

Analyse numérique des EDP elliptiques

Franck Boyer

Master 2ième année - Mathématiques et Applications

Université Paul Cézanne

1^{er} décembre 2009

TABLE DES MATIÈRES

Chapitre I	Prérequis. Introduction aux EDP elliptiques	1
1	Espaces de Sobolev	1
1.1	Espaces $H^k(\Omega)$, $H_0^k(\Omega)$, $H^{-k}(\Omega)$	1
1.2	Inégalités/Injections de Sobolev	4
1.3	Opérateur de traces. Espaces de traces	4
1.4	Inégalités de Poincaré	6
1.5	Champs L^2 à divergence L^2 . Trace normale	7
2	Formulations faibles de problèmes elliptiques	10
2.1	Introduction	10
2.2	Théorème de Lax-Milgram	11
2.3	Exemples de base	12
2.4	Exemples plus sophistiqués	15
3	Le principe du maximum faible	18
4	Théorème de Lax-Milgram généralisé : formulations mixtes	19
4.1	Enoncé et preuve du théorème général	19
4.2	Problèmes de type point-selle ou formulations variationnelles sous contrainte	20
4.3	Exemples	26
Chapitre II	Approximations de Galerkin. Méthode des éléments finis	29
1	Principes généraux de l'approximation de Galerkin	29
1.1	Cadre coercif. Approximation de Galerkin.	30
1.2	Cadre coercif. Approximation de Petrov-Galerkin.	31
1.3	Cadre d'un problème mixte	33
1.4	Choix des espaces d'approximation et de la base	35
2	Exemples d'espaces éléments finis	36
2.1	Le cas 1D	36
2.2	Quelques notions en dimension supérieure	45
2.3	Autres éléments finis utiles	57
3	Problèmes de type point-selle. Couples d'éléments inf-sup stables	58
3.1	Lemme de Fortin	59
3.2	Le problème de Stokes	60

Chapitre III	Discrétisation Volumes Finis de problèmes elliptiques	67
1	Le problème de Laplace sur maillages orthogonaux admissibles	68
1.1	Maillages orthogonaux admissibles	68
1.2	Construction du schéma numérique	69
1.3	Premières propriétés : existence et unicité de la solution - stabilité	70
1.4	Compacité - Convergence	73
1.5	Estimation de l'erreur en dimension 2	78
2	Quelques problèmes plus généraux	80
2.1	Le problème de diffusion isotrope hétérogène	80
2.2	Le problème de convection-diffusion	80
2.3	Le problème anisotrope hétérogène	80

Chapitre I

PRÉREQUIS. INTRODUCTION AUX EDP ELLIPTIQUES

1 Espaces de Sobolev

L'essentiel du contenu de ce paragraphe provient des références [Bre83] et [Ada75].

1.1 Espaces $H^k(\Omega)$, $H_0^k(\Omega)$, $H^{-k}(\Omega)$

Définition I.1

Soit Ω un ouvert de \mathbb{R}^d . Pour tout entier $k \geq 0$, on définit l'espace de Sobolev $H^k(\Omega)$ de la façon suivante

$$H^k(\Omega) = \{u \in L^2(\Omega), \partial^\alpha u \in L^2(\Omega), \forall \alpha \in \mathbb{N}^d, |\alpha| \leq k\},$$

muni de la norme

$$\|u\|_{H^k} = \left(\sum_{|\alpha| \leq k} \|\partial^\alpha u\|_{L^2}^2 \right)^{\frac{1}{2}}.$$

De façon équivalente, on a $H^0(\Omega) = L^2(\Omega)$ et $H^{k+1}(\Omega) = \{u \in H^k(\Omega), \nabla u \in (H^k(\Omega))^d\}$.

On définit enfin $H_0^k(\Omega)$, l'adhérence dans $H^k(\Omega)$ de $\mathcal{D}(\Omega)$, ensemble des fonctions de classe C^∞ et à support compact dans Ω .

Proposition I.2

Les espaces $H^k(\Omega)$ et $H_0^k(\Omega)$ ainsi définis sont des espaces de Hilbert et de plus, si Ω est Lipschitzien, alors l'ensemble $C^\infty(\bar{\Omega})$ des fonctions régulières jusqu'au bord de Ω est dense dans $H^k(\Omega)$.

Il est possible de caractériser le dual topologique de l'espace $H_0^k(\Omega)$ de la façon suivante.

Définition I.3

Pour tout $k \geq 1$, on définit l'espace de distributions suivant

$$H^{-k}(\Omega) = \left\{ f \in \mathcal{D}'(\Omega), \quad f = \sum_{|\alpha| \leq k} \partial^\alpha f_\alpha, \quad \text{avec } f_\alpha \in L^2(\Omega) \right\},$$

muni de la norme

$$\|f\|_{H^{-k}} = \inf \left(\sum_{|\alpha| \leq k} \|f_\alpha\|_{L^2}^2 \right)^{\frac{1}{2}},$$

l'infimum étant pris sur toutes les décompositions possibles de f sous la forme intervenant dans la définition de H^{-k} .

Proposition I.4

Pour tout $k \geq 1$, l'espace $H^{-k}(\Omega)$ ainsi défini est un espace de Hilbert qui, de plus, est isomorphe au dual topologique de $H_0^k(\Omega)$. De façon plus précise, le crochet de dualité s'exprime de la façon suivante

$$\langle f, u \rangle_{H^{-k}, H_0^k} = \sum_{|\alpha| \leq k} (-1)^{|\alpha|} \int_{\Omega} f_\alpha \partial^\alpha u \, dx, \quad \forall f \in H^{-k}(\Omega), \forall u \in H_0^k(\Omega).$$

On vérifie bien sûr que cette formule ne dépend pas de la décomposition de f en somme des $\partial^\alpha f_\alpha$.

Attention, le dual topologique de $H^k(\Omega)$ n'est pas un espace de distributions et ne possède donc pas de caractérisation aussi simple.

Remarque I.5

L'exemple le plus simple à retenir, et le plus utile pour la suite de ce cours, c'est que tout élément f de $H^{-1}(\Omega)$ s'écrit, au sens des distributions, sous la forme

$$f = u + \operatorname{div} G, \quad \text{avec } u \in L^2(\Omega) \text{ et } G \in (L^2(\Omega))^d.$$

Concluons ce paragraphe par le théorème de dérivation des fonctions composées dans les espaces de Sobolev.

Théorème I.6

Soit Ω un ouvert de \mathbb{R}^d . Pour toute fonction $u \in H^1(\Omega)$ et toute fonction $T : \mathbb{R} \mapsto \mathbb{R}$ de classe \mathcal{C}^1 à dérivée bornée nous avons

$$T(u) \in H^1(\Omega), \quad \text{et } \nabla T(u) = T'(u) \nabla u. \tag{I.1}$$

De plus,

$$u \in H^1(\Omega) \mapsto T(u) \in H^1(\Omega),$$

est continue.

Preuve :

1. On raisonne par densité en approchant $u \in H^1(\Omega)$, par une suite $(u_n)_n$ de fonctions régulières. Pour ces fonctions u_n , le théorème de dérivation des fonctions composées s'applique et on a donc $\nabla(T(u_n)) =$

$T'(u_n)\nabla u_n$. En prenant une fonction test $\Phi \in (\mathcal{C}_c^\infty(\Omega))^d$, on a donc

$$\int_{\Omega} T(u_n) \operatorname{div} \Phi \, dx = - \int_{\Omega} T'(u_n) \nabla u_n \cdot \Phi \, dx.$$

Il faut donc vérifier que l'on peut passer à la limite dans cette formule pour montrer le résultat.

- Tout d'abord, comme T est à dérivée bornée :

$$|T(u_n) - T(u)| \leq \|T'\|_{\infty} |u_n - u|,$$

et donc $T(u_n)$ converge vers $T(u)$ dans $L^2(\Omega)$, ce qui permet de passer à la limite dans le premier terme.

- Pour le second terme, on utilise le fait que, quitte à extraire une sous-suite de $(u_n)_n$ on peut supposer que u_n converge presque partout vers u (et donc $T'(u_n)$ converge presque partout vers $T'(u)$ puisque T' est continue). Par convergence dominée, on en déduit que $T(u_n)$ converge vers $T(u)$ dans $L^2(\Omega)$, et comme par ailleurs ∇u_n converge vers ∇u dans $L^2(\Omega)$, on a bien justifié le passage à la limite dans l'égalité proposée.

2. Soit $(u_n)_n$ une suite de fonctions de $H^1(\Omega)$ qui converge vers $u \in H^1(\Omega)$.

- On a tout d'abord

$$\|T(u_n) - T(u)\|_{L^2} \leq \|T'\|_{\infty} \|u_n - u\|_{L^2},$$

et donc $T(u_n)$ converge vers $T(u)$ dans L^2 .

- Il faut maintenant regarder les gradients.

$$\|\nabla T(u_n) - \nabla T(u)\|_{L^2}^2 \leq \|T'\|_{\infty}^2 \|\nabla u_n - \nabla u\|_{L^2}^2 + \int_{\Omega} |T'(u_n) - T'(u)|^2 |\nabla u|^2 \, dx.$$

Le premier terme tend clairement vers 0. Quant au second terme, on va montrer que sa limite supérieure α est nulle. Pour cela, on considère une sous-suite $u_{\varphi(n)}$ qui réalise la valeur α . Comme $(u_n)_n$ converge dans $L^2(\Omega)$ et quitte à extraire encore une fois une sous-suite, on peut supposer que $u_{\varphi(n)}$ converge vers u presque partout. Dans ces conditions, on peut appliquer le théorème de convergence dominée et obtenir

$$\alpha = \lim_{n \rightarrow \infty} \int_{\Omega} |T'(u_{\varphi(n)}) - T'(u)|^2 |\nabla u|^2 \, dx = 0.$$

En réalité, on peut montrer que cette propriété est encore vraie si T est seulement Lipschitzienne et \mathcal{C}^1 par morceaux par exemple. Dans ce cas, T' n'est pas définie de manière univoque en certains points, mais le théorème est quand même vrai et implique en particulier que la valeur que l'on donne en T' en ces points n'intervient pas. ■

Exemple I.7

Si $T(x) = x^+ = \max(x, 0)$, on prend $T'(x) = 1$ si $x > 0$ et $T'(x) = 0$ si $x \leq 0$. On obtient alors

$$\nabla(u^+) = 1_{u>0} \nabla u.$$

Mais en affectant la valeur 1 à T' en $x = 0$, on trouve aussi

$$\nabla(u^+) = 1_{u \geq 0} \nabla u.$$

Le théorème précédent dit que ces deux formules sont exactes, ce qui implique que

$$\nabla u = 0, \text{ presque partout sur l'ensemble } \{u = 0\},$$

et en effet, nous avons :

$$\nabla u = 0, \text{ presque partout sur l'ensemble } \{u = \alpha\},$$

pour toute valeur de $\alpha \in \mathbb{R}$.

1.2 Inégalités/Injections de Sobolev

On se contente ici de donner les résultats concernant l'espace $H^1(\Omega)$ même si des résultats similaires peuvent être démontrés pour les espaces $H^k(\Omega)$ ou les espaces $W^{1,p}(\Omega)$, avec $p \neq 2$.

Théorème I.8

Soit Ω un ouvert borné Lipschitzien de \mathbb{R}^d .

- Si $d = 1$: on a l'injection continue de $H^1(\Omega)$ dans l'espace de Hölder $C^{0,\frac{1}{2}}(\Omega)$.
- Si $d = 2$: on a l'injection continue de $H^1(\Omega)$ dans l'espace $L^p(\Omega)$ pour tout $p < +\infty$. Par contre, on **n'a pas** l'injection continue de $H^1(\Omega)$ dans $L^\infty(\Omega)$.
- Si $d \geq 3$: on a l'injection continue de $H^1(\Omega)$ dans l'espace $L^{p^*}(\Omega)$ avec $p^* = \frac{2d}{d-2}$.

De plus, les injections **non critiques** sont compactes.

Dans le cas non borné, les résultats sont essentiellement les mêmes (à part quelques variantes du fait que les espaces L^p ne sont pas emboîtés dans ce cas). La différence essentielle est que la compacité des injections de Sobolev n'est plus vérifiée.

Comme on va s'intéresser par la suite à la discrétisation de problème aux dérivées partielles, le cadre de domaines bornés suffira à notre étude.

1.3 Opérateur de traces. Espaces de traces

On sait qu'une fonction lipschitzienne définie sur un ouvert Ω peut se prolonger par continuité jusqu'au bord de Ω , ce qui permet de donner un sens à la notion de *trace* de la fonction sur le bord. On va montrer que l'on peut, dans une certaine mesure, faire une construction équivalente dans le cas de fonctions moins régulières.

Théorème I.9

Soit Ω un ouvert borné et lipschitzien de \mathbb{R}^d . L'application

$$\gamma_0 : u \in C^\infty(\bar{\Omega}) \mapsto \gamma_0(u) = u|_{\partial\Omega} \in L^2(\partial\Omega),$$

se prolonge de manière unique, et de façon continue à l'espace de Sobolev $H^1(\Omega)$. On appelle l'opérateur γ_0 ainsi obtenu : l'application de traces.

L'opérateur γ_0 n'est pas surjectif sur $L^2(\partial\Omega)$. L'image de γ_0 est un espace de Sobolev fractionnaire appelé $H^{\frac{1}{2}}(\partial\Omega)$ et qui est un Hilbert pour la norme

$$\|v\|_{H^{\frac{1}{2}}(\partial\Omega)} = \inf_{u \in H^1(\Omega), \gamma_0 u = v} \|u\|_{H^1(\Omega)}.$$

Dans ces conditions, il existe un opérateur linéaire continu $R_0 : H^{\frac{1}{2}}(\partial\Omega) \mapsto H^1(\Omega)$, dit de relèvement, qui vérifie

$$\gamma_0 \circ R_0 = \text{Id}_{\partial\Omega}.$$

N.B. : Dans le théorème précédent, l'espace $L^2(\partial\Omega)$ est défini par le biais de cartes locales.

Proposition I.10

L'opérateur de traces ainsi construit permet d'écrire une formule d'intégration par parties

$$\forall u \in H^1(\Omega), \forall \Phi \in (H^1(\Omega))^d, \int_{\Omega} u \operatorname{div} \Phi \, dx + \int_{\Omega} \nabla u \cdot \Phi \, dx = \int_{\partial\Omega} (\gamma_0 u)(\gamma_0 \Phi) \cdot \nu \, ds.$$

Exemple I.11

Il est instructif de prendre l'exemple du carré unité pour comprendre la démarche. Soit donc $\Omega =]0, 1[^2$ et $u \in C^\infty(\bar{\Omega})$. D'après la formule de Taylor nous avons

$$u^2(x, 0) = u^2(x, y) - 2 \int_0^y u(x, z) \partial_y u(x, z) dz, \quad \forall x, y \in]0, 1[.$$

On prend la valeur absolue, puis on intègre cette inégalité par rapport à y entre 0 et 1 :

$$u^2(x, 0) \leq \int_0^1 u^2(x, y) dy + 2 \int_0^1 |u(x, z)| |\partial_y u(x, z)| dz.$$

En intégrant par rapport à x , puis en utilisant l'inégalité de Cauchy-Schwarz il vient

$$\int_0^1 u^2(x, 0) dx \leq \int_\Omega u^2 dx dy + 2 \|u\|_{L^2} \|\partial_y u\|_{L^2},$$

et in fine

$$\|u\|_{L^2(\{0,1\} \times \{0\})} \leq C \|u\|_{L^2}^{\frac{1}{2}} \|\nabla u\|_{L^2}^{\frac{1}{2}} \leq \frac{C}{2} (\|u\|_{L^2} + \|\nabla u\|_{L^2}).$$

Ceci montre la continuité de l'application γ_0 sur l'espace $C^\infty(\bar{\Omega})$ pour la topologie de $H^1(\Omega)$. Son prolongement continu est alors immédiat par un argument de densité.

D'une certaine façon, l'exposant 1/2 qui apparaît sur les dérivées de u dans la première des inégalités précédentes, explique (ou en tout cas est en lien très fort avec) l'exposant 1/2 dans la définition/notation de l'espace de traces $H^{\frac{1}{2}}$.

Dans le cas d'ouverts suffisamment réguliers, il existe une autre caractérisation utile de l'espace $H_0^1(\Omega)$:

Proposition I.12

Soit Ω un ouvert borné et Lipschitzien de \mathbb{R}^d , et $u \in H^1(\Omega)$, on a l'équivalence

$$u \in H_0^1(\Omega) \iff \bar{u} \in H^1(\mathbb{R}^d),$$

où \bar{u} désigne le prolongement par 0 de u à \mathbb{R}^d tout entier.

On peut également regarder ce que deviennent les traces de fonctions H^1 quand on applique à u une application non-linéaire Lipschitzienne T .

Proposition I.13

Soit Ω ouvert Lipschitzien de \mathbb{R}^d et $T : \mathbb{R} \mapsto \mathbb{R}$ une fonction de classe C^1 à dérivée bornée. Pour tout $u \in H^1(\Omega)$, on a

$$\gamma_0(Tu) = T(\gamma_0 u).$$

Cette propriété reste valable si T est C^1 par morceaux avec un nombre au plus dénombrable de points de discontinuité de T' .

Preuve :

Soit $u \in H^1(\Omega)$. D'après le théorème de densité, il existe une suite $(u_n)_n$ de fonctions de $C^\infty(\bar{\Omega})$ qui converge vers u dans $H^1(\Omega)$ et d'après le théorème I.6, nous savons que $T(u_n)$ converge vers $T(u)$ dans $H^1(\Omega)$ et donc par continuité de l'opérateur de traces, nous avons

$$\gamma_0 T(u_n) \xrightarrow{n \rightarrow \infty} \gamma_0 T(u).$$

Comme u_n est régulière, $\gamma_0(u_n)$ n'est autre que la restriction de u_n sur $\partial\Omega$. De même, $T(u_n)$ est une fonction continue et donc $\gamma_0 T(u_n)$ est égale à la restriction de $T(u_n)$ sur $\partial\Omega$. Ainsi, on a

$$\gamma_0 T(u_n) = T(\gamma_0 u_n), \quad \forall n \in \mathbb{N}.$$

Enfin, nous avons $\gamma_0 u_n$ qui converge dans $H^{\frac{1}{2}}(\partial\Omega)$ et donc dans $L^2(\partial\Omega)$ vers $\gamma_0 u$. Par convergence dominée (en utilisant la sous-linéarité de T), on a donc convergence de $T(\gamma_0 u_n)$ vers $T(\gamma_0 u)$.

Le résultat est donc démontré en mettant bout à bout toutes les convergences obtenues ci-dessus. ■

Ce résultat est fort utile pour démontrer que toutes les propriétés raisonnables de la trace sont satisfaites. Ainsi, en choisissant convenablement T on démontre que si $u \in H^1(\Omega)$ vérifie

$$u(x) \geq \alpha, \quad \text{pour presque tout } x, \text{ alors } \gamma_0(u) \geq \alpha \text{ presque partout.}$$

En particulier, cela prouve que

$$\gamma_0(H^1(\Omega) \cap L^\infty(\Omega)) \subset H^{\frac{1}{2}}(\partial\Omega) \cap L^\infty(\partial\Omega),$$

et on peut même montrer l'égalité entre ces deux espaces.

1.4 Inégalités de Poincaré

L'opérateur de traces défini précédemment permet de caractériser de façon plus claire l'espace $H_0^1(\Omega)$ défini plus haut.

Proposition I.14

Soit Ω un ouvert borné et Lipschitzien de \mathbb{R}^d . On a la propriété

$$H_0^1(\Omega) = \text{Ker } \gamma_0,$$

autrement dit, l'espace $H_0^1(\Omega)$ est exactement l'ensemble des fonctions de $H^1(\Omega)$ qui sont nulles au bord, au sens de l'opérateur de traces.

Reprenons maintenant l'exemple I.11 du carré unité et supposons que u soit nulle au bord. Un calcul tout à fait similaire donne maintenant

$$\int_{\Omega} u^2(x, y) \, dx \, dy \leq 2 \|u\|_{L^2} \|\nabla u\|_{L^2},$$

ce qui amène à

$$\|u\|_{L^2} \leq 2 \|\nabla u\|_{L^2}.$$

Cette inégalité montre que la norme L^2 de u est contrôlée par la norme L^2 de ses dérivées, ceci n'étant bien sûr valable ici que pour les fonctions nulles au bord. Ce résultat se généralise de la façon suivante

Théorème I.15 (Inégalité de Poincaré)

Soit Ω un ouvert borné et lipschitzien de \mathbb{R}^d . Il existe une constante $C > 0$ telle que

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

En conséquence de quoi, l'application $u \mapsto \|\nabla u\|_{L^2}$ est une norme sur $H_0^1(\Omega)$, équivalente à la norme usuelle. Il existe de nombreuses variantes de ce résultat, dont les plus utiles sont peut-être les suivantes :

Théorème I.16

Soit Ω un ouvert borné et Lipschitzien de \mathbb{R}^d .

1. Soit Γ_D une partie fermée d'intérieur non vide dans $\partial\Omega$ (pour la topologie induite). On note

$$H_D^1(\Omega) = \{u \in H^1(\Omega), \gamma_0(u) = 0, \text{ sur } \Gamma_D\}.$$

Alors il existe $C > 0$ telle que

$$\|u\|_{L^2} \leq C \|\nabla u\|_{L^2}, \quad \forall u \in H_D^1(\Omega).$$

2. On note

$$H_m^1(\Omega) = \{u \in H^1(\Omega), \int_{\Omega} u \, dx = 0\}.$$

Alors il existe $C > 0$ telle que

$$\|u\|_{L^2} \leq C \|\nabla u\|_{L^2}, \quad \forall u \in H_m^1(\Omega).$$

1.5 Champs L^2 à divergence L^2 . Trace normale

On définit l'espace

$$H_{\text{div}}(\Omega) = \{v \in (L^2(\Omega))^d, \text{div } v \in L^2(\Omega)\},$$

muni de la norme

$$\|v\|_{H_{\text{div}}} = (\|v\|_{L^2}^2 + \|\text{div } v\|_{L^2}^2)^{\frac{1}{2}}.$$

Cet espace est un espace de Hilbert et nous avons le résultat de densité suivant :

Théorème I.17

Soit Ω un ouvert borné régulier. L'ensemble des champs de vecteurs de classe $(C^\infty(\overline{\Omega}))^d$ est dense dans $H_{\text{div}}(\Omega)$.

Pour démontrer ce résultat on utilise une conséquence du théorème de Hahn-Banach.

Lemme I.18

Soit E un espace de Banach et F un sous-espace de E . Alors F est dense dans E si et seulement si toute forme linéaire continue sur E qui s'annule sur F est identiquement nulle.

Preuve (du théorème):

Soit donc Φ une forme linéaire continue sur H_{div} qui s'annule sur $(C^\infty(\overline{\Omega}))^d$. Comme $H_{\text{div}}(\Omega)$ est un espace de Hilbert, le théorème de représentation de Riesz nous dit que Φ se représente comme un élément f de $H_{\text{div}}(\Omega)$ par

$$\langle \Phi, u \rangle_{H_{\text{div}}, H_{\text{div}}} = \int_{\Omega} f \cdot u \, dx + \int_{\Omega} (\text{div } f)(\text{div } u) \, dx.$$

Par hypothèse, on a donc

$$\int_{\Omega} f \cdot \varphi \, dx + \int_{\Omega} (\text{div } f)(\text{div } \varphi) \, dx = 0,$$

pour tout $\varphi \in (C^\infty(\overline{\Omega}))^d$. Si on note \bar{f} le prolongement par 0 de f à \mathbb{R}^d et $\overline{(\text{div } f)}$ celui de $\text{div } f$, on obtient

$$\int_{\mathbb{R}^d} \bar{f} \cdot \varphi \, dx + \int_{\mathbb{R}^d} \overline{(\text{div } f)}(\text{div } \varphi) \, dx = 0,$$

pour tout $\varphi \in (\mathcal{D}(\mathbb{R}^d))^d$. Ceci montre en particulier que $\overline{(\text{div } f)}$ est dans $H^1(\mathbb{R}^d)$ et que l'on a

$$\nabla \overline{(\text{div } f)} = \bar{f}.$$

Ainsi $\overline{(\operatorname{div} f)}$ est dans $H^1(\mathbb{R}^d)$, donc la fonction $g \equiv \operatorname{div} f$ est dans $H_0^1(\Omega)$ et son gradient n'est autre que f . Ainsi la forme linéaire F se réécrit

$$\langle F, u \rangle_{H'_{\operatorname{div}}, H_{\operatorname{div}}} = \int_{\Omega} u \cdot \nabla g \, dx + \int_{\Omega} (\operatorname{div} u) g \, dx,$$

où g est une fonction de $H_0^1(\Omega)$. Or, $\mathcal{D}(\Omega)$ est dense dans $H_0^1(\Omega)$, il existe donc une suite $g_n \in \mathcal{D}(\Omega)$ qui converge vers g dans $H_0^1(\Omega)$. Mais pour tout n , on a

$$\int_{\Omega} u \cdot \nabla g_n \, dx + \int_{\Omega} (\operatorname{div} u) g_n \, dx = 0,$$

par définition de la divergence au sens des distributions. On peut donc passer à la limite par rapport à n et obtenir le résultat $F = 0$, ce qui prouve le résultat de densité attendu. ■

On rappelle que l'espace $H^{\frac{1}{2}}(\partial\Omega)$ est l'image de l'application *trace* des fonctions de $H^1(\Omega)$ sur $L^2(\partial\Omega)$ et que la norme sur cet espace peut être définie par

$$\|\varphi\|_{H^{\frac{1}{2}}(\partial\Omega)} = \inf \{ \|u\|_{H^1}, u \in H^1(\Omega), \text{ tq } u = \varphi \text{ sur } \partial\Omega \text{ au sens des traces} \}.$$

On note dorénavant $H^{-\frac{1}{2}}(\partial\Omega)$, par définition, le dual de $H^{\frac{1}{2}}(\partial\Omega)$ après identification de $L^2(\partial\Omega)$ à son dual.

Proposition I.19

Soit Ω un ouvert lipschitzien de \mathbb{R}^d . L'application γ_{ν} qui à $u \in (C^\infty(\overline{\Omega}))^d$ associe $(u \cdot \nu)|_{\partial\Omega}$ (où ν est la normale sortante au bord de l'ouvert Ω) se prolonge en une application linéaire continue de $H_{\operatorname{div}}(\Omega)$ dans $H^{-\frac{1}{2}}(\partial\Omega)$ et on a la formule de Stokes suivante :

$$\forall u \in H_{\operatorname{div}}(\Omega), \forall w \in H^1(\Omega), \int_{\Omega} u \cdot \nabla w \, dx + \int_{\Omega} w \operatorname{div} u \, dx = \langle \gamma_{\nu} u, \gamma_0 w \rangle_{H^{-\frac{1}{2}}, H^{\frac{1}{2}}},$$

où γ_0 est l'opérateur de trace de $H^1(\Omega)$ dans $H^{\frac{1}{2}}(\partial\Omega)$.

Remarque I.20

Si $u \in (H^1(\Omega))^d$, on a en particulier $u \in H_{\operatorname{div}}$ et donc on a, a priori, deux définitions possibles de la trace normale. En réalité, les deux définitions coïncident et on a donc

$$\gamma_0(u) \cdot \nu = \gamma_{\nu} u.$$

Dans ce cas, $\gamma_{\nu} u$ est un élément de $H^{\frac{1}{2}}(\partial\Omega)$.

Preuve :

On rappelle qu'il existe un opérateur de relèvement linéaire continu de $H^{\frac{1}{2}}(\partial\Omega)$ dans $H^1(\Omega)$ noté R_0 qui vérifie par définition

$$\gamma_0 \circ R_0 = \operatorname{Id}_{H^{\frac{1}{2}}(\partial\Omega)}.$$

voir le théorème I.9.

- Pour tout $u \in H_{\operatorname{div}}(\Omega)$ et $\varphi \in H^{\frac{1}{2}}(\partial\Omega)$, on pose

$$X_u(\varphi) = \int_{\Omega} (R_0(\varphi) \operatorname{div} u + u \cdot \nabla R_0(\varphi)) \, dx.$$

Il vient

$$|X_u(\varphi)| \leq \|R_0(\varphi)\|_{H^1} \|u\|_{H_{\operatorname{div}}} \leq C \|\varphi\|_{H^{\frac{1}{2}}(\partial\Omega)} \|u\|_{H_{\operatorname{div}}},$$

par continuité de l'opérateur de relèvement R_0 . Ceci montre, u étant fixé, que X_u est une forme linéaire continue sur $H^{\frac{1}{2}}(\partial\Omega)$ qui vérifie en outre

$$\|X_u\|_{H^{-\frac{1}{2}}(\partial\Omega)} \leq C \|u\|_{H_{\operatorname{div}}}.$$

Cette dernière inégalité exprime que l'application qui à $u \in H_{\text{div}}(\Omega)$ associe $X_u \in H^{-\frac{1}{2}}(\partial\Omega)$ est linéaire et continue. On notera

$$\gamma_{\nu}(u) = X_u.$$

Il reste à vérifier que cet opérateur γ_{ν} vérifie la propriété annoncée.

- Si $u \in (C^{\infty}(\overline{\Omega}))^d$, et si $w_1, w_2 \in H^1(\Omega)$ sont tels que $\gamma_0(w_1) = \gamma_0(w_2)$, alors on a par intégration par parties

$$\int_{\Omega} ((w_1 - w_2) \operatorname{div} u + u \cdot \nabla (w_1 - w_2)) dx = \int_{\partial\Omega} \gamma_0(w_1 - w_2) (u \cdot \nu) d\sigma = 0.$$

Ceci montre en particulier que pour tout $w \in H^1(\Omega)$, en prenant $w_1 = w$ et $w_2 = R_0(\gamma_0 w)$, on a

$$\begin{aligned} \langle (u \cdot \nu), \gamma_0 w \rangle &= \int_{\Omega} (w \operatorname{div} u + u \cdot \nabla w) dx \\ &= \int_{\Omega} (R_0(\gamma_0(w)) \operatorname{div} u + u \cdot \nabla R_0(\gamma_0(w))) dx \\ &= X_u(\gamma_0(w)) = \langle \gamma_{\nu} u, \gamma_0 w \rangle_{H^{-\frac{1}{2}}, H^{\frac{1}{2}}}. \end{aligned}$$

Ceci étant vrai pour toute fonction régulière, on conclut par densité grâce à la continuité de l'opérateur γ_{ν} . ■

Théorème I.21

Soit Ω un ouvert borné lipschitzien de \mathbb{R}^d et soit $H_{0,\text{div}}(\Omega)$ la fermeture de $(\mathcal{D}(\Omega))^d$ dans $H_{\text{div}}(\Omega)$. Alors :

$$H_{0,\text{div}}(\Omega) = \operatorname{Ker}(\gamma_{\nu}).$$

Preuve :

Il est clair que $H_{0,\text{div}}$ est inclus dans $\operatorname{Ker}(\gamma_{\nu})$, car $(\mathcal{D}(\Omega))^d \subset \operatorname{Ker}(\gamma_{\nu})$ par définition de l'opérateur de traces.

Montrons maintenant la seconde inclusion. Soit donc $u \in H_{\text{div}}(\Omega)$, tel que $\gamma_{\nu}(u) = 0$. Comme Ω est Lipschitzien, on peut montrer qu'il existe un recouvrement ouvert fini de $\overline{\Omega}$ par des ouverts ω_i tels que les $\omega_i \cap \Omega$ soient étoilés. On introduit alors une partition de l'unité $(\alpha_i)_i$ associée à ce recouvrement.

Soit maintenant $\varphi \in \mathcal{D}(\mathbb{R}^d)$, que l'on restreint à Ω , on applique la formule de Stokes donnée dans la proposition I.19, ce qui donne

$$\int_{\Omega} \varphi \operatorname{div} u dx + \int_{\Omega} u \cdot \nabla \varphi dx = 0.$$

Remplaçons dans cette formule, φ par $\alpha_i \varphi$, on obtient

$$\int_{\Omega \cap \omega_i} \alpha_i \varphi \operatorname{div} u dx + \int_{\Omega \cap \omega_i} u \cdot (\varphi \nabla \alpha_i + \alpha_i \nabla \varphi) dx = 0.$$

Ecrivons cela de la façon suivante

$$\int_{\mathbb{R}^d} (1_{\Omega \cap \omega_i} \alpha_i \operatorname{div} u + 1_{\Omega \cap \omega_i} u \cdot \nabla \alpha_i) \varphi dx + \int_{\mathbb{R}^d} (1_{\Omega \cap \omega_i} u \alpha_i) \cdot \nabla \varphi dx = 0.$$

Ceci étant valable pour toute fonction test φ , exprime le fait que la fonction $v_i = 1_{\Omega \cap \omega_i} u \alpha_i$ (qui est bien sûr dans $(L^2(\mathbb{R}^d))^d$) a une divergence au sens des distributions donnée par $1_{\Omega \cap \omega_i} \alpha_i \operatorname{div} u + 1_{\Omega \cap \omega_i} u \cdot \nabla \alpha_i$ qui est également dans $L^2(\mathbb{R}^d)$. Ainsi la fonction v_i est dans l'espace $H_{\text{div}}(\mathbb{R}^d)$ et a son support dans $\Omega \cap \omega_i$ qui est un ouvert étoilé par hypothèse. Pour simplifier les notations, on suppose que cet ouvert est étoilé par rapport à 0. On introduit alors la famille de fonctions définie par homothétie pour tout $0 < \theta < 1$, par

$$v_i^{\theta}(x) = v_i\left(\frac{x}{\theta}\right).$$

Comme le support de v_i est étoilé, les fonctions v_i^{θ} ont un support compact dans $\Omega \cap \omega_i$. Il est clair par ailleurs que v_i^{θ} converge vers v_i dans $H_{\text{div}}(\mathbb{R}^d)$ (et donc dans $H_{\text{div}}(\Omega)$) quand θ tend vers 1. Si $\varepsilon > 0$ est fixé, on peut donc trouver $0 < \theta < 1$ tel que

$$\|v_i - v_i^{\theta}\|_{H_{\text{div}}} \leq \varepsilon.$$

Mais comme v_i^θ a un support compact dans $\Omega \cap \omega_i$, on peut maintenant utiliser un procédé de convolution pour régulariser cette fonction sans que le support de la fonction ne sorte de l'ouvert $\Omega \cap \omega_i$, ainsi il existe $\eta > 0$ tel que

$$\rho_\eta \star v_i^\theta \in (\mathcal{D}(\Omega \cap \omega_i))^d,$$

et

$$\|\rho_\eta \star v_i^\theta - v_i^\theta\|_{H_{\text{div}}} \leq \varepsilon.$$

Au final, cela donne

$$\|v_i - \rho_\eta \star v_i^\theta\|_{H_{\text{div}}} \leq 2\varepsilon.$$

Ceci permet donc d'approcher chaque v_i par des fonctions de $(\mathcal{D}(\Omega))^d$ pour la norme de $H_{\text{div}}(\Omega)$. Comme les α_i forment une partition de l'unité, on peut donc écrire

$$u = \sum_i u \alpha_i = \sum_i v_i,$$

et comme la somme est finie, le processus d'homothétie/régularisation employé sur chaque v_i permet de conclure.

■

2 Formulations faibles de problèmes elliptiques

2.1 Introduction

Quand on cherche la solution d'une EDP, on peut être confrontés aux problèmes suivants :

- La solution ou les coefficients de l'EDP peuvent ne pas être assez réguliers pour donner un sens classique à l'équation.
- La solution peut être régulière mais l'espace de régularité en question n'a pas les bonnes propriétés pour obtenir directement l'existence de la solution dans cet espace.

On va donc essayer de donner un sens plus faible à la notion de solution, sans perdre de vue si possible les notions les plus naturelles. Donner un sens faible à une EDP signifie :

1. Chercher une solution dans un espace de régularité plus faible que souhaité.
2. Etablir, en général à l'aide de fonction tests et d'intégrations par parties "formelles", une série d'équations que doit vérifier cette solution.
3. Cette série d'équations doit être construite (si possible) de telle sorte que les éventuelles solutions régulières soient aussi solutions faibles et, a contrario, que des solutions faibles qui seraient régulières soient aussi solutions classiques du problème initial.

Les difficultés d'une telle approche sont nombreuses :

- Le choix de l'espace fonctionnel dans lequel on va chercher la solution est crucial, pas toujours unique et parfois non trivial.
- Si on affaiblit trop la notion de solution, il sera a priori plus facile de démontrer l'existence mais les propriétés d'unicité seront plus difficiles à obtenir.
- Si on n'affaiblit pas suffisamment l'équation, l'existence sera ardue à prouver mais l'unicité sera plus simple.

Il s'agit donc de trouver un juste équilibre ... Les équations elliptiques, les seules que nous allons considérer dans ce cours, ont pour propriété, en général, de mettre en jeu des dérivées d'ordre pair de la solution dans le terme principal de l'équation. L'idée première pour résoudre ces problèmes aussi bien d'un point de vue théorique que numérique va être de multiplier formellement l'équation par une fonction test régulière, puis d'intégrer sur le domaine et enfin d'intégrer par parties un nombre suffisant de fois pour faire porter autant de dérivées sur la solution que sur les fonctions tests. On pourra ainsi envisager d'utiliser le même espace fonctionnel pour la solution et les fonctions tests.

2.2 Théorème de Lax-Milgram

L'un des outils qui rend cette approche utilisable est le théorème central suivant :

Théorème I.22 (Lax-Milgram)

Soit H un espace de Hilbert réel, a une forme bilinéaire sur H , L une forme linéaire sur H . On suppose que

1. a est continue :

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

2. a est coercive : il existe $\alpha > 0$ tel que

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

3. L est continue :

$$\forall u \in H, |L(u)| \leq \|L\| \|u\|_H.$$

Alors, il existe un unique u dans H qui vérifie

$$\forall v \in H, a(u, v) = L(v), \quad (\text{I.2})$$

et celui-ci vérifie

$$\|u\|_H \leq \frac{\|L\|}{\alpha}.$$

Si de plus a est symétrique, alors u est aussi l'unique élément de H qui minimise la fonctionnelle

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

Ce théorème nous donne donc des conditions suffisantes (mais non nécessaires comme on le verra plus tard), pour trouver une solution (unique !) à un problème abstrait de la forme (I.2).

Dans les problèmes d'équations aux dérivées partielles qui nous occupent, H sera un espace fonctionnel (L^2 , H^1 , H_0^1 , etc ...), u sera la solution (faible !) que l'on cherche et les fonctions v seront des fonctions test pour notre problème. La forme bilinéaire a s'exprimera en général sous la forme d'intégrales contenant des dérivées de u et de v , la forme linéaire L contiendra les termes sources et certains termes de bord.

Quelques remarques s'imposent :

- Pour un problème donné, le bon choix de a et de H est une question cruciale pour plusieurs raisons :
 - Comme on le verra, une partie de la prise en compte des conditions aux limites du problème de départ se fera *via* le choix de l'espace fonctionnel.
 - La régularité des fonctions dans l'espace H ne peut être complètement déconnectée de la forme bilinéaire a : si les fonctions de H sont trop régulières, on n'aura pas la propriété de coercivité, alors que si elles ne sont pas assez régulières, on perdra la continuité de a . Exemples :
 - * Prenons $a(u, v) = \int_{\Omega} u(-\Delta v) dx$. Pour que celle-ci ait un sens il faut pouvoir définir Δv , ce qui incite à choisir $H = H^2(\Omega) \cap H_0^1(\Omega)$ (l'espace $H_0^1(\Omega)$ est juste là pour imposer une trace nulle au bord). Ceci fournira bien une forme bilinéaire continue, d'après l'inégalité de Cauchy-Schwarz

$$|a(u, v)| \leq \|u\|_{L^2} \|\Delta v\|_{L^2} \leq \|u\|_{L^2} \|v\|_{H^2} \leq \|u\|_H \|v\|_H.$$

Par contre, nous n'avons pas la coercivité. En effet, si $u \in H$, on a

$$a(u, u) = \int_{\Omega} u(-\Delta u) dx = \int_{\Omega} u \operatorname{div}(-\nabla u) dx = \int_{\partial\Omega} u(-\nabla u) \cdot \boldsymbol{\nu} dx + \int_{\Omega} |\nabla u|^2 dx = \int_{\Omega} |\nabla u|^2 dx.$$

Il est clair qu'il ne peut exister de $\alpha > 0$ tel que

$$\|\nabla u\|_{L^2}^2 \geq \alpha \|u\|_H^2,$$

ce qui impliquerait que $H^2(\Omega) \cap H_0^1(\Omega)$ serait égal à $H_0^1(\Omega)$.

* Prenons maintenant $a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx$ et $H = H^1(\Omega)$. Dans ce cas, la continuité est assurée mais la coercivité est en défaut car $a(u, u) = 0$ si u est constante. Le problème n'est donc pas coercif.

- Pour pouvoir appliquer le théorème de Lax-Milgram, il faut que u et v soient dans le même espace fonctionnel, ce qui explique que lors de la fabrication de la formulation faible à partir d'une EDP, on essaie d'équilibrer les dérivées entre la fonction solution et la fonction test.

2.3 Exemples de base

Problème de Poisson avec condition aux limites de Dirichlet homogène : Soit à résoudre

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega, \\ u = 0, & \text{sur } \partial\Omega, \end{cases} \quad (\text{I.3})$$

avec $f \in L^2(\Omega)$.

Prenons une fonction test v suffisamment régulière pour le moment et effectuons tous les calculs de façon formelle. On multiplie l'équation par v , on l'intègre sur Ω puis on effectue une intégration par parties.

$$\int_{\Omega} f v \, dx = \int_{\Omega} \nabla u \cdot \nabla v \, dx - \int_{\partial\Omega} v(\nabla u \cdot \nu) \, dx. \quad (\text{I.4})$$

Regardons de plus près la chose :

- Si on ne regarde que les termes de volume, une seule dérivée de u et v est présente. Ainsi, si on veut la continuité et la coercivité de la forme bilinéaire, l'espace H sur lequel on va travailler risque fort d'être un sous-espace de $H^1(\Omega)$.
- On n'a pas de condition *a priori* sur $\nabla u \cdot \nu$. Si on laisse le second terme tel quel, il faudra donc pouvoir contrôler ce terme en fonction de la norme de u dans H . On a vu que ceci n'est possible que si ∇u est dans L^2 et si sa divergence est dans L^2 . Ceci suggère donc de prendre pour H le sous-espace de $H^1(\Omega)$ constitué des fonctions dont le laplacien est dans L^2 et muni de la norme $\|u\|_H^2 = \|u\|_{H^1}^2 + \|\Delta u\|_{L^2}^2$. On peut alors vérifier que la coercivité n'est pas satisfaite dans ce cadre.
- De toutes les façons l'approche tentée précédemment ne permet pas, *a priori*, de prendre en compte les conditions aux limites du problème.

La bonne démarche consiste à se rappeler que l'on souhaite *in fine* prendre u et v dans le même espace. Comme on veut que u soit nulle au bord, l'idée est de considérer un espace de fonctions tests nulles au bord. Ceci permet du même coup de faire disparaître (formellement) le second terme dans la formulation faible.

Bilan des courses : on va choisir $H = H_0^1(\Omega)$ et

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

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

C'est maintenant un simple exercice de vérifier que les hypothèses du théorème de Lax-Milgram sont satisfaites (on utilise ici l'inégalité de Poincaré !!).

Exercice I.23

Soit $f \in L^p(\Omega)$ avec $p \in [1, \infty]$. Pour quelles valeurs de p peut-on encore mettre en oeuvre la démarche précédente ?

Que fait-on une fois que l'on a obtenu une fonction $u \in H = H_0^1(\Omega)$, vérifiant $a(u, v) = L(v)$ pour tout $v \in H$? Tout d'abord, $\nabla u \in (L^2(\Omega))^d$ et on a $\mathcal{D}(\Omega) \subset H$, on peut donc prendre $v = \varphi \in \mathcal{D}(\Omega)$ comme fonction test dans le problème. On obtient donc

$$\int_{\Omega} \nabla u \cdot \nabla \varphi \, dx = \int_{\Omega} f \varphi \, dx.$$

Ceci, par définition, montre que la divergence de ∇u au sens des distributions est égale à $-f$. La fonction u est donc bien solution de $-\Delta u = f$ au sens des distributions.

On peut en réalité aller encore plus loin grâce au résultat (assez difficile) suivant.

Théorème I.24 (propriété de régularité elliptique)

On suppose que Ω est un ouvert de classe C^2 de \mathbb{R}^d . Soit $f \in L^2(\Omega)$ et $u \in H_0^1(\Omega)$ l'unique solution variationnelle de (I.3).

On a $u \in H^2(\Omega)$ et il existe $C > 0$ ne dépendant que de Ω telle que

$$\|u\|_{H^2} \leq C \|f\|_{L^2}.$$

Dans les conditions de ce théorème, toutes les dérivées partielles secondes (faibles) de u sont donc des fonctions de L^2 . Il faut remarquer que l'hypothèse de régularité de l'ouvert Ω est importante. Elle peut être affaiblie (par exemple si l'ouvert est polygonal convexe) mais il y a des contre-exemples. Typiquement, si l'ouvert Ω possède un coin rentrant, alors il existe des fonctions harmonique qui se comportent en $r^{\frac{\pi}{\alpha}}$ au voisinage du coin en question avec $\alpha \in]\pi, 2\pi[$. On peut vérifier qu'une telle fonction n'est pas dans $H^2(\Omega)$.

Problème de Poisson avec condition aux limites de Dirichlet non-homogène Soit à résoudre

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega, \\ u = g, & \text{sur } \partial\Omega, \end{cases} \quad (\text{I.5})$$

avec $f \in L^2(\Omega)$ et $g \in H^{1/2}(\partial\Omega)$. Pourquoi choisit-on un g dans cet espace ? la raison est que l'on sait par expérience du cas précédent que la solution que l'on va obtenir sera probablement dans l'espace $H^1(\Omega)$. Sa trace sur le bord de Ω sera donc un objet de $H^{1/2}(\partial\Omega)$. Si on voulait résoudre ce même problème avec une donnée au bord moins régulière (ce qui peut tout à fait arriver !), on sortirait du cadre purement variationnel à la Lax-Milgram et les affaires se corseraient sérieusement.

Si on suppose donc $g \in H^{1/2}(\partial\Omega)$, d'après le théorème de traces, on peut trouver un relèvement de cette trace que l'on note $R_0g \in H^1(\Omega)$; il vérifie donc par définition $\gamma_0 R_0g = g$. On cherche maintenant u sous la forme $u = \tilde{u} + R_0g$, ce qui ramène le problème à trouver \tilde{u} vérifiant

$$\begin{cases} -\Delta \tilde{u} = f + \Delta(R_0g), & \text{dans } \Omega, \\ \tilde{u} = 0, & \text{sur } \partial\Omega. \end{cases}$$

Il se trouve que, R_0g étant un élément de $H^1(\Omega)$, nous avons $\nabla(R_0g) \in L^2(\Omega)^d$ et donc $\Delta(R_0g) = \operatorname{div}(\nabla(R_0g)) \in H^{-1}(\Omega)$ d'après la proposition I.4 et la remarque I.5. Ainsi, le problème ci-dessus relève encore du théorème de Lax-Milgram avec $H = H_0^1(\Omega)$ et

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

$$L(\tilde{v}) = \int_{\Omega} f \tilde{v} \, dx + \langle \Delta(R_0g), \tilde{v} \rangle_{H^{-1}, H_0^1}.$$

On vérifie que le théorème de Lax-Milgram s'applique dans ce contexte et donc que l'on a bien existence et unicité de \tilde{u} et donc de $u = \tilde{u} + R_0g$.

Remarquons tout de même que, d'après la caractérisation de $H^{-1}(\Omega)$, on peut écrire le terme de dualité dans L de la façon suivante :

$$\langle \Delta(R_0g), \tilde{v} \rangle_{H^{-1}, H_0^1} = - \int_{\Omega} \nabla(R_0g) \cdot \nabla \tilde{v} \, dx.$$

Ainsi, en introduisant l'espace $H_g = \{u \in H^1(\Omega), \gamma_0 u = g\}$, la solution $u = \tilde{u} + R_0g$ est l'unique fonction vérifiant

$$u \in H_g, \quad \text{et} \quad \int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx, \quad \forall v \in H_0^1(\Omega). \quad (\text{I.6})$$

On ne pouvait néanmoins pas appliquer directement le théorème de Lax-Milgram à ce problème car l'espace H_g^1 n'est pas un espace vectoriel et donc encore moins un espace de Hilbert !

Il n'en reste pas moins que la formulation (I.6) a l'immense avantage de ne pas faire apparaître de relèvement de la donnée au bord, dont le calcul serait fort pénible en pratique. On verra que c'est cette formulation qui est utilisée d'un point de vue numérique.

Le théorème de régularité elliptique I.24 reste valable à condition de supposer que la donnée au bord g est dans l'espace $H^{3/2}(\partial\Omega)$, qui est l'image par l'opérateur de traces de l'espace $H^2(\Omega)$.

Problème de Poisson avec conditions aux limites de Neumann : Soit à résoudre

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega, \\ \nabla u \cdot \nu = g, & \text{sur } \partial\Omega, \end{cases} \quad (\text{I.7})$$

avec $f \in L^2(\Omega)$ et $g \in H^{-1/2}(\partial\Omega)$. La régularité demandée à g est, là-aussi, assez naturelle. En effet, on s'attend à trouver u dans l'espace de Sobolev $H^1(\Omega)$ (ou dans un de ces sous-espaces ...) de telle sorte que $\nabla u \in (L^2(\Omega))^d$. Comme de plus, on demande que $\operatorname{div}(\nabla u) = \Delta u = -f \in L^2(\Omega)$, on voit que si la solution existe, on aura nécessairement $\nabla u \in H_{\operatorname{div}}(\Omega)$, ce qui permet de donner un sens faible à la trace sur le bord de $(\nabla u) \cdot \nu$ dans l'espace $H^{-\frac{1}{2}}(\partial\Omega)$.

On peut ici faire deux remarques préalables.

1. Si une solution u de (I.7) existe, alors on peut intégrer l'équation sur Ω et obtenir (si toutes les données sont régulières) :

$$\int_{\Omega} f \, dx = \int_{\Omega} (-\Delta u) \, dx = - \int_{\partial\Omega} (\nabla u) \cdot \nu \, dx = - \int_{\partial\Omega} g \, dx.$$

Ceci prouve qu'il existe une condition dite **de compatibilité** entre les données qu'il est nécessaire de vérifier pour espérer obtenir l'existence d'une solution. Dans le cas où g est seulement dans $H^{-\frac{1}{2}}(\partial\Omega)$, celle-ci s'écrit

$$\int_{\Omega} f \, dx = -\langle g, 1 \rangle_{H^{-1/2}, H^{1/2}}. \quad (\text{I.8})$$

2. Il n'y a aucune chance pour que la solution soit unique car l'ajout d'une constante à toute solution éventuelle en fournit une nouvelle. Il faudra donc imposer une condition supplémentaire sur u pour espérer montrer l'unicité de la solution. Une façon naturelle de procéder est d'imposer

$$\int_{\Omega} u \, dx = 0. \quad (\text{I.9})$$

Reprenons la démarche générale : on choisit une fonction test v a priori quelconque et on reprend le calcul effectué en (I.4). Cette fois-ci, on a une condition sur $\nabla u \cdot \nu$ qui doit être égal à g . En revanche, on n'a plus aucune condition sur la valeur de la solution au bord, on n'a donc aucune raison d'imposer une quelconque condition à la fonction test v . Par contre, on a imposé en (I.9) que la moyenne de u soit nulle, il nous faut donc imposer la même condition sur les fonctions tests.

On va donc choisir $H = H_m^1(\Omega) = \{v \in H^1(\Omega), \int_{\Omega} v \, dx = 0\}$ et

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

$$L(v) = \int_{\Omega} f v \, dx + \langle g, v \rangle_{H^{-1/2}, H^{1/2}}.$$

On peut vérifier aisément, en utilisant l'inégalité de Poincaré-moyenne, que ces définitions permettent l'utilisation du théorème de Lax-Milgram.

En réalité, une fois que l'on a obtenu $u \in H_m^1(\Omega)$ solution du problème abstrait, on peut retrouver une formulation avec toutes les fonctions tests de $H^1(\Omega)$. En effet, si $v \in H^1(\Omega)$ est quelconque, on pose $\tilde{v} = v - m(v)$, où $m(v)$ désigne la moyenne de v sur Ω . On a alors

$$\int_{\Omega} \nabla u \cdot \nabla \tilde{v} \, dx = \int_{\Omega} f \tilde{v} \, dx + \langle g, \tilde{v} \rangle_{H^{-1/2}, H^{1/2}},$$

ce qui donne, en faisant apparaître v (qui vérifie $\nabla v = \nabla \tilde{v}$) :

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx - m(v) \left(\int_{\Omega} f \, dx \right) + \langle g, v \rangle_{H^{-1/2}, H^{1/2}} - m(v) \langle g, 1 \rangle_{H^{-1/2}, H^{1/2}},$$

mais d'après la condition de compatibilité (I.8) on voit que les termes en $m(v)$ disparaissent, et il reste finalement

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx + \langle g, v \rangle_{H^{-1/2}, H^{1/2}}, \quad (\text{I.10})$$

ce calcul étant valable pour toute fonction test de $v \in H^1(\Omega)$.

On peut commencer à prendre des fonctions test $v = \varphi \in \mathcal{D}(\Omega)$ comme pour le problème de Dirichlet, ce qui a pour effet de supprimer le terme de bord dans la formulation. On obtient donc que l'équation $-\Delta u = f$ est vérifiée au sens des distributions. Comme on sait que $\nabla u \in (L^2(\Omega))^d$ et que $f \in L^2(\Omega)$, l'équation montre que l'on a $\nabla u \in H_{\text{div}}(\Omega)$. Dans ces conditions la trace de $\nabla u \cdot \nu$ a un sens dans $H^{-1/2}(\partial\Omega)$ et on a vu que la formule de Stokes suivante est vraie

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\Omega} \operatorname{div}(\nabla u) v \, dx = \langle \nabla u \cdot \nu, v \rangle_{H^{-1/2}, H^{1/2}}.$$

En comparant cette formule à l'équation (I.10), et comme $f = -\Delta u$, on a bien obtenu

$$\nabla u \cdot \nu = g, \quad \text{sur } \partial\Omega,$$

en un sens faible.

Exercice I.25

Ecrire et étudier une formulation variationnelle du problème de Poisson avec des conditions aux limites de Fourier (ou Robin) :

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega, \\ \nabla u \cdot \nu + \alpha u = g, & \text{sur } \partial\Omega, \end{cases}$$

où $\alpha > 0$ est un paramètre donné.

2.4 Exemples plus sophistiqués

Les éléments illustrés précédemment dans les exemples de base permettent ensuite d'analyser bon nombre de problèmes issus de la physique. On présente ici, sans justification détaillée, certains de ces problèmes et leur formulation faible que l'on peut analyser par le théorème de Lax-Milgram.

Les opérateurs elliptiques linéaires généraux : Soit $x \in \Omega \mapsto A(x)$ une application mesurable bornée à valeurs matricielles. On suppose que les matrices $A(x)$ sont symétriques définies positives et uniformément elliptiques :

$$\exists \alpha > 0, \quad (A(x)\xi, \xi) \geq \alpha |\xi|^2, \quad \forall \xi \in \mathbb{R}^d, \text{ pour presque tout } x \in \Omega.$$

On s'intéresse au problème

$$\begin{cases} -\operatorname{div}(A(x)\nabla u) = f, & \text{dans } \Omega, \\ u = 0, & \text{sur } \partial\Omega, \end{cases}$$

Pour des conditions aux limites de Dirichlet homogènes, la formulation faible est donnée par $H = H_0^1(\Omega)$ et

$$a(u, v) = \int_{\Omega} (A(x)\nabla u, \nabla v) \, dx, \quad L(v) = \int_{\Omega} f v \, dx, \quad \forall u, v \in H.$$

L'adaptation au cas de condition non-homogène se fait dans les mêmes conditions que dans le paragraphe précédent. Pour les conditions de type Neumann, la seule difficulté consiste à comprendre que la condition s'écrit

$$A(x)\nabla u \cdot \nu = g.$$

Enfin, pour démontrer un théorème de régularité elliptique pour ce problème, il faut supposer, au minimum, que l'application A est Lipschitzienne par rapport à x .

Réaction - diffusion : Il s'agit ici d'étudier un problème de la forme $-\Delta u + \eta u = f$ avec $\eta > 0$. Notons que la condition $\eta > 0$ est importante ici. Il existe des valeurs de $\eta < 0$ et des seconds membres f pour lesquelles le problème n'admet pas de solution (c'est le problème des valeurs propres de l'opérateur $-\Delta$).

Pour les conditions de Dirichlet, on utilise exactement la même méthode que pour le Laplacien et ça ne change rien (à part un terme supplémentaire dans la formulation bien évidemment).

Pour les conditions de type Neumann

$$\begin{cases} -\Delta u + \eta u = f, & \text{dans } \Omega, \\ \nabla u \cdot \nu = g, & \text{sur } \partial\Omega, \end{cases} \quad (\text{I.11})$$

la situation est différente. En effet, si une solution u existe, on peut intégrer l'équation sur Ω et obtenir

$$\int_{\Omega} f \, dx = \eta \int_{\Omega} u \, dx - \int_{\partial\Omega} g \, dx,$$

ce qui montre qu'il n'y a *a priori* pas de condition de compatibilité sur les données pour ce problème et que de plus, contrairement au cas précédent, la moyenne de u est imposée et ne peut être fixée.

Ceci étant dit, la formulation faible s'obtient en prenant $H = H^1(\Omega)$ et

$$a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx + \eta \int_{\Omega} uv \, dx, \quad L(v) = \int_{\Omega} fv \, dx.$$

Convection - diffusion : Soit Ω un ouvert borné régulier de \mathbb{R}^d , $f \in L^2(\Omega)$ et un champ de vecteurs $b \in (C^1(\overline{\Omega}))^d$. On s'intéresse au problème suivant

$$\begin{cases} -\Delta u + b \cdot \nabla u = f, & \text{dans } \Omega, \\ u = 0, & \text{sur } \partial\Omega. \end{cases}$$

Tout d'abord, on constate que le terme de convection $b \cdot \nabla u$ contient moins de dérivées que le terme de diffusion. On a donc envie de prendre $H = H_0^1(\Omega)$ et

$$a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\Omega} (b \cdot \nabla u)v \, dx, \quad L(v) = \int_{\Omega} fv \, dx.$$

La continuité de a (qui n'est pas symétrique) ne fait aucun doute. En revanche, la coercivité est un problème plus délicat. En effet, on a

$$a(u, u) = \|\nabla u\|_{L^2}^2 + \int_{\Omega} (b \cdot \nabla u)u \, dx,$$

et il faut donc arriver à contrôler le second terme par le premier pour établir la coercivité.

- Premier calcul possible :

$$\left| \int_{\Omega} b \cdot \nabla uu \, dx \right| \leq \|b\|_{\infty} \|\nabla u\|_{L^2} \|u\|_{L^2} \leq C_{\Omega} \|b\|_{\infty} \|\nabla u\|_{L^2}^2,$$

où C_{Ω} est la constante de Poincaré sur le domaine Ω . Ainsi, si $\|b\|_{\infty} C_{\Omega} < 1$ on a bien la coercivité de la forme a .

- Deuxième calcul possible, en intégrant par parties :

$$- \int_{\Omega} b \cdot \nabla uu \, dx = \int_{\Omega} (\operatorname{div} b) \frac{u^2}{2} \, dx \leq \frac{1}{2} \|(\operatorname{div} b)^+\|_{L^{\infty}} C_{\Omega}^2 \|\nabla u\|_{L^2}^2,$$

et donc cette fois si $\|(\operatorname{div} b)^+\|_{L^{\infty}} C_{\Omega}^2 < 2$, on a bien la coercivité de a . Un cas particulier important est celui où la divergence de b est négative ou nulle.

Dans les deux cas, on peut donc appliquer le théorème de Lax-Milgram et ainsi obtenir l'existence et l'unicité de la solution.

Le bilaplacien : Soit Ω un ouvert borné régulier de \mathbb{R}^d et $f \in L^2(\Omega)$. On s'intéresse au problème suivant

$$\begin{cases} \Delta^2 u = f, & \text{dans } \Omega, \\ u = \nabla u \cdot \nu = 0, & \text{sur } \partial\Omega. \end{cases}$$

Sa formulation faible *naturelle* s'obtient en prenant $H = H_0^2(\Omega)$ (l'adhérence de $\mathcal{D}(\Omega)$ dans $H^2(\Omega)$) et

$$a(u, v) = \int_{\Omega} \Delta u \Delta v \, dx, \quad L(v) = \int_{\Omega} f v \, dx.$$

On peut montrer que $H_0^2(\Omega) = \{v \in H^2(\Omega), \gamma_0 v = 0, \text{ et } \gamma_0(\nabla v \cdot \nu) = 0\}$, la deuxième condition ayant un sens car $\nabla u \in H^1(\Omega)$! Il reste à démontrer la coercivité de la forme bilinéaire a . Autrement dit, il faut montrer qu'il existe $\alpha > 0$ telle que

$$\|\Delta u\|_{L^2} \geq \alpha \|u\|_{H^2}, \quad \forall u \in H_0^2(\Omega).$$

Il se trouve que cette inégalité n'est rien d'autre que la propriété de régularité elliptique du Laplacien avec conditions aux limites de Dirichlet (ou de Neumann d'ailleurs !).

On peut aussi s'intéresser à la même équation mais avec un autre jeu de conditions aux limites. Par exemple, considérons le problème suivant

$$\begin{cases} \Delta^2 u = f, & \text{dans } \Omega, \\ \nabla u \cdot \nu = \nabla(\Delta u) \cdot \nu = 0, & \text{sur } \partial\Omega, \end{cases}$$

pour lequel on voit immédiatement qu'il est nécessaire que la condition de compatibilité

$$\int_{\Omega} f \, dx = 0,$$

soit vérifiée. De même, on observe que l'équation est invariante par ajout d'une constante à la solution. Il faut donc travailler dans l'espace $H = \{v \in H^2(\Omega), m(v) = 0, \nabla v \cdot \nu = 0\}$ et prendre à nouveau

$$a(u, v) = \int_{\Omega} \Delta u \Delta v \, dx, \quad L(v) = \int_{\Omega} f v \, dx.$$

La propriété de coercivité est assurée encore une fois par les propriétés de régularité elliptique du problème de Neumann pour le Laplacien.

L'élasticité linéaire : On s'intéresse à l'équation décrivant le déplacement u d'un solide élastique quand on le soumet à un champ de forces. On peut montrer que le problème se modélise de la façon suivante :

Le domaine $\Omega \subset \mathbb{R}^d$ ($d = 3$ est la dimension physique, $d = 2$ est un cas plus simple) représente le solide au repos. L'inconnue $u : \Omega \mapsto \mathbb{R}^d$ est un champ de vecteurs représentant le déplacement du point matériel situé initialement en x sous l'effet des forces extérieures. Autrement dit, le point qui est en x au repos est déplacé en $x + u(x)$.

On introduit la jacobienne ∇u du champ de vecteurs u et $D(u) = \frac{1}{2}(\nabla u + (\nabla u)^t)$ le tenseur des déformations. On peut démontrer ensuite que le tenseur des contraintes σ dans le solide est donné par une loi de comportement de la forme

$$\sigma = \lambda \text{Tr}(D(u)) \text{Id} + 2\mu D(u) = \lambda(\text{div } u) \text{Id} + 2\mu D(u),$$

où λ et μ sont des coefficients appelés : coefficients de Lamé et qui décrivent le comportement élastique du solide. On fera l'hypothèse que $\lambda \geq 0$ et $\mu > 0$.

Il reste ensuite à écrire l'équation d'équilibre du solide. On montre, par des considérations de mécanique simples, que celle-ci s'écrit

$$-\text{div}(\sigma) = f.$$

Ainsi, si on rassemble les deux équations précédentes, on trouve

$$-2\mu \text{div}(D(u)) - \lambda \nabla(\text{div } u) = f.$$

Noter que l'on peut formellement écrire ce problème sous la forme suivante

$$-\mu \Delta u - (\lambda + \mu) \nabla(\text{div } u) = f,$$

mais celle-ci n'est pas adaptée aux conditions aux limites naturelles en contrainte pour ce problème.

En effet, deux conditions aux limites sont envisageables pour ce problème : une condition de Dirichlet $u = 0$ sur une partie Γ_D du bord qui correspond à une hypothèse d'adhérence complète du solide sur Γ_D ou une condition de type "sans contrainte" (de type Neumann) qui s'écrit $\sigma \cdot \nu = 0$.

On peut donc introduire l'espace fonctionnel

$$H_D = \{u \in (H^1(\Omega))^d \text{ t.q. } u = 0 \text{ sur } \Gamma_D\}.$$

et la forme bilinéaire

$$a(u, v) = \int_{\Omega} 2\mu D(u) : D(v) dx + \int_{\Omega} \lambda (\operatorname{div} u)(\operatorname{div} v) dx,$$

et la forme linéaire

$$L(v) = \int_{\Omega} f \cdot v dx.$$

Cette formulation est bien posée grâce à l'inégalité de Korn qui s'énonce ainsi :

Proposition I.26 (Inégalité de Korn)

Si Ω est suffisamment régulier et $|\Gamma_D| > 0$, il existe une constante $C > 0$ telle que pour tout $u \in H_D$, on a

$$\|\nabla u\|_{L^2} \leq C \|D(u)\|_{L^2}.$$

Preuve :

On va seulement montrer l'inégalité dans le cas où $\Gamma_D = \partial\Omega$. Dans ce cas, $H_D = (H_0^1(\Omega))^d$ et il vient

$$\int_{\Omega} |D(u)|^2 dx = \int_{\Omega} D(u) : \nabla u dx = \frac{1}{2} \int_{\Omega} |\nabla u|^2 dx + \frac{1}{2} \int_{\Omega} (\nabla u)^t : \nabla u dx.$$

Le résultat sera démontré si on établit que ce dernier terme est positif. Pour cela, on écrit

$$\int_{\Omega} (\nabla u)^t : \nabla u dx = \int_{\Omega} \sum_{i,j} \partial_i u_j \partial_j u_i dx,$$

puis on intègre par parties en utilisant les conditions aux limites

$$\int_{\Omega} (\nabla u)^t : \nabla u dx = - \int_{\Omega} \sum_{i,j} \partial_i \partial_j u_j u_i dx = - \int_{\Omega} \nabla(\operatorname{div}(u)) \cdot u dx = \int_{\Omega} (\operatorname{div} u)^2 dx \geq 0.$$

■

L'inégalité de Korn fournit la coercivité de la forme bilinéaire a et la continuité ne fait guère de doute. Le théorème de Lax-Milgram peut donc s'appliquer et on obtient l'existence d'une solution faible à cette formulation variationnelle. On peut alors vérifier qu'on a bien résolu le problème proposé.

3 Le principe du maximum faible

Pour les problèmes elliptiques du second ordre, on peut démontrer que leurs solutions faibles vérifient le principe du maximum suivant.

Théorème I.27

Soit $A(x)$ une famille mesurable et bornée de matrices uniformément elliptiques. Soit $f \in L^2(\Omega)$ et $g \in H^{\frac{1}{2}}(\partial\Omega)$. On considère l'unique solution variationnelle u dans $H^1(\Omega)$ du problème

$$-\operatorname{div}(A(x)\nabla u) = f, \text{ dans } \Omega, \quad u = g, \text{ sur } \partial\Omega.$$

Alors, si f et g sont positives presque partout, u est également positive presque partout.

Preuve :

Soit $T : \mathbb{R} \mapsto \mathbb{R}$ une fonction de classe \mathcal{C}^1 , croissante, à dérivée bornée telle que $T(s) = 0$ si et seulement si $s \geq 0$. Comme u est dans $H^1(\Omega)$, on a vu que $T(u)$ est aussi dans $H^1(\Omega)$, de plus la trace de $T(u)$ est égale à $T(\gamma_0 u) = T(g)$. Comme on a supposé que g est positive presque partout, on a donc $\gamma_0(T(u)) = T(g) = 0$ presque partout. Ainsi $T(u)$ est un élément de $H_0^1(\Omega)$.

On peut donc prendre $v = T(u)$ comme fonction test dans la formulation variationnelle du problème, ce qui donne

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

Nous avons de plus $\nabla T(u) = T'(u) \nabla u$, ce qui donne

$$\int_{\Omega} \underbrace{T'(u)}_{\geq 0} \underbrace{(A(x) \nabla u, \nabla u)}_{\geq \alpha |\nabla u|^2 \geq 0} \, dx = \int_{\Omega} \underbrace{f}_{\geq 0} \underbrace{T u}_{\leq 0} \, dx.$$

A la vue des signes des différents termes, on voit que les deux termes de l'égalité doivent être nuls. En particulier, cela implique que $\nabla T(u) = T'(u) \nabla u$ doit être nul. Ainsi $T(u)$ est un élément de $H_0^1(\Omega)$ dont le gradient est nul, cela implique que $T(u) = 0$. Ceci montre bien que u est positif presque partout. ■

On verra par la suite que ce résultat admet quelques équivalents discrets.

4 Théorème de Lax-Milgram généralisé : formulations mixtes

Dans cette section on va essayer de considérer des problèmes plus généraux que ceux que l'on peut traiter par le théorème de Lax-Milgram. De plus, le résultat qui suit va donner des conditions nécessaires et suffisantes de résolubilité d'un problème variationnel.

4.1 Enoncé et preuve du théorème général

Théorème I.28 (Banach-Necas-Babuska)

Soient V et W deux espaces de Hilbert, a une forme bilinéaire continue sur $V \times W$. Alors les propositions suivantes sont équivalentes :

1) Pour toute forme linéaire continue L sur W , il existe un unique $u \in V$ tel que

$$a(u, w) = L(w), \quad \forall w \in W. \tag{I.12}$$

2) Les deux conditions suivantes sont vérifiées :

$$\exists \alpha > 0, \quad \inf_{v \in V} \left(\sup_{w \in W} \frac{a(v, w)}{\|v\|_V \|w\|_W} \right) \geq \alpha, \tag{I.13}$$

$$(\forall v \in V, a(v, w) = 0) \Rightarrow w = 0. \tag{I.14}$$

Dans le cas où ces deux propositions sont vérifiées, l'unique solution u de (I.12) vérifie en sus

$$\|u\|_V \leq \frac{\|L\|_{W'}}{\alpha}.$$

Preuve :

On introduit l'opérateur A défini de V dans W' par

$$\langle Av, w \rangle_{W', W} = a(v, w).$$

Comme a est une forme bilinéaire continue, l'opérateur A est linéaire et continu de norme égale à $\|a\|$.

- Si 1) est vérifiée, alors l'opérateur A est bijectif. D'après le théorème de l'application ouverte, cela implique que A^{-1} est un opérateur continu. Ainsi, il existe une constante $C > 0$ telle que pour tout $L \in W'$, l'unique solution de (I.12) vérifie

$$\|u\|_V \leq C \|L\|_{W'}.$$

Soit maintenant $v \in V$ quelconque. On pose $L(w) = a(v, w)$ de sorte que v est exactement la solution de (I.12) pour cette forme linéaire. D'après l'inégalité précédente, on a

$$\|v\|_V \leq C \sup_{w \in W} \frac{a(v, w)}{\|w\|_W}.$$

Ceci étant vrai pour tout $v \in V$, on a bien prouvé (I.13) avec $\alpha = 1/C$.

De même si $a(v, w) = 0$ pour tout $v \in V$. On introduit la forme linéaire continue $L(w') = (w, w')_W$ sur W . On applique alors l'hypothèse à $v = A^{-1}L$ de sorte que

$$0 = a(A^{-1}L, w) = \langle AA^{-1}L, w \rangle_{W', W} = L(w) = \|w\|_W^2,$$

et donc $w = 0$, ce qui prouve (I.14).

- Supposons (I.13) et (I.14) vérifiées. L'hypothèse (I.14) implique que l'opérateur adjoint $A' : W \mapsto V'$ est injectif, ce qui est équivalent à la densité de l'image de l'opérateur A . En effet, utilisons pour cela la caractérisation de la densité d'un sous-espace donnée dans le lemme I.18. Pour cela, on considère une forme linéaire continue $\varphi \in (W')'$ qui s'annule sur l'image de A . Comme W est un Hilbert, il est réflexif et donc φ se représente par un élément $w \in W$ par $\varphi(L) = \langle L, w \rangle_{W', W}$. L'hypothèse implique que pour tout $v \in V$, on a $\varphi(Av) = \langle Av, w \rangle_{W', W} = a(v, w) = 0$. D'après l'hypothèse (I.14), ceci implique que $w = 0$ et donc que $\varphi = 0$, ce qui montre bien que l'image de A est dense.

Revenons maintenant à l'hypothèse (I.13) qui s'écrit aussi

$$\|Av\|_{W'} \geq \alpha \|v\|_V,$$

ce qui implique en particulier que A est injectif et que l'image de A est fermée. Comme cette image est également dense, on obtient que A est surjectif ce qui prouve bien que le problème (I.12) est bien posé. ■

Remarque I.29

Dans le cas particulier où $V = W$ et où a est coercive, on retrouve bien comme cas particulier le théorème de Lax-Milgram.

4.2 Problèmes de type point-selle ou formulations variationnelles sous contrainte

Soient deux espaces de Hilbert X et M , a une forme bilinéaire continue sur $X \times X$, b une forme bilinéaire continue sur $X \times M$. Pour toutes formes linéaires $L \in X'$ et $G \in M'$, on cherche $u \in X$ et $p \in V$ tels que

$$\begin{cases} a(u, v) + b(v, p) = L(v), \quad \forall v \in X, \\ b(u, q) = G(q), \quad \forall q \in M. \end{cases} \quad (\text{I.15})$$

La deuxième équation est souvent appelée **la contrainte** et l'inconnue p **le multiplicateur de Lagrange** associé à cette contrainte. Cette dénomination vient du fait que, si a est symétrique, le problème (I.15) est exactement l'équation d'Euler-Lagrange associé au problème de minimisation de la fonctionnelle $J(v) = \frac{1}{2}a(v, v) - L(v)$ sur le sous-espace affine Z_G de X défini par $Z = \{v \in X, b(v, q) = G(q), \quad \forall q \in M\}$.

- On reprend la notation A pour l'opérateur linéaire continu de X dans X' défini par

$$\langle Au, v \rangle_{X', X} = a(u, v).$$

On peut également introduire un opérateur linéaire continu $B : X \mapsto M'$ défini par

$$\langle Bv, p \rangle_{M', M} = b(v, p).$$

Le problème (I.15) peut alors s'écrire en termes d'opérateurs de la façon suivante : trouver $(u, p) \in X \times M$ tel que

$$\begin{cases} Au + B'p = L, \\ Bu = G, \end{cases}$$

où $B' : M \mapsto X'$ désigne l'opérateur adjoint de B , étant entendu qu'on a identifié le bidual de M à M lui-même.

- Si on suppose que A est auto-adjoint, le problème d'optimisation associé à cette formulation s'écrit

$$\text{Trouver } u \in Z_G \text{ qui minimise } J(v) = \frac{1}{2} \langle Av, v \rangle_{X', X} - \langle L, v \rangle_{X', X}, \text{ sur } Z_G,$$

$$Z_G = \{v \in X, Bv = G\}.$$

- Remarquons que le problème (I.15) peut se mettre sous la forme générale d'un problème variationnel usuel en constatant qu'il est équivalent à

$$a(u, v) + b(v, p) - b(u, q) = L(v) - G(q), \quad \forall v \in X, \forall q \in M. \quad (\text{I.16})$$

Ainsi, en posant $V = W = X \times M$ et $c((u, p), (v, q)) = a(u, v) + b(v, p) - b(u, q)$, on déduit du théorème BNB que le problème proposé est bien posé si et seulement si les deux conditions du théorème sont vérifiées par la forme c .

Notons que l'on a

$$c((u, p), (u, p)) = a(u, u),$$

et donc c n'est jamais coercive sur $X \times M$ car ce terme ne donne aucun contrôle sur le multiplicateur p .

Théorème I.30

Si a est coercive sur X tout entier (avec constante de coercivité $\alpha > 0$), alors le problème (I.15) est bien posé si et seulement si la condition suivante est vérifiée :

$$\exists \beta > 0, \quad \inf_{p \in M} \left(\sup_{v \in X} \frac{b(v, p)}{\|v\|_X \|p\|_M} \right) \geq \beta. \quad (\text{I.17})$$

De plus, l'unique solution (u, p) de (I.15) vérifie

$$\|u\|_X \leq \frac{\|L\|_{X'}}{\alpha} + \frac{1}{\beta} \left(1 + \frac{\|a\|}{\alpha} \right) \|G\|_{M'}, \quad (\text{I.18})$$

$$\|p\|_M \leq \left(1 + \frac{\|a\|}{\alpha} \right) \left(\frac{1}{\beta} \|L\|_{X'} + \frac{\|a\|}{\beta^2} \|G\|_{M'} \right). \quad (\text{I.19})$$

Preuve :

- Supposons que le problème (I.15) soit bien posé. D'après le théorème de l'application ouverte, cela implique en particulier que la solution (u, p) du problème dépend continument des données L et G . Ainsi, il existe une constante $C > 0$ telle que

$$\|u\|_X + \|p\|_M \leq C(\|L\|_{X'} + \|G\|_{M'}).$$

Soit $\tilde{p} \in M$. Prenons $L = 0$ et G définie par $G(q) = (\tilde{p}, q)_M$ et dont la norme M' n'est autre que la norme de \tilde{p} dans M . On obtient en particulier l'existence d'un $u \in X$ tel que

$$b(u, q) = (\tilde{p}, q), \quad \forall q \in M, \quad \text{et } \|u\|_X \leq C\|\tilde{p}\|_M.$$

On a alors

$$\|\tilde{p}\|_M^2 = (\tilde{p}, \tilde{p})_M = b(u, \tilde{p}) \leq C \frac{b(u, \tilde{p})}{\|u\|_X} \|\tilde{p}\|_M,$$

ce qui donne

$$\frac{1}{C} \|\tilde{p}\|_M \leq \sup_{v \in X} \frac{b(v, \tilde{p})}{\|v\|_X},$$

c'est-à-dire exactement la condition (I.17) avec $\beta = 1/C$.

- Nous allons montrer que les hypothèses du théorème impliquent que la forme bilinéaire $c((u, p), (v, q)) = a(u, v) + b(v, p) - b(u, q)$ vérifie les hypothèses (I.13)-(I.14) du théorème de Lax-Milgram généralisé (ou théorème BNB).

– Commençons par vérifier (I.14) qui est plus facile. On suppose donc que $(v, q) \in X \times M$ est tel que

$$c((u, p), (v, q)) = 0, \quad \forall (u, p) \in X \times M. \quad (\text{I.20})$$

On veut montrer que $(v, q) = (0, 0)$. Pour cela on commence par prendre $(u, p) = (v, q)$ dans (I.20), ce qui fournit $a(v, v) = 0$ et donc $v = 0$ car a est coercive sur X . Ainsi, (I.20) s'écrit maintenant $b(u, q) = 0$ pour tout $u \in X$, ce qui n'est possible, d'après (I.17), que si q est nul.

– Montrons maintenant l'hypothèse (I.13). Soit donc $(u, p) \in X \times M$. D'après l'hypothèse (I.17), il existe $\tilde{u} \in X$ tel que

$$\frac{b(\tilde{u}, p)}{\|\tilde{u}\|_X} \geq \beta \|p\|_M.$$

De plus, il est clair qu'on peut choisir la norme de \tilde{u} et ainsi prendre par exemple

$$\|\tilde{u}\|_X = \|p\|_M.$$

On pose maintenant $(v, q) = (u + \gamma \tilde{u}, p)$ et on calcule $c((u, p), (v, q))$:

$$\begin{aligned} c((u, p), (v, q)) &= a(u, u) + \gamma a(u, \tilde{u}) + \gamma b(\tilde{u}, p) \geq \alpha \|u\|_X^2 - \gamma \|a\| \|u\|_X \|\tilde{u}\|_X + \gamma \beta \|\tilde{u}\|_X^2 \\ &\geq \frac{\alpha}{2} \|u\|_X^2 - \frac{\gamma^2 \|a\|^2}{2\alpha} \|\tilde{u}\|_X^2 + \gamma \beta \|\tilde{u}\|_X^2. \end{aligned}$$

Ainsi, en choisissant $\gamma < \frac{\alpha\beta}{\|a\|^2}$, il reste

$$c((u, p), (v, q)) \geq \frac{\alpha}{2} \|u\|_X^2 + \frac{\gamma\beta}{2} \|\tilde{u}\|_X^2 = \frac{\alpha}{2} \|u\|_X^2 + \frac{\gamma\beta}{2} \|p\|_M^2 \geq \delta \|(u, p)\|_{X \times M}^2,$$

où δ ne dépend que de γ, β, α et $\|a\|$. De plus, nous avons

$$\|(v, q)\|_{X \times M} = \|v\|_X + \|q\|_M \leq \|u\|_X + \gamma \|\tilde{u}\|_X + \|p\|_M \geq \|u\|_X + (1 + \gamma) \|p\|_M,$$

et donc

$$\|(v, q)\|_{X \times M} \leq (1 + \gamma) \|(u, p)\|_{X \times M}.$$

On a donc finalement obtenu

$$\frac{c((u, p), (v, q))}{\|(v, q)\|_{X \times M}} \geq \frac{\delta}{1 + \gamma} \|(u, p)\|_{X \times M},$$

ce qui démontre bien que l'hypothèse (I.13) est vérifiée.

- Montrons maintenant les estimations de stabilité (I.18)-(I.19). Tout d'abord, on fixe G et on prend momentanément $L = 0$ et $a = (\cdot, \cdot)_X$. Comme le problème de point-selle proposé est bien posé pour toutes données, il existe un unique couple $(u_G, p_G) \in X \times M$ tel que

$$(u_G, v)_X + b(v, p_G) = 0, \quad \forall v \in X, \text{ et } b(u_G, q) = G(q), \quad \forall q \in M.$$

Un tel u_G vérifie donc (en prenant $v = u_G$)

$$\|u_G\|_X^2 = -G(p_G) \leq \|G\|_{M'} \|p_G\|_M.$$

De plus, d'après l'inégalité inf-sup, nous avons

$$\beta \|p_G\|_M \leq \sup_{v \in X} \frac{b(v, p_G)}{\|v\|_X} = \sup_{v \in X} \frac{(u_G, v)_X}{\|v\|_X} = \|u_G\|_X.$$

En rassemblant les deux inégalités, on a donc obtenu

$$\beta \|u_G\|_X \leq \|G\|_{M'}. \quad (\text{I.21})$$

Notons que l'application $G \in M' \mapsto u_G$ est linéaire et constitue un inverse à droite continu de l'opérateur B défini plus haut.

Reprenons maintenant (u, p) solution du problème de point-selle qui nous intéresse. On prend $v = u - u_G$ comme fonction test dans la première équation :

$$a(u, u - u_G) + b(u - u_G, p) = L(u - u_G),$$

On remarque le terme en b est nul car on a $b(u, p) = G(p)$ et $b(u_G, p) = G(p)$, il reste donc

$$a(u - u_G, u - u_G) = L(u - u_G) - a(u_G, u - u_G).$$

On utilise maintenant la coercivité de a et les estimations sur u_G pour obtenir

$$\alpha \|u - u_G\|_X \leq \|L\|_{X'} + \frac{\|a\|}{\beta} \|G\|_{M'}.$$

Il vient donc

$$\|u\|_X \leq \|u - u_G\|_X + \|u_G\|_X \leq \frac{\|L\|_{X'}}{\alpha} + \frac{1}{\beta} \left(1 + \frac{\|a\|}{\alpha}\right) \|G\|_{M'},$$

ce qui est le résultat attendu.

Pour l'estimation sur p , on utilise bien entendu à nouveau l'inégalité inf-sup qui donne

$$\beta \|p\|_M \leq \sup_v \frac{b(v, p)}{\|v\|} = \sup_v \frac{a(u, v) - L(v)}{\|v\|} \leq \|a\| \|u\|_X + \|L\|_{X'},$$

et en remplaçant l'estimation obtenue sur u dans cette formule, le résultat attendu suit. ■

On peut proposer une méthode alternative pour prouver directement que le problème est bien posé. Cette démonstration a pour avantage de ne pas utiliser le théorème BNB et de ne s'appuyer seulement que sur le théorème de Lax-Milgram. De plus, la méthode de preuve suggère une méthode numérique (peu efficace en réalité) de résolution de tels systèmes.

Preuve (alternative "constructive"):

Supposons donc a coercive sur X et que la condition (I.17) est satisfaite. En s'inspirant de (I.16), on introduit le problème approché suivant : trouver $(u_\varepsilon, p_\varepsilon) \in X \times M$ tel que

$$a(u_\varepsilon, v) + b(v, p_\varepsilon) - b(u_\varepsilon, q) + \varepsilon(p_\varepsilon, q)_M = L(v) - G(q), \quad \forall v \in X, \forall q \in M.$$

La forme bilinéaire c_ε qui définit ce problème sur $X \times M$ est bien entendu continue et vérifie

$$c_\varepsilon((v, q), (v, q)) = a(v, v) + \varepsilon(p, p)_M,$$

elle est donc bien coercive, d'après le théorème de Lax-Milgram. Il existe donc bien une unique solution $(u_\varepsilon, p_\varepsilon)$ au problème approché.

On peut remarquer qu'en termes d'opérateurs ce problème approché s'écrit de la façon suivante :

$$\begin{cases} Au_\varepsilon + B'p_\varepsilon = L, \\ Bu_\varepsilon + \varepsilon p_\varepsilon = G, \end{cases}$$

Utilisons maintenant l'inégalité (I.17), on a

$$\|p_\varepsilon\|_M \leq \frac{1}{\beta} \sup_{v \in X} \frac{b(v, p_\varepsilon)}{\|v\|_X} = \frac{1}{\beta} \sup_{v \in X} \frac{-a(u_\varepsilon, v) + L(v)}{\|v\|_X} \leq \frac{\|a\|}{\beta} \|u_\varepsilon\|_X + \frac{\|L\|_{X'}}{\beta}.$$

On prend maintenant $v = u_\varepsilon$ et $q = p_\varepsilon$ dans le problème approché, il vient

$$a(u_\varepsilon, u_\varepsilon) + \varepsilon \|p_\varepsilon\|_M^2 = L(u_\varepsilon) - G(p_\varepsilon),$$

et donc par coercivité de a

$$\alpha \|u_\varepsilon\|_X^2 \leq \|L\|_{X'} \|u_\varepsilon\|_X + \|G\|_{M'} \|p_\varepsilon\|_M \leq \left(\|L\|_{X'} + \frac{\|a\| \|G\|_{M'}}{\beta} \right) \|u_\varepsilon\|_X + \frac{\|L\|_{X'}^2}{\beta}.$$

Ainsi, $(u_\varepsilon)_\varepsilon$ et $(p_\varepsilon)_\varepsilon$ sont bornées respectivement dans X et dans M . On peut donc trouver une suite $(\varepsilon_k)_k$ qui tend vers 0 et telle que $u_k = u_{\varepsilon_k}$ et $p_k = p_{\varepsilon_k}$ convergent faiblement vers des limites notées u et p dans X et M respectivement.

On voit alors qu'on peut aisément passer à la limite dans le problème approché et obtenir que u et p sont solutions du problème de départ.

L'unicité découle immédiatement des hypothèses en prenant tout d'abord (u, p) comme fonctions tests dans la formulation faible sans terme source, ce qui montre que $u = 0$, puis l'inégalité (I.17) implique alors que nécessairement $p = 0$. ■

En fait, par des techniques similaires, on peut démontrer que toute la famille $(u_\varepsilon, p_\varepsilon)$ converge vers (u, p) . Plus précisément, on a le résultat suivant :

Proposition I.31

Il existe $C > 0$ qui dépend des données telle que

$$\|u - u_\varepsilon\|_X + \|p_\varepsilon - p\|_M \leq C\varepsilon, \quad \forall \varepsilon > 0.$$

Preuve :

On regarde le problème satisfait par la différence $(u_\varepsilon - u, p_\varepsilon - p) \in X \times M$:

$$a(u_\varepsilon - u, v) + b(v, p_\varepsilon - p) - b(u_\varepsilon - u, q) + \varepsilon (p_\varepsilon, q)_M = 0, \quad \forall v \in X, \forall q \in M.$$

On commence par prendre $q = 0$ et à utiliser la condition inf-sup :

$$\beta \|p_\varepsilon - p\|_M \leq \sup_{v \in X} \frac{b(v, p_\varepsilon - p)}{\|v\|_X} = \sup_{v \in X} \frac{a(u_\varepsilon - u, v)}{\|v\|_X} \leq \|a\| \|u_\varepsilon - u\|_X.$$

On prend ensuite $v = u_\varepsilon - u$ et $q = p_\varepsilon - p$ dans la formulation ci-dessus et on utilise la coercivité de a

$$\alpha \|u_\varepsilon - u\|_X^2 + \varepsilon (p_\varepsilon - p, p_\varepsilon - p)_M = -\varepsilon (p, p_\varepsilon - p)_M,$$

d'où

$$\|u_\varepsilon - u\|_X \leq \varepsilon \frac{\|a\|}{\alpha \beta} \|p\|_M.$$

et

$$\|p_\varepsilon - p\|_M \leq \varepsilon \frac{\|a\|^2}{\alpha \beta^2} \|p\|_M. \quad \blacksquare$$

Définition I.32

On appelle noyau de la forme bilinéaire b , le sous-espace fermé Z de X défini par

$$Z = \{v \in X, \quad b(v, q) = 0, \forall q \in M\}. \quad (\text{I.22})$$

Corollaire I.33

Soit L une forme linéaire continue sur X telle que $L(v) = 0$ pour tout $v \in Z$, alors il existe un unique $p \in M$ tel que

$$L(v) = b(v, p), \quad \forall v \in X,$$

et

$$\|p\|_M \leq \frac{\|L\|_{X'}}{\beta}.$$

Preuve :

D'après le théorème précédent, le problème suivant admet une unique solution

$$(u, v)_X + b(v, p) = L(v), \quad \forall v \in X,$$

$$b(u, q) = 0, \quad \forall q \in M.$$

La seconde équation dit exactement que u appartient à Z , donc si on prend $v = u$ dans la première équation, on obtient en utilisant l'hypothèse $L|_Z = 0$, que $\|u\|_X^2 = 0$ et donc $u = 0$. Il reste bien $b(v, p) = L(v)$ pour tout $v \in X$. De plus, par l'inégalité inf-sup on a

$$\beta \|p\|_M \leq \sup_{v \in X} \frac{b(v, p)}{\|v\|_X} = \sup_{v \in X} \frac{L(v)}{\|v\|_X} = \|L\|_{X'}.$$

En réalité on a un théorème plus fort que le précédent. Il dit que le résultat reste vrai si la forme bilinéaire a est seulement coercive sur le noyau de b . ■

Théorème I.34

Le même énoncé que le théorème I.30 est valable si on suppose seulement que a est coercive sur Z , i.e. sur le noyau de b défini par (I.22).

Preuve :

On a vu dans la preuve du théorème précédent que, G étant fixé, il existe $u_G \in X$ vérifiant

$$\|u_G\|_X \leq \frac{\|G\|_{X'}}{\beta}, \quad \text{et } b(u_G, q) = G(q), \quad \forall q \in M.$$

On va donc chercher (u, p) solution de notre problème de point-selle sous la forme $u = \tilde{u} + u_G$ où \tilde{u} est donc maintenant un élément du noyau Z , par définition de u_G .

On cherche donc $\tilde{u} \in Z$ qui vérifie

$$a(\tilde{u}, \tilde{v}) = L(\tilde{v}) - a(u_G, \tilde{v}), \quad \forall \tilde{v} \in Z.$$

Comme Z est un sous-espace fermé de X , c'est un Hilbert et l'existence et l'unicité d'un tel \tilde{u} est donc la conséquence immédiate du théorème de Lax-Milgram. On pose maintenant

$$\tilde{L}(v) = -a(u_G + \tilde{u}, v) + L(v), \quad \forall v \in X.$$

Par définition de \tilde{u} , la forme \tilde{L} s'annule sur Z , et d'après le corollaire précédent, il existe donc un unique $p \in M$ tel que

$$\tilde{L}(v) = b(v, p), \quad \forall v \in X.$$

Ceci exprime exactement que $(u = \tilde{u} + u_G, p)$ est solution de notre problème. L'estimation sur les normes de u et p s'obtient exactement comme dans le théorème précédent (où on avait utilisé seulement la coercivité de a sur Z !!) ■

4.3 Exemples

4.3.1 Un problème de diffusion scalaire en formulation mixte

On va, dans cette section, reconsidérer sous un angle nouveau le problème $-\operatorname{div}(A(x)\nabla u) = f$. On va, en effet, l'écrire sous la forme

$$\begin{cases} -\operatorname{div} \sigma = f, \\ A(x)\nabla u - \sigma = 0, \end{cases}$$

et on va maintenant chercher un couple (u, σ) solution de ce problème.

Pour essayer de retrouver le cadre général précédent, on va plutôt l'écrire sous la forme équivalente suivante

$$\begin{cases} A(x)^{-1}\sigma - \nabla u = 0, \\ \operatorname{div} \sigma = -f, \end{cases}$$

Plusieurs cadres fonctionnels sont possibles :

1. On cherche $u \in H_0^1(\Omega)$ et $\sigma \in (L^2(\Omega))^d$. En prenant des fonctions tests v et τ dans ces mêmes espaces, la formulation variationnelle peut s'écrire

$$\int_{\Omega} A(x)^{-1}\sigma \cdot \tau \, dx - \int_{\Omega} \nabla u \cdot \tau \, dx + \int_{\Omega} \sigma \cdot \nabla v \, dx = \int_{\Omega} f v \, dx.$$

ATTENTION : dans cette formulation $X = (L^2(\Omega))^d$ et $M = H_0^1(\Omega)$ et c'est donc σ qui joue le rôle de u et u qui joue le rôle de p dans la formulation générale ... On introduit donc

$$a(\sigma, \tau) = \int_{\Omega} A(x)^{-1}\sigma \cdot \tau \, dx,$$

$$b(\sigma, u) = - \int_{\Omega} \sigma \cdot \nabla u \, dx.$$

Comme A est uniformément bornée et uniformément coercive, a est bien une forme bilinéaire continue et coercive sur X et b est bien une forme bilinéaire continue sur $X \times M$. Pour prouver que le problème ainsi formulé est bien posé, il suffit de montrer l'inégalité Inf-Sup (I.17). Dans le cadre proposé, celle-ci s'écrit :

$$\sup_{\sigma \in (L^2(\Omega))^d} \frac{\int_{\Omega} \sigma \cdot \nabla u \, dx}{\|\sigma\|_{L^2}} \geq \beta \|u\|_{H_0^1(\Omega)}.$$

Or ceci est clairement vérifié avec $\beta = 1$ en prenant $\sigma = \nabla u$ dans le quotient proposé. En effet, ce quotient prend alors pour valeur

$$\frac{\|\nabla u\|_{L^2}^2}{\|u\|_{H_0^1}^2} = \|u\|_{H_0^1}.$$

Il existe donc une unique solution au problème proposé. Il va de soi que dans ces conditions, on retrouve exactement la même solution que celle obtenue en utilisant le théorème de Lax-Milgram pour la formulation primitive du problème.

2. Deuxième cadre fonctionnel possible : On cherche $u \in L^2(\Omega)$ et $\sigma \in H_{\operatorname{div}}(\Omega)$. Les conditions aux limites sur u sont prises en compte dans les intégrations par parties. On pose alors

$$a(\sigma, \tau) = \int_{\Omega} A(x)^{-1}\sigma \cdot \tau \, dx,$$

$$b(\sigma, u) = \int_{\Omega} u \operatorname{div} \sigma \, dx.$$

Cette fois, on voit que la forme bilinéaire a n'est plus coercive sur X tout entier car elle ne permet pas de contrôler la norme L^2 de $\operatorname{div} \sigma$. En revanche, elle est bien coercive sur le noyau Z de la forme bilinéaire b . En effet ce noyau Z n'est autre que $H_{0,\operatorname{div}}(\Omega) = \{\sigma \in H_{\operatorname{div}}(\Omega), \operatorname{div} \sigma = 0\}$ et sur Z la norme L^2 et la norme H_{div} coïncident.

Il faut cette fois vérifier l'inégalité Inf-Sup (I.17) suivante :

$$\sup_{\sigma \in H_{\text{div}}(\Omega)} \frac{\int_{\Omega} u \operatorname{div} \sigma \, dx}{\|\sigma\|_{H_{\text{div}}}} \geq \beta \|u\|_{L^2(\Omega)}.$$

Pour la démontrer, on considère la solution $\varphi \in H_0^1(\Omega)$ du problème $-\Delta\varphi = u$ et on pose $\sigma = -\nabla\varphi$ qui est bien un élément de $H_{\text{div}}(\Omega)$ et on a

$$\|\sigma\|_{H_{\text{div}}}^2 = \|\nabla\varphi\|_{L^2}^2 + \|u\|_{L^2}^2.$$

D'après l'inégalité de Poincaré, on a

$$\int_{\Omega} |\nabla\varphi|^2 \, dx = \int_{\Omega} u\varphi \, dx \leq \|u\|_{L^2} \|\varphi\|_{L^2} \leq C \|u\|_{L^2} \|\nabla\varphi\|_{L^2},$$

et il reste donc $\|\sigma\|_{H_{\text{div}}} \leq \sqrt{1+C^2} \|u\|_{L^2}$ et il reste

$$\frac{\int_{\Omega} u \operatorname{div} \sigma \, dx}{\|\sigma\|_{H_{\text{div}}}} \geq \frac{1}{\sqrt{1+C^2}} \|u\|_{L^2},$$

et l'inégalité inf-sup est donc démontrée.

4.3.2 Le problème de Stokes

Le problème de Stokes est le cas typique de problème de type point-selle qui apparaît naturellement. Il s'agit de trouver un champ de vecteurs u et un champ scalaire p vérifiant

$$\begin{cases} -\Delta u + \nabla p = f, \\ \operatorname{div} u = 0, \end{cases} \quad (\text{I.23})$$

avec les conditions aux limites $u = 0$ et la condition de moyenne nulle $\int_{\Omega} p \, dx = 0$. On supposera dans la suite que Ω est connexe.

Le cadre fonctionnel choisi est $X = (H_0^1(\Omega))^d$ et $M = L_0^2$ (l'ensemble des fonctions L^2 à moyenne nulle). On pose alors

$$a(u, v) = \int_{\Omega} \nabla u : \nabla v \, dx, \quad b(v, p) = - \int_{\Omega} p \operatorname{div} v \, dx.$$

Les formes bilinéaires a et b sont continues et a est coercive sur X . Il reste donc à démontrer l'inégalité Inf-Sup (I.17) qui s'écrit ici :

$$\sup_{v \in X} \frac{\int_{\Omega} p \operatorname{div} v \, dx}{\|v\|_{H^1}} \geq \beta \|p\|_{L^2}, \quad \forall p \in L_0^2(\Omega).$$

Cette inégalité inf-sup provient du résultat (assez difficile dans le cas général) suivant

Lemme I.35

Pour toute fonction $p \in L_0^2(\Omega)$, il existe une fonction $v \in (H_0^1(\Omega))^d$ telle que $\operatorname{div} v = p$. Par ailleurs, on peut choisir v tel que

$$\|v\|_{H_0^1} \leq C \|p\|_{L^2},$$

où $C > 0$ ne dépend que de Ω .

Le problème de Stokes est donc bien posé et on verra plus loin comment en faire une discrétisation raisonnable.

Chapitre II

APPROXIMATIONS DE GALERKIN. MÉTHODE DES ÉLÉMENTS FINIS

Notation : Dans toute la suite, on notera $|v|_{H^k}$ la semi-norme H^k , c'est-à-dire

$$|v|_{H^k} = \left(\sum_{|\alpha|=k} \|\partial^\alpha v\|_{L^2}^2 \right)^{\frac{1}{2}}.$$

1 Principes généraux de l'approximation de Galerkin

Etant donné un espace de Hilbert V , une forme bilinéaire continue a et une forme linéaire L continue. On suppose qu'il existe une unique solution $u \in V$ du problème

$$a(u, v) = L(v), \quad \forall v \in V. \quad (\text{II.1})$$

On cherche une approximation de cette solution u en introduisant un sous-espace de dimension finie $V_h \subset V$ et en cherchant une solution $u_h \in V_h$ du problème approché

$$a(u_h, v_h) = L(v_h), \quad \forall v_h \in V_h. \quad (\text{II.2})$$

Quelques remarques s'imposent :

1. L'indice h dans la notation V_h fait référence à "une taille de maillage" ou plus précisément à la qualité de l'approximation de V par V_h .
2. En général, il n'y a aucune raison pour que le problème approché admette une unique solution. Si c'est le cas, le problème obtenu est un système linéaire carré de dimension finie qu'il "suffira" de résoudre pour obtenir la solution approchée u_h . Nous reviendrons sur ces aspects plus tard.

3. D'autres variantes existent : on peut également approcher la forme bilinéaire a par une forme a_h , on peut également considérer des espaces d'approximation V_h qui ne sont pas inclus dans V . On parle alors d'une approximation non conforme.
4. On peut également considérer deux espaces de dimensions finies V_h et W_h (de même dimensions) et chercher u_h dans V_h et les fonctions tests v_h dans W_h . On parle alors d'approximation de Petrov-Galerkin.

Supposons maintenant qu'il existe une solution u_h du problème approché (II.2), de sorte qu'en comparant (II.1) et (II.2), on obtient

$$\forall v_h \in V_h \subset V, a(u, v_h) = L(v_h) = a(u_h, v_h),$$

ou encore

$$a(u - u_h, v_h) = 0.$$

L'erreur $e_h = u - u_h$ est donc a -orthogonale à l'espace V_h . C'est la formule de base qui va nous permettre d'analyser la convergence de la méthode.

1.1 Cadre coercif. Approximation de Galerkin.

1.1.1 Existence et convergence

Le premier cadre que nous allons considérer est celui où la forme bilinéaire a est coercive sur V . Dans ces conditions, la restriction de a à $V_h \times V_h$ est également coercive et donc le problème approché admet une unique solution u_h qui vérifie

$$\|u_h\|_V \leq \frac{\|a\|}{\alpha} \|L\|_{V'}. \quad (\text{II.3})$$

Proposition II.1

On suppose que pour tout $v \in V$, $d(v, V_h) \rightarrow 0$ quand $h \rightarrow 0$. Alors $(u_h)_h$ converge vers u dans V quand h tend vers 0.

Preuve :

De la borne (II.3), on déduit que la famille $(u_h)_h$ d'éléments de V est bornée. On peut donc trouver une sous-suite $(h_n)_n$ qui tend vers 0 et un élément $\tilde{u} \in V$ tels que $(u_{h_n})_n$ converge faiblement vers \tilde{u} . On va montrer que \tilde{u} est égal à u .

Pour cela, on prend $v \in V$ quelconque. Par hypothèse, il existe une suite $(v_h)_h$ d'éléments de V telle que $v_h \in V_h$ pour tout $h > 0$ et telle que $v_h \rightarrow v$ dans V , quand $h \rightarrow 0$.

Prenons v_h comme fonction test dans le problème discret

$$a(u_h, v_h) = L(v_h).$$

Par convergence fort-faible, on peut passer à la limite dans le premier terme, et bien entendu dans le second. On obtient donc

$$a(\tilde{u}, v) = L(v).$$

Ceci étant vrai pour tout $v \in V$, et par unicité de la solution du problème initial, on obtient bien que $\tilde{u} = u$.

Ceci montre en particulier que la famille $(u_h)_h$ admet une unique valeur d'adhérence faible dans V (qui est u !) et donc, en particulier, elle converge faiblement vers u .

Montrons maintenant la convergence forte de u_h vers u . Pour cela, on prend $v_h = u_h$ dans le problème discret et on constate que

$$a(u_h, u_h) = L(u_h) \xrightarrow{h \rightarrow 0} L(u) = a(u, u).$$

Il vient alors

$$a(u_h - u, u_h - u) = a(u_h, u_h) - a(u_h, u) - a(u, u_h) + a(u, u) \xrightarrow{h \rightarrow 0} 0.$$

Par coercivité de a , cela implique bien que $\|u_h - u\|_V \rightarrow 0$. ■

1.1.2 Estimation de l'erreur

On peut en fait estimer l'erreur d'approximation de la façon suivante

Proposition II.2 (Lemme de Céa)

L'erreur e_h vérifie

$$\|e_h\|_V \leq \frac{\|a\|}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V = \frac{\|a\|}{\alpha} d(u, V_h).$$

Preuve :

On prend $v_h - u_h$ dans l'équation d'orthogonalité de l'erreur et on trouve

$$0 = a(e_h, v_h - u_h) = a(e_h, v_h - u) + a(e_h, e_h).$$

On obtient

$$\alpha \|e_h\|_V^2 \leq \|a\| \|e_h\|_V \|v_h - u\|_V,$$

ceci étant vrai pour tout $v_h \in V_h$, on obtient bien l'estimation demandée. ■

Ce lemme montre donc que l'erreur d'approximation e_h se mesure (en norme V) en fonction de l'erreur d'approximation entre l'espace V_h et V , c'est-à-dire la distance entre ces deux espaces. En fait, c'est même mieux que cela, car pour obtenir la convergence (et l'estimation d'erreur) il suffit de savoir approcher la solution u par des fonctions de V_h .

Le résultat précédent peut être amélioré dans le cas symétrique.

Proposition II.3 (Lemme de Céa - cas symétrique)

Si on suppose que a est symétrique, alors on a l'estimation

$$\|e_h\|_V \leq \sqrt{\frac{\|a\|}{\alpha}} \inf_{v_h \in V_h} \|u - v_h\|_V.$$

Preuve :

La relation d'orthogonalité de l'erreur $a(u_h - u, v_h) = 0$ implique que u_h minimise dans V_h la fonctionnelle

$$E(v_h) = \frac{1}{2} a(v_h - u, v_h - u), \quad \forall v_h \in V_h.$$

En particulier on a

$$\alpha \|u_h - u\|_V^2 \leq a(u_h - u, u_h - u) \leq a(v_h - u, v_h - u) \leq \|a\| \|v_h - u\|_V^2, \quad \forall v_h \in V_h,$$

ce qui donne le résultat. ■

Dans toute la suite, on va donc chercher à construire des espaces V_h convenables et à estimer la distance entre la solution u et V_h .

1.2 Cadre coercif. Approximation de Petrov-Galerkin.

On considère une méthode un peu plus générale, où on considère deux sous-espaces de V de dimensions finies V_h et W_h et de mêmes dimensions. On s'intéresse alors au problème suivant : trouver $u_h \in V_h$ tel que

$$a(u_h, w_h) = L(w_h), \quad \forall w_h \in W_h. \quad (\text{II.4})$$

1.2.1 Existence and convergence

La coercivité de a sur $V \times V$ ne suffit pas à démontrer l'existence et l'unicité d'une solution à ce problème. Pour obtenir cette propriété, il faut vérifier les deux conditions du théorème BNB. En réalité, dans le cas de la dimension finie, les deux conditions sont équivalentes. Donc le problème discret ci-dessus est bien posé, si et seulement si il existe $\alpha_h > 0$ tel que

$$\inf_{u_h \in V_h} \sup_{w_h \in W_h} \frac{a(u_h, w_h)}{\|u_h\|_V \|w_h\|_V} \geq \alpha_h.$$

Dans ces conditions la solution u_h de (II.4) existe de façon unique et vérifie

$$\|u_h\|_V \leq \frac{\|L\|_{V'}}{\alpha_h}.$$

On voit donc déjà que la solution approchée ne sera bornée *a priori* que si la constante α_h de la condition inf-sup ne tend pas vers 0 quand h tend vers 0.

Proposition II.4

On suppose qu'il existe $\alpha^ > 0$ tel que $\alpha_h \geq \alpha^*$ pour tout $h > 0$ et que pour tout $v \in V$, on a $d(v, W_h) \rightarrow 0$ quand $h \rightarrow 0$. Alors u_h converge vers u dans V quand $h \rightarrow 0$.*

La démonstration est similaire au cas précédent, avec toutefois une petite subtilité due au fait que V_h et W_h sont deux espaces différents.

1.2.2 Estimation de l'erreur

Sous les hypothèses précédentes, l'erreur $e_h = u - u_h$, vérifie la condition d'orthogonalité

$$a(e_h, w_h) = 0, \quad \forall w_h \in W_h.$$

Proposition II.5 (Lemme de Strang - cas simplifié)

L'erreur e_h vérifie l'estimation

$$\|e_h\|_V \leq \left(1 + \frac{\|a\|}{\alpha_h}\right) \inf_{v_h \in V_h} \|u - v_h\|_V.$$

Preuve :

Soit $v_h \in V_h$. On a

$$e_h = u - u_h = (u - v_h) + (v_h - u_h).$$

Or d'après la condition inf-sup

$$\|v_h - u_h\|_V \leq \frac{1}{\alpha_h} \sup_{w_h \in W_h} \frac{a(v_h - u_h, w_h)}{\|w_h\|_V} = \frac{1}{\alpha_h} \sup_{w_h \in W_h} \frac{a(u - v_h, w_h)}{\|w_h\|_V} \leq \frac{\|a\|}{\alpha_h} \|u - v_h\|_V.$$

Il vient donc

$$\|e_h\|_V \leq \left(1 + \frac{\|a\|}{\alpha_h}\right) \|u - v_h\|_V.$$

Ceci étant vrai pour tout $v_h \in V_h$, on obtient le résultat annoncé. ■

1.2.3 Bilan et exemple

Dans le cas de l'approximation de Petrov-Galerkin, il faut donc :

- Contrôler l'erreur d'approximation associée à l'espace V_h .
- Contrôler la constante α_h de l'estimation inf-sup entre les espaces V_h et W_h . Il ne faut pas que ces espaces soient trop orthogonaux au sens de a quand h tend vers 0.

Prenons un exemple en dimension 1 d'espace : $V = H_0^1(]0, 1[)$, $a(u, v) = \int_0^1 u'v' dx$. On prend pour V_h l'espace de dimension 1 engendré par la fonction $p(x) = 1 - |2x - 1|$ dont le gradient vaut 2 sur $[0, 1/2]$ et -2 sur

$[1/2, 1]$. On prend ensuite pour W_h l'espace de dimension 1 engendré par $q(x) = \sin(4\pi x)$ dont le gradient vaut $4\pi \cos(4\pi x)$. On va montrer que la condition inf-sup n'est pas satisfaite pour ce problème car en effet on a

$$a(p, q) = \int_0^1 p'q' = 8\pi \left(\int_0^{\frac{1}{2}} \sin(4\pi x) dx - \int_{\frac{1}{2}}^1 \sin(4\pi x) dx \right) = 0.$$

N.B. : même si a est symétrique et que le schéma de Petrov-Galerkin est bien posé, il ne conduit pas nécessairement à un système discret symétrique à la fin.

1.3 Cadre d'un problème mixte

Soient X et M deux espaces de Hilbert indépendants. On s'intéresse au problème de type point-selle suivant :

$$\begin{cases} a(u, v) + b(v, p) = L(v), \forall v \in X, \\ b(u, q) = G(q), \forall q \in M. \end{cases}$$

où L et G sont des formes linéaires continues respectivement sur X et M . On a vu qu'un tel problème est bien posé dès lors que a est coercive sur X (ou sur le noyau de b) et que b vérifie la condition inf-sup adéquate

$$\inf_{p \in M} \sup_{v \in X} \frac{b(v, p)}{\|v\|_X \|p\|_M} \geq \beta > 0.$$

Soient maintenant X_h et M_h deux sous-espaces de dimensions finies respectivement de X et M et on considère le problème approché suivant : trouver $u_h \in X_h$ et $p_h \in M_h$ tels que

$$\begin{cases} a(u_h, v_h) + b(v_h, p_h) = L(v_h), \forall v_h \in X_h, \\ b(u_h, q_h) = G(q_h), \forall q_h \in M_h. \end{cases} \quad (\text{II.5})$$

Notons qu'on aurait pu envisager une méthode de type Petrov-Galerkin en choisissant v_h et q_h dans d'autres espaces que X_h et M_h mais cela compliquerait bien entendu l'analyse qui suit.

1.3.1 Existence et convergence

Que peut-on dire de ce problème approché ?

- Si a est coercive sur X tout entier, alors a est aussi coercive sur X_h avec la même constante de coercivité. Pour assurer que le problème (II.5) est bien posé il faut donc vérifier une condition inf-sup sur b qui s'écrit

$$\inf_{p_h \in M_h} \sup_{v_h \in X_h} \frac{b(v_h, p_h)}{\|p_h\|_M \|v_h\|_X} \geq \beta_h > 0.$$

Comme nous sommes en dimension finie, on voit par un argument de compacité que cette condition est vérifiée si et seulement si on a

$$p_h \in M_h \text{ et } b(v_h, p_h) = 0, \forall v_h \in X_h \implies p_h = 0.$$

Si on regarde la restriction B'_h de l'opérateur $B' : M \mapsto X'$ comme un opérateur de M_h dans X'_h , alors cette condition dit que cet opérateur doit être injectif. En particulier, la dimension de M_h doit être inférieure ou égale à celle de X_h . **Moralité :** la condition inf-sup sera en défaut si M_h est trop gros par rapport à X_h .

Quoi qu'il en soit, si cette condition inf-sup est vérifiée, il existe une unique solution $(u_h, p_h) \in X_h \times M_h$ du problème approché et on a les bornes suivantes

$$\begin{aligned} \|u_h\|_X &\leq \frac{\|L\|_{X'}}{\alpha} + \frac{1}{\beta_h} \left(1 + \frac{\|a\|}{\alpha} \right) \|G\|_{M'}, \\ \|p_h\|_M &\leq \left(1 + \frac{\|a\|}{\alpha} \right) \left(\frac{1}{\beta_h} \|L\|_{X'} + \frac{\|a\|}{\beta_h^2} \|G\|_{M'} \right). \end{aligned}$$

- Si a est seulement coercive sur le noyau de b , alors rien n'assure que sa restriction au noyau Z_h de la restriction $b : X_h \times M_h \mapsto \mathbb{R}$. En effet, on n'a pas en général l'inclusion $Z_h \subset Z$, car un élément $v_h \in Z_h$ est tel que $b(v_h, p_h) = 0$ pour tout $p_h \in M_h$ mais rien ne dit que $b(v_h, p) = 0$ pour $p \in M \setminus M_h$.

Dans ces conditions, il faut donc rajouter la coercivité de a restreinte à Z_h comme condition supplémentaire de résolubilité du système.

Théorème II.6

S'il existe $\beta^ > 0$ telle que $\beta_h \geq \beta^*$ pour tout $h > 0$ et si $d(v, X_h) \rightarrow 0$ et $d(q, M_h) \rightarrow 0$ pour tout $v \in X$ et $q \in M$, alors $(u_h, p_h) \rightarrow (u, p)$ dans $X \times M$, quand $h \rightarrow 0$.*

1.3.2 Estimation de l'erreur

Essayons maintenant de produire une estimation de l'erreur dans ce contexte. Pour cela, on introduit $e_h = u - u_h$ et $\pi_h = p - p_h$ et on observe que l'on a

$$\begin{cases} a(e_h, v_h) + b(v_h, \pi_h) = 0, \forall v_h \in X_h, \\ b(e_h, q_h) = 0, \forall q_h \in M_h. \end{cases}$$

1. On récrit la première équation comme suit :

$$a(e_h, v_h) + b(v_h, p - q_h) = b(v_h, p_h - q_h), \forall v_h \in X_h.$$

Ainsi, grâce à la condition inf-sup discrète, on trouve

$$\|q_h - p_h\|_M \leq \frac{1}{\beta_h} \sup_{v_h \in X_h} \frac{a(e_h, v_h) + b(v_h, p - q_h)}{\|v_h\|_X} \leq \frac{1}{\beta_h} (\|a\| \|e_h\|_X + \|b\| \|p - q_h\|_M).$$

L'inégalité triangulaire donne, pour tout $q_h \in M_h$:

$$\|\pi_h\|_M \leq \|p - q_h\|_M + \|q_h - p_h\|_M,$$

et donc

$$\|\pi_h\|_M \leq \left(1 + \frac{\|b\|}{\beta_h}\right) d(p, M_h) + \frac{\|a\|}{\beta_h} \|e_h\|_X.$$

2. Soit $v_h \in X_h$ quelconque. D'après la condition inf-sup discrète, il existe $r_h \in X_h$ tel que

$$b(r_h, q_h) = b(u - v_h, q_h), \quad \forall q_h \in M_h, \quad \text{et} \quad \|r_h\|_X \leq \frac{\|b\|}{\beta_h} \|u - v_h\|_X.$$

On constate donc que pour tout $q_h \in M_h$ on a

$$b(u - v_h - r_h, q_h) = 0,$$

or $b(e_h, q_h) = 0$ donc on a

$$b(u_h - (v_h + r_h), q_h) = 0.$$

On va donc prendre $(v_h + r_h) - u_h$ comme fonction test dans la première équation. Il vient

$$a(e_h, (v_h + r_h) - u_h) + b((v_h + r_h) - u_h, \pi_h) = 0.$$

Comme $u_h - (v_h + r_h)$ est orthogonal à M_h , on peut remplacer π_h dans le second terme par $p - q_h$ pour tout q_h . On obtient

$$\alpha \|(v_h + r_h) - u_h\|_X^2 \leq \|a\| \|u - (v_h + r_h)\|_X \|(v_h + r_h) - u_h\|_X + \|b\| \|v_h + r_h - u_h\|_X \|p - q_h\|_X,$$

D'où

$$\|(v_h + r_h) - u_h\|_X \leq \frac{\|a\|}{\alpha} \|u - (v_h + r_h)\|_X + \frac{\|b\|}{\alpha} d(p, X_h),$$

puis

$$\|u - u_h\|_X \leq \|u - (v_h + r_h)\| + \|(v_h + r_h) - u_h\|_X \leq \left(1 + \frac{\|a\|}{\alpha}\right) \|u - (v_h + r_h)\|_X + \frac{\|b\|}{\alpha} d(p, M_h).$$

Par ailleurs, on a

$$\|u - (v_h + r_h)\|_X \leq \|u - v_h\|_X + \|r_h\|_X \leq \left(1 + \frac{\|b\|}{\beta_h}\right) \|u - v_h\|_X.$$

Ceci étant vrai pour tout $v_h \in X_h$, on trouve

$$\|u - u_h\|_X \leq \left(1 + \frac{\|a\|}{\alpha}\right) \left(1 + \frac{\|b\|}{\beta_h}\right) d(u, X_h) + \frac{\|b\|}{\alpha} d(p, M_h).$$

L'estimation sur la pression s'obtient alors en utilisant l'estimation ci-dessus.

En résumé, on doit cette fois estimer l'erreur d'approximation pour chacun des espaces X_h et M_h et également estimer la dépendance de la constante de l'inégalité inf-sup en fonction du paramètre de discrétisation.

On va attaquer ces deux types de questions séparément. Etudions tout d'abord la façon dont les espaces V_h et X_h/M_h peuvent être construits.

1.4 Choix des espaces d'approximation et de la base

Quelques principes guident le choix des espaces d'approximation V_h et d'une base $(\varphi_i)_i$ d'un tel espace. Remarquons que trouver la solution $u_h = \sum_j u_j \varphi_j$ du problème variationnel revient à résoudre un système linéaire dont la matrice A a pour coefficients : $a_{i,j} = a(\varphi_j, \varphi_i)$. On doit donc pouvoir :

- Calculer ces coefficients.
- Obtenir une matrice A avec une structure la plus simple possible, c'est-à-dire qu'on puisse inverser aisément.

Pour tout cela, le choix de l'espace et de la base sur laquelle on va travailler est cruciale.

Exemples :

- Si on prend pour V_h l'ensemble des fonctions polynômes de degré inférieur ou égal à N , avec N grand et la base formée des monômes, alors il y a de fortes chances que la matrice A soit pleine (penser à $a(u, v) = \int_{\Omega} u'v' dx$). Bien entendu si on prend une base de polynômes orthonormés pour le produit scalaire défini par a (si cette forme a est symétrique par exemple), alors la matrice A est diagonale et le calcul de la solution approchée est trivial.
- Le meilleur espace d'approximation est l'espace de dimension 1 engendré par la solution exacte. Bien entendu, on ne connaît pas la solution exacte, donc cette construction n'est pas utilisable en pratique.
- Une construction qui peut s'avérer intelligente et de prendre pour espace d'approximation, l'espace engendré par les fonctions propres de l'opérateur mis en jeu dans l'équation. Si la solution est régulière on peut démontrer que, sous de bonnes hypothèses, la méthode est extrêmement efficace. On parle de **méthode spectrale**.

Ainsi, si on note V_N l'espace engendré par les N premières fonctions propres du Laplacien-Dirichlet en dimension 1 d'espace (i.e. les $\sin(k\pi x)$ sur $]0, 1[$), alors on a le résultat suivant (identité de Parseval) :

$$\forall u \in H^m(\Omega), \quad d(u, V_N) \leq \frac{C_k}{N^m} \|u\|_{H^m}.$$

L'inconvénient de cette méthode est qu'on ne sait calculer les fonctions propres de l'opérateur que dans des cas très particuliers (il y a même de nombreux cas où cette notion n'est pas bien définie). Dans les autres cas, on peut utiliser les fonctions propres d'un autre opérateur, en espérant que la méthode va bien fonctionner. On se retrouve alors dans une situation où la matrice du système est pleine et donc difficile à inverser.

- La dernière méthode que je présente et qui va faire l'objet de la suite de ce chapitre, c'est la méthode dite des **éléments finis**. L'idée est de découper le domaine en petits morceaux (maillage) et de considérer comme espace d'approximation un ensemble de fonctions régulières par morceaux et ayant une structure particulière sur chaque morceau. Typiquement, on considérera des fonctions polynomiales par morceaux.

2 Exemples d'espaces éléments finis

2.1 Le cas 1D

2.1.1 L'espace d'approximation \mathbb{P}_1

On se place en dimension 1 d'espace, sur l'intervalle $\Omega =]0, 1[$ et on cherche un espace d'approximation raisonnable pour $V = H^1(\Omega)$. Pour cela, on commence à **mailler** le domaine Ω . Dans le cas monodimensionnel, cela revient à découper l'intervalle Ω en segments $[x_i, x_{i+1}]$ avec $i \in \{0, \dots, N\}$ de telle sorte que $x_i < x_{i+1}$ et $x_0 = 0, x_{N+1} = 1$. On pose alors $h = \sup_i |x_{i+1} - x_i|$, que l'on appelle le pas du maillage. On notera $K_i = [x_i, x_{i+1}]$ les cellules de ce maillage.

Ceci étant fait on se propose de considérer l'ensemble des fonctions affines par morceaux

$$V_h = \{u \in V, \forall i, u|_{K_i} \in \mathbb{P}_1\},$$

où \mathbb{P}_1 désigne l'ensemble des polynômes de degré inférieur ou égal à 1. Notons que l'on considère bien ici un espace d'approximation conforme puisque, par construction, on aura $V_h \subset V$.

Lemme II.7

L'espace V_h peut aussi se définir par

$$V_h = \{u \in C^0(\bar{\Omega}), \forall i, u|_{K_i} \in \mathbb{P}_1\},$$

et de plus, l'application

$$\Phi : u \in V_h \mapsto (u(x_0), \dots, u(x_{N+1})) \in \mathbb{R}^{N+2},$$

est un isomorphisme. En particulier $\dim V_h = N + 2$.

Preuve :

On doit montrer deux inclusions. La première \subset est triviale car on sait qu'en dimension 1 les fonctions de H^1 sont continues. Pour montrer la seconde inclusion, il s'agit juste d'une intégration par parties ...

Le fait que l'application proposée est linéaire et bijective est assez clair et consiste juste à remarquer qu'une fonction affine par morceaux est définie de manière unique par les valeurs qu'elle prend aux noeuds. ■

Le fait qu'on recherche une approximation conforme de l'espace V n'est pas sans conséquence : supposons que l'on se soit intéressés à l'espace $V = H^2(\Omega)$, alors on peut montrer que l'espace d'approximation \mathbb{P}_1 noté V_h construit de façon analogue à ce qui précède est de dimension 2 et seulement constitué des fonctions affines sur tout le domaine Ω . On comprend bien que cet espace V_h n'a guère d'intérêt en vue de l'approximation numérique d'un problème variationnel posé dans $H^2(\Omega)$.

Comme l'application Φ est bijective, on voit que tout élément $u \in V_h$ peut s'identifier à $\Phi(u) \in \mathbb{R}^{n+2}$. Les éléments de $\Phi(u)$ sont appelés les **degrés de liberté** dans l'espace d'approximation V_h .

On dit qu'on a affaire à un élément fini de Lagrange quand les degrés de liberté sont tous de la forme $u \mapsto u(a)$ pour un certain point $a \in \bar{\Omega}$. Les points a_i ainsi définis sont tous distincts et sont appelés les noeuds associés à l'espace d'approximation. Dans le cas présent, les noeuds sont exactement les points x_i qui définissent le maillage.

Définition II.8

On note e_i pour $i = 0, \dots, n + 1$ les vecteurs de la base canonique de \mathbb{R}^{n+2} . On appelle fonction de forme associée au noeud x_i , et on note φ_i , l'antécédent de e_i par l'application Φ . Autrement dit, les $\varphi_i \in V_h$ sont définis par

$$\varphi_i(x_j) \delta_{ij}, \quad \forall i, j \in \{0, \dots, n + 1\}.$$

Dans le cadre des éléments finis \mathbb{P}_1 ces fonctions de forme (ou fonctions de base) sont aussi appelées **fonctions chapeau** pour des raisons relativement évidentes.

On va chercher maintenant à estimer l'erreur d'approximation associée à ces espaces V_h . Pour ce faire, on va introduire la notion d'opérateur d'interpolation.

Définition II.9

On appelle opérateur d'interpolation de V dans V_h , l'opérateur \mathcal{I}_h^1 défini par

$$\forall u \in V, \quad \mathcal{I}_h^1 u = \Phi^{-1}(u(x_0), \dots, u(x_{n+1})),$$

ou de façon plus claire

$$\mathcal{I}_h^1 u(x) = \sum_{i=0}^{n+1} u(x_i) \varphi_i(x).$$

Remarquons que si $u \in H_0^1(\Omega)$, alors $\mathcal{I}_h^1 u \in H_0^1(\Omega)$ et donc $V_{h,0} = V_h \cap H_0^1(\Omega)$ est un espace d'approximation conforme de $H_0^1(\Omega)$.

On remarque aussi que \mathcal{I}_h^1 est un projecteur (**non orthogonal !**) : $\mathcal{I}_h^1 \circ \mathcal{I}_h^1 = \mathcal{I}_h^1$. Enfin, comme $H^1(\Omega)$ s'injecte de façon continue dans $C^0(\bar{\Omega})$, il est clair que \mathcal{I}_h^1 est un opérateur continu de $H^1(\Omega)$ dans lui-même. A ce stade il n'est pas clair que sa norme est indépendante de h (i.e. de n).

Nous allons donc maintenant montrer des estimations sur \mathcal{I}_h^1 qui feront intervenir des constantes indépendantes de h .

Théorème II.10 (Propriétés d'interpolation de \mathcal{I}_h^1)

1. Il existe $C > 0$ telle que

$$\forall u \in V, \quad \|\mathcal{I}_h^1 u - u\|_{L^2} \leq Ch|u|_{H^1}, \quad (\text{II.6})$$

$$\forall u \in V, \quad |\mathcal{I}_h^1 u - u|_{H^1} \leq C|u|_{H^1}. \quad (\text{II.7})$$

2. Si $u \in V \cap H^2(\Omega)$, alors on a

$$|\mathcal{I}_h^1 u - u|_{H^1} \leq Ch|u|_{H^2}, \quad (\text{II.8})$$

$$\|\mathcal{I}_h^1 u - u\|_{L^2} \leq Ch^2|u|_{H^2}, \quad (\text{II.9})$$

où C ne dépend que de Ω .

3. On a

$$\forall u \in V, \quad \lim_{h \rightarrow 0} \|\mathcal{I}_h^1 u - u\|_{H^1} = 0.$$

En particulier, pour tout $u \in V$, $d(u, V_h) \rightarrow 0$ quand $h \rightarrow 0$.

Remarque : on ne peut pas améliorer l'estimation (II.8) même si u est plus régulière que $H^2(\Omega)$.

Toutes les estimations des points 1 et 2 peuvent s'écrire :

$$\forall m \in \{0, 1\}, \forall u \in H^{m+1}(\Omega), \quad \|\mathcal{I}_h^1 u - u\|_{L^2} + h|\mathcal{I}_h^1 u - u|_{H^1} \leq Ch^{m+1}|u|_{H^{m+1}}.$$

Preuve :

L'idée de base est de regarder ce qui se passe sur chaque maille (ou élément) de la discrétisation. Soit donc $i \in \{0, \dots, n\}$. Prenons $x \in K_i$.

1. Regardons la situation sur un des intervalles $[x_i, x_{i+1}]$ du maillage.

$$\mathcal{I}_h^1 u(x) = \frac{u_i(x_{i+1} - x) + u_{i+1}(x - x_i)}{h_i}, \quad (\text{II.10})$$

où on a noté $u_i = u(x_i)$ et $h_i = x_{i+1} - x_i$, pour simplifier les notations.

On obtient

$$\mathcal{I}_h^1 u(x) - u(x) = \frac{(u_i - u(x))(x_{i+1} - x) + (u_{i+1} - u(x))(x - x_i)}{h_i},$$

et donc

$$\begin{aligned} |\mathcal{I}_h^1 u(x) - u(x)| &\leq |u_i - u(x)| + |u_{i+1} - u(x)| \\ &\leq \left| \int_{x_i}^x u'(t) dt \right| + \left| \int_x^{x_{i+1}} u'(t) dt \right| \leq 2h_i^{\frac{1}{2}} \left(\int_{x_i}^{x_{i+1}} |u'(t)|^2 dt \right)^{\frac{1}{2}}, \end{aligned}$$

et donc

$$\int_{x_i}^{x_{i+1}} |\mathcal{I}_h^1 u - u|^2 dx \leq 2h_i^2 \int_{x_i}^{x_{i+1}} |u'(t)|^2 dt,$$

soit en sommant :

$$\|\mathcal{I}_h^1 u - u\|_{L^2} \leq \sqrt{2}h \|u'\|_{L^2} = \sqrt{2}h |u|_{H^1}, \quad (\text{II.11})$$

et ainsi

$$\|\mathcal{I}_h^1 u\|_{L^2} \leq C \|u\|_{H^1}.$$

On revient à (II.10) et il vient

$$|(\mathcal{I}_h^1 u)'(x)| = \left| \frac{u_{i+1} - u_i}{h_i} \right| \leq \frac{1}{h_i^{\frac{1}{2}}} \left(\int_{x_i}^{x_{i+1}} |u'(t)|^2 dt \right)^{\frac{1}{2}},$$

d'où

$$\int_{x_i}^{x_{i+1}} |(\mathcal{I}_h^1 u)'(x)|^2 dx \leq \int_{x_i}^{x_{i+1}} |u'(t)|^2 dt,$$

et ainsi

$$\|(\mathcal{I}_h^1 u)'\|_{L^2} \leq \|u'\|_{L^2},$$

soit

$$|\mathcal{I}_h^1 u|_{H^1} \leq |u|_{H^1}.$$

En rassemblant les deux résultats précédents, on obtient la première estimation.

Remarque II.11

Noter que cette estimation permet déjà de démontrer que si $u \in V$, alors $\mathcal{I}_h^1 u$ converge faiblement vers u dans $H^1(\Omega)$. On va voir par la suite que la convergence est forte mais il faut retenir l'idée de la preuve. On sait en effet que $\mathcal{I}_h^1 u$ est bornée dans H^1 , on peut donc trouver une suite $h_n \rightarrow 0$ telle que $\mathcal{I}_{h_n}^1 u$ converge faiblement dans H^1 . D'après (II.11), la limite forte dans L^2 de $\mathcal{I}_{h_n}^1 u$ n'est autre que u . Par unicité de la limite au sens des distributions, par exemple, on conclut que la limite faible H^1 de cette suite est aussi égale à u . Par un argument d'unicité de la valeur d'adhérence, on obtient bien la convergence faible annoncée.

2. On a déjà obtenu (II.11), il reste donc à étudier le gradient de $\mathcal{I}_h^1 u - u$. Pour cela on écrit

$$(\mathcal{I}_h^1 u)'(x) - u'(x) = \frac{u_{i+1} - u_i}{h_i} - u'(x).$$

Ecrivons les deux formules de Taylor suivantes :

$$u(x_i) = u(x) + (x_i - x)u'(x) + \underbrace{\frac{1}{2} \int_0^1 (1-t)u''(x+t(x_i-x))(x_i-x)^2 dt}_{\stackrel{\text{def}}{=} R_i(x)},$$

$$u(x_{i+1}) = u(x) + (x_{i+1} - x)u'(x) + \underbrace{\frac{1}{2} \int_0^1 (1-t)u''(x+t(x_{i+1}-x))(x_{i+1}-x)^2 dt}_{\stackrel{\text{def}}{=} R_{i+1}(x)}.$$

On en déduit les estimations suivantes :

- On écrit $\mathcal{I}_h^1 u - u$ au point x :

$$\begin{aligned} & \frac{(u_i - u(x))(x_{i+1} - x) + (u_{i+1} - u(x))(x - x_i)}{h_i} \\ &= \underbrace{\frac{(x_i - x)u'(x)(x_{i+1} - x) + (x_{i+1} - x)u'(x)(x - x_i)}{h_i}}_{=0} + \frac{R_i(x)(x_{i+1} - x) + R_{i+1}(x)(x - x_i)}{h_i}. \end{aligned}$$

D'où (Jensen ...)

$$|\mathcal{I}^1 h u(x) - u(x)|^2 \leq |R_i(x)|^2 + |R_{i+1}(x)|^2,$$

et donc

$$\int_{x_i}^{x_{i+1}} |\mathcal{I}_h^1 u - u|^2 \leq \int_{x_i}^{x_{i+1}} |R_i|^2 + \int_{x_i}^{x_{i+1}} |R_{i+1}|^2. \quad (\text{II.12})$$

Estimons maintenant la norme L^2 de R_i . On a tout d'abord

$$|R_i(x)|^2 \leq \frac{1}{2} \int_0^1 |u''(x + t(x_i - x))|^2 (x_i - x)^4 dt \leq \frac{h_i^3}{2} \int_{x_i}^x |u''|^2 \leq \frac{h_i^3}{2} \int_{x_i}^{x_{i+1}} |u''|^2,$$

ce qui donne

$$\int_{x_i}^{x_{i+1}} |R_i(x)|^2 dx \leq \frac{h_i^4}{2} \int_{x_i}^{x_{i+1}} |u''|^2 dx.$$

En reportant dans (II.12), il vient

$$\int_{x_i}^{x_{i+1}} |\mathcal{I}_h^1 u - u|^2 dx \leq h_i^4 \int_{x_i}^{x_{i+1}} |u''|^2 dx,$$

ce qui donne le résultat attendu en sommant sur i .

- On écrit maintenant la dérivée de $\mathcal{I}^1 h u - u$ au point x :

$$\frac{u_{i+1} - u_i}{x_{i+1} - x_i} - u'(x) = \frac{R_{i+1}(x) - R_i(x)}{h_i},$$

et donc

$$\left| \frac{u_{i+1} - u_i}{x_{i+1} - x_i} - u'(x) \right|^2 \leq \frac{2}{h_i^2} (|R_i(x)|^2 + |R_{i+1}|^2) \leq h_i \int_{x_i}^x |u''(t)|^2 dt + h_i \int_x^{x_{i+1}} |u''(t)|^2 dt.$$

Ainsi

$$\int_{x_i}^{x_{i+1}} \left| \frac{u_{i+1} - u_i}{x_{i+1} - x_i} - u'(x) \right|^2 dx \leq h_i^2 \int_{x_i}^{x_{i+1}} |u''(t)|^2 dt.$$

Si on somme, il vient

$$|\mathcal{I}_h^1 u - u|_{H^1} \leq h |u|_{H^2}.$$

3. Le dernier résultat se déduit des deux premiers par densité de $H^2(\Omega)$ dans $H^1(\Omega)$. ■

Il est instructif, en vue de résultats plus généraux, de faire une preuve par changement de variable. On considère le segment unité $\hat{K} = [0, 1]$ et on note \mathcal{T}_0^1 l'opérateur d'interpolation de Lagrange de degré 1 sur cette intervalle. On suppose qu'il existe $C > 0$ telle que

$$\forall v \in H^1(\hat{K}), \|\mathcal{T}_0^1 v - v\|_{L^2(\hat{K})} \leq C |v|_{H^1(\hat{K})}, \quad (\text{II.13})$$

$$\forall v \in H^1(\hat{K}), |\mathcal{T}_0^1 v - v|_{H^1(\hat{K})} \leq C |v|_{H^1(\hat{K})}, \quad (\text{II.14})$$

$$\forall v \in H^2(\hat{K}), |\mathcal{T}_0^1 v - v|_{H^1(\hat{K})} \leq C |v|_{H^2(\hat{K})}, \quad (\text{II.15})$$

$$\forall v \in H^2(\hat{K}), \|\mathcal{T}_0^1 v - v\|_{L^2(\hat{K})} \leq C |v|_{H^2(\hat{K})}. \quad (\text{II.16})$$

Alors, on peut en déduire les résultats du théorème précédent. En effet, soit $i \in \{0, \dots, n\}$, on note $K_i = [x_i, x_{i+1}]$. On introduit le changement de variable affine $T_i : \hat{K} \mapsto K_i$ défini par

$$T_i(t) = (1 - t)x_i + tx_{i+1}.$$

On a alors, pour toute fonction $\varphi \in H^1(K_i)$,

$$\|\varphi \circ T_i\|_{L^2(\hat{K})} = \frac{1}{\sqrt{h_i}} \|\varphi\|_{L^2(K_i)},$$

$$|\varphi \circ T_i|_{H^1(\hat{K})} = \frac{1}{\sqrt{h_i}} h_i |\varphi|_{H^1(K_i)}.$$

On constate enfin que l'on a

$$(\mathcal{I}_h^1 u) \circ T_i = \mathcal{I}_0^1(u \circ T_i).$$

- Appliquons à $\varphi = u \circ T_i$ l'inégalité (II.13) et utilisons l'identité de commutation des opérateurs de changement de variable et d'interpolation :

$$\|(\mathcal{I}_h^1 u) \circ T_i - u \circ T_i\|_{L^2(\hat{K})} \leq C |u \circ T_i|_{H^1(\hat{K})},$$

puis les changements de variable et il vient

$$\|\mathcal{I}_h^1 u - u\|_{L^2(K_i)} \leq C h_i |u|_{H^1(K_i)}.$$

On peut ensuite sommer sur i et obtenir le résultat attendu.

- On effectue les autres démonstrations de la même façon.

Moralité : il suffit de s'être estimé les erreurs d'interpolation sur un intervalle fixe \hat{K} pour en déduire tous les résultats locaux sur les intervalles du maillage.

2.1.2 Approximation \mathbb{P}_1 d'un problème elliptique en dimension 1

Présentation de l'exemple Soit à résoudre le problème $-\partial_x(k(x)\partial_x u) + \alpha u = f$, dans $]0, 1[$ avec conditions aux limites de Dirichlet homogène. On suppose $\alpha \geq 0$, $f \in L^2(\Omega)$, $k \in L^\infty(\Omega)$ et $\inf k > 0$.

On a vu que la formulation variationnelle d'un tel problème a lieu dans l'espace $V = H_0^1(\Omega)$ avec les formes bilinéaires a et linéaire L définies par

$$a(u, v) = \int_0^1 k(x) u'(x) v'(x) dx + \alpha \int_0^1 u(x) v(x) dx,$$

$$L(v) = \int_0^1 f(x) v(x) dx.$$

On note $u \in V$, l'unique solution du problème variationnel associé

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

Résolution Revenons au cadre général que nous avons vu plus haut. Si on considère l'espace d'approximation \mathbb{P}_1 introduit ci-dessus et noté $V_{h,0}$, on note $u_h \in V_{h,0}$ l'unique solution approchée du problème discret.

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

D'après le théorème II.10, on a bien $d(u, V_{h,0}) \leq \|u - \mathcal{I}_h^1 u\|_V \rightarrow 0$ quand $h \rightarrow 0$ et donc le théorème général nous garantit la convergence dans H^1 de u_h vers u .

Théorème II.12

Si on suppose que la solution $u \in V$ du problème initial est régulière, i.e. $u \in H^2(\Omega)$, alors on a l'estimation d'erreur

$$\|u_h - u\|_V \leq C \frac{\|a\|}{\alpha} h |u|_{H^2}.$$

On a vu que dans certains cas, le problème elliptique que l'on considère possède des propriétés de régularité qui impliquent que si le second membre est L^2 alors la solution est H^2 . Dans le cas qui nous intéresse, ce sera le cas dès que la fonction k sera régulière (Lipschitzienne suffit), on aura alors la propriété de régularité elliptique suivante : il existe $C > 0$ telle que la solution u du problème variationnel est dans $H^2(\Omega)$ et vérifie

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

Dans ces conditions, l'estimation d'erreur précédente s'applique et on obtient

$$\|u_h - u\|_V \leq Ch\|f\|_{L^2}.$$

Profitons de cette discussion pour obtenir une estimation d'erreur dans une norme plus faible (la norme L^2) mais qui sera d'ordre plus élevé.

Pour ce faire, on a besoin de supposer que le problème adjoint vérifie une propriété de régularité elliptique qui s'écrit de la façon suivante :

Il existe $C > 0$ telle que pour tout $v \in L^2(\Omega)$, l'unique solution $\varphi_v \in V$ du problème

$$a(w, \varphi_v) = (v, w)_{L^2}, \quad \forall w \in V, \quad (\text{II.17})$$

vérifie

$$\varphi_v \in H^2(\Omega), \quad \text{et} \quad \|\varphi_v\|_{H^2} \leq C\|v\|_{L^2}.$$

Dans l'exemple présenté ici, cette propriété est bien entendu vérifiée car le problème adjoint est identique au problème initial par symétrie de la forme bilinéaire a .

Théorème II.13 (Astuce d'Aubin-Nitsche)

Sous les hypothèses précédentes, on a l'estimation suivante

$$\|u - u_h\|_{L^2} \leq Ch\|u - u_h\|_V \leq C'h^2|u|_{H^2}.$$

Preuve :

Notons $e_h = u - u_h \in V$. Si on prend $v = w = e_h$ dans le problème adjoint (II.17), on obtient

$$\|e_h\|_{L^2}^2 = a(e_h, \varphi_{e_h}).$$

Mais, d'après la propriété de a -orthogonalité de l'erreur e_h au sous-espace V_h , on en déduit

$$\|e_h\|_{L^2}^2 = a(e_h, \varphi_{e_h} - \mathcal{I}_h^1 \varphi_{e_h}) \leq \|a\| \|e_h\|_V \|\varphi_{e_h} - \mathcal{I}_h^1 \varphi_{e_h}\|_{H^1} \leq Ch\|e_h\|_V \|\varphi_{e_h}\|_{H^2} \leq Ch\|e_h\|_V \|e_h\|_{L^2}.$$

D'où le résultat. ■

En réalité, ce résultat est assez général et n'est pas spécifique à l'approximation \mathbb{P}_1 , ni au problème considéré.

Quelques mots sur les systèmes linéaires sous-jacents Pour approcher le problème précédent, nous avons N degrés de liberté (car on est dans H_0^1 et donc les deux noeuds du bord ne sont pas des degrés de liberté), correspondant aux coordonnées de la solution u_h dans la base $(\varphi_i)_{1 \leq i \leq N}$ de $V_{h,0}$.

D'un point de vue pratique, on va chercher u_h sous la forme

$$u_h = \sum_{j=1}^N u_j \varphi_j.$$

Tout revient donc à chercher un vecteur $U = (u_j)_{1 \leq j \leq N} \in \mathbb{R}^N$ tel que u_h définie ci-dessus vérifie le système discret. Comme on connaît une base de $V_{h,0}$, il est clair que le système discret est équivalent à

$$\forall i \in \{1, \dots, N\}, \quad a(u_h, \varphi_i) = L(\varphi_i),$$

ou encore, par bilinéarité de a :

$$\forall i \in \{1, \dots, N\}, \quad \sum_{j=1}^N u_j a(\varphi_j, \varphi_i) = L(\varphi_i).$$

Ces équations forment donc un système linéaire de la forme $AU = F$, avec

$$A = (a(\varphi_j, \varphi_i))_{1 \leq i, j \leq N}, \quad F = (L(\varphi_i))_{1 \leq i \leq N}.$$

Revenons à l'exemple considéré plus haut : les coefficients de la matrice A sont donnés par

$$a_{i,j} = \underbrace{\int_{\Omega} k(x) \varphi'_i(x) \varphi'_j(x) dx}_{=k_{i,j}} + \alpha \underbrace{\int_{\Omega} \varphi_i \varphi_j dx}_{=m_{i,j}}.$$

Traditionnellement on découpe ces coefficients en deux parties $k_{i,j}$ et $m_{i,j}$ comme présenté ci-dessus, de sorte que la matrice A s'écrit

$$A = K + \alpha M.$$

La matrice K s'appelle **la matrice de rigidité du système** et la matrice M s'appelle **la matrice de masse**. Dans le cadre de l'élément fini \mathbb{P}_1 , on voit que les intégrales qui interviennent dans $k_{i,j}$ et $m_{i,j}$ sont nulles dès que $|i-j| > 1$ car les supports de φ_i et φ_j sont alors disjoints. C'est un point crucial de la méthode des éléments finis !

Il reste donc à calculer

$$\begin{aligned} k_{i,i} &= \frac{1}{h_{i-1}^2} \int_{x_{i-1}}^{x_i} k(x) dx + \frac{1}{h_i^2} \int_{x_i}^{x_{i+1}} k(x) dx, \quad \text{et} \quad m_{i,i} = \int_{x_{i-1}}^{x_i} \varphi_i(x)^2 dx = \frac{h_{i-1} + h_i}{3}, \\ k_{i,i+1} &= \frac{-1}{h_i^2} \int_{x_i}^{x_{i+1}} k(x) dx, \quad \text{et} \quad m_{i,i+1} = \int_{x_i}^{x_{i+1}} \varphi_i(x) \varphi_{i+1}(x) dx = \frac{h_i}{6}, \\ k_{i,i-1} &= \frac{-1}{h_{i-1}^2} \int_{x_{i-1}}^{x_i} k(x) dx, \quad \text{et} \quad m_{i,i-1} = \int_{x_{i-1}}^{x_i} \varphi_{i-1}(x) \varphi_i(x) dx = \frac{h_{i-1}}{6}, \end{aligned}$$

avec les adaptations immédiates pour les cas $i = 1$ et $i = N$.

Proposition II.14

La matrice A ainsi construite est tridiagonale, symétrique et définie positive. De plus, elle vérifie le principe du maximum discret

$$A^{-1} \geq 0.$$

Preuve :

La première partie du résultat est évidente car il s'agit de la matrice de Gram d'une famille libre de V pour un certain produit scalaire. Pour montrer le principe du maximum discret, on vérifie que A est une M-matrice. ■

Remarquons que pour calculer effectivement cette matrice, il est nécessaire de calculer, *a priori* de façon exacte, des intégrales qui font intervenir les coefficients du problème. En pratique, dans les cas complexes, on remplace ces calculs exacts par des calculs approchés d'intégrales par formules de quadrature. Ceci implique en particulier que le problème discret effectivement résolu n'est pas exactement celui analysé précédemment. Pour être certain de ne pas perdre de précision à ce stade, il est nécessaire de faire une analyse de l'erreur induite par ces formules de quadrature. Nous en reparlerons dans la suite.

- Si le second membre est de la forme $L(v) = \int_{\Omega} f v dx$, alors le vecteur F est donné par les intégrales $\int_{x_{i-1}}^{x_{i+1}} f \varphi_i dx$ qu'il convient aussi de calculer de façon exacte ou approchée. Dans ce dernier cas, il faut en tenir compte dans l'analyse d'erreur.

- Si $a(u, v) = \int_{\Omega} u' v' dx + \int_{\Omega} u' v dx$, alors la forme a n'est plus symétrique, la matrice A n'est alors plus symétrique non plus. Elle n'en demeure pas moins inversible car a est tout de même coercive sur V .

2.1.3 L'élément fini \mathbb{P}_2

Supposons que la solution u du problème que l'on cherche à calculer soit beaucoup plus régulière ($u \in \mathcal{C}^{\infty}$ par exemple). Alors, la méthode précédente, ne permet pas d'obtenir une meilleure précision. La raison en est que l'opérateur d'interpolation naturellement associé à l'espace d'approximation introduit n'est pas plus précis quand on augmente la régularité de la fonction interpolée.

Pour profiter de la régularité de la solution et gagner en précision (sans changer le maillage de départ), alors il faut changer l'espace d'approximation. Dans le paragraphe précédent nous avons considéré un espace de fonctions affines par morceaux, nous allons maintenant considérer un espace de fonctions quadratiques par morceaux.

On repart du maillage précédent et on introduit les centres des cellules $x_{i+1/2} = \frac{x_i + x_{i+1}}{2}$ pour $i = 0, \dots, N$ et on forme l'espace suivant

$$V_h = \{u \in V, \forall i, u|_{K_i} \in \mathbb{P}_2\},$$

où \mathbb{P}_2 désigne l'ensemble des polynômes de degré inférieur ou égal à 2.

Lemme II.15

L'espace V_h peut aussi se définir par

$$V_h = \{u \in C^0(\bar{\Omega}), \forall i, u|_{K_i} \in \mathbb{P}_2\}.$$

De plus, l'application

$$\Phi : u \in V_h \mapsto (u(x_0), u(x_{1/2}), u(x_1), u(x_{3/2}), \dots, u(x_{N+1/2}), u(x_N)) \in \mathbb{R}^{2N+3},$$

est un isomorphisme. En particulier $\dim(V_h) = 2N + 3$.

Preuve :

Le premier résultat est exactement identique au précédent. L'injectivité de Φ est due au fait qu'un polynôme de degré inférieur ou égal à 2 ayant 3 racines est nécessairement nul. La surjectivité de Φ provient de l'existence du polynôme d'interpolation de Lagrange sur chacun des K_i . ■

Bien que nous ayons augmenté le degré des polynômes en question, nous n'avons pas construit un espace d'approximation conforme H^2 , pour cela, il aurait fallu imposer la continuité des dérivées aux interfaces, ce qui n'est pas le cas ici. Chacune des fonctions coordonnées de Φ est une forme linéaire sur V_h qu'on appelle degré de liberté. Elles sont toutes de la forme "évaluation de u en un point", c'est donc qu'on a affaire à un élément fini de Lagrange. Les noeuds de cet élément sont exactement les points x_0, \dots, x_{N+1} et les points $x_{1/2}, \dots, x_{N+1/2}$. On voit donc ici que la notion de "noeud" et celle de sommets du maillage sont bien distinctes.

Les fonctions de forme associées sont les $(\varphi_i)_i$ et les $(\varphi_{i+1/2})$ définis par les relations

$$\varphi_i(x_j) = \delta_{ij}, \forall j \in \{0, \dots, N+1\}, \quad \varphi_i(x_{j+1/2}) = 0, \forall j \in \{0, \dots, N\},$$

$$\varphi_{i+1/2}(x_j) = 0, \forall j \in \{0, \dots, N+1\}, \quad \varphi_{i+1/2}(x_{j+1/2}) = \delta_{ij}, \forall j \in \{0, \dots, N\}.$$

Quelques remarques s'imposent :

- Le support de φ_i est égal à la réunion de K_{i-1} et K_i .
- Le support de $\varphi_{i+1/2}$ est exactement la cellule K_i .

On peut maintenant introduire l'opérateur d'interpolation associé à cet élément fini.

Définition II.16

On appelle opérateur d'interpolation de V dans V_h , l'opérateur \mathcal{I}_h^2 défini par

$$\mathcal{I}_h^2 u(x) = \sum_{i=0}^{N+1} u(x_i) \varphi_i(x) + \sum_{i=0}^N u(x_{i+1/2}) \varphi_{i+1/2}(x).$$

On remarque que si $u \in H_0^1(\Omega)$, alors $\mathcal{I}_h^2 u \in H_0^1(\Omega)$ de sorte que l'espace $V_h \cap H_0^1(\Omega)$ muni du même opérateur d'interpolation est un espace d'approximation conforme de $H_0^1(\Omega)$.

Comme précédemment, on constate que \mathcal{I}_h^2 est un projecteur continu de $H^1(\Omega)$ dans lui-même et on peut montrer les propriétés d'interpolation suivantes.

Théorème II.17

Il existe $C > 0$ telle que pour tout $m \in \{0, 1, 2\}$,

$$\forall u \in H^{m+1}(\Omega), \|\mathcal{I}_h^2 u - u\|_{L^2} + h|\mathcal{I}_h^2 u - u|_{H^1} \leq Ch^{m+1}|u|_{H^{m+1}}.$$

Remarquons que pour les fonction $u \in H^2(\Omega)$, le résultat d'interpolation n'améliore pas celui obtenu pour l'élément fini \mathbb{P}_1 . L'utilisation d'un tel élément n'a donc d'intérêt que si la solution que l'on cherche à approcher est au moins dans $H^3(\Omega)$.

Preuve :

On peut faire la démonstration de ce théorème à la main comme on l'a faite pour l'élément \mathbb{P}_1 , mais cela devient plus pénible à écrire. On peut toutefois utiliser la technique de changement de variable et donc ramener toute l'étude des propriétés d'interpolation au segment de référence $[0, 1]$.

On va voir dans la suite que ce type de résultat peut se démontrer de façon essentiellement automatique sans grand effort. ■

2.1.4 Approximation \mathbb{P}_2 d'un problème elliptique en dimension 1

On considère l'espace d'approximation \mathbb{P}_2 noté V_h construit plus haut et sa version $V_{h,0} = V_h \cap H_0^1(\Omega)$. On obtient la convergence dans le cas général car $d(u, V_{h,0}) \rightarrow 0$ quand $h \rightarrow 0$ pour tout $u \in V$.

Théorème II.18

On suppose que la solution $u \in V$ du problème initial satisfait $u \in H^3(\Omega)$, alors on a les estimations d'erreur

$$\begin{aligned} \|u - u_h\|_{H^1} &\leq Ch^2 |u|_{H^3}, \\ \|u - u_h\|_{L^2} &\leq Ch^3 |u|_{H^3}. \end{aligned}$$

Preuve :

La première propriété se déduit immédiatement de l'analyse générale effectuée plus haut et la seconde s'obtient par l'astuce d'Aubin-Nitsche. ■

2.1.5 Quelques mots sur les systèmes linéaires sous-jacents

En pratique, il faut numéroter les inconnues d'une façon "continue" entre 1 et $2N + 1$ (on rappelle que les deux degrés de liberté en x_0 et x_{N+1} ne sont pas présents à cause des conditions aux limites).

Pour cela, tout $k \in \{1, \dots, 2N + 1\}$ s'écrit $k = 2i + p$ avec $i \in \{0, \dots, N\}$ et $p \in \{0, 1\}$, on notera $k = (i, p)$. Dans ces conditions on peut numéroter les inconnues et les fonctions de base comme suit

$$\psi_k = \psi_{(i,p)} = \begin{cases} \varphi_i, & \text{si } p = 0, \\ \varphi_{i+1/2}, & \text{si } p = 1. \end{cases}$$

De même on notera par la même convention les inconnues u_k de la solution u_h dans la base ψ_k .

Si on reprend le premier exemple traité par éléments finis \mathbb{P}_1 , la matrice A du nouveau système s'écrit comme précédemment $A = (a_{k,l})_{kl}$ avec

$$a_{kl} = \int_{\Omega} \psi'_k \psi'_l dx.$$

Considérons les différents cas :

- Si $k = (i, 0)$, le support de ψ_k est l'union des mailles K_{i-1} et K_i , les seuls coefficients l pour lesquels $a_{k,l}$ est (éventuellement) non nul sont :
 - Si $l = k = (i, 0)$, on obtient $a_{k,k} = \int_{x_{i-1}}^{x_{i+1}} k(x) (\varphi'_i)^2 dx$.
 - Si $l = k + 1 = (i, 1)$, on obtient $a_{k,k+1} = \int_{x_i}^{x_{i+1}} k(x) \varphi'_i \varphi'_{i+1/2} dx$.
 - Si $l = k + 2 = (i + 1, 0)$, on obtient $a_{k,k+2} = \int_{x_i}^{x_{i+1}} k(x) \varphi'_i \varphi'_{i+1} dx$.
 - Si $l = k - 1$ ou $l = k - 2$, on a des formules symétriques.
- Si $k = (i, 1)$, le support de ψ_k est la maille K_i , les seuls coefficients l pour lesquels $a_{k,l}$ est (éventuellement) non nul sont :
 - Si $l = k = (i, 1)$, on obtient $a_{k,k} = \int_{x_i}^{x_{i+1}} k(x) (\varphi'_{i+1/2})^2 dx$.
 - Si $l = k + 1 = (i + 1, 0)$, on obtient $a_{k,k+1} = \int_{x_i}^{x_{i+1}} k(x) (\varphi'_{i+1/2}) (\varphi_{i+1})' dx$.
 - Si $l = k - 1 = (i, 0)$, on obtient $a_{k,k-1} = \int_{x_i}^{x_{i+1}} k(x) (\varphi'_{i+1/2}) (\varphi_i)' dx$.

Ainsi, la matrice A est pentadiagonale, elle est donc potentiellement plus complexe à résoudre que la matrice du cas \mathbb{P}_1 . Elle est toujours symétrique définie positive, mais ce n'est plus une M-matrice et le principe du maximum discret n'est vraisemblablement plus vérifié.

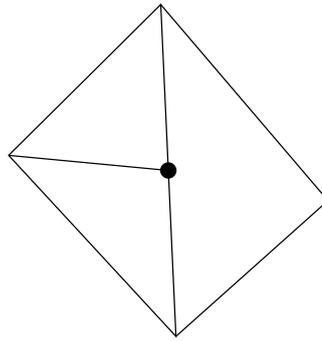


FIGURE II.1 – Un exemple de maillage non géométriquement conforme

2.2 Quelques notions en dimension supérieure

2.2.1 Maillages

On suppose dans toute la suite que Ω est un domaine borné, connexe et polygonal (ou polyhédral en 3D ...). Si l'ouvert a un bord courbe il faut bien sûr en tenir compte dans la mise en place des schémas et dans l'analyse de ceux-ci. Il y a essentiellement deux façons de s'y prendre : ou bien définir des éléments courbes ou bien approcher le domaine original par un domaine polygonal.

Proposition II.19

Soit \mathcal{T} un maillage polygonal de Ω , c'est-à-dire un ensemble de cellules polygonales $(K)_{K \in \mathcal{T}}$ tel que

$$\bar{\Omega} = \bigcup_{K \in \mathcal{T}} \bar{K}, \quad \overset{\circ}{K} \cap \overset{\circ}{L} = \emptyset \text{ si } K \neq L.$$

Soit $m \geq 1$ un entier et u une fonction définie sur Ω . Les deux propriétés suivantes sont équivalentes :

1. u est dans $H^m(\Omega)$.
2. Pour tout $K \in \mathcal{T}$, $u|_K \in H^m(K)$ et pour tout $K \neq L$ tels que $\bar{K} \cap \bar{L}$ soit de co-dimension 1 alors la trace de $\partial^\alpha u|_K$ et celle de $\partial^\alpha u|_L$ coïncident sur $\partial K \cap \partial L$, pour tout $|\alpha| \leq m - 1$.

Dans le cas où ces propriétés sont vérifiées, on a pour tout $|\alpha| \leq m$:

$$\|\partial^\alpha u\|_{L^2(\Omega)}^2 = \sum_{K \in \mathcal{T}} \|\partial^\alpha u\|_{L^2(K)}^2.$$

Ce résultat permet de comprendre comment on peut construire des espaces éléments finis conformes dans H^m en dimension supérieure à 1. On prendra des fonctions régulières par morceaux tout en assurant la nullité du saut (on dit dès fois à tort : "la continuité") de u sur les interfaces du maillage.

On comprend alors comment intervient la notion de conformité géométrique des maillages. En effet, dans la situation de la Figure II.1, on voit que le noeud atypique au milieu de la figure ne peut pas être un noeud associé à un degré de liberté de l'espace d'approximation. Supposons en effet que l'on considère une approximation \mathbb{P}_1 par morceaux, alors la valeur de la solution approchée en ce noeud particulier est liée aux valeurs au sommet des inconnues du grand triangle.

On va dorénavant considérer un maillage de Ω constitué de simplexes (on parle de façon un peu abusive de : triangulation de Ω). Les simplexes sont les enveloppes convexes de $d + 1$ points non contenus dans un hyperplan. C'est en fait un mot savant pour parler de triangles en dimension 2 et de tétraèdres en dimension 3. Pour l'instant on ne fait aucune hypothèse sur le maillage.

2.2.2 Élément fini simplicial de référence. Élément de Lagrange simplicial \mathbb{P}_k

Comme on l'a vu précédemment, afin de construire l'espace d'approximation et d'analyser l'opérateur d'interpolation associé, on est amené à se baser sur un élément particulier que l'on appelle *élément de référence*. Toutes les quantités géométriques du maillage et de l'espace d'approximation vont être définies à partir de cet objet.

On note donc \hat{K} le simplexe unité (engendré par les éléments de la base affine canonique de \mathbb{R}^d) qui servira de référence pour toute la construction. Par définition on a

$$\hat{K} = \{x \in \mathbb{R}^d, \forall i, x_i \geq 0, \sum_i x_i \leq 1\}.$$

Proposition II.20

Tout simplexe $K = \text{conv}(a_0, \dots, a_d)$ du maillage est l'image de \hat{K} par une transformation affine $T : \hat{K} \mapsto K$ de la forme

$$T(\hat{x}) = a_0 + B_K \hat{x},$$

où la i -ième colonne de B_K contient les coordonnées de $a_i - a_0$ dans la base canonique de \mathbb{R}^d .

On a les propriétés suivantes de la matrice B_K :

1. $|\det(B_K)| = d!|K|$.
2. $\|B_K\|_2 \leq \frac{h_K}{\rho_{\hat{K}}}$.
3. $\|B_K^{-1}\|_2 \leq \frac{h_{\hat{K}}}{\rho_K}$.

Dans ces formules h_K et $h_{\hat{K}}$ désignent les diamètres de K et \hat{K} et ρ_K et $\rho_{\hat{K}}$ les "rondeurs" de K et \hat{K} , c'est-à-dire le diamètre de la plus grande boule inscrite dans K et \hat{K} respectivement.

Preuve :

La démonstration de la première propriété est élémentaire.

Démontrons seulement les estimations sur B_K . Pour la première propriété, on utilise le théorème de changement de variable :

$$|K| = \int_K 1 dx = \int_{\hat{K}} 1 |\det B_K| d\hat{x} = |\det B_K| |\hat{K}|,$$

or le volume du simplexe de référence est donné par $|\hat{K}| = \frac{1}{d!}$.

Par définition de la norme $\|\cdot\|_2$, on a

$$\|B_K\|_2 = \sup_{\|\hat{x}\|=\rho_{\hat{K}}} \frac{\|B_K \hat{x}\|}{\rho_{\hat{K}}}.$$

Ce supremum étant atteint (en dimension finie !), il existe $a, b \in \hat{K}$ tel que $\|a - b\| = \rho_{\hat{K}}$. On a alors

$$\|B_K(a - b)\| = \|Ta - Tb\| \leq h_K,$$

d'où le résultat. L'autre inégalité se démontre en échangeant les rôles de K et \hat{K} . ■

Dans ce résultat il faut comprendre que $h_{\hat{K}}$ et $\rho_{\hat{K}}$ ne dépendent que de la dimension (ce sont des constantes).

Théorème II.21

Soit $v : K \mapsto \mathbb{R}$ une fonction définie sur K . On a $v \in H^m(K)$ si et seulement si $\hat{v} = v \circ T \in H^m(\hat{K})$. De plus, pour tout $0 \leq k \leq m$, on a

$$|v|_{H^k(K)} \leq C \frac{|K|^{\frac{1}{2}}}{\rho_K^k} |\hat{v}|_{H^k(\hat{K})},$$

$$|\hat{v}|_{H^k(\hat{K})} \leq C \frac{h_K^k}{|K|^{\frac{1}{2}}} |v|_{H^k(K)}.$$

Preuve :

Effectuons la démonstration pour $m = 1$. Le cas général se démontre de la même façon par un calcul à peine plus compliqué.

Dans le cas $k = 0$, il s'agit d'estimer la norme L^2 . On utilise la formule de changement de variable

$$\|v\|_{L^2(K)}^2 = \int_K |v(x)|^2 dx = \int_{\hat{K}} |\hat{v}(\hat{x})|^2 |\det B_K| d\hat{x} = |K| \|\hat{v}\|_{L^2(\hat{K})}^2,$$

ce qui donne le résultat.

Dans le cas $k = 1$, on a

$$\nabla \hat{v} = B_K^t (\nabla v) \circ T = B_K^t \widehat{\nabla} v,$$

d'où avec le calcul précédent

$$\|\nabla \hat{v}\|_{L^2(\hat{K})}^2 \leq \frac{h_K^2}{\rho_{\hat{K}}^2} \|\widehat{\nabla} v\|_{L^2(\hat{K})}^2 = \frac{h_K^2}{\rho_{\hat{K}}^2 |K|} \|\nabla v\|_{L^2(K)}^2.$$

Cela donne le second résultat. L'autre inégalité s'obtient en inversant les rôles de K et \hat{K} . ■

Proposition et Définition II.22 (Élément fini simplicial de Lagrange \mathbb{P}_k)

On considère l'ensemble $\Sigma \subset \hat{K}$ des points définis de la façon suivante

$$\left(\frac{i_1}{k}, \dots, \frac{i_d}{k} \right), \quad \forall (i_1, \dots, i_d) \in \mathbb{N}^d, i_1 + \dots + i_d \leq k.$$

Le cardinal de cet ensemble est donné par $|\Sigma| = C_{d+k}^k = \frac{(k+d)!}{d!k!}$. On note $(\hat{a}_j)_{1 \leq j \leq |\Sigma|}$ les éléments de cet ensemble.

Alors l'application $p \in \mathbb{P}_k \mapsto (p(\hat{a}_j))_{1 \leq j \leq |\Sigma|} \in \mathbb{R}^{|\Sigma|}$ est un isomorphisme, en particulier \mathbb{P}_k et Σ ont le même cardinal.

- Le triplet $(\hat{K}, \mathbb{P}_k, \Sigma)$ est appelé : élément fini simplicial de Lagrange \mathbb{P}_k .
- L'ensemble des points $(\hat{a}_j)_j$ est l'ensemble des noeuds de cet élément fini et la forme linéaire qui à toute fonction continue v associe $v(\hat{a}_j)$ est appelée le degré de liberté associé au noeud \hat{a}_j .
- Pour tout $1 \leq i \leq |\Sigma|$, il existe une unique fonction $\theta_i \in \mathbb{P}_k$ telle que

$$\theta_i(\hat{a}_j) = \delta_{ij}, \quad \forall 1 \leq j \leq |\Sigma|.$$

Ces fonctions sont appelées : fonctions de forme locales

Exemple II.23

- L'élément fini \mathbb{P}_1 : les noeuds sont exactement les sommets de \hat{K} .
- L'élément fini \mathbb{P}_2 : les noeuds sont les sommets de \hat{K} et les milieux des arêtes.
- L'élément fini \mathbb{P}_3 en 2D : les noeuds sont : les sommets, le centre de gravité et les points aux $1/3$ et $2/3$ de chaque arête du triangle.

Définition II.24 (Maillage simplicial géométriquement conforme)

Soit \mathcal{T} un maillage de Ω formé de cellules K simpliciales. On dit que le maillage est géométriquement conforme si et seulement si pour tout couple (K, L) de cellules distinctes, l'intersection $F = \overline{K} \cap \overline{L}$ vérifie l'une des propriétés suivantes :

- ou bien F est de dimension inférieure ou égale à $d - 2$ (en particulier F peut être vide).
- ou bien F est à la fois une face de K et une face de L .

Proposition et Définition II.25 (Espace d'approximation \mathbb{P}_k associé à un maillage simplicial)

Soit \mathcal{T} un maillage simplicial géométriquement conforme du domaine Ω .

- On note $\Sigma_h = \bigcup_K T_K(\Sigma)$ l'ensemble des noeuds de la discrétisation.
- On appelle espace d'approximation \mathbb{P}_k sur le maillage \mathcal{T} , l'espace

$$V_h = \{v \in C^0(\overline{\Omega}), \forall K \in \mathcal{T}, v|_K \in \mathbb{P}_k\}.$$

- Pour tout $a \in \Sigma_h$, il existe une unique fonction dans V_h , notée φ_a telle que

$$\varphi_a(a) = 1, \quad \varphi_a(b) = 0, \forall b \in \Sigma_h \setminus \{a\}.$$

Ces fonctions s'appellent les fonctions de forme globales de l'espace d'approximation. Elles forment une base de V_h .

- Pour tout $a \in \Sigma_h$, l'application $v \in C^0(\overline{\Omega}) \mapsto v(a)$ est une forme linéaire continue appelée : degré de liberté associé au noeud a .

Proposition et Définition II.26

L'espace V_h vérifie

$$V_h = \{v \in H^1(\Omega), \forall K \in \mathcal{T}, v|_K \in \mathbb{P}_k\}.$$

De plus, pour tout $a \in \Sigma_h$ et tout $K \in \mathcal{T}$, tel que $a \in K$, il existe un unique indice $J(a, K) \in \{1, \dots, |\Sigma|\}$ tel que $a = T_K(\hat{a}_{J(a, K)})$. On a alors :

$$(\varphi_a)|_K \circ T_K = \theta_{J(a, K)}.$$

On dira alors que $K \in \text{Supp } \varphi_a$.

Pour les cellules K qui ne contiennent pas a , on a $(\varphi_a)|_K = 0$ et on dira que $K \notin \text{Supp } \varphi_a$.

Preuve :

- Le point délicat dans la première partie du résultat concerne l'existence de $J(a, K)$. L'unicité est triviale car tous les noeuds locaux sont distincts et T_K est bijectif.

D'où vient alors la difficulté ? Si $a \in \Sigma_h$, c'est qu'*a priori* il existe un élément $L \in \mathcal{T}$ tel que $a \in T_L(\Sigma)$. Si maintenant $K \in \mathcal{T}$ est un autre élément contenant a , il faut vérifier que a est aussi un élément de $T_K(\Sigma)$.

Regardons seulement la démonstration en dimensions 2 et 3 et remarquons que le seul cas intéressant est celui où $a \in \partial L$.

Soit alors $F = K \cap L$. Par hypothèse de conformité géométrique du maillage, F est soit un sommet de K et L , soit une arête commune soit une face commune à K et L .

On note alors $F_K = T_K^{-1}(F)$ et $F_L = T_L^{-1}(F)$, de sorte que F_K et F_L sont ou bien des sommets, ou bien des arêtes ou bien des faces de l'élément de référence \hat{K} .

Or, on peut constater (et c'est une propriété importante de l'élément de Lagrange simplicial \mathbb{P}_k), que la configuration géométrique des noeuds sur les sommets, arêtes et faces de l'élément ne dépendent pas du sommet, de l'arête ou de la face considérée. Cela signifie qu'il existe une bijection $\Phi_{L,K}$ **affine** de F_L sur F_K qui envoie les noeuds de $F_L \cap \Sigma$ exactement sur les noeuds de $F_K \cap \Sigma$.

On voit qu'en fait, on peut changer Φ_{LK} , en composant par une bijection affine qui laisse F_K invariante, pour faire en sorte que $T_K \circ \Phi_{LK} \circ T_L^{-1}$ soit exactement l'identité sur F , ce qui montre le résultat.

- Définissons φ_a par les formules données dans l'énoncé et vérifions que c'est bien la fonction de forme que l'on cherche. Tout d'abord, on constate que comme T_K est affine (son inverse aussi), on a bien $(\varphi_a)|_K \in \mathbb{P}_k$ pour toute cellule $K \in \mathcal{T}$.

Soient maintenant K et L deux éléments qui partagent une face F en commun. On a alors que $(\varphi_a)|_{K \cap F}$ est un élément de \mathbb{P}_k sur la face F en coordonnées barycentriques qui est nul sur tous les sommets de $F \cap \Sigma_h$ sauf éventuellement en a (si $a \in F$). De même $(\varphi_a)|_{L \cap F}$ vérifie les mêmes propriétés. Or, il existe un unique élément de \mathbb{P}_k sur F qui prend ces valeurs sur les noeuds (on considère F comme un élément fini simplicial de Lagrange en dimension $d - 1$!). On en déduit donc que

$$(\varphi_a)|_{K \cap F} = (\varphi_a)|_{L \cap F},$$

et donc que φ_a est une fonction continue et même un élément de $H^1(\Omega)$.

Soit maintenant $b \in \Sigma_h$ un autre noeud de la discrétisation ($a \neq b$). S'il existe $K \neq L$ tels que $a \in K$ et $b \in L$, alors par définition $(\varphi_a)|_L = 0$ et donc $\varphi_a(b) = 0$.

Si a et b appartiennent au même élément K , alors il existe $J' \neq J(a, K)$ tel que $b = T_K(\hat{a}_{J'})$ et donc $\varphi_a(b) = \theta_{J(a, K)}(\hat{a}_{J'}) = 0$, par définition des fonctions de forme locale.

■

La conséquence de ce résultat, c'est que toutes les quantités importantes que l'on doit calculer pour mettre en place la méthode des éléments finis sont issues d'informations locales sur l'élément de référence.

2.2.3 Opérateur d'interpolation local. Opérateur d'interpolation global

Définition II.27 (Opérateur d'interpolation local)

On appelle **opérateur d'interpolation local**, l'opérateur qui à toute fonction \hat{v} continue sur \hat{K} associe $\mathcal{I}_0^k \hat{v}$ définie par

$$\mathcal{I}_0^k \hat{v}(\hat{x}) = \sum_{i=1}^{|\Sigma|} \hat{v}(\hat{a}_i) \theta_i(\hat{x}).$$

C'est donc aussi l'unique polynôme de \mathbb{P}_k qui coïncide avec \hat{v} sur les noeuds de l'élément de référence.

Définition II.28 (Opérateur d'interpolation global)

On appelle **opérateur d'interpolation global**, l'opérateur qui à toute fonction v continue sur $\bar{\Omega}$ associe $\mathcal{I}_h^k v$ définie par

$$\mathcal{I}_h^k v(x) = \sum_{a \in \Sigma_h} v(a) \varphi_a(x).$$

C'est donc aussi l'unique fonction de V_h qui coïncide avec v sur les noeuds de la discrétisation.

Ces deux opérateurs sont liés par la formule suivante

Lemme II.29

- Pour tout $K \in \mathcal{T}$, on a

$$(\mathcal{I}_h^k v)|_K \circ T_K = \mathcal{I}_0^k(v \circ T_K).$$

- Si v est une fonction continue, nulle sur le bord de Ω . Alors $\mathcal{I}_h^1 v$ est également nulle sur le bord de Ω . En particulier, \mathcal{I}_h^1 est aussi un opérateur d'interpolation que l'on peut utiliser dans $V_{h,0} = V_h \cap H_0^1(\Omega)$.

Preuve :

- Il suffit de constater que $(\mathcal{I}_h^k v)|_K \circ T_K$ et $\mathcal{I}_0^k(v \circ T_K)$ sont deux éléments de \mathbb{P}_k dont les valeurs coïncident sur tous les noeuds de l'élément de référence. Par unicité du polynôme d'interpolation de Lagrange, la propriété est démontrée.
- On a déjà vu plus haut que la restriction de $\mathcal{I}_h^k v$ à une face du maillage ne dépend que des degrés de liberté associés aux noeuds sur la face F . Donc si les valeurs de ces degrés de liberté sont nulles, alors la restriction de $\mathcal{I}_h^k v$ sur la face F est identiquement nulle. Le résultat demandé est donc établi en prenant une face F incluse dans $\partial\Omega$. ■

Ces opérateurs ont pour "défaut" de n'être définis que pour des fonctions continues. Or, il se trouve que les fonctions de $H^1(\Omega)$ ne sont pas continues en dimension $d \geq 2$. En revanche, en dimensions $d = 2$ et $d = 3$, l'espace $H^2(\Omega)$ s'injecte bien dans l'espace des fonctions continues et donc on peut utiliser cet opérateur.

2.2.4 Analyse de l'erreur d'interpolation

Commençons par deux outils généraux.

Théorème II.30 (Lemme de Deny-Lions)

Soit U un ouvert borné lipschitzien de \mathbb{R}^d et $k \in \mathbb{N}$. On note \mathbb{P}_k l'ensemble des polynômes de degré inférieur ou égal à k sur U .

Il existe une constante $C > 0$ telle que

$$\forall u \in H^{k+1}(\Omega), \quad \inf_{\pi \in \mathbb{P}_k} \|u - \pi\|_{H^{k+1}(U)} \leq C |u|_{H^{k+1}(U)}.$$

Preuve :

Pour tout multi-indice $\alpha \in \mathbb{N}^d$ tel que $|\alpha| \leq k$, on note f_α la forme linéaire définie sur $H^{k+1}(\Omega)$ par

$$f_\alpha(v) = \int_U \partial^\alpha v \, dx.$$

L'ensemble de tels multi-indices a exactement la dimension de \mathbb{P}_k (il y en a autant que de monômes distincts dans cet espace).

- On constate tout d'abord que l'application linéaire

$$\pi \in \mathbb{P}_k \mapsto (f_\alpha(\pi))_{|\alpha| \leq k} \in \mathbb{R}^{(\dim \mathbb{P}_k)}, \quad (\text{II.18})$$

est bijective. En effet, pour des raisons de dimension, il suffit de montrer l'injectivité de cette application. Si on suppose qu'un polynôme non nul $\pi \in \mathbb{P}_k$ annule cette application, on peut trouver un monôme non nul d'indice maximal α dans π , mais alors on voit que $\partial^\alpha \pi$ est une constante non nulle dont l'intégrale ne peut pas s'annuler.

- On montre maintenant qu'il existe $C > 0$ telle que

$$\forall u \in H^{k+1}(U), \quad \|u\|_{H^{k+1}} \leq C \left(|u|_{H^{k+1}} + \sum_{|\alpha| \leq k} |f_\alpha(u)| \right). \quad (\text{II.19})$$

Supposons que cette propriété soit fautive, il existe alors une suite $(u_n)_n$ d'éléments de $H^{k+1}(U)$ telle que

$$\|u_n\|_{H^{k+1}(U)} = 1, \quad (\text{II.20})$$

$$|u_n|_{H^{k+1}} + \sum_{|\alpha| \leq k} |f_\alpha(u_n)| \leq \frac{1}{n}. \quad (\text{II.21})$$

D'après ces inégalités, on peut extraire une sous-suite $(u_{\varphi(n)})_n$ qui converge faiblement dans H^{k+1} vers une fonction u . De plus, par compacité, la convergence de $u_{\varphi(n)}$ vers u est forte dans $H^k(U)$.

D'après (II.21), toutes les dérivées d'ordre exactement $k+1$ de $(u_{\varphi(n)})_n$ tendent vers 0. Ceci prouve que les dérivées partielles d'ordre exactement $k+1$ de u sont nulles et donc $u \in \mathbb{P}_k$. De plus, en passant également à la limite dans les $f_\alpha(u_{\varphi(n)})$, on obtient que $f_\alpha(u) = 0$ pour tout $|\alpha| \leq k$. D'après le premier point, ceci prouve que $u = 0$.

Si on reprend maintenant (II.20), on trouve que $\|u_{\varphi(n)}\|_{H^k} \rightarrow 1$ et par convergence forte on en déduit que $\|u\|_{H^k} = 1$, ce qui contredit le fait que u soit nulle.

- Soit maintenant $u \in H^{k+1}(U)$ quelconque. Par surjectivité de l'application (II.18), il existe un polynôme $\tilde{\pi} \in \mathbb{P}_k$ tel que pour tout $|\alpha| \leq k$, on a $f_\alpha(u - \tilde{\pi}) = 0$.

On en déduit alors

$$\inf_{\pi \in \mathbb{P}_k} \|u - \pi\|_{H^{k+1}} \leq \|u - \tilde{\pi}\|_{H^{k+1}} \leq C \left(|u - \tilde{\pi}|_{H^{k+1}} + \sum_{|\alpha| \leq k} \underbrace{|f_\alpha(u - \tilde{\pi})|}_{=0} \right) \leq C |u - \tilde{\pi}|_{H^{k+1}} = C |u|_{H^{k+1}}.$$

■

Théorème II.31 (Lemme de Bramble-Hilbert)

Soit U un ouvert borné lipschitzien de \mathbb{R}^d , $k \in \mathbb{N}$ et Φ un opérateur linéaire continu de $H^{k+1}(U)$ dans un espace de Banach E . Si Φ s'annule sur \mathbb{P}_k , alors il existe $C > 0$ telle que

$$\forall u \in H^{k+1}(U), \quad \|\Phi u\|_E \leq C |u|_{H^{k+1}(U)}.$$

Preuve :

Pour tout $u \in H^{k+1}(U)$ et tout $\pi \in \mathbb{P}_k$ nous avons

$$\|\Phi u\|_E = \|\Phi(u - \pi)\|_E \leq \|\Phi\| \|u - \pi\|_{H^{k+1}(U)}.$$

En prenant l'infimum sur π et en utilisant le lemme de Deny-Lions, il vient

$$\|\Phi u\|_E \leq C |u|_{H^{k+1}(U)}.$$

■

On peut déduire de ces résultats généraux, le résultat d'interpolation suivant :

Théorème II.32 (Propriété de l'opérateur d'interpolation)

Sous les hypothèses précédentes, on se donne un entier $0 \leq m \leq k$ et on suppose que $m + 1 > d/2$, de sorte que $H^{m+1}(\Omega) \subset C^0(\bar{\Omega})$.

Pour $p \in \{0, 1\}$, il existe $C > 0$ telle que

$$\forall K \in \mathcal{T}, \forall v \in H^{m+1}(K), \quad |v - \mathcal{I}_h^k v|_{H^p(K)} \leq C \frac{h_K^{m+1}}{\rho_K^p} |v|_{H^{m+1}(K)}. \quad (\text{II.22})$$

Ainsi, on a l'estimation globale

$$\forall v \in H^{m+1}(\Omega), \quad |v - \mathcal{I}_h^k v|_{H^p(\Omega)} \leq C \sigma_{\mathcal{T}}^p h_K^{m+1-p} |v|_{H^{m+1}(\Omega)}, \quad (\text{II.23})$$

où

$$\sigma_{\mathcal{T}} = \sup_{K \in \mathcal{T}} \frac{h_K}{\rho_K}.$$

Définition II.33

On dit qu'une famille $(\mathcal{T}_h)_h$ de maillages de Ω est régulière si et seulement si il existe une constante $C > 0$ telle que

$$\sigma_{\mathcal{T}_h} \leq C, \quad \forall h > 0.$$

On remarque que cette propriété est équivalente à demander que toutes les cellules aient un volume d'ordre h_K^d :

$$\exists C > 0, \forall h > 0, \forall K \in \mathcal{T}_h, \quad |K| \geq Ch_K^d.$$

Grâce à cette définition, on constate que l'estimation d'interpolation globale (II.23) fournit l'estimation de l'erreur d'approximation optimale attendue pour l'espace d'approximation V_h construit sur une triangulation régulière. On en déduit donc le résultat d'estimation d'erreur suivant :

Théorème II.34 (Estimation d'erreur pour l'approximation \mathbb{P}_k d'un problème elliptique)

Soit a une forme bilinéaire continue et coercive sur $H_0^1(\Omega)$ et L une forme linéaire continue sur $H_0^1(\Omega)$. Soit $(\mathcal{T}_h)_h$ une famille régulière de maillages simpliciaux de Ω et $(V_h)_h$ les espaces d'approximation Lagrange \mathbb{P}_k construits sur ces maillages et $V_{h,0} = V_h \cap H_0^1(\Omega)$ leurs pendents dans $H_0^1(\Omega)$. On suppose que la solution $u \in H_0^1(\Omega)$ du problème elliptique suivant

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

est dans $H^{m+1}(\Omega)$, pour un certain $m \leq k$. On note u_h la solution du problème approché construit sur l'espace $V_{h,0}$. Il existe une constante $C > 0$ qui dépend de Ω , de a , de $\sup_h(\sigma_{\mathcal{T}_h})$ telle que

$$\|u - u_h\|_{H^1} \leq Ch^{m+1} |u|_{H^{m+1}(\Omega)}.$$

Si de plus, le problème adjoint admet une propriété de régularité elliptique, alors on a l'estimation en norme L^2 :

$$\|u - u_h\|_{L^2} \leq Ch^{m+2} |u|_{H^{m+1}(\Omega)}.$$

Preuve :

Ce théorème se déduit immédiatement de l'étude générale donnée au début du chapitre et des propriétés de l'opérateur d'interpolation \mathcal{I}_h^k obtenues plus haut, au moins dans le cas où $m + 1 \geq d/2$. Dans le cas contraire, il faut construire un autre opérateur d'interpolation. On le verra par la suite.

Pour le second point, il suffit d'appliquer la technique d'Aubin-Nitsche. ■

Démontrons maintenant les propriétés de \mathcal{I}_h^k .

Preuve (du Théorème II.32):

- L'estimation globale (II.23) se déduit immédiatement des estimations locales, d'après la proposition II.19. C'est ici qu'intervient la restriction $p \in \{0, 1\}$ car les fonctions de V_h ne sont pas dans $H^p(\Omega)$ pour $p \geq 2$. En revanche, l'estimation locale est vraie pour tout $p \leq m + 1$.
- Démontrons donc l'estimation locale pour $p \leq m + 1$. Commençons par utiliser le lemme de Bramble-Hilbert (Théorème II.31). Comme l'opérateur d'interpolation local \mathcal{I}_0^k laisse invariant les éléments de \mathbb{P}_k , on en déduit que l'application $\Phi : \hat{v} \in H^{m+1}(\hat{K}) \mapsto \hat{v} - \mathcal{I}_0^k \hat{v} \in H^p(\hat{K})$ s'annule sur \mathbb{P}_k . De plus, c'est un opérateur linéaire et continu (car $H^{m+1}(\hat{K})$ s'injecte continûment dans $C^0(\hat{K})$). D'après le lemme de Bramble-Hilbert, il existe $C > 0$ telle que

$$\forall \hat{v} \in H^{m+1}(\hat{K}), \quad \|\hat{v} - \mathcal{I}_0^k \hat{v}\|_{H^p(\hat{K})} \leq C |\hat{v}|_{H^{m+1}(\hat{K})}.$$

On utilise maintenant le Théorème de changement de variable II.21. Soit $v \in H^{m+1}(K)$. On pose $\hat{v} = v \circ T_K$ et on a alors

$$\begin{aligned} |v - \mathcal{I}_h^k v|_{H^p(K)} &\leq C \frac{|K|^{\frac{1}{2}}}{\rho_K^p} |v \circ T_K - \mathcal{I}_h^k v \circ T_K|_{H^p(\hat{K})} \\ &= C \frac{|K|^{\frac{1}{2}}}{\rho_K^p} |\hat{v} - \mathcal{I}_0^k \hat{v}|_{H^p(\hat{K})} \\ &\leq C' \frac{|K|^{\frac{1}{2}}}{\rho_K^p} |\hat{v}|_{H^{m+1}(\hat{K})} \\ &\leq C'' \frac{h_K^{m+1}}{\rho_K^p} |v|_{H^{m+1}(K)}. \end{aligned}$$

Ceci donne exactement le résultat escompté. ■

2.2.5 Compléments : inégalités de Sobolev inverses

Proposition II.35 (Inégalité de Sobolev inverse locale)

Soit $0 \leq l \leq m$. Il existe une constante $C > 0$ telle que

$$\forall v \in V_h, \forall K \in \mathcal{T}, \quad |v|_{H^m(K)} \leq C \frac{1}{\rho_K^{m-l}} |v|_{H^l(K)}.$$

Preuve :

Quitte à appliquer le résultat aux dérivées partielles de v_h , il suffit de montrer le résultat pour $l = 0$. On prend maintenant $\hat{v} = v \circ T_K \in H^l(\hat{K})$. Dans \mathbb{P}_k (qui est de dimension finie), toutes les normes sont équivalentes, en particulier il existe $C > 0$ tel que

$$|\hat{v}|_{H^m(\hat{K})} \leq C \|\hat{v}\|_{L^2(\hat{K})}, \quad \forall \hat{v} \in \mathbb{P}_k.$$

Ainsi, on a

$$|v \circ T_K|_{H^m(\hat{K})} \leq C \|v \circ T_K\|_{L^2(\hat{K})},$$

et donc en utilisant les inégalités du théorème II.21, on trouve

$$|v|_{H^m(K)} \leq \frac{C}{\rho_K^m} |v|_{L^2(K)}.$$
■

Théorème II.36 (Inégalité de Sobolev inverse globale)

Soit $0 \leq l \leq m$. Il existe $C_\sigma > 0$ dépendant de $\sigma = \sup_K \frac{h_K}{\rho_K}$ telle que

$$\forall v \in V_h, \forall K \in \mathcal{T}, \quad \sum_K ((h_K^{m-l}|v|_{H^m(K)})^2 \leq C_\sigma \sum_K |v|_{H^l(K)}^2.$$

Si de plus, on suppose que $m \leq 1$ et que l'on introduit la constante d'uniformité du maillage

$$\tau = \sup_K \frac{h}{h_K},$$

alors il existe $C_{\sigma,\tau} > 0$ telle que

$$\forall v \in V_h, \|v_h\|_{H^m(\Omega)} \leq C_{\sigma,\tau} h^{l-m} \|v_h\|_{H^l(\Omega)}.$$

Preuve :

Il suffit de voir que dans ce cas, les semi-normes brisées se recollent bien en des normes sur tout Ω . ■

On dira d'une famille de triangulations \mathcal{T}_n qu'elle est régulière et quasi-uniforme si les constantes σ_n et τ_n associées à chacun des maillages restent bornées quand n tend vers l'infini. Dans ces conditions, on dispose de l'inégalité de Sobolev inverse

$$\forall v_h \in V_h, \quad \|v_h\|_{H^1(\Omega)} \leq \frac{C}{h} \|v_h\|_{L^2(\Omega)}.$$

2.2.6 Autre opérateur d'interpolation

L'opérateur d'interpolation de Lagrange \mathcal{I}_h^k étudié précédemment est assez pratique car facile à construire et à étudier. De plus, il respecte les conditions aux limites de Dirichlet homogène et l'estimation de l'erreur d'interpolation est purement locale. On a vu qu'on ne pouvait le définir que pour des fonctions de $H^{m+1}(\Omega)$ car on a besoin de la continuité de ces fonctions pour bien faire.

En réalité, on peut construire d'autres opérateurs "d'interpolation" qui passent outre ces difficultés. Nous allons en donner un exemple ici.

Soit \mathcal{T} un maillage simplicial de Ω . A tout noeud $a \in \Sigma_h$, on associe un simplexe de dimension d ou $d-1$, noté K_a défini par

- Si a est à l'intérieur d'une cellule K , i.e. $a \in \overset{\circ}{K}$, on pose $K_a = K$.
- Si a est sur le bord d'une cellule K , on pose $K_a = F$ ou F est une face du maillage qui contient a .

Si a est à l'intérieur d'une face, alors dans le second cas, on a unicité du choix de K_a , c'est l'unique face à laquelle a appartient.

En revanche, si a est un sommet du maillage (par exemple) ou s'il appartient à une arête en dimension 3, alors il y a plusieurs choix possibles pour K_a . Si $a \in \partial\Omega$, on choisit K_a de telle sorte que $K_a \subset \partial\Omega$, ce qui est toujours possible. On fait ce choix une bonne fois pour toute pour chaque noeud a de la discrétisation.

On introduit ensuite $P_a = \{(\varphi_b)|_{K_a}, \forall b \in \Sigma_h\} \setminus \{0\}$, l'ensemble des traces (non nulles) des fonctions de base sur K_a . On note cet ensemble

$$P_a = \{\varphi_i^a, i = 1, \dots, n_a\},$$

qui est une famille libre, en supposant par convention que φ_1^a est l'unique fonction de cet ensemble qui vaut 1 au point a , c'est-à-dire la restriction à K_a de la fonction de forme φ_a . Enfin, on notera

$$Q_a = \{\psi_i^a, i = 1, \dots, n_a\},$$

la base $L^2(K_a)$ -orthogonale de P_a , définie par

$$\psi_i^a \in \text{Vect}(P_a), \forall i = 1, \dots, n_a,$$

$$\int_{K_a} \varphi_i^a \psi_j^a dx = \delta_{ij}.$$

Définition II.37 (Opérateur d'interpolation de Scott-Zhang)

Pour tout $v \in H^1(\Omega)$ et tout noeud $a \in \Sigma_h$, on note

$$\pi_a(v) = \int_{K_a} v(x) \psi_1^a(x) dx.$$

On définit alors, l'opérateur d'interpolation de Scott-Zhang Π_h^{SZ} par la formule

$$\Pi_h^{SZ} v = \sum_{a \in \Sigma_h} \pi_a(v) \varphi_a.$$

Théorème II.38

L'opérateur de Scott-Zhang vérifie les propriétés suivantes :

- C'est un opérateur d'interpolation au sens suivant

$$\forall v \in \mathbb{P}_k, \quad \Pi_h^{SZ} v = v.$$

- Il préserve les conditions aux limites de Dirichlet homogène :

$$\Pi_h^{SZ}(H_0^1(\Omega)) \subset H_0^1(\Omega).$$

- On a les propriétés de stabilité et approximation suivantes : pour tout $0 \leq m \leq k$ et tout $0 \leq p \leq m$

$$\forall v \in H^{m+1}(\Omega), \quad \|v - \Pi_h^{SZ} v\|_{H^p(\Omega)} \leq C_\sigma h^{m-p+1} |v|_{H^{m+1}(\Omega)},$$

où C_σ dépend de σ qui est n'importe quel réel tel que

$$\sup_{K \in \mathcal{T}} \frac{h_K}{\rho_K} \leq \sigma.$$

2.2.7 Assemblage - Quadratures - Conditions aux limites

Considérons le même problème que celui considéré en dimension 1 d'espace

$$-\operatorname{div}(k(x)\nabla u) + \alpha u = f,$$

avec des conditions aux limites de Dirichlet homogène.

On se place donc dans $V = H_0^1(\Omega)$ et on pose

$$a(u, v) = \int_{\Omega} k(x) \nabla u \cdot \nabla v dx + \alpha \int_{\Omega} uv dx, \quad L(v) = \int_{\Omega} fv dx.$$

On choisit maintenant un maillage simplicial géométriquement conforme \mathcal{T} de Ω et on construit l'espace d'approximation $V_{h,0}$ basé sur l'élément de Lagrange \mathbb{P}_k . Soit $\Sigma_{h,0} = \Sigma_h \cap \Omega$ l'ensemble des noeuds intérieurs au domaine et soit N le cardinal de cet ensemble. On numérote de 1 à N les noeuds a_j de cet ensemble et on note, pour simplifier, $\varphi_j = \varphi_{a_j}$ la fonction de forme associé à un tel noeud.

La matrice A du système approché est alors donnée par $(a(\varphi_i, \varphi_j))_{1 \leq i, j \leq N}$ et de façon plus précise

$$a(\varphi_i, \varphi_j) = \int_{\Omega} k(x) \nabla \varphi_i \cdot \nabla \varphi_j dx + \alpha \int_{\Omega} \varphi_i \varphi_j dx.$$

Découpons chaque intégrale en intégrales sur les cellules :

$$a(\varphi_i, \varphi_j) = \sum_{K \in \mathcal{T}} \left(\int_K k(x) \nabla \varphi_i \cdot \nabla \varphi_j dx \right) + \alpha \sum_{K \in \mathcal{T}} \left(\int_K \varphi_i \varphi_j dx \right).$$

On voit que les intégrales élémentaires (sur chaque K) ne sont éventuellement non nulles que si elles sont simultanément dans le support de φ_i et de φ_j . On en déduit que

$$a(\varphi_i, \varphi_j) = \sum_{K \in \text{Supp } \varphi_i \cap \text{Supp } \varphi_j} \left(\int_K k(x) \nabla \varphi_i \cdot \nabla \varphi_j dx \right) + \alpha \sum_{K \in \text{Supp } \varphi_i \cap \text{Supp } \varphi_j} \left(\int_{\Omega} \varphi_i \varphi_j dx \right).$$

Or, si $K \in \text{Supp } \varphi_i \cap \text{Supp } \varphi_j$, cela signifie que les noeuds a_i et a_j appartiennent tous les deux à l'élément K . En particulier, $J(a_i, K)$ et $J(a_j, K)$ sont bien définis tous les deux et on a

$$\varphi_i \circ T_K = \theta_{J(a_i, K)}, \quad \varphi_j \circ T_K = \theta_{J(a_j, K)},$$

et donc

$$(B_K^t)(\nabla \varphi_i) \circ T_K = \nabla \theta_{J(a_i, K)}, \quad (B_K^t)(\nabla \varphi_j) \circ T_K = \nabla \theta_{J(a_j, K)}.$$

Ainsi par changement de variable, on a

$$\begin{aligned} \int_K k(x) \nabla \varphi_i \cdot \nabla \varphi_j dx &= |\det B_K| \int_{\hat{K}} k(T_K(\hat{x})) \left((B_K^t)^{-1} \nabla \theta_{J(a_i, K)} \right) \cdot \left((B_K^t)^{-1} \nabla \theta_{J(a_j, K)} \right) d\hat{x} \\ &= |\det B_K| \int_{\hat{K}} k(T_K(\hat{x})) \left((B_K^t B_K)^{-1} \nabla \theta_{J(a_i, K)} \right) \cdot \nabla \theta_{J(a_j, K)} d\hat{x}, \end{aligned}$$

et

$$\int_K \varphi_i \varphi_j dx = |\det B_K| \int_{\hat{K}} \theta_{J(a_i, K)} \theta_{J(a_j, K)} dx.$$

Ainsi, pour construire la matrice A , il suffit de savoir calculer des intégrales de ce type directement sur l'élément de référence. Pour le second type d'intégrale, le calcul exact se fait une bonne fois pour toute et on utilise ensuite les valeurs ainsi calculées directement dans le code. Pour le premier type, à cause de la présence du coefficient k variable, il faut recalculer à chaque fois ces intégrales.

Pour cela, on peut utiliser des formules de quadrature numérique les plus précises possibles (de type méthode de Gauss). Ces formules de quadrature doivent être programmées uniquement sur l'élément de référence. Il faut alors faire attention au phénomène de sous-intégration. En effet, si on utilise l'élément \mathbb{P}_k , alors les gradients des fonctions de forme sont des polynômes de degré au plus $k - 1$ et le produit de deux telles fonctions sont des polynômes de degré au plus $2k - 2$. Si k est une fonction régulière et qu'on veut une précision suffisante, il faudra donc utiliser une formule de quadrature qui soit exacte sur les polynômes de degré au minimum $2k$ par exemple.

En pratique, les choses ne se passent pas exactement comme cela. On construit plutôt la matrice globale A en parcourant les éléments. Cela est suffisant car, comme on l'a vu toutes les intégrales en jeu sont définies au niveau de l'élément par des noeuds de cet élément. Le pseudo-algorithme pour construire la matrice de rigidité par exemple est donc le suivant

- Initialiser la matrice A à 0.
- Parcourir tous les éléments $K \in \mathcal{T}$:
 - Pour tout couple de noeuds locaux $(p_1, p_2) \in \{1, \dots, |\Sigma|\}^2$, calculer l'intégrale

$$\hat{A}_{p_1, p_2} = |\det B_K| \int_{\hat{K}} k(T_K(\hat{x})) \left((B_K^t B_K)^{-1} \nabla \theta_{p_1} \right) \cdot \nabla \theta_{p_2} d\hat{x}.$$

- Ajouter la valeur de \hat{A}_{p_1, p_2} au coefficient $A_{i, j}$ de A , où i et j sont les numéros globaux des noeuds $T_K(\hat{a}_{p_1})$ et $T_K(\hat{a}_{p_2})$, autrement dit les uniques $i, j \in \{1, \dots, N\}$ tels que $J(a_i, K) = p_1$ et $J(a_j, K) = p_2$.

N.B. : Si l'un des noeuds en question est sur le bord de Ω , alors on ne fait rien.

On procède de la même façon pour calculer le second membre du système linéaire à résoudre.

2.3 Autres éléments finis utiles

2.3.1 L'élément de Lagrange quadrangulaire \mathbb{Q}_k

Si on considère maintenant un maillage de Ω constitué de quadrangles en 2D (ou de parallépipèdes en 3D), quel type d'élément fini peut-on construire ?

On ne peut pas espérer avoir un degré de liberté sur chaque sommet du maillage en demandant à ce que les fonctions approchées soit \mathbb{P}_1 par morceaux. Il faut donc agrandir un peu l'espace de fonctions polynômes qu'on considère.

C'est la raison pour laquelle on définit l'espace \mathbb{Q}_k des fonctions polynômes dont le degré partiel par rapport à chaque variable est au plus égal à k .

$$\mathbb{Q}_k(\mathbb{R}^d) = \left\{ u = \sum_{\substack{\alpha \in \mathbb{N}^d \\ \sup \alpha_i \leq k}} a_\alpha x_1^{\alpha_1} \cdots x_d^{\alpha_d} \right\} = \mathbb{P}_k(\mathbb{R}) \otimes \cdots \otimes \mathbb{P}_k(\mathbb{R}).$$

Bien entendu, en dimension 1, on a $\mathbb{Q}_1 = \mathbb{P}_1$.

On constate qu'un élément de \mathbb{Q}_1 est défini par 4 coefficients en dimension 2 et par 8 coefficients en dimension 3. Cela correspond au nombre de sommets d'un quadrangle ou d'un parallépipède, ce qui laisse espérer que cet espace de fonctions va convenir aux maillages que l'on veut considérer (il faudra bien sûr le démontrer).

De façon, un peu plus précise on a

$$\mathbb{Q}_1(\mathbb{R}^2) = \text{Vect}(1, x_1, x_2, x_1x_2),$$

$$\mathbb{Q}_1(\mathbb{R}^3) = \text{Vect}(1, x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3).$$

On considère alors l'élément de référence $\hat{K} = [0, 1]^d$ comme étant le cube unité de \mathbb{R}^d .

Proposition et Définition II.39 (Élément fini simplicial de Lagrange \mathbb{Q}_k)

On considère l'ensemble $\Sigma \subset \hat{K}$ des points définis de la façon suivante

$$\left(\frac{i_1}{k}, \dots, \frac{i_d}{k} \right), \quad \forall (i_1, \dots, i_d) \in \mathbb{N}^d, \forall 1 \leq p \leq d, i_p \leq k.$$

Le cardinal de cet ensemble est donné par $|\Sigma| = (k+1)^d$. On note $(\hat{a}_j)_{1 \leq j \leq |\Sigma|}$ les éléments de cet ensemble.

Alors l'application $p \in \mathbb{Q}_k \mapsto (p(\hat{a}_j))_{1 \leq j \leq |\Sigma|} \in \mathbb{R}^{|\Sigma|}$ est un isomorphisme, en particulier \mathbb{P}_k et Σ ont le même cardinal.

- Le triplet $(\hat{K}, \mathbb{P}_k, \Sigma)$ est appelé : élément fini quadrangulaire de Lagrange \mathbb{Q}_k .
- L'ensemble des points $(\hat{a}_j)_j$ est l'ensemble des noeuds de cet élément fini et la forme linéaire qui à toute fonction continue v associe $v(\hat{a}_j)$ est appelée le degré de liberté associé au noeud \hat{a}_j .
- Pour tout $1 \leq i \leq |\Sigma|$, il existe une unique fonction $\theta_i \in \mathbb{Q}_k$ telle que

$$\theta_i(\hat{a}_j) = \delta_{ij}, \quad \forall 1 \leq j \leq |\Sigma|.$$

Ces fonctions sont appelées : fonctions de forme locales

La notion de régularité du maillage est légèrement modifiée par rapport aux maillages simpliciaux. En particulier, en dimension 2, la notion de rondeur d'un quadrangle K est légèrement différente. On la définit par

$$\rho_K = \min(\rho_{T_1}, \rho_{T_2}, \rho_{T_3}, \rho_{T_4}),$$

où les T_i sont les quatre triangles obtenus en joignant trois des sommets du quadrangle.

Les transformations T_K qui amènent de l'élément de référence à chaque cellule K ne sont d'ailleurs plus affines, sauf si les cellules sont des parallélogrammes. Ces transformations sont quadratiques : $T_K \in \mathbb{Q}_1^d$. En particulier leur Jacobien n'est plus une constante.

Dans ces conditions, l'espace d'approximation est défini par

$$V_h = \{v \in C^0(\bar{\Omega}), \forall K \in \mathcal{T}, v|_K \in \mathbb{Q}_k\},$$

ce qui n'est pas équivalent à demander que $v|_K$ soit dans \mathbb{Q}_k pour tout K , sauf si les cellules sont des parallélogrammes (auquel cas T_K est affine).

L'opérateur d'interpolation se définit alors comme précédemment pour l'élément \mathbb{P}_k . De plus, sous les hypothèses ainsi formulées de régularité du maillage, on obtient le résultat d'interpolation suivant :

Proposition II.40

Pour une famille de maillages quadrangles régulière, il existe une constante $C > 0$ telle que pour tout $0 \leq m \leq k$, et tout $0 \leq p \leq m$

$$\forall v \in H^{m+1}(\Omega), \quad \|v - \mathcal{I}_h^k v\|_{H^p(\Omega)} \leq C_\sigma h^{m-p+1} |v|_{H^{m+1}(\Omega)}.$$

2.3.2 Un élément conforme L^2 : l'élément totalement discontinu $\mathbb{P}_0 = \mathbb{Q}_0$

On note $\mathbb{P}_0 = \mathbb{Q}_0$ l'ensemble des fonctions constantes (qui sont bien des polynômes de degré total inférieur à 1 et de degrés partiels inférieurs à 1).

On peut alors aisément construire, à partir d'un maillage quadrangulaire ou simplicial, l'ensemble des fonctions constantes par cellules. Cet espace est bien conforme dans $L^2(\Omega)$ (mais certainement pas dans $H^1(\Omega)$). Pour les fonctions suffisamment régulières, on peut prendre comme degré de liberté dans une cellule la valeur de la fonction au centre de gravité de la cellule.

Se faisant on peut définir un opérateur d'interpolation de Lagrange \mathcal{I}_h^0 comme ci-dessus et obtenir les bonnes estimations. Néanmoins, pour cet espace d'approximation particulier, il existe un opérateur d'interpolation plus simple qui s'applique à toutes les fonctions de L^2 .

Définition II.41

Pour toute fonction $v \in L^2(\Omega)$ et tout maillage comme ci-dessus, on définit une fonction constante par morceaux $\mathcal{I}_h^{0,m} v$ constituée par la moyenne de v sur chaque cellule K .

Proposition II.42

L'opérateur $\mathcal{I}_h^{0,m}$ est l'opérateur de projection orthogonale sur l'espace d'approximation V_h . De plus, sans hypothèse de régularité du maillage, pour tout $0 \leq m \leq 1$, il existe $C > 0$ telle que

$$\forall v \in H^m(\Omega), \quad \|v - \mathcal{I}_h^{0,m} v\|_{L^2(\Omega)} \leq Ch^m |v|_{H^m(\Omega)}.$$

Cet élément est notamment utilisé pour discrétiser la pression dans le problème de Stokes.

3 Problèmes de type point-selle. Couples d'éléments inf-sup stables

On a construit dans le paragraphe précédent des espaces d'approximation de type éléments finis et on en a étudié les diverses propriétés.

Néanmoins, on a vu que pour les problèmes non coercifs de type point-selle, il est nécessaire d'assurer que les espaces d'approximation X_h et M_h soient compatibles, au sens où une condition inf-sup uniforme doit être satisfaite par les couples d'espaces.

Comme on l'a déjà vu, en dimension finie, le fait que la constante de l'inégalité inf-sup soit strictement positive est équivalent à l'injectivité de l'opérateur $B'_h : M_h \mapsto X'_h$ (qui dans le cadre de Stokes n'est autre que l'opérateur

gradient discret). Dans ces conditions, on peut définir

$$\beta_h = \inf_{q_h \in M_h} \left(\sup_{v_h \in X_h} \frac{b(v_h, q_h)}{\|v_h\|_X \|q_h\|_M} \right) > 0.$$

Il faut donc être capable de montrer l'injectivité de B'_h d'une part et le fait que β_h ne tend pas vers 0 avec h , d'autre part.

3.1 Lemme de Fortin

Bien souvent, on considère des approximations conformes d'espaces X et M qui vérifient la condition inf-sup. Dans ce cas, la validité d'une condition inf-sup discrète uniforme est donnée par le résultat suivant, dû à Fortin.

Théorème II.43 (Lemme de Fortin)

Soit $b : X \times M \mapsto \mathbb{R}$ une forme bilinéaire continue et $X_h \subset X$, $M_h \subset M$ deux sous-espaces de dimension finie. On suppose que b vérifie la condition inf-sup sur $X \times M$:

$$\exists \beta > 0, \inf_{q \in M} \sup_{v \in X} \frac{b(v, q)}{\|v\|_X \|q\|_M} \geq \beta.$$

Alors b vérifie une condition inf-sup uniforme sur $X_h \times M_h$ si et seulement si, il existe un opérateur linéaire continu $\Pi_h : X \mapsto X_h$ et une constante $C > 0$ tel que

$$\forall v \in X, \quad \|\Pi_h v\|_X \leq C \|v\|_X,$$

et

$$\forall q_h \in M_h, \quad b(v, q_h) = b(\Pi_h v, q_h).$$

Preuve :

Si un tel opérateur existe, pour tout $q_h \in M_h$, on a

$$\begin{aligned} \beta \|q_h\|_M &\leq \sup_{v \in X} \frac{b(v, q_h)}{\|v\|_X} = \sup_{v \in X} \frac{b(\Pi_h v, q_h)}{\|v\|_X} \leq \frac{1}{C} \sup_{v \in X} \frac{b(\Pi_h v, q_h)}{\|\Pi_h v\|_X} \\ &\leq \frac{1}{C} \sup_{v_h \in X_h} \frac{b(v_h, q_h)}{\|v_h\|_X}, \end{aligned}$$

ce qui démontre la condition inf-sup uniforme pour b sur $X_h \times M_h$, avec une constante βC .

Inversement, supposons la condition inf-sup uniforme est vérifiée avec une constante $\tilde{\beta}$. On sait alors que le problème

$$\begin{cases} (u_h, v_h)_X + b(v_h, p_h) = 0, & \forall v_h \in X_h, \\ b(u_h, q_h) = G(q_h), & \forall q_h \in M_h, \end{cases}$$

admet une unique solution $(u_h, q_h) \in X_h \times M_h$ qui dépend continument de la donnée de G avec l'estimation

$$\|u_h\|_X \leq \frac{2}{\tilde{\beta}} \|G\|_{X'}.$$

Soit $v \in X$, on prend pour G , la forme linéaire $G(q) = b(v, q)$, dont la norme est majorée par $\|b\| \|v\|_X$. On trouve alors un élément $u_h \in X_h$, qui dépend linéairement de v , tel que

$$\|u_h\|_X \leq \frac{2\|b\|}{\tilde{\beta}} \|v\|_X,$$

et

$$b(u_h, q_h) = b(v, q_h), \quad \forall q_h \in M_h.$$

On a donc bien construit un opérateur $\Pi_h : v \mapsto u_h$ qui vérifie les conditions du théorème. ■

3.2 Le problème de Stokes

Pour des raisons de simplicité, on va se restreindre ici au cas bidimensionnel.

3.2.1 Discrétisation $\mathbb{Q}_1/\mathbb{P}_0$

On considère un domaine $\Omega =]0, 1[^2$ carré muni d'un maillage rectangle uniforme. Si besoin les cellules seront notées $K_{i,j}$ avec $1 \leq i \leq N$ et $1 \leq j \leq M$. On suppose que $N = 4n$ et $M = 4m$ et on note $h = 1/N$ et $k = 1/M$. Ceci permet de simplifier les calculs.

On introduit l'espace d'approximation \mathbb{Q}_1 pour la vitesse défini par

$$X_h = \{v \in (H_0^1(\Omega))^2, \forall K \in \mathcal{T}, v|_K \in (\mathbb{Q}_1)^2\},$$

et l'espace d'approximation \mathbb{P}_0 pour la pression

$$M_h = \{v \in L_0^2(\Omega), \forall K \in \mathcal{T}, p|_K \in \mathbb{P}_0\}.$$

On va montrer que la condition inf-sup discrète n'est certainement pas vérifiée par ce couple d'espaces d'approximation.

Proposition II.44

Le noyau de B'_h est de dimension 1.

Preuve :

On notera $p_{i,j}$ l'inconnue en pression dans chaque maille et $(u_{i-1/2,j-1/2}, v_{i-1/2,j-1/2})$ l'inconnue en vitesse au sommet de "coordonnée" $(i-1/2, j-1/2)$.

Dans ces conditions, on peut calculer

$$b(v_h, p_h) = \int_{\Omega} (\operatorname{div} v_h) p_h \, dx = \sum_{i=1}^N \sum_{j=1}^M p_{i,j} \left(\int_{K_{i,j}} \operatorname{div} v_h \right).$$

Or sur chaque maille on a

$$\begin{aligned} \int_{K_{i,j}} \operatorname{div} v_h &= \int_{\partial K_{i,j}} v_h \cdot \nu \\ &= \frac{hk}{2} \left(\frac{v_{i-1/2,j+1/2} - v_{i-1/2,j-1/2}}{k} + \frac{v_{i+1/2,j+1/2} - v_{i+1/2,j-1/2}}{k} \right. \\ &\quad \left. + \frac{u_{i+1/2,j+1/2} - u_{i-1/2,j+1/2}}{h} + \frac{u_{i+1/2,j-1/2} - u_{i-1/2,j-1/2}}{h} \right). \end{aligned}$$

Comme v_h est nulle au bord, on peut regrouper les termes dans la somme et obtenir

$$b(v_h, p_h) = - \sum_{i,j} hk v_{i-1/2,j-1/2} (\delta_y p_h)_{i-1/2,j-1/2} - \sum_{i,j} hk u_{i-1/2,j-1/2} (\delta_x p_h)_{i-1/2,j-1/2},$$

où

$$\begin{aligned} (\delta_x p_h)_{i-1/2,j-1/2} &= \frac{1}{2} \left(\frac{p_{i,j} - p_{i-1,j}}{h} + \frac{p_{i,j-1} - p_{i-1,j-1}}{h} \right), \\ (\delta_y p_h)_{i-1/2,j-1/2} &= \frac{1}{2} \left(\frac{p_{i,j} - p_{i,j-1}}{k} + \frac{p_{i-1,j} - p_{i-1,j-1}}{k} \right). \end{aligned}$$

On en déduit que $b(v_h, p_h) = 0$ pour tout $v_h \in X_h$ si et seulement si $\delta_x p = \delta_y p = 0$ identiquement.

On voit bien que ceci est vrai si et seulement si

$$p_{i,j} = p_{i-1,j-1}, \quad \forall i, j,$$

$$p_{i-1,j} = p_{i,j-1}, \quad \forall i, j.$$

Comme il faut que $p_h \in M_h$, cela impose que la moyenne de p_h soit nulle et on voit donc que cela implique aussi que

$$p_{i,j} = -p_{i-1,j}.$$

On obtient donc un mode *parasite* très fortement oscillant, que l'on appelle le mode en damier. Ce mode engendre exactement le noyau de B'_h . Il s'écrit

$$\psi_h = \sum_i \sum_j 1_{K_{i,j}} (-1)^{i+j}.$$

Si on note ψ_h un générateur du noyau de B'_h , on peut alors envisager de remplacer l'espace M_h par l'espace \tilde{M}_h obtenu à partir de M_h en retirant l'espace engendré par ψ_h plus exactement en prenant le complémentaire orthogonal de ψ_h dans M_h .

Les espaces ainsi obtenus vérifient la condition inf-sup discrète mais malheureusement, elle n'est pas uniforme.

Théorème II.45

Pour tout h on a

$$\beta_h \stackrel{\text{def}}{=} \inf_{p_h \in \tilde{M}_h} \sup_{v_h \in X_h} \frac{b(v_h, p_h)}{\|v_h\|_{H^1} \|p_h\|_{L^2}} > 0.$$

De plus, il existe deux constantes c_1 et c_2 telles que

$$c_1 h \leq \beta_h \leq c_2 h.$$

Preuve :

Le fait que $\beta_h > 0$ vient du fait qu'on a justement choisi \tilde{M}_h pour que l'opérateur B'_h soit injectif, ce qui, en dimension finie, assure que $\beta_h > 0$.

On va maintenant seulement démontrer la partie vraiment intéressante de l'inégalité : $\beta_h \leq c_2 h$, qui montre que la constante inf-sup tend vers 0 quand $h \rightarrow 0$.

Pour cela, on va construire une fonction q_h qui va vérifier les bonnes propriétés. Plus précisément, on pose

$$q_h = \sum_{i=1}^{2n} \sum_{j=1}^{2m} 1_{K_{i,j}} (-1)^{i+j} (E((i-1)/2) - (n-1)/2).$$

On voit immédiatement que q_h est à moyenne nulle par construction. De plus, on a

$$\begin{aligned} \int_{\Omega} q_h \psi_h \, dx &= hk \sum_{i,j} (E((i-1)/2) - (n-1)/2) \\ &= hkM \sum_{i=1}^{2n} (E((i-1)/2) - (n-1)/2) = 2hkM \sum_{p=0}^{n-1} (p - (n-1)/2) = 0. \end{aligned}$$

Donc, q_h est bien dans \tilde{M}_h .

- On calcule la norme L^2 de q_h :

$$\|q_h\|_{L^2}^2 = hk \sum_{i,j} \left(E((i-1)/2) - (n-1)/2 \right)^2 = 2hkM \sum_{p=0}^{n-1} (p - (n-1)/2)^2 \sim Chn^3 \sim C'/h^2.$$

Donc $\|q_h\|_{L^2}$ est de l'ordre de $1/h$.

- Il faut maintenant évaluer le terme $b(v_h, q_h)$. Pour cela, on reprend les notations précédentes et on calcule $\delta_x q_h$ et $\delta_y q_h$. On observe tout d'abord que $\delta_x q_h = 0$. Ensuite, on a

$$\begin{aligned} (\delta_y q_h)_{i-1/2, j-1/2} &= 0, \quad \text{si } i \text{ est pair,} \\ (\delta_y q_h)_{i-1/2, j-1/2} &= 2 \frac{(-1)^{j+1}}{k}, \quad \text{si } i = 2p + 1 \text{ est impair.} \end{aligned}$$

Il vient

$$b(v_h, q_h) = 2 \sum_{p=0}^{n-1} \sum_{j=1}^{2m} h(-1)^{j+1} v_{2p+1/2, j-1/2} = 2 \sum_{p=0}^{n-1} \sum_{r=1}^m h(v_{2p+1/2, 2r-3/2} - v_{2p+1/2, 2r-1/2}).$$

D'où

$$\begin{aligned} |b(v_h, q_h)| &\leq 2h \sum_{p=0}^{n-1} \int_0^1 \left| \frac{\partial v_h}{\partial y}((2p+1)h, y) \right| dy \leq 2h \left(\int_0^1 \left(\sum_{p=0}^{n-1} \left| \frac{\partial v_h}{\partial y}((2p+1)h, y) \right| \right)^2 dy \right)^{\frac{1}{2}} \\ &\leq 2h\sqrt{n} \left(\int_0^1 \sum_{p=0}^{n-1} \left| \frac{\partial v_h}{\partial y}((2p+1)h, y) \right|^2 dy \right)^{\frac{1}{2}}. \end{aligned}$$

Comme v_h est \mathbb{Q}_1 , $x \mapsto \frac{\partial v}{\partial y}(x, y)$ est continue et affine par morceaux pour presque tout y . On utilise alors l'inégalité suivante

$$a^2 + b^2 \leq 6 \int_0^1 |ax + b(1-x)|^2 dx,$$

qui se transpose de la façon suivante sur un intervalle de longueur h

$$a^2 + b^2 \leq \frac{6}{h} \int_0^h \left| a \frac{x}{h} + b \left(1 - \frac{x}{h}\right) \right|^2 dx.$$

On a alors dans ces conditions :

$$\sum_{p=0}^{n-1} \left| \frac{\partial v_h}{\partial y}((2p+1)h, y) \right|^2 \leq \frac{6}{h} \int_0^1 \left| \frac{\partial v_h}{\partial y}(x, y) \right|^2 dx,$$

et donc on a, en intégrant par rapport à y ,

$$|b(v_h, q_h)| \leq 2\sqrt{h}\sqrt{n} \|\nabla v_h\|_{L^2}.$$

Ainsi, on a obtenu

$$\frac{|b(v_h, q_h)|}{\|v_h\|_{H^1}} \leq C.$$

Comme on a vu plus haut que la norme de q_h est en C/h , on a bien montré que $\beta_h \leq C'h$. ■

3.2.2 Discrétisation $\mathbb{P}_1/\mathbb{P}_0$

On reprend le maillage précédent mais en découpant les rectangles en deux triangles. On considère alors la discrétisation \mathbb{P}_1 de la vitesse et \mathbb{P}_0 pour la pression.

$$X_h = \{v \in (H_0^1(\Omega))^2, \forall K \in \mathcal{T}, v \circ T_K \in (\mathbb{P}_1)^2\}.$$

$$M_h = \{q \in L_0^2(\Omega), \forall K \in \mathcal{T}, q \circ T_K \in \mathbb{P}_0\}.$$

Quelques remarques :

- Il y a $2NM$ éléments et donc $2NM - 1$ degrés de liberté en pression.
- Il y a $(N+1)(M+1)$ sommets et $(N-1)(M-1)$ sommets intérieurs et donc $2(N-1)(M-1)$ degrés de liberté en vitesse.

Ainsi, le noyau de $B'_h : M_h \mapsto X'_h$ a ainsi une dimension au moins égale à

$$2NM - 1 - 2(N - 1)(M - 1) = 2(N + M) - 3.$$

Ainsi, le noyau de cet opérateur est très gros !

En fait, on peut même montrer que le noyau de B_h (qui contient les fonctions approchées à divergence discrète nulle) est réduit à 0 ! Autrement dit, la seule vitesse $v_h \in X_h$ susceptible d'être solution du système est la fonction nulle.

Pour cela, on constate que $v_h \in \text{Ker } B_h$ si $\int_{\Omega} (\text{div } v_h) q_h dx = 0$ pour tout $q_h \in M_h$. En réalité, comme v_h est nul au bord, cette condition sera aussi vérifiée par les pressions \mathbb{P}_0 dont la moyenne est non nulle. En pratique, on a donc une condition par élément qui s'écrit

$$\forall K \in \mathcal{T}, \int_K (\text{div } v_h) dx = 0.$$

Mais comme v_h est \mathbb{P}_1 sur K , $\text{div } v_h$ est une constante et donc, on obtient que $v_h \in \text{Ker } B_h$ si et seulement si $\text{div } v_h = 0$ au sens classique (ceci n'est pas vrai pour n'importe quelle discrétisation de la pression !).

Essayons maintenant de comprendre ce que cela signifie pour v_h . Soit K un triangle dont les sommets sont numérotés 1, 2 et 3 et v une fonction $(\mathbb{P}_1)^2$ sur K . Par la formule de Stokes, la condition de divergence nulle sur ce triangle, s'écrit

$$\sum_{\sigma} \int_{\sigma} (v \cdot \nu_{\sigma}) = 0.$$

Comme v est affine, l'intégrale sur l'arête est égale à la valeur de la fonction au milieu de l'arête qui est elle-même égale à la demi-somme des degrés de liberté aux sommets correspondants. Si on note m_i et ν_i la mesure et la normale sortante à l'arête opposée au sommet i , ceci s'écrit

$$m_3 \frac{v_1 + v_2}{2} \cdot \nu_3 + m_2 \frac{v_1 + v_3}{2} \cdot \nu_2 + m_1 \frac{v_2 + v_3}{2} \cdot \nu_1 = 0,$$

ou encore

$$v_1 \cdot (m_2 \nu_2 + m_3 \nu_3) + v_2 \cdot (m_1 \nu_1 + m_3 \nu_3) + v_3 \cdot (m_1 \nu_1 + m_2 \nu_2) = 0.$$

Or, on peut vérifier que dans un triangle la relation suivante est vérifiée

$$m_1 \nu_1 + m_2 \nu_2 + m_3 \nu_3 = 0,$$

ce qui ramène la condition de divergence nulle sur K à

$$m_1 v_1 \cdot \nu_1 + m_2 v_2 \cdot \nu_2 + m_3 v_3 \cdot \nu_3 = 0.$$

Ceci nous dit en particulier que si v_1 et v_2 sont nuls, alors v_3 est orthogonal à ν_3 . Ainsi si K est un triangle du bord dont l'un seulement de ces degrés de liberté est à l'intérieur, on obtient que la valeur de ce degré de liberté est orientée parallèlement au bord. Si ce même degré de liberté appartient à un autre triangle du bord, on obtient une double condition d'orthogonalité qui prouve qu'en réalité la vitesse v est nulle en ce point.

Ce raisonnement fonctionne en particulier sur le maillage triangle obtenu en découpant des rectangles en deux, ce qui prouve l'effet dit *de locking*.

3.2.3 Discrétisation $\mathbb{P}_2/\mathbb{P}_1$ (ou élément de Taylor-Hood)

On considère maintenant un maillage simplicial \mathcal{T} de Ω et la discrétisation $\mathbb{P}_2 - \mathbb{P}_1$ de ce problème.

$$X_h = \{v \in (H_0^1(\Omega))^2, \forall K \in \mathcal{T}, v \circ T_K \in (\mathbb{P}_2)^2\}.$$

$$M_h = \{q \in L_0^2(\Omega), \forall K \in \mathcal{T}, q \circ T_K \in \mathbb{P}_1\}.$$

Pour simplifier l'analyse, on va se placer en 2D.

Théorème II.46

*Si on suppose que tout élément $K \in \mathcal{T}$ a au plus une seule arête sur le bord de Ω .
Les espaces X_h et M_h ainsi définis vérifient la condition inf-sup uniforme, avec une constante qui dépend de la constante de régularité du maillage.*

Remarque : l'hypothèse géométrique sur le maillage n'est pas très restrictive, vu qu'on peut toujours s'y ramener quitte à découper en trois certains triangles.

Preuve :

Soit $q_h \in M_h$ non nul. Comme q_h est à moyenne nulle, on a vu qu'il existe $v \in (H_0^1(\Omega))^2$ tel que

$$\operatorname{div} v = q_h, \quad \|v\|_{H^1} \leq C \|q_h\|_{L^2}.$$

On prend maintenant $v_h = \mathcal{I}_h^{SZ} v$ l'interpolé de Scott-Zhang de cette fonction v . On va estimer le quotient suivant

$$\frac{b(v_h, q_h)}{\|v_h\|_{H^1}} = \frac{\int_{\Omega} (\operatorname{div} \mathcal{I}_h^{SZ} v) q_h \, dx}{\|\mathcal{I}_h^{SZ} v\|_{H^1}} \geq C \frac{\int_{\Omega} (\operatorname{div} \mathcal{I}_h^{SZ} v) q_h \, dx}{\|v\|_{H^1}} \geq C' \|q_h\|_{L^2} + C \frac{\int_{\Omega} (\operatorname{div} \mathcal{I}_h^{SZ} v - \operatorname{div} v) q_h \, dx}{\|v\|_{H^1}}.$$

Etudions ce dernier terme en intégrant par parties (on utilise que l'interpolé de Scott-Zhang préserve les conditions aux limites !)

$$\begin{aligned} \frac{\int_{\Omega} (\operatorname{div} \mathcal{I}_h^{SZ} v - \operatorname{div} v) q_h \, dx}{\|v\|_{H^1}} &= - \frac{\int_{\Omega} (\mathcal{I}_h^{SZ} v - v) \nabla q_h \, dx}{\|v\|_{H^1}} \\ &\leq \frac{\|\mathcal{I}_h^{SZ} v - v\|_{L^2} \|\nabla q_h\|_{L^2}}{\|v\|_{H^1}} \leq C \left(\sum_K h_K^2 |q_h|_{H^1(K)}^2 \right)^{\frac{1}{2}}, \end{aligned}$$

d'après les propriétés d'interpolation de l'opérateur de Scott-Zhang.

Pour chaque arête σ dans le maillage, on note τ_{σ} un vecteur unité de cette arête et $|\sigma|$ sa longueur. On construit alors l'unique élément $v_h \in X_h$ par ses valeurs dans les noeuds définies comme suit :

- $v_h(a) = 0$ si a est un sommet du maillage.
- $v_h(a) = -|\sigma|^2 \tau_{\sigma} \cdot (\nabla q_h) \cdot \tau_{\sigma}$, si a est le milieu d'une arête **interne** σ .
- $v_h(a) = 0$ si a est le milieu d'une arête du bord.

Quelques remarques s'imposent :

- Par construction, v_h est bien nulle au bord.
- La définition de $v_h(a)$, si a est le milieu d'une arête interne est cohérente. En effet, *a priori* le gradient de q_h est constant par maille et donc n'a pas de trace bien définie sur les arêtes. Or, comme q_h est bien continu à travers l'arête, le gradient tangentiel de q_h est, lui, défini de manière unique sur l'arête.

Nous utilisons maintenant la formule de quadrature suivante, valable pour les éléments de \mathbb{P}_2 :

$$\forall \pi \in \mathbb{P}_2, \forall K \in \mathcal{T}, \quad \frac{1}{|K|} \int_K \pi(x) \, dx = \sum_{a \text{ milieu de l'arête}} \frac{\pi(a)}{5} - \sum_{a \text{ sommet de } K} \frac{\pi(a)}{20}.$$

Celle-ci se démontre sur l'élément de référence puis par changement de variable sur n'importe quel élément.

On trouve ainsi

$$b(v_h, q_h) = - \int_{\Omega} v_h \nabla q_h \, dx = - \sum_K \left(\int_K v_h \nabla q_h \, dx \right) = \frac{1}{5} \sum_K |K| \left(\sum_{\substack{\sigma \in \partial K \\ \sigma \not\subset \partial \Omega}} |\sigma|^2 (\nabla q_h \cdot \tau_{\sigma})^2 \right)$$

Soit un triangle K et une fonction affine sur K , on note u_1, u_2 et u_3 les trois valeurs de u aux sommets M_1, M_2 et M_3 de K . On souhaite obtenir un contrôle du gradient de u en fonction des $u_i - u_j$. Pour cela, on constate que le vecteur $\rho_K \frac{\nabla u}{|\nabla u|}$ est de longueur ρ_K . Il existe donc deux points X et Y de K tels que

$$X - Y = \rho_K \frac{\nabla u}{|\nabla u|}.$$

On note x_i et y_i les coordonnées barycentriques (positives !) de X et Y dans le triangle K et on observe que, comme u est affine, on a

$$u(X) - u(Y) = (\nabla u) \cdot (X - Y) = \rho_K |\nabla u|,$$

mais aussi

$$u(X) - u(Y) = \sum_{i,j} x_i y_j (u(M_i) - u(M_j)) = \sum_{i,j} x_i y_j (u_i - u_j).$$

Ainsi, par Jensen,

$$\rho_K^2 |\nabla u|^2 \leq \sum_{i,j} x_i y_j |u_i - u_j|^2 \leq |u_1 - u_2|^2 + |u_1 - u_3|^2 + |u_2 - u_3|^2,$$

Comme on a par ailleurs

$$|u_1 - u_2|^2 \leq (|u_1 - u_3| + |u_2 - u_3|)^2 \leq 2|u_1 - u_3|^2 + 2|u_2 - u_3|^2,$$

on voit que l'estimation persiste (avec la constante 2 au lieu de 1) si l'on ne dispose que de deux des trois termes.

Si on revient à nos moutons, nous sommes exactement dans cette situation puisqu'on a supposé qu'au bord les triangles ne pouvaient avoir qu'une seule arête dans $\partial\Omega$.

On a donc obtenu l'estimation :

$$b(v_h, q_h) \geq C \sum_K |K| \rho_K^2 |\nabla q_h|^2 = \sum_K \rho_K^2 |q_h|_{H^1(K)}^2 \geq C_\sigma \sum_K h_K^2 |q_h|_{H^1(K)}^2,$$

dès lors que la triangulation est régulière de constante de régularité σ .

Il reste à estimer la norme de v_h dans H^1 . Pour cela, on choisit $K \in \mathcal{T}$ et φ_i une fonction de forme dont le support intersecte K . D'après le Théorème II.21, on a

$$|\varphi_i|_{H^1(K)} \leq C \frac{|K|^{\frac{1}{2}}}{\rho_K},$$

ainsi en revenant à la définition de v_h , on a

$$|v_h|_{H^1(K)}^2 \leq \sum_{a \text{ milieu d'arête}} C \frac{|K|}{\rho_K^2} |\sigma|^4 |\nabla q_h|^2 \leq C |K| \frac{h_K^2}{\rho_K^2} h_K^2 |\nabla q_h|^2,$$

de sorte que si σ est la constante de régularité du maillage

$$|v_h|_{H^1(K)}^2 \leq C_\sigma h_K^2 |q_h|_{H^1(K)}^2,$$

et donc

$$|v_h|_{H^1(\Omega)}^2 \leq C_\sigma \sum_K h_K^2 |q_h|_{H^1(K)}^2.$$

In fine, on a obtenu

$$\left(\sum_K h_K^2 |q_h|_{H^1(K)}^2 \right)^{\frac{1}{2}} \leq C \frac{b(v_h, q_h)}{\|v_h\|_{H^1}}.$$

Tout ceci mis bout à bout fournit l'inégalité inf-sup attendue. ■

Comme on connaît les résultats d'interpolation pour les espaces X_h et M_h , on déduit de ce qui précède, le résultat suivant d'estimation d'erreur.

Théorème II.47 (Estimation d'erreur pour Stokes pour l'élément de Taylor-Hood)

Soit une famille régulière $(\mathcal{T}_k)_k$ de maillages vérifiant l'hypothèse géométrique sur les arêtes du bord donnée précédemment et $X_{h_k} \times M_{h_k}$ les couples d'espaces éléments finis définis ci-dessus. On suppose que la solution (u, p) du problème de Stokes est dans $(H^{m+1}(\Omega))^d \times H^m(\Omega)$ avec $m = 1$ ou $m = 2$, alors on a l'estimation d'erreur suivante

$$\|u - u_{h_k}\|_{H^1(\Omega)} + \|p - p_{h_k}\|_{L^2(\Omega)} \leq Ch_k^m (\|u\|_{H^{m+1}(\Omega)} + \|p\|_{H^m(\Omega)}).$$

Si de plus, le problème adjoint (qui est aussi le problème de Stokes dans ce cas) possède la propriété de régularité elliptique dans Ω , alors, on a l'estimation

$$\|u - u_{h_k}\|_{L^2(\Omega)} \leq Ch_k^{m+1} (\|u\|_{H^{m+1}(\Omega)} + \|p\|_{H^m(\Omega)}).$$

3.2.4 Discrétisation \mathbb{P}_1 -iso $\mathbb{P}_2/\mathbb{P}_1$

L'inconvénient de la discrétisation $\mathbb{P}_2/\mathbb{P}_1$ (ou Taylor-Hood) est qu'elle amène à inverser des matrices avec des profils assez larges, ce qui peut présenter certaines difficultés (notamment en 3D).

L'idée est de considérer une discrétisation moins coûteuse ayant exactement les mêmes degrés de liberté que la discrétisation $\mathbb{P}_2/\mathbb{P}_1$. Ainsi, chaque triangle K (idem en 3D), est divisé en quatre triangles en utilisant les milieux des côtés de K . On obtient ainsi un nouveau maillage \mathcal{T}' raffiné une fois par rapport au précédent. On prend alors pour X_h l'ensemble des champs de vecteurs continus, \mathbb{P}_1 par morceaux sur le maillage fin \mathcal{T}' et nuls au bord et pour M_h l'ensemble des fonctions continues, \mathbb{P}_1 par morceaux sur le maillage plus grossier \mathcal{T} et à moyenne nulle.

On peut alors démontrer par une preuve semblable à la précédente, que cette discrétisation, appelée \mathbb{P}_1 -iso $\mathbb{P}_2/\mathbb{P}_2$, est inf-sup stable.

Bien entendu, si la solution attendue est très régulière, on ne peut pas profiter de l'ordre élevé de la discrétisation \mathbb{P}_2 de la vitesse. On peut également reprocher à cette méthode de ne pas être optimale pour la pression.

Chapitre III

DISCRÉTISATION VOLUMES FINIS DE PROBLÈMES ELLIPTIQUES

Remarque préliminaire. Ce chapitre est très incomplet. Le lecteur peut se référer au livre [EGH00] ou à la très abondante littérature pour compléter ses connaissances.

On pourra également consulter les documents suivants qui sont les transparents utilisés lors d'un mini-cours de 6h donné en Août 2009 dans une école d'été :

http://www.cmi.univ-mrs.fr/~fboyer/exposes/Frejus09_part1.pdf

http://www.cmi.univ-mrs.fr/~fboyer/exposes/Frejus09_part2.pdf

http://www.cmi.univ-mrs.fr/~fboyer/exposes/Frejus09_part3.pdf

Contrairement à la méthode des éléments finis qui consiste à “discrétiser” la formulation variationnelle d'un problème elliptique comme on l'a vu précédemment, la méthode des volumes finis a pour point de départ, les notions de **flux** et de **bilan de flux**.

Ainsi, de façon très générale, supposons que l'on recherche une fonction u (suffisamment régulière pour l'instant) qui satisfait une équation aux dérivées partielles sous forme divergence

$$-\operatorname{div}(\varphi(x, u, \nabla u)) = f(x), \text{ dans } \Omega,$$

où φ est une application de $\Omega \times \mathbb{R} \times \mathbb{R}^d$ dans \mathbb{R}^d suffisamment régulière.

Si \mathcal{K} est un sous-domaine ouvert borné et Lipschitzien de Ω , on peut intégrer l'équation ci-dessus sur \mathcal{K} et utiliser la formule de Stokes. Il vient

$$\int_{\mathcal{K}} f(x) dx = - \int_{\mathcal{K}} \operatorname{div}(\varphi(x, u, \nabla u)) dx = - \int_{\partial \mathcal{K}} \varphi(x, u, \nabla u) \cdot \nu_{\mathcal{K}} dx.$$

La quantité $\varphi(x, u, \nabla u) \cdot \nu_{\mathcal{K}}$ est appelé le *flux sortant de \mathcal{K} au point x* . On peut même montrer que l'EDP de départ est, en un certain sens, équivalente à ce bilan des flux sur tout $\mathcal{K} \subset \Omega$.

Supposons maintenant que \mathcal{K} soit un ouvert polygonal (en 2D) ou polyédral (en 3D). L'intégrale de bord sur $\partial \mathcal{K}$ s'écrit alors comme une somme d'intégrales sur chaque arête (resp. face) de \mathcal{K} . L'équation s'écrit alors

$$\int_{\mathcal{K}} f(x) dx = - \sum_{\sigma \in \mathcal{E}_{\mathcal{K}}} \int_{\sigma} \varphi(x, u, \nabla u) \cdot \nu_{\mathcal{K}} dx,$$

ou encore, en notant $\bar{F}_{\mathcal{K},\sigma}$ l'intégrale $-\int_{\sigma} \varphi(x, u, \nabla u) \cdot \nu_{\mathcal{K}} dx$ et $f_{\mathcal{K}}$ la moyenne de f sur \mathcal{K} , on obtient

$$|\mathcal{K}|f_{\mathcal{K}} = \sum_{\sigma \in \mathcal{E}_{\mathcal{K}}} \bar{F}_{\mathcal{K},\sigma}.$$

Supposons maintenant que l'on dispose d'une partition de Ω par un ensemble de telles mailles \mathcal{K} , le principe de la méthode des volumes finis consiste alors à écrire une approximation du flux $\bar{F}_{\mathcal{K},\sigma}$ en fonction d'un certain nombre d'inconnues (i.e. de degrés de liberté) du système.

Le choix de base pour ces inconnues est d'en prendre une pour chaque maille \mathcal{K} . Pour ce faire, on va se donner un point $x_{\mathcal{K}}$ dans chaque maille \mathcal{K} (pour l'instant où l'on veut) et on va considérer une inconnue $u_{\mathcal{K}}$ correspondant à ce point $x_{\mathcal{K}}$. Ainsi, si u est suffisamment régulière, on espère que $u_{\mathcal{K}} \sim u(x_{\mathcal{K}})$. A partir de ces inconnues à notre disposition, on essaiera de construire une approximation $F_{\mathcal{K},\sigma}$ de chacun des flux.

Une propriété essentielle des flux du problème continu est la propriété de conservativité (locale) : si \mathcal{K} et \mathcal{L} sont deux mailles qui partagent une arête $\sigma = \mathcal{K}|\mathcal{L}$ en commun, alors le flux sortant de \mathcal{K} à travers σ , i.e. $\bar{F}_{\mathcal{K},\sigma}$, est l'exact opposé du flux sortant de \mathcal{L} à travers σ , i.e. $\bar{F}_{\mathcal{L},\sigma}$. Ceci s'écrit donc

$$\bar{F}_{\mathcal{K},\sigma} = -\bar{F}_{\mathcal{L},\sigma}, \text{ si } \sigma = \mathcal{K}|\mathcal{L}.$$

En demandant à ces flux approchés de vérifier le même bilan de flux que les vrais flux du problème continu mais aussi de vérifier la propriété de conservativité locale, on va obtenir un système de N équations à N inconnues que l'on pourra résoudre puis analyser.

1 Le problème de Laplace sur maillages orthogonaux admissibles

On s'intéresse dans cette première partie au cas de l'équation de Laplace avec conditions de Dirichlet homogène sur le bord du domaine. Dans ces conditions la fonction φ qui apparaît dans le cadre général esquissé plus tôt n'est rien d'autre que $\varphi(x, u, \nabla u) = \nabla u$. Les flux exacts sont alors donnés par

$$\bar{F}_{\mathcal{K},\sigma} = -\int_{\sigma} (\nabla u) \cdot \nu_{\mathcal{K}} dx.$$

1.1 Maillages orthogonaux admissibles

Définition III.1

Soit Ω un ouvert borné polygonal connexe de \mathbb{R}^d . On se donne un maillage \mathcal{T} de Ω constitué d'une famille de sous-domaines compacts, polygonaux, convexes, non vides (resp. polyédraux) \mathcal{K} de Ω , appelés **volumes de contrôle** tels que

- Si $\mathcal{K} \neq \mathcal{L}$, on a $\overset{\circ}{\mathcal{K}} \cap \overset{\circ}{\mathcal{L}} = \emptyset$.
- $\bar{\Omega} = \bigcup_{\mathcal{K} \in \mathcal{T}} \mathcal{K}$.

On dit que ce maillage est admissible au sens des volumes finis s'il existe de plus, une famille de points $(x_{\mathcal{K}})_{\mathcal{K} \in \mathcal{T}}$ associés à tous les volumes de contrôle (qu'on appellera centres) tels que

- Pour tout $\mathcal{K} \in \mathcal{T}$, $x_{\mathcal{K}} \in \overset{\circ}{\mathcal{K}}$.
- Pour tout $\mathcal{K}, \mathcal{L} \in \mathcal{T}$, $\mathcal{K} \neq \mathcal{L}$, si $\mathcal{K} \cap \mathcal{L}$ est de dimension $d-1$, alors c'est une face de \mathcal{K} et de \mathcal{L} , notée $\mathcal{K}|\mathcal{L}$ et qui de plus, vérifie la **condition d'orthogonalité**

$$[x_{\mathcal{K}}, x_{\mathcal{L}}] \perp \sigma. \tag{III.1}$$

Notons que l'ensemble des centres $(x_{\mathcal{K}})_{\mathcal{K}}$ fait partie de la donnée du maillage même si par abus de notation on écrit souvent $\mathcal{T} = (\mathcal{K})_{\mathcal{K} \in \mathcal{T}}$.

Etant donné un maillage quelconque de Ω , l'existence de la famille de points $(x_{\mathcal{K}})_{\mathcal{K}}$ qui convienne n'est pas du tout assurée. Prenons quelques exemples :

- Si \mathcal{T} est un maillage conforme (au sens EF du terme) constitué de rectangles (ce qui restreint beaucoup la géométrie de Ω), alors on peut prendre pour $x_{\mathcal{K}}$ le centre de gravité de \mathcal{K} .
- Si \mathcal{T} est un maillage conforme de triangles, alors on peut prendre pour $x_{\mathcal{K}}$ le centre du cercle circonscrit à \mathcal{K} . Il faut quand même faire attention car dans ce cas, il est possible que $x_{\mathcal{K}}$ ne soit pas dans \mathcal{K} (ce qui n'est pas si grave) mais il se peut surtout que les centres de deux mailles voisines \mathcal{K} et \mathcal{L} ne soient pas *rangés dans le bon ordre* : on peut par exemple avoir $x_{\mathcal{K}} \in \mathcal{L}$ et $x_{\mathcal{L}} \in \mathcal{K}$ ce qui pose de nombreux problèmes !
- Une façon de construire un maillage admissible à partir des points $x_{\mathcal{K}}$ est la construction de Voronoï : Soit $(x_i)_{1 \leq i \leq N}$ est une famille de points deux à deux distincts dans Ω . Pour tout $1 \leq i \leq N$, on note

$$\mathcal{K}_i = \{x \in \bar{\Omega}, \quad |x - x_i| \leq |x - x_j|, \quad \forall j \neq i\}.$$

Alors la famille $(\mathcal{K}_i)_{1 \leq i \leq N}$ munie des centres $(x_i)_i$ associés, constitue un maillage admissible de Ω .

- Il est possible de généraliser une partie des résultats au cas de mailles non convexes mais nous ne traiterons pas ce problème ici.

Il faut maintenant introduire un certain nombre de notations :

- On note \mathcal{E}_{ext} l'ensemble des arêtes extérieures (celles incluses dans le bord de Ω) et \mathcal{E}_{int} l'ensemble des arêtes intérieures.
- On note $d_{\mathcal{K}\mathcal{L}}$ la distance entre $x_{\mathcal{K}}$ et $x_{\mathcal{L}}$, et $d_{\mathcal{K}\sigma}$ la distance entre $x_{\mathcal{K}}$ et l'arête σ .
- On note $\nu_{\mathcal{K}\mathcal{L}}$ la normale unitaire à l'arête $\mathcal{K}|\mathcal{L}$ orientée de \mathcal{K} vers \mathcal{L} .

1.2 Construction du schéma numérique

On va donc chercher une solution approchée sous la forme d'un ensemble fini de valeurs réelles notées $u^T = (u_{\mathcal{K}})_{\mathcal{K} \in \mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$. Pour déterminer cette solution u^T , il faut y associer des valeurs des flux approchés $F_{\mathcal{K},\sigma}$ pour tout \mathcal{K} et toute arête σ de \mathcal{K} . Pour ce faire, il nous faut une approximation de la dérivée normale $\nabla u \cdot \nu_{\mathcal{K}}$ de la solution sur l'arête σ . Deux cas sont à considérer :

- Si $\sigma \in \mathcal{E}_{ext}$, alors, on sait que la solution exacte recherchée est nulle sur le bord de Ω , elle est donc nulle sur le projeté orthogonal x_{σ} de $x_{\mathcal{K}}$ sur (la droite qui contient) σ . Comme on dispose par ailleurs d'une valeur approchée de u en $x_{\mathcal{K}}$, on en déduit qu'une approximation raisonnable de $\nabla u \cdot \nu_{\mathcal{K}}$ sur σ est donnée par

$$(\nabla u) \cdot \nu_{\mathcal{K}} \sim \frac{u(x_{\sigma}) - u(x_{\mathcal{K}})}{d_{\mathcal{K}\sigma}} = \frac{-u(x_{\mathcal{K}})}{d_{\mathcal{K}\sigma}},$$

de sorte que l'on a envie de considérer la définition suivante du flux approché

$$F_{\mathcal{K},\sigma}(u^T) = -|\sigma| \frac{0 - u_{\mathcal{K}}}{d_{\mathcal{K}\sigma}}. \quad (\text{III.2})$$

- Si $\sigma \in \mathcal{E}_{int}$, alors il existe un volume de contrôle \mathcal{L} tel que $\sigma = \mathcal{K}|\mathcal{L}$. Dans ces conditions, on a, par définition, l'orthogonalité du segment $[x_{\mathcal{K}}x_{\mathcal{L}}]$ avec l'arête σ , de sorte que

$$x_{\mathcal{L}} - x_{\mathcal{K}} = d_{\mathcal{K}\mathcal{L}} \nu_{\mathcal{K}\mathcal{L}},$$

ainsi, si u est suffisamment régulière, on a sur l'arête σ

$$(\nabla u) \cdot \nu_{\mathcal{K}\mathcal{L}} \sim \frac{u(x_{\mathcal{L}}) - u(x_{\mathcal{K}})}{d_{\mathcal{K}\mathcal{L}}}.$$

On s'inspire de cette formule pour définir un flux numérique

$$F_{\mathcal{K},\sigma}(u^T) = -|\sigma| \frac{u_{\mathcal{L}} - u_{\mathcal{K}}}{d_{\mathcal{K}\mathcal{L}}}. \quad (\text{III.3})$$

Remarquons que cette construction assure automatiquement la conservativité locale du flux numérique :

$$\forall u^T \in \mathbb{R}^T, \quad F_{\kappa,\sigma}(u^T) = -F_{\mathcal{L},\sigma}(u^T), \quad (\text{III.4})$$

cette valeur commune sera d'ailleurs notée $F_{\kappa,\mathcal{L}}(u^T)$ qui est donc aussi égale à $F_{\mathcal{L},\kappa}(u^T)$.

On peut maintenant construire le schéma volumes finis correspondant aux définitions ci-dessus : on cherche $u^T \in \mathbb{R}^T$ tel que

$$\forall \mathcal{K} \in \mathcal{T}, \quad |\mathcal{K}|f_{\mathcal{K}} = \sum_{\sigma \in \mathcal{E}_{\mathcal{K}}} F_{\kappa,\sigma}(u^T), \quad (\text{III.5})$$

où les flux numériques sont donnés par (III.2) et (III.3).

On obtient donc une famille de N équations linéaires (car les flux numériques sont linéaires en u^T) à N inconnues, où N est le nombre de volumes de contrôle dans le maillage.

Ce schéma est parfois appelé : TPFA (pour Two Point Flux Approximation) ou VF4 (car la largeur du stencil du schéma est 4 sur les maillages triangles en 2D).

1.3 Premières propriétés : existence et unicité de la solution - stabilité

Toute l'analyse des méthodes volumes finis reposent sur un calcul assez simple qui revient essentiellement à effectuer une intégration par parties discrète dans le schéma. On obtient ainsi une sorte de formulation variationnelle non conforme du schéma.

Notations :

- Pour toute arête $\sigma \in \mathcal{E}$, on note $F_{\sigma}(u^T)$ l'un quelconque des (au plus deux) flux numériques définis à travers σ . Autrement dit :

- Si $\sigma \in \mathcal{E}_{ext}$, on pose :

$$F_{\sigma} = F_{\kappa,\sigma},$$

où $\mathcal{K} \in \mathcal{T}$ est l'unique volume de contrôle dans \mathcal{T} dont σ est une arête.

- Si $\sigma \in \mathcal{E}_{int}$, on pose :

$$F_{\sigma} = F_{\kappa,\sigma},$$

où $\mathcal{K} \in \mathcal{T}$ est l'un des deux volumes de contrôle dans \mathcal{T} dont σ une arête. Ce choix est fait de façon arbitraire une bonne fois pour toute et ne changera plus dans toute la suite. Tous les résultats énoncés sont bien entendu indépendants de ce choix, à cause de la propriété de conservativité locale.

- Pour tout couple de volumes de contrôle voisins \mathcal{K}, \mathcal{L} , on notera

$$D_{\mathcal{K}\mathcal{L}}(u^T) = \frac{u_{\mathcal{L}} - u_{\mathcal{K}}}{d_{\mathcal{K}\mathcal{L}}},$$

le quotient différentiel qui intervient dans la définition du flux numérique. On utilise les notations formellement semblables pour $D_{\kappa\sigma}(u^T)$ et $D_{\sigma}(u^T)$, etc ...

Lemme III.2

Soit $u^T \in \mathbb{R}^T$, une solution (si elle existe !) du schéma VF4 (i.e. vérifiant (III.5), (III.2) et (III.3)). Alors pour tout $v^T \in \mathbb{R}^T$, on a

$$\sum_{\sigma \in \mathcal{E}} d_{\mathcal{K}\mathcal{L}}|\sigma|D_{\sigma}(u^T)D_{\sigma}(v^T) = \sum_{\kappa \in \mathcal{T}} |\mathcal{K}|v_{\kappa}f_{\kappa},$$

ce qui s'écrit encore

$$[u^T, v^T]_{1,\mathcal{T}} \stackrel{\text{def}}{=} \sum_{\sigma=\mathcal{K}|\mathcal{L} \in \mathcal{E}_{int}} d_{\mathcal{K}\mathcal{L}}|\sigma| \frac{u_{\mathcal{L}} - u_{\mathcal{K}}}{d_{\mathcal{K}\mathcal{L}}} \frac{v_{\mathcal{L}} - v_{\mathcal{K}}}{d_{\mathcal{K}\mathcal{L}}} + \sum_{\sigma \in \mathcal{E}_{ext}} d_{\mathcal{K}\sigma}|\sigma| \frac{u_{\mathcal{K}} - 0}{d_{\mathcal{K}\sigma}} \frac{v_{\mathcal{K}} - 0}{d_{\mathcal{K}\sigma}} = \sum_{\kappa \in \mathcal{T}} |\mathcal{K}|v_{\kappa}f_{\kappa}.$$

Preuve :

On multiplie l'équation du schéma correspondant au volume de contrôle \mathcal{K} par $v_{\mathcal{K}}$, puis on somme sur tous les volumes de contrôle et on trouve

$$T \stackrel{\text{def}}{=} \sum_{\mathcal{K} \in \mathcal{T}} |\mathcal{K}| f_{\mathcal{K}} v_{\mathcal{K}} = \sum_{\mathcal{K} \in \mathcal{T}} v_{\mathcal{K}} \left(\sum_{\sigma \in \mathcal{E}_{\mathcal{K}}} F_{\mathcal{K},\sigma}(u^T) \right) = \sum_{\mathcal{K} \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_{\mathcal{K}}} v_{\mathcal{K}} F_{\mathcal{K},\sigma}(u^T).$$

Dans cette dernière somme, les arêtes du bord apparaissent une fois et une seule, tandis que les arêtes intérieures apparaissent exactement deux fois (une fois pour chaque volume de contrôle qui la contient). On trouve donc :

$$T = \sum_{\sigma \in \mathcal{E}_{ext}} v_{\mathcal{K}} |\sigma| \frac{u_{\mathcal{K}}}{d_{\mathcal{K}\sigma}} + \sum_{\sigma = \mathcal{K} | \mathcal{L} \in \mathcal{E}_{int}} (v_{\mathcal{K}} F_{\mathcal{K},\sigma}(u^T) + v_{\mathcal{L}} F_{\mathcal{L},\sigma}(u^T)).$$

A ce stade, on utilise la conservativité du schéma, de sorte que les deux flux qui apparaissent dans les termes de la deuxième somme sont exactement opposés. On peut donc écrire

$$\begin{aligned} T &= \sum_{\sigma \in \mathcal{E}_{ext}} v_{\mathcal{K}} |\sigma| \frac{u_{\mathcal{K}}}{d_{\mathcal{K}\sigma}} + \sum_{\sigma = \mathcal{K} | \mathcal{L} \in \mathcal{E}_{int}} (v_{\mathcal{K}} - v_{\mathcal{L}}) F_{\mathcal{K},\sigma}(u^T) \\ &= \sum_{\sigma \in \mathcal{E}_{ext}} d_{\mathcal{K}\sigma} |\sigma| \frac{u_{\mathcal{K}}}{d_{\mathcal{K}\sigma}} \frac{v_{\mathcal{K}}}{d_{\mathcal{K}\sigma}} + \sum_{\sigma = \mathcal{K} | \mathcal{L} \in \mathcal{E}_{int}} \frac{d_{\mathcal{K}\mathcal{L}}}{|\sigma|} F_{\mathcal{K},\sigma}(v^T) F_{\mathcal{K},\sigma}(u^T) \end{aligned}$$

■

Proposition et Définition III.3

La forme bilinéaire

$$(u^T, v^T) \in \mathbb{R}^T \times \mathbb{R}^T \mapsto [u^T, v^T]_{1,T},$$

*est symétrique définie positive. C'est donc un produit scalaire sur \mathbb{R}^T appelé le **produit scalaire H_0^1 discret**.*

*La norme associée à ce produit scalaire est notée $\|u^T\|_{1,T}$ et appelée la **norme H_0^1 discrète**.*

Preuve :

La seule chose à démontrer, c'est le caractère défini du produit scalaire. Pour cela, on constate que si $u^T \in \mathbb{R}^T$ est tel que $[u^T, u^T]_{1,T} = 0$ alors cela signifie que $u_{\mathcal{K}}$ est nul pour tout volume de contrôle \mathcal{K} situé au bord du domaine (c'est-à-dire qui a une arête entièrement incluse dans $\partial\Omega$) et de plus pour tout couple de volumes \mathcal{K}, \mathcal{L} qui partagent une arête, $u_{\mathcal{K}} = u_{\mathcal{L}}$.

Ceci montre bien (par exemple par récurrence sur le nombre de cellules intérieures, en enlevant couches par couches les cellules du bord) que tous les $u_{\mathcal{K}}$ sont nuls. ■

Théorème III.4

Pour toute donnée $f \in L^2(\Omega)$, le schéma VF4 admet une unique solution u^T et de plus, nous avons

$$\|u\|_{1,T}^2 \leq \|u^T\|_{L^2} \|f^T\|_{L^2} \leq \|u^T\|_{L^2} \|f\|_{L^2}.$$

Preuve :

D'après le résultat précédent, en prenant $v^T = u^T$, et en utilisant l'inégalité de Cauchy-Schwarz l'estimation attendue. Celle-ci prouve que si $f^T = 0$, alors $u^T = 0$ est la seule solution du système linéaire étudié. Ce système étant carré, on a bien montré l'existence et l'unicité de la solution pour toute donnée.

L'inégalité $\|f^T\|_{L^2} \leq \|f\|_{L^2}$ provient, elle, de l'inégalité de Jensen. ■

A priori, il semble que le schéma numérique puisse nous fournir une information sur la norme H_0^1 discrète de la solution (ce qui n'est pas très surprenant !). Pour ce faire, il faut quand même majorer la norme L^2 par la norme H_0^1 discrète. Toutes les normes étant équivalentes en dimension finie, ceci est bien sûr possible mais, pour que le résultat soit exploitable, il faut pouvoir estimer la constante.

Pour cela, on a besoin d'une inégalité de Poincaré discrète qui permet justement de répondre à cette question.

Théorème III.5 (Inégalité de Poincaré discrète)

Pour tout maillage admissible \mathcal{T} , on a

$$\forall u^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}, \quad \|u^{\mathcal{T}}\|_{L^2} \leq \text{diam}(\Omega) \|u^{\mathcal{T}}\|_{1, \mathcal{T}}.$$

Preuve :

Pour toute arête $\sigma \in \mathcal{E}$, on définit la fonction $\chi_{\sigma}(x, y)$ qui vaut 1 si $[x, y] \cap \sigma \neq \emptyset$ et 0 sinon.

Soit ξ un vecteur unitaire de \mathbb{R}^d . Pour tout $x \in \mathcal{K}$, on considère $y(x)$ la “projection” de x sur $\partial\Omega$ dans la direction ξ , c’est-à-dire le premier point de la demi-droite $x + \mathbb{R}^+\xi$ qui rencontre le bord.

On considère maintenant l’ensemble $\tilde{\Omega}$ des $x \in \Omega$, tels que $[x, y(x)]$ ne rencontre aucun sommet du maillage et ne contienne aucune arête du maillage. Comme il y a un nombre fini de sommets et d’arêtes, le complémentaire de cet ensemble est une union finie d’ensembles de mesure nulle, c’est donc un ensemble de mesure nulle.

Soit maintenant $\mathcal{K} \in \mathcal{T}$ et $x \in \mathcal{K} \cap \tilde{\Omega}$. En suivant le segment $[x, y(x)]$ en partant de x , on va successivement croiser des volumes de contrôle noté $(\mathcal{K}_i)_{1 \leq i \leq m}$ avec $\mathcal{K}_1 = \mathcal{K}$ et \mathcal{K}_m est un volume de contrôle du bord tel que $y(x)$ est dans une arête du bord de Ω . De plus, par construction de $\tilde{\Omega}$, on croise ces volumes de contrôle de façon non tangentielle. On écrit alors

$$u_{\mathcal{K}} = u_{\mathcal{K}_1} = \sum_{i=1}^{m-1} (u_{\mathcal{K}_i} - u_{\mathcal{K}_{i+1}}) + u_{\mathcal{K}_m}.$$

Ainsi, on a

$$|u_{\mathcal{K}}| \leq \sum_{i=1}^{m-1} |u_{\mathcal{K}_i} - u_{\mathcal{K}_{i+1}}| + |u_{\mathcal{K}_m} - 0|.$$

Comme l’interface $\sigma = \mathcal{K}_i | \mathcal{K}_{i+1}$ intervient dans cette somme si et seulement si $\chi_{\sigma}(x, y(x)) = 1$, on obtient

$$|u_{\mathcal{K}}| \leq \sum_{\sigma \in \mathcal{E}} d_{\sigma} |D_{\sigma}(u^{\mathcal{T}})| \chi_{\sigma}(x, y(x)).$$

On note c_{σ} le produit scalaire $c_{\sigma} = |\nu \cdot \xi|$, où ν est la normale à l’arête σ . Comme $x \in \tilde{\Omega}$, toute arête σ telle que $\chi_{\sigma}(x, y(x)) = 1$ est telle que $c_{\sigma} > 0$, on peut donc écrire

$$|u_{\mathcal{K}}| \leq \sum_{\sigma \in \mathcal{E}} d_{\sigma} \frac{1}{\sqrt{c_{\sigma}}} |D_{\sigma}(u^{\mathcal{T}})| \sqrt{c_{\sigma}} \chi_{\sigma}^2(x, y(x)).$$

En utilisant l’inégalité de Cauchy-Schwarz

$$|u_{\mathcal{K}}|^2 \leq \left(\sum_{\sigma \in \mathcal{E}} \frac{d_{\sigma}}{c_{\sigma}} |D_{\sigma}(u^{\mathcal{T}})|^2 \chi_{\sigma}(x, y(x)) \right) \left(\sum_{\sigma \in \mathcal{E}} d_{\sigma} c_{\sigma} \chi_{\sigma}(x, y(x)) \right).$$

Evaluons le second terme pour tout $x \in \tilde{\Omega}$. On note $\tilde{\sigma}$ l’unique arête du bord qui contient $y(x)$ et $x_{\tilde{\sigma}}$ son *centre* (i.e. la projection orthogonale du centre de la maille sur $\tilde{\sigma}$). Par construction, nous avons

$$x_{\mathcal{K}} - x_{\tilde{\sigma}} = \sum_{i=1}^{m-1} (x_{\mathcal{K}_i} - x_{\mathcal{K}_{i+1}}) + x_{\mathcal{K}_m} - x_{\tilde{\sigma}},$$

de sorte que

$$(x_{\mathcal{K}} - x_{\tilde{\sigma}}) \cdot \xi = \sum_{i=1}^{m-1} \underbrace{(x_{\mathcal{K}_i} - x_{\mathcal{K}_{i+1}})}_{=d_{\mathcal{K}_i, \mathcal{K}_{i+1}} \nu_{\mathcal{K}_i, \mathcal{K}_{i+1}}} \cdot \xi + d_{\mathcal{K}_m, \tilde{\sigma}} \nu_{\mathcal{K}_m, \tilde{\sigma}} \cdot \xi,$$

et comme toutes les normales sont orientées dans le même sens, nous avons donc

$$\sum_{\sigma \in \mathcal{E}} d_{\sigma} c_{\sigma} \chi_{\sigma}(x, y(x)) = |(x_{\mathcal{K}} - x_{\tilde{\sigma}}) \cdot \xi| \leq \text{diam}(\Omega).$$

In fine, on a obtenu

$$|u_\kappa|^2 \leq \text{diam}(\Omega) \left(\sum_{\sigma \in \mathcal{E}} \frac{d_\sigma}{c_\sigma} |D_\sigma(u^\mathcal{T})|^2 \chi_\sigma(x, y(x)) \right),$$

Si on intègre sur Ω , on trouve

$$\sum_\kappa |\mathcal{K}| |u_\kappa|^2 \leq \text{diam}(\Omega) \sum_{\sigma \in \mathcal{E}} \frac{d_\sigma}{c_\sigma} |D_\sigma(u^\mathcal{T})|^2 \left(\int_\Omega \chi_\sigma(x, y(x)) dx \right).$$

Il reste à évaluer les intégrales des fonctions χ_σ . Pour cela on constate que si σ est fixée, $\chi_\sigma(x, y(x))$ ne peut valoir 1 que dans une bande de largeur $|\sigma|c_\sigma$ de sorte que

$$\int_\Omega \chi_\sigma(x, y(x)) dx \leq \text{diam}(\Omega) |\sigma| c_\sigma.$$

Mettant tout cela bout à bout, on trouve

$$\|u^\mathcal{T}\|_{L^2}^2 \leq \text{diam}(\Omega)^2 \sum_{\sigma \in \mathcal{E}} d_\sigma |\sigma| |D_\sigma(u^\mathcal{T})|^2 = \text{diam}(\Omega)^2 \|u^\mathcal{T}\|_{1, \mathcal{T}}^2.$$

■

1.4 Compacité - Convergence

Pour tout maillage admissible \mathcal{T} , on a donc une solution $u^\mathcal{T}$ du schéma de volumes finis. De plus, la semi-norme H_0^1 discrète de cette solution reste bornée quand on raffine le maillage.

Etant donnée une famille $(\mathcal{T}_n)_n$ de maillages dont le pas $\text{size}(\mathcal{T}_n)$ tend vers 0 quand n tend vers l'infini, on peut donc se demander si la suite de fonctions $(u^{\mathcal{T}_n})_n$ converge, en un certain sens, vers la solution du problème continu.

Pour cela, on va commencer par définir un gradient discret pour les fonctions (constantes par morceaux). Pour tout $u^\mathcal{T} \in \mathbb{R}^\mathcal{T}$, on note ainsi $\nabla^\mathcal{T} u^\mathcal{T}$ la fonction vectorielle constante par diamant, dont la valeur sur tout diamant \mathcal{D} est notée $\nabla_{\mathcal{D}}^\mathcal{T} u^\mathcal{T}$ et définie par

$$\nabla_{\mathcal{D}}^\mathcal{T} u^\mathcal{T} = \begin{cases} d \frac{u_{\kappa_\mathcal{L}} - u_{\kappa_\mathcal{R}}}{d_{\kappa_\mathcal{L}\kappa_\mathcal{R}}} \nu_{\kappa_\mathcal{L}\kappa_\mathcal{R}} = d D_\sigma(u^\mathcal{T}) \nu_\sigma, & \text{si } \mathcal{D} \text{ est un diamant intérieur } \mathcal{D}_{\kappa_\mathcal{L}\kappa_\mathcal{R}}, \\ d \frac{0 - u_\kappa}{d_{\kappa\sigma}} \nu_{\kappa\sigma} = d D_\sigma(u^\mathcal{T}) \nu_\sigma, & \text{si } \mathcal{D} \text{ est un diamant du bord } \mathcal{D}_\sigma. \end{cases}$$

ATTENTION : Noter la présence du facteur d (la dimension de l'espace) dans la formule. On verra plus loin pourquoi ce facteur doit effectivement être présent.

On peut maintenant démontrer un résultat de compacité.

Théorème III.6 (Compacité)

Soit $(\mathcal{T}_n)_n$ une suite de maillages admissibles dont le pas tend vers 0 et $(u^{\mathcal{T}_n})_n$ une famille de fonctions discrètes définies sur chacun des maillages telle que

$$\sup_n \|u^{\mathcal{T}_n}\|_{1, \mathcal{T}_n} < +\infty. \quad (\text{III.6})$$

Alors, on a les propriétés suivantes :

- Il existe une fonction $u \in L^2(\Omega)$, et une sous-suite $(u^{\mathcal{T}_{\varphi(n)}})_n$ qui converge fortement vers u dans $L^2(\Omega)$.
- La fonction u est dans $H_0^1(\Omega)$.
- La suite des gradients discrets $(\nabla_{\mathcal{T}_{\varphi(n)}}^\mathcal{T} u^{\mathcal{T}_{\varphi(n)}})_n$ converge faiblement vers ∇u dans $(L^2(\Omega))^d$.

La preuve du théorème ci-dessus repose sur le critère de compacité de Kolmogorov (appelé aussi Riesz-Fréchet-Kolmogoroff)

Théorème III.7 (Kolmogorov)

Soit \mathcal{F} une partie **bornée** de $L^p(\Omega)$, avec $1 \leq p < +\infty$. On suppose que

1. Pour tout $\varepsilon > 0$, et pour tout ouvert ω borné, il existe $\delta > 0$, tel que

$$\forall f \in \mathcal{F}, \forall \eta \in \mathbb{R}^d, |\eta| \leq \delta \implies \|\tau_\eta f - f\|_{L^p(\omega)} \leq \varepsilon.$$

2. Pour tout $\varepsilon > 0$, il existe un ouvert borné ω , tel que

$$\forall f \in \mathcal{F}, \|f\|_{L^p(\mathbb{R}^d \setminus \omega)} \leq \varepsilon.$$

Alors \mathcal{F} est relativement compact dans $L^p(\mathbb{R}^d)$.

Dans ce théorème $\tau_\eta f$ désigne la fonction “translatée” définie par

$$\tau_\eta f(x) = f(x + \eta), \quad \forall x \in \mathbb{R}^d.$$

Ce résultat est essentiellement une conséquence, ou une généralisation, du théorème d’Ascoli. On admet donc ce résultat classique et on passe à la démonstration du théorème de compacité pour les éléments de \mathbb{R}^T .

Preuve (du théorème III.6):

- On va tout d’abord démontrer la convergence forte dans $L^2(\Omega)$ à sous-suite près. Pour cela, pour tout $n \geq 0$, on note \tilde{u}_n le prolongement par 0 de u^{T_n} à \mathbb{R}^d tout entier.

D’après l’inégalité de Poincaré et la borne (III.6), la suite u^{T_n} est bornée dans $L^2(\Omega)$ et donc la suite $(u_n)_n$ est bornée dans $L^2(\mathbb{R}^d)$. On va donc essayer de vérifier que $\mathcal{F} = \{u_n, n \in \mathbb{N}\}$ vérifie les hypothèses du théorème de Kolmogorov.

La seconde hypothèse est triviale car tous les u_n sont nuls en dehors de Ω qui est un ouvert borné. Il ne reste donc qu’à démontrer l’estimation sur les translations.

Soit donc $\eta \in \mathbb{R}^d$ et notons $\tau_\eta u_n$ la fonction translatée de u_n . On introduit à nouveau la fonction $\chi_\sigma(x, y)$ qui vaut 1 si et seulement si le segment $[x, y]$ rencontre l’arête σ , puis on note $c_\sigma = |\nu \cdot \frac{\eta}{|\eta|}|$.

Par l’argument de sommes télescopiques, on trouve, pour presque tout $x \in \mathbb{R}^d$:

$$\begin{aligned} |u_n(x + \eta) - u_n(x)| &\leq \sum_{\sigma \in \mathcal{E}_{int}} \chi_\sigma(x, x + \eta) |u_\kappa^n - u_\zeta^n| + \sum_{\sigma \in \mathcal{E}_{ext}} \chi_\sigma(x, x + \eta) |u_\kappa^n| \\ &\leq \sum_{\sigma \in \mathcal{E}} \chi_\sigma(x, x + \eta) d_\sigma |D_\sigma(u^{T_n})|. \end{aligned}$$

Par l’inégalité de Cauchy-Schwarz, on trouve

$$|u_n(x + \eta) - u_n(x)|^2 \leq \left(\sum_{\sigma \in \mathcal{E}} \chi_\sigma(x, x + \eta) \frac{d_\sigma}{c_\sigma} |D_\sigma(u^{T_n})|^2 \right) \left(\sum_{\sigma \in \mathcal{E}} \chi_\sigma(x, x + \eta) d_\sigma c_\sigma \right).$$

On montre maintenant que, pour presque tout $x \in \Omega$, on a

$$\left(\sum_{\sigma \in \mathcal{E}} \chi_\sigma(x, x + \eta) d_\sigma c_\sigma \right) \leq C(|\eta| + \text{size}(\mathcal{T}_n)).$$

Si on suppose que Ω est convexe (sinon on peut généraliser la preuve), on voit tout d’abord qu’on peut se ramener au cas où $[x, x + \eta] \subset \bar{\Omega}$ en remplaçant le segment $[x, x + \eta]$ en le plus grand segment $[y, z] \subset$

$[x, x + \eta] \cap \bar{\Omega}$. Ensuite, pour presque tout x , on peut trouver \mathcal{K} et \mathcal{L} dans le maillage tel que $x \in \mathcal{K}$ et $x + \eta \in \mathcal{L}$. Dans ces conditions on a

$$\left(\sum_{\sigma \in \mathcal{E}} \chi_{\sigma}(x, x + \eta) d_{\sigma} c_{\sigma} \right) = |(x_{\mathcal{L}} - x_{\mathcal{K}}) \cdot \frac{\eta}{|\eta|}| \leq |x_{\mathcal{K}} - x_{\mathcal{L}}| \leq |x - (x + \eta)| + |x - x_{\mathcal{K}}| + |x + \eta - x_{\mathcal{L}}| \leq |\eta| + 2\text{size}(\mathcal{T}_n).$$

De plus, comme on l'a déjà vu on a

$$\int_{\mathbb{R}^d} \chi_{\sigma}(x, x + \eta) dx \leq |\eta| |\sigma| c_{\sigma}.$$

Ainsi, en intégrant la première inégalité obtenue, on a

$$\|\tau_{\eta} u_n - u_n\|_{L^2}^2 \leq |\eta| (|\eta| + 2\text{size}(\mathcal{T}_n)) \left(\sum_{\sigma \in \mathcal{E}} d_{\sigma} |\sigma| |D_{\sigma}(u^{\mathcal{T}_n})|^2 \right) \leq C |\eta| (|\eta| + \text{size}(\mathcal{T}_n)) \|u^{\mathcal{T}_n}\|_{1, \mathcal{T}_n}^2.$$

On obtient bien une estimation sur les translattées de u_n qui permet de valider la première hypothèse du théorème de Kolmogorov.

Ainsi, on peut trouver une sous-suite (que l'on notera encore $(u^{\mathcal{T}_n})_n$) et une fonction $u \in L^2(\Omega)$ telle que $u^{\mathcal{T}_n}$ converge vers u **fortement** dans L^2 . On peut par ailleurs supposer que $(\nabla^{\mathcal{T}_n} u^{\mathcal{T}_n})_n$ converge faiblement vers un certain G dans $(L^2(\Omega))^d$.

- Il faut maintenant démontrer que la fonction u obtenue ci-dessus est dans $H_0^1(\Omega)$ et que son gradient est précisément G . Pour cela on se donne un champ de vecteurs Φ de classe \mathcal{C}^{∞} sur \mathbb{R}^d tout entier et on considère l'intégrale

$$I_n = \int_{\Omega} u^{\mathcal{T}_n} (\text{div} \Phi) dx.$$

Par convergence de $u^{\mathcal{T}_n}$ vers u dans L^2 , cette intégrale converge, quand $n \rightarrow \infty$, vers $\int_{\Omega} u (\text{div} \Phi) dx$. Essayons maintenant de calculer cette intégrale d'une autre manière, en utilisant la définition de $u^{\mathcal{T}_n}$:

$$I_n = \sum_{\mathcal{K} \in \mathcal{T}_n} u_{\mathcal{K}}^n \left(\int_{\mathcal{K}} \text{div} \Phi dx \right) = \sum_{\mathcal{K} \in \mathcal{T}_n} u_{\mathcal{K}}^n \sum_{\sigma \in \mathcal{E}_{\mathcal{K}}} \left(\int_{\sigma} \Phi \cdot \nu_{\mathcal{K}\sigma} dx \right).$$

Pour une arête intérieure $\sigma = \mathcal{K}|\mathcal{L}$, on a $\nu_{\mathcal{K}\sigma} = -\nu_{\mathcal{L}\sigma}$ et donc, on peut réécrire cette somme comme une somme sur les arêtes.

$$\begin{aligned} I_n &= \sum_{\sigma \in \mathcal{E}_{int}} (u_{\mathcal{K}}^n - u_{\mathcal{L}}^n) \left(\int_{\sigma} \Phi \cdot \nu_{\mathcal{K}\mathcal{L}} \right) + \sum_{\sigma \in \mathcal{E}_{ext}} u_{\mathcal{K}}^n \left(\int_{\sigma} \Phi \cdot \nu_{\mathcal{K}\sigma} \right) \\ &= \sum_{\sigma \in \mathcal{E}_{int}} (u_{\mathcal{K}}^n - u_{\mathcal{L}}^n) \nu_{\mathcal{K}\mathcal{L}} \cdot \left(\int_{\sigma} \Phi \right) + \sum_{\sigma \in \mathcal{E}_{ext}} u_{\mathcal{K}}^n \nu_{\mathcal{K}\sigma} \cdot \left(\int_{\sigma} \Phi \right) \\ &= - \sum_{\sigma \in \mathcal{E}_{int}} \frac{|\sigma| d_{\mathcal{K}\mathcal{L}}}{d} \left(d \frac{u_{\mathcal{K}}^n - u_{\mathcal{L}}^n}{d_{\mathcal{K}\mathcal{L}}} \nu_{\mathcal{K}\mathcal{L}} \right) \cdot \frac{1}{|\sigma|} \left(\int_{\sigma} \Phi \right) + \sum_{\sigma \in \mathcal{E}_{ext}} \frac{|\sigma| d_{\mathcal{K}\sigma}}{d} \left(d \frac{u_{\mathcal{K}}^n}{d_{\mathcal{K}\sigma}} \nu_{\mathcal{K}\sigma} \right) \cdot \frac{1}{|\sigma|} \left(\int_{\sigma} \Phi \right) \\ &= - \sum_{\mathcal{D} \in \mathfrak{D}} |\mathcal{D}| \nabla_{\mathcal{D}}^{\mathcal{T}_n} u^{\mathcal{T}_n} \cdot \left(\frac{1}{|\sigma|} \int_{\sigma} \Phi \right). \end{aligned}$$

Or, comme Φ est de classe \mathcal{C}^{∞} , on a sur chaque diamant \mathcal{D} et arête σ correspondante :

$$\left| \frac{1}{|\sigma|} \int_{\sigma} \Phi - \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} \Phi \right| \leq \|\nabla \Phi\|_{\infty} \text{size}(\mathcal{T}_n).$$

Donc, comme la suite $\nabla^{\mathcal{T}_n} u^{\mathcal{T}_n}$ est bornée dans L^2 , on a

$$I_n = - \sum_{\mathcal{D} \in \mathfrak{D}} |\mathcal{D}| \nabla_{\mathcal{D}}^{\mathcal{T}_n} u^{\mathcal{T}_n} \cdot \left(\frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} \Phi \right) + O(\text{size}(\mathcal{T}_n)),$$

ce qui s'écrit aussi

$$I_n = - \sum_{\mathcal{D} \in \mathfrak{D}} (\nabla_{\mathcal{D}}^{\mathcal{T}_n} u^{\mathcal{T}_n}) \cdot \left(\int_{\mathcal{D}} \Phi \right) + O(\text{size}(\mathcal{T}_n)),$$

et donc

$$I_n = - \int_{\Omega} \nabla^{\mathcal{T}_n} u^{\mathcal{T}_n} \cdot \Phi \, dx + O(\text{size}(\mathcal{T}_n)).$$

Par convergence faible de $\nabla^{\mathcal{T}_n} u^{\mathcal{T}_n}$ vers G , on en déduit que

$$I_n \xrightarrow{n \rightarrow \infty} - \int_{\Omega} G \cdot \Phi \, dx.$$

On a donc montré que pour tout $\Phi \in (\mathcal{C}^\infty(\mathbb{R}^d))^d$, on a

$$\int_{\Omega} u(\text{div} \Phi) \, dx = - \int_{\Omega} G \cdot \Phi \, dx,$$

ce qui prouve bien que $u \in H_0^1(\Omega)$ et que $G = \nabla u$ est bien la limite faible des gradients discrets. ■

On peut maintenant énoncer et démontrer le théorème de convergence du schéma volumes finis VF4.

Théorème III.8 (Convergence du schéma VF4)

Soit $f \in L^2(\Omega)$. Soit $(\mathcal{T}_n)_n$ une famille de maillages dont le pas tend vers 0. Pour tout n , on note $u^{\mathcal{T}_n} \in \mathbb{R}^{\mathcal{T}_n}$ l'unique solution du schéma sur le maillage \mathcal{T}_n pour la donnée f . On note u la solution exacte du problème sur Ω .

On a :

- *La suite $(u^{\mathcal{T}_n})_n$ converge fortement vers u dans $L^2(\Omega)$.*
- *La suite $(\nabla^{\mathcal{T}_n} u^{\mathcal{T}_n})_n$ converge faiblement vers ∇u dans $(L^2(\Omega))^d$.*

On va utiliser la définition suivante dans la preuve.

Définition III.9 (Projection de fonctions régulières)

Pour toute fonction v continue sur $\overline{\Omega}$ et nulle au bord, on pose

$$\mathbb{P}^{\mathcal{T}} v = (v(x_{\kappa}))_{\kappa \in \mathcal{T}}.$$

Preuve :

• **Déroulement de la preuve :**

D'après le théorème de compacité et la borne *a priori* sur les solutions approchées, on peut considérer tout d'abord une sous-suite, toujours notée $(u^{\mathcal{T}_n})_n$ qui converge fortement dans $L^2(\Omega)$ vers un $v \in H_0^1(\Omega)$ et telle que $\nabla^{\mathcal{T}_n} u^{\mathcal{T}_n}$ converge faiblement vers ∇v . On va montrer que cette fonction v vérifie la formulation faible du problème de Poisson, ce qui montrera, par unicité, que $v = u$.

La limite v obtenue étant unique (elle ne dépend pas de la sous-suite convergente choisie au départ), un résultat classique de compacité fournit la convergence de la suite entière vers la solution u du problème dans le sens écrit dans l'énoncé du théorème.

- Soit donc $\varphi \in \mathcal{C}_c^\infty(\Omega)$ une fonction test. On va prendre $\mathbb{P}^{\mathcal{T}_n} \varphi \in \mathbb{R}^{\mathcal{T}_n}$ comme fonction-test discrète dans le schéma volumes finis, ce qui donne d'après le Lemme III.2.

$$\sum_{\sigma \in \mathcal{E}} d_{\sigma} |\sigma| D_{\sigma}(u^{\mathcal{T}_n}) \frac{\varphi(x_{\mathcal{L}}) - \varphi(x_{\mathcal{K}})}{d_{\mathcal{K}\mathcal{L}}} = \sum_{\kappa \in \mathcal{T}} |\mathcal{K}| f_{\kappa} \varphi(x_{\kappa}).$$

En utilisant la définition du gradient discret, on trouve

$$\sum_{\sigma \in \mathcal{E}} \frac{d_\sigma |\sigma|}{d} \nabla_{\mathcal{D}}^T(u^{\mathcal{T}_n}) \cdot \left(\frac{\varphi(x_{\mathcal{L}}) - \varphi(x_{\mathcal{K}})}{d_{\mathcal{K}\mathcal{L}}} \boldsymbol{\nu}_{\mathcal{K}\mathcal{L}} \right) = \sum_{\kappa \in \mathcal{T}} |\mathcal{K}| f_{\kappa} \varphi(x_{\kappa}).$$

D'après la condition d'orthogonalité, la mesure du diamant correspondant à σ est exactement égale à $\frac{1}{d} |\sigma| d_\sigma$. De plus, comme $\nabla_{\mathcal{D}}^T u^{\mathcal{T}}$ est orienté selon la normale à l'arête σ , on peut rajouter au terme contenant φ , n'importe quel élément tangent à σ . Ainsi, en réécrivant tous les termes comme des intégrales, on trouve

$$\begin{aligned} \sum_{\sigma \in \mathcal{E}} \int_{\mathcal{D}} \nabla_{\mathcal{D}}^T(u^{\mathcal{T}_n}) \cdot \left(\nabla \varphi(x) + \underbrace{\left(\frac{\varphi(x_{\mathcal{L}}) - \varphi(x_{\mathcal{K}})}{d_{\mathcal{K}\mathcal{L}}} \boldsymbol{\nu}_{\mathcal{K}\mathcal{L}} - (\nabla \varphi(x) \cdot \boldsymbol{\nu}_{\mathcal{K}\mathcal{L}}) \boldsymbol{\nu}_{\mathcal{K}\mathcal{L}} \right)}_{=R_{\nabla \varphi}(x)} \right) dx \\ = \sum_{\kappa \in \mathcal{T}} \int_{\mathcal{K}} f(x) \left(\varphi(x) + \underbrace{(\varphi(x_{\mathcal{K}}) - \varphi(x))}_{=R_{\varphi}(x)} \right) dx, \quad (\text{III.7}) \end{aligned}$$

avec les modifications qui s'imposent pour les termes du bord. En particulier, on utilise ici le fait que φ est nulle au bord. On peut également éviter le problème du bord en disant que, si φ est fixée et à support compact, alors pour n assez grand, les termes de bord sont tous nuls car les volumes de contrôle correspondants sont en dehors du support de φ .

Quoi qu'il en soit, comme φ est régulière, on vérifie, par développement de Taylor très usuels, que l'on a

$$\forall \mathcal{D} \in \mathfrak{D}, \forall x \in \mathcal{D}, |R_{\nabla \varphi}(x)| \leq \|D^2 \varphi\|_{L^\infty} d_{\mathcal{D}} \leq 2 \|D^2 \varphi\|_{L^\infty} \text{size}(\mathcal{T}_n),$$

$$\forall \mathcal{K} \in \mathcal{T}, \forall x \in \mathcal{K}, |R_{\varphi}(x)| \leq \|\nabla \varphi\|_{L^\infty} \text{diam}(\mathcal{K}) \leq \|\nabla \varphi\|_{L^\infty} \text{size}(\mathcal{T}_n).$$

Comme $\nabla^T(u^{\mathcal{T}_n})$ est bornée dans L^2 , on a donc obtenu finalement

$$\int_{\Omega} \nabla_{\mathcal{D}}^T(u^{\mathcal{T}_n}) \cdot \nabla \varphi(x) dx = \int_{\Omega} f(x) \varphi(x) dx + O(\text{size}(\mathcal{T}_n)).$$

Mais, par définition, $\nabla_{\mathcal{D}}^T(u^{\mathcal{T}_n})$ converge faiblement vers ∇v dans $(L^2(\Omega))^d$, on peut passer à la limite et obtenir

$$\int_{\Omega} \nabla v \cdot \nabla \varphi dx = \int_{\Omega} f(x) \varphi(x) dx.$$

Ceci étant vrai pour toute fonction $\mathcal{C}_c^2(\Omega)$, c'est encore vrai pour toute fonction $\varphi \in H_0^1(\Omega)$ par densité et donc v est l'unique solution variationnelle du problème de Laplace, c'est-à-dire que $v = u$ et le théorème de convergence est démontré. ■

Il faut remarquer une certaine similitude de la démonstration avec la preuve de convergence de l'approximation de Galerkin, où on utilisait le fait que $d(v, V_h) \rightarrow 0$ quand $h \rightarrow 0$ pour toute fonction test $v \in H_0^1(\Omega)$.

Dans ce même théorème, nous avons conclu à la convergence forte de la solution dans H^1 en démontrant la convergence de la norme du gradient. On va voir qu'ici la situation est très différente.

Théorème III.10 (Non convergence forte du gradient approché)

Dans les conditions du théorème précédent, et si la dimension d est au moins égale à 2, la convergence forte de la suite des gradients approchés $(\nabla^{\mathcal{T}_n} u^{\mathcal{T}_n})_n$ vers ∇u n'a lieu que dans le cas où $f = u = 0$.

Remarquons qu'en dimension 1, on a bien la convergence forte du gradient discret.

Preuve :

Supposons f non identiquement nulle, de sorte que u est également non nulle. Si on choisit $v^{\mathcal{T}} = u^{\mathcal{T}_n}$ comme fonction test dans le schéma, on voit que le lemme III.2 donne

$$\sum_{\sigma \in \mathcal{E}} d_\sigma |\sigma| D_\sigma(u^{\mathcal{T}_n})^2 = \sum_{\kappa \in \mathcal{T}} |\mathcal{K}| f_{\kappa} u_{\kappa},$$

où encore, d'après la définition du gradient discret :

$$\frac{1}{d} \|\nabla^{\mathcal{T}} u^{\mathcal{T}n}\|_{L^2}^2 = \frac{1}{d} \sum_{\sigma \in \mathcal{E}} |\mathcal{D}| |\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}n}|^2 = \sum_{\kappa \in \mathcal{T}} |\mathcal{K}| f_{\kappa} u_{\kappa}.$$

Comme on a la convergence forte de $(u^{\mathcal{T}n})_n$ vers u dans L^2 , on peut passer à la limite dans le second membre et ainsi obtenir

$$\lim_{n \rightarrow \infty} \|\nabla^{\mathcal{T}} u^{\mathcal{T}n}\|_{L^2}^2 = d \int_{\Omega} f(x) u(x) dx = d \|\nabla u\|_{L^2}^2.$$

Ainsi, comme u n'est pas constante, la norme L^2 du gradient approché ne converge pas vers la norme L^2 du gradient exact (qui est la limite faible des gradients approchés) et donc la convergence n'est certainement pas forte. ■

1.5 Estimation de l'erreur en dimension 2

Pour estimer l'erreur d'approximation du schéma VF4, il faut bien évidemment supposer plus de régularité sur la solution $u \in H_0^1(\Omega)$ du problème étudié. Classiquement (comme dans le chapitre précédent), nous allons supposer que $u \in H^2(\Omega)$ (bien que l'on puisse considérer bien d'autres cadres fonctionnels, en particulier le cas $u \in \mathcal{C}^2(\overline{\Omega})$ est beaucoup plus simple). Le résultat qui suit est présenté en dimension 2 pour simplifier, mais il est vrai en toute dimension quitte à modifier légèrement la preuve et la définition de la constante de régularité du maillage.

Dans ces conditions, u est une fonction continue et on peut donc considérer sa projection sur le maillage $\mathbb{P}^{\mathcal{T}} u = (u(x_{\kappa}))_{\kappa}$. On peut alors définir l'erreur comme le vecteur $e^{\mathcal{T}} = u^{\mathcal{T}} - \mathbb{P}^{\mathcal{T}} u$.

De plus, comme $u \in H^2(\Omega)$, on peut réellement appliquer la formule de Stokes sur chaque volume de contrôle (avec des vraies intégrales sur le bord car $\nabla u \cdot \nu_{\kappa}$ est dans $H^{\frac{1}{2}}(\partial\mathcal{K})$). Il vient

$$|\mathcal{K}| f_{\kappa} = \int_{\mathcal{K}} f(x) dx = \sum_{\sigma \in \mathcal{E}_{\kappa}} \int_{\sigma} -\nabla u \cdot \nu_{\kappa\sigma} dx,$$

ce que l'on va écrire

$$|\mathcal{K}| f_{\kappa} = \sum_{\sigma \in \mathcal{E}_{\kappa}} -|\sigma| \frac{u(x_{\mathcal{L}}) - u(x_{\kappa})}{d_{\kappa\mathcal{L}}} + \sum_{\sigma \in \mathcal{E}_{\kappa}} |\sigma| R_{\sigma}(u),$$

où $R_{\sigma}(u)$ est l'erreur de consistance du flux numérique (normalisé) définie par

$$R_{\sigma}(u) = \frac{u(x_{\mathcal{L}}) - u(x_{\kappa})}{d_{\kappa\mathcal{L}}} - \frac{1}{|\sigma|} \int_{\sigma} \nabla u \cdot \nu_{\kappa\mathcal{L}} dx.$$

Si on soustrait, la définition du schéma à cette formule, on trouve que l'erreur $e^{\mathcal{T}}$ vérifie

$$\sum_{\sigma \in \mathcal{E}} d_{\sigma} |\sigma| D_{\sigma}(e^{\mathcal{T}}) D_{\sigma}(v^{\mathcal{T}}) = \sum_{\sigma \in \mathcal{E}} d_{\sigma} |\sigma| R_{\sigma}(u) D_{\sigma}(v^{\mathcal{T}}).$$

Ainsi, si on prend $v^{\mathcal{T}} = e^{\mathcal{T}}$ dans cette formule, puis qu'on utilise l'inégalité de Cauchy-Schwarz on trouve

$$\|e^{\mathcal{T}}\|_{1,\mathcal{T}}^2 \leq \sum_{\sigma \in \mathcal{E}} d_{\sigma} |\sigma| |R_{\sigma}(u)|^2. \quad (\text{III.8})$$

Ainsi, pour estimer l'erreur, il suffit de savoir estimer l'erreur de consistance $R_{\sigma}(u)$ en fonction du pas de maillage et de la norme H^2 de u . Pour cela, on a besoin d'une hypothèse de régularité sur le maillage. Celle-ci va consister à demander que, dans chaque diamant \mathcal{D} , les distances $d_{\kappa\sigma}$ et $d_{\mathcal{L}\sigma}$ soient de taille comparables à la longueur de l'arête correspondante (on rappelle que $d_{\kappa\mathcal{L}} = d_{\kappa\sigma} + d_{\mathcal{L}\sigma}$). Pour chaque maillage admissible, on introduit donc une mesure de régularité

$$\text{reg}(\mathcal{T}) = \sup_{\mathcal{D} \in \mathfrak{D}} \left(\frac{|\sigma|}{d_{\kappa\sigma}} + \frac{|\sigma|}{d_{\mathcal{L}\sigma}} \right).$$

Lemme III.11

Il existe $C > 0$ ne dépendant que de $\text{reg}(T)$ telle que

$$\forall \sigma \in \mathcal{E}, \quad |R_\sigma(u)| \leq C d_{\mathcal{D}} \left(\frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} |D^2 u|^2 dx \right)^{\frac{1}{2}}.$$

Preuve :

Les quantités qui interviennent dans l'inégalité attendue sont continues par rapport à la norme H^2 de u , il suffit donc, par un argument de densité maintenant classique, que cette estimation est vraie pour toute fonction u très régulière (par exemple $u \in C^2(\overline{\Omega})$).

Soit $x \in \sigma$. On effectue, un développement de Taylor de u au point x évalué en x_κ (resp. en $x_\mathcal{L}$). Il vient

$$u(x_\kappa) = u(x) + \nabla u(x) \cdot (x_\kappa - x) + \int_0^1 (1-t) D^2 u(x + t(x_\kappa - x)) \cdot (x_\kappa - x)^2 dt,$$

$$u(x_\mathcal{L}) = u(x) + \nabla u(x) \cdot (x_\mathcal{L} - x) + \int_0^1 (1-t) D^2 u(x + t(x_\mathcal{L} - x)) \cdot (x_\mathcal{L} - x)^2 dt.$$

On soustrait alors ces deux égalités et on utilise la condition d'orthogonalité qui dit exactement $x_\mathcal{L} - x_\kappa = d_{\kappa\mathcal{L}} \nu_{\kappa\mathcal{L}}$ pour obtenir

$$u(x_\mathcal{L}) - u(x_\kappa) = d_{\kappa\mathcal{L}} \nabla u(x) \cdot \nu_{\kappa\mathcal{L}} + \int_0^1 (1-t) D^2 u(x + t(x_\mathcal{L} - x)) \cdot (x_\mathcal{L} - x)^2 dt - \int_0^1 (1-t) D^2 u(x + t(x_\kappa - x)) \cdot (x_\kappa - x)^2 dt.$$

On divise ensuite par $d_{\kappa\mathcal{L}}$ et on prend la moyenne (en x) sur σ , ce qui donne

$$R_\sigma(u) = \underbrace{\frac{1}{d_{\kappa\mathcal{L}} |\sigma|} \int_\sigma \int_0^1 (1-t) D^2 u(x + t(x_\mathcal{L} - x)) \cdot (x_\mathcal{L} - x)^2 dt dx}_{=T_1} - \underbrace{\frac{1}{d_{\kappa\mathcal{L}} |\sigma|} \int_\sigma \int_0^1 (1-t) D^2 u(x + t(x_\kappa - x)) \cdot (x_\kappa - x)^2 dt dx}_{=T_2}.$$

On va maintenant traiter le terme T_1 , le second terme étant en tout point similaire.

$$|T_1|^2 \leq \frac{1}{d_{\kappa\mathcal{L}}^2 |\sigma|} \int_\sigma \int_0^1 |1-t|^2 |D^2 u(x + t(x_\mathcal{L} - x))|^2 |x_\mathcal{L} - x|^4 dt dx.$$

On considère maintenant le changement de variable

$$(t, x) \in [0, 1] \times \sigma \mapsto y = x + t(x_\mathcal{L} - x) \in \mathcal{D}_\mathcal{L},$$

où $\mathcal{D}_\mathcal{L}$ est le demi-diamant du côté \mathcal{L} . Il s'agit bien d'une transformation bijective dont le jacobien vaut $(1-t)(x_\mathcal{L} - x) \cdot \nu$, on obtient donc

$$|T_1|^2 \leq \frac{d_{\mathcal{D}}^4}{d_{\kappa\mathcal{L}}^2 d_{\mathcal{L},\sigma} |\sigma|} \int_{\mathcal{D}_\mathcal{L}} |D^2 u(y)|^2 dy \leq C(\text{reg}(T)) \frac{d_{\mathcal{D}}^2}{|\mathcal{D}|} \int_{\mathcal{D}} |D^2 u(y)|^2 dy.$$

Un calcul similaire pour T_2 conclut la preuve. ■

En mettant bout à bout l'inégalité (III.8) et le lemme précédent, on obtient le résultat suivant

$$\|e^{\mathcal{T}}\|_{1,\mathcal{T}}^2 \leq C(\text{reg}(T)) \sum_{\mathcal{D} \in \mathfrak{D}} d_{\mathcal{D}}^2 \left(\int_{\mathcal{D}} |D^2 u|^2 dx \right) C(\text{reg}(T)) \text{size}(T)^2 \int_{\Omega} |D^2 u|^2 dx,$$

ce qui fournit une estimation d'erreur d'ordre 1 en norme H^1 discrète

$$\|e^{\mathcal{T}}\|_{1,\mathcal{T}} \leq \tilde{C}(\text{reg}(T)) \text{size}(T) |u|_{H^2(\Omega)}.$$

Par l'inégalité de Poincaré discrète on déduit aussi une estimation d'erreur en norme L^2

$$\|e^{\mathcal{T}}\|_{L^2} \leq \bar{C}(\text{reg}(T)) \text{size}(T) |u|_{H^2(\Omega)}.$$

2 Quelques problèmes plus généraux

Ce paragraphe est dans un état embryonnaire ... il a vocation à être détaillé dès que possible

2.1 Le problème de diffusion isotrope hétérogène

On s'intéresse dans cette partie à l'extension du schéma VF4 au cas isotrope et hétérogène. C'est-à-dire que l'on remplace le problème $-\Delta u = f$ par le problème $-\operatorname{div}(k(x)\nabla u) = f$ avec $k(x)$ une fonction **scalaire** strictement positive et bornée.

Dans ces conditions, les flux exacts que l'on doit approcher sont de la forme

$$\int_{\sigma} (k(x)\nabla u) \cdot \nu.$$

Il est tentant de les approcher par une formule à deux points du type

$$F_{\mathcal{K}\mathcal{L}} = k_{\sigma} |\sigma| \frac{u_{\mathcal{L}} - u_{\mathcal{K}}}{d_{\mathcal{K}\mathcal{L}}},$$

où k_{σ} est un coefficient de diffusion "moyen" sur l'arête. La question est de savoir quelle valeur on va choisir pour ce coefficient.

1. Premier cas : la fonction k de départ est suffisamment régulière (disons Lipschitzienne), alors la solution u du problème est régulière (au moins si le domaine est convexe) et dans ce cas tout choix raisonnable de k_{σ} convient, comme par exemple

$$k_{\sigma} = k(x_{\mathcal{K}}), \quad k_{\sigma} = k(x_{\mathcal{L}}), \quad k_{\sigma} = \frac{1}{2}(k(x_{\mathcal{K}}) + k(x_{\mathcal{L}})), \quad k_{\sigma} = \frac{1}{|\sigma|} \int_{\sigma} k(x) dx, \dots$$

On montre alors dans tous ces cas, l'existence et l'unicité de la solution approchée, la convergence de celle-ci vers la solution exacte du problème et enfin l'estimation d'erreur d'ordre 1 en norme H^1 discrète.

2. Second cas : la fonction k n'est pas suffisamment régulière (typiquement discontinue), alors d'une part, on ne peut plus supposer que u est dans $H^2(\Omega)$, car cela est grossièrement faux à la traversée des surfaces de discontinuités de k .

Regardons le cas où k est constante sur chaque volume de contrôle par exemple. On notera $k_{\mathcal{K}}$ la valeur sur la maille \mathcal{K} .

L'idée est de dire que, comme sur le problème continu on a $k(x)\nabla u \in H_{\operatorname{div}}(\Omega)$, la quantité qui est "continue" au travers des interfaces dans ce problème c'est le vrai flux $(k\nabla u) \cdot \nu$.

Pour calculer un bon flux numérique on va introduire momentanément une nouvelle inconnue "artificielle" notée u_{σ} qui correspond à une approximation de la solution à l'interface, au point x_{σ} .

De ce fait, on peut raisonnablement construire une approximation de $\nabla u \cdot \nu$ du côté \mathcal{K} et une autre approximation du côté \mathcal{L} données respectivement par

$$\frac{u_{\mathcal{L}} - u_{\sigma}}{d_{\mathcal{L}\sigma}}, \quad \text{et} \quad \frac{u_{\sigma} - u_{\mathcal{K}}}{d_{\mathcal{K}\sigma}}.$$

On demande ensuite la continuité du flux autrement dit, on écrit :

$$F_{\mathcal{L},\sigma} = k_{\mathcal{L}} \frac{u_{\mathcal{L}} - u_{\sigma}}{d_{\mathcal{L}\sigma}} = k_{\mathcal{K}} \frac{u_{\sigma} - u_{\mathcal{K}}}{d_{\mathcal{K}\sigma}} = F_{\mathcal{K},\sigma}.$$

L'inégalité centrale permet de déterminer u_{σ} en fonction de $u_{\mathcal{K}}$ et $u_{\mathcal{L}}$, puis de déterminer enfin la valeur du flux numérique en fonction de $u_{\mathcal{L}}$ et $u_{\mathcal{K}}$.

2.2 Le problème de convection-diffusion

2.3 Le problème anisotrope hétérogène

BIBLIOGRAPHIE

- [Ada75] R. A. Adams, *Sobolev spaces*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975, Pure and Applied Mathematics, Vol. 65.
- [ADN59] S. Agmon, A. Douglis, and L. Nirenberg, *Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I*, Comm. Pure Appl. Math. **12** (1959), 623–727.
- [ADN64] ———, *Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions II*, Comm. Pure Appl. Math. **17** (1964), 35–92.
- [BF06] F. Boyer and P. Fabrie, *Eléments d'analyse pour l'étude de quelques modèles d'écoulements de fluides visqueux incompressibles*, Mathématiques et Applications, vol. 52, Springer, 2006.
- [Bre83] H. Brezis, *Analyse fonctionnelle*, Masson, 1983.
- [EG04] A. Ern and J.-L. Guermond, *Theory and practice of finite elements*, Applied Mathematical Sciences, vol. 159, Springer-Verlag, New York, 2004. MR MR2050138 (2005d :65002)
- [EGH00] Robert Eymard, Thierry Gallouët, and Raphaële Herbin, *Finite volume methods*, Handbook of numerical analysis, Vol. VII, Handb. Numer. Anal., VII, North-Holland, Amsterdam, 2000, pp. 713–1020.
- [FG07] Demengel F. and Demengel G., *Espaces fonctionnels. utilisation dans la résolution des équations aux dérivées partielles*, Savoirs actuels, EDP Sciences, 2007.