9.5}$$

Démonstration : Soit (u, λ) solution du problème $\mathcal{P}'$. On a alors, d'après la proposition précédente, $Bu = z$. D'autre part, comme

$$L(u, \lambda) \leq L(v, \lambda) \quad \forall v \in V,$$

on a, d'après le théorème de Lax-Milgram,

$$a(u, v) = \langle B^\star\lambda - \varphi \, , v\rangle \quad \forall v \in V,$$

c'est-à-dire

$$Au + B^\star\lambda = \varphi.$$

Réciproquement supposons le couple (u, λ) solution de $\mathcal{P}''$. Comme $Bu = z$ on a tout de suite

$$L(u, \mu) = J(u) = L(u, \lambda) \quad \forall \mu \in \Lambda.$$

D'autre part, d'après le théorème de Lax-Milgram, $Au + B^\star\lambda = \varphi$ implique que u minimise sur V la fonctionnelle

$$J(v) + (v, B^\star\mu) - \langle \mu \, , z\rangle,$$

d'où la seconde inégalité caractérisant le point-selle. □

La continuité de B ne suffit pas à assurer l'existence d'un multiplicateur de Lagrange associé à la contrainte. L'équivalence entre les deux formulations nécessite l'ajout d'une condition supplémentaire sur B.

Théorème 9.8. On se place dans le cadre des hypothèses (9.4). On suppose de plus que B est à image fermée dans Λ. Alors il existe $\lambda \in \Lambda$ tel que (u, λ) soit solution de $(\mathcal{P}')$, où u est la solution du problème $\mathcal{P}$.

Démonstration : Remarquons tout d'abord que, d'après la formulation variationnelle du problème de minimisation $\mathcal{P}$ écrit dans l'espace admissible K, $\varphi - Au$ est dans K_0°. Comme B est à image fermée dans Λ, l'image de $B^\star$ s'identifie à l'orthogonal du noyau de B (voir proposition 7.18, page 83), c'est-à-dire K_0°. Il existe donc $\lambda \in \Lambda'$ que tel $B^\star\lambda = \varphi - Au$. □

On notera que le théorème précédent assure l'existence d'un multiplicateur de Lagrange, mais pas son unicité. On a unicité du λ au prix d'une hypothèse supplémentaire sur B :

Théorème 9.9. On se place dans le cadre des hypothèses (9.4). Si B est surjectif, alors le problème $\mathcal{P}'$ admet une unique solution $(u, \lambda) \in V \times \Lambda'$.

Démonstration : Si B est surjectif, alors $B^\star$ est injectif[3], d'où l'unicité du λ. □

3. C'est une conséquence directe de la proposition 7.20, page 84. Mais on peut aussi le démontrer directement en une ligne, sans utiliser la proposition 7.20.

Dans le cas où B est surjective (ou, de façon équivalente lorsque la condition inf-sup est satisfaite), on a une estimation sur la norme du multiplicateur de Lagrange.

Proposition 9.10. On se place dans le cadre des hypothèses (9.4). On suppose que B est surjective, et l'on note (u, λ) la solution du problème de point-selle $\mathcal{P}'$. On a l'estimation

$$|\lambda| \leq \frac{1}{\beta}\left(1 + \frac{\|a\|}{\alpha}\right)\|\varphi\| + \frac{1}{\beta}\frac{C^2}{\alpha}|u_0|\,.$$

Démonstration : On a, d'après la proposition 7.20, page 84,

$$|\lambda| \leq \frac{1}{\beta}\sup_v \frac{|(\lambda, Bv)|}{|v|} \leq \frac{1}{\beta}\left(\|a\|\,|u| + \|\varphi\|\,,\right)$$

d'où l'on conclut grâce à la proposition 9.1. □

Proposition 9.11. On suppose $f = 0$. On a

$$|u| \leq \frac{\|a\|}{\beta\alpha}|z|\,.$$

Démonstration : La démarche menée au début de la démonstration précédente donne

$$|\lambda| \leq \frac{1}{\beta}\|B^\star\lambda\| \leq \frac{\|a\|}{\beta}|u|\,.$$

On a donc

$$\alpha|u|^2 \leq a(u,u) = |(\lambda, Bu)| \leq |\lambda|\,|z| \leq \frac{\|a\|}{\beta}|u|\,|z|\,,$$

d'où l'estimation sur $|u|$. □

Dans le cas le plus général (B est simplement supposée linéaire continue), on peut résoudre de façon approchée le problème $\mathcal{P}''$ au sens suivant :

Proposition 9.12. On suppose B linéaire continue (non nécessairement à image fermée). Pour tout $\varepsilon > 0$, il existe $\lambda_\varepsilon \in \Lambda$ tel que

$$Au + B^\star\lambda_\varepsilon = \varphi + h_\varepsilon,$$

avec $\|h_\varepsilon\| < \varepsilon$.

Démonstration : On note $\xi = \varphi - Au$, de telle sorte que

$$Au + \xi = \varphi.$$

D'après la formulation variationnelle du problème initial $\mathcal{P}$, l'élément g est dans K_0°. Or K_0° est exactement l'adhérence de l'image de $B^\star$ (voir proposition 7.18, page 83). Il existe donc λ_ε dans Λ tel que $\|\xi - B^\star\lambda_\varepsilon\| < \varepsilon$. □

Remarque 9.13. On peut ainsi construire une suite (λ_ε) dont l'image par $B^\star$ converge vers g. Mais en général la suite (λ_ε) n'est pas bornée dans Λ.

Remarque 9.14. Précisons qu'il existe toujours une formulation point-selle bien posée (avec existence et unicité du multiplicateur de Lagrange) du problème de minimisation initial. On peut voir en effet la formulation variationnelle de la proposition9.2, qui définit ξ, comme une formulation point-selle particulière, basée sur l'application B définie comme la projection sur K_0°. Cette application étant surjective, on a bien existence et unicité d'un multiplicateur de Lagrange $\lambda \in (K_0^\circ)$On a donc

$$a(u,v) + (\lambda, P_{K_0^\circ} v) = \langle \varphi \, , v \rangle,$$

d'où l'existence et l'unicité du ξ définit par

$$\langle \xi \, , v \rangle = (z, P_{K_0^\circ} v) = (z, v).$$

Méthode du Lagrangien augmenté. Nous introduisons ici une nouvelle formulation du type point-selle, qui rentre parfaitement dans le cadre théorique exposé dans la section précédente, mais qui peut permettre d'améliorer les propriétés de convergence des algorithmes numériques. On considère le Lagrangien suivant,

$$\begin{array}{rcl} L_r \, : \, V \times \Lambda & \longrightarrow & \mathbb{R} \\ (v, \mu) & \longmapsto & L_r(v,\mu) = J(v) + \dfrac{r}{2} \left| Bv \right|^2 + (\mu, Bv), \end{array}$$

où r est une constante positive. Précisons tout de suite que, malgré une certaine analogie formelle avec la méthode de pénalisation, le paramètre r n'est pas destiné à tendre vers $+\infty$. On s'intéresse au problème de la recherche d'un point selle pour L_r. Cela revient à appliquer la démarche de la section précédente à la minimisation de la fonctionnelle

$$v \longmapsto J_r(v) = J(v) + \frac{r}{2} \left| Bv \right|^2 ,$$

sur l'espace admissible $K = \{ v \in V \, , \ Bv = 0 \}$. Comme J_r s'identifie à J sur V, le problème de minimisation est inchangé. De plus cette nouvelle fonctionnelle vérifie les mêmes conditions que J. La constante de coercivité est inchangée (tout du moins elle n'est pas inférieure à celle de J), et la constante de continuité peut augmenter mais reste finie. Tout ce qui a été établi à la question précédente reste donc valable pour cette nouvelle formulation. On se reportera à [**10**] pour une analyse détaillée de cette méthode.

Algorithme d'Uzawa. L'algorithme d'Uzawa est destiné à résoudre effectivement les problèmes de point-selle. Il est en général présenté dans la littérature sous forme matricielle (*i.e.* pour des problèmes de dimension finie). Nous montrons ici qu'il a un sens en dimension infinie et que, sous la simple hypothèse d'existence d'un point-selle, il converge.

L'algorithme d'Uzawa consiste en la construction d'une suite dans l'espace Λ selon le procédé suivant : on se donne ρ un paramètre strictement positif, λ^0 un élément de Λ (on prendra en général $\lambda^0 = 0$), et l'on construit (λ^k, u^k) selon la procédure

$$\begin{array}{rcl} u^{k+1} & = & A^{-1} \left(\varphi - B^\star \lambda^k \right) \\ \lambda^{k+1} & = & \lambda^k + \rho B u^{k+1}. \end{array}$$

On notera que u^k n'intervient pas dans la relation de récurrence. La composante primale du couple (u, λ) apparaît ainsi simplement comme une variable auxilliaire. On peut écrire l'algorithme $\lambda^{k+1} = T_\rho \lambda^k$, où T_ρ est l'application de Λ dans lui-même définie par

$$T_\rho = \mathrm{Id} + \rho B A^{-1} \left(\varphi - B^\star \right).$$

Remarque 9.15. L'algorithme d'Uzawa est en fait un algorithme de gradient à pas fixe pour la fonctionnelle

$$\mu \longmapsto \tilde{J}(\mu) = \frac{1}{2}\big\langle \varphi - B^\star\mu \, , A^{-1}(\varphi - B^\star\mu)\big\rangle,$$

dont le gradient s'écrit

$$\nabla\tilde{J}(\mu) = -BA^{-1}\left(\varphi - B^\star\mu\right).$$

On notera que, dans le cas où B n'est pas à image fermée, cette fonctionnelle n'est pas coercive, ni même en général sa restriction à l'orthogonal du noyau de $B^\star$. Elle est en revanche minorée (par 0), donc elle admet un infimum (qui n'est pas nécessairement atteint).

Remarque 9.16. L'algorithme d'Uzawa peut être vu comme une discrétisation explicite en temps de l'équation d'évolution

$$\frac{d\lambda}{dt} = BA^{-1}\left(f - B^\star\lambda\right) = -\nabla\tilde{J}(\lambda).$$

On parle de *flot-gradient*. Le paramètre ρ joue ainsi le rôle d'un pas de temps.

Proposition 9.17. On suppose $\rho > 0$. Les assertions suivantes sont équivalentes :
(i) λ est point fixe de T_ρ, *i.e.* $T_\rho\lambda = \lambda$.
(ii) Le couple (u, λ), avec $u = A^{-1}\left(\varphi - B^\star\lambda\right)$, est solution de $\mathcal{P}''$.

Démonstration : λ est point fixe de T_ρ si et seulement si

$$BA^{-1}\left(f - B^\star\lambda\right) = 0,$$

ce qui est équivalent à : $u = A^{-1}\left(\varphi - B^\star\lambda\right)$ vérifie $Bu = 0$ et $Au + B^\star\lambda = \varphi$. □

Proposition 9.18. On suppose que le Lagrangien L admet un point-selle (u, λ). Alors la suite $u^k = A^{-1}(f - B^\star\lambda^{k-1})$ converge vers u, solution du problème de minimisation $(\mathcal{P})$, dès que

$$0 < \rho < \frac{2\alpha}{\|B\|^2}, \tag{9.6}$$

où α est la constante de coercivité de $a(\cdot,\cdot)$.

Démonstration : On a

$$\begin{aligned} \lambda^{k+1} &= \lambda^k + \rho B u^k \\ \lambda &= \lambda + \rho B u \end{aligned}$$

d'où

$$\lambda^{k+1} - \lambda = \lambda^k - \lambda + \rho B(u^{k+1} - u).$$

On en déduit

$$\begin{aligned} \left|\lambda^{k+1} - \lambda\right|^2 &= \left|\lambda^k - \lambda\right|^2 + 2\rho\left(u^{k+1} - u, B^\star(\lambda^k - \lambda)\right) + \rho^2\left|B(u^{k+1} - u)\right|^2 \\ &= \left|\lambda^k - \lambda\right|^2 - 2\rho\left(u^{k+1} - u, A(u^{k+1} - u)\right) + \rho^2\left|B(u^{k+1} - u)\right|^2 \\ &\leq \left|\lambda^k - \lambda\right|^2 - \rho\left(2\alpha - \rho\|B\|^2\right)\left|u^{k+1} - u\right|^2. \end{aligned}$$

Si la condition sur ρ est vérifiée, alors la suite $\left|\lambda^k - \lambda\right|$ est décroissante positive, donc converge, et par suite u^k converge vers u. □

Convergence de la suite des multiplicateurs de Lagrange (Uzawa). Nous démontrons ici une propriété peu documentée dans la littérature, qui est que la suite des multiplicateurs de Lagrange construite par l'algorithme d'Uzawa converge faiblement, même dans le cas de non-unicité du point-selle. La démonstration est basée sur la proposition suivante (voir par exemple [**12**]) :

Proposition 9.19. (Lemme d'Opial)
Soit Λ un espace de Hilbert, $\tilde{\Lambda}$ un sous-ensemble non vide de Λ, et (λ^k) une suite d'éléments de Λ telle que

(i) pour tout $\mu \in \tilde{\Lambda}$, la suite $\left|\lambda^k - \mu\right|$ converge,

(ii) si une sous-suite $(\lambda^{\varphi(k)})$ converge faiblement vers un élément μ de Λ, alors $\mu \in \tilde{\Lambda}$.

Alors la suite (λ^k) converge faiblement vers un élément de $\tilde{\Lambda}$.

Démonstration : D'après (i), la suite (λ^k) est bornée. Il suffit donc de vérifier que deux sous-suites qui convergent faiblement ont la même limite. On considère donc deux sous-suites (λ^{m_k}) et (λ^{n_k}) qui convergent faiblement vers λ_1 et λ_2, respectivement. On introduit les limites

$$\ell_1 = \lim \left|\lambda^k - \lambda_1\right| \ , \ \ell_2 = \lim \left|\lambda^k - \lambda_2\right|$$

qui sont bien définies par hypothèse. On écrit simplement

$$\left|\lambda^k - \lambda_1\right|^2 - \left|\lambda^k - \lambda_2\right|^2 = \left(\lambda_2 - \lambda_1, 2\lambda^k - \lambda_1 - \lambda_2\right)$$

On passe à la limite dans l'identité précédente pour la sous-suite (λ^{m_k}), puis pour (λ^{n_k}). Il vient

$$|\ell_1|^2 - |\ell_2|^2 = -\left|\lambda_2 - \lambda_1\right|^2 \text{ et } |\ell_1|^2 - |\ell_2|^2 = \left|\lambda_2 - \lambda_1\right|^2 .$$

On a donc nécessairement $|\lambda_2 - \lambda_1| = 0$, d'où le résultat. □

Proposition 9.20. On suppose que le Lagrangien L admet un point-selle (u, λ) (non nécessairement unique). Alors, sous l'hypothèse (9.6), la suite (λ^k) converge faiblement vers $\mu \in \Lambda$, tel que (u, μ) est point-selle pour L.

Démonstration : On note $\tilde{\Lambda} \subset \Lambda$ l'ensemble des μ tels que (u, μ) est point-selle pour L (ou de façon équivalente solution du problème $\mathcal{P}''$), et l'on se propose de vérifier que la suite (λ^k) rentre dans le cadre du lemme d'Opial. L'hypothèse (i) est vérifiée, comme on l'a vu lors de la démonstration de la proposition 9.18. Considérons maintenant une sous-suite, que nous notons encore (λ^k) pour alléger l'écriture, qui converge faiblement vers $\mu \in \Lambda$. On a

$$Au^k + B^\star \lambda^k = f$$

pour tout k. Or Au^k converge vers Au, et $B^\star \lambda^k$ converge faiblement vers $B^\star \mu$ (d'après la proposition 6.31, page 74). On a donc par passage à la limite (faible)

$$Au + B^\star \mu = f,$$

et ainsi (u, μ) est point-selle de L c'est-à-dire $\mu \in \tilde{\Lambda}$. Le lemme d'Opial ci-dessus permet de conclure. □

Remarque 9.21. Dans le cas de la dimension finie (donc notamment lors de la résolution numérique de problèmes discrétisés en espace), on aura donc convergence forte de la suite des multiplicateurs de Lagrange. On distinguera bien cette propriété de convergence pour le problème discrétisé (à paramètre h de discrétisation fixé), de la propriété de convergence éventuelle de la suite des limites vers "quelque chose" quand le paramètre de discrétisation h tend vers 0.

9.3. Pénalisation

Le principe de la méthode de pénalisation consiste à remplacer le problème sous contrainte par un problème de minimisation sans contrainte, en rajoutant à la fonctionnelle de départ un terme dit de pénalisation.

On considère le jeu d'hypothèses suivant

$$\left|\begin{array}{l} V \text{espace de Hilbert,} \\ a(\cdot,\cdot) \text{ bilinéaire symétrique coercive continue, } \varphi \in V' \\ b(\cdot,\cdot) \text{ bilinéaire symétrique continue positive } (i.e.\ b(v,v) \geq 0 \quad \forall v \in V), \\ K = u_0 + K_0 = u_0 + \{u \in V\ ,\ b(u,u) = 0\} = u_0 + \ker b, \\ J(v) = \dfrac{1}{2}a(v,v) - \langle \varphi\ , v\rangle, \quad u = \arg\min_K J, \\ J_\varepsilon(v) = \dfrac{1}{2}a(v,v) + \dfrac{1}{2\varepsilon}b(v-u_0, v-u_0) - \langle \varphi\ , v\rangle, \quad u^\varepsilon = \arg\min_V J. \end{array}\right. \tag{9.7}$$

On appelera $\mathcal{P}$ (resp. $\mathcal{P}_\varepsilon$) le problème de minimisation sur K dont u est solution (resp. sur V dont u^ε est solution).

Théorème 9.22. Sous les hypothèses (11.1), la suite (u^ε) converge vers u quand ε tend vers 0.

Démonstration : On écrit la formulation variationnelle associée au problème pénalisé :

$$a(u^\varepsilon, v) + \frac{1}{\varepsilon}b(u^\varepsilon - u_0, v) = \langle \varphi\ , v\rangle \quad \forall v \in V.$$

Prenant $v = u^\varepsilon - u_0$ dans ce qui précède, et utilisant la positivité de la forme bilinéaire b, on obtient

$$a(u^\varepsilon, u^\varepsilon) \leq C_0 + C_1 |u^\varepsilon| \Longrightarrow \alpha |u^\varepsilon|^2 \leq C_0 + C_1 |u^\varepsilon|$$

d'où l'on déduit que $|u^\varepsilon|$ est majoré. La suite (u^ε) étant bornée, on peut en extraire une sous-suite, que l'on notera toujours (u^ε), qui converge faiblement vers $z \in V$. Or, du fait que $J_\varepsilon \geq J$ et que la solution du problème initial u vérifie exactement la contrainte (*i.e.* $b(u-u_0, u-u_0) = 0$), on a, pour tout $\varepsilon > 0$,

$$J(u^\varepsilon) \leq J_\varepsilon(u^\varepsilon) \leq J_\varepsilon(u) = J(u), \tag{9.8}$$

et ainsi, d'après le théorème 6.39, page 77,

$$J(z) \leq \liminf J(u^\varepsilon) \leq J(u).$$

Il suffit donc de vérifier que la limite z vérifie la contrainte pour l'identifier à l'unique solution u du problème $\mathcal{P}$. Pour cela on montre tout d'abord que $b(u^\varepsilon - u_0, u^\varepsilon - u_0)$ tend vers 0. En effet, la majoration $J_\varepsilon(u^\varepsilon) \leq J(u)$ s'écrit

$$J(u^\varepsilon) + \frac{1}{2\varepsilon} b(u^\varepsilon - u_0, u^\varepsilon - u_0) \leq J(u),$$

d'où $b(u^\varepsilon - u_0, u^\varepsilon - u_0)/\varepsilon$ est borné, et donc $b(u^\varepsilon - u_0, u^\varepsilon - u_0)$ tend vers 0 quand ε tend vers 0. On utilise alors une nouvelle fois le théorème 6.39, qui assure que

$$0 \leq b(z,z) \leq \liminf b(u^\varepsilon - u_0, u^\varepsilon - u_0) = 0,$$

d'où $z \in K$, et par suite $z = u$. Cette démonstration pouvant s'effectuer pour n'importe quelle sous-suite extraite, on a bien convergence faible de l'ensemble de la suite (u^ε) vers u.

Pour montrer que la convergence est forte, il suffit de montrer que l'on a convergence de u^ε vers u_0 pour la norme associée au produit scalaire défini par a :

$$(v,w)_a = a(v,w)\,,\ |v|_a = \sqrt{a(v,v)},$$

car cette norme Hilbertienne est équivalente à la norme de départ du fait de la continuité et la coercivité de a. On a tout d'abord $a(u^\varepsilon, v) \longrightarrow a(u,v)$, car $a(\cdot, v)$ est dans V', pour tout $v \in V$. D'autre part $|u|_a \leq \liminf |u^\varepsilon|_a$, et enfin, d'après (9.8),

$$\frac{1}{2} a(u^\varepsilon, u^\varepsilon) - \langle \varphi\,, u^\varepsilon \rangle \leq \frac{1}{2} a(u,u) - \langle \varphi\,, u \rangle,$$

d'où $\limsup |u^\varepsilon|_a \leq |u|_a$. On a donc convergence faible de u^ε vers u dans V muni de $(\cdot,\cdot)_a$, et convergence des normes $|u^\varepsilon|_a$ vers la norme de la limite, d'où, d'après la proposition 6.30, u^ε converge fortement vers u pour $|\cdot|_a$, et donc pour la norme initiale. □

Remarque 9.23. On peut préciser le comportement asymptotique du terme $b(u^\varepsilon, v)/\varepsilon$ de la formulation variationnelle pénalisée. On note ξ la forme linéaire définie par la proposition 9.2. On a

$$a(u,v) + \langle \xi\,, v \rangle = \langle \varphi\,, v \rangle \quad \forall v \in V,$$

et d'autre part

$$a(u^\varepsilon, v) + \frac{1}{\varepsilon} b(u^\varepsilon - u_0, v) = \langle \varphi\,, v \rangle \quad \forall v \in V.$$

On a donc en soustrayant ces deux identités

$$\left| \frac{1}{\varepsilon} b(u^\varepsilon - u_0, v) - \langle \xi\,, v \rangle \right| \leq |a(u - u^\varepsilon, v)| \leq C\,|u - u^\varepsilon|\,|v|\,.$$

Comme u^ε converge vers u dans V, on a ainsi convergence (forte dans V') de la suite des formes linéaires

$$v \longmapsto \frac{1}{\varepsilon} b(u^\varepsilon - u_0, v)$$

vers la forme linéaire ξ.

Exercice 9.1. Étendre les résultats du théorème 9.22 au cas où $a(\cdot,\cdot)$ est simplement supposée coercive sur $K^\perp$:

$$a(v,v) \geq \alpha\,|v|^2 \quad \forall v \in K^\perp.$$

La démonstration qui précède, qui utilise un argument de compacité, ne donne aucune indication sur la vitesse de convergence de u^ε vers u. Nous énonçons ici un certain nombre de propriétés qui nous permettront dans certains cas d'estimer cette vitesse de convergence.

Le résultat que nous présentons ici, strictement plus fin que les résultats précédents (voir remarque 9.27 à la fin de cette section), s'inspire d'un travail déjà ancien de Babuška [**2**].

Proposition 9.24. On se place dans le cadre des hypothèses (11.1), page 141. On suppose qu'il existe $\tilde{\xi} \in V$ tel que

$$b(\tilde{\xi}, v) = \langle \xi \, , \, v \rangle \quad \forall v \in V,$$

où ξ est la forme linéaire définie par (9.3), page 116. On a alors

$$|u - u^\varepsilon| \leq C\varepsilon.$$

Démonstration : Notons dans un premier temps qu'il est licite de prendre $\tilde{\xi}$ dans $K^\perp$ (sinon, on le remplace par sa projection sur $K^\perp$). On introduit maintenant la nouvelle fonctionnelle

$$R_\varepsilon(v) = \frac{1}{2}a(u - v, u - v) + \frac{1}{2\varepsilon}b(\varepsilon\tilde{\xi} - v, \varepsilon\tilde{\xi} - v).$$

Montrons que minimiser J_ε revient à minimiser R_ε. On a

$$R_\varepsilon(v) = \frac{1}{2}a(u, u) + \frac{\varepsilon}{2}b(\tilde{\xi}, \tilde{\xi}) + \frac{1}{2}a(v, v) + \frac{\varepsilon}{2}b(v, v) - a(u, v,) - b(\tilde{\xi}, v).$$

Or on a par hypothèse $b(\tilde{\xi}, v) = \langle \xi \, , \, v \rangle$, d'où

$$-a(u, v) - b(\tilde{\xi}, v) = - < \varphi, v >,$$

et par suite $R_\varepsilon(v)$ est égal à $J_\varepsilon(v)$ plus une quantité qui ne dépend pas de v.

On introduit maintenant l'élément $w = \varepsilon\tilde{\xi} + u$. On a

$$R_\varepsilon(w) = \frac{\varepsilon^2}{2}a(\tilde{\xi}, \tilde{\xi}) + 0 \quad \text{car } u \in K = \ker b,$$

et ainsi $|R_\varepsilon(w)| \leq C\varepsilon^2$. Comme u^ε minimise R_ε, on a

$$0 \leq R_\varepsilon(u^\varepsilon) = \frac{1}{2}a(u - u^\varepsilon, u - u^\varepsilon) + \frac{1}{2\varepsilon}b(\varepsilon\tilde{\xi} - u^\varepsilon, \varepsilon\tilde{\xi} - u^\varepsilon) \leq C\varepsilon,$$

dont on déduit, d'après la coercivité de $a(\cdot, \cdot)$, la majoration de l'erreur en $\mathcal{O}(\varepsilon)$. □

Remarque 9.25. Si l'on note $\overline{K^\perp}^b$ le complété de $K^\perp$ pour la norme $b(\cdot, \cdot)^{1/2}$ (qui est bien une norme, en général non complète, sur $K^\perp = (\ker b)^\perp$), on a un triplet de Gelfand (voir section 13.1, page 161) en considérant

$$K^\perp \subset \overline{K^\perp}^b \subset \left(K^\perp\right)'.$$

La proposition 13.4, page 162 assure donc l'équivalence suivante, pour tout $\xi \in (K^\perp)'$

$$\exists \overline{\xi} \in \overline{K^\perp}^b \, , \; b\left(\overline{\xi}, v\right) = \langle \xi \, , \, v \rangle \quad \forall v \in V \Longleftrightarrow \exists C > 0 \, , \; |\langle \xi \, , \, v \rangle| \leq C b(v, v)^{1/2} \quad \forall v \in V. \tag{9.9}$$

On prendra garde que cette équivalence ne permet pas de remplacer l'hypothèse de la proposition précédente (existence d'un $\tilde{\xi} \in V'$) par une simple inégalité faisant intervenir $b(\cdot, \cdot)$, car le $\overline{\xi}$ de l'équivalence ci-dessus n'est pas nécessairement dans $K^\perp$, mais dans son complété.

On utilisera souvent l'estimation de la proposition 9.24 sous la forme du corollaire suivant (strictement plus faible, comme le précise la remarque 9.27 ci-après) :

Corollaire 9.26. Under assumptions (11.1), we assume in addition that $b(\cdot,\cdot)$ can be written $b(u,v) = (Bu, Bv)$, where B is a linear continuous operator onto a Hilbert space Λ, with closed range. Then $|u^\varepsilon - u| = \mathcal{O}(\varepsilon)$.

Démonstration : Let us show that the assumption of Prop. 9.24 is met. It is sufficient to prove that any $\xi \in V'$ which vanishes over K identifies through $b(\cdot,\cdot)$ to some $\tilde{\xi} \in V$, i.e. there exists $\tilde{\xi} \in V$ such that

$$\langle \xi \, , v\rangle = b(\tilde{\xi}, v) \quad \forall v \in V.$$

Note that, as ξ vanishes over K, it can be seen as a linear functional defined on $K^\perp$, so that it is equivalent to establish that $T \, : \, V \longrightarrow (K^\perp)'$ defined by

$$\tilde{\xi} \longmapsto \zeta \, : \, \langle \zeta \, , v\rangle = b(\tilde{\xi}, v) \quad \forall v \in K^\perp,$$

is surjective. We denote by $T^\star \in \mathcal{L}\,(K^\perp, V)$ the adjoint of T. For all $w \in K^\perp$,

$$|T^\star w| = \sup_{v\neq 0} \frac{(T^\star w, v)}{|v|} = \sup_{v\neq 0} \frac{b(w,v)}{|v|} = \sup_{v\neq 0} \frac{(Bw, Bv)}{|v|} \geq \frac{|Bw|^2}{|w|}.$$

As B has closed range, $|Bw| \geq C\,|w|$ for all w in $(\ker B)^\perp = K^\perp$, so that

$$|T^\star w| \geq C^2\,|w| \quad \forall w \in K^\perp,$$

from which we conclude that T is surjective. □

Remarque 9.27. Le corollaire qui précède est strictement plus faible que le résultat général de convergence à l'ordre 1. Dans le cas où l'on cherche à imposer des conditions de Dirichlet homogènes par exemple, on cherche u^ε dans $H^1(\Omega)$ tel que

$$\int_\Omega (u^\varepsilon v + \nabla u^\varepsilon \cdot \nabla v) + \frac{1}{\varepsilon}\int_\Gamma u^\varepsilon v = \int_\Omega f v \quad \forall v \in H^1(\Omega).$$

La proposition 9.24 assure la convergence vers la solution du problème avec conditions de Dirichlet homogènes en $\mathcal{O}(\varepsilon)$, alors que la forme $b(\cdot,\cdot)$ ne vérifie pas la condition du corollaire 9.26 (l'espace $H^{1/2}$ des traces de fonctions de H^1 n'est pas fermé dans $L^2(\partial\Omega)$).

Nous terminons par une suite de propriétés qui conduisent à une démonstration alternative de la convergence à l'ordre 1 dans le cas où $b(u,v)$ se met sous la forme (Bu, Bv), avec B à image fermée (corollaire 9.26). Les étapes de ce cheminement pourront s'avérer utile pour les estimations d'erreur du problème discrétisé en espace.

Proposition 9.28. On note $\|a\|$ la constante de continuité de a, et $\mathrm{dist}(u^\varepsilon, K)$ la distance de u^ε à K. On a alors

$$|u^\varepsilon - u| \leq \sqrt{\frac{\|a\|}{\alpha}}\,\mathrm{dist}(u^\varepsilon, K).$$

Démonstration : Les formulations variationnelles des problèmes $\mathcal{P}$ et $\mathcal{P}_\varepsilon$ s'écrivent, respectivement,

$$a(u,v) = \langle \varphi \, , v\rangle \quad \forall v \in K_0,$$
$$a(u^\varepsilon, v) + \frac{1}{\varepsilon} b(u^\varepsilon - u_0, v) = \langle \varphi \, , v\rangle \quad \forall v \in V.$$

Comme $b(u^\varepsilon - u_0, v)$ est nul pour tout v dans K_0, on a

$$a(u^\varepsilon - u, v) = 0 \quad \forall v \in K_0,$$

ce qui exprime que u minimise la quantité $a(u^\varepsilon - v, u^\varepsilon - v)$ sur K (en d'autres termes, u est la projection sur K de u^ε pour la norme associée au produit scalaire $a(\cdot,\cdot)$). On a ainsi

$$\alpha \left|u^\varepsilon - u\right|^2 \leq a(u^\varepsilon - u, u^\varepsilon - u) = \min_{v\in K} a(u^\varepsilon - v, u^\varepsilon - v) \leq \|a\| \min_{v\in K} |u^\varepsilon - v|^2 = \|a\| \operatorname{dist}(u^\varepsilon, K)^2,$$

d'où la majoration de $|u^\varepsilon - u|$ annoncée. □

L'estimation précédente ne permet pas toujours d'accéder à une estimation de l'erreur en fonction de ε. Il existe pourtant une situation où l'on peut exprimer cette erreur, situation qui fait l'objet de la proposition ci-dessous, dont la démonstration repose sur les deux lemmes suivants.

Lemme 9.29. Il existe une constante C telle que

$$\frac{1}{\varepsilon} b(u^\varepsilon - u_0, u^\varepsilon - u_0) \leq C \left|u - u^\varepsilon\right|.$$

Démonstration : On utilise la chaîne d'inégalités (9.8), page 124, qui s'écrit

$$\frac{1}{2}a(u^\varepsilon, u^\varepsilon) - \langle \varphi\,, u^\varepsilon\rangle \leq \frac{1}{2}a(u^\varepsilon, u^\varepsilon) - \langle \varphi\,, u^\varepsilon\rangle + \frac{1}{2\varepsilon} b(u^\varepsilon - u_0, u^\varepsilon - u_0) \leq \frac{1}{2}a(u,u) - \langle \varphi\,, u\rangle,$$

d'où l'on déduit

$$0 \leq \frac{1}{2\varepsilon} b(u_\varepsilon - u_0, u_\varepsilon - u_0) \leq \frac{1}{2}a(u,u) - \frac{1}{2}a(u^\varepsilon, u^\varepsilon) + \langle \varphi\,, u^\varepsilon - u\rangle.$$

Comme la forme quadratique $v \mapsto a(v,v)$ est Lipschitzienne sur tout borné, il existe une constante telle que le membre de droite de l'inégalité ci-dessus se majore par une constante multipliée par $|u^\varepsilon - u|$. □

Lemme 9.30. On suppose que $b(\cdot,\cdot)$ est de la forme

$$b(u,v) = (Bu, Bv)\,,$$

où B est application linéaire continue de V dans un espace de Hilbert Λ, à image fermée. Il existe alors une constante C telle que

$$|u^\varepsilon - u| \leq C\sqrt{b(u^\varepsilon - u_0, u^\varepsilon - u_0)}.$$

Démonstration : Comme B est à image fermée, il existe une constante $\beta > 0$ telle que, pour tout $\varepsilon > 0$, il existe un antécédent w^ε à $B(u^\varepsilon - u_0)$ avec

$$|w^\varepsilon| \leq \beta \left|B(u^\varepsilon - u_0)\right|.$$

Comme $Bw^\varepsilon = Bu^\varepsilon - Bu_0$, on a $u^\varepsilon - w^\varepsilon \in K$, et ainsi

$$\operatorname{dist}(u^\varepsilon, K) \leq |u^\varepsilon - (u^\varepsilon - w^\varepsilon)| = |w^\varepsilon| \leq \beta \left|B(u^\varepsilon - u_0)\right| = \beta\sqrt{b(u^\varepsilon - u_0, u^\varepsilon - u_0)}.$$

On conclut grâce à la proposition 9.28. □

Proposition 9.31. On suppose comme précédemment que $b(\cdot,\cdot)$ est de la forme

$$b(u,v) = (Bu, Bv)\,,$$

où B est linéaire continue de V dans un espace de Hilbert Λ, à image fermée. Il existe alors une constante C telle que

$$|u^\varepsilon - u| \leq C\varepsilon.$$

Démonstration : La démonstration est une conséquence directe des lemmes précédents :

$$|u^\varepsilon - u| \leq C_1\sqrt{b(u^\varepsilon - u_0, u^\varepsilon - u_0)} \leq C_2\sqrt{\varepsilon}\sqrt{|u^\varepsilon - u|},$$

d'où l'on déduit l'estimation en $\mathcal{O}(\varepsilon)$. □

Remarque 9.32. (Autre démonstration de 9.31)
La propriété précédente est en général démontrée d'une façon très différente (voir par exemple Brezzi [**4**]), par introduction d'une nouvelle variable qui s'apparente au multiplicateur de Lagrange de la section suivante. On vérifie aisément (en prenant $\lambda^\varepsilon = 1/\varepsilon Bu^\varepsilon$) que le problème $\mathcal{P}_\varepsilon$ est équivalent à trouver $(u^\varepsilon, \lambda^\varepsilon) \in V \times \Lambda$ tel que

$$\begin{array}{rcrcl} Au^\varepsilon & + & B^\star\lambda^\varepsilon & = & \varphi \\ Bu^\varepsilon & - & \varepsilon\lambda^\varepsilon & = & 0. \end{array}$$

Une propriété des systèmes variationnels de ce type (voir Brezzi [**4**, th. 1.2, page 47]) permet alors de conclure à la convergence de u^ε vers u en $\mathcal{O}(\varepsilon)$ dans le cas où B est à image fermée.

Chapitre 10

Méthode des éléments finis : aspects théoriques

10.1. Approximation de Lagrange

Nous établissons dans cette section des résultats d'approximation d'une fonction u. On se propose ici de préciser comment une fonction u sur un domaine peut être approchée par une fonction u_h continue, dont la restriction à chaque élément d'un maillage donné est un polynôme de degré fixé.

10.1.1. Préliminaires. Dans la suite K désigne un simplexe de $\mathbb{R}^N$ non dégénéré (*i.e.* de volume non nul). On désignera par $\tilde{K}$ le simplexe de référence, défini par

$$\tilde{K} = \left\{(x_1, \dots, x_N) \in \mathbb{R}^N_+ ,\ x_1 + \cdots + x_N \leq 1\right\}.$$

NOTATION 10.1. Pour toute fonction w définie sur K (ou sur tout autre domaine), on notera (lorsque ces quantités sont définies)

$$|w|_{0,K} = \|w\|_{L^2(K)} ,\ |w|_{1,K} = \|\nabla w\|_{L^2(K)^2} ,\ |w|_{2,K} = \left\|D^2 w\right\|_{L^2(K)^{N^2}} = \left(\sum_{i,j} |\partial_{ij} u|^2\right)^{1/2}.$$

NOTATION 10.2. On note $P^k(K)$ l'espace des fonctions polynômiales sur K, de degré total inférieur ou égal à k. Ainsi $P^1(K)$ désigne l'espace des fonctions affines sur K, de dimension $N+1$, et $P^0(K)$ la droite des fonctions constantes.

Lemme 10.3. Soit I_K un opérateur linéaire continu de $H^{k+1}(K)$ dans $H^m(K)$, pour m entier positif ou nul, inférieur ou égal à k. On suppose que I_K laisse invariant tous les éléments de P^k. Alors il existe une constante C telle que

$$|v - I_K v|_{m,K} \leq C\, |v|_{k+1,K} \quad \forall v \in H^{k+1}(K).$$

DÉMONSTRATION : On raisonne par l'absurde, en supposant l'existence d'une suite (v_n) telle que

$$|v_n - I_K v_n|_{m,K} > nC\, |v_n|_{k+1,K}.$$

On choisit de prendre v_n dans l'orthogonal de P^k (ce qui est possible, quitte à corriger par un polynôme de degré k, ce qui ne change aucun des membres), et de norme 1 dans H^{k+1}. Cette suite est bornée dans H^{k+1}, on peut donc en extraire une sous-suite qui converge faiblement vers $u \in H^{k+1}$. Cette sous-suite (toujours notée v_n) converge fortement dans H^k par injection compacte, et donc fortement en fait dans H^{k+1} car, $|v_n|_{k+1,K}$ tendant vers 0, elle y est de Cauchy. Elle converge donc fortement vers u. La limite u est dans l'orthogonal de P^k, mais toutes ses dérivées à l'ordre $k+1$ sont nulles : il s'agit donc d'un polynôme de degré au plus k. On a donc $u = 0$, ce qui absurde car u est de norme 1 dans H^{k+1}. □

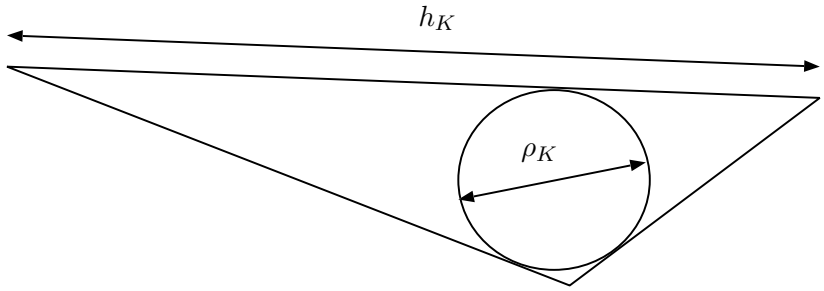


FIGURE 1. Définition de h et ρ pour un triangle

Remarque 10.4. On notera que la démonstration est dans son esprit très proche de celle de l'inégalité de Poincaré généralisée (proposition 8.51, page 103). De fait, dans le cas particulier $m = k = 0$, avec I_K projection L^2 sur les fonctions constantes (i.e. opérateur qui à une fonction associe sa valeur moyenne), le lemme ci-dessus est conséquence de cette inégalité de Poincaré généralisée, en prenant pour T l'opérateur I_K lui-même. On applique l'inégalité à $v - I_K v$:

$$|v - I_K v|_0 \leq C \left(\|I_K(v - I_K v)\|_{H^1} + |\nabla(v - I_K v)|_0 \right).$$

Comme $I_K(I_K(v)) = I_K(v)$, le premier terme du membre de droite s'annule, et comme d'autre part $\nabla I_K(v) = 0$, on obtient l'inégalité annoncée.

10.1.2. Approximation sur un simplexe (éléments d'ordre 1).

Définition 10.5. (Opérateur d'interpolation)
On définit l'opérateur d'interpolation I_K comme l'application de $C(K)$ (ensemble des applications continues de K dans $\mathbb{R}$) dans $P^1(K)$ qui à $u \in C(K)$ associe la fonction $I_K u$ affine sur K qui prend la valeur $u(\mathbf{x})$ en chaque sommet $\mathbf{x}$ de K. On définit de même I_K^0 l'application de L^1 dans $P^0(K)$ qui à une fonction associe la fonction constante sur K, de même valeur moyenne.

NOTATION 10.6. On note h_K la longueur de la plus longue arête de K, et ρ_K le diamètre de la plus grande sphère contenue dans K (voir figure 1). On a ainsi $h_K/\rho_K \geq 1$. On notera $\tilde{h}$ et $\tilde{\rho}$ les quantités associées au simplexe de référence.

Lemme 10.7. Soit Φ l'application affine qui envoie $\tilde{K}$ dans K (noter que l'on peut choisir Φ linéaire si l'on suppose que 0 est un sommet de chacun des simplexes) :

$$\tilde{\mathbf{x}} \longmapsto \mathbf{x} = \Phi(\tilde{\mathbf{x}}) = B\tilde{\mathbf{x}} + \mathbf{b}$$

On a

$$\|\nabla\Phi\| = \left\|{}^t\nabla\Phi\right\| = \|B\| \leq \frac{1}{\tilde{\rho}} h_K\,, \ \left\|\nabla\Phi^{-1}\right\| = \left\|{}^t\nabla\Phi^{-1}\right\| = \left\|B^{-1}\right\| \leq \frac{1}{\rho_K}\tilde{h}.$$

DÉMONSTRATION : Soit $\tilde{\xi} \in \mathbb{R}^N$ de norme $\tilde{\rho}$. Il existe $\tilde{\mathbf{x}}_1$ et $\tilde{\mathbf{x}}_2$ dans $\tilde{K}$ tels que $\tilde{\xi} = \tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1$. On a donc

$$B\tilde{\xi} = B\tilde{\mathbf{x}}_2 - B\tilde{\mathbf{x}}_1 = \Phi\tilde{\mathbf{x}}_2 - \Phi\tilde{\mathbf{x}}_1 = \mathbf{x}_2 - \mathbf{x}_1,$$

qui est de norme inférieure à h_K par définition. On en déduit la première inégalité. La seconde se montre de la même manière en considérant $\xi = \mathbf{x}_2 - \mathbf{x}_1$ de norme ρ_K. □

Théorème 10.8. On suppose $N = 1$, 2, ou 3, de telle sorte que $H^2(K)$ s'injecte de façon continue dans $C^0(\overline{K})$. Il existe une constante C universelle telle que, pour tout triangle K du plan, non dégénéré, on a

$$\begin{aligned} |I_K u - u|_{1,K} &\leq C\frac{h_K^2}{\rho_K}\,|u|_{2,K} \quad \forall u \in H^2(K) \\ |I_K u - u|_{0,K} &\leq C h_K^2\,|u|_{2,K} \quad \forall u \in H^2(K) \\ \left|I_K^0 u - u\right|_{0,K} &\leq C h_K\,|u|_{1,K} \quad \forall u \in H^1(K) \end{aligned}$$

Démonstration : On se reportera à Raviart [**13**] pour une démonstration détaillée de ces inégalités, pour des approximations d'ordre arbitrairement élevé. La première propriété est obtenue de la manière suivante : on écrit l'inégalité du lemme 10.3 pour $k = 1$, $m = 1$,

$$\left|\tilde{u} - I_{\tilde{K}}\tilde{u}\right|_{1,\tilde{K}} \leq C\,|\tilde{u}|_{2,\tilde{K}}\,,$$

où $\tilde{K}$ est le simplexe de référence défini par

$$\tilde{K} = \left\{(x_1,\ldots,x_N) \in \mathbb{R}_+^N\,,\ x_1 + \cdots + x_N \leq 1\right\}$$

de telle sorte que la constante C ci-dessus est une constante universelle. Pour un simplexe K non dégénéré, on introduit ensuite l'application affine Φ qui envoie $\tilde{K}$ dans K (voir lemme 10.7 ci-dessus). Pour tout $u \in H^2(K)$, on définit $\tilde{u} = u \circ \Phi \in H^2(\tilde{K})$, et l'on écrit l'inégalité d'interpolation pour $\tilde{u}$. Il reste à effectuer un changement de variable pour remplacer les intégrales sur $\tilde{K}$ par des intégrales sur K. Noter en premier lieu que le Jacobien de la transformation $|\nabla\Phi|$ se trouve de part et d'autre de l'inégalité, et ne va donc jouer aucun rôle dans le transport de l'inégalité. En revanche le fait de dériver en espace à des échelles différentes introduit des modifications. Plus précisément on a, pour toute fonction $\tilde{v} = v \circ \Phi \in H^1(\tilde{K})$,

$$\nabla\tilde{v} = {}^t\nabla\Phi \cdot \nabla v\,,\ \nabla v = {}^t\nabla\Phi^{-1} \cdot \nabla\tilde{v}$$

On a ainsi

$$|v|_{1,K}^2 = \int_K |\nabla v|^2\,dx = \int_{\tilde{K}} \left|{}^t\nabla\Phi^{-1} \cdot \nabla\tilde{v}\right|^2 |\nabla\Phi|\,d\tilde{x} \leq C\left\|B^{-1}\right\|^2 |\tilde{v}|_{1,\tilde{K}}^2\,|\nabla\Phi|\,.$$

Comme Φ est affine et envoie les sommets de $\tilde{K}$ sur ceux de K, on a $I_K u \circ \Phi = I_{\tilde{K}}\tilde{u}$. En appliquant l'inégalité ci-dessus à $u - I_K u$, en utilisant l'inégalité sur $\tilde{K}$, on obtient donc

$$|u - I_K u|_{1,K} \leq \left\|B^{-1}\right\| |\nabla\Phi|^{1/2} \left|\tilde{u} - I_{\tilde{K}}\tilde{u}\right|_{1,\tilde{K}} \leq C\left\|B^{-1}\right\| |\nabla\Phi|^{1/2}\,|\tilde{u}|_{2,\tilde{K}}\,.$$

On effectue maintenant le changement de variable $\tilde{\mathbf{x}} \mapsto \mathbf{x} = \Phi(\tilde{\mathbf{x}})$ pour se ramener à une intégrale sur K. Comme la semi norme $|\cdot|_2$ fait intervenir des dérivées secondes, on obtient

$$|u - I_K u|_{1,K} \leq C\left\|B^{-1}\right\| \|B\|^2\,|\nabla\Phi|^{1/2}\,|\nabla\Phi|^{-1/2}\,|u|_{2,K}\,.$$

Le lemme 10.7 permet de conclure (les quantités $\tilde{\rho}$ et $\tilde{h}$ sont intégrées à la constante). Les autres inégalités se démontrent de la même manière. □

10.1.3. Approximation sur un domaine.

Définition 10.9. (Triangulation)
Soit Ω un domaine polygonal du plan. On appelle triangulation de Ω une famille T_h de simplexes non dégénérés [1] deux à deux disjoints telle que

$$\overline{\Omega} = \bigcup_{K \in T_h} \overline{K},$$

et telle que toute face d'un polyèdre K de T_h est la face d'un autre polyèdre K' de T_h, ou alors est inclus dans la frontière de Ω. Les sommets des polyèdres de T_h sont appelés les nœuds de la triangulation.

Définition 10.10. (Opérateur d'interpolation)
Soit Ω un domaine polygonal du plan, et T_h une triangulation de Ω en simplexes (segments, triangles, ou tétraèdres pour $N = 1$, 2, 3, respectivement). On définit l'opérateur d'interpolation I_h comme l'application de $C(\overline{\Omega})$ (ensemble des applications continues de $\overline{\Omega}$ dans $\mathbb{R}$) qui à $u \in C(\overline{\Omega})$ associe la fonction u_h affine sur chaque $K \in T_h$ qui prend la valeur $u(\mathbf{x})$ en chaque sommet $\mathbf{x}$ de T_h.

Le paramètre h joue un rôle un peu ambigu dans ce qui suit : il désigne à la fois l'indice d'un membre d'une famille de triangulations (c'est donc le *label* d'une triangulation), et ce qu'il est convenu d'appeler le diamètre de la triangulation , c'est à dire le sup de h_K pour $K \in T_h$, qui est un nombre réel. C'est évidemment un abus de notation, puisque deux triangulations peuvent avoir le même diamètre sans être identiques. Nous conservons néammoins cet usage qui ne pose pas de problème en pratique.

Définition 10.11. (Famille régulière de triangulations)
Soit Ω un domaine polygonal. On appelle famille régulière de triangulations une famille (T_h) telle que

(i) il existe une constante σ telle que $\sup_h \sup_{K \in T_h} (h_K/\rho_K) \leq \sigma$,

(ii) le diamètre de T_h tend vers 0, c'est-à-dire que $\sup_{K \in T_h} h_K \longrightarrow 0$.

Théorème 10.12. Soit Ω un domaine polygonal, et (T_h) une famille régulière de triangulations de Ω. Pour tout $u \in H^2(\Omega)$, on a

$$|u - I_h u|_{1,\Omega} \leq C\sigma h\, |u|_{2,\Omega}\ ,\ |u - I_h u|_{0,\Omega} \leq Ch^2\, |u|_{2,\Omega}$$

Démonstration : On a

$$\int_\Omega |u - I_h u|^2 = \sum_{K \in T_h} \int_K |u - I_h u|^2 \leq C^2 h^4 \sum_{K \in T_h} |u|^2_{2,K} \leq C^2 h^2\, |u|^2_{2,\Omega}\,.$$

On raisonne de la même manière pour estimer $|u - I_h u|_{1,\Omega}$. □

1. La notion de triangulation s'étend à des décompositions en polyèdres quelconques, mais nous nous limiterons ici à des triangulations au sens premier du terme, c'est-à-dire des décompositions en triangles ou tétraèdre.

10.1.4. Approximation d'ordres supérieurs, généralisations.

Définition 10.13. (N-simplexe)
On appelle N-simplexe de $\mathbb{R}^N$ l'enveloppe convexe de $N+1$ points de $\mathbb{R}^N$. On dira que le N-simplexe K est non dégénéré si le volume engendré est non nul, on de façon équivalente si les $N+1$ points n'appartiennent pas à un même hyperplan de $\mathbb{R}^N$.

Définition 10.14. (Maillage)
Soit Ω un ouvert polyédrique de $\mathbb{R}^N$. Un maillage (triangulaire) de Ω est un ensemble T_h de N-simplexes non dégénérés inclus dans $\overline{\Omega}$, dont la réunion recouvre $\overline{\Omega}$, et tel que l'intersection de 2 quelconques de ces simplexes est un k simplexe, avec $0 \leq k \leq N$, dont tous les sommets sont des sommets de T_h.

Théorème 10.15. Soit $k \geq 1$ (de telle sorte que $H^{k+1}(K)$ s'injecte de façon continue dans $C^0(\overline{K})$ pour les dimensions physiques), $m \leq K$, et I_K un opérateur d'interpolation d'ordre k. Il existe une constante C universelle telle que, pour tout triangle K du plan, non dégénéré, on a

$$|I_K u - u|_{m,K} \leq C \frac{h^{k+1}}{\rho^m} |u|_{k+1,K} \quad \forall u \in H^{k+1}(K).$$

Démonstration : La démonstration est semblable à celle du théorème 10.12. Le principe à retenir, peu surprenant, est que plus on est exigeant sur l'erreur, plus l'estimation est mauvaise, et que plus la solution est régulière, meilleure est l'estimation (sous réserve que l'on utilise des éléments d'ordre suffisamment élevés pour profiter de la régularité de cette solution). Plus précisément, lorsque l'on exige une erreur portant sur un ordre de dérivée m, on le paye par un facteur $1/\rho^m$ (qui tend vers $+\infty$ quand h tend vers 0) dans la constante. Si l'on dispose d'un majorant de la dérivée $k+1$-ième de u, et que l'on interpole par des polynômes d'ordre k, alors on gagne un facteur h^{k+1} dans la constante. Ces deux facteurs, l'un défavorable, l'autre favorable, apparaissent dans la démonstration sous la forme de $\left\|\nabla\Phi^{-1}\right\|^m$ et $\|\nabla\Phi\|^{k+1}$, respectivement, lors des changements de variables qui permettent de passer de l'élément de référence à l'élément réel. On se reportera à [**13**] pour une démonstration détaillée. □

Définition formelle d'un élément fini. La définition la plus précise de ce que l'on appelle *un élément fini*, telle qu'elle s'est imposée dans la littérature, est la suivante :

un élément fini [2] sur $\mathbb{R}^N$ est la donnée d'un triplet (K, P, Σ), plus précisément

(1) Une partie K compacte, connexe, et d'intérieur non vide de $\mathbb{R}^N$;

(2) Un espace P de dimension finie n de fonctions de K dans $\mathbb{R}$;

(3) Un ensemble $\Sigma = (\sigma_1, \ldots, \sigma_n)$ de formes linéaires sur P (appelés degrés de liberté de l'élément), tels que l'application

$$p \in P \longmapsto (\sigma_1(p), \ldots, \sigma_n(p)) \in \mathbb{R}^n,$$

soit bijective.

2. Certains auteurs parlent d'élément unisolvent : la donnée des n réels $\sigma_i(p)$ détermine de façon unique la fonction $p \in P$.

10.2. Principes abstraits

10.2.1. Approche directe. Nous commençons cette section par une propriété abstraite qui nous permettra d'établir des estimations d'erreur pour un problème de minimisation résolu directement (c'est-à-dire en se plaçant dans l'espace admissible), ce qui revient à un problème de minimisation sans contrainte. On considère $a(\cdot,\cdot)$ une forme bilinéaire symétrique coercive sur V, de constante de coercivité α et de constante de continuité $\|a\|$, et $f \in V'$. On note u l'élément de V qui minimise la fonctionnelle

$$v \in V \longmapsto J(v) = \frac{1}{2}a(v,v) - \langle \varphi\,,\, v\rangle.$$

Dans le cadre de la discrétisation en espace qui sera présentée dans les sections suivantes, on utilisera la notation V_h pour représenter un espace d'approximation de dimension finie, h étant un paramètre associé au maillage sur lequel cette discrétisation s'effectue. Dans la proposition abstraite qui suit, à la base de la méthode des éléments finis, V_h désigne simplement un sous-espace fermé de V.

Proposition 10.16. (Lemme de Céa (cas symétrique))
Soit $a(\cdot,\cdot)$ une forme bilinéaire symétrique coercive sur V, de constante de coercivité α et de constante de continuité $\|a\|$, et $\varphi \in V'$. On note u l'élément de V qui minimise la fonctionnelle

$$v \in V \longmapsto J(v) = \frac{1}{2}a(v,v) - \langle \varphi\,,\, v\rangle.$$

Soit V_h un sous-espace fermé de V. On note u_h l'élément de V_h qui minimise J sur V_h. alors

$$|u_h - u| \leq \sqrt{\frac{\|a\|}{\alpha}} \inf_{v_h \in V_h} |v_h - u|\,.$$

Démonstration : On écrit les formulations variationnelles associées aux problèmes de minimisation sur V et sur V_h, respectivement,

$$a(u,v) = \langle \varphi\,,\, v\rangle \quad \forall v \in H,$$
$$a(u_h,v_h) = \langle \varphi\,,\, v_h\rangle \quad \forall v_h \in V_h.$$

On a donc

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

ce qui exprime que u_h minimise la fonctionnelle $v \mapsto a(v_h - u, v_h - u)$ sur V_h. On a donc, en utilisant la coercivité et la continuité de $a(\cdot,\cdot)$,

$$\alpha\,|u_h - u|^2 \leq a(u_h - u, u_h - u) \leq \inf_{v_h \in V_h} a(v_h - u, v_h - u) \leq \|a\| \inf_{v_h \in V_h} |v_h - u|^2\,,$$

d'où l'inégalité annoncée. □

La propriété demeure (avec une constante plus grande) pour une forme non symétrique, comme l'exprime le lemme de Céa général :

Proposition 10.17. (Lemme de Céa)
Soit $a(\cdot,\cdot)$ une forme bilinéaire (non nécessairement symétrique) coercive sur V, de constante de coercivité α et de constante de continuité $\|a\|$, et $\varphi \in V'$. Soit V_h un sous-espace de V. On note u et u_h les élements de V et V_h, respectivement, qui vérifient

$$a(u,v) = \langle \varphi\,,\, v\rangle \quad \forall v \in V,$$

$$a(u_h, v_h) = \langle \varphi \, , \, v_h \rangle \quad \forall v_h \in V_h.$$

Alors

$$|u_h - u| \leq \frac{\|a\|}{\alpha} \inf_{v_h \in V_h} |v_h - u| \, .$$

DÉMONSTRATION : On utilise comme précédemment

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

dont on déduit que $a(u_h - u, u_h - u) = a(u_h - u, v_h - u)$, pour tout $v_h \in V_h$, d'où

$$\alpha \, |u_h - u|^2 \leq a(u_h - u, u_h - u) \leq |a(u_h - u, v_h - u)| \leq \|a\| \, |u - u_h| \inf_{v_h \in V_h} |v_h - u| \, ,$$

d'où l'on déduit l'inégalité en prenant l'infimum en v_h. □

Il pourra être utilie (dans le cas de conditions aux limites de Dirichlet non homogènes par exemple) d'utiliser la forme affine de ces lemmes. Nous l'explicitons ici dans le cas général (non symétrique).

Proposition 10.18. (Lemme de Céa (sur un espace affine))
Soit V un espace de Hilbert, $a(\cdot, \cdot)$ une forme bilinéaire continue sur V, et $\varphi \in V'$. Soit V^1 un sous-espace affine fermé de V et V^0 le sous-espace vectoriel associé. Soit V_h^1 un sous-espace affine fermé de V^1, et V_h^0 le sous-espace vectoriel sous-jacent. On suppose $a(\cdot, \cdot)$ coercive sur V^0, et l'on note u et u_h les élements de V^1 et V_h^1, respectivement, qui vérifient

$$a(u, v) = \langle \varphi \, , \, v \rangle \quad \forall v \in V^0,$$
$$a(u_h, v_h) = \langle \varphi \, , \, v_h \rangle \quad \forall v_h \in V_h^0.$$

Alors

$$|u_h - u| \leq \frac{\|a\|}{\alpha} \inf_{v_h \in V_h^1} |v_h - u| \, .$$

DÉMONSTRATION : On introduit $u_h^1 \in V_h^1$. On a, pour tout $v_h \in V_h^0$,

$$\begin{aligned} a(u - u_h, u - u_h) = a(u - u_h, (u - u_h^1) - \underbrace{(u_h - u_h^1)}_{\in V_h^0}) &= a(u - u_h, (u - u_h^1) - v_h) \\ &= a(u - u_h, u - (v_h + u_h^1)), \end{aligned}$$

d'où l'inégalité proposée ($v_h + u_h^1$ parcourt V_h^1 quand v_h parcourt V_h^0). □

10.3. Résolution de problèmes elliptiques par éléments finis

Proposition 10.19. Soit Ω un domaine polyédrique convexe, et (T_h) une famille régulière de triangulations de Ω. On note V_h l'ensemble des fonctions de $H_0^1(\Omega)$ dont la restriction à chaque triangle de T_h est affine. Pour $f \in L^2(\Omega)$, on note $u \in H_0^1(\Omega)$ la solution faible de

$$-\triangle u = f,$$

et u_h la solution du problème discrétisé

$$\int_\Omega \nabla u_h \cdot \nabla v_h = \int_\Omega f v_h \quad \forall v_h \in V_h.$$

Il existe une constante $C > 0$ telle que

$$|u - u_h|_{\Omega,1} \leq Ch\,|f|_{\Omega,0}\,.$$

DÉMONSTRATION : C'est une application directe de la proposition 8.63, page 107 (ou plus précisément de la proposition 8.65 qui s'applique au cas d'un polyèdre convexe), et du lemme de Céa 10.16. □

Proposition 10.20. (Lemme de Aubin-Nitsche)
Sous les hypothèses de la proposition précédente, il existe une constante $C > 0$ telle que

$$|u - u_h|_{\Omega,0} \leq Ch^2\,|f|_{\Omega,0}\,.$$

DÉMONSTRATION : On considère le problème aux limites suivant

$$-\triangle w = u - u_h.$$

On prend $u - u_h$ comme fonction-test dans la formulation variationnelle de ce problème. Il vient

$$\int_\Omega |u - u_h|^2 = \int_\Omega \nabla w \cdot \nabla(u - u_h) = \int_\Omega \nabla(w - I_h w) \cdot \nabla(u - u_h)$$

car $\int \nabla(u - u_h) \cdot \nabla v_h = 0$ pour tout $v_h \in V_h$. On a donc

$$|u - u_h|_0^2 \leq |w - I_h w|_1\,|u - u_h|_1\,.$$

Le premier facteur du produit se majore de la façon suivante

$$|w - I_h w|_1 \leq Ch\,|w|_2 \leq Ch\,|u - u_h|_0\,,$$

et le second par $C_4 h\,|f|_0$, d'où l'estimation en $\mathcal{O}(h^2)$ sur la norme L^2 de l'erreur. □

10.4. Approximation des valeurs propres

On s'intéresse ici à l'approximation des valeurs propres d'une forme bilinéaire du type $\int \nabla u \cdot \nabla v$.

Théorème 10.21. On se place dans le cadre du théorème 13.5, page 162. On introduit une suite d'espaces d'approximation (V_h) de V, et l'on note (u_h^i, λ_h^i) les solutions du problème aux valeurs propres sur V_h :

$$a(u_h^i, v) = \lambda_h^i (u_h^i, v),$$

où $(\cdot,\cdot)$ est le produit scalaire sur H.

On a alors, pour tout i, convergence de λ_h^i vers λ^i quand h tend vers 0.

DÉMONSTRATION : On note N_h la dimension de V_h. Notons tout d'abord que le principe du min-max

$$\lambda^i = \min_{W \in E^i} \max_{w \in W \setminus \{0\}} R(w)\,,\ \lambda_h^i = \min_{W \in E_h^i} \max_{w \in W \setminus \{0\}} R(w)$$

où E^i (respectivement E_h^i désigne l'ensemble des sous-espaces vectoriels de V (resp. V_h) de dimension i, implique $\lambda^i \leq \lambda_h^i$ pour tout $i \leq N_h$. Notons Π_h la projection de V sur V_h

pour le produit scalaire associé à $a(\cdot,\cdot)$, et W_i l'espace vectoriel engendré par les i premiers vecteurs propres de $a(\cdot,\cdot)$. Pour tout $u \in W_i$, on a

$$u = \sum_{k=1}^{i} \beta^k u_k,$$

et ainsi

$$\|\Pi_h u - u\|_V = \left| \sum_{k=1}^{i} \beta^k (\Pi_h u_k - u_k) \right| \leq \left(\sum_{k=1}^{i} \left| \beta^k \right|^2 \right)^{1/2} \left(\sum_{k=1}^{i} \|\Pi_h u_k - u_k\|_V^2 \right)^{1/2}$$
$$= |u| \left(\sum_{k=1}^{i} \|\Pi_h u_k - u_k\|_V^2 \right)^{1/2}.$$

On a donc

$$\lim_{h \to 0} \sup_{u \in W_i} \frac{|\Pi_h u - u|_V}{|u|} = 0$$

Par ailleurs, on a $a(\Pi_h u, \Pi_h u) \leq a(u,u)$, pour tout $u \in V$. Le principe du min-max permet pour finir d'écrire que

$$\lambda_h^i \leq \max_{w \in W_h \setminus \{0\}} R(w),$$

pour tout W_h de dimension i. Prenant $W_h = \Pi_h(W_i)$, il vient

$$\lambda_h^i \leq \max_{u \in W_i \setminus \{0\}} \frac{a(\Pi_h u, \Pi_h u)}{|\Pi_h u|^2} \leq \max_{u \in W_i \setminus \{0\}} \frac{a(u,u)}{|\Pi_h u|^2} \leq \lambda_i \max_{u \in W_i \setminus \{0\}} \frac{|u|^2}{|\Pi_h u|^2}.$$

Mais, d'après ce qui précède, on a

$$|\Pi_h u| = |u| + \mathcal{O}(|\Pi_h u - u|) = |u| + \mathcal{O}(\|\Pi_h u - u\|_V) = |u| \, (1 + o(h))$$

d'où l'on déduit, pour tout i, la convergence de λ_h^i vers λ^i quand h tend vers 0. □

Chapitre 11

Méthode des éléments finis pour les problèmes sous contrainte

11.1. Penalty and FEM

As in Section 9.3, we consider the following set of assumptions

$$\left|\begin{array}{l} V \text{ is a Hilbert space}, \varphi \in V', \\ a(\cdot,\cdot) \text{ bilinear, symmetric, continuous, elliptic } (a(v,v) \geq \alpha\,|v|^2), \\ b(\cdot,\cdot) \text{ bilinear, symmetric, continuous, non-negative}, \\ K = \{u \in V\ ,\ b(u,u) = 0\} = \ker b, \\ J(v) = \dfrac{1}{2}a(v,v) - \langle \varphi\ , v\rangle, \quad u = \arg\min_K J, \\ J_\varepsilon(v) = \dfrac{1}{2}a(v,v) + \dfrac{1}{2\varepsilon}b(v,v) - \langle \varphi\ , v\rangle, \quad u^\varepsilon = \arg\min_V J_\varepsilon. \end{array}\right\} \tag{11.1}$$

We introduce now a family $(V_h)_h$ of inner approximation spaces ($V_h \subset V$), and the associated penalized/discretized problems expressed in their variational form :

$$\text{Find } u_h^\varepsilon \in V_h \text{ such that } J^\varepsilon(u_h^\varepsilon) = \inf_{v_h \in V_h} J^\varepsilon(v_h), \tag{11.2}$$

Our objective is to establish that u_h^ε tends to u as h and ε go to zero (in a manner which has to be made precise).

We shall need the following lemma :

Lemme 11.1. Under assumptions (11.1), there exists $C > 0$ such that

$$b(u^\varepsilon, u^\varepsilon) \leq C\varepsilon\,|u - u^\varepsilon|\,.$$

Démonstration. By definition of u^ε,

$$J_\varepsilon(u^\varepsilon) = \frac{1}{2}a(u^\varepsilon,u^\varepsilon) - \langle \varphi\ , u^\varepsilon\rangle + \frac{1}{2\varepsilon}b(u^\varepsilon,u^\varepsilon) \leq J_\varepsilon(u) = \frac{1}{2}a(u,u) - \langle \varphi\ , u\rangle,$$

so that

$$\begin{aligned} 0 \leq \frac{1}{2\varepsilon}b(u_\varepsilon,u_\varepsilon) &\leq \frac{1}{2}a(u,u) - \frac{1}{2}a(u^\varepsilon,u^\varepsilon) + \langle \varphi\ , u^\varepsilon - u\rangle \\ &\leq \frac{1}{2}a(u+u^\varepsilon, u-u^\varepsilon) + \langle \varphi\ , u^\varepsilon - u\rangle, \end{aligned}$$

which yields the estimate by continuity of $a(\cdot,\cdot)$ and φ.

□

Proposition 11.2. Under assumptions (11.1), we denote by u_h^ε the solution to Problem (11.2). Then

$$|u_h^\varepsilon - u| \leq C \left(\min_{v_h \in V_h \cap K} |v_h - u| + \sqrt{|u^\varepsilon - u|} \right).$$

Démonstration. As u_h^ε minimizes $a(v-u^\varepsilon, v-u^\varepsilon) + b(v-u^\varepsilon, v-u^\varepsilon)/\varepsilon$ over V_h,

$$\begin{aligned}
\alpha\, |u_h^\varepsilon - u^\varepsilon|^2 &\leq a(u_h^\varepsilon - u^\varepsilon, u_h^\varepsilon - u^\varepsilon) \\
&\leq a(u_h^\varepsilon - u^\varepsilon, u_h^\varepsilon - u^\varepsilon) + \frac{1}{\varepsilon} b(u_h^\varepsilon - u^\varepsilon, u_h^\varepsilon - u^\varepsilon) \\
&\leq \min_{v_h \in V_h} \left(a(v_h - u^\varepsilon, v_h - u^\varepsilon) + \frac{1}{\varepsilon} b(v_h - u^\varepsilon, v_h - u^\varepsilon) \right) \\
&\leq \min_{v_h \in V_h \cap K} \left(a(v_h - u^\varepsilon, v_h - u^\varepsilon) + \frac{1}{\varepsilon} b(v_h - u^\varepsilon, v_h - u^\varepsilon) \right).
\end{aligned}$$

As v_h is in K, the second term is $b(u^\varepsilon, u^\varepsilon)/\varepsilon$, which is bounded by $C\,|u^\varepsilon - u|$ (by Lemma 11.1). Finally we get

$$|u_h^\varepsilon - u^\varepsilon| \leq C \left(\min_{v_h \in V_h \cap K} |v_h - u^\varepsilon| + \sqrt{|u^\varepsilon - u|} \right),$$

from which we conclude. □

Proposition 11.3. Under assumptions (11.1), it holds

$$|u_h^\varepsilon - u| \leq \frac{C}{\sqrt{\varepsilon}} \inf_{v_h \in V_h} |u^\varepsilon - v_h| + |u^\varepsilon - u|,$$

where u_h^ε is the solution to (11.2).

Démonstration. One has

$$|u_h^\varepsilon - u| \leq |u_h^\varepsilon - u^\varepsilon| + |u^\varepsilon - u|,$$

and we control the first term by Céa's Lemma applied to the bilinear form $a + b/\varepsilon$, whose norm behaves like $1/\varepsilon$. □

11.2. FEM and saddle-point formulation

11.2.1. Approximation interne des multiplicateurs de Lagrange. On considère la formulation variationnelle du problème de point-selle

$$(\mathcal{P}'') \qquad \left\{ \begin{array}{lcll} a(u,v) + \langle B^\star \lambda\,, v \rangle &=& \langle \varphi\,, v \rangle & \forall v \in V \\ (\mu, Bu) &=& 0 & \forall \mu \in \Lambda. \end{array} \right. \tag{11.3}$$

dont on notera (u, λ) une solution quand elle existe. Anticipant sur la démarche de discrétisation en espace, nous ferons référence à ce problème en tant que problème continu. On utilisera la notation $b(v, \mu) = (\lambda, Bv)$.

On considère maintenant deux sous-espaces de dimensions finies (donc fermés) V_h et Λ_H de V et Λ, respectivement, et l'on s'intéresse au problème suivant

$$(\mathcal{P}''_h) \qquad \left\{ \begin{array}{rcll} a(u_h, v_h) & + & (\lambda_H, Bv_h) & = \langle \varphi\,, v_h\rangle \quad \forall v_h \in V_h \\ (\mu_H, Bu_h) & & & = 0 \quad \forall \mu_h \in \Lambda_H. \end{array} \right. \tag{11.4}$$

On définira l'opérateur B_H de V dans Λ_H par

$$(B_H v, \mu_H) = (Bv, \mu_H) \quad \forall \mu_H \in \Lambda_H.$$

Bien que certains des objets indicés par h ou H dépendent en fait simultanément des deux paramètres, nous allégerons les notations en se limitant au paramètre associé à l'espace dans lequel vivent des différents objets.

On introduit l'espace

$$K_h^H = \{v_h \in V_h\,,\ (Bv_h, \mu_H) = 0 \quad \forall \mu_H \in \Lambda_H\} = \ker B_H \cap V_h.$$

On prendra garde au fait que, en général, l'espace K_h^H ne s'identifie pas à $K \cap V_h$ (il le contient toujours, mais peut être strictement plus grand).

La question est bien entendu de savoir si une solution (u_h, λ_h) de $\mathcal{P}''_h$ est une approximation d'une (ou la) solution de $\mathcal{P}''$, lorsque V_h et Λ_H « approchent » V et Λ, respectivement. Nous démontrons ici trois résultats qui peuvent s'énoncer de la manière informelle (le sens que l'on donne à la notion de sous-espaces proches sera précisé par la suite) qui suit.

(1) Si on a existence d'un point-selle pour $\mathcal{P}''$, si V_h et Λ_h sont des approximations de V et Λ, respectivement, si K_h^H approche correctement K, alors u_h est une approximation de u.

(2) Dans le cadre des hypothèses qui assurent l'existence et l'unicité d'un point-selle, si l'on suppose de plus que les espaces V_h et Λ_h vérifient une propriété du type condition inf-sup, alors (u_h, λ_h) est une approximation de (u, λ).

(3) Sous les hypothèses les plus faibles sur B (existence d'un multiplicateur de Lagrange non assurée), si V_h approche V et si $B^\star \Lambda_H$ approche $K^\perp$ alors u_h est une approximation de u.

Proposition 11.4. On suppose que le problème (11.3) admet une solution (u, λ), et on note (u_h, λ_H) une solution du problème (11.4). On a

$$|u - u_h| \le \left(1 + \frac{\|a\|}{\alpha}\right) \inf_{w_h \in K_h^H} |u - w_h| + \frac{\|B\|}{\alpha} \inf_{\mu_H \in \Lambda_H} |\mu_H - \lambda|$$

Démonstration : Comme u_h minimise J sur K_h^H, on a

$$a(u_h, v_h) = \langle \varphi\,, v_h\rangle \quad \forall v_h \in K_h^H.$$

Pour tout $w_h \in K_h^H$, on note $v_h = u_h - w_h \in K_h^H$. On a

$$a(v_h, v_h) = \langle \varphi\,, v_h\rangle - a(w_h, v_h).$$

Comme u est solution de $\mathcal{P}''$, on a notamment (prendre $v = v_h$)

$$a(u, v_h) + (Bv_h, \lambda) = \langle \varphi\,, v_h\rangle,$$

d'où

$$a(v_h, v_h) = a(u - w_h, v_h) + (Bv_h, \lambda - \mu_H),$$

pour tout $\mu_H \in \Lambda_H$. On a donc

$$\alpha \left|u_h - w_h\right| \leq \|a\| \left|w_h - u\right| + \|B\| \left|\lambda - \mu_H\right|,$$

d'où

$$|u - u_h| \leq \left(1 + \frac{\|a\|}{\alpha}\right) |w_h - u| + \frac{\|B\|}{\alpha} |\lambda - \mu_H|,$$

pour tous $w_h \in K_h^H$, $\mu_H \in \Lambda_H$. □

Proposition 11.5. On suppose que B est surjective, et que la condition inf-sup discrète est satisfaite : il existe $\beta > 0$ telle que

$$\inf_{\mu_H \in \Lambda_H} \sup_{v_h \in V_h} \frac{(\mu_H, B v_h)}{|\mu_H| \, |v_h|} \geq \beta. \tag{11.5}$$

On note (u, λ) la solution du problème $\mathcal{P}''$, et (u_h, λ_H) la solution du problème $\mathcal{P}''_h$. On a

$$|u - u_h| + |\lambda - \lambda_H| \leq C \left(\inf_{v_h \in V_h} |u - v_h| + C_2 \inf_{\mu_H \in \Lambda_h} |\lambda - \mu_H| \right)$$

Démonstration : Montrons dans un premier temps que le min sur K_h^H dans l'estimation de la proposition 11.4 peut être remplacé par un min sur V_h. On montre pour cela que, quand la condition inf-sup discrète est vérifiée, pour tout $v_h \in V_h$ approchant u on peut construire $w_h \in K_h^H$ approchant u aussi bien (à une constante multiplicative près) que v_h.

Soit $v_h \in V_h$. On note z_h l'élément de V_h de norme minimale tel que

$$B_H z_h = -B_H v_h$$

Ce z_h est donc la partie primale du problème de point-selle

$$\begin{aligned} (z_h, y_h) + (\eta_H, B_H y_h) &= 0 \quad \forall y_h \in V_h \\ (\mu_H, B_H z_h) &= -(\mu_H, B_H v_h) \quad \forall \mu_H \in \Lambda_H. \end{aligned}$$

On a d'une part $|\eta_H| \leq |z_h| / \beta$ (voir proposition 9.10). D'autre part, prenant $y_h = z_h$, il vient

$$|z_h|^2 \leq |(\eta_H, B_H z_h)| \leq |\eta_H| \, |B_H v_h| \leq \frac{\|a\|}{\beta} |z_h| \, |B_H (u - v_h)|,$$

d'où finalement

$$|z_h| \leq \frac{\|B\|}{\beta} \beta \, |u - v_h|.$$

On a $w_h = z_h + v_h \in K_h^H$, et

$$|u - w_h| \leq |u - v_h| + |z_h|,$$

d'où

$$\inf_{w_h \in K_h^H} |u - w_h| \leq \left(1 + \frac{C_B}{\beta}\right) \inf_{v_h \in V_h} |u - v_h|. \tag{11.6}$$

Pour l'estimation d'erreur sur le multiplicateur de Lagrange, on écrit

$$(\lambda_H, B v_h) = a(u - u_h, v_h) + (\lambda, B v_h),$$

d'où

$$(\lambda_H - \mu_H, B v_h) = a(u - u_h, v_h) + (\lambda - \mu_H, B v_h) \quad \forall \mu_H \in \Lambda_H.$$

La condition inf-sup discrète implique donc

$$\begin{aligned} |\lambda_H - \mu_H| &\leq \frac{1}{\beta} \sup_{v_h \in V_h} \frac{|a(u-u_h, v_h) + (\lambda - \mu_H, Bv_h)|}{|v_h|} \\ &\leq \frac{1}{\beta} \left(\|a\| \, |u-u_h| + C_B \, |\lambda - \mu_H| \right), \end{aligned}$$

d'où

$$|\lambda - \lambda_H| \leq \frac{\|a\|}{\beta} |u-u_h| + \left(1 + \frac{C_B}{\beta}\right) \inf_{\mu_H \in \Lambda_H} |\lambda - \mu_H| . \tag{11.7}$$

Les estimations (11.6), (11.7) et la proposition 11.4 permettent de conclure. □

La dernière proposition correspond au cas du problème de recherche de point-selle mal posé dans sa version continue.

Proposition 11.6. On ne suppose pas ici B à image fermée, de telle sorte que (11.3) peut ne pas avoir de solution. On a alors

$$|u-u_h| \leq \left(1 + \frac{\|a\|}{\alpha}\right) \inf_{w_h \in K_h^H} |u - w_h| + \frac{1}{\alpha} \inf_{\mu_H \in \Lambda_H} \|\xi - B^\star \mu_H\|_{V'} ,$$

où ξ est la forme linéaire sur V définie par

$$a(u,v) + \langle \xi \, , v \rangle = \langle \varphi \, , v \rangle \quad \forall v \in V.$$

Démonstration : La forme ξ est telle que

$$a(u,v) + (\xi, \mu) = \langle f \, , v \rangle \quad \forall v \in V.$$

On peut reproduire la démonstration de la proposition 11.4 en remplaçant l'expression $(Bv_h, \lambda - \mu_H)$ par $\langle \xi \, , v_h \rangle - \langle B^\star \mu_H \, , v_h \rangle$, d'où l'on déduit le résultat. □

11.2.2. Cas général. Nous considérons pour finir une situation plus générale : V_h est toujours un sous-espace de dimension finie de V, mais la contrainte s'exprime $B_H u_h = 0$, où $B_H \in \mathcal{L}(V, \Lambda_H)$ n'est pas supposé égal à B, et Λ_H n'est pas nécessairement inclus dans Λ. On s'intéresse au problème suivant

$$(\mathcal{P}_h'') \qquad \left\{ \begin{array}{rcccll} a(u_h, v_h) & + & \langle B_H^\star \lambda_H \, , v_h \rangle & = & \langle \varphi \, , v_h \rangle & \forall v_h \in V_h \\ (\mu_H, B_H u_h) & & & = & 0 & \forall \mu_h \in \Lambda_H. \end{array} \right. \tag{11.8}$$

où B_H est un opérateur linéaire continu de V vers Λ_H. On définit, de façon analogue à ce qui précède,

$$K_{hH} = \{ v_h \in V_h \, , \ (B_H v_h, \mu_H) = 0 \quad \forall \mu_H \in \Lambda_H \} .$$

Comme B_H est à image fermée (car Λ_H est de dimension finie), le problème $\mathcal{P}_h''$ admet une solution $(u_h, \lambda_H) \in V_h \times \Lambda_H$. La partie primale $u_h \in V_h$ de cette solution est définie de façon unique comme l'élément de K_h^H qui minimise la fonctionnelle

$$v_h \longmapsto J(v_h) = \frac{1}{2} a(v_h, v_h) - \langle \varphi \, , v_h \rangle$$

sur K_h^H.

Proposition 11.7. On a

$$|u-u_h| \leq \left(1+\frac{\|a\|}{\alpha}\right) \inf_{w_h \in K_h^H} |u-w_h| + \frac{1}{\alpha} \inf_{\mu_H \in \Lambda_H} \|\xi - B_H^\star \mu_H\|_{V'},$$

où ξ est la forme linéaire sur V définie par

$$a(u,v) + \langle \xi\,,\, v\rangle = \langle \varphi\,,\, v\rangle \quad \forall v \in V.$$

Démonstration : La démonstration est très proche de celle de la proposition 11.6. Néammoins, comme il s'agit d'un résultat très utile en pratique, nous la développons dans son intégralité.

Comme u_h minimise J sur K_h^H, on a

$$a(u_h, v_h) = \langle \varphi\,,\, v_h\rangle \quad \forall v_h \in K_h^H.$$

Pour tout $w_h \in K_h^H$, on note $v_h = u_h - w_h \in K_h^H$. On a

$$a(v_h, v_h) = \langle \varphi\,,\, v_h\rangle - a(w_h, v_h).$$

Comme u est solution de $\mathcal{P}''$, on a notamment (prendre $v = v_h$)

$$a(u, v_h) + \langle \xi\,,\, v_h\rangle = \langle \varphi\,,\, v_h\rangle,$$

d'où

$$\begin{aligned} a(v_h, v_h) &= a(u-w_h, v_h) + \langle \xi\,,\, v_h\rangle - (B_H v_h, \mu_H) \\ &= a(u-w_h, v_h) + \langle \xi - B_H^\star \mu_H\,,\, v_h\rangle \end{aligned}$$

pour tout $\mu_H \in \Lambda_H$. On a donc

$$\alpha\,|u_h - w_h| \leq \|a\|\,|w_h - u| + \|\xi - B_H^\star \mu_H\|_{V'},$$

d'où

$$|u-u_h| \leq \left(1+\frac{\|a\|}{\alpha}\right)|w_h-u| + \frac{1}{\alpha}\,\|\xi - B_H^\star \mu_H\|_{V'} \quad \forall w_h \in K_h^H,\ \mu_H \in \Lambda_H,$$

d'où l'estimation annoncée. □

11.3. Condition inf-sup discrète

Proposition 11.8. (Lemme de Fortin)
On suppose que $B \in \mathcal{L}(V, \Lambda)$ vérifie la condition inf-sup continue. Alors la suite d'espaces (V_h, Λ_H) vérifie la condition inf-sup discrète uniforme si et seulement s'il existe un opérateur continu $\Pi_h \in \mathcal{L}(V, V_h)$ tel que

$$b(\Pi_h v - v, \mu_H) = 0 \quad \forall (v, \mu_H) \in V \times \Lambda_H,$$

et tel qu'il existe une constante (indépendante de h), telle que

$$|\Pi_h v| \leq C\,|v|.$$

DÉMONSTRATION : Condition nécessaire : pour $v \in V$, tout on construit l'élément v_h de norme minimale tel que $Bv_h = Bv$ (comme au début de la démonstration de la proposition 11.5). La formulation point-selle de ce problème s'écrit

$$\begin{aligned}(v_h, w_h) + (\lambda_H, Bw_h) &= 0 \quad \forall w_h \in V_h \\ (\mu_H, Bv_h) &= (\mu_H, Bv) \quad \forall \mu_H \in \Lambda_H.\end{aligned}$$

La condition inf-sup permet d'écrire $\lambda_H \leq |v_h| / \beta$, et l'on prend ensuite $w_h = v_h$ dans la première ligne pour établir $|v_h| \leq C\,|v|$. On définit $\Pi_h v = v_h$ (projection orthogonale sur l'image $B^\star(\Lambda_H)$.

Pour la condition suffisante, on se donne $\mu_H \in \Lambda_H$, et l'on note v son antécédent de norme minimale dans V, norme contrôlée par celle de μ_H du fait que B est à image fermée. On a alors

$$\sup \frac{b(v_h, \mu_H)}{|v_h|} \geq \frac{b(\Pi_h v, \mu_H)}{|\Pi_h v|} \geq \frac{1}{C}\,|\mu_H|\,,$$

d'où le résultat. □

Chapitre 12

Éléments d'analyse numérique matricielle

12.1. Définitions, préliminaires

Conditionnement. La notion de conditionnement d'une matrice (on parle aussi de conditionnement d'un système linéaire) joue un rôle très important dans l'étude de la résolution de systèmes linéaires. Nous verrons plus loin que ce conditionnement intervient notamment de façon essentielle dans le vitesse de convergence de méthodes de résolution itératives.

Le conditionnement d'une matrice apparaît de façon naturelle lorsque l'on cherche à estimer la stabilité de la résolution d'un système linéaire par rapport aux données, indépendamment de la méthode numérique utilisée effectivement pour résoudre le système. Considérons une matrice $A \in \mathcal{M}_n(\mathbb{R})$ inversible, un second membre $b \in \mathbb{R}^n$, et le système linéaire

$$Au = b.$$

Le conditionnement quantifie la confiance que l'on peut avoir dans la solution (exacte) de ce système en fonction de la confiance que l'on a dans les données (en l'occurrence le second membre b), qui sont susceptibles d'être entachées d'erreurs de mesure, d'erreurs liées au stockage sur ordinateur avec une précision finie. Dans ce qui suit nous considérons la norme matricielle $\|A\|_2$, notée simplement $\|A\|$, subordonnée à la norme euclidienne sur $\mathbb{R}^n$. On considère ainsi une perturbation δb du second membre, et l'on cherche à estimer la variation δu induite sur la solution :

$$A(u + \delta u) = b + \delta b.$$

On a donc $\delta u = A^{-1}\delta b$, d'où $|\delta u| \leq \left\|A^{-1}\right\| |\delta b|$. D'autre part $b = Au$ implique $|b| \leq \|A\| |\delta b|$, d'où finalement

$$\frac{|\delta u|}{|u|} \leq \left\|A^{-1}\right\| \|A\| \frac{|\delta b|}{|b|}.$$

Définition 12.1. (Conditionnement)
Soit A une matrice inversible. On appelle nombre de conditionnement de A le réel

$$\kappa = \left\|A^{-1}\right\| \|A\|.$$

La quantité κ mesure donc le rapport entre l'erreur relative maximale sur la solution et l'erreur relative sur les données. Cette quantité sans dimension est toujours supérieure ou égale à 1 ($1 = \|\mathrm{Id}\| = \left\|AA^{-1}\right\| \leq \kappa$). Pour $\kappa \gg 1$, le problème est très instable par rapport aux données.

Remarque 12.2. On peut aussi se demander quel est l'effet sur la solution d'une perturbation de la matrice elle-même :

$$(A + \delta A)(u + \delta u) = b.$$

On obtient au premier ordre (on néglige le terme en $\delta A \delta u$) une formule analogue à la précedente, qui fait intervenir le κ comme un majorant du facteur d'amplification de l'erreur relative :

$$\frac{|\delta u|}{|u|} \leq \left\|A^{-1}\right\| \|A\| \frac{\|\delta A\|}{\|A\|}.$$

Conditionnement des matrices s.d.p. Dans le cas où A est symétrique définie positive, de valeurs propres

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n,$$

le conditionnement s'écrit $\kappa = \lambda_n/\lambda_1$.

Exemple 12.3. Considérons la matrice du Laplacien discret donnée dans la section 13.4.2, dont les valeurs propres sont connues. Le conditionnement de cette matrice est donc

$$\kappa = \lambda_{N-1}/\lambda_1 = \frac{\sin^2\left(\frac{(N-1)\pi}{2N}\right)}{\sin^2\left(\frac{\pi}{2N}\right)} \sim 4N^2 \quad \text{quand } N \to +\infty.$$

Définition 12.4. Soit $A = (a_{ij})$ une matrice. On dit que A est une matrice-bande s'il existe ℓ tel que $a_{ij} = 0$ dès que $|j - i| > \ell$. Bien sûr cette notion n'a d'intérêt que si ℓ est significativement plus petit que n.

12.2. Méthodes directes

On s'intéresse dans cette section à la résolution d'un système linéaire $Au = b$ bien posé (matrice A inversible).

Décomposition LU. La décomposition LU est basée sur la méthode du pivot de Gauss. Elle consiste à effectuer une factorisation dite LU de la matrice (L pour *low*, U pour *low* :

$$A = LU$$

, où L (resp. U) est une matrice triangulaire inférieure (resp. supérieure), et L ne contient que des 1 sur la diagonale. Une fois que cette décomposition est réalisée, la solution s'obtient par résolution de 2 systèmes triangulaires.

Il peut être intéressant de choisir le pivot à chaque étape (pour éviter par exemple d'inverser des nombres trop petits). Il s'agit alors de la décomposition avec permutation :

$$A = PLU,$$

où P est une matrice de permutation (les éléments sont des 0 ou des 1, et chaque ligne et chaque colonne contient exactement un 1.

Méthode de Cholesky. La méthode de Cholesky est une forme particulière de décomposition LU tirant partie du caractère symétrique d'une matrice. Cette méthode consiste à décomposer une matrice symétrique définie positive en un produit de 2 matrices triangulaires transposées l'une de l'autre.

Algorithme 12.5. (Cholesky)
Soit $A = (a_{ij})$ une matrice symétrique définie positive de $\mathcal{M}_n(\mathbb{R})$. Alors la matrice triangulaire inférieure $L = (b_{ij})_{j \leq i}$ définie par

$$b_{11} = \sqrt{a_{11}}\,,\ b_{21} = a_{21}/b_{11}\,,\ \ldots\,,\ b_{n1} = a_{n1}/b_{11},$$

et, pour $j = 2, \dots, n$,

$$b_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} b_{jk}^2}\,,\ b_{ij} = \frac{a_{ij} - \sum_{k=1}^{j-1} b_{jk} b_{ik}}{b_{jj}}\,,\ i = j+1, \dots, n,$$

est telle que $A = L\,{}^tL$.

Le système $Au = b$ est alors résolu par la résolution successive des deux systèmes triangulaires

$$Lw = b\,,\ {}^tLu = w.$$

Proposition 12.6. La décomposition d'une matrice A s.d.p. de taille $n \times n$ par la méthode de Cholesky nécessite n extractions de racines, et un équivalent de $n^3/6$ divisions ou multiplications.

La résolution du système linéaire $Au = b$ par cette méthode nécessite en outre, pour la résolution des deux systèmes triangulaires, l'équivalent de n^2 opérations élémentaires (multiplications ou divisions).

Démonstration : Le nombre d'extraction de racines est bien égal à n. Pour le nombre de multiplications/divisions, on cherche directement un équivalent. La première étape n'est donc pas prise en compte. Le gros du coût est dans le calcul de chacun des éléments extradiagonaux b_{ij}, au nombre de $n-j$ pour j fixé, qui nécessite (on ne garde que l'essentiel) j mutiplications. La complexité est donc en

$$\sum_j (n-j)j,$$

qui est un $\mathcal{O}(n^3)$, avec le coefficient $1/6$ (penser à $\int x(1-x) = 1/6$).

La résolution d'un système triangulaire consiste à effectuer, pour tout $j = 1, \dots, n$, j multiplications et une division. On a donc une complexité en $n^2/2$ pour chacun des systèmes triangulaires. □

Remarque 12.7. La complexité réelle est en général très inférieure (tout du moins si l'écriture du programme informatique est adaptée à la situation), notamment dans le cas des matrices-bande (voir définition 12.4 ci-dessus), ce qui est souvent le cas des matrices résultants de la discrétisation par éléments finis d'un opérateur elliptique. Dans ce cas, on peut montrer que la matrice L associée possède la même structure de matrice bande. En conséquence, pour j allant de 2 à n, le nombre d'éléments extradiagonaux b_{ij} chute de $n - j$ à ℓ, tout comme le nombre d'opérations nécessaire. La complexité descend donc à $n\ell^2$. Noter que la résolution des 2 systèmes triangulaires, dont la complexité chute à $n\ell$, reste d'un coût négligeable par rapport à la factorisation (au moins dans de la cas d'un seul système, voir à ce sujet la remarque 12.8). Dans le cas du Laplacien discret en dimension 1, la largeur de bande est 2, d'où une complexité de l'ordre de n, le nombre de points (nous ne précisons pas la constante, car la petite largeur de bande rend significatives des opérations dont nous avions négligé le nombre). En dimension 2, pour un problème scalaire sur un maillage $\sqrt{n} \times \sqrt{n}$, la matrice est de taille n, et de largeur de bande $\sqrt{n}$, d'où une compléxité en $n^2/6$.

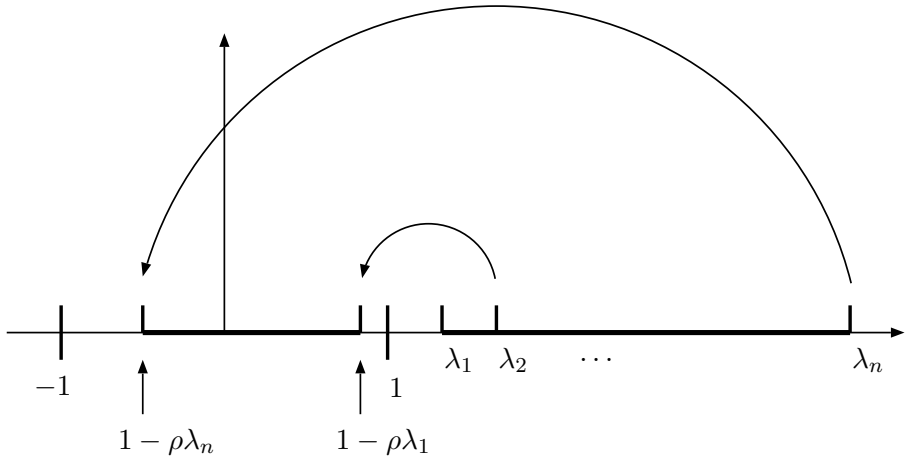


FIGURE 1. Spectre de $\mathrm{Id} - \rho A$

Remarque 12.8. Cette méthode peut être particulièrement performante lorsque l'on souhaite résoudre un grand nombre de fois un système [1] impliquant une matrice donnée A (pour des seconds membres distincts). Notons M ce nombre de systèmes à résoudre. la complexité totale est de $n^3/6 + Mn^2$, de telle sorte que dans la situation extrême où n devient négligeable devant M, on a une complexité asymptotique de la méthode en n^2 (coût unitaire d'une résolution de système).

12.3. Méthodes itératives

12.3.1. Méthode du gradient à pas fixe (Richardson).

Algorithme 12.9. Soit A une matrice symétrique définie positive de $\mathcal{M}_n(\mathbb{R})$. L'algorithme du gradient à pas fixe est basé sur la construction suivante : on se donne $\rho > 0$, un vecteur initial $u^0 \in \mathbb{R}^n$, et l'on construit

$$u^{k+1} = u^k - \rho(Au^k - b).$$

Proposition 12.10. L'algorithme du gradient à pas fixe converge dès que $\rho \in]0, 2/\lambda_n[$, où λ_n est la plus grande valeur propre de A

DÉMONSTRATION : On note $e^k = u^k - u$ l'erreur, qui vérifie $e^{k+1} = (\mathrm{Id} - \rho A)e^k$. Cette erreur converge dès que les valeurs propres de $\mathrm{Id} - \rho A$ sont de module strictement inférieur à 1. L'opération $A \mapsto \mathrm{Id} - \rho A$ renverse le spectre de A comme illustré sur la figure 1. Les valeurs propres de la nouvelle matrice sont donc de module strictement inférieur à 1 si et seulement si $1 - \rho\lambda_n > -1$, c'est à dire $0 < \rho < 2/\lambda_n$. □

Remarque 12.11. Bien que la notion de choix optimal pour ρ soit sujette à caution, on notera que le choix

$$\rho = 2/(\lambda_1 + \lambda_n)$$

1. Cette situation se rencontre par exemple dans le cadre de la discrétisation en temps d'un problème d'évolution par une méthode implicite, qui se ramène à chaque pas de temps à la résolution d'un système pour une même matrice mais des seconds membres différents.

minimise le rayon spectral de $\mathrm{Id} - \rho A$. Pour ce choix, le rapport géométrique de convergence est $1 - 2\lambda_1/(\lambda_1 + \lambda_n)$, donc de l'ordre de $1 - 2\kappa^{-1}$ pour κ grand. La convergence sera donc d'autant plus lente que le conditionnement κ est grand.

12.3.2. Méthode du gradient à pas optimal. La méthode du gradient à pas optimal est basée sur un calcul explicite du pas ρ de l'algorithme de gradient ci-dessus, de façon à minimiser la valeur de la fonctionnelle J sur la droite $\left\{u^k - \rho(Au^k - b)\,,\ \rho \in \mathbb{R}\right\}$. Un simple calcul permet d'exprimer ce ρ optimal à chaque itération :

Algorithme 12.12. Soit A une matrice symétrique définie positive de $\mathcal{M}_n(\mathbb{R})$. L'algorithme du gradient à pas optimal est basé sur la construction suivante : on se donne un vecteur initial $u^0 \in \mathbb{R}^n$, et l'on construit

$$u^{k+1} = u^k - \rho_k(Au^k - b)\,,\ \rho_k = \frac{\left|b - Au^k\right|^2}{\left|b - Au^k\right|_A^2}\,,\quad \text{avec } |v|_A^2 = (Av, v)\,.$$

Remarque 12.13. Noter que ρ_k est minoré et majoré, pour toute matrice s.d.p. A donnée.

12.3.3. Méthode du gradient conjugué. La méthode du gradient conjugué permet d'approcher numériquement la solution de problèmes du type $Ax = b$, où A est une matrice symétrique définie positive. Nous verrons qu'en fait il s'agit d'une méthode exacte (qui converge en un nombre d'itérations fini égal à la dimension de l'espace), mais elle est dans la pratique utilisée comme un algorithme itératif.

Algorithme 12.14. Soit A une matrice symétrique définie positive de $\mathcal{M}_n(\mathbb{R})$. L'algorithme du gradient conjugué est basé sur la constrution itérative suivante, à partir d'un vecteur initial $u_0 \in \mathbb{R}^n$. On définit tout d'abord le résidu initial correspondant $r_0 = b - Au_0$, et l'on pose $p_0 = r_0$,

$$\begin{aligned}
\alpha_k &= \frac{|r_k|^2}{(Ap_k, p_k)}\\
u_{k+1} &= u_k + \alpha_k p_k\\
r_{k+1} &= r_k - \alpha_k A p_k\\
\beta_{k+1} &= |r_{k+1}|^2 \,/\, |r_k|^2\\
p_{k+1} &= r_{k+1} + \beta_{k+1} p_k.
\end{aligned}$$

Proposition 12.15. Les suites (r_k), (p_k) construites selon l'algorithme du gradient conjugué 12.14 vérifient les propriétés suivantes :

$$(r_k, p_i) = (r_k, r_i) = 0 \quad \forall i \leq k-1\,,\ (p_k, Ap_i) = 0 \quad \forall i \leq k-1\,,\ |r_{k+1}|_{A^{-1}} \leq |r_k|_{A^{-1}}\ ,$$
$$|r_k|_{A^{-1}} = \min_{F_k} |b - Au|_{A^{-1}}\ ,\ F_k = u_0 + \mathrm{vect}(p_0, \ldots, p_{k-1}).$$

Démonstration : On démontre ces propriétés par récurrence. On a

$$(r_{k+1}, r_k) = |r_k|^2 - \alpha_k(r_k, Ap_k) = |r_k|^2 - \alpha_k(p_k - \beta_k p_{k-1}, Ap_k) = |r_k|^2 - \alpha_k(p_k -, Ap_k) = 0.$$

Pour tout $i \leq k-1$, on a

$$(r_{k+1}, r_i) = (r_k - \alpha_k Ap_k, r_i) = -\alpha_k(Ap_k, r_i).$$

Comme $r_i = p_i - \beta_i p_{i-1}$, le produit scalaire est nul du fait que les directions p_j sont deux à deux conjuguées pour $j \leq k$ (hypothèse de récurrence).

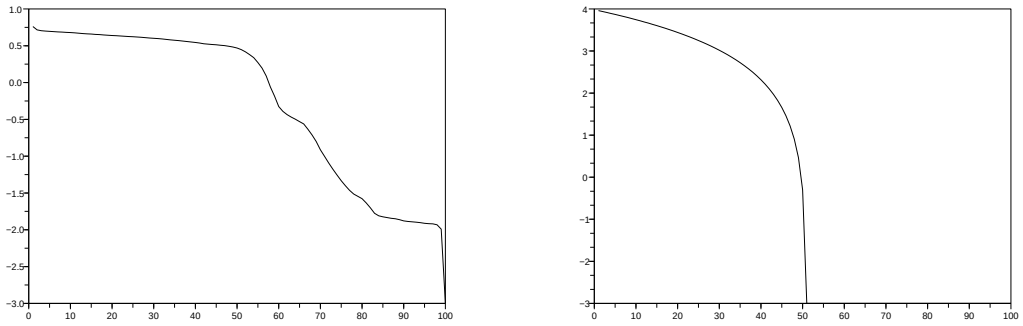


FIGURE 2. Log du résidu au cours des itérations

On a de même $(r_{k+1}, p_i) = 0$ pour tout $i \leq k$, car p_i s'exprime en fonctions des r_j, pour $j \leq i$.

Pour la conjugaison des directions de descente, on a

$$(p_{k+1}, Ap_k) = (r_{k+1} + \beta_{k+1} p_k, (r_k - r_{k+1})/\alpha_k),$$

ce qui donne (on utilise $(r_{k+1}, r_k) = (p_k, r_{k+1}) = 0$)

$$(p_{k+1}, Ap_k) = -\frac{1}{\alpha_k}\left(|r_{k+1}|^2 + \beta_{k+1}(p_k, r_k)\right) = 0$$

car $(p_k, r_k) = |r_k|^2$, et $\beta_{k+1} = |r_{k+1}|^2 / |r_k|^2$.

□

Proposition 12.16. Soit A une matrice symétrique définie positive, et (u_k) une suite d'itérés produite par l'algorithme du gradient conjugué 12.14. On note $|\cdot|_A$ la norme associée à la matrice A, et $\kappa = \lambda_n/\lambda_1$ le conditionnement de A. On a

$$|u_k - u|_A \leq 4\,|u_0 - u|_A \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k.$$

Corollaire 12.17. La norme de l'erreur vérifie

$$|u_k - u| \leq 4\kappa\,|u_0 - u| \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k.$$

Remarque 12.18. Pour de grands nombres de conditionnement, on a une convergence géométrique de rapport voisin de $1 - 2/\sqrt{\kappa}$. On remarquera que ce taux est bien meilleur que celui trouvé pour la méthode de gradient à pas fixe (égal à $1-2/\kappa$, voir remarque 12.11).

Remarque 12.19. La convergence étant géométrique de rapport $1 - 2/\sqrt{\kappa}$, le nombre d'itérations à réaliser pour être sûr d'avoir une précision donnée ε est de l'ordre de $k_\varepsilon = \sqrt{\kappa}\ln(1/\varepsilon)$, contre $\kappa\ln(1/\varepsilon)$ pour le gradient à pas fixe. Le gain potentiel en termes de temps de calcul est donc considérable. Pour la résolution du Laplacien en dimension 1, avec $N = 100$ points, le conditionnement est de l'ordre de 10^4 (voir exemple 12.3, page 150), et le calcul par gradient conjugué va 100 fois plus vite que le calcul par gradient simple.

Le comportement effectif du gradient conjugué dépend très sensiblement de la matrice bien sûr, mais aussi du second membre considéré. La figure 2 représente le logarithme de l'erreur au cours des itérations, pour la matrice du Laplacien discret d'ordre 100, pour un second membre obtenu comme N réalisations indépendantes d'une variable aléatoire de loi uniforme sur $[0,1]$ (figure de gauche), puis pour un second membre dont tous les éléments sont égaux à 1 (figure de droite). Dans le premier cas, sur la première moitié du parcours, la convergence est géométrique de rapport $1-0.014$. Le conditionnement de la matrice est de l'ordre de 10^4, ce qui donne un ordre théorique de $1-0.02$, proche de l'ordre effectif. Noter qu'en revanche après l'itération 50 la convergence est beaucoup plus rapide. Ce phénomène est encore plus net pour un second membre « non quelconque », puisqu'on obtient la précision machine après 50 itérations. Par ailleurs, si la pente pour les premières itérations correspond à peu près à la pente théorique, la convergence ne cesse d'accélérer.

12.4. Méthodes rapides

Le terme de méthode rapide fait référence à des algorithmes particuliers permettant de limiter le nombre d'opérations élémentaires pour réaliser (sans approximation) un calcul donné.

L'exemple le plus simple est le calcul d'une puissance entière d'un nombre réel (ou entier). Calculer x à la puissance 8 requiert a priori 7 multiplications. Mais on peut aussi calculer x^2, multiplier le résultat par lui-même, et encore une fois le résultat par lui-même, pour calculer le même nombre en 3 multiplications.

Dans le même esprit, le calcul de la valeur d'un polynôme

$$a_0 + a_1 X + \cdots + a_n X^n$$

en un point x peut s'écrire

$$(\ldots((a_n x + a_{n-1})x + a_{n-2}) + \cdots + a_1)x + a_0,$$

ce qui permet de limiter le nombre de multiplications à n (algorithme de Horner).

Tranformée de Fourier rapide (dimension 1). Pour ce qui concerne la résolution de problèmes du type de ceux rencontrés, nous nous contentons de donner ici le principe[2] d'une méthode permettant de résoudre rapidement (dans un sens que nous préciserons) des systèmes linéaires du type de ceux résultants de la discrétisation du Laplacien sur un maillage cartésien. Il s'agit de la méthode de transformée de Fourier rapide (*Fast Fourier Transform*). En dimension 1, la discrétisation en espace du problème de Poisson avec condition de Dirichlet homogène

$$-u'' = f\,,\ u(0) = u(1) = 0,$$

conduit à un système linéaire du type

$$Au = b,$$

2. De nombreuses améliorations sont possibles, qui permettent d'accélérer encore le calcul, mais l'approche basique que nous présentons ici donne l'ordre de grandeur de la complexité, c'est à dire du nombre d'opérations nécessaire à la résolution du problème.

où A est à une constante multiplicative près ($1/h = N$ en l'occurrence) la matrice du Laplacien discret (voir (13.4), page 165). Cette matrice est symétrique, donc diagonalisable dans une base orthogonale de vecteurs propres. On peut expliciter les éléments propres de cette matrice (voir section 13.4.2), ce qui permet d'écrire

$$A = PDP^t\,,\ D = \operatorname{diag}\left(4\sin^2\left(\frac{k\pi}{2N}\right)\right)_{k=1,\dots,N-1},$$

et

$$P = \sqrt{\frac{2}{N}}\begin{pmatrix} \sin\left(\frac{\pi}{N}\right) & \sin\left(\frac{2\pi}{N}\right) & \sin\left(\frac{3\pi}{N}\right) & \cdot & \cdot & \sin\left(\frac{(N-1)\pi}{N}\right) \\ \sin\left(\frac{2\pi}{N}\right) & \sin\left(\frac{4\pi}{N}\right) & \sin\left(\frac{6\pi}{N}\right) & \cdot & \cdot & \cdot \\ \sin\left(\frac{3\pi}{N}\right) & \sin\left(\frac{6\pi}{N}\right) & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & & \cdot \\ \sin\left(\frac{(N-1)\pi}{N}\right) & \cdot & \cdot & \cdot & \sin\left(\frac{(N-2)(N-1)\pi}{N}\right) & \sin\left(\frac{(N-1)^2\pi}{N}\right) \end{pmatrix}$$

La résolution du problème $Au = b$ se ramène donc (on utilise $P = P^t = P^{-1}$) au calcul de $u = PD^{-1}Pb$. Il s'agit donc de 2 produits matrice-vecteur et de la multiplication par une matrice diagonale. Le cœur de la méthode réside dans la manière d'effectuer le produit Pb (et de la même manière Pc avec $c = D^{-1}Pb$). On introduit le vecteur $\tilde{b} = \mathbb{R}^{2N}$ construit de la façon suivante

$$\tilde{b} = \left(\tilde{b}_0, \dots, \tilde{b}_{2N-1}\right) = (0, b_1, b_2, \dots, b_{N-1}, 0, -b_{N-1}, -b_{N-2}, \dots, -b_1).$$

On a

$$\begin{aligned}\sqrt{\frac{N}{2}}\,(Pb)_k &= \sum_{\ell=1}^{N-1}\sin\left(\frac{kl\pi}{N}\right)b_\ell = \frac{1}{2}\left(\sum_{\ell=1}^{N-1}\sin\left(\frac{2kl\pi}{2N}\right)b_\ell - \sum_{\ell=1}^{N-1}\sin\left(\frac{2k(2N-\ell)\pi}{2N}\right)b_\ell\right)\\ &= \frac{1}{2}\sum_{\ell=0}^{2N-1}\sin\left(\frac{2kl\pi}{2N}\right)\tilde{b}_\ell = \frac{i}{2}\sum_{\ell=0}^{2N-1}\exp\left(-\frac{2ikl\pi}{2N}\right)\tilde{b}_\ell = \frac{i}{2}\sum_{\ell=0}^{2N-1}\omega_{2N}^{k\ell}\tilde{b}_\ell,\end{aligned}$$

avec

$$\omega_{2N} = \exp\left(-\frac{2i\pi}{2N}\right).$$

Le k-ième coefficient de Pb (au facteur $\sqrt{N/2}$ près) est donc le k-ième coefficient de ce que l'on appelle la transformée de Fourier discrète (d'ordre $2N$, avec indexation de 0 à $2N-1$) du vecteur $\tilde{b}$. On note $\mathcal{F}$ cette transformée de Fourier discrète, de telle sorte que

$$\sqrt{\frac{N}{2}}\,(Pb)_k = \left(\mathcal{F}_{2N}(\tilde{b})\right)_k.$$

La somme ci-dessus peut se décomposer de la façon suivante (on sépare les termes impairs et les termes pairs) :

$$\begin{aligned}\sum_{\ell=0}^{2N-1}\omega_{2N}^{k\ell}\tilde{b}_\ell &= \sum_{\ell=0}^{N-1}\omega_{2N}^{2\ell k}\tilde{b}_{2\ell} + \sum_{\ell=0}^{N-1}\omega_{2N}^{(2\ell+1)k}\tilde{b}_{2\ell+1} = \sum_{\ell=0}^{N-1}\omega_N^{\ell k}\tilde{b}_{2\ell} + \omega_{2N}^k\sum_{\ell=0}^{N-1}\omega_N^{\ell k}\tilde{b}_{2\ell+1}\\ &= \mathcal{F}_N(\tilde{b}^0)_k + \omega_{2N}^{-k}\mathcal{F}_N(\tilde{b}^1)_k.\end{aligned}$$

où b^0 (resp. b^1) est le vecteur des termes pairs (resp. impairs) de $\tilde{b}$. Précisons que si k est plus grand que N (c'est a priori inutile ici, mais c'est utile pour la suite), on obtient

$$\mathcal{F}_N(\tilde{b}^0)_{k-N} + \omega_{2N}^{-k}\mathcal{F}_N(\tilde{b}^1)_{k-N}.$$

Supposons que l'on sache calculer tous les termes des deux transformées ci-dessus (vecteurs de taille N). On doit effectuer de l'ordre de N multiplications complexes (on néglige ici les constantes multiplicatives). Si N est une puissance de 2, on peut ainsi récursivement calculer les TFD aux différentes échelles, le coût du passage d'une étape à l'autre étant à chaque fois de l'ordre de $2N$. Le nombre d'étape étant de l'ordre de $\log_2 N$, le coût total est de l'ordre de $N \log_2 N$.

Le lecteur avide de curiosités pourra se reporter à la section 13.13 pour une présentation de ces principes dans le cadre de la transformée de Fourier sur l'espace $\mathbb{Z}_2$ des entiers dyadiques.

Tranformée de Fourier rapide (dimension 2). On considère maintenant le problème de Poisson en dimension 2 sur un maillage cartésien du carré unité, avec $N+1$ points dans chaque direction, y compris les points au bord, donc au total $(N-1)^2$ degrés de liberté. On note u_{ij} la valeur de la solution approchée au point (ih, jh) (avec $h = 1/N$). Le système résultant de la discrétisation par éléments finis du problème s'écrit

$$Au = b,$$

où $A \in \mathcal{M}_{(N-1)^2}(\mathbb{R})$ peut s'écrire par blocs (avec $B \in \mathcal{M}_{N-1}(\mathbb{R})$)

$$A = \begin{pmatrix} C & -\mathrm{Id} & 0 & \cdot & \cdot & 0 \\ -\mathrm{Id} & C & -\mathrm{Id} & 0 & \cdot & 0 \\ 0 & -\mathrm{Id} & C & -\mathrm{Id} & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & & -\mathrm{Id} \\ 0 & \cdot & \cdot & 0 & -\mathrm{Id} & C \end{pmatrix}, \; C = \begin{pmatrix} 4 & -1 & 0 & \cdot & \cdot & 0 \\ -1 & 4 & -1 & 0 & \cdot & 0 \\ 0 & -1 & 4 & -1 & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & & -1 \\ 0 & \cdot & \cdot & 0 & -1 & 4 \end{pmatrix},$$

et u est le vecteur des inconnues

$$u = (u_{11}, u_{21}, \ldots, u_{N-1,1}, u_{1,2}, \ldots, u_{N-1,N-1})^T$$

On introduit les vecteurs colonne u_i correspondant aux inconnues sur la ligne verticale $x = ih$, et les vecteurs ligne u^j, correspondant aux inconnues sur la ligne horizontale $y = jh$ (voir figure 3), ce qui permet d'écrire le vecteur u sous la forme d'une matrice $(u_1, \ldots, u_{N-1}$ (on a une écriture analogue en lignes).

On introduit maintenant la matrice du Laplacien discret (voir (13.4) que l'on note ici Λ. On a (en utilisant une indexation (i,j) pour représenter les vecteurs de $\mathbb{R}^{(N-1)^2}$)

$$(Au)_{i,j} = (\Lambda u_i)_j + \left(\Lambda u^j\right)_i. \tag{12.1}$$

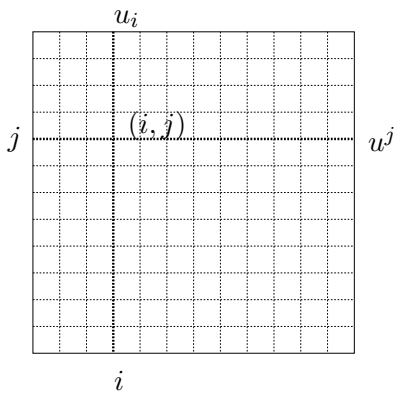


FIGURE 3. Maillage cartésien

On cherche à réécrire le système de façon plus ramassée en écrivant le vecteur des inconnues sous forme de matrice (deux écriture sont possibles, en colonnes et en lignes)

$$U = (u_1, \dots, u_{N-1}) = \begin{pmatrix} u^1 \\ u^2 \\ \vdots \\ u^{N-1} \end{pmatrix},$$

on écrit de la même manière le second membre sous la forme d'une matrice B, et l'on remarque

$$\Lambda U = (\Lambda u_1, \dots, \Lambda u_{N-1}) \ , \ U\Lambda = (\Lambda^T U^T)^T = \begin{pmatrix} \Lambda u^1 \\ \Lambda u^2 \\ \vdots \\ \Lambda u^{N-1} \end{pmatrix}.$$

Le système $(Au)_{i,j} = B_{i,j}$ peut donc s'écrire, d'après (12.1), sous la forme suivante :

$$\Lambda U + U\Lambda = B.$$

Or on a vu que la matrice Λ est diagonalisable (avec une matrice de passage orthogonale et symétrique : $\Lambda = PDP$. On a donc (en multipliant à gauche et à droite par P, et en utilisant $P^2 = \mathrm{Id}$)

$$DPUP + PUPD = PBP.$$

On introduit la matrice $W = PUP$. On s'est finalement ramené au calcul de $B' = PBP$, de la résolution d'un problème du type

$$DW + WD = B' \Longleftrightarrow W_{ij} = \frac{1}{\lambda_i + \lambda_j} B'_{ij},$$

où les λ_i sont connus (voir section 13.4.2), et finalement de $U = PWP$. En dehors de l'étape centrale, pour laquelle on a une formule explicite, il s'agit donc d'effectuer des produits matrice-vecteur du type PX ou XP. Le premier produit consiste en le calcul de la transformée de Fourier discrète (donc potentiellement *rapide*) des vecteurs colonnes de X, et le second $XP = (PX^T)^T$ la TFD des vecteurs lignes de X. Dans les deux cas le calcul par FFT donne une complexité de l'ordre de $N \times N \log_2 N$. On a donc finalement un nombre d'opérations de l'ordre de $m \log_2 m$, où $m = (N-1)^2$ est le nombre d'inconnues.

12.5. Préconditionnement

Les sections précédentes mettent en évidence l'importance du conditionnement dans la rapidité de résolutions des systèmes linéaires, lorsque l'on utilise des méthodes itératives (les plus utilisées dans le cas de grand systèmes linéaires). Il peut être très efficace de remplacer le système $Au = b$ par un système dit préconditionné

$$C^{-1}Au = C^{-1}b.$$

On pourra améliorer très significativement la vitesse de convergence des méthodes si l'on est capable de trouver une matrice C spectralement proche de 1, de telle sorte que le conditionnement de $C^{-1}A$ est très inférieur à celui de A. Pour que cette approche soit efficace, il faut bien sûr que la matrice C soit plus facile à inverser que A.

Un très grand nombre de stratégies sont possibles, parmi lesquelles

(1) Préconditionnement diagonal. On prend pour C la matrice diagonale constituée des éléments diagonaux de A. L'inversion de C est alors immédiate, mais l'on vérifie aisément que cette approche est sans intérêt dans certaines situations, par exemple si A est la matrice du Laplacien discrétisé sur maillage cartésien (C est alors proportionnelle à l'identité, de telle sorte que l'on ne change pas le conditionnement de la matrice. En revanche, cette approche peut être féconde dans le cas de maillage très irréguliers, en particuliers lorsque la matrice à inverser est du type $\alpha M + A$, où M est la matrice de masse. Cette approche simpliste peut aussi être efficace dans le cas où la matrice A résulte de la discrétisation par élements finis d'une formulation pénalisée d'un problème sous contrainte.

(2) Décomposition incomplète. Dans ce cas, C est construit en effectuant de façon incomplète la décomposition (par exemple de Cholesky) de la matrice A.

Chapitre 13

Compléments

13.1. Triplet de Gelfand

Nous considérons ici la situation suivante :

$$\left|\begin{array}{l} (V, \|\cdot\|) \text{ et } (H, |\cdot|) \text{ espaces de Hilbert} \\ V \subset H \text{ avec injection continue} \\ V \text{ strictement inclus dans } H, \text{ dense dans } H \\ T \ : \ u \in H \longmapsto Tu \in V' \text{ défini par } \langle Tu \, , w\rangle = (u, w) \quad \forall w \in V. \end{array}\right. \tag{13.1}$$

On rencontrera souvent dans la pratique des situations où l'inclusion de V dans H est de plus compacte (la caractère strict de l'inclusion en est alors une conséquence), mais nous n'aurons pas besoin de cette hypothèse ici.

On notera H' le dual topologique de H pour la norme $|\ |$, et $(\cdot, \cdot)$ le produit scalaire sur H. Nous n'introduisons pas de notation spécifique pour le produit scalaire sur V, le principe même de la présente démarche étant la description du dual de V sans utiliser ce produit scalaire. Ce qui suit a pour but de donner un sens à la chaîne d'inclusions

$$V \subset H \subset V'.$$

La continuité de l'injection de V dans H se traduit par l'existence d'une constante C telle que

$$|x| \leq C \, \|x\| \quad \forall x \in X.$$

Proposition 13.1. Sous les hypothèses (13.1), l'application T est linéaire continue de H dans V', et injective.

Démonstration : Pour tout u dans H, w dans V, on a

$$|\langle Tu \, , w\rangle| = |(u, w)| \leq |u| \, |w| \leq C \, |u| \, \|w\| \, ,$$

d'où $Tu \in V'$. D'autre part $Tu = 0$ implique $(u, w) = 0$ pour tout w dans V, d'où $u = 0$ car V est dense dans H.

Noter que T est l'adjoint de l'opérateur d'injection de V dans H (où l'on a identifié H avec son dual). □

Proposition 13.2. Sous les hypothèses (13.1), l'application T n'est pas surjective.

Démonstration : D'après la définition de T, l'adjoint de T est l'injection de V dans H. D'après la proposition 7.20, la surjectivité de T entrainerait l'existence d'une constante α

telle que

$$\|w\| \leq \alpha \, |T^\star w| = \alpha \, |w| \, ,$$

d'où l'on déduirait que les normes $\|\cdot\|$ et $|\cdot|$ sont équivalentes sur V, et qu'ainsi V est à la fois fermé et dense dans H, ce qui est en contradiction avec l'hypothèse de stricte inclusion.

Proposition 13.3. Sous les hypothèses (13.1), l'image de T est dense dans V'.

Démonstration : On utilise la caractérisation 6.18, page 71. Soit $w \in V$ tel que

$$\langle Tu \, , w \rangle = (u, w) = 0 \quad \forall u \in H.$$

On a nécessairement $w = 0$, d'où la densité de $T(H)$.

La proposition suivante permet de caractériser les éléments de l'image de T.

Proposition 13.4. On se place toujours sous les hypothèses (13.1). Soit $\psi \in X'$. On a

$$\psi \in T(H) \Longleftrightarrow \exists C' > 0 \, , \ |\langle \psi \, , w \rangle| \leq C' \, |w| \quad \forall w \in w.$$

Démonstration : La condition nécessaire est immédiate. Réciproquement, si ψ vérifie l'inégalité, alors elle se prolonge par densité en une forme linéaire continue sur H, que l'on peut ainsi identifier à un élément $u \in H$, dont ψ est l'image par construction. □

Exercice 13.1. On considère

$$H = \ell^2 \, , \ V = \left\{ u = (u_n)_{n \geq 1} \in \ell^2 \, , \ \sum n u_n^2 \right\} .$$

Montrer que l'on vérifie bien le jeu d'hypothèses (13.1) (en précisant le produit scalaire dont on munit V). Vérifier que l'injection $V \subset H$ est en outre compacte. Construire un élément de V' qui n'est pas dans $T(H)$.

13.2. Éléments d'analyse spectrale, équations d'évolution

Théorème 13.5. Soient V et H deux espaces de Hilbert, de dimension infinie, avec injection $V \subset H$ compacte et dense. Soit $a(\cdot, \cdot)$ une forme bilinéaire symétrique continue et coercive. Le problème de recherche d'un couple $(u, \lambda) \in H \times \mathbb{R}$ tel que

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

admet une infinité de solutions. Les λ solutions, appelées valeurs propres de a, forment une suite

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \ldots$$

qui tend vers l'infini. Les fonctions propres (u_k) associées, normalisée à 1 pour H, forment une base Hilbertienne de H. De plus, $(u/\sqrt{\lambda_k})$ est une base Hilbertienne de V pour le produit scalaire associé à $a(\cdot, \cdot)$.

Dans le contexte du théorème précédent, on définit pour tout $v \in V$ le quotient de Rayleigh par

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

Théorème 13.6. (Courant-Fisher)
On se place dans les hypothèse du théorème 13.5. On note E_k l'ensemble des sous-espaces vectoriels de V de dimension k. On a

$$\lambda_k = \min_{W\in E_k} \max_{w\in W\setminus\{0\}} R(w) = \max_{W\in E_{k-1}} \min_{w\in W^\perp\setminus\{0\}} R(w).$$

Théorème 13.7. Soient V et H deux espaces de Hilbert, de dimension infinie, avec injection $V \subset H$ compacte et dense. Soit $a(\cdot,\cdot)$ une forme bilinéaire symétrique continue et coercive, et $f \in L^2(]0,T[,H)$ un terme source. On se donne une donnée initiale $u_0 \in H$. Le problème

$$\frac{d}{dt}(u(t),v) + a(u(t),v) = (f(t),v) \quad \forall v \in V\,,\ 0<t<T$$

avec condition initiale $u(0)=u_0$, a une unique solution $u \in L^2(]0,T[,V)) \cap C([0,T],H)$.

Théorème 13.8. Soient V et H deux espaces de Hilbert, de dimension infinie, avec injection $V \subset H$ compacte et dense. Soit $a(\cdot,\cdot)$ une forme bilinéaire symétrique continue et coercive sur V, et $f \in L^2(]0,T[,H)$ un terme source. On se donne une donnée initiale $(u_0,u_1) \in V\times H$. Le problème

$$\frac{d^2}{dt^2}(u(t),v) + a(u(t),v) = (f(t),v) \quad \forall v \in V\,,\ 0<t<T$$

avec conditions initiales $u(0)=u_0$, $du/dt(0)=u_1$, a une unique solution $u \in C([0,T],V))\cap C^1([0,T],H)$.

13.3. Schémas numériques pour les équations d'évolution

On s'intéresse ici aux problèmes d'évolution d'ordre 1 en temps du type

Trouver $u \in C([0,T],H)$ solution [1] de

$$\frac{du}{dt} + Au = f, \tag{13.2}$$

avec condition initiale $u(0)=u_0 \in H$, où H est un espace de Hilbert, et A est un opérateur linéaire (opérateur différentiel pour les cas qui nous intéressent ici). On s'intéressera au cas où le problème est bien posé (on peut par exemple supposer que l'on se place dans le cadre du théorème 13.7 ci-dessus).

On considère maintenant une classe générale de schéma d'approximation en temps et en espace : on notera δt le pas de temps et (V_h) une suite de sous-espaces de H, avec $V_h \subset h$ pour tout h. On suppose que les espaces V_h approchent H au sens suivant : il existe un opérateur de projection R_h tel que, pour tout $u \in H$,

$$|\ R_h u - u|_H \longrightarrow 0 \text{ quand } \ h \to 0.$$

On considère la classe générale de procédés constructifs suivante :

$$U_h^{n+1} = C_h(\delta t)U_h^n + \delta t f_h^n\,,\ U_h^0 = U_{0,h}, \tag{13.3}$$

1. Il s'agit de solutions dans un sens faible : on demande que

$$-\int_0^T u(s)\varphi'(s)\,ds - u_0\varphi(0) + \int_0^T \varphi(s)Au(s) = \int_0^T f(s)\varphi(s)\,ds,$$

pour toute fonction C^∞ à support compact dans $[0,T[$.

où $C_h(\delta t) \in \mathcal{L}(V_h)$.

Définition 13.9. (Stabilité)
On dit que le schéma (13.3) est (inconditionnellement) stable s'il existe une constante K telle que

$$\|(C_h(\delta t))^n R_h\|_{\mathcal{L}(H)} \leq K \quad \forall n\,,\ \delta t \text{ avec } \ n\delta t \leq T.$$

Dans le cas où l'inégalité n'est vérifiée que sous certaines contraintes sur h et δt, on dira que le schéma est conditionnellement stable.

Définition 13.10. (Consistence)
On dit que le schéma (13.3) est consistent s'il existe un sous-espace $D \subset V$ dense dans V tel que, pour toute solution u de (13.2) avec $u_0 \in D$, on a

$$\lim_{h,\delta t\to 0} \sup_t \left| \frac{u(t+\delta t) - C_h(\delta t) R_h u(t)}{\delta t} \right|_H = 0$$

Définition 13.11. (Convergence)
On suppose que le problème (13.2) est bien posé et l'on note $u(t)$ sa solution. On dit que le schéma (13.3) est convergent si la convergence de $U_{0,h}$ vers u_0 implique la convergence des approximations H_h^n, i.e.

$$U_h^n \longrightarrow u(t) \text{ quand } \delta t\,,\ h \to 0 \text{ avec } \ n\delta t \to t.$$

Théorème 13.12. On suppose que le problème (13.2) est bien posé. Alors le schéma (13.3) est convergent si et seulement s'il est stable et consistent.

13.4. Valeurs propres, vecteurs propres

13.4.1. Estimation des valeurs propres.

Théorème 13.13. (Courant-Fisher)
Soit A une matrice symétrique, de valeurs propres

$$\lambda_1 \leq \cdots \leq \lambda_N.$$

On a

$$\lambda_k = \inf_{\dim E = k} \sup_{v\in E} \frac{(Av, v)}{|v^2|}.$$

En particulier,

$$\lambda_1 = \inf_{v\in\mathbb{R}^N} \frac{(Av, v)}{|v^2|}\,,\ \lambda_N = \sup_{v\in\mathbb{R}^N} \frac{(Av, v)}{|v^2|}.$$

Théorème 13.14. (Gerschgorin)
Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{C})$. Soit $\mathrm{Sp}(A)$ l'ensemble des valeurs propres de A. On a

$$\mathrm{Sp}(A) \subset \bigcup_{i=1}^n D(a_{ii}, r_i)\,,\ r_i = \sum_{j\neq i} |\ a_{ij}|\ ,$$

où $D(a, r) \subset \mathbb{C}^2$ désigne le disque fermé de centre a et de rayon r.

13.4.2. Spectre du Laplacien discret. La matrice

$$A = \begin{pmatrix} 2 & -1 & 0 & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot \\ 0 & -1 & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & 2 & -1 \\ 0 & \cdot & \cdot & 0 & -1 & 2 \end{pmatrix} \in \mathcal{M}_{N-1}(\mathbb{R}) \tag{13.4}$$

possède $N-1$ valeurs propres distinctes

$$\lambda_k = 4\sin^2\left(\frac{k\pi}{2N}\right),\ k = 1,\ \ldots,\ N-1.$$

Le vecteur propre associé à la valeur propre λ_k s'écrit

$$\mathbf{u}_k = {}^t\left(\sin\left(\frac{k\pi}{N}\right),\ \sin\left(\frac{2k\pi}{N}\right),\ \ldots,\ \sin\left(\frac{(N-1)k\pi}{N}\right)\right).$$

13.4.3. Valeurs propres du Laplacien.

Proposition 13.15. Les valeurs propres du Laplacien avec conditions de Dirichlet homogènes dans un parallélépipède $L_1 \times \cdot L_n$ de $\mathbb{R}^3$ sont

$$\lambda_{k_1 \ldots k_n} = \pi^2\left(\left(\frac{k_1}{L_1}\right)^2 + \cdots + \left(\frac{k_n}{L_n}\right)^2\right).$$

13.5. Assemblage des matrices éléments finis

13.5.1. Intégrale de fonctions barycentriques dans un simplexe. Soit K un simplexe de $\mathbb{R}^n$ non dégérénéré, c'est à dire l'enveloppe convexe de $n+1$ points dont les combinaisons barycentriques engendrent l'espace. On note $\lambda_i(x)$ la i-ème coordonnée barycentrique d'un point x de K. On a

$$\int_K \lambda_1^{\alpha_1} \ldots \lambda_{n+1}^{\alpha_{n+1}} = |K|\,\frac{\alpha_1! \ldots \alpha_{n+1}!}{(\alpha_1 + \cdots + \alpha_{n+1} + n)!}, \tag{13.5}$$

où $|K|$ est le volume de K .

13.6. Réseaux résistifs

On définit un réseau résistifs Λ comme la donnée d'un ensemble V de sommets, d'un ensemble $E \subset V \times V$ d'arêtes, symétrique ($(x,y) \in E \Longrightarrow (y,x) \in E$), et de résistances définies sur les arêtes e de E ($(r(x,y) = r(y,x) > 0$ pour $(x,y) \in E$). On suppose V fini.

La loi d'Ohm [2]s'écrit

$$u(x) - u(y) = r(x,y)j(x,y),$$

2. On peut aussi donner une interprétation fluide du système, en considérant que Λ modélise un réseau de tuyaux au travers desquels s'écoule un fluide visqueux. La loi de Poiseuille (voir section 13.10) assure la proportionalité du saut de pression entre l'entrée et la sortie d'un tuyau et le flux qui le traverse. Le

où $u(\cdot)$ représente le potentiel au point considéré, et $j(x,y)$ l'intensité du courant de x vers y.

Si l'on note $J(x)$ le flux injecté dans le réseau au point x, la loi des nœuds s'écrit, en tout point $x \in V$,

$$\sum_{y,(x,y)\in E} j(x,y) = J(x),$$

qui s'écrit d'après la loi d'Ohm (avec $c(x,y) = r(x,y)^{-1}$)

$$\sum_{y,(x,y)\in E} c(x,y)(u(x)-u(y)) = J(x),$$

que l'on peut écrire de façon plus ramassée

$$-\triangle u = J,$$

où $\triangle$ est le Laplacien discret associé au réseau.

Remarque 13.16. Dans le cas d'un réseau cartésien (avec une numérotation (i,j) des sommets), avec $r \equiv 1$, on retrouve la discrétisation du Laplacien usuel par différences finies :

$$4u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1} = J_{i,j}.$$

Puissance dissipée. On a une formule Green discrète

$$\sum_x (-\Delta u)(x)u(x) = \sum_x \sum_{y,(x,y)\in E} c(x,y)(u(x)-u(y))u(x) = \sum_e c(x,y)\,|u(x)-u(y)|^2$$

qui peut s'écrire $\sum_e r(e)\,|j(e)|^2$, ce qui correspond à la puissance dissipée dans le réseau. Noter que dans le cas d'un réseau connexe, si l'on fixe la valeur en un point x_0, alors l'opérateur $-\Delta$ restreint à $\Lambda \setminus \{x_0\}$ est inversible (la forme bilinéraire associée est définie positive).

On peut s'étonner de l'absence de termes de bords dans la formule de Green ci-dessus. En fait, tous les points du réseau jouent le rôle de points intérieurs. On peut retrouver une formule avec termes de bords en choisissant un sous ensemble Λ_0 de points (typiquement les points reliés à un seul autre sommet, mais pas forcément, la notion de bord étant ici arbitraire), et en sommant sur le complémentaire de Λ_0.

Problème de Dirichlet et principe du maximum. On considère un réseau connexe, et l'on note Λ_0 l'ensemble des bouts du réseaux, c'est à dire les points qui ne sont reliés qu'à un autre point, et $\Lambda^i = \Lambda \setminus \Lambda_0$ l'ensemble des points intérieurs. On se donne une collection de valeurs U sur Λ_0, et l'on considère le problème

$$-\Delta u(x) = 0 \quad \forall x \in \Lambda^i\,,\ u(x) = U(x) \quad \forall x \in \Lambda_0.$$

Le champ de potentiels u vérifie alors le principe du maximum :

$$\sup_{x\in\Lambda} u(x) = \sup_{x\in\Lambda^0} u(x).$$

En effet, en tout point intérieur, $u(x)$ est combinaison convexe des valeurs de u aux voisins.

champ u est alors interprété comme un champ de pressions aux points de raccord, et $j(e)$ est simplement le flux de fluide aux travers de l'arête e.

13.7. Formules d'intégration par partie

Proposition 13.17. Soient $\mathbf{u}$ et : $\mathbf{v}$ deux champs réguliers sur Ω. On a

$$\int_\Omega \nabla\mathbf{u} : {}^t\nabla\mathbf{v} = \int_\Omega (\nabla\cdot\mathbf{u})\,(\nabla\cdot\mathbf{v}) - \int_\Gamma (\nabla\cdot\mathbf{u})\mathbf{v}\cdot\mathbf{n} + \int_\Gamma (\nabla\mathbf{u}\cdot\mathbf{v})\cdot\mathbf{n}$$

Démonstration : On a

$$\begin{aligned}
\int_\Gamma (\nabla\mathbf{u}\cdot\mathbf{v})\cdot\mathbf{n} &= \int_\Omega \nabla\cdot(\nabla\mathbf{u}\cdot\mathbf{v}) \\
&= \int_\Omega \sum_i \partial_i \sum_j v_j\partial_j u_i \\
&= \int_\Omega \sum_i \sum_j ((\partial_i\partial j u_i)v_j + \partial_j u_i \partial_i v_j) \\
&= \int_\Omega \mathbf{v}\,(\nabla\nabla\cdot\mathbf{u}) + \int_\Omega \nabla\mathbf{u} : {}^t\nabla\mathbf{v} \\
&= \int_\Omega (\nabla\cdot\mathbf{u})\,\mathbf{v}\cdot\mathbf{n} - \int_\Omega (\nabla\cdot\mathbf{u})\,(\nabla\cdot\mathbf{v}) + \int_\Omega \nabla\mathbf{u} : {}^t\nabla\mathbf{v}
\end{aligned}$$

13.8. Opérateurs différentiels en coordonnées curvilignes

La divergence d'un champ de vecteur $\mathbf{u} = (u_r, u_\theta, u_z)$ en coordonnées polaires s'écrit

$$\nabla\cdot\mathbf{u} = \frac{\partial u_r}{\partial r} + \frac{u_r}{r} + \frac{1}{r}\frac{\partial u_\theta}{\partial\theta} + \frac{\partial u_z}{\partial z}.$$

Le Laplacien d'une fonction u du plan écrite en coordonnées cylindriques (r,θ,z) s'écrit

$$\triangle u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial\theta^2} + \frac{\partial^2 u}{\partial z^2}.$$

Le Laplacien d'une fonction u de l'espace écrite en coordonnées sphériques (r,θ,Φ) s'écrit

$$\triangle u = \frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2\sin^2\Phi}\frac{\partial^2 u}{\partial\theta^2} + \frac{\cos\Phi}{r^2\sin\Phi}\frac{\partial u}{\partial\Phi} + \frac{1}{r^2}\frac{\partial^2 u}{\partial\Phi^2}$$

13.9. Solutions particulières de l'équation de Poisson

En dimension 2, la fonction (exprimée en coordonnées polaires) $u(r,\theta) = u(r) = \ln r$ est harmonique sur $\mathbb{R}^2\setminus\{0\}$, et vérifie

$$-\Delta u = \delta,$$

au sens des distributions sur $\mathbb{R}^2$.

Pour tout entier $m \geq 1$, la fonction (exprimée en coordonnées polaires)

$$u(r,\theta) = r^m \sin(m\theta)$$

est harmonique sur $\mathbb{R}^2$.

En dimension $d \geq 3$, la fonction (exprimée en coordonnées polaires) $u(r,\theta) = u(r) = r^{d-2}$ est harmonique sur $\mathbb{R}^d \setminus \{0\}$, et vérifie

$$-\Delta u = \delta,$$

au sens des distributions sur $\mathbb{R}^2$.

13.10. Solutions particulières pour Stokes

Écoulement de Poiseuille bidimensionnel. On considère l'écoulement bidimensionnel d'un fluide visqueux incompressible entre deux parois parallèles, en l'absence de forces de masse. On considère plus précisément le domaine $]0, L[\times]-a, +a[$. Pour tout $\mathrm{U}_0 > 0$ (vitesse maximale du fluide en entrée), les champs

$$\mathbf{u}(x,y) = \mathrm{U}_0 \left(1 - \frac{y^2}{a^2}\right) \mathbf{e}_x \,,\ \mathrm{p}(x,y) = -2\frac{\mu \mathrm{U}_0}{a^2} x,$$

sont solutions des équations de Stokes. On peut en déduire une formule très importante en pratique qui relie le débit Q, la viscosité μ, le saut de pression entre l'entrée et la sortie, et la longueur du domaine. Le débit (volume de fluide traversant une section transverse par unité de temps) étant égal à

$$Q = \frac{4}{3} a \mathrm{U}_0,$$

cette relation s'écrit

$$Q = \frac{2}{3}\frac{a^3}{\mu L}(\mathrm{p}_{\text{entrée}} - \mathrm{p}_{\text{sortie}}).$$

ou, par analogie avec la loi d'Ohm qui gère la circulation de courant électrique dans un fil de résistance R,

$$(\mathrm{p}_{\text{entrée}} - \mathrm{p}_{\text{sortie}}) = RQ \,,\ R = \frac{3}{2}\frac{\mu L}{a^3} \tag{13.6}$$

Écoulement de Poiseuille tridimensionnel. On s'intéresse ici à l'écoulement d'un fluide visqueux incompressible à travers un cylindre de section droite Ω, où Ω est un domaine bidimensionnel de forme quelconque. Nous allons vérifier que, comme en dimension deux, il existe une solution telle que la pression est constante dans chaque section du tube, et la vitesse est invariante par translation le long de la direction génératrice du cylindre. On suppose que l'écoulement se fait dans la direction z, et l'on introduit le champ scalaire $\Phi(x,y)$ solution de

$$\left\{ \begin{array}{rcll} -\triangle \Phi & = & 1 & \text{dans} \quad \Omega \\ \Phi & = & 0 & \text{sur} \quad \partial\Omega. \end{array} \right.$$

Alors les champs

$$\mathbf{u}(x,y,z) = \lambda \Phi(x,y) \mathbf{e}_z \,,\ \mathrm{p}(x,y,z) = -\mu\lambda z,$$

sont solutions des équations de Stokes tridimensionnelles.

Écoulement de Poiseuille dans un tube de section circulaire. On peut calculer explicitement les vitesses et pressions correspondant à l'écoulement d'un fluide visqueux incompressible à travers un tube de section circulaire. En effet, dans le cas où Ω est un

cercle, on peut exprimer la fonction Φ (introduite ci-dessus) explicitement. En coordonnées polaires (Ω est un cercle de rayon a, centré en 0), elle s'exprime

$$\Phi(r,\theta) = \Phi(r) = \frac{1}{4}(a^2 - r^2).$$

La solution peut donc s'écrire

$$\mathbf{u}(x,y,z) = \mathrm{U}_0 \left(1 - \frac{r^2}{a^2}\right) \mathbf{e}_z\,,\ \mathrm{p}(x,y,z) = -4\frac{\mu \mathrm{U}_0}{a^2}(z - z_0),$$

où U_0 est la vitesse maximale, réalisée sur l'axe du cylindre. On a effet pour la formule du Laplacien en coordonnées cylindriques)

$$\Delta \mathbf{u} = (\Delta u_z)\mathbf{e}_z = -4\frac{U_0}{a^2},$$

de telle sorte que les équations de Stokes sont vérifiées en tout point.

On peut en déduire une formule analogue à celle trouvée pour l'écoulement bidimensionnel, pour une conduite de longueur L. Le débit vaut en effet (Loi de Poiseuille)

$$Q = \mathrm{U}_0 \pi \frac{a^2}{2} = \frac{\pi}{8}\frac{a^4}{\mu L}(\mathrm{p}_{\text{entrée}} - \mathrm{p}_{\text{sortie}}), \tag{13.7}$$

soit une résistance qui s'exprime en fonction du diamètre $D = 2a$

$$R = \frac{\mu}{128\pi}\frac{L}{D^4}.$$

Cette résistance exprime la proportionnalité entre saut de pression et débit, et s'exprime donc en $\mathrm{Pa}\,s\,m^{-3}$.

Écoulement autour d'une sphère. On peut décrire explicitement le champ de vitesse correspondant à l'écoulement d'un fluide visqueux en milieu infini autour d'une sphère fixe. On considère une sphère de rayon a centrée à l'origine d'un repère $(O, \mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$, et l'on se place dans le système de coordonnées sphériques (O, r, θ, ϕ) : pour tout point de $\mathbb{R}^3$ représenté par son rayon vecteur $\mathbf{r} = (x, y, z)$, r est le module de $\mathbf{r}$, θ est l'angle que fait $(x, y, 0)$ avec $\mathbf{e}_x$ (longitude, comprise entre 0 et 2π), et Φ est l'angle que fait $\mathbf{r}$ avec l'axe des z (latitude, comprise entre 0 et π). On suppose que la vitesse à l'infini est égale à $\mathrm{U}_0\mathbf{e}_z$. Les vecteurs unitaires associés à ce système de coordonnées qui vont nous servir à exprimer le champ des vitesses sont

$$\mathbf{e}_r = \frac{\mathbf{r}}{r}\,,\ \mathbf{e}_\Phi = \frac{1}{r}\frac{\partial \mathbf{r}}{\partial \Phi}.$$

On peut vérifier que tout couple $(\mathbf{u}, \mathrm{p})$ défini par

$$\mathbf{u} = u_r \mathbf{e}_r + u_\Phi \mathbf{e}_\Phi\,,\ u_r = \mathrm{U}_0 \cos\Phi \left(1 - \frac{3a}{2r} + \frac{a^3}{2r^3}\right)\,,\ u_\Phi = -\mathrm{U}_0 \sin\Phi \left(1 - \frac{3a}{4r} - \frac{a^3}{4r^3}\right),$$

$$\mathrm{p} - \mathrm{p}_0 = -\frac{3}{2}\frac{\mu \mathrm{U}_0 a}{r^2}\cos\Phi,$$

où p_0 est une constante arbitraire, est solution des équations de Stokes dans le domaine $\mathbb{R}^3 \setminus B(0,a)$, avec des conditions d'adhérence ($\mathbf{u} = 0$) sur la sphère $\{r = a\}$, et des conditions à l'infini

$$\lim_{r\to+\infty} \mathbf{u}(r,\theta,\Phi) = \mathrm{U}_0 \mathbf{e}_z.$$

On en déduit l'expression du module de la force exercée par le fluide sur la sphère :

$$F = 6\pi\mu a \mathrm{U}_0. \tag{13.8}$$

13.11. Coefficient de Poisson, module d'Young, et paramètres de Lamé

13.11.1. Définitions, relations. Les paramètres caractérisant un matériau élastique en régime linéaire les plus couramment utilisés par les ingénieurs (car directement accessibles à la mesure) sont le module d'Young E, qui quantifie le rapport entre l'allongement et la contrainte exercée, et le coefficient de Poisson ν, qui quantifie le rapport entre la déformation dans la direction transverse à l'étirement et la déformation selon la direction de l'effort.

Les paramètres intervenant dans les équations de l'élasticité linéaires sont appelés coefficients de Lamé, notés en général μ et λ. Ils précisent le lien entre le tenseur des contraintes et le tenseur des taux de déformation :

$$\boldsymbol{\sigma} = \mu\left(\nabla\mathbf{u} + {}^t\nabla\mathbf{u}\right) + \lambda\left(\nabla\cdot\mathbf{u}\right)\mathrm{Id}$$

Ces familles de paramètres sont liées par les relations suivantes

$$\mu = \frac{E}{2(1+\nu)}\,,\ \lambda = \frac{\nu E}{(1+\nu)(1-2\nu)}\,, \tag{13.9}$$

et inversement

$$E = \frac{\mu(3\lambda+2\mu)}{\lambda+\mu}\,,\ \nu = \frac{\lambda}{2(\lambda+\mu)}. \tag{13.10}$$

Ces relations peuvent être démontrées en considérant la situation d'un échantillon de matériaux occupant le parallélépipède

$$]-a/2, a/2[\times]-a/2, a/2[\times]0, b[.$$

On cherche un champ de déplacement associé à un étirement selon $\mathbf{e}_3$ sous la forme suivante (ε représente la déformation selon la direction vecticale)

$$u^1 = -\nu\varepsilon x_1\,,\ u^2 = -\nu\varepsilon x_2\,,\ u^3 = \varepsilon x_3.$$

Le tenseur des contraintes associé est constant égal à

$$\boldsymbol{\sigma} = \begin{pmatrix} -2\mu\nu\varepsilon + \lambda\varepsilon(1-2\nu) & 0 & 0 \\ 0 & -2\mu\nu\varepsilon + \lambda\varepsilon(1-2\nu) & 0 \\ 0 & 0 & 2\mu\varepsilon + \lambda\varepsilon(1-2\nu) \end{pmatrix}$$

Un tel champ vérifie les équations de l'élasticité au sein du matériau (le tenseur des contraintes, constant, a une divergence nulle). Sur les bords latéraux, le tenseur normal des contraintes est nul dès que

$$\sigma_{11} = \sigma_{22} = -2\mu\nu\varepsilon + \lambda\varepsilon(1-2\nu) = 0 \Longleftrightarrow \nu = \frac{\lambda}{2(\lambda+\mu)}.$$

L'effort par unité de surface exercé sur l'échantillon au niveau du bord supérieur s'écrit

$$\sigma_{33} = 2\mu\varepsilon + \lambda\varepsilon(1-2\nu),$$

qui vaut $E\varepsilon$ par définition du module d'Young. On en déduit donc (en remplaçant ν par son expression trouvée ci dessus) la première équation de 13.10

Remarque 13.18. Pour fixer les idées, remarquons que $10^{-6}E$ correspond à la force (exprimée en Newton) qu'il faut exercer sur un échantillon de section égale à 1 cm^2 pour obtenir un allongement de 1%. Ainsi pour du verre ($E = 60$ GPa), cette force est de $60.10^9.10^{-6} = 6.10^4$ N. Pour un échantillon de même section qui supporte une masse de 1 kg, l'allongement est de $\varepsilon = 10^5/E = 1.6\,10^{-4}\%$.

Les valeurs approximatives de E et ν pour un certain nombre de matériaux sont données dans le tableau suivant (valeurs tirées de [**6**]) :

	E (en GPa)	ν	ρ	λ/μ
acier	200	0.3	7.8	5
verre	60	0.25	2.8	4
bois	7	0.2	0.4	3.3
élastomère	0.2	0.5	1	$+\infty$

13.12. Elasticité bi-dimensionnelle

Cette section précise le sens que l'on peut donner aux problèmes d'élasticité en dimension 2. Une première manière de voir les choses (qui correspond à ce qui se passe en mécanique des fluides) est de considérer que l'on s'intéresse à un échantillon tridimensionnel cylindrique (axe selon $\mathbf{e}_3$), soumis à des sollicitations telles que le déplacement selon $\mathbf{e}_3$ est nul, et que toutes les quantités sont invariantes par translation selon $\mathbf{e}_3$. Dans ce cadre, le matériau bidimensionnel possède les mêmes coefficients de Lamé que le matériau réel tridimensionnel. Mais cette manière de voir les choses est peu réaliste, notamment car même un effort purement radial est susceptible d'entraîner des déformations dans la direction orthogonale, sauf dans des situations très particulières.

Une manière alternative de donner un sens au modèle 2D est de considérer une plaque élastique d'épaisseur constante occupant un domaine

$$\omega = \Omega \times]0, \eta[,$$

où Ω est un domaine de $\mathbb{R}^2$.

On suppose cette plaque soumise à des sollicitations ou des contraintes sur le déplacement sur son bord latéral $\partial\Omega\times]0,\eta[$, et des conditions libres sur $\Omega \times \{0\}$ et $\Omega \times \{\eta\}$ Si l'on suppose que le tenseur des contraintes est constant dans la direction x_3, les conditions de traction libre au bord imposent

$$\partial_1 u_3 + \partial_3 u_1 \equiv 0\,,\ \partial_2 u_3 + \partial_3 u_2 \equiv 0\,,\ 2\mu\partial_3 u_3 + \lambda(\partial_1 u_1 + \partial_2 u_2 + \partial_3 u_3) \equiv 0,$$

On a ainsi

$$\partial_3 u_3 = -\lambda \frac{\partial_1 u_1 + \partial_2 u_2}{2\mu + \lambda},$$

ce qui permet de se ramèner donc à un problème d'élasticité linéaire pour un matériau bidimensionnel de coefficients de Lamé

$$\mu^\star = \mu\,,\ \lambda^\star = \frac{2\lambda\mu}{\lambda + 2\mu}.$$

On cherche maintenant à donner un sens à E et ν pour les cas bidimensionnel. La démarche menée ci-dessus conduit aux coefficients :

$$\nu^\star = \frac{\lambda^\star}{\lambda^\star + 2\mu}\,,\ E^\star = \frac{4\mu(\lambda^\star + \mu)}{\lambda^\star + 2\mu}.$$

13.13. Transformée de Fourier sur $\mathbb{Z}_2$ et FFT

$\mathbb{Q}_2$ est l'ensemble des rationnels dyadiques, que l'on peut voir comme l'ensemble des nombres de la forme

$$\xi = \sum_{j=-n}^{+\infty} \xi_j 2^j$$

où n est un entier ≥ 1. C'est un espace métrique complet pour la distance

$$d(\xi, \xi') = \left|\xi' - \xi\right|_2 = \frac{1}{2^v}\,,\ v = \min\left\{n\,,\ \xi_m = \xi'_m \quad \forall m < n\right\}$$

$\mathbb{Z}_2$ est le sous-ensemble des nombres sans partie rationnelle, qu'on peut aussi définir comme la boule unité fermée de $\mathbb{Q}_2$. On peut définir une tranformée de Fourier sur $\mathbb{Z}_2$ de la façon suivante : pour tout fonction u intégrable sur $\mathbb{Z}_2$, on définit

$$\hat{u}(\xi) = \int_{\mathbb{Z}_2} e^{-2i\pi\xi x} u(x)\, dx.$$

On vérifie immédiatement que l'expression ne dépend pas de la partie entière de ξ.

On considère ξ de norme $\leq 2^n$, c'est à dire de la forme

$$\xi = \sum_{j=-n}^{+\infty} \xi_j 2^j.$$

On a

$$\hat{u}(\xi) = \int_{\mathbb{Z}_2} e^{-2i\pi\xi x} u(x)\, dx = \sum_{a=0}^{2^n-1} \int_{a+2^n\mathbb{Z}_2} e^{-2i\pi\xi x} u(x)\, dx = \sum_{a=0}^{2^n-1} e^{-2i\pi\xi a} \int_{a+2^n\mathbb{Z}_2} u(x)\, dx.$$

On a donc

$$\hat{u}(\xi) = \sum_{a=0}^{2^n-1} e^{-2i\pi\xi a} u_n^a = \sum_{a=0}^{2^n-1} \omega^{ka} u_n^a$$

avec

$$u_n^a = \int_{a+2^n\mathbb{Z}_2} u(x)\, dx\,,\quad \omega_n = e^{-2i\pi/2^n} \text{ et } k = \sum_{j=-n}^{-1} \xi_j 2^{n+j}.$$

La quantité $\hat{u}(\xi)$ peut prendre au plus 2^n valeurs différentes, correspondant aux différentes valeurs du k si dessus pour des ξ de norme $\leq 2^n$. Calculer la transformée de Fourier sur $\overline{B}(0, 2^n)$ est donc analogue à calculer la transformée de Fourier discrète du vecteur $u_n = (u_n^0, \ldots, u_n^{2^n-1})$, où u_n^a est l'intégrale de u sur $a + 2^n\mathbb{Z}_2$.

Transformée de Fourier rapide. Le principe de la transformée de Fourier rapide peut s'exprimer comme suit

$$\hat{u}(\xi) = \int_{\mathbb{Z}_2} e^{-2i\pi\xi x} u(x)\, dx = \int_{2\mathbb{Z}_2} e^{-2i\pi\xi x} u(x)\, dx + \int_{1+2\mathbb{Z}_2} e^{-2i\pi\xi x} u(x)\, dx$$

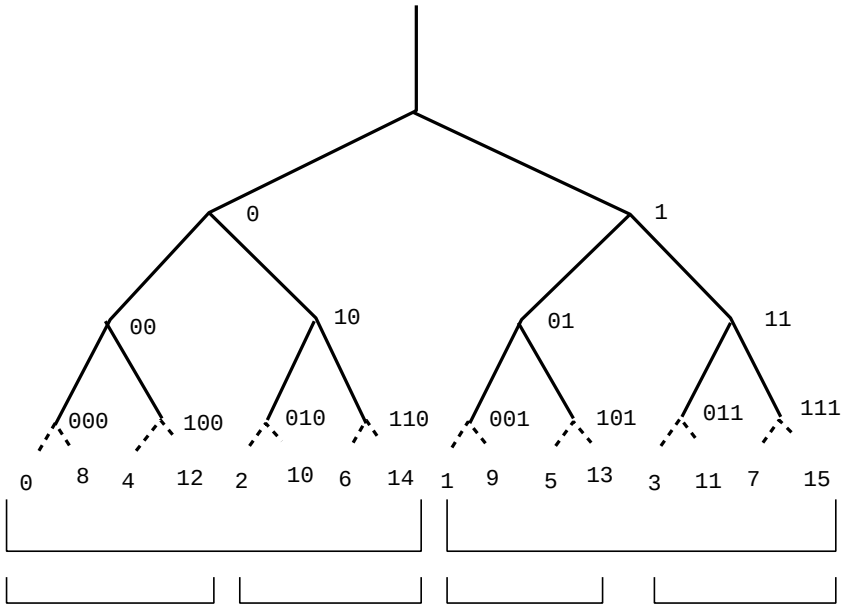


FIGURE 1. Représentation de $\mathbb{Z}_2$ en arbre

$$= \frac{1}{2}\left(\int_{\mathbb{Z}_2} e^{-2i\pi\xi 2y} u(2y)\,dy + \int_{\mathbb{Z}_2} e^{-2i\pi\xi(2y+1)} u(2y+1)\,dy\right)$$
$$= \frac{1}{2}\left(\hat{u}_0(2\xi) + e^{-2i\pi\xi}\hat{u}_1(2\xi)\right)$$

avec $u_0(y) = u(2y)$ et $u_1(y) = u(2y+1)$.

La numérotation naturelle des feuilles de l'arbre tronqué induite par la construction de $\mathbb{Z}_2$ est très adaptée à la formulation de la transformée de Fourier rapide : les indices pairs correspondent à la première moitié. Si l'on divise ces indices par deux, alors les indices pairs au niveau inférieur correspondent toujours à la première moitié. Les transformées de Fourier successives de sous-vecteurs que l'on calcule en effectuant la FFT correspondent ainsi à des suites d'indices consécutifs dans cette numérotation.

Bibliographie

1. Grégoire Allaire, *Analyse numérique et optimisation*, Publications Ecole Polytechnique, vol. 15, Ellipses, Paris, 2005.
2. Ivo Babuška, *The finite element method with penalty*, Math. Comp. **27** (1973), 221–228.
3. H. Brezis, *Analyse fonctionnelle, théorie et applications*, Masson, Paris, 1983.
4. Franco Brezzi and Michel Fortin, *Mixed and hybrid finite element methods*, Springer Series in Computational Mathematics, vol. 15, Springer-Verlag, New York, 1991.
5. Martin Costabel and Monique Dauge, *Edge singularities for elliptic boundary value problems*, Journées équations aux dérivées partielles (1992), 1–12.
6. G. Duvaut, *Mécanique des milieux continus*, Masson, Paris, 1990.
7. G. Duvaut and J. L. Lions, *Les inéquations en mécanique et en physique*, Dunod, Paris, 1972.
8. Howard C. Elman and Gene H. Golub, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal. **31** (1994), no. 6, 1645–1661. MR MR1302679 (95f :65065)
9. V. Girault and Raviart, *Finite element methods for navier-stokes equations- theory and algorithms*, Springer-Verlag, Berlin Heidelberg New York, 1979.
10. R. Glowinski, *Augmented lagrangian and operator-splitting methods in nonlinear mechanics*, SIAM Studies in Applied Mechanics, Philadelphia, 1989.
11. P. Grisvart, *Elliptic problems in nonsmooth domains*, Monographs and Studies in Mathematics, Pitman, Boston London melbourne, 1985.
12. A. Haraux, *Nonlinear evolution equations - global behaviour of solutions*, Springer Verlag, Berlin-Heidelberg-New York, 1981.
13. P.-A. Raviart and J.M. Thomas, *Introduction à l'analyse numérique des équations aux dérivées partielles*, Masson, Paris, 1983.

Index