Université Ahmed ZABANA Relizane

Département de Physique

Master 1ere Année

METHODES DES ELEMENTS FINIS

Cours et exercices

Table des matières

Préface

L'objet de ce cours est d'introduire les notions de base de résolution des équations aux dérivées partielles par la méthode des éléments finis. Depuis son introduction au milieu du XX$^{\text{ème}}$ siècle, cette méthode est devenue l'outil de base dans la résolution des équations aux dérivées partielles qui interviennent dans les études scientifiques ou techniques. Conçue initialement comme un procédé de calcul en mécanique des structures, c'est sa formalisation qui a permis de l'étendre efficacement à des domaines complètement différents comme la mécanique des fluides ou l'électromagnétisme. Nous prenons en quelque sorte l'édifice sous sa forme achevée : nous introduisons cette méthode sous sa forme de procédé général de résolution.

La formalisation ci-dessus est basée sur une formulation variationnelle des problèmes d'équations aux dérivées partielles, posées sur un domaine de $\mathbb{R}^N$ ($N \leq 3$ généralement dans les problèmes d'ingénierie), qui avec des conditions appropriées sur la solution au bord de ce domaine, sont nommés problèmes aux limites. La méthode des éléments finis apparait alors comme une méthode de Galerkin particulière. Ce formalisme nous permettra, non seulement d'avoir un cadre général pour la description de cette méthode, mais en plus de bien comprendre ses limites de validité : en particulier, les situations où elle engendre un procédé de résolution instable.

Contrairement, cependant, à une présentation orientée essentiellement vers la description des propriétés mathématiques de cette méthode, nous détaillerons la mise en œuvre numérique et les principes de programmation des algorithmes induits qui sont aussi importants en ingénierie que ses performances en tant que procédé d'approximation.

Nous commencerons par les problèmes dits elliptiques qui, en ingénierie, modélisent des phénomènes où le temps t n'est pas un paramètre : ce sont les problèmes statiques, c'est à dire, décrits par des variables constantes au cours du temps, ou des phénomènes stationnaires où la dépendance en temps est connue a priori. Nous verrons ensuite comment étendre cette méthode aux problèmes évoluant au cours du temps à partir d'une condition initiale.

Notre objectif dans ce cours a été de ménager un équilibre entre les aspects mathématiques, qui sont importants pour une réelle compréhension de la méthode et son utilisation avec efficacité dans des situations non cataloguées *a priori*, et les aspects d'implémentation numérique dont l'importance est de premier plan. Les retours sur cet enseignement, qui sont fortement sollicités, permettront de mesurer le niveau de réalisation de ce programme.

CHAPITRE 1

Notions élémentaires sur la théorie des problèmes aux limites elliptiques

Nous introduisons dans ce chapitre la notion de problème aux limites elliptique. Même en se limitant aux problèmes scalaires, i.e. dont l'inconnue est une fonction à valeurs scalaires, cette classe de problèmes intervient dans un grand nombre de situations physiques en sciences de l'ingénieur comme le montreront les quelques exemples que nous considèrerons. La résolution numérique standard de ces problèmes est basée sur l'utilisation d'une méthode d'éléments finis. Cette méthode est basée sur une formulation variationnelle de ces problèmes et apparaît alors comme une méthode de Galerkin particulière. Nous nous concentrerons sur cet aspect dans ce chapitre.

1. Problème en dimension un

1.1. Position générale du problème aux limites. La forme générale des problèmes aux limites elliptiques est la suivante : il s'agit de déterminer une fonction inconnue u sur l'intervalle $]0, L[$, qui est le **domaine** où est posé le problème aux limites elliptique, qui satisfait les conditions suivantes.

- **Une équation aux dérivées partielles** (équation différentielle ici puisqu'il n'y a qu'**une variable de dérivation dans un problème en dimension un**)

$$-\partial_x(a(x)\partial_x u)(x) + a_0(x)u(x) = f(x), \quad x \in]0, L[\,. \tag{1.1}$$

- **Des conditions aux limites** sur la frontière du domaine (ici les points $x = 0$ et $x = L$). Nous prendrons ici pour décrire les différentes situations qui peuvent se présenter en pratique :

(1) – **condition de Dirichlet** en $\{x = 0\}$

$$u(0) = g_0. \tag{1.2}$$

– **condition de Fourier-Robin** (pour $\lambda \neq 0$), **de Neumann** (pour $\lambda = 0$), en $\{x = L\}$

$$(a\partial_x u)(L) + \lambda u(L) = h_0. \tag{1.3}$$

Les données peuvent être classées suivant les rubriques suivantes.

- **Coefficients de l'EDP** : ce sont les deux fonctions a et a_0, (éventuellement définies seulement presque partout sur $]0, L[$). Les propriétés physiques du système étudié assurent que le coefficient a vérifie la propriété suivante (qui caractérise le caractère elliptique du problème et qui sera fondamentale aussi bien d'un point de vue théorique que pour les propriétés des schémas d'approximation numérique) : il existe deux constantes α et β telles que

$$0 < \alpha \leq a(x) \leq \beta, \quad \text{p. p. tout } x \in]0, L[\,. \tag{1.4}$$

Le coefficient a_0 est généralement nul. Il apparait lorsqu'on utilise une semi-discrétisation en temps pour un problème évoluant avec le temps ou s'il y a une absorption

d'énergie par des forces de frottement. Il vérifie : il existe une constante β_0 telle que

$$0 \leq a_0(x) \leq \beta_0, \quad \text{p. p. tout } x \in]0, L[\,. \tag{1.5}$$

- **Second membre de l'EDP** : il est donné par la fonction f qui peut être définie seulement presque partout sur $]0, L[$. Nous préciserons la classe fonctionnelle auquelle elle appartient lors de l'étude de l'existence-unicité d'une solution au problème aux limites.
- **Condition aux limites de Dirichlet** : elle correspond physiquement à une condition imposée, à une contrainte sur le système. On *connait* la solution au point $\{x = 0\}$. Sa valeur est donnée et égale ici à $g_{0.}$.
- **Condition de Neumann ou de Fourier-Robin** : elle correspond au cas où on laisse le système physique évoluer librement. Le paramètre $\lambda \geq 0$ traduit une réaction de l'extérieur et, plus précisément une absorption de l'énergie par l'extérieur.

1.2. Quelques exemples de problèmes physiques. Nous donnons ci-après sous forme d'un tableau quelques exemples de situations physiques modélisées par un problème aux limites elliptique du second ordre. Cette liste montre, que même dans un cadre simple, on peut décrire, de cette façon, plusieurs situations significatives en ingénierie. Dans tous ces exemples, la fonction a_0 et la constante λ sont nulles. La variable primaire est l'inconnue par rapport à laquelle on résout le problème. La variable secondaire est une quantité qui fixe la donnée de la condition de Neumann ou qu'il est important de calculer par un post-traitement une fois l'inconnue primaire déterminée.

Problème	**Variable primaire** u	**Loi de comportement** a	**Terme source** f	**Variable secondaire** $a\partial_x u$
Déflexion d'un cable	Déflexion transverse	Tension dans le cable	Densité de chargement transverse	Force axiale (généralement, inconnue)
Transfert thermique	Température	Conductivité thermique	Apport calorifique	Flux thermique
Ecoulement dans une conduite	Vitesse	Viscosité	Gradient de pression	Contrainte axiale
Ecoulement dans un milieu poreux	Vitesse	Coefficient de perméablilité	Injection ou extraction	Flux (filtration)
Electrostatique	Potentiel électrique	Permittivité diélectrique	Densité de charges	Flux du champ électrique

1.3. Formulation variationnelle. L'étude de l'existence-unicité d'une solution pour le problème (1.1, 1.2, 1.3) et la mise en oeuvre d'un schéma pour sa résolution numérique passent par sa formulation sous forme d'un problème variationnel. La technique de base pour l'effectuer repose sur une formule d'intégration par parties. Pour cela, on considère une fonction test v, quelconque pour l'instant. On écrit

$$-\int_0^L \partial_x(a\partial_x u)v\,dx + \int_0^L a_0 uv\,dx = \int_0^L fv\,dx \tag{1.6}$$

et ensuite

$$-\int_0^L \partial_x(a\partial_x u)v\,dx = (a\partial_x u)(L)v(L) - (a\partial_x u)(0)v(0) + \int_0^L a\partial_x u\partial_x v\,dx. \tag{1.7}$$

En imposant à la fonction test de s'annuler là où est donnée une condition de Dirichlet, on obtient le problème variationnel

$$\left\{\begin{array}{l} u(0) = g_0, \quad \forall v,\ v(0) = 0, \\ \displaystyle\int_0^L (a\partial_x u \partial_x v + a_0 uv)\ dx + \lambda u(L)v(L) = \int_0^L fv\,dx + h_0 v(L). \end{array}\right. \tag{1.8}$$

Reamarquons que, si les conditions de **Dirichlet sont imposées** dans la formulation, celles de **Neumann sont implicites**.

1.4. Cadre fonctionnel de la formulation variationnelle. Le cadre fonctionnel, comme la formulation variationnelle, est important non seulement pour établir rigoureusement un résultat d'existence-unicité de la solution mais aussi pour comprendre la dérivation des schémas numériques d'approximation de celle-ci.

Pour que la formulation variationnelle ait un sens, il faudrait d'abord que toutes les intégrales existent. Comme a et a_0 sont bornées, i.e. dans $L^\infty(]0,L[)$, ceci revient à exiger que u, $\partial_x u$, v, $\partial_x v$, et f soient dans $L^2(]0,L[)$, les dérivations étant prises au sens des distributions.

On voit donc que l'espace fonctionnel, où on doit chercher la solution et faire varier la fonction test, est l'**espace de Sobolev** suivant

$$H^1(]0,L[) := \left\{v \in L^2(]0,L[)\,;\ \partial_x v \in L^2(]0,L[)\right\}. \tag{1.9}$$

L'importance de ce cadre fonctionnel vient de la propriété suivante.

THÉORÈME 1.1. *L'espace $H^1(]0,L[)$ est un espace de Hilbert pour le produit scalaire*

$$(u,v)_{H^1(]0,L[)} = \int_0^L (\partial_x u \partial_x v + uv)\,dx. \tag{1.10}$$

DÉMONSTRATION. Il est immédiat de vérifier que l'expression de l'énoncé est un produit scalaire.

Introduisons quelques notations

$$|v|_{0,]0,L[} := \left\{\int_0^L uv\,dx\right\}^{1/2}, \qquad |v|_{1,]0,L[} := \left\{\int_0^L \partial_x u \partial_x v\,dx\right\}^{1/2} \tag{1.11}$$

respectivement la norme dans $L^2(]0,L[)$ et la semi-norme d'ordre 1. La norme associée au produit scalaire est ainsi

$$\|v\|_{1,]0,L[} = \left\{|v|^2_{1,]0,L[} + |v|^2_{0,]0,L[}\right\}^{1/2}. \tag{1.12}$$

Il suffit ainsi de vérifier que si $\{v_n\}_{n\geq 0}$ est telle que

$$\lim_{n,m\to\infty} \|v_n - v_m\|_{1,]0,L[} = 0,$$

alors il existe $v \in H^1(]0,L[)$ telle que

$$\lim_{n\to\infty} \|v_n - v\|_{1,]0,L[} = 0. \tag{1.13}$$

Mais comme

$$|v_n - v_m|_{0,]0,L[} \leq \|v_n - v_m\|_{1,]0,L[}, \quad |v_n - v_m|_{1,]0,L[} \leq \|v_n - v_m\|_{1,]0,L[},$$

et que l'espace $L^2(]0,L[)$ est **complet**, il existe v et g dans $L^2(]0,L[)$ tels que

$$\lim_{n\to\infty} |v_n - v|_{0,]0,L[} = 0 \text{ et } \lim_{n\to\infty} |\partial_x v_n - g|_{0,]0,L[} = 0.$$

La convergence dans L^2 entrainant la convergence au sens des distributions, on a ainsi

$$\lim_{n\to\infty} \partial_x v_n = \partial_x v \text{ au sens des distributions.}$$

Comme $\lim_{n\to\infty} \partial_x v_n = g$ dans L^2, et donc au sens des distributions, il vient

$$\partial_x v = g.$$

On a donc bien vérifié (1.13). □

Il reste à s'assurer que les conditions de Dirichlet ont un sens de même que le terme $\lambda u(L)v(L)$. Le fait qu'on soit en **dimension un** permet de répondre relativement facilement à cette question. En effet, l'espace $H^1(]0,L[)$ possède les propriétés importantes suivantes qui résultent de celles de l'intégrale de Lebesgue :

– tout $v \in H^1(]0,L[)$ est continu sur $[0,L]$

$$H^1(]0,L[) \subset \mathcal{C}^0([0,L]) \ ; \tag{1.14}$$

– de plus, la formule d'intégration par parties suivante est vérifiée

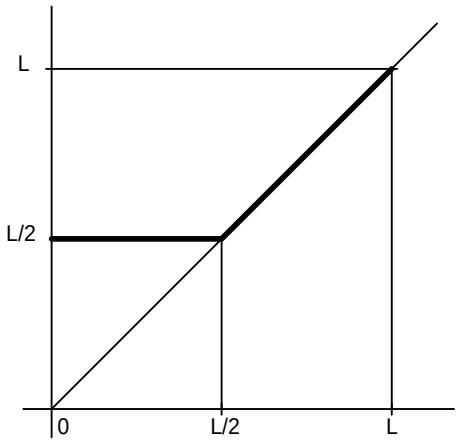
$$\left\{ \begin{array}{l} \text{pour } 0 \le a < b \le L \text{ et } u,\ v \text{ dans } H^1(]0,L[), \\ \displaystyle\int_a^b (u\partial_x v + v\partial_x u)\, dx = u(b)v(b) - u(a)v(a) \end{array} \right. \tag{1.15}$$

En particulier, en prenant $v = 1$, on déduit de (1.15), la formule suivante

$$u(x) = u(a) + \int_a^x \partial_x u(t)\, dt. \tag{1.16}$$

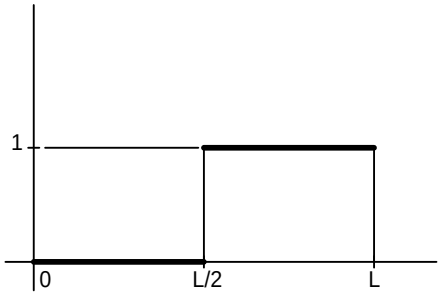
REMARQUE 1.1. *Bien sûr, l'espace $H^1(]0,L[)$ ne se limite pas aux fonctions dans $\mathcal{C}^1([0,L])$. La formule des sauts montre que la fonction*

$$u(x) = \left\{ \begin{array}{ll} L/2, & \textit{pour } 0 < x \le L/2, \\ x, & \textit{pour } L/2 \le x < L, \end{array} \right.$$



a pour dérivée

$$\partial_x u = \left\{ \begin{array}{ll} 0, & \textit{pour } 0 < x < L/2, \\ 1, & \textit{pour } 0 < x < L, \end{array} \right.$$



qui est dans $L^\infty(]0,L[)$, et donc dans $L^2(]0,L[)$. Remarquer que $\partial_x u$ n'est définie que p. p. sur $]0,L[$.

Les deux membres de la formulation variationnelle (1.8) ont ainsi un sens. Pour préparer l'étude de cette formulation, nous l'écrivons à l'aide des notations suivantes

$$a(u,v) := \int_0^L (a\partial_x u\partial_x v + a_0 uv)\, dx + \lambda u(L)v(L), \tag{1.17}$$

$$Lv := \int_0^L fv\, dx + h_0 v(L) \tag{1.18}$$

$$V := \left\{ v \in H^1(]0,L[)\,;\ v(0) = 0 \right\} \tag{1.19}$$

$$\begin{cases} u \in H^1(]0,L[) \text{ tel que } u - g_0 \in V, \quad \forall v \in V, \\ a(u,v) = Lv. \end{cases} \tag{1.20}$$

1.5. Propriétés de la formulation variationnelle. Nous allons maintenant mettre en évidence les propriétés de la formulation variationnelle (1.20) qui permettent de vérifier qu'elle admet une solution et une seule et de développer les procédés d'approximation numérique de sa solution.

Les **propriétés algébriques** suivantes sont de vérification immédiate.

Forme bilinéaire: L'application

$$(u,v) \in H^1(]0,L[) \longmapsto a(u,v) \in \mathbb{R}$$

est bilinéaire, i.e., $u \longmapsto a(u,v)$, pour v fixé, et $v \longmapsto a(u,v)$, pour u fixé, sont toutes les deux linéaires.

Forme linéaire: L'application

$$v \in H^1(]0,L[) \longmapsto Lv \in \mathbb{R}$$

est linéaire.

Ces formes ont en outre des **propriétés de continuité**. Ces propriétés sont importantes en pratique. Elles résultent directement du fait que les applications précédentes sont bien définies.

Continuité de la forme bilinéaire: Elle consiste à s'assurer qu'il existe une constante M, indépendante de u et v dans $H^1(]0,L[)$ telle que

$$|a(u,v)| \le M \|u\|_{1,]0,L[} \|v\|_{1,]0,L[}\,. \tag{1.21}$$

On va faire cette vérification ici pour voir sur un exemple comment elle résulte du sens qu'on a donné ci-dessus à l'écriture de la formulation variationnelle. En pratique, on ne s'attarde sur cette étude que si elle fait apparaître une *instabilité*, i.e. lorsque M dépend d'un paramètre et tend vers $+\infty$ lorsque ce pramètre tend vers une valeur limite.

Ici, on a clairement une décomposition de $a(u,v)$ sous la forme d'une somme de trois termes qu'on traite un à un.

On a d'abord par (1.4)

$$\begin{aligned} \left| \int_0^L a\partial_x u\partial_x v\, dx \right| &\le \int_0^L |a|\, |\partial_x u|\, |\partial_x v|\ dx \le \beta\, |u|_{1,]0,L[}\, |v|_{1,]0,L[} \\ &\le \beta\, \|u\|_{1,]0,L[}\, \|v\|_{1,]0,L[}\,. \end{aligned}$$

On a de même

$$\left|\int_0^L a_0 uv\,dx\right| \leq \beta_0\,|u|_{0,]0,L[}\,|v|_{0,]0,L[} \leq \beta_0\,\|u\|_{1,]0,L[}\,\|v\|_{1,]0,L[}\,.$$

Pour le troisième terme, on utilise la formule (1.16) qui donne

$$\begin{aligned}|u(L)| = \left|u(x) + \int_x^L \partial_x u(t)\,dt\right| &\leq |u(x)| + \int_x^L |\partial_x u(t)|\,dt\\ &\leq |u(x)| + \sqrt{L-x}\,|u|_{1,]0,L[}\,.\end{aligned}$$

On intègre une nouvelle fois entre 0 et L pour obtenir, par l'inégalité de Cauchy-Schwarz, continue puis discrète, une nouvelle fois,

$$L\,|u(L)| \leq \sqrt{L}\,|u|_{0,]0,L[} + \tfrac{2}{3}L^{\frac{3}{2}}\,|u|_{1,]0,L[} \leq \sqrt{L}\left(1+\tfrac{4}{9}L^2\right)^{1/2}\|u\|_{1,]0,L[}$$

soit l'estimation

$$|u(L)| \leq \left(\tfrac{4}{9}L + \frac{1}{L}\right)^{1/2}\|u\|_{1,]0,L[}\,.$$

A titre d'exercice, on pourra améliorer cette estimation pour $L \gg 1$.
On a alors (1.21) avec $M = \beta + \beta_0 + \lambda\left(\frac{4}{9}L + \frac{1}{L}\right)^{1/2}$.

Continuité de la forme linéaire: Elle consiste à s'assurer de l'existence d'une constante C, indépendant de v dans $H^1(]0,L[)$ telle que

$$|Lv| \leq C\,\|v\|_{1,]0,L[}\,. \tag{1.22}$$

Cette vérification s'effectue comme pour la forme bilinéaire ci-dessus.

En fait la propriété, **généralement non immédiate**, qui est **cruciale** et dont **il faut s'assurer**, est la coercivité. C'est une des caractéristiques des problèmes elliptiques.

La **coercivité** consiste à vérifier la propriété suivante dans V (et non dans $H^1(]0,L[)$ tout entier où elle peut être fausse !)

$$\exists\gamma > 0 : a(v,v) \geq \gamma\,\|v\|^2_{1,]0,L[}\,, \quad \forall v \in V. \tag{1.23}$$

Remarquons que cette inégalité n'est pas triviale. Par exemple, si v est constant, le premier membre peut être nul (pour a_0 et λ nuls !), alors que le second membre est > 0. Le fait que v soit dans V est fondamental.

Prenons donc $v \in V$. On a

$$a(v,v) = \int_0^L a\,|\partial_x v|^2\,dx + \int_0^L a_0\,|v|^2\,dx + \lambda\,|v(L)|^2 \geq \alpha\,|v|^2_{1,]0,L[} \tag{1.24}$$

car on a juste $a_0 \geq 0$ et $\lambda \geq 0$. Le second membre de (1.24) dépend seulement de la semi-norme d'ordre 1 et non la norme dans $H^1(]0,L[)$. Il s'annule sur les constantes par exemple. Le point fondamental dans la vérification de la coercivité est l'inégalité établie dans la proposition suivante.

PROPOSITION 1.1 (Inégalité de Poincaré). *Il existe une constante $C > 0$ indépendante de $v \in V$ telle que*

$$|v|_{0,]0,L[} \leq C\,|v|_{1,]0,L[}\,. \tag{1.25}$$

DÉMONSTRATION. On utilise là encore la relation (1.16) pour écrire

$$v(x) = \int_0^x \partial_x v(t)\, dt.$$

En utilisant l'inégalité de Cauchy-Schwarz, on obtient

$$|v(x)|^2 \leq x\ |v|^2_{1,]0,L[}\,.$$

En intégrant cette inégalité entre 0 et L, il vient alors

$$|v|^2_{0,]0,L[} \leq \frac{L^2}{2}\,|v|^2_{1,]0,L[}\,.$$

D'où l'inégalité avec $C = L/\sqrt{2}$. □

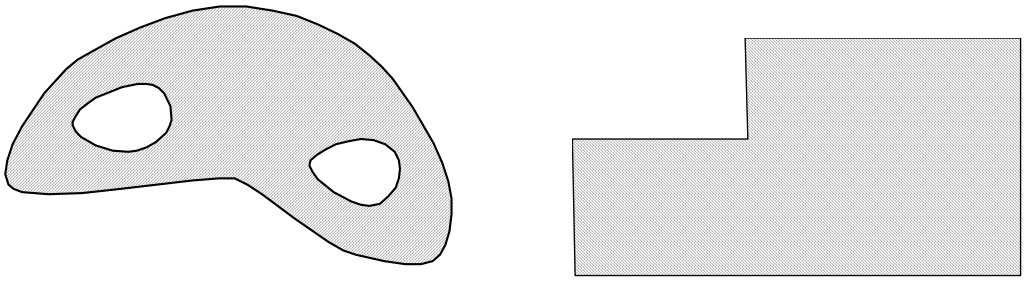
On déduit alors immédiatement la coercivité. A partir de (1.24), on écrit

$$a(v,v) \geq \frac{\alpha}{2}\,|v|^2_{1,]0,L[} + \frac{\alpha}{2}\,|v|^2_{1,]0,L[} \geq \frac{\alpha}{2}\,|v|^2_{1,]0,L[} + \frac{\alpha}{L^2}\,|v|^2_{0,]0,L[} = \alpha \min(\frac{1}{2}, \frac{1}{L^2})\,\|v\|^2_{1,]0,L[}$$

obtenant ainsi l'inégalité (1.23) avec $\gamma = \alpha \min(\frac{1}{2}, \frac{1}{L^2})$.

2. Problèmes en dimension supérieure

2.1. Position générale des problèmes aux limites. Le domaine où est posé le problème aux limites est maintenant un domaine borné Ω de $\mathbb{R}^N$, $N = 2$ ou 3. La géométrie peut être maintenant extrêmement variée et complexe comme le suggère les deux exemples suivants



Deux types de géométrie

La forme de la seconde figure se compose de segments parallèles aux axes. Elle peut être décrite assez simplement. La première forme est bien plus complexe.

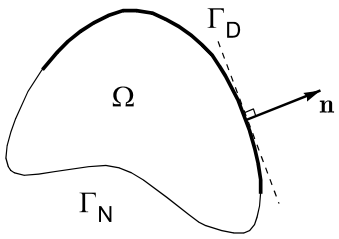
La position générale des problèmes aux limites est alors la suivante.

– **Equation aux dérivées partielles**

$$-\sum_{i,j=1}^{N} \partial_{x_i}\left(a_{ij}\partial_{x_j} u\right) + a_0 u = f \quad \text{dans } \Omega \tag{2.1}$$

où $\sum_{i,j=1}^N$ est la somme double $\sum_{i=1}^N \sum_{j=1}^N$.

– **Conditions aux limites**. Elles sont posées sur la frontière Γ du domaine Ω. On effectue une partition sans recouvrement de Γ en Γ_D et Γ_N (éventuellement Γ_D ou Γ_N peut être vide)



– *Conditions de Dirichlet*

$$u = g \qquad \text{sur } \Gamma_D. \tag{2.2}$$

– *Conditions de Neumann* (ou de Fourier-Robin pour $\lambda \neq 0$)

$$\sum_{i,j=1}^{N} a_{ij}\, \partial_{x_j} u\, n_i = h \qquad \text{sur } \Gamma_N, \tag{2.3}$$

où n_i est la composante relativement à l'axe x_i de la normale $\mathbf{n}$ unitaire à Γ orientée vers l'extérieur de Ω.

Pour les données, nous reprenons la classification de la dimension un mais maintenant g, h et λ sont des fonctions.

– **Coefficients de l'EDP**. Ce sont maintenant N^2+1 fonctions : a_{ij}, $i,j = 1,\ldots,N$, et a_0 qu'on supposera dans $L^\infty(\Omega)$ pour couvrir les besoins des applications en pratique. Nous supposerons que a_0 vérifie l'analogue de condition (1.5) en replaçant $]0,L[$ par le domaine Ω. L'inégalité (1.4) se généralise par l'**inégalité d'ellipticité suivante** :

$$\exists \alpha > 0 : \sum_{i,j=1}^{N} a_{ij}(x)\xi_j\xi_i \geq \alpha \sum_{i=1}^{N} \xi_i^2 \tag{2.4}$$

pour presque tout $x \in \Omega$ et tout système $\xi_1,\ldots,\xi_N$ de N nombres réels.

– **Coefficient dans la condition de Fourier-Robin.** La fonction λ joue un rôle sur Γ_N analogue à celui de a_0 sur Ω. On suppose qu'elle est dans $L^\infty(\Gamma_N)$ et qu'elle vérifie

$$\lambda(x) \geq 0, \qquad \text{pour presque tout } x \in \Gamma_N. \tag{2.5}$$

– **Second membres**.
  – *Second membre de l'EDP.* Comme en dimension un, on supposera simplement que f est dans $L^2(\Omega)$.
  – *Second membre de la condition de Neumann.* Si on n'a pas besoin de travailler avec les conditions minimales de régularité, il suffit de supposer que $h \in L^2(\Gamma_N)$.
  – *Second membre de la condition de Dirichlet.* La condition de Dirichlet est plus difficile à poser car la solution du problème aux limites n'est plus forcément continue comme en dimension un. Cependant, ces difficultés n'apparaissent pas au niveau du schéma de résolution numérique. Comme en dimension un, où la condition a été exprimée par $u - g_0$ s'annule en $\{x = 0\}$, nous supposerons que la fonction g est en fait définie sur Ω et qu'elle est dans la même classe fonctionnelle que la solution u. La condition $u = g$ sur Γ_D est alors prise au sens où $u - g$ s'annule sur Γ_D dans une signifiaction qui sera précisée par la suite.

Pour terminer sur la position du problème aux limites, notons qu'on condense souvent l'écriture en utilisant la notation matricielle. Ceci donne une forme pour l'équation (2.1)

et la condition aux limites (2.3) anolgue respectivement à l'équation (1.1) et à la condition (1.3). On note donc

- **matrice** A : matrice $N \times N$, de coefficients a_{ij},
- **gradient** de la solution : c'est le vecteur ∇u colonne de composante $\partial_{x_j} u$

$$\nabla u = \begin{bmatrix} \partial_{x_1} u \\ \vdots \\ \partial_{x_N} u \end{bmatrix},$$

- **divergence** $\nabla \cdot \mathbf{v}$ d'un vecteur colonne $\mathbf{v}$ à N composantes $v_1, \ldots, v_N$

$$\nabla \cdot \mathbf{v} := \partial_{x_1} v_1 + \cdots + \partial_{x_N} v_N,$$

- **produit scalaire** de deux vecteurs colonne $\mathbf{v}$ et $\mathbf{w}$, de composnates respectives $v_1, \ldots, v_N$ et $w_1, \ldots, w_N$,

$$\mathbf{v} \cdot \mathbf{w} = u_1 v_1 + \cdots + u_N v_N.$$

A l'aide de ces notations, le problème aux limites se réécrit

$$\begin{cases} -\nabla \cdot A \nabla u + a_0 uv = f \quad \text{dans } \Omega, \\ u = g \quad \text{sur } \Gamma_D, \qquad A\nabla u \cdot \mathbf{n} + \lambda u = h \quad \text{sur } \Gamma_N. \end{cases} \tag{2.6}$$

2.2. Quelques exemples de problèmes physiques. Nous regroupons dans la liste suivante quelques exemples de situations physiques en dimension 2 ou 3 modélisées par un problème aux limites elliptique du 2$^{\text{nd}}$ ordre.

Problème	Variable primaire u	Loi de comportement a	Terme source f	Variable secondaire $a\partial_x u$
Transfert thermique Dim. 2 ou 3	Température T	Conductivité thermique*	Apport calorifique Q	Flux thermique q : deux termes conduction $k\partial_{\mathbf{n}}T$ convection $h(T-T_\infty)$
Ecoulement irrotationnel d'un fluide parfait Dim. 2	Fonction courant ψ	1	création de masse normalement 0	Vitesse $v=(-\partial_y\psi,\partial_x\psi)$
Ecoulement irrotationnel d'un fluide parfait Dim. 2 ou 3	Potentiel de vitesse φ	1	Forces extérieures	Vitesse $v=(\partial_x\varphi,\partial_y\varphi)$
Ecoulement dans un milieu poreux Dim. 2 ou 3	Vitesse de filtration φ	perméablilité généralement matrice K	Injection ou extraction	Flux $q=K\nabla u\cdot n$ Filtration $v=-K\nabla u$
Electrostatique Dim. 2 ou 3	Potentiel électrique	Permittivité diélectrique ε^*	Densité de charges ϱ	Flux du champ électrique
Magnétostatique Dim. 2 ou 3	Potentiel magnétique	Perméabilité magnétique μ^*	Densité de charges magnétiques normalement 0	Flux du champ magnétique

* scalaire ou matrice dans le cas anisotrope.

2.3. Formulation variationnelle. On utilise là aussi une intégration par parties basée sur la formule de Green suivante

$$\int_\Omega \partial_{x_i} w\,dx = \int_\Gamma w\,n_i\,d\Gamma \tag{2.7}$$

qui donne ici

$$-\int_\Omega \partial_{x_i}\left(a_{ij}\partial_{x_j}u\right)v dx = \int_\Omega a_{ij}\partial_{x_j}u\partial_{x_i}v\,dx - \int_\Gamma v a_{ij}\partial_{x_j}u\,n_i\,d\Gamma.$$

En multipliant donc l'équation (2.1) par la fonction test v et en intégrant par parties, on obtient

$$\sum_{i,j=1}^N \int_\Omega a_{ij}\partial_{x_j}u\partial_{x_i}v\,dx + \int_\Omega a_0 uv\,dx - \sum_{i,j=1}^N \int_\Gamma v a_{ij}\partial_{x_j}u\,n_i\,d\Gamma = \int_\Omega fv\,dx.$$

En imposant $v = 0$ sur Γ_D et en utilisant la condition (2.3), on arrive à l'équation variationnelle

$$a(u, v) = Lv \tag{2.8}$$

avec

$$a(u, v) := \sum_{i,j=1}^{N} \int_\Omega a_{ij} \partial_{x_j} u \partial_{x_i} v \, dx + \int_\Omega a_0 uv \, dx + \int_{\Gamma_N} \lambda uv \, d\Gamma \tag{2.9}$$

$$Lv := \int_\Omega fv \, d\Omega + \int_{\Gamma_N} hv \, d\Gamma \tag{2.10}$$

2.4. Cadre fonctionnel de la formulation variationnelle. Tout comme en dimension un, pour que les intégrales sur Ω soient définies, il suffit que u et v soient dans l'espace de Sobolev

$$H^1(\Omega) := \left\{ v \in L^2(\Omega) \,;\; \partial_{x_j} v \in L^2(\Omega),\ j = 1, \ldots, N \right\}. \tag{2.11}$$

En adaptant la démonstration de la dimension un, on obtient directement que $H^1(\Omega)$ est un espace de Hilbert pour le produit scalaire

$$(u, v)_{H^1(\Omega)} := \sum_{j=1}^{N} \int_\Omega \partial_{x_j} u \partial_{x_j} v \, dx + \int_\Omega uv \, dx. \tag{2.12}$$

On note là encore

$$|v|_{0,\Omega} := \left\{ \int_\Omega |u|^2 \, dx \right\}^{1/2}, \tag{2.13}$$

$$|v|_{1,\Omega} := \left\{ \sum_{j=1}^{N} \int_\Omega \left| \partial_{x_j} u \right|^2 dx \right\}^{1/2}, \tag{2.14}$$

$$\|v\|_{1,\Omega} := \left\{ |v|_{1,\Omega}^2 + |v|_{0,\Omega}^2 \right\}^{1/2}. \tag{2.15}$$

On doit maintenant définir la condition aux limites de Dirichlet ou les intégrales sur Γ_N. Malheureusement, la propriété $H^1(\Omega) \subset \mathcal{C}^0(\overline{\Omega})$ n'est pas vérifiée en dimension > 1. On a cependant le théorème, dit de trace suivant, qui permet de donner un sens à la condition $v = 0$ sur Γ_D et à l'intégrale sur Γ_N. Les conditions d'application de ce théorème nécessitent un minimum de régularité pour la frontière Γ de Ω. Ces conditions sont difficiles à décrire en général bien qu'elles soient satisfaites dans la quasi-totalité des situations pratiques et plus encore dans celles où une résolution numérique est envisageable. C'est pourquoi nous supposerons implicitement qu'elles sont vérifiées.

THÉORÈME 1.2 (Théorème de trace). *L'application, définie pour $v \in H^1(\Omega) \cap \mathcal{C}^0(\overline{\Omega})$ par $v|_\Gamma$ se prolonge en une application linéaire, continue, de $H^1(\Omega)$ dans $L^2(\Gamma)$.*

La démonstration de théorème, assez compliquée, sera admise. En gros, ce théorème énonce que, pour v dans $H^1(\Omega)$, $v|_\Gamma$ a un sens comme fonction de $L^2(\Gamma)$. Observons qu'on ne peut pas définir $v|_\Gamma$ à partir de la seule donnée de v dans $L^2(\Omega)$.

Si v est dans $H^1(\Omega)$, on peut ainsi imposer à la fonction $v|_\Gamma$, qui est dans $L^2(\Gamma)$, de s'annuler sur Γ_D (presque partout pour la mesure $d\Gamma$). On peut ainsi définir le sous-espace

$$V := \left\{ v \in H^1(\Omega) \,;\; v|_{\Gamma_D} = 0 \right\}. \tag{2.16}$$

On peut alors démontrer le résultat suivant.

PROPOSITION 1.2. *Le sous-espace V est fermé dans $H^1(\Omega)$.*

DÉMONSTRATION. Elle est une simple conséquence du théorème 1.2. Si $\{v_n\}_{n\geq 0}$ est une suite contenue dans V telle que $\lim_{n\to\infty} v_n = v$ dans $H^1(\Omega)$, ce théorème donne

$$\int_{\Gamma_D} |v|^2 \, d\Gamma = \int_{\Gamma_D} |v_n - v|^2 \, d\Gamma \leq \int_{\Gamma} |v_n - v|^2 \, d\Gamma$$
$$\leq C^2 \, \|v_n - v\|^2_{1,\Omega} \, .$$

En passant à la limite, pour $n \to \infty$, sur le dernier terme, on obtient

$$\int_{\Gamma_D} |v|^2 \, d\Gamma = 0.$$

Ceci exprime exactement que $v \in V$. □

La condition de Dirichlet sera prise au sens suivant : il existe une fonction, toujours notée g, dans $H^1(\Omega)$ telle que

$$u - g \in V. \tag{2.17}$$

On peut donner alors la formulation variationnelle dans le cadre précis suivant

$$\begin{cases} u \in H^1(\Omega), \; u - g \in V, \quad \forall v \in V, \\ a(u, v) = Lv. \end{cases} \tag{2.18}$$

Cette formulation a les mêmes propriétés de continuité que la formulation en dimension un. Une extension de l'inégalité de Poincaré, qui demande maintenant, pour être établie, des éléments d'analyse fonctionnelle qui seront abordés dans les chapitres suivants, permet de s'assurer là aussi que la formulation est coercive.

CHAPITRE 2

Problèmes variationnels abstraits

Nous montrons dans une première section comment le cadre général fourni par le théorème de Lax-Milgram permet de démontrer un résultat d'existence et d'unicité pour la solution des problèmes aux limites du chapitre précédent. Nous présentons ensuite la méthode de Galerkin qui fournit le cadre abstrait d'étude de la convergence de l'approximation de la solution de ces problèmes aux limites par la méthode des éléments finis.

1. Théorie élémentaire des problèmes variationnels

1.1. Position du problème variationnel. Les formulations variationnelles du chapitre précédent nous ont permis de poser le problème aux limites de dimension un et en dimension supérieure sous la forme générale suivante

$$\begin{cases} u \in X,\ u - g \in V, \quad \forall v \in V, \\ a(u,v) = Lv. \end{cases} \tag{1.1}$$

où

- X est un espace de Hilbert dont nous noterons respectivement la norme et le produit scalaire par
$$\|u\|_X \quad \text{et} \quad (u,v)_X \quad \text{pour } u \text{ et } v \text{ dans } X,$$
- V est un sous-espace **fermé** de X,
- g est un élément donné dans X,
- $(u,v) \mapsto a(u,v)$ est une forme bilinéaire sur $X \times X$ vérifiant
  - **continuité sur** X : il existe une constante M telle que
$$|a(u,v)| \leq M \|u\|_X \|v\|_X, \quad \text{pour tout } u \text{ et tout } v \text{ dans } X, \tag{1.2}$$
  - **coercivité sur** V :
$$\exists \gamma > 0 : a(v,v) \geq \gamma \|v\|_X^2, \quad \text{pour tout } v \text{ dans } V, \tag{1.3}$$
- L est une forme linéaire continue sur X, autrement dit L est un élément donné dans le dual topologique X' de X.

On rappelle que X' est un espace de Hilbert pour la norme (dite duale)

$$\|L\|_{X'} = \sup_{\|v\|_{X'}=1} |Lv|. \tag{1.4}$$

Le théorème de Riesz affirme que l'application

$$J_X : X \longrightarrow X' \tag{1.5}$$

définie pour $z \in X$ par

$$J_X z(v) := (z,v)_X, \quad \text{pour tout } v \text{ dans } X, \tag{1.6}$$

est une **isométrie** de X sur X', i.e.,

$$J_X \text{ est un isomorphisme algébrique de } X \text{ sur } X', \tag{1.7}$$

qui vérifie

$$\|J_X z\|_{X'} = \|z\|_X \quad \text{pour tout } z \text{ dans } X. \tag{1.8}$$

1.2. Unicité et stabilité de la solution du problème variationnel. Le théorème suivant établit une inégalité vérifiée par une solution éventuelle du problème variationnel (1.1). Une telle inégalité est appelée **inégalité a priori**. De telles inégalités sont souvent à la base des propriétés d'existence, d'unicité et de dépendance continue par rapport aux données du second membre.

THÉORÈME 2.1. *Toute solution u du problème (1.1) vérifie*

$$\|u\|_X \leq \frac{1}{\gamma}\|L\|_{X'} + \left(1 + \frac{M}{\gamma}\right)\|g\|_X \tag{1.9}$$

et constitue donc l'unique solution de ce problème.

DÉMONSTRATION. On a, si u est une solution du problème (1.1),

$$a(u-g,v) = a(u,v) - a(g,v) = Lv - a(g,v), \quad \forall v \in V.$$

Comme $u - g \in V$, en utilisant successivement la coercivité, la définition de la norme duale et la continuité de la forme bilinéaire $a(\cdot,\cdot)$, on obtient

$$\begin{aligned}\gamma\|u-g\|_X^2 &\leq a(u-g,u-g) = L(u-g) - a(g,u-g)\\ &\leq \|L\|_{X'}\|u-g\|_X + M\|g\|_X\|u-g\|_X.\end{aligned}$$

Si $u = g$, l'inégalité (1.9) est trivialement vérifiée. Sinon, en divisant par $\gamma\|u-g\|_X$, on arrive à

$$\|u-g\|_X \leq \frac{1}{\gamma}\|L\|_{X'} + \frac{M}{\gamma}\|g\|_X.$$

L'inégalité triangulaire donne alors

$$\|u\|_X = \|u-g+g\|_X \leq \|u-g\|_X + \|g\|_X$$

et conduit ainsi à l'inégalité (1.9).

Supposons maintenant que u_1 et u_2 sont des solutions correspondant respectivement aux données L_1, g_1 et L_2, g_2. Il est immédiat que $u_1 - u_2$ est solution du problème (1.9) relativement aux données $L_1 - L_2$ et $g_1 - g_2$. L'inégalité (1.9) montre alors que

$$\|u_1 - u_2\|_X \leq \frac{1}{\gamma}\|L_1 - L_2\|_{X'} + \left(1 + \frac{M}{\gamma}\right)\|g_1 - g_2\|_X. \tag{1.10}$$

Le second membre de l'inégalité (1.10) est nul pour $L_1 = L$, $g_1 = g$, $L_2 = L$ et $g_2 = g$ et conduit ainsi à

$$\|u_1 - u_2\|_X \leq 0$$

qui induit $u_1 = u_2$. □

REMARQUE 2.1. *Prenons $L_1 = L$, $L_2 = L + \Delta L$, $g_1 = g$ et $g_2 = g + \Delta g$ où ΔL et Δg sont deux perturbations des données. L'inégalité (1.9) et son écriture (1.10) donnent, si on note $u_1 = u$ et $u_2 = u + \Delta u$,*

$$\|\Delta u\|_X \leq \frac{1}{\gamma}\|\Delta L\|_{X'} + \left(1 + \frac{M}{\gamma}\right)\|\Delta g\|_X.$$

Autrement dit, la perturbation Δu résultant de petite perturbations ΔL sur L et Δg sur g reste petite si la constante de coercivité γ n'est pas trop faible et si la constante de continuité n'est pas trop forte. Les deux constantes M et $1/\gamma$ mesurent ainsi la stabilité du problème. Si l'une de ces constantes vient à exploser, la résolution devient délicate.

1.3. Existence d'une solution au problème variationnel. On va se ramener au conditions d'application duthéorème de Lax-Milgram. Pour celà, on va faire le changement d'inconnue suivant

$$u_0 = u - g. \tag{1.11}$$

On est ainsi ramené à étudier l'existence d'une solution pour le problème variationnel

$$\begin{cases} u_0 \in V, \quad \forall v \in V, \\ a(u_0, v) = L_0 v, \end{cases} \tag{1.12}$$

où L_0 est la forme linéaire continue sur X, et à fortiori sur V, donnée par

$$L_0 v := Lv - a(g, v), \quad \forall v \in X. \tag{1.13}$$

Comme V est un sous-espace fermé de X, c'est un espace de Hilbert pour le produit scalaire $(\cdot,\cdot)_V$ induit par celui $(\cdot,\cdot)_X$ de X. On notera aussi provisoirement par $\|\cdot\|_V$ la norme induite par celle de X. L'existence d'une solution au problème (1.12) constitue le théorème de Lax-Milgram.

THÉORÈME 2.2 (Théorème de Lax-Milgram). *Sous les conditions de continuité (1.2) et de coercivité (1.3) pour la forme bilinéaire $a(\cdot,\cdot)$, pour toute forme linéaire continue L_0 sur V, le problème variationnel (1.12) admet une solution et une seule.*

DÉMONSTRATION. Comme $(u, v) \mapsto a(u, v)$ est une forme bilinéaire continue sur $V \times V$, le théorème de Riesz montre que l'identification

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

définit un opérateur linéaire continu de V dans V.

La coercivité et l'inégalité de Cauchy-Schwarz donnent

$$\gamma \|v\|_V^2 \leq a(v, v) = (Av, v)_V \leq \|Av\|_V \|v\|_V ,$$

d'où

$$\gamma \|v\|_V \leq \|Av\|_V . \tag{1.14}$$

On va montrer que cette inégalité entraine que l'opérateur est injectif et que son image

$$R(A) := \{w \in V \,;\, \exists v \in V,\ Av = w\}$$

est fermée.

- A est injectif : si $Av = 0$ alors $\|Av\|_V = 0$ et donc, par (1.14), $\|v\|_V = 0$ et ainsi $v = 0$.
- $R(A)$ est fermé : supposons que $\lim_{n\to\infty} Av_n = z$; la suite est donc une suite de Cauchy. Appliquons alors l'inégalité (1.14) à l'élément $v_n - v_m$

  $$\|v_n - v_m\|_V \leq \frac{1}{\gamma} \|A(v_n - v_m)\|_V = \frac{1}{\gamma} \|Av_n - Av_m\|_V .$$

  On en déduit alors que la suite $\{v_n\}_{n\geq 0}$ est elle-même une suite de Cauchy. Comme V est complet, il existe ainsi $v \in V$ tel que $\lim_{n\to\infty} v_n = v$. La continuité de A entraine que

  $$\lim_{n\to\infty} Av_n = Av$$

  et donc que $Av = z$. Ceci montre que $z \in R(A)$ et donc que $R(A)$ est fermé.

Le problème (1.12) admettra une solution si $R(A) = V$. Comme $R(A)$ est fermé, il suffit de montrer que si $w \in V$ est orthogonal à tout $R(A)$, il est obligatoirement égal à 0. Soit donc un tel w

$$(Av, w)_V = 0, \quad \forall v \in V.$$

En particulier, on a

$$(Aw, w)_V = a(w, w) = 0.$$

La coercivité donne alors $w = 0$. □

2. La méthode de Galerkin

La méthode des éléments finis est une méthode de Galerkin particulière mais avec des propriétés spécifiques qui la rendent spécialement attractive dans la résolution des problèmes aux limites. Nous allons la présenter ici comme un cadre général abstrait permettant l'étude de la convergence de l'approximation par éléments finis qui sont données dans le chapitre suivant.

2.1. Principe de base d'une méthode de Galerkin. Reprenons le problème abstrait (1.1). Le but est de ramener sa résolution à celle d'un **problème discret**, équivalente à la recherche d'un **nombre fini** de paramètres réels.

Le principe d'une méthode de Galerkin est d'approcher les éléments de l'espace X par ceux d'un sous-espace X^h de **dimension finie**. C'est donc, à la base, la construction d'un processus d'approximation des éléments d'un espace de Hilbert X. L'exposant h est un paramètre réel > 0 caractérisant la discrétisation et tendant vers 0 au fur et à mesure que la discrétisation devient de plus en plus fine. Cette convention est utile pour les méthodes d'éléments finis où ce paramètre a un sens géométrique précis.

La propriété d'approximation est décrite par la condition suivante dit d'**approximation interne**

$$\forall v \in X, \quad \lim_{h \to 0} \inf_{v^h \in X^h} \left\| v - v^h \right\|_X = 0. \tag{2.1}$$

La quantité $\inf_{v^h \in X^h} \left\| v - v^h \right\|_X$ est la distance de v à X^h. Comme X^h est un espace de dimension finie, dès que X est un **espace normé** (même s'il n'est pas complet), cette distance est caractérisée de la façon suivante.

– Il existe $z^h \in X^h$ tel que

$$\left\| v - z^h \right\|_X = \inf_{v^h \in X^h} \left\| v - v^h \right\|_X \tag{2.2}$$

et z^h est appelé **la meilleure approximation** de v par les éléments de X^h.

– La propriété d'approximation interne peut ainsi être décrite de façon équivalente :

$$\forall v \in X, \quad \exists \left\{ w^h \right\}_{h>0} : \begin{cases} w^h \in X^h, \quad \forall h > 0 \\ \lim\limits_{h \to 0} w^h = v \text{ dans } X \end{cases} \tag{2.3}$$

– Si X est un **espace préhilbertien**, i.e. si sa norme est associée à un produit scalaire, la meilleure approximation est unique : c'est la **projection** de v sur X^h. Elle est caractérisée par les **équations** dites **normales**

$$\begin{cases} z^h \in X^h, \quad \forall v^h \in X^h, \\ \left(z^h, v^h \right)_X = \left(v, v^h \right)_X . \end{cases} \tag{2.4}$$

Le problème variationnel fait intervenir aussi le sous-espace vectoriel V de X. Une façon simple, mais qui ne fonctionne pas toujours, est d'introduire le sous-espace de X^h

$$V^h := X^h \cap V. \tag{2.5}$$

Pour $v \in V$, on a cependant seulement

$$\inf_{v^h \in X^h} \left\|v - v^h\right\|_X \leq \inf_{v^h \in V^h} \left\|v - v^h\right\|_X$$

puisqu'on prend la borne inférieure sur un espace plus restreint. On est amené à faire l'hypothèse supplémentaire que la famille $\left\{V^h\right\}_{h>0}$ constitue **une approximation interne** de V

$$\lim_{h\to 0} \inf_{v^h \in V^h} \left\|v - v^h\right\|_X . \tag{2.6}$$

2.2. Le problème discret. En supposant donc que pour tout $h > 0$, on se donne $g^h \in X^h$ tel que

$$\lim_{h\to 0} g^h = g \text{ dans } X, \tag{2.7}$$

la méthode de Galerkin consiste simplement à mimer le problème (1.1) dans les sous-espaces de dimension finie X^h et V^h

$$\begin{cases} u^h \in X^h,\ u^h - g^h \in V^h, & \forall v^h \in V, \\ a(u^h, v^h) = Lv^h. \end{cases} \tag{2.8}$$

Le théorème de Lax-Milgram assure bien sûr là aussi que ce problème possède une solution et une seule avec **la propriété de stabilité** suivante

$$\left\|u^h\right\|_X \leq \frac{1}{\gamma} \|L\|_{X'} + (\frac{M}{\gamma} + 1) \left\|g^h\right\|_X . \tag{2.9}$$

L'intérêt de la méthode de Galerkin vient du fait que la résolution du problème variationnel peut être ramenée à celle d'**un système linéaire**. Pour celà, fixons une base $\left\{B_i^h\right\}_{i=}^{i=N^h}$ de V^h. On voit immédiatement pourquoi le problème (2.8) est un **système discret** : sa résolution équivaut à la détermination des coefficients $x_1, \ldots, x_{N^h}$ de $u^h - g^h$ dans la base $\left\{B_i^h\right\}_{i=1}^{i=N^h}$

$$u^h = g^h + x_1 B_1^h + \cdots + x_{N^h} B_{N^h}^h. \tag{2.10}$$

Il sera commode d'organiser les coefficients $x_1, \ldots, x_{N^h}$ suivant un vecteur colonne noté x. De même, la donnée de v^h dans V^h équivaut à celle de ses composantes $y_1, \ldots, y_{N^h}$, organisées suivant un vecteur colonne y

$$v^h = y_1 B_1^h + \cdots + y_{N^h} B_{N^h}^h. \tag{2.11}$$

La restriction de $a(\cdot, \cdot)$ à $V^h \times V^h$ est bien sûr une forme bilinéaire sur ce dernier. Comme V^h est de dimension finie, toute forme bilinéaire sur $V^h \times V^h$ est complètement caractérisée par une matrice. Dans le cas présent, cette matrice A est une matrice $N^h \times N^h$ dont les coefficients sont donnés par

$$A_{ij} = a(B_j^h, B_i^h). \tag{2.12}$$

On alors

$$a(x_1 B_1^h + \cdots + x_{N^h} B_{N^h}^h, y_1 B_1^h + \cdots + y_{N^h} B_{N^h}^h) = \sum_{i,j=1}^{N^h} A_{ij} x_j y_i = y^\top A x. \tag{2.13}$$

De même, la forme linéaire L restreinte à V^h est complètement caractérisée par un vecteur colonne ℓ de composantes

$$\ell_i := LB_i^h, \quad i = 1, \ldots, N^h, \tag{2.14}$$

et la relation

$$Lv^h = y^T \ell. \tag{2.15}$$

Exactement, selon la même démarche, le nombre $a(g^h, v^h)$ s'écrit de façon matricielle sous la forme

$$a(g^h, v^h) = y^T g \tag{2.16}$$

où g est le vecteur colonne à N^h composantes

$$a(g^h, B_i^h) = g_i \quad i = 1, \ldots, N^h. \tag{2.17}$$

Compte tenu des précédentes relations, en posant en outre

$$b = \ell - g, \tag{2.18}$$

le problème (2.8) se réécrit sous forme matricielle

$$\left\{ \begin{array}{l} x \in \mathbb{R}^{N^h}, \quad \forall y \in \mathbb{R}^{N^h}, \\ y^\top A x = y^\top b. \end{array} \right. \tag{2.19}$$

Clairement, en prenant $y = Ax - b$, on vérifie que le problème variationnel (2.19) est équivalent au système linéaire

$$Ax = b. \tag{2.20}$$

On alors le théorème suivant qui est de première importance pour les applications.

THÉORÈME 2.3. *Sous la condition de coercivité (1.3), la matrice A du système linéaire (2.20) est inversible. De plus, elle admet la décomposition*

$$A = LU \tag{2.21}$$

où L est une matrice triangulaire avec des coefficients diagonaux égaux à 1 et U une matrice triangulaire. De plus, si la forme $a(\cdot, \cdot)$ est symétrique, i.e. vérifie

$$a(u, v) = a(v, u), \quad \forall u, \forall v \text{ dans } X, \tag{2.22}$$

alors, la matrice A est symétrique définie positive.

DÉMONSTRATION. Montrer que A est inversible et admet la décomosition (2.21) revient à s'assurer que les sous-matrices principales $A^{(p)} \in \mathbb{R}^{p \times p}$ données par

$$A_{ij}^{(p)} = A_{ij}, \quad i, j = 1, \ldots p,$$

sont toutes inversibles. Soit $y^{(p)} \in \mathbb{R}^p$ tel que $A^{(p)} y^{(p)} = 0$. En complétant éventuellement les composantes de $y^{(p)}$ par des 0, on a donc

$$\left(y^{(p)}\right)^\top A^{(p)} y^{(p)} = \left[\begin{array}{cc} \left(y^{(p)}\right)^\top & 0 \end{array} \right] A \left[\begin{array}{c} y^{(p)} \\ 0 \end{array} \right] = 0.$$

Notant $v^h = \sum_{j=1}^{j=N^h} y_j B_j^h$, on a donc

$$a(v^h, v^h) = \left[\begin{array}{cc} \left(y^{(p)}\right)^\top & 0 \end{array} \right] A \left[\begin{array}{c} y^{(p)} \\ 0 \end{array} \right] = 0.$$

La coercivité donne alors $v^h = 0$ et par suite $y_j = 0$ pour $j = 1, \ldots, p$, et par suite $y^{(p)} = 0$.

Si $(u,v) \mapsto a(u,v)$ est symétrique, on a

$$A_{ij} = a(B_j^h, B_i^h) = a(B_i^h, B_j^h) = A_{ji}$$

et donc A est symétrique. La coercivité donne une nouvelle fois

$$y^\top A y = a(v^h, v^h) \geq \gamma \left\|v^h\right\|_X^2 > 0 \text{ si } y \neq 0.$$

□

REMARQUE 2.2. *La propriété (2.21) équivaut au fait que le système linéaire (2.20) peut être résolu par l'algorithme de Gauss sans échange de ligne ou de colonne. Cette propriété sera cruciale pour résoudre à moindre coût dans certaines situations les systèmes résultant de la discrétisation par une méthode d'éléments finis.*

2.3. Convergence d'une méthode de Galerkin. Le point-clé dans l'étude de la convergence d'une méthode de Galerkin est le résultat simple mais important suivant.

LEMME 2.1 (Lemme de Céa). *Soient u et u^h les solutions respectives des problèmes continu (1.1) et discret (2.8) ; alors,*

$$a(u^h - u, w^h) = 0, \quad \forall w^h \in V^h. \tag{2.23}$$

DÉMONSTRATION. Soit $w^h \in V^h$, comme $V^h \subset V$, on a simplement

$$a(u^h, w^h) = Lw^h = a(u, w^h).$$

□

PROPOSITION 2.1 (Estimation de l'erreur de résolution). *L'erreur de résolution vérifie l'estimation suivante*

(2.24)

$$\underbrace{\left\|u - u^h\right\|_X}_{\substack{\textit{Erreur de} \\ \textit{résolution}}} \leq \left(1 + \frac{M}{\gamma}\right) \left(\underbrace{\inf_{v^h \in V^h} \left\|u - g - v^h\right\|_X}_{\substack{\textit{Erreur d'approximation} \\ \textit{de la solution}}} + \underbrace{\left\|g - g^h\right\|_X}_{\substack{\textit{Erreur sur la} \\ \textit{donnée de Dirichlet}}} \right)$$

La solution du problème discret converge donc si la famille $\left\{V^h\right\}_{h>0}$ est une approximation interne de V et si

$$\lim_{h \to 0} g^h = g \text{ dans } X.$$

DÉMONSTRATION. Soit $v^h \in V^h$. La coercivité de $a(\cdot,\cdot)$ sur V^h donne

$$\gamma \left\|u^h - g^h - v^h\right\|_X^2 \leq a(u^h - g^h - v^h, u^h - g^h - v^h). \tag{2.25}$$

L'égalité (2.23), démontrée ci-dessus dans le lemme de Céa, permet alors d'écrire

$$a(u^h - g^h - v^h, u^h - g^h - v^h) = a(u - g^h - v^h, u^h - g^h - v^h),$$

soit en ajoutant et en retranchant g

$$\begin{aligned} a(u^h - g^h - v^h, u^h - g^h - v^h) &= a(u - g - v^h, u^h - g^h - v^h) \\ &\quad + a(g - g^h, u^h - g^h - v^h). \end{aligned} \tag{2.26}$$

La continuité de la forme bilinéaire $a(\cdot,\cdot)$ donne alors

$$a(u-g-v^h, u^h-g^h-v^h) \leq M \left\|u-g-v^h\right\|_X \left\|u^h-g^h-v^h\right\|_X , \tag{2.27}$$

$$a(g-g^h, u^h-g^h-v^h) \leq M \left\|g-g^h\right\|_X \left\|u^h-g^h-v^h\right\|_X . \tag{2.28}$$

On déduit alors de (2.25), (2.26), (2.27) et (2.28),

$$\left\|u^h-g^h-v^h\right\|_X \leq \frac{M}{\gamma}\left(\left\|u-g-v^h\right\|_X + \left\|g-g^h\right\|_X\right). \tag{2.29}$$

L'inégalité triangulaire permet alors d'écrire

$$\begin{aligned}\left\|u-u^h\right\|_X &= \left\|u-g-v^h+g-g^h-(u^h-g^h-v^h)\right\|_X \\ &\leq \left\|u-g-v^h\right\|_X + \left\|g-g^h\right\|_X + \left\|u^h-g^h-v^h\right\|_X .\end{aligned}$$

L'inégalité (2.29) permet alors d'arriver à (2.24).

La convergence suit de l'approximation de $u-g$ par les éléments de V^h et de g par g^h. □

REMARQUE 2.3. *Si la constante M/γ devient grande devant 1, l'erreur de résolution peut être importante même si l'erreur d'approximation de g et de $u-g$ est faible. Typiquement, on observe alors un cas d'instabilité de la méthode.*

CHAPITRE 3

Introduction à la méthode des éléments finis

Nous allons dans ce chapitre présenter les méthodes d'éléments finis les plus simples, qui restent aussi les plus utilisées, pour la résolution des problèmes aux limites des chapitres précédents. Cela nous permettra d'introduire les principes permettant de construire des méthodes de ce type plus générales.

1. Principes de base

1.1. Elément géométrique. La résolution des problèmes (1.20) et (2.18) du chapitre I par une méthode d'éléments finis est basée sur la construction d'une **approximation interne** de $H^1(\Omega)$ constituée de **fonctions polynomiales par morceaux**. Plus précisément, on utilise une partition finie sans recouvrement $\mathcal{T}^h$ (la signification du paramètre $h > 0$ sera donnée plus loin) du domaine de calcul Ω en sous-domaines **ouverts** T de "**forme simple**"

$$\overline{\Omega} = \bigcup_{T \in \mathcal{T}^h} \overline{T} \tag{1.1}$$

$$T \cap L = \emptyset \text{ si } T \neq L \tag{1.2}$$

Cette partition est sujette à une condition de compatibilité qui sera précisée plus loin. Elle est appelée **maillage** de Ω. On note

$$h_T := \operatorname{diam} T \tag{1.3}$$

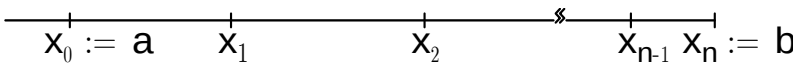
le diamètre de T (i.e. $h_T := \sup_{x,y \in T} |x - y|$), et

$$h := \max_{T \in \mathcal{T}^h} h_T \tag{1.4}$$

le pas du maillage.

EXEMPLE 3.1. *Maillages d'un intervalle et d'un pavé*

(1) ***Maillage d'un intervalle de*** $\mathbb{R}$ *:* $\Omega :=]a, b[$



Un maillage en dimension 1 correspond seulement à la donnée d'une grille de points

$$\{x_0 := a < x_1 < \cdots < x_{n-1} < x_n := b\}$$

non nécessairement répartie de façon uniforme dans $[a, b]$*. Les éléments géométriques sont des sous-intervalles*

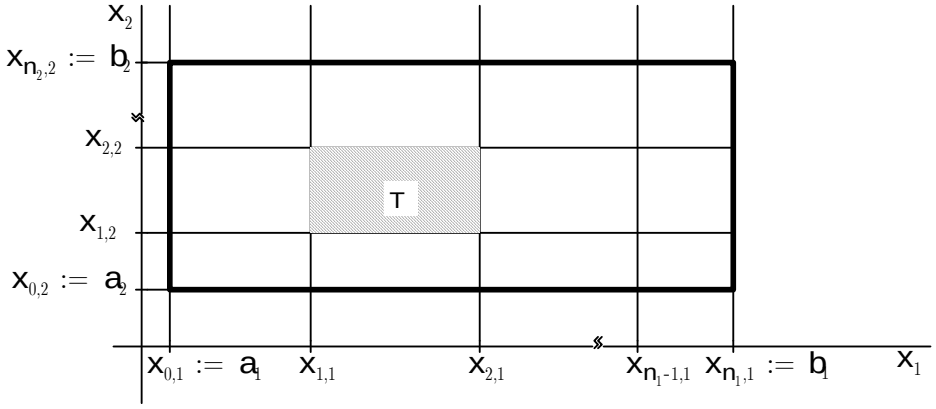
$$T :=]x_i, x_{i+1}[\,.$$

(2) ***Maillage d'un pavé de*** $\mathbb{R}^N$ *:* $\Omega := \prod_{k=1}^{N}]a_k, b_k[$. *Pour chaque* k*, on se donne une grille sur l'axe des coordonnées* x_k

$$\{x_{0,k} := a_k < x_{1,k} < \cdots < x_{n_k-1,k} < x_{n_k,k} := b_k\}.$$

On obtient un maillage par

$$T := \prod_{k=1}^{N}]x_{i_k,k}, x_{i_k+1,k}[$$



D'autres maillages, de forme plus générale, seront introduits par la suite.

1.2. Fonctions de forme. La seconde étape est la construction d'une approximation interne de $H^1(\Omega)$ en considérant des fonctions approchantes $v \in L^\infty(\Omega)$ de forme "simple" sur chaque élément géométrique

$$v|_T = v_T \in \mathbb{P}_T, \quad \forall T \in \mathcal{T}^h, \tag{1.5}$$

où $\mathbb{P}_T$ est un espace de dimension finie de fonctions dans $\mathcal{C}^\infty(\overline{T})$. Dans la quasi-totalité des cas, $\mathbb{P}_T$ est un espace de polynômes. Le plus souvent, on utilise l'espace $\mathbb{P}_m^{(N)}$ des polynômes à N indéterminées —N est la dimension du problème, i.e., le nombre de variables des fonctions à approcher—

$$\mathbb{P}_m^{(N)} := \left\{ p \in \mathcal{C}^\infty(\mathbb{R}^N);\ p(x) = \textstyle\sum_{|\alpha|\le m} a_\alpha x^\alpha,\ a_\alpha \in \mathbb{R} \right\}$$

avec les notations suivantes :
– multi-indice : $\alpha := (\alpha_1, \ldots, \alpha_N)$, $\alpha_i \in \mathbb{N}$ $(i = 1, \ldots, N)$,
– monôme : $x^\alpha := x_1^{\alpha_1} \cdots x_N^{\alpha_N}$.

EXEMPLE 3.2. *Polynômes à une, deux et trois indéterminées.*

(1) ***Polynômes à une indéterminée***

$$p(x) = a_0 + a_1 x + \cdots + a_m x^m, \quad a_i \in \mathbb{R}.$$

On a, de façon élémentaire,

$$\dim \mathbb{P}_m^{(1)} = m + 1.$$

(2) ***Polynômes à deux indéterminées***

$$p(x) = \textstyle\sum_{i+j\le N} a_{ij} x_1^i x_2^j.$$

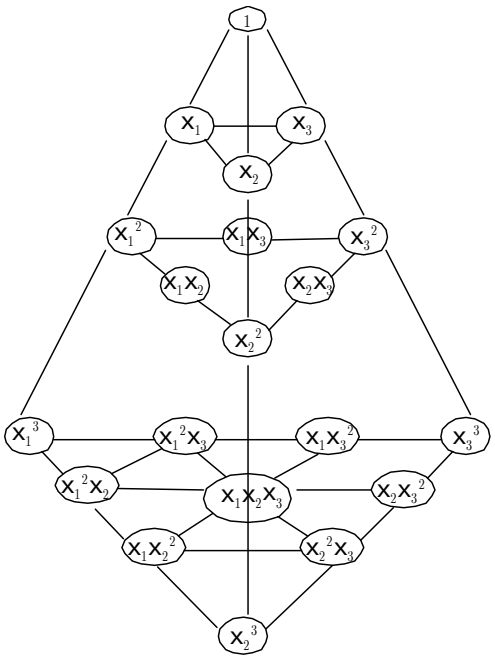


FIG. 1. Représentation des monômes à trois indéterminées à l'aide d'un tétraèdre

Les différents monômes peuvent être décrits par un tableau triangulaire (de type triangle de Pascal)

degré					
0	*1*				
1	x_1	x_2			
2	x_1^2	x_1x_2	x_2^2		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	
m	x_1^m	$x_1^{m-1}x_2$	$x_1^{m-2}x_2^2$	$\cdots$	x_2^m

On a directement en comptant le nombre de monômes du tableau précédent

$$\dim \mathbb{P}_m^{(2)} = (m+1)\,(m+2)\,/2.$$

(3) ***Polynômes à trois indéterminées***

$$p(x) = \textstyle\sum_{i+j+k\le N} a_{ijk} x_1^i x_2^j x_3^k.$$

Il y a autant de monômes $x_1^i x_2^j x_3^k$ de degré total $\ell = i+j+k$ que de monômes $x_1^i x_2^j$ degré $i+j \le \ell$. On a donc la formule suivante

$$\dim \mathbb{P}_m^{(3)} = \textstyle\sum_{\ell=0}^N \dim \mathbb{P}_\ell^{(2)} = \frac{1}{2}\sum_{\ell=0}^N (\ell+1)\,(\ell+2) = \frac{1}{6}(m+1)(m+2)(m+3) \tag{1.6}$$

qu'on pourra vérifier facilement par récurrence. La figure (Fig. 1) montre que les monômes à trois indéterminées peuvent être représentés à l'aide d'un tétraèdre qui est la généralisation à l'espace tridimensionnel du triangle.

1.3. Conditions de raccord. Il faut que les fonctions définies par les conditions (1.5) satisfassent des propriétés de continuité globales, appelées **conditions de raccord**, pour que la fonction $v \in L^\infty(\Omega)$ correspondante soit dans $H^1(\Omega)$. Par exemple, les fonctions en escalier

$$v|_{]x_i,x_{i+1}[} = v_i \in \mathbb{R}, \quad \text{pour } i = 0,\dots,m-1,$$

relativement au maillage ci-dessus du segment $]a, b[$ ont pour dérivée

$$v' = \sum_{i=1}^{m-1}(v_{i+1} - v_i)\delta_{x_i}, \tag{1.7}$$

où δ_{x_i} est la masse de Dirac au point x_i, et ne sont donc pas dans $H^1(]a, b[)$.

Afin de donner le théorème qui précise les conditions de raccord que doivent satisfaire les différentes expressions v_T d'une fonction v donnée par (1.5) pour donner lieu à une fonction dans $H^1(\Omega)$, nous avons besoin auparavant de la **notion d'interface** entre deux éléments géométriques.

Pour T et L deux éléments dans $\mathcal{T}^h$, on appelle interface entre T et L, l'ensemble $G' := \overline{T} \cap \overline{L}$ lorsqu'il vérifie la condition suivante

(1) $G' = \{a\}$ en dimension $N = 1$ (i.e. si Ω est un intervalle de $\mathbb{R}$),

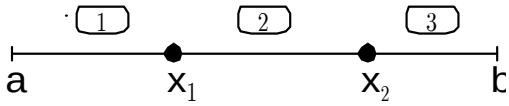
(2) $|G'| > 0$ en dimension $N > 1$, où $|G'|$ est la mesure de G' comme partie du bord T' de T (ou L' de L).

REMARQUE 3.1. *Comme le maillage $\mathcal{T}^h$ de Ω est une partition sans recouvrement, on a*

$$\overline{T} \cap \overline{L} = T' \cap L'.$$

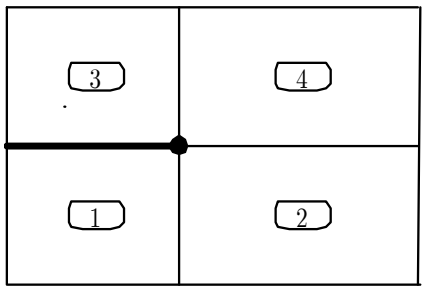
EXEMPLE 3.3. *Interfaces de sous-domaines dans les cas mono et multidimensionnels.*

(1) ***Interface entre intervalles.***



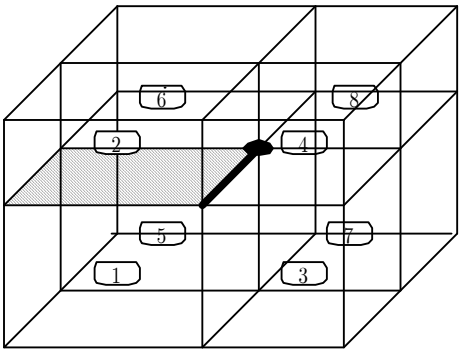
Interface entre l'élément $\boxed{1}$ *et* $\boxed{2}$ *:* $\{x_1\}$
Interface entre l'élément $\boxed{1}$ *et* $\boxed{3}$ *:* $\{x_2\}$

(2) ***Interface entre deux polygones.***



Segment épais : interface entre l'élément $\boxed{1}$ *et* $\boxed{3}$.
Le point • *n'est pas une interface entre l'élément* $\boxed{1}$ *et* $\boxed{4}$.

(3) ***Interface entre deux polyèdres.***



Face hachurée : interface entre l'élément $\boxed{1}$ *et* $\boxed{2}$.
L'arête (segment épais) n'est pas une interface entre l'élément $\boxed{1}$ *et* $\boxed{4}$.
Le point $\bullet$ *n'est pas une interface entre* $\boxed{1}$ *et* $\boxed{8}$.

THÉORÈME 3.1. *Soit* $v \in L^\infty(\Omega)$ *vérifiant (1.5) ; alors,* $v \in H^1(\Omega)$ *si et seulement si* $v \in \mathcal{C}^0(\overline{\Omega})$.

DÉMONSTRATION. La démonstration, dans le cas $N = 1$, s'obtient simplement à l'aide de la formule des sauts (de façon analogue à (1.7)). Nous pouvons donc nous concentrer sur le cas $N \geq 2$.

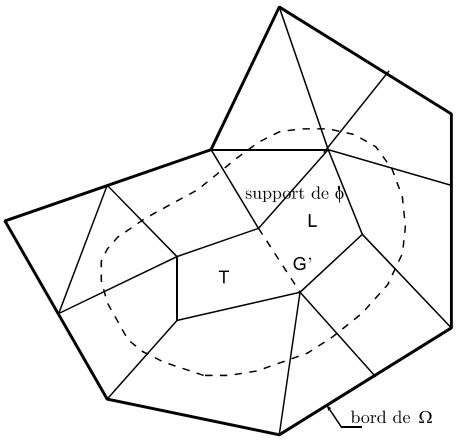
Comme $v \in L^\infty(\Omega)$ et que Ω est borné, on a directement que $v \in L^2(\Omega)$. On va calculer $\partial_{x_j} v$ au sens des distributions et montrer qu'il est dans $L^2(\Omega)$. Soit $\varphi \in \mathcal{D}(\Omega)$. On a

$$\langle \partial_{x_j} v, \varphi \rangle = - \langle v, \partial_{x_j} \varphi \rangle = - \int_\Omega v \partial_{x_j} \varphi \, d\Omega = - \sum_{T \in \mathcal{T}^h} \int_T v_T \partial_{x_j} \varphi \, dT. \tag{1.8}$$

Or v_K et φ sont dans $\mathcal{C}^1(\overline{T})$. La formule de Green dans T donne alors (n_j est la composante j de la normale $\mathbf{n}$ à T' orientée vers l'extérieur de T)

$$- \int_T v_T \partial_{x_j} \varphi \, dT = - \int_{T'} v_T \, dT' + \int_T \partial_{x_j} v_T \, \varphi \, dT. \tag{1.9}$$

Comme $\varphi = 0$ sur le bord de Ω, l'intégrale sur T' se réduit seulement à l'intégrale sur les interfaces séparant T de ses éléments adjacents.



D'où si l'on définit presque partout $g_j \in L^\infty(\Omega)$ par

$$g_j|_T = \partial_{x_j} v_T, \quad \text{pour tout } T \in \mathcal{T}^h,$$

on peut réécrire (1.8) en utilisant (1.9) sous la forme

$$\langle \partial_{x_j} v, \varphi \rangle = \int_\Omega g_j \varphi \, d\Omega - \sum_{\text{interfaces}} \int_{G'} (v_T n_j^T + v_L n_j^L) \varphi \, dG'.$$

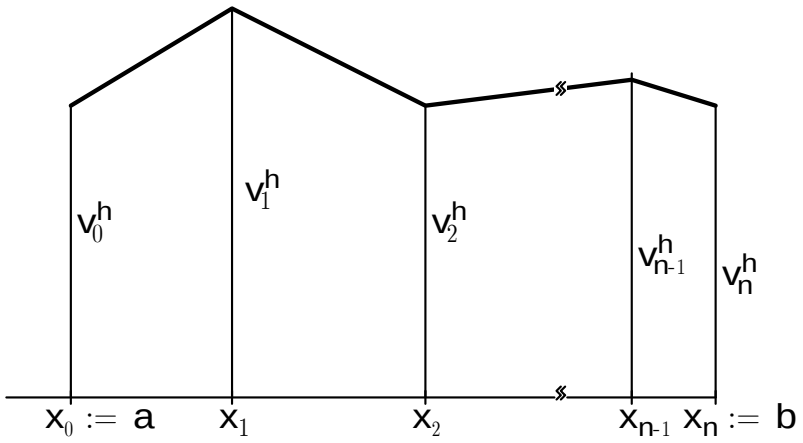


FIG. 2. Graphe d'une fonction générique de X^h.

On a ainsi que $\partial_{x_j} v \in L^2(\Omega)$ si seulement si les intégrales sur chaque interface sont nulles pour tout φ dans $\mathcal{D}(\Omega)$. Or sur chaque interface G' séparant T de L, on a $\mathbf{n}^T + \mathbf{n}^L$. Comme $(v_T)\,|_{G'} = (v_L)\,|_{G'}$ si et seulement si v est globalement continue sur $\overline{\Omega}$, ceci termine la démonstration. □

2. Eléments finis usuels de plus bas degré

Nous allons dans cette section présenter les éléments finis les plus simples pour résoudre les problèmes variationnels du chapitre I. Cette simplicité n'est cependant aucunement en rapport avec une efficacité limitée en tant que procédé de résolution. Il est souvent inutile de recourir à des méthodes plus élaborées en pratique.

2.1. Eléments finis de plus bas degré monodimensionnels. Le théorème 3.1 montre qu'on doit choisir des fonctions globalement continues pour construire un sous-espace de $H^1(]a,b[)$. En se limitant à des fonctions polynômiales sur chaque élément, on est amené à considérer ainsi successivement des fonctions $\mathbb{P}_0^{(1)}$ sur chaque élément, puis $\mathbb{P}_1^{(1)}$, etc. pour cette construction. On a déjà vu que le choix $\mathbb{P}_0^{(1)}$ comme espace de fonctions de forme ne peut être retenu car le raccord de fonctions en escalier de façon à constuire une fonction globalement continue donne seulement des fonctions constantes sur tout l'intervalle $]a,b[$. Ceci ne peut bien sûr être considéré comme un procédé efficace d'approximation de fonctions.

On est donc amené de façon naturelle à considérer

$$X^h := \left\{ v^h \in \mathcal{C}^0([a,b])\,;\ v^h|_{T^{[e]}} \in \mathbb{P}_1^{(1)},\ \text{pour } e = 0,\dots,n-1 \right\} \tag{2.1}$$

où

$$T^{[e]} :=]x_e, x_{e+1}[\tag{2.2}$$

est l'élément $[e]$ du maillage. Comme tout polynôme p de degré ≤ 1 est complètement déterminé dans $K^{[e]}$ à partir de ses valeurs $p_1^{[e]} := p(x_e)$ et $p_2^{[e]} := p(x_{e+1})$, on obtient donc un sous-espace de fonctions suffisamment riche pour approcher les éléments de $H^1(]a,b[)$ (voir Fig. 2).

Les propriétés de cet espace d'éléments finis, appelés $\mathbb{P}_1$-continus, sont décrites dans le théorème suivant dont la démonstration est immédiate.

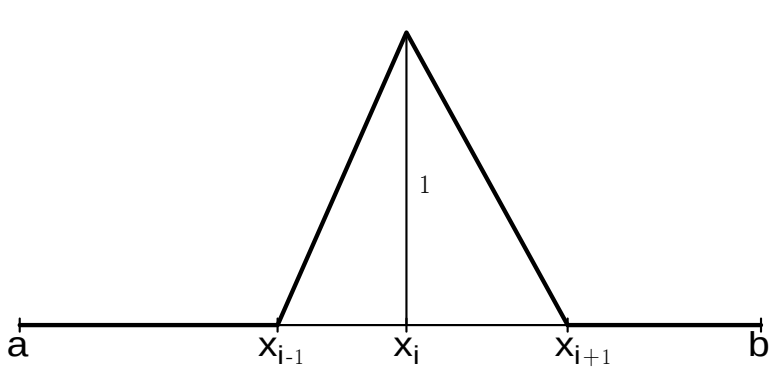


FIG. 3. Fonction chapeau (avec des modifications adéquates pour les cas $i = 0$ et $i = m$).

THÉORÈME 3.2. *Chaque fonction $v^h \in X^h$ est complètement déterminée par ses* ***degrés de liberté*** *(ici, ses* ***valeurs nodales****)*

$$v_i^h := v^h(x_i), \quad \textit{pour } i = 0, \ldots, n. \tag{2.3}$$

Comme de plus, $v^h(x) = \sum_{i=0}^{i=n} v_i^h B_i^h(x)$ où B_i^h est la fonction de X^h définie par

$$B_i^h(x_j) = \delta_{ij}, \quad i, j = 0, \ldots, n$$

X^h est un sous-espace de dimension $N_S := n + 1$, le nombre de sommets du maillage.

Les fonctions $\left\{B_i^h\right\}_{i=0}^{i=N}$ sont les fonctions de base de cette méthode d'éléments finis. On les appelle ici fonctions chapeau par référence à leur graphe (voir Fig. 3)

2.2. Eléments finis bidimensionnels. En prenant le triangle comme élément géométrique, on peut généraliser la construction précédente aux cas des fonctions à deux variables en s'appuyant sur deux propriétés.

La première de ces propriétés est donnée par le résultat suivant.

PROPOSITION 3.1 (Coordonnées barycentriques). *Soit T un triangle non dégénéré du plan de sommets $a_j^T := (a_{1j}^T, a_{2j}^T)$ $(j = 1, 2, 3)$. Les conditions suivantes*

$$\begin{cases} \lambda_i^T \in \mathbb{P}_1^{(2)} \ (i = 1, 2, 3), \\ \lambda_i^T(a_j^T) = \delta_{ij} \ (i, j = 1, 2, 3), \end{cases} \tag{2.4}$$

où δ_{ij} est le symbole de Kronecker, caractérisent une base de $\mathbb{P}_1^{(2)}$ permettant de déterminer tout polynôme $p \in \mathbb{P}_1^{(2)}$ à l'aide de ses valeurs nodales : $p_i^T := p(a_i^T)$ $(i = 1, 2, 3)$ (valeurs aux sommets de K) par

$$p(x) = \sum_{i=1}^{3} p_i^T \lambda_i^T(x) \tag{2.5}$$

DÉMONSTRATION. On va déterminer les λ_i^T à partir des conditions nécessaires qu'ils vérifient. Ainsi donc, les λ_i^T, s'ils existent, déterminent en particulier les polynômes 1, x_1 et x_2. Ils satisfont dès lors les conditions suivantes

$$\begin{cases} \lambda_1^T(x) & + & \lambda_2^T(x) & + & \lambda_3^T(x) & = & 1 \\ a_{11}^T \lambda_1^T(x) & + & a_{12}^T \lambda_2^T(x) & + & a_{13}^T \lambda_2^T(x) & = & x_1 \\ a_{21}^T \lambda_1^T(x) & + & a_{22}^T \lambda_2^T(x) & + & a_{23}^T \lambda_2^T(x) & = & x_2 \end{cases} \tag{2.6}$$

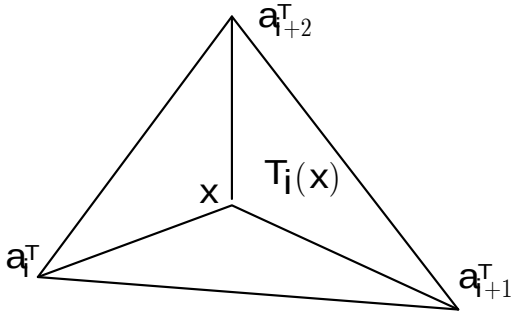


FIG. 4. Triangle intervenant dans le calcul de la coordonnée barycentrique i

Le déterminant de ce système est donné par

$$\det \begin{bmatrix} 1 & 1 & 1 \\ a_{11}^T & a_{12}^T & a_{13}^T \\ a_{21}^T & a_{22}^T & a_{23}^T \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 0 \\ a_{11}^T & a_{12}^T - a_{11}^T & a_{13}^T - a_{11}^T \\ a_{21}^T & a_{22}^T - a_{21}^T & a_{23}^T - a_{21}^T \end{bmatrix}$$
$$= \det \begin{bmatrix} a_{12}^T - a_{11}^T & a_{13}^T - a_{11}^T \\ a_{22}^T - a_{21}^T & a_{23}^T - a_{21}^T \end{bmatrix} = 2\varepsilon |T|$$

où $|T| > 0$ est l'aire du triangle T et $\varepsilon = \pm 1$ suivant qu'on tourne, en suivant la numérotation des sommets, dans le sens direct ou non. Ceci montre déjà que les λ_i^T sont déterminés de façon unique. On n'a pas encore montré, cependant, qu'ils répondent à la question. Pour celà, on résout le système

(2.6) à l'aide des formules de Cramer, soit, en utilisant les propriétés bien connues de permuation des colonnes des déterminants et une permutation circulaire des indices,

$$\lambda_i^T(x) = \frac{1}{2\varepsilon |T|} \det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & a_{1i+1}^T & a_{1i+2}^T \\ x_2 & a_{2i+1} & a_{2i+2}^T \end{bmatrix}, \quad \text{pour } i = 1, 2, 3.$$

En développant le déterminant du numérateur par rapport à la première colonne, on montre d'abord que λ_i^T est bien un polynôme dans $\mathbb{P}_1^{(2)}$. La même réduction que celle effectuée pour le déterminant de la matrice du système montre ensuite que, pour $x \in T$, on a

$$\lambda_i^T(x) = \frac{2\varepsilon |T_i(x)|}{2\varepsilon |T|} = \frac{|T_i(x)|}{|T|} \tag{2.7}$$

où $T_i(x)$ est le triangle de sommets x, a_{i+1}^T, a_{i+2}^T (voir Fig. 4). On termine la démonstration en observant que la formule (2.7) donne $\lambda_i^T(a_j^T) = \delta_{ij}$ où δ_{ij} est le symbole de Kronecker. □

REMARQUE 3.2. *On déduit du théorème précédent et de sa démonstration les propriétés suivantes.*

(1) *Les relations (2.6) peuvent être réécrites*

$$\lambda_1^T(x) + \lambda_2^T(x) + \lambda_3^T(x) = 1$$
$$\lambda_1^T(x) a_1^T + \lambda_2^T(x) a_2^T + \lambda_3^T(x) a_3^T = x$$

Ces relations expriment que x est le barycentre des sommets du triangle T affectés respectivement des poids $\lambda_1^T(x)$, $\lambda_2^T(x)$ et $\lambda_3^T(x)$; d'où la terminologie coordonnées barycentriques pour les fonctions $\lambda_1^T(x)$, $\lambda_2^T(x)$ et $\lambda_3^T(x)$. Ces fonctions

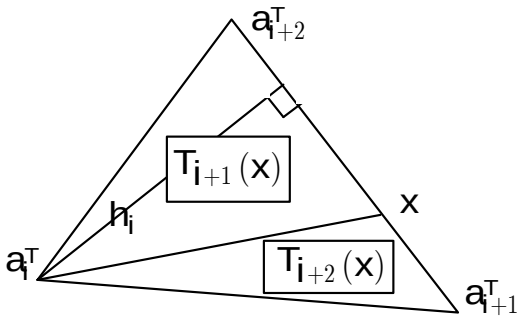


FIG. 5. Triangles $T_i(x)$, $T_{i+1}(x)$ et $T_{i+2}(x)$ pour $x \in \left[a_{i+1}^T, a_{i+2}^T\right]$.

interviennent dans un grand nombre de questions et en particulier dans l'étude des méthodes d'éléments finis plus compliquées.

(2) *Le triangle $T_i(x)$ dégénère lorsque x est sur l'arête $\left[a_{i+1}^T, a_{i+2}^T\right]$. On obtient donc*

$$\lambda_i^T(x) = 0 \tag{2.8}$$

comme équation cartésienne de la droite supportant l'arête opposée au sommet a_i^T.

Pour énoncer la seconde propriété permettant d'étendre la construction de l'approximation par éléments finis en dimension 1, nous avons besoin du lemme suivant.

LEMME 3.1. *Lorsque x est sur l'arête $\left[a_{i+1}^T, a_{i+2}^T\right]$, les coordonnées barycentriques $\lambda_{i+1}(x)$ et $\lambda_{i+2}(x)$ sont données par*

$$\lambda_{i+1}(x) = \left|x - a_{i+2}^T\right| / \left|a_{i+1}^T - a_{i+2}^T\right|, \quad \lambda_{i+2}(x) = \left|x - a_{i+1}^T\right| / \left|a_{i+2}^T - a_{i+1}^T\right| \tag{2.9}$$

où $|x-y|$ désigne la distance de x à y.

DÉMONSTRATION. Notons par h_i la hauteur relativement à la base $\left|a_{i+1}^T - a_{i+2}^T\right|$ de T ; h_i est aussi hauteur de $T_{i+1}(x)$ relativement à la base $\left|x - a_{i+2}^T\right|$ et de $T_{i+2}(x)$ relativement à la base $\left|x - a_{i+1}^T\right|$ (voir Fig. 5) On a donc $|T_{i+1}(x)| = h_i \left|x - a_{i+2}^T\right| /2$ et $|T_{i+2}(x)| = h_i \left|x - a_{i+1}^T\right| /2$; d'où la formule (2.9) en utilisant (2.7). □

On est alors en mesure de montrer la seconde propriété, nécessaire à l'extension évoquée ci-dessus, qui permet de raccorder deux polynômes dans $\mathbb{P}_1^{(2)}$ sur toute une arête en raccordant seulement leurs valeurs aux sommets de cette arête. Plus précisément, on a la proposition suivante.

PROPOSITION 3.2. *Soient deux triangles T et L non dégénérés du plan partageant une arête en commun $G' = [a, b]$. Soit aussi une fonction v dans $L^\infty(D)$ où D est le quadrilatère dont T et L assurent une partition sans recouvrement telle que*

$$v|_T = p \in \mathbb{P}_1^{(2)} \text{ et } v|_L = q \in \mathbb{P}_1^{(2)}$$

alors, $v \in \mathcal{C}^0(\overline{D})$ si et seulement si ces deux fonctions se raccordent aux points a et b (i.e. v est globalement continue sur $\overline{D}$ si et seulement si elle est continue au points a et b).

DÉMONSTRATION. Il est clair que la continuité aux points a et b est nécessaire. Montrons qu'elle est suffisante. Notons par c et d les sommets respectifs de T et L non sur l'arête G' (voir Fig. 6). Notons par λ_a^T, λ_b^T, λ_c^T les coordonnées barycentriques relatives au triangle T et λ_a^L, λ_b^L, λ_d^L celles relatives au triangle L. Pour $x \in G'$, on a $\lambda_c^T(x) = \lambda_d^L(x) = 0$,

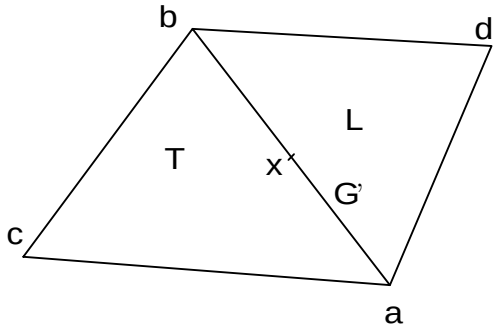


FIG. 6. Triangles partageant une arête en commun

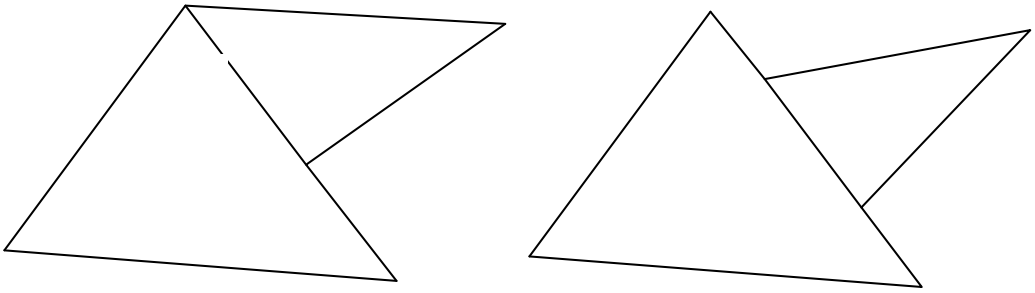


FIG. 7. Situations interdites

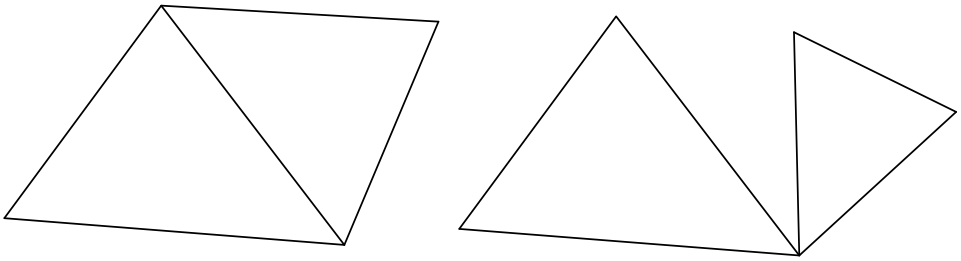


FIG. 8. Situations permises

$\lambda_a^T(x) = \lambda_a^L(x) = |x-b| \,/\, |a-b|$ et $\lambda_b^T(x) = \lambda_b^L(x) = |x-a| \,/\, |a-b|$. Il résulte donc des relations ci-dessus que pour $x \in G'$

$$\begin{aligned} p(x) &= v(a)\lambda_a^T(x) + v(b)\lambda_b^T(x) \\ &= v(a)\lambda_a^L(x) + v(b)\lambda_b^L(x) = q(x), \end{aligned}$$

d'où la proposition. □

A partir des deux propositions précédentes, nous sommes donc amenés à mailler Ω en triangles vérifiant la **condition de compatibilité** suivante

(2.10) L'interface G' entre deux triangles T et L ne peut être qu'une arête commune

Autrement dit, pour deux triangles T et L distincts, $\overline{T}$ et $\overline{L}$ ne peuvent avoir qu'une intersection vide, un sommet commun ou toute une arête en commun. Les figures (Fig. 7) et (Fig. 8) donnent des exemples de situations conduisant ou non à des maillages utilisables.

A partir des propositions 3.1 et 3.2, on déduit immédiatement la proposition suivante.

PROPOSITION 3.3. *Soit un maillage $\mathcal{T}^h$ en triangles d'un domaine Ω polygonal du plan satisfaisant les conditions (1.1), (1.2) et 2.10) ; alors, l'espace*

$$X^h := \left\{ v^h \in \mathcal{C}^0(\overline{\Omega})\,;\ v^h|_T \in \mathbb{P}_1^{(2)},\ \forall T \in \mathcal{T}^h \right\} \tag{2.11}$$

est un sous-espace de $H^1(\Omega)$ de dimension N_S le nombre de sommets du maillage. (Toute fonction v^h de X^h est complètement définie à l'aide de ses degrés de liberté (ici, valeurs nodales) $v_i^h := v^h(a_i)$ $(i = 1, \ldots, N_S)$, $\{a_i\}_{i=1}^{i=N_S}$ étant les sommets du maillage.)

On peut étendre la construction précédente aux maillages de domaines polyèdriques Ω de $\mathbb{R}^3$ en tétraèdres. Cette classe d'éléments finis est appelée méthode d'éléments finis $\mathbb{P}_1$-continus.

Nous admettrons enfin le théorème suivant dont la démonstration demande certains développements d'analyse fonctionnelle.

THÉORÈME 3.3. *La méthode d'éléments finis $\mathbb{P}_1$-continus, sur des maillages en segments en dimension un, en triangles en dimension deux et en tétraèdres en dimension trois, donne des approximations internes X^h de $H^1(\Omega)$.*

3. Notion d'élément fini

Jusqu'à maintenant nous avons parlé d'éléments finis sans vraiment définir cette notion. A partir des constructions précédentes, nous allons dégager la notion d'élément fini. Elle nous permettra par la suite de construire des schémas numériques pour des problèmes plus complexes que les problèmes du second ordre ou d'améliorer certaines caractéristiques du procédé d'approximation comme la précision, le volume de calculs, le stockage, etc.

Une méthode d'éléments finis est caractérisée par les trois données suivantes et les relations qui les lient :

(1) **Domaine géométrique** T, comme par exemple dans les constructions précédentes
   - Segment $T := \left]a_1^T, a_2^T\right[$ en dimension un,
   - Triangle T de sommets $\left\{a_1^T, a_2^T, a_3^T\right\}$ en dimension deux,
   - Tétraèdre T de sommets $\left\{a_1^T, a_2^T, a_3^T, a_4^T\right\}$ en dimension trois.

(2) **Espace de fonctions de forme** (forme de fonctions approchantes) : espace $\mathbb{P}_T$ **de dimension finie** fonctions régulières définies et régulières sur $\overline{T}$, comme par exemple ci-dessus,
   - Espace $\mathbb{P}_1^{(1)}$ des polynômes à une indéterminée de degré ≤ 1 de dimension 2,
   - Espace $\mathbb{P}_1^{(2)}$ des polynômes à deux indéterminées de degré ≤ 1 de dimension 3,
   - Espace $\mathbb{P}_1^{(3)}$ des polynômes à trois indéterminées de degré ≤ 1 de dimension 4.

(3) **Système de degrés de liberté.** On se limite dans ce cours aux degrés de liberté qui sont des **valeurs nodales** qui sont les valeurs d'une fonction ou de certaines de ses dérivées en des points de $\overline{T}$ appelés **noeuds**. Le système de degrés de liberté doit être **unisolvant** : la donnée des degrés de liberté $\ell(p)$ doit déterminer une fonction de forme $p \in \mathbb{P}_T$ et en plus de façon unique.

Dans les exemples précédents, on avait
- $\Sigma_1^{(1)} := \left\{v(a_1^T), v(a_2^T)\right\}$

– $\Sigma_1^{(2)} := \big\{v(a_1^T), v(a_2^T), v(a_3^T)\big\}$
– $\Sigma_1^{(3)} := \big\{v(a_1^T), v(a_2^T), v(a_3^T), v(a_4^T)\big\}$

On peut aussi avoir des valeurs nodales qui sont données par la valeur de dérivées ou d'une combinaison de dérivées. On a ainsi successivement lorsque l'élément est un segment, puis un triangle

– Espace $\mathbb{P}_3^{(1)}$, $\mathcal{H}_3^{(1)} := \big\{v(a_1^T), v'(a_1^T), v(a_2^T), v'(a_2^T)\big\}$
– Espace $\mathbb{P}_3^{(2)}$, $\mathcal{H}_3^{(2)} := \big\{v(a_j^T), \partial_{x_1}v(a_j^T), \partial_{x_2}v(a_j^T)\,(j=1,2,3),\ v(a_0^T)\big\}$ où $a_0^T := (a_1^T + a_2^T + a_3^T)/3$ est le centre de gravité de T.

Observons qu'un degré de liberté est une forme linéaire sur un espace de fonctions définies régulières sur $\overline{T}$ contenant $\mathbb{P}$.

Une méthode d'éléments finis $\{T, \mathbb{P}_T, \Sigma\}$ est dite de classe $\mathcal{C}^k$ si les dérivées partielles $\partial^\alpha p$ ($|\alpha| \leq k$) d'un élément p quelconque de $\mathbb{P}_T$ sur une interface quelconque ne dépendent que que des degrés de liberté sur cette interface. Par exemple, les éléments finis construits ci-dessus sont de classe $\mathcal{C}^0$ mais non de classe $\mathcal{C}^1$.

L'outil algébrique suivant est très utile pour vérifier l'insolvance.

PROPOSITION 3.4. *Soient $\mathbb{P}$ un espace de dimension finie n et Σ un sytème de m degrés de liberté. Alors, Σ est $\mathbb{P}$-unisolvant si et seulement si les deux conditions suivantes sont vérifiées*

(1) *La dimension n de $\mathbb{P}$ et le nombre de degrés de liberté m de Σ sont égaux.*
(2) *L'unique $p \in \mathbb{P}$ qui a tous ses degrés de liberté $\ell_i(p) = 0$ $(i = 1, \ldots, m)$, est $p = 0$.*

DÉMONSTRATION. Soit une base $\{B_j\}_{j=1}^{j=m}$ de $\mathbb{P}$. Par définition, dire que le système Σ est $\mathbb{P}$-unisolvant revient à dire que pour chaque système de nombres $\{b_i\}_{i=1}^{i=n}$ (valeurs des degrés de liberté), il existe un et un seul système de nombres $\{x_j\}_{j=1}^{j=m}$ (coefficients d'un élément $p \in \mathbb{P}$ dans la base $\{B_j\}_{j=1}^{j=m}$) tel que

$$\begin{cases} \ell_1(x_1B_1 + \cdots + x_nB_n) = b_1 \\ \qquad\qquad \vdots \\ \ell_m(x_1B_1 + \cdots + x_nB_n) = b_m \end{cases}$$

Le système précédent est en fait un système linéaire

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \qquad\qquad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases}$$

avec $a_{ij} := \ell_i(B_j)$. L'unisolvance est ainsi équivalente au fait que le système linéaire précédent est inversible. On sait que ceci équivaut à $m = n$ et que le système homogène associé n'admet que la solution nulle. Ceci démontre la proposition. □

REMARQUE 3.3. *De même, pour s'assurer qu'un élément fini $\{T, \mathbb{P}, \Sigma\}$ est de classe $\mathcal{C}^k$, il suffit de vérifier que si $p \in \mathbb{P}$ a tous ses degrés de liberté sur une interface G' nuls, alors toutes les dérivées jusqu'à l'ordre k sont nulles sur tout G'.*

CHAPITRE 4

Mise en oeuvre de la méthode d'éléments finis

Nous examinons dans ce chapitre les lignes directrices permettant d'effectuer une implantation informatique de la méthode des éléments finis. Nous prenons comme fil conducteur la résolution des problèmes aux limites des chapitres précédents par la méthode des éléments finis de plus bas degré. Nous commençons par la première étape : la donnée du maillage. Nous abordons ensuite la discrétisation, la formation des équations et enfin la résolution.

1. Structure de données d'un maillage

La donnée d'un maillage consiste au minimum en celle de :
– la liste des sommets,
– la table de connectivité.

(1) **Liste des sommets** : caractérisée par les deux données
- N_S : nombre des sommets,
- Pour chaque sommet $\boxed{i} = 1, \ldots, N_S$
  - les coordonnées du sommet $\boxed{i}$
- **Exemples.**

(a) Dimension 1 : $x_0 := a < x_1 < x_2 < \cdots < x_n := b$
- $N_S = n + 1$
- Tableau à une entrée de longueur N_S

$\boxed{i}$	1	2	$\cdots$	N_S
X	x_0	x_1	$\cdots$	x_n

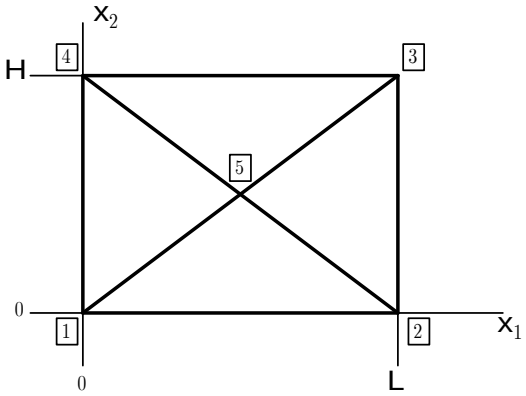
Il ne faut pas toujours stocker cette structure de données. Si la grille est uniforme, elle est donnée de façon implicite par

$$N_S = n + 1; \quad x_0 := a;\ x_i := x_0 + ih,\ i = 1, \ldots, (N_S - 1).$$

(b) Dimension 2 : Tableau des coordonnées $\{(x_{1,i}, x_{2,i})\}_{i=1}^{i=N_S}$
- Nombre de sommets : N_S
- Tableau à deux lignes et N_S colonnes

$\boxed{i}$	1	2	$\cdots$	N_S
X	$\boxed{x_{1,1}}$ $\boxed{x_{2,1}}$	$\boxed{x_{1,2}}$ $\boxed{x_{2,2}}$	$\cdots$	$\boxed{x_{1,N_S}}$ $\boxed{x_{2,N_S}}$

Dans l'exemple suivant, on a $N_S = 5$



i	1	2	3	4	5
X	$\begin{matrix}0\\0\\1\end{matrix}$	$\begin{matrix}L\\0\\1\end{matrix}$	$\begin{matrix}L\\H\\1\end{matrix}$	$\begin{matrix}0\\H\\1\end{matrix}$	$\begin{matrix}L/2\\H/2\\0\end{matrix}$

On peut ajouter un numéro de référence pour les sommets. Ici, il sert à distinguer les sommets sur la frontière du sommet interne.

(2) **Table de connectivité** : caractérisée par les données

- N_E : nombre d'éléments
- Pour chaque élément $\boxed{e} = 1, \ldots, N_E$
  - $n_{e,j}$: numéro du sommet j de l'élément $\boxed{e}$
- **Exemples.**

(a) Dimension 1 :

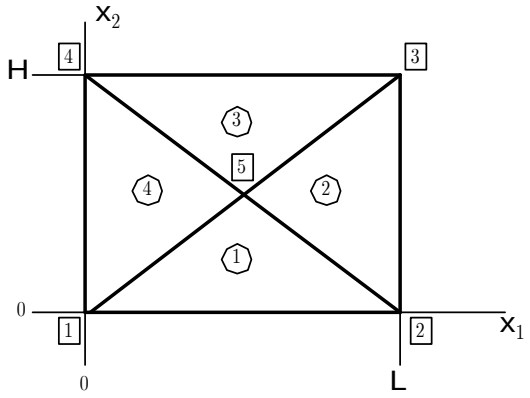
- $N_E = N_S - 1$
- Pour chaque élément $\boxed{e} = 1, \ldots, N_E$

$$n_{e,1} := e, \quad n_{e,2} := e + 1$$

Pour le cas monodimensionnel, la table de connectivité est implicite à partir de la liste des sommets. Il est inutile par conséquent de la stocker.

(b) Dimension 2 : Pour un maillage en triangles,

- N_E : nombre de triangles
- $n_{e,j}$: $e = 1, \ldots, N_E,\ j = 1, 2, 3$



$\boxed{e}$	1	2	3	4
1	5	5	3	1
2	1	2	4	5
3	2	3	5	4

L'ordre dans lequel sont donnés les numéros de sommet n'est pas important. Si on peut, il faut respecter un sens comme ici le sens trigonométrique. Cela peut faciliter pour certains problèmes quelques points de programmation.
Souvent on ajoute un numéro de référence pour pouvoir introduire

une caractérisation des équations au niveau de chaque élément. La table

$\boxed{e}$	1	2	3	4
1	5	5	3	1
2	1	2	4	5
3	2	3	5	4
4	1	2	1	2

indique que les éléments $\boxed{1}$ et $\boxed{3}$ ont les mêmes caractéristiques, de même que les éléments $\boxed{2}$ et $\boxed{4}$.

2. Maillage et conditions aux limites

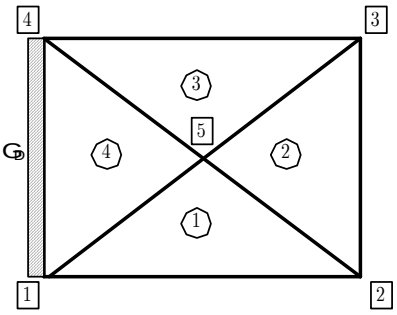
On se limite au cas où la dimension du problème aux limites est $N \leq 2$. Pour $N = 2$, on suppose que Ω est un domaine polygonal. On se donne un maillage $\mathcal{T}^h$ en segments si $N = 1$ et en triangles si $N = 2$.

On note par I_D les numéros de sommets de $\mathcal{T}^h$ qui appartiennent à l'adhérence $\overline{\Gamma_D}$ de Γ_D ; I_D est généralement donné à l'aide des numéros de référence de sommets.

EXEMPLE 4.1. *On examine successivement les cas de la dimension $N = 1$ puis $N = 2$.*

(1) *Dimension 1.* $a := x_0 < \cdots < x_n := b$; $\overline{\Gamma_D} = \{a\}$; $I_D := \{1\}$.

(2) *Dimension 2.*



$I_D := \{1, 4\}$

Sachant que X^h est l'espace d'éléments finis introduit au chapitre précédent, on pose alors

$$V^h := \left\{ v^h \in X^h \,;\; v_i^h = 0,\; i \in I_D \right\} \tag{2.1}$$

(ou encore V^h est le sous-espace de X^h dont les valeurs nodales sur $\overline{\Gamma_D}$ sont nulles).

Si en dimension deux on suppose que **le maillage est compatible avec le changement de type des conditions aux limites**, i.e. si les points de jonction de Γ_D et de Γ_N sont des sommets du maillage, on a alors la proposition suivante.

PROPOSITION 4.1. *Sous la condition de compatibilité précédente sur le maillage, on a*

$$V^h = X^h \cap V. \tag{2.2}$$

DÉMONSTRATION. En dimension $N = 1$, il n'y a rien à démontrer. En dimension $N = 2$, il suffit de remarquer qu'une fonction de $\mathbb{P}_1^{(2)}$ est nulle sur une arête d'un triangle si et seulement si elle s'annule aux extrémités de cette arête. □

On admettra la proposition suivante dont le résultat est intuitif.

PROPOSITION 4.2. *Les espaces X^h et V^h constituent respectivement des approximations internes de $H^1(\Omega)$ et de V*

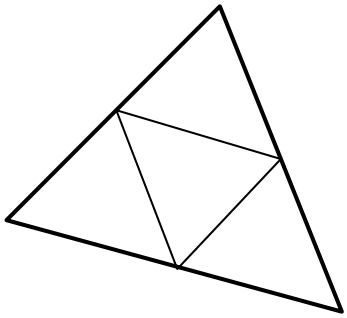
(1) *sans aucune restriction si* $\Omega =]0, L[$,

(2) *sous la condition d'angle suivante si* Ω *est un domaine polygonal de* $\mathbb{R}^2$

$$\left\{ \begin{array}{l} \theta_T \text{ est le plus petit angle de } T \in \mathcal{T}^h,\ \theta^h = \min_{T\in\mathcal{T}^h} \theta_T \\ \theta^h \geq \theta_0 > 0 \text{ indépendamment de } h \end{array} \right. \tag{2.3}$$

La condition d'angle exprime que le maillage ne doit pas comporter de triangles aplatis comme

C'est pourquoi souvent pour raffiner les maillages, on opère de la façon suivante : on partage chaque triangle en 4 en joignant les milieux de ses arêtes



3. L'assemblage

Sous le vocable d'assemblage, on désigne le procédé de formation des équations du système linéaire résultant de l'application d'une méthode d'éléments finis. L'efficacité de ce procédé, généralement non disponible pour les autres méthodes de Galerkin, participe pour une bonne part à la popularité de la méthode des éléments finis.

Auparavant, on rappelle l'obtention des problèmes discrets associés aux formulations variationnelles (1.20) et (2.18) du chapitre I construits à l'aide des espaces d'éléments finis X^h et V^h. On se donne donc g^h une approximation de g dans X^h. En supposant que la fonction $g|_{\Gamma_D} \in \mathcal{C}^0(\overline{\Gamma_D})$, on peut construire de façon pratique la fonction g^h de la façon suivante (dite par interpolation) en se donnant ses valeurs nodales g_i^h sur les sommets a_i^h du maillage $\mathcal{T}^h$

$$g_i^h = g(a_i^h), \quad i \in I_D, \quad g_i^h \text{ quelconque si } i \notin I_D.$$

Le problème discret s'écrit alors

$$\left\{ \begin{array}{l} u^h \in X^h, \quad u^h - g^h \in V^h, \\ a(u^h, v^h) = Lv^h, \quad \forall v^h \in V^h. \end{array} \right. \tag{3.1}$$

Les fonctions de base $\left\{B_j^h\right\}_{j=1}^{j=N_S}$ ont été déterminées par leur restriction à tout élément au chapitre précédent. Elles sont définies par les conditions

$$B_j^h \in X^h, \quad B_j^h(a_i^h) = \delta_{ij}, \quad i, j = 1, \ldots, N_S.$$

Le système variationnel (3.1) est alors équivalent à un système linéaire dont on s'intéresse maintenant à la formation des coefficients. Ces coefficients sont obtenus par la contribution de chaque élément du maillage. Chacune de ces contributions est décrite par des matrices de petite taille : les matrices élémentaires. Le procédé d'assemblage permet d'accumuler toutes ces contributions.

3.1. Matrices élémentaires. La restriction de la forme bilinéaire $a(\cdot,\cdot)$ et de la forme linéaire $L(\cdot)$ au sous-espace X^h est décrite dans la base des B_j^h par **les matrices totales**

$$\left[\begin{array}{ccc} v_1^h & \cdots & v_{N_S}^h \end{array}\right] A^{\#} \left[\begin{array}{c} u_1^h \\ \vdots \\ u_{N_S}^h \end{array}\right] = a(u^h, v^h), \quad u^h, v^h \in X^h \tag{3.2}$$

$$\left[\begin{array}{ccc} v_1^h & \cdots & v_{N_S}^h \end{array}\right] L^{\#} = Lv^h, \quad v^h \in X^h \tag{3.3}$$

où u_j^h (resp. v_j^h) sont les valeurs nodales de u^h (resp. v^h).

Le procédé d'assemblage s'appuie sur le fait que les intégrales

$$a(u^h, v^h) = \int_\Omega \sum_{i,j=1}^N a_{ij}\partial_j u^h \partial_i v^h \, d\Omega + \int_\Omega a_0 u^h v^h \, d\Omega + \int_{\Gamma_N} \lambda uv \, d\Gamma \tag{3.4}$$

$$Lv^h = \int_\Omega f v^h \, d\Omega + \int_{\Gamma_N} hv \, d\Gamma \tag{3.5}$$

avec les adaptations adéquates pour le cas unidimensionnel, peuvent être **décomposées en une somme d'intégrales sur chaque élément**

$$a(u^h, v^h) = \sum_{T^{[e]} \in \mathcal{T}^h} \int_{T^{[e]}} \sum_{i,j=1}^N a_{ij}\partial_j u^{[e]} \partial_i v^{[e]} \, dT^{[e]} + \sum_{T^{[e]} \in \mathcal{T}^h} \int_{T^{[e]}} a_0 u^{[e]} v^{[e]} \, dT^{[e]} + \sum_{T^{[e]} \in \mathcal{T}^h} \sum_{\left(T^{[e]}\right)' \in T^{[e]}} \int_{\left(T^{[e]}\right)' \cap \Gamma_N} \lambda u^{[e]} v^{[e]} \, d\left(T^{[e]}\right)'$$

$$Lv^h = \sum_{T^{[e]} \in \mathcal{T}^h} \int_{T^{[e]}} f v^{[e]} \, dT^{[e]} + \sum_{T^{[e]} \in \mathcal{T}^h} \sum_{\left(T^{[e]}\right)' \in T^{[e]}} \int_{\left(T^{[e]}\right)' \cap \Gamma_N} h v^{[e]} \, d\left(T^{[e]}\right)' \tag{3.6}$$

où la notation $\sum_{\left(T^{[e]}\right)' \in T^{[e]}}$ indique la somme sur les faces (arêtes dans le cas $N = 2$) composant la frontière de $T^{[e]}$ et $u^{[e]} := u^h|_{T^{[e]}} \in \mathbb{P}_1^{(N)}$ (resp. $v^{[e]} := v^h|_{T^{[e]}} \in \mathbb{P}_1^{(N)}$) les expressions polynomiales de u^h (resp. v^h) dans l'élément $T^{[e]}$. Dans le cas unidimensionnel, le dernier terme des sommes ci-dessus est donné par $\lambda u_{N_S} v_{N_S}$ et par $h v_{N_S}$ respectivement. Ils sont décrits globalement de façon directe comme nous le verrons par la suite.

Chacune des intégrales ci-dessus ne dépend donc que des valeurs nodales de u^h et v^h (ou seulement de v^h) relatives à un noeud dans $\overline{T^{[e]}}$. On peut donc les décrire à l'aide de matrices $\ell \times \ell$ (ou seulement $\ell \times 1$) où ℓ est le nombre de noeud par élément : $\ell = 3$ en dimension $N = 2$, $\ell = 2$ en dimension $N = 1$. Ce sont les matrices élémentaires. On en donne la liste ci-dessous.

1. **Matrice raideur élémentaire**

$$\int_{T^{[e]}} \sum_{i,j=1}^N a_{ij}\partial_j u^{[e]} \partial_i v^{[e]} \, dT^{[e]} = \left[\begin{array}{ccc} v_1^{[e]} & \cdots & v_\ell^{[e]} \end{array}\right] K^{[e]} \left[\begin{array}{c} u_1^{[e]} \\ \vdots \\ u_\ell^{[e]} \end{array}\right] \tag{3.7}$$

2. **Matrice masse élémentaire**

$$\int_{T^{[e]}} a_0 u^{[e]} v^{[e]} \, dT^{[e]} = \left[\begin{array}{ccc} v_1^{[e]} & \cdots & v_\ell^{[e]} \end{array}\right] M^{[e]} \left[\begin{array}{c} u_1^{[e]} \\ \vdots \\ u_\ell^{[e]} \end{array}\right] \tag{3.8}$$

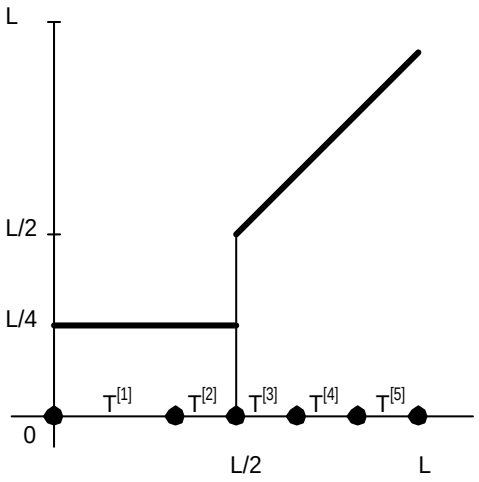


FIG. 1. Maillage et discontinuités des données

3. **Matrice chargement élémentaire de volume**

$$\int_{T^{[e]}} f v^{[e]}\, dT^{[e]} = \left[\begin{array}{ccc} v_1^{[e]} & \cdots & v_\ell^{[e]} \end{array}\right] F^{[e]} \tag{3.9}$$

4. **Matrice raideur élémentaire concentrée sur le bord**

$$\int_{\left(G^{[a]}\right)'} \lambda u^{[a]} v^{[a]}\, d\left(G^{[a]}\right)' = \left[\begin{array}{ccc} v_1^{[a]} & \cdots & v_{\ell'}^{[a]} \end{array}\right] K_C^{[a]} \left[\begin{array}{c} u_1^{[a]} \\ \vdots \\ u_{\ell'}^{[a]} \end{array}\right] \tag{3.10}$$

où $\left(G^{[a]}\right)'$ est une face (arête en dimension deux) sur Γ_N et ℓ' est le nombre de valeurs nodales sur $\left(G^{[a]}\right)'$ et $u^{[a]}$ et $v^{[a]}$ sont les restrictions respectives de u^h et v^h sur $\left(G^{[a]}\right)'$,

5. **Matrice chargement élémentaire concentrée sur le bord**

$$\int_{\left(G^{[a]}\right)'} h v^{[e]}\, d\left(G^{[a]}\right)' = \left[\begin{array}{ccc} v_1^{[a]} & \cdots & v_{\ell'}^{[a]} \end{array}\right] F_C^{[a]} \tag{3.11}$$

Les matrices et forces concentrées en dimension $N = 1$ sont introduites à l'initialisation ou à la fin du processus d'assemblage.

3.2. Matrices élémentaires en dimension un. On a ici $T^{[e]} = \left]a_1^{[e]}, a_2^{[e]}\right[$. En prenant **un maillage compatible avec les discontinuités** éventuelles des données, c'est-à-dire pour lequel les discontinuités des données sont toujours aux interfaces entre deux deux éléments, on peut supposer que $a|_{T^{[e]}} = a^{[e]}$, $a_0|_{T^{[e]}} = a_0^{[e]}$ et $f|_{T^{[e]}} = f^{[e]}$ sont tels que

$$a^{[e]},\ a_0^{[e]} \text{ et } f^{[e]} \in \mathcal{C}^\infty(\left]a_1^{[e]}, a_2^{[e]}\right[).$$

Par exemple, pour la fonction représentée dans la figure Fig. 1, on a

$$\begin{array}{l} a^{[1]}(x) = L/4, \quad a^{[2]}(x) = L/4, \\ a^{[3]}(x) = x, \quad a^{[4]}(x) = x, \quad a^{[5]}(x) = x. \end{array}$$

La particularisation de la situation générale au cas de la dimension $N = 1$ donne alors.

1. **Matrice raideur élémentaire.**
On a simplement ici

$$\left(u^{[e]}\right)' = \frac{u_2^{[e]} - u_1^{[e]}}{\left|T^{[e]}\right|}, \quad \left(v^{[e]}\right)' = \frac{v_2^{[e]} - v_1^{[e]}}{\left|T^{[e]}\right|},$$

où $\left|T^{[e]}\right|$ est la longueur de l'élément $T^{[e]}$. On a alors

$$\begin{aligned} \left[\begin{array}{cc} v_1^{[e]} & v_2^{[e]} \end{array}\right] K^{[e]} \left[\begin{array}{c} u_1^{[e]} \\ u_2^{[e]} \end{array}\right] &= \int_{a_1^{[e]}}^{a_2^{[e]}} a^{[e]} \left(u^{[e]}\right)' \left(v^{[e]}\right)' dx \\ &= \frac{\overline{a^{[e]}}}{\left|T^{[e]}\right|}(u_2^{[e]} - u_1^{[e]})(v_2^{[e]} - v_1^{[e]}) \end{aligned} \tag{3.12}$$

où $\overline{a^{[e]}}$ est la moyenne de la fonction $a^{[e]}$ sur l'élément $T^{[e]}$

$$\overline{a^{[e]}} := \frac{1}{|T^{[e]}|} \int_{a_1^{[e]}}^{a_2^{[e]}} a^{[e]}\, dT^{[e]}$$

qu'on peut approcher par la formule des trapèzes ou du point milieu

$$\overline{a^{[e]}} \approx \frac{1}{2}(a^{[e]}(a_1^{[e]}) + a^{[e]}(a_2^{[e]})) \approx a^{[e]}(\frac{1}{2}(a_1^{[e]} + a_2^{[e]})). \tag{3.13}$$

On identifie les coefficients de $K^{[e]}$ par la relation (3.12) : $K_{ij}^{[e]}$ est le coefficient de $u_j^{[e]} v_i^{[e]}$

$$K^{[e]} := \frac{\overline{a^{[e]}}}{\left|T^{[e]}\right|} \left[\begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array}\right]$$

2. **Matrice masse condensée.**
Généralement, il est suffisant de calculer l'intégrale (3.8) de façon approchée par la méthode des trapèzes

$$\begin{aligned} \left[\begin{array}{cc} v_1^{[e]} & v_2^{[e]} \end{array}\right] M^{[e]} \left[\begin{array}{c} u_1^{[e]} \\ u_2^{[e]} \end{array}\right] &= \int_{a_1^{[e]}}^{a_2^{[e]}} a_0^{[e]} u^{[e]} v^{[e]}\, dx \\ &\approx \frac{\left|T^{[e]}\right|}{2} \left(\left(a_0^{[e]}\right)_1 u_1^{[e]} v_1^{[e]} + \left(a_0^{[e]}\right)_2 u_2^{[e]} v_2^{[e]}\right) \end{aligned} \tag{3.14}$$

avec $\left(a_0^{[e]}\right)_j := a_0^{[e]}(a_j^{[e]})$, $j = 1, 2$. Ceci donne une matrice diagonale

$$M^{[e]} := \frac{\left|T^{[e]}\right|}{2} \left[\begin{array}{cc} \left(a_0^{[e]}\right)_1 & 0 \\ 0 & \left(a_0^{[e]}\right)_2 \end{array}\right] \tag{3.15}$$

3. **Matrice chargement élémentaire.**
On peut là encore utiliser la formule des trapèzes pour approcher l'intégrale (3.9)

$$\begin{aligned} \left[\begin{array}{cc} v_1^{[e]} & v_2^{[e]} \end{array}\right] F^{[e]} &= \int_{a_1^{[e]}}^{a_2^{[e]}} f^{[e]} v^{[e]}\, dx \\ &\approx \frac{\left|T^{[e]}\right|}{2} \left(f_1^{[e]} v_1^{[e]} + f_2^{[e]} v_2^{[e]}\right) \end{aligned} \tag{3.16}$$

avec $f_j^{[e]} := f^{[e]}(a_j^{[e]})$, $j = 1, 2$. D'où la matrice élémentaire

$$F^{[e]} := \frac{|T^{[e]}|}{2} \begin{bmatrix} f_1^{[e]} \\ f_2^{[e]} \end{bmatrix} \tag{3.17}$$

4. **Matrices concentrées sur le bord.**
C'est ici des matrices à un seul coefficient non nul qui sont données directement à partir de

$$\lambda u_{N_S}^h v_{N_S}^h \quad \text{et} \quad h v_{N_S}^h$$

par

$$(K_C^{\#})_{N_S N_S} = \lambda \quad \text{et} \quad (F_C^{\#})_{N_S} = h. \tag{3.18}$$

3.3. Matrices élémentaires en dimension deux. On note de la même façon qu'en dimension un les expressions des restrictions $a_{ij}^{[e]} := a_{ij}|_{T^{[e]}}$, $A^{[e]}$ la matrice 2×2 dont les coefficients sont les $a_{ij}^{[e]}$, $a_0^{[e]} := a_0|_{T^{[e]}}$ et $f^{[e]} := f|_{T^{[e]}}$ qu'on peut supposer dans $\mathcal{C}^\infty(\overline{T^{[e]}})$ si le maillage est compatible avec les discontinuités des données.

On peut alors donner l'expression des matrices élémentaires.

1. **Matrice raideur élémentaire.**
La méthode suivante de détermination de la matrice élémentaire est en fait générale. Elle peut être utilisée pour la détermination de n'importe quelle matrice élémentaire et dans le cadre de n'importe quelle méthode d'éléments finis.
On a par définition

$$\begin{bmatrix} v_1^{[e]} & v_2^{[e]} & v_3^{[e]} \end{bmatrix} K^{[e]} \begin{bmatrix} u_1^{[e]} \\ u_2^{[e]} \\ u_3^{[e]} \end{bmatrix} = \int_{T^{[e]}} \begin{bmatrix} \partial_1 v^{[e]} & \partial_2 v^{[e]} \end{bmatrix} A^{[e]} \begin{bmatrix} \partial_1 u^{[e]} \\ \partial_2 u^{[e]} \end{bmatrix} dT^{[e]} \tag{3.19}$$

D'où si l'on note par $\overline{A^{[e]}}$ la moyenne de $A^{[e]}$ sur $T^{[e]}$

$$\begin{bmatrix} v_1^{[e]} & v_2^{[e]} & v_3^{[e]} \end{bmatrix} K^{[e]} \begin{bmatrix} u_1^{[e]} \\ u_2^{[e]} \\ u_3^{[e]} \end{bmatrix} = \begin{bmatrix} \partial_1 v^{[e]} & \partial_2 v^{[e]} \end{bmatrix} \left(|T^{[e]}| \, \overline{A^{[e]}} \right) \begin{bmatrix} \partial_1 u^{[e]} \\ \partial_2 u^{[e]} \end{bmatrix} \tag{3.20}$$

On exprime maintenant $u^{[e]}$ et $v^{[e]}$ à l'aide de leurs coefficients polynomiaux et d'une écriture matricielle

$$u^{[e]}(x) = \alpha_0^{[e]} + \alpha_1^{[e]} x_1 + \alpha_2^{[e]} x_2 = \begin{bmatrix} 1 & x_1 & x_2 \end{bmatrix} \begin{bmatrix} \alpha_0^{[e]} \\ \alpha_1^{[e]} \\ \alpha_2^{[e]} \end{bmatrix} \tag{3.21}$$

$$v^{[e]}(x) = \beta_0^{[e]} + \beta_1^{[e]} x_1 + \beta_2^{[e]} x_2 = \begin{bmatrix} 1 & x_1 & x_2 \end{bmatrix} \begin{bmatrix} \beta_0^{[e]} \\ \beta_1^{[e]} \\ \beta_2^{[e]} \end{bmatrix} \tag{3.22}$$

On peut alors calculer

$$\left[\begin{array}{c}\partial_1 u^{[e]}\\ \partial_2 u^{[e]}\end{array}\right] = \left[\begin{array}{ccc}0&1&0\\0&0&1\end{array}\right] \left[\begin{array}{c}\alpha_0^{[e]}\\ \alpha_1^{[e]}\\ \alpha_2^{[e]}\end{array}\right] = B^{[e]} \left[\begin{array}{c}\alpha_0^{[e]}\\ \alpha_1^{[e]}\\ \alpha_2^{[e]}\end{array}\right] \tag{3.23}$$

$$\left[\begin{array}{c}\partial_1 v^{[e]}\\ \partial_2 v^{[e]}\end{array}\right] = \left[\begin{array}{ccc}0&1&0\\0&0&1\end{array}\right] \left[\begin{array}{c}\beta_0^{[e]}\\ \beta_1^{[e]}\\ \beta_2^{[e]}\end{array}\right] = B^{[e]} \left[\begin{array}{c}\beta_0^{[e]}\\ \beta_1^{[e]}\\ \beta_2^{[e]}\end{array}\right] \tag{3.24}$$

On utilise alors la relation qui permet de relier les coefficients polynomiaux aux valeurs nodales

$$\left\{\begin{array}{l}\alpha_0^{[e]} + \alpha_1^{[e]} a_{11}^{[e]} + \alpha_2^{[e]} a_{21}^{[e]} = u_1^{[e]}\\ \alpha_0^{[e]} + \alpha_1^{[e]} a_{12}^{[e]} + \alpha_2^{[e]} a_{22}^{[e]} = u_2^{[e]}\\ \alpha_0^{[e]} + \alpha_1^{[e]} a_{13}^{[e]} + \alpha_2^{[e]} a_{23}^{[e]} = u_3^{[e]}\end{array}\right.$$

qu'on écrit sous forme matricielle

$$P^{[e]}\left[\alpha^{[e]}\right] = \left[u^{[e]}\right], \quad P^{[e]} := \left[\begin{array}{ccc}1&a_{11}^{[e]}&a_{21}^{[e]}\\1&a_{12}^{[e]}&a_{22}^{[e]}\\1&a_{13}^{[e]}&a_{23}^{[e]}\end{array}\right], \quad \left[\alpha^{[e]}\right] := \left[\begin{array}{c}\alpha_0^{[e]}\\ \alpha_1^{[e]}\\ \alpha_2^{[e]}\end{array}\right], \quad \left[u^{[e]}\right] := \left[\begin{array}{c}u_1^{[e]}\\ u_2^{[e]}\\ u_3^{[e]}\end{array}\right] \tag{3.25}$$

et de même

$$P^{[e]}\left[\beta^{[e]}\right] = \left[v^{[e]}\right], \quad \left[\beta^{[e]}\right] := \left[\begin{array}{c}\beta_0^{[e]}\\ \beta_1^{[e]}\\ \beta_2^{[e]}\end{array}\right], \quad \left[v^{[e]}\right] := \left[\begin{array}{c}v_1^{[e]}\\ v_2^{[e]}\\ v_3^{[e]}\end{array}\right] \tag{3.26}$$

On inverse alors $H^{[e]} = \left(P^{[e]}\right)^{-1}$ pour exprimer les coefficients polynomiaux à l'aide des valeurs nodales

$$\left[\alpha^{[e]}\right] = H^{[e]}\left[u^{[e]}\right], \quad \left[\beta^{[e]}\right] = H^{[e]}\left[v^{[e]}\right]. \tag{3.27}$$

Cette inversion constitue en fait la détermination des fonctions de base sur $T^{[e]}$ de l'élément fini. Comme on le voit cette détermination peut être effectuée (pour cet élément et pour n'importe quel élément fini) de façon complètement implicite par programmation.
En utilisant toutes les relations précédentes, on obtient

$$\left[v^{[e]}\right]^{\mathrm{T}} K^{[e]}\left[u^{[e]}\right] = \left[v^{[e]}\right]^{\mathrm{T}} \left(H^{[e]}\right)^{\mathrm{T}} \left(B^{[e]}\right)^{\mathrm{T}} \left(\left|T^{[e]}\right| \overline{A^{[e]}}\right) B^{[e]} H^{[e]}\left[u^{[e]}\right]. \tag{3.28}$$

D'où la matrice raideur élémentaire par un simple calcul matriciel

$$K^{[e]} := \left(H^{[e]}\right)^{\mathrm{T}} \left(B^{[e]}\right)^{\mathrm{T}} \left(\left|T^{[e]}\right| \overline{A^{[e]}}\right) B^{[e]} H^{[e]}. \tag{3.29}$$

La moyenne de $A^{[e]}$ peut être obtenue par une formule d'intégration approchée comme en dimension un

$$\overline{A^{[e]}} \approx \frac{1}{3}\left(A_1^{[e]} + A_2^{[e]} + A_3^{[e]}\right) \approx A_0^{[e]}$$

avec $A_j^{[e]}$ valeur de $A^{[e]}$ au sommet $a_j^{[e]}$, $j = 1, 2, 3$ et $A_0^{[e]}$ valeur de $A^{[e]}$ au centre de gravité de $T^{[e]}$: $(a_1^{[e]} + a_2^{[e]} + a_3^{[e]})/3$.

2. **Matrice raideur concentrée sur le bord.**

On dispose généralement à partir des données fournies par le mailleur (sinon on peut la construire par programmation) de la liste des arêtes sur la frontière munies chacune d'un

numéro de référence permettant entre autres de savoir si l'arête est contenue dans Γ_N ou ne recouvre pas Γ_N. On peut donc supposer qu'on dispose des données suivantes :

(1)
- N_a : nombre d'arêtes sur la frontière Γ,
- pour chaque arête $\boxed{a} = 1, \ldots, N_a$
  - n_{a1} : numéro du 1er sommet $a_1^{[a]}$
  - n_{a2} : numéro du 2nd sommet $a_2^{[a]}$
  - n_{a2} : numéro de référence de l'arête

On peut donc parcourir les arêtes et calculer la matrice élémentaire donnant la contribution à la raideur concentrée sur Γ_N

$$\left[\begin{array}{cc} v_1^{[a]} & v_2^{[a]} \end{array}\right] K_C^{[a]} \left[\begin{array}{c} u_1^{[a]} \\ u_2^{[a]} \end{array}\right] = \int_{\left(G^{[a]}\right)'} \lambda^{[a]} u^{[a]} v^{[a]} \, d\left(G^{[a]}\right)'. \tag{3.30}$$

où $\lambda^{[a]} := \lambda|_{\left(G^{[a]}\right)'}$ est la restriction de λ à l'intérieure de l'arête $\left(G^{[a]}\right)'$. Si on approche cette intégrale par la formule des trapèzes, on a

$$K_C^{[a]} := \frac{\left|\left(G^{[a]}\right)'\right|}{2} \left[\begin{array}{cc} \lambda_1^{[a]} & 0 \\ 0 & \lambda_2^{[a]} \end{array}\right] \tag{3.31}$$

où $\left|\left(G^{[a]}\right)'\right|$ est la longueur de l'arête et $\lambda_j^{[a]} := \lambda^{[a]}(a_j^{[a]})$, $j = 1, 2$, sont les valeurs de $\lambda^{[a]}$ aux extrémités de l'arête.

2. **Matrice masse condensée.**

La matrice masse condensée sans recourir à la matrice $H^{[e]}$ en utilisant la formule d'intégration approchée à partir des sommets

$$\int_{T^{[e]}} a_0^{[e]} u^{[e]} v^{[e]} \, dT^{[e]} = \frac{\left|T^{[e]}\right|}{3} \left(\left(a_0^{[e]}\right)_1 u_1^{[e]} v_1^{[e]} + \left(a_0^{[e]}\right)_2 u_2^{[e]} v_2^{[e]} + \left(a_0^{[e]}\right)_3 u_3^{[e]} v_3^{[e]} \right) \tag{3.32}$$

où $\left(a_0^{[e]}\right)_j := a_0^{[e]}(a_j^{[e]})$, $j = 1, 2, 3$, sont les valeurs de $a_0^{[e]}$ aux sommets de $T^{[e]}$. Il vient par suite

$$M^{[e]} := \frac{\left|T^{[e]}\right|}{3} \left[\begin{array}{ccc} \left(a_0^{[e]}\right)_1 & 0 & 0 \\ 0 & \left(a_0^{[e]}\right)_2 & 0 \\ 0 & 0 & \left(a_0^{[e]}\right)_3 \end{array}\right] \tag{3.33}$$

Le fait que la matrice masse condenée est diagonale sera déterminant dans certaines résolutions de problèmes en dynamique, c'est-à-dire dépendant en outre du temps t.

4. **Matrices chargement élémentaires.**

On a de même que pour la matrice chargement élémentaire de volume en dimension un

$$F^{[e]} := \frac{\left|T^{[e]}\right|}{3} \left[\begin{array}{c} f_1^{[e]} \\ f_2^{[e]} \\ f_3^{[e]} \end{array}\right], \qquad f_j^{[e]} := f^{[e]}(a_j^{[e]}),\ j = 1, 2, 3. \tag{3.34}$$

La matrice chragement élémentaire concentrée sur le bord est donnée de la même façon par

$$F_C^{[a]} := \frac{\left|\left(G^{[a]}\right)'\right|}{2} \left[\begin{array}{c} h_1^{[a]} \\ h_2^{[a]} \end{array}\right], \qquad h_j^{[a]} := h^{[a]}(a_j^{[a]}),\ j = 1, 2. \tag{3.35}$$

3.4. Assemblage. La **matrice raideur totale** est définie par

(3.36)
$$\left[\begin{array}{ccc} v_1^h & \cdots & v_{N_S}^h \end{array}\right] K^{\#} \left[\begin{array}{c} u_1^h \\ \vdots \\ u_{N_S}^h \end{array}\right] = \int_\Omega \left[\begin{array}{cc} \partial_1 v^h & \partial_h v^h \end{array}\right] A \left[\begin{array}{c} \partial_1 u^h \\ \partial_2 u^h \end{array}\right] d\Omega, \quad u^h, v^h \in X^h.$$

En utilisant la décomposition de l'intégrale sur Ω

$$\int_\Omega \{\cdots\}\, d\Omega = \sum_{T^{[e]} \in \mathcal{T}^h} \int_{T^{[e]}} \{\cdots\}\, dT^{[e]}$$

et la définition (3.7) de la matrice raideur élémentaire, on obtient la relation suivante entre la matrice raideur totale et les matrices raideur élémentaires

$$\left[\begin{array}{ccc} v_1^h & \cdots & v_{N_S}^h \end{array}\right] K^{\#} \left[\begin{array}{c} u_1^h \\ \vdots \\ u_{N_S}^h \end{array}\right] = \sum_{T^{[e]} \in \mathcal{T}^h} \left[\begin{array}{ccc} v_1^{[e]} & \cdots & v_\ell^{[e]} \end{array}\right] K^{[e]} \left[\begin{array}{c} u_1^{[e]} \\ \vdots \\ u_\ell^{[e]} \end{array}\right]. \tag{3.37}$$

On identifie alors le coefficient de $u_j^h v_i^h$ dans les deux membres de cette realation pour obtenir

$$\left[K^{\#}\right]_{ij} = \sum_{\left\{T^{[e]} \in \mathcal{T}^h,\, n_{[e]i_e} = i,\, n_{[e]j_e} = j\right\}} \left[K^{[e]}\right]_{i_e j_e}. \tag{3.38}$$

Conséquences. La formule (3.38) résume le procédé d'assemblage. Elle a plusieurs conséquences.

(1) Un coefficient $\left[K^{\#}\right]_{ij}$ **est nul a priori** si a_i^h et a_j^h ne sont pas les sommets d'un même élément. Autrement dit, les degrés de liberté u_j^h et v_i^h ne sont couplés (i.e. $\left[K^{\#}\right]_{ij} \neq 0$ ou encore l'équation relative à v_i^h dépend de l'inconnue u_j^h) que si u_j^h et v_i^h appartiennent à un même élément.

(2) Pour former $K^{\#}$, il suffit de parcourir les éléments en ajoutant successivement la contribution de chaque de chaque matrice élémentaire $\left[K^{[e]}\right]_{i_e j_e}$ à la matrice totale $\left[K^{\#}\right]_{ij}$ avec $i = n_{[e]i_e}$, $j = n_{[e]j_e}$.

Le pseudo-code FORTRAN 90 (Alg. 1) donne l'organigramme général d'un assemblage.

Algorithm 1 Algorithme général schématisant un assemblage

```
K# := 0 ; {Initialisation des éléments de K# stockés en mémoire}
do e = 1, N_E  {Boucle sur les éléments}
  K_elem := form(e) ; {Formation de la matrice élémentaire}
  do i_e = 1, ℓ  {Boucle sur les lignes de la matrice élémentaire}
    i := n(e, i_e) ; {Indice de ligne du coefficient à incrémenter}
    do j_e = 1, ℓ  {Boucle sur les colonnes de la matrice élémentaire}
      j := n(e, j_e) ; {Indice de colonne du coefficient à incrémenter}
      K#(i, j) := K#(i, j) + K_elem(i_e, j_e) ; {Assemblage}
    enddo
  enddo
enddo
```

EXEMPLE 4.2. ***Problème en dimension un.***
Dans le cas de la forme (1.17) du problème (1.20) du chapitre I, on a deux propriétés importantes

– *la matrice $A^{\#}$ est symétrique*
– *chaque élément $\boxed{i}$ ne contient que les noeuds a_i^h et a_{i+1}^h.*

On en déduit les propriétés suivantes

$$\left[A^{\#}\right]_{ij} = \left[A^{\#}\right]_{ji}. \tag{3.39}$$

Il suffit donc de ne stocker que la partie triangulaire supérieure de $A^{\#}$, i.e. les coefficients $\left[A^{\#}\right]_{ij}$ pour $i \leq j$. De plus,

$$\left[A^{\#}\right]_{ij} = 0 \quad \textit{pour } |i-j| \geq 2. \tag{3.40}$$

On peut donc stocker cette matrice à l'aide uniquement de 2 vecteurs : la diagonale principale $\left[A^{\#}\right]_{ii}$ $(i = 1, \ldots, N_S)$, et la première codiagonale $\left[A^{\#}\right]_{i,i+1}$ $(i = 1, \ldots, N_S - 1)$.

En dimension supérieure, la situation n'est plus aussi favorable que dans l'exemple ci-dessus. Cependant, en numérotant convenablement, ou en utilisant l'algorithme de Cuthill-MacKee pour renuméroter, les degrés de liberté, on peut miminiser l'entier ℓ_B pour que la matrice $A^{\#}$ vérifie

$$\left[A^{\#}\right]_{ij} = 0 \quad \text{pour } |i-j| \geq \ell_B. \tag{3.41}$$

Dans le cas où $A^{\#}$ est symétrique, on dit qu'elle est **une matrice bande** avec une demi-largeur de bande égale à ℓ_B. Elle peut être stockée à l'aide d'un tableau à ℓ_B lignes et N_S colonnes. La correspondance entre les coefficients de $A^{\#}$ et du tableau servant à la stocker est schématisée comme suit pour $\ell_B = 3$

$$\left[\begin{array}{ccccccc}
a_{11} & a_{12} & a_{13} & & & & \\
\times & a_{22} & a_{23} & & & & \\
\times & \times & a_{33} & \vdots & a_{i-2,i} & & \\
 & \times & \times & \vdots & a_{i-1,i} & \vdots & \\
 & & & \vdots & a_{ii} & \vdots & a_{N_S-2,N_S} \\
 & & & & & \vdots & a_{N_S-1,N_S} \\
 & & & & & & a_{N_SN_S}
\end{array}\right]$$

$$\left[\begin{array}{ccccccc}
* & * & a_{13} & \vdots & a_{i-2,i} & \vdots & a_{N_S-2,N_S} \\
* & a_{12} & a_{23} & \vdots & a_{i-1,i} & \vdots & a_{N_S-1,N_S} \\
a_{11} & a_{22} & a_{33} & \vdots & a_{ii} & \vdots & a_{N_SN_S}
\end{array}\right]$$

L'étoile $*$ indique les entrées du tableau non utilisées, la croix $\times$ le coefficient a_{ij} avec $i > j$ répéré par la valeur a_{ji} qui est stockée.

Les formules suivantes permettent de passer du coefficient a_{ij} ($i \leq j$ et $j - i \leq \ell_B$) à l'entrée du tableau $A(i_0, j_0)$ qui sert à le stocker

$$j_0 = j \text{ (même colonne)}, \tag{3.42}$$

$$i_0 = \ell_B + i - j. \tag{3.43}$$

On peut se rappeler les formules précédentes à partir des deux repères suivants

– les indices de colonnes sont les mêmes
– les coefficients de la diagonale de la matrice sont sur la dernière ligne du tableau (ligne ℓ_B).

3.5. Le système linéaire final. Après assemblage des différentes matrices , on obtient la matrice totale

$$A^{\#} = K^{\#} + K_C^{\#} + M^{\#}. \tag{3.44}$$

Avec des notations évidentes, on a aussi

$$L^{\#} = F^{\#} + F_C^{\#}. \tag{3.45}$$

Le système variationnel (3.1) peut être réécrit alors

$$\left\{ \begin{array}{ll} \left[\begin{array}{c} u_1^h \\ \vdots \\ u_{N_S}^h \end{array} \right] \in \mathbb{R}^{N_S}, & \forall \left[\begin{array}{c} v_1^h \\ \vdots \\ v_{N_S}^h \end{array} \right] \in \mathbb{R}^{N_S}, \\ u_i^h = g_i^h := g(a_i^h),\ i \in I_D, & v_i^h = 0,\ i \in I_D, \\ \left[\begin{array}{ccc} v_1^h & \cdots & v_{N_S}^h \end{array} \right] A^{\#} \left[\begin{array}{c} u_1^h \\ \vdots \\ u_{N_S}^h \end{array} \right] & = \left[\begin{array}{ccc} v_1^h & \cdots & v_{N_S}^h \end{array} \right] L^{\#} \end{array} \right. \tag{3.46}$$

Renumérotons les sommets $a_1^h, \ldots a_{N_S}^h$ de la façon suivante
- l'ordre initial n'est pas modifié,
- les sommets correspondant à des degrés de liberté bloqués (i.e. pour $i \in I_D$) sont numérotés en dernier.

En pratique, on peut utiliser l'algorithme (Alg. 2) pour effectuer cette renumérotation.

Algorithm 2 Renumérotation des degrés de liberté

$N := 0$; {Compteur des degrés de libertés libres}
IDL $:= 0$; {Vecteur de longueur N_S}
do $i = 1, N_S$
  if $i \notin I_D$ **then**
    $N := N + 1$; IDL$(i) := N$;
  end if
enddo

A la fin de la boucle, N contiendra le nombre de degrés de liberté non bloqués, $IDL(i) = 0$ si le degré de liberté i est bloqué et IDL$(i) > 0$ donne son numéro dans la nouvelle numérotation sinon.

Le système (3.46) pourra être ainsi réécrit dans cette nouvelle numérotation

$$\left\{ \begin{array}{l} \left[\begin{array}{c} u_{i_1}^h \\ \vdots \\ u_{i_N}^h \\ g_{i_{N+1}}^h \\ \vdots \\ g_{i_{N_S}}^h \end{array} \right] \in \mathbb{R}^{N_S}, \quad \forall \left[\begin{array}{c} v_{i_1}^h \\ \vdots \\ v_{i_N}^h \\ 0 \end{array} \right] \in \mathbb{R}^{N_S}, \\ \left[\begin{array}{cccc} v_{i_1}^h & \cdots & v_{i_{N_S}}^h & 0 \end{array} \right] \left[\begin{array}{cc} A & A_{12} \\ A_{21} & A_{22} \end{array} \right] \left[\begin{array}{c} u_{i_1}^h \\ \vdots \\ u_{i_N}^h \\ g_{i_{N+1}}^h \\ \vdots \\ g_{i_{N_S}}^h \end{array} \right] = \left[\begin{array}{cccc} v_{i_1}^h & \cdots & v_{i_{N_S}}^h & 0 \end{array} \right] \left[\begin{array}{c} L \\ L_2 \end{array} \right] \end{array} \right. \tag{3.47}$$

En effectuant ce produit matriciel par blocs, on trouve

$$\left[\begin{array}{ccc} v^h_{i_1} & \cdots & v^h_{i_{N_S}} \end{array}\right] Ax = \left[\begin{array}{ccc} v^h_{i_1} & \cdots & v^h_{i_{N_S}} \end{array}\right] b,\ \forall \left[\begin{array}{c} v^h_{i_1} \\ \vdots \\ v^h_{i_N} \end{array}\right] \in \mathbb{R}^N, \tag{3.48}$$

avec

$$x := \left[\begin{array}{c} u^h_{i_1} \\ \vdots \\ u^h_{i_N} \end{array}\right], \quad b := L - A_{12} \left[\begin{array}{c} g^h_{i_{N+1}} \\ \vdots \\ g^h_{i_{N_S}} \end{array}\right] \tag{3.49}$$

ou encore

$$Ax = b. \tag{3.50}$$

Dans ce système,

- A est la **matrice finale,** obtenue à partir de la matrice totale par suppression des lignes et des colonnes correspondant à des degrés de liberté bloqués,
- L est le vecteur chargement final, obtenu de même par suppression des lignes correspondant à des degrés de liberté bloqués.

De façon pratique, le système (3.50) est obtenu comme suit.

(1) On forme les matrices totales $A^\#$ et $L^\#$. (Cette formation peut être explicite seulement pour les matrices finales.)

(2) On multiplie les colonnes de $A^\#$ correspondant à des degrés de liberté bloqués par les valeurs de ces degrés de liberté et on les retranche de $L^\#$. (Cette opération peut être effectuée en cours d'assemblage.)

(3) On supprime les lignes et colonnes corespondant à des degrés de liberté bloqués.

Le fragment de pseudo-code FORTRAN 90 (Alg. 3) illustre les points précédents. Il faudra, en pratique, ajouter une structure de données permettant de prendre en compte le caractère creux de la matrice finale A. De même, il faudra ajouter les contributions des raideurs ou des chargements concentrés éventuels par une boucle sur les arêtes frontière comme celle effectuée ici sur les éléments.

Algorithm 3 Formation des matrices finales

$A := 0$; $L := 0$ {Initialisation}
do $e = 1, N_E$ {Boucle sur les éléments}
  $A_{\text{elem}} :=$ formK(e) ; $L_{\text{elem}} :=$ formL(e) ; {Formation des matrices élémentaires}
  do $i_e = 1, \ell$ {Boucle sur les lignes de la matrice élémentaire}
    $i_t := n(e, i_e)$; $i := IDL(i_t)$;
    if $i \neq 0$ **then**
      do $j_e = 1, \ell$ {Boucle sur les colonnes de la matrice élémentaire}
        $j_t := n(e, j_e)$; $j := IDL(j_t)$;
        if $j \neq 0$ **then**
          $A(i, j) := A(i, j) + A_{\text{elem}}(i_e, j_e)$;
        else
          $L(i) := L(i) - A_{\text{elem}}(i_e, j_e) * G(i_t)$;
        end if
      enddo
      $L(i) := L(i) + L_{\text{elem}}(i_e)$;
    end if
  enddo
enddo

CHAPITRE 5

Eléments d'analyse fonctionnelle

Dans la plupart des problèmes, des estimations explicites comme l'inégalité de Poincaré au chapitre I ne sont pas disponibles. De même, certaines équations variationnelles ne possèdent pas de propriétés de coercivité. Le but de ce chapitre est de donner quelques compléments d'analyse fonctionnelle qui permettent de traiter ce type de situation. La notion essentielle introduite ici est celle de convergence faible. Le théorème de Riesz, établi ci-dessous, permet de justifier l'introduction de cette notion. On se limite pour simplifier l'exposé et les démonstrations aux espaces de Hilbert mais les résultats s'étendent au cas des espaces de Banach dits réflexifs (voir le livre de Haim Brezis "*Analyse fonctionnelle*", Masson).

Tous les éléments d'analyse fonctionnelle, que nous allons énoncer, sont valables, avec les adaptations requises, dans un espace de Banach, c'est-à-dire un espace muni d'une norme (espace normé) complet dont la norme n'est pas nécessairement associée à un produit scalaire. Cependant, les démonstrations sont considérablement simplifiées dans le cadre hilbertien. C'est pourquoi, sauf si la structure hilbertienne n'amène aucune simplification spécifique, nous énoncerons ces éléments d'analyse fonctionnelle pour un espace de Hilbert. Nous noterons X cet espace de Hilbert, $(\cdot,\cdot)_X$ son produit scalaire et $\|\cdot\|_X$ le produit scalaire associé. Lorsque le cadre hilbertien n'amène rien de particulier, nous utiliserons un espace de Banach E dont nous noterons la norme de façon analogue à celle de X.

1. Le théorème de Riesz

On a d'abord le résultat suivant qui établit en fait que tous les espaces normés de dimension finie peuvent être considérés comme des copies de $\mathbb{R}^m$.

THÉORÈME 5.1. *Soit E un espace normé de dimension finie m et $e_1,\dots, e_m$ une base de E. Alors, il existe deux constantes γ_0 et $\gamma_1 > 0$ telles que*

$$\gamma_0 \left(\sum_{i=1}^m |x_i|^2\right)^{1/2} \leq \|x\|_E \leq \gamma_1 \left(\sum_{i=1}^m |x_i|^2\right)^{1/2} \qquad si \quad x = x_1 e_1 + \cdots + x_m e_m.$$

DÉMONSTRATION. L'estimation de droite s'obtient directement par d'abord

$$\|x\|_E = \|x_1 e_1 + \cdots + x_m e_m\|_E \leq |x_1| \, \|e_1\|_E + \cdots + |x_m| \, \|e_m\|_E$$

puis par l'inégalité de Cauchy-Schwarz discrète

$$\|x\| \leq \underbrace{\left(\sum_{i=1}^m \|e_i\|_E^2\right)^{1/2}}_{=\gamma_1} \left(\sum_{i=1}^m |x_i|^2\right)^{1/2}. \tag{1.1}$$

L'estimation de gauche

$$\gamma_0 \left(\sum_{i=1}^m |x_i|^2\right)^{1/2} \leq \|x\|_E \tag{1.2}$$

s'obtient de façon indirecte en utilisant un raisonnement par l'absurde. Si elle était fausse, il existerait une suite $\left\{x^{(n)} = x_1^{(n)} e_1 + \cdots + x_m^{(n)} e_m\right\}_{n\geq 0}$ telle que

$$\sum_{i=1}^m \left|x_i^{(n)}\right|^2 = 1 \text{ et } \lim_{n\to 0} \left\|x^{(n)}\right\|_E = 0.$$

On déduit de

$$\left|x_\ell^{(n)}\right| \leq \left(\sum_{i=1}^m \left|x_i^{(n)}\right|^2\right)^{1/2} = 1 \quad (\ell = 1, \ldots, m)$$

que toutes les suites numériques $\left\{x_\ell^{(n)}\right\}_{n\geq 0}$ sont bornées. Le théorème de Bolzano-Weierstrass assure alors que, quitte à passer à une suite extraite, on peut supposer que

$$\lim_{n\to\infty} x_\ell^{(n)} = x_\ell \ (\ell = 1, \ldots, m) \ \text{ avec } \sum_{i=1}^m |x_i|^2 = 1.$$

Il s'ensuit donc que

$$\lim_{n\to\infty} x^{(n)} = x \neq 0.$$

Ceci est en contradiction avec l'hypothèse $\lim_{n\to 0} \left\|x^{(n)}\right\|_E = 0$ et établit de ce fait (1.2). □

REMARQUE 5.1. *La démonstration précédente est exemplaire des deux classes de méthodes permettant d'établir des estimations. L'estimation (1.1) s'obtient de façon directe avec une évaluation explicite de γ_1. L'estimation (1.2) est démontrée par un raisonnement par l'absurde à l'aide d'une constante γ_0 dont l'existence est établie uniquement de façon théorique.*

On déduit immédiatement de ce théorème.

COROLLAIRE 5.1. *Soit E un espace normé de dimension finie. Alors, de toute suite bornée $\left\{x^{(n)}\right\}_{n\geq 0}$ de E, i.e. telle que*

$$\exists M : \left\|x^{(n)}\right\|_E \leq M \quad \forall n \geq 0,$$

on peut extraire une sous-suite convergente.

DÉMONSTRATION. A démontrer à titre d'exercice. □

On dit qu'un sous-ensemble K d'un espace normé E est compact si de toute suite $\left\{x^{(n)}\right\}_{n\geq 0} \subset K$, on peut extraire une suite convergente vers un élément x dans K. Désignons par

$$B_r(x) := \{v \in E\,;\ \|v - x\|_E \leq r\}$$

la boule centrée en x de rayon r d'un espace normé E. La boule unité de E est alors simplement la boule centrée en 0 de rayon 1. Il est facile de voir que si la boule unité d'un espace normé est compacte, alors tout sous-ensemble fermé est compact. Le résultat du corollaire précédent exprime que la boule unité d'un espace normé de dimension finie est compact.

Le théorème de Riesz établit que ce résultat n'est vrai que pour les espaces de dimension finie. Cette propriété, qui peut sembler être un résultat négatif à première vue, est cependant à la base d'une collection d'outils d'analyse fonctionnelle très puissants incluant l'alternative de Fredholm et la théorie spectrale des opérateurs compacts. Nous aurons besoin de quelques préparatifs pour établir ce théorème.

On a besoin aussi du lemme suivant caractérisant un espace normé de dimension infinie.

LEMME 5.1. *Soit X un espace préhilbertien (espace muni d'un produit scalaire non nécessairement complet). Si E est de dimension infinie. Alors, il existe une suite de points $\left\{x^{(n)}\right\}_{n\geq 0}$ telle que*

$$\left\|x^{(n)}\right\|_X = 1 \quad n \geq 0, \tag{1.3}$$

$$\left\|x^{(n)} - x^{(m)}\right\|_X \geq 1 \quad n \neq m. \tag{1.4}$$

DÉMONSTRATION. L'espace X est de dimension infinie. Il possède donc une famille libre $\left\{u^{(m)}\right\}_{m\geq 0}$ ayant un nombre dénombrable d'éléments. On orthonormalise cette suite par l'algorithme de Schmidt qu'on rappelle ci-dessous

$$\begin{array}{l} x^{(0)} := u^{(0)} / \left\|u^{(0)}\right\|_X \\ \left\{ \begin{array}{l} \ell = 0, 1, \ldots \\ w^{(\ell+1)} := u^{(\ell+1)} - \sum_{j=0}^{j=\ell} (u^{(\ell+1)}, x^{(\ell)})_E x^{(\ell)} \\ x^{(\ell+1)} := w^{(\ell+1)} / \left\|w^{(\ell+1)}\right\|_X \end{array} \right. \end{array}$$

On peut démontrer facilement par récurrence que la suite $\left\{x^{(m)}\right\}_{m\geq 0}$ est bien définie, qu'elle engendre le même sous-espace que $\left\{u^{(m)}\right\}_{m\geq 0}$ et que c'est une famille orthonormée

$$(x^{(n)}, x^{(m)})_X = \delta_{nm} = \left\{ \begin{array}{ll} 1 & \text{si } n = m, \\ 0 & \text{sinon} \end{array} \right.$$

Pour terminer la démonstration, il suffit d'écrire que pour $n \neq m$

$$\begin{aligned} \left\|x^{(n)} - x^{(m)}\right\|_X &= \sqrt{(x^{(n)} - x^{(m)}, x^{(n)} - x^{(m)})_X} = \sqrt{\|x^{(n)}\|_X^2 + \|x^{(m)}\|_X^2} \\ &= \sqrt{2} \geq 1. \end{aligned}$$

□

THÉORÈME 5.2 (Théorème de Riesz). *Soit un espace préhilbertien X. Si la boule unité de X est compacte, alors X est de dimension finie.*

DÉMONSTRATION. C'est une simple conséquence du lemme précédent qui permet d'affirmer que, si X est de dimension infinie, il possède une suite bornée telle que

$$\left\|x^{(m)} - x^{(n)}\right\|_X \geq 1.$$

Aucune suite extraite ne peut ainsi être une suite de Cauchy et ne peut donc être convergente. □

REMARQUE 5.2. *Le lemme et le théorème précédent sont valables pour un espace normé mais la démonstration est plus compliquée (voir J. Dieudonné,* "Eléments d'Analyse, *Tome 1").*

2. Convergence faible

2.1. Définition et premières propriétés. Le théorème de Riesz montre que si $\left\{x^{(n)}\right\}_{n\geq 0}$ est une suite bornée de X, on n'est pas sûr qu'on puisse extraire une sous-suite $\left\{y^{(n)}\right\}_{n\geq 0}$ de $\left\{x^{(n)}\right\}_{n\geq 0}$ qui soit convergente. Rappelons que $\left\{y^{(n)}\right\}_{n\geq 0}$ est obtenue à l'aide de $\left\{x^{(n)}\right\}_{n\geq 0}$ par $y^{(n)} = x^{(g(n))}$ où g est une application strictement croissante de $\mathbb{N}$ dans $\mathbb{N}$. Pour pallier à cette insuffisance, on introduit la notion de convergence faible.

On dit que la suite $\left\{x^{(n)}\right\}_{n\geq 0} \subset X$ **converge faiblement** vers x dans X si

$$\forall y \in X, \quad \lim_{n\to\infty} (x^{(n)}, y)_X = (x, y)_X. \tag{2.1}$$

On note alors

$$\text{w-}\lim_{n\to\infty} x^{(n)} = x. \tag{2.2}$$

Pour distinguer entre cette convergence et la convergence au sens usuel pour la norme de X, on dira que $\left\{x^{(n)}\right\}_{n\geq 0}$ converge fortement lorsqu'elle converge pour la norme.

REMARQUE 5.3. *Considérons une base hilbertienne $\left\{e^{(n)}\right\}_{n\geq 0}$ d'un espace de Hilbert X, i.e. $(e^{(n)}, e^{(m)})_E = \delta_{nm}$ et*

$$\lim_{m\to\infty} \sum_{n=0}^{m} (e^{(n)}, v)_X \, e^{(n)} = v$$

fortement dans X pour tout $v \in X$. (L'algorithme de Gram-Schmidt, donné ci-dessus, assure qu'une telle base existe toujours si X est séparable, i.e. s'il contient une famille dénombrable partout dense.) L'égalité de Bessel-Parseval montre alors que

$$\sum_{n=0}^{\infty} \left|(e^{(n)}, v)_X\right|^2 = \|v\|_X^2 .$$

Le terme général de la série numérique convergente tend donc vers 0. On a ainsi

$$\lim_{n\to\infty} (e^{(n)}, v)_X = 0 \quad \forall v \in X.$$

Ceci exprime exactement que

$$w\text{-}\lim_{n\to\infty} e^{(n)} = 0.$$

On a immédiatement la propriété suivante sans laquelle la notion de convergence faible aurait été inutilisable.

PROPOSITION 5.1 (Unicité de la limite faible). *Soit $\left\{x^{(n)}\right\}_{n\geq 0}$ une suite de X. Alors*

$$w\text{-}\lim_{n\to\infty} x^{(n)} = x \quad \textit{et} \quad w\text{-}\lim_{n\to\infty} x^{(n)} = y \quad \textit{entrainent } x = y. \tag{2.3}$$

DÉMONSTRATION. Sous les conditions de la proposition, on a

$$\lim_{n\to\infty} (x^{(n)}, x - y)_X = (x, x - y)_X = (y, x - y)_X,$$

et donc par différence $(x - y, x - y)_E = 0$. D'où la proposition. □

Les propriétés suivantes sont analogues aux propriétés usuelles et se vérifient directement.

PROPOSITION 5.2. *La propriété suivante est vérifiée*

$$w\text{-}\lim_{n\to\infty} (\lambda x^{(n)} + \mu y^{(n)}) = \lambda \, w\text{-}\lim_{n\to\infty} x^{(n)} + \mu \, w\text{-}\lim_{n\to\infty} y^{(n)}. \tag{2.4}$$

Enfin, comme son nom l'indique la convergence faible est plus faible que la convergence forte.

PROPOSITION 5.3. *On a*

$$\lim_{n\to\infty} x^{(n)} = x \quad \textit{entraine} \quad w\text{-}\lim_{n\to\infty} x^{(n)} = x. \tag{2.5}$$

DÉMONSTRATION. Il suffit d'observer que pour tout $v \in X$

$$\left|(x^{(n)}, v)_E - (x, v)_E\right| = \left|(x^{(n)} - x, v)_E\right| \leq \left\|x^{(n)} - x\right\|_E \|v\|_E$$

pour établir la proposition. □

2.2. Toute suite faiblement convergente est bornée. On sait qu'une suite fortement convergente est bornée. De façon assez surprenante, la même propriété est vérifiée par une suite qui est seulement faiblement convergente. Cependant, la démonstration de ce fait utilise un outil puissant d'analyse fonctionnelle : le théorème de Banach-Steinhauss.

Le théorème de Banach-Steinhauss s'obtient à l'aide du théorème de Baire relatif à la topologie des espaces métriques complets qu'on énonce ici dans le contexte des espaces de Banach.

THÉORÈME 5.3 (Version simplifiée du théorème de Baire). *Soit E un espace de Banach. Si $\{X_n\}_{n\geq 0}$ est une suite de sous-ensembles fermés de E tels que*

$$\bigcup_{n\geq 0} X_n = E.$$

Alors, il existe $r > 0$, u et $n_0 \geq 0$ tel que

$$B_r(u) \subset X_{n_0}.$$

DÉMONSTRATION. Voir par exemple H. Brezis, "Analyse fonctionnelle. Théorie et applications", Masson (1983). □

On a alors le résultat fondamental suivant.

THÉORÈME 5.4 (Théorème de Banach-Steinhauss). *Soient E un espace de Banach, F un espace normé et $\{T_a\}_{a\in A}$ une famille d'opérateurs linéaires continus de E dans F vérifiant*

$$\sup_{a\in A} \|T_a(v)\|_F < +\infty \quad \forall v \in E. \tag{2.6}$$

Alors,

$$\sup_{a\in A} \|T_a\|_{\mathcal{L}(E,F)} < +\infty \tag{2.7}$$

avec

$$\|T_a\|_{\mathcal{L}(E,F)} = \sup_{\|v\|_E \leq 1} \|T_a v\|_F .$$

DÉMONSTRATION. Notons pour $n \geq 0$

$$X_n := \left\{ v \in E;\ \sup_{a\in A} \|T_a(v)\|_F \leq n \right\}.$$

On établit facilement (exercice) que X_n est fermé. Il résulte de (2.6) que

$$\bigcup_{n\geq 0} X_n = E.$$

Le théorème de Baire assure alors qu'il existe x_0, $r_0 > 0$ et n_0 tel que

$$\sup_{a\in A} \|T_a(x_0 + r_0 v)\|_F \leq n_0 \quad \forall v \text{ vérifiant } \|v\|_E < 1.$$

Cela induit en particulier que

$$\sup_{a\in A} \|T_a(x_0)\|_F \leq n_0.$$

Il vient alors pour tout v vérifiant $\|v\|_E < 1$

$$\begin{aligned} r_0 \|T_a(v)\|_F &= \|T_a(x_0 + r_0 v) - T_a x_0\|_F \\ &\leq n_0 + \|T_a(x_0)\|_F \leq 2n_0 \quad \forall a \in A, \ . \end{aligned}$$

On montrera à titre d'exercice que l'inégalité précédente est en fait vraie pour $\|v\|_E \leq 1$. Ceci donne alors

$$\|T_a\|_{\mathcal{L}(E,F)} \leq \frac{2n_0}{r_0}$$

et ainsi

$$\sup_{a\in A}\|T_a\|_{\mathcal{L}(E,F)} \leq \frac{2n_0}{r_0}.$$

D'où (2.7). □

On peut alors démontrer le résultat annoncé.

PROPOSITION 5.4. *Toute suite $\left\{x^{(n)}\right\}_{n\geq 0}$ d'un espace de Hilbert X, faiblement convergente, est bornée.*

DÉMONSTRATION. On définit $T_n \in X'$ par $T_n := J_X x^{(n)}$ où J_X est l'isométrie naturelle de X dans son dual X'. Rappelons que $T_n := J_X x^{(n)}$ est défini par

$$T_n v = (x^{(n)}, v)_X, \quad \forall v \in X$$

et qu'on a

$$\sup_{\|v\|_X\leq 1} |T_n v| = \left\|x^{(n)}\right\|_X.$$

Comme la suite numérique $\{T_n v\}_{n\geq 0}$ est convergente, on a donc

$$\sup_{n\geq 0} |T_n v| < +\infty.$$

Le théorème de Banach-Steinhauss donne alors

$$\sup_{n\geq 0}\sup_{\|v\|_E\leq 1} |T_n v| = \sup_{n\geq 0}\left\|x^{(n)}\right\|_X < +\infty.$$

Ceci établit la proposition. □

On en déduit les résultats suivants qui sont très importants pour les applications.

PROPOSITION 5.5. *Les propriétés suivantes sont vérifiées :*

(1) $\lim_{n\to\infty} \ell^{(n)} = \ell$ *fortement dans* X' *et* $w\text{-}\lim_{n\to\infty} x^{(n)} = x$ *dans* X *entrainent*

$$\lim_{n\to\infty} \ell^{(n)}(x^{(n)}) = \ell(x), \tag{2.8}$$

(2) *Soient X et Y deux espaces de Hilbert et une application bilinéaire a, continue de $X \times Y$ dans $\mathbb{R}$. Alors,* $\lim_{n\to\infty} x^{(n)} = x$ *fortement dans X et* $w\text{-}\lim_{n\to\infty} y^{(n)} = y$ *dans Y entrainent*

$$\lim_{n\to\infty} a(x^{(n)}, y^{(n)}) = a(x, y). \tag{2.9}$$

DÉMONSTRATION. On écrit

$$\begin{aligned}\left|\ell^{(n)}(x^{(n)}) - \ell(x)\right| &= \left|\ell^{(n)}(x^{(n)}) - \ell(x^{(n)}) + \ell(x^{(n)}) - \ell(x)\right| \\ &\leq \left|\ell^{(n)}(x^{(n)}) - \ell(x^{(n)})\right| + \left|\ell(x^{(n)}) - \ell(x)\right| \\ &\leq \left\|\ell^{(n)} - \ell\right\|_{X'}\left\|x^{(n)}\right\|_X + \left|\ell(x^{(n)}) - \ell(x)\right|.\end{aligned}$$

On utilise alors l'isométrie J_X pour écrire

$$\left|\ell^{(n)}(x^{(n)}) - \ell(x)\right| \leq \left\|\ell^{(n)} - \ell\right\|_{X'}\left\|x^{(n)}\right\|_X + \left|(x^{(n)}, J_X^{-1}\ell)_X - (x, J_X^{-1}\ell)_X\right|.$$

Comme la proposition précédente donne que $\left\{x^{(n)}\right\}_{n\geq 0}$ est bornée, on obtient

$$\left|\ell^{(n)}(x^{(n)}) - \ell(x)\right| \leq M\left\|\ell^{(n)} - \ell\right\|_{E'} + \left|(x^{(n)}, J_X^{-1}\ell)_X - (x, J_X^{-1}\ell)_X\right|.$$

D'où (2.8) en passant à la limite sur le second membre.

La forme bilinéaire continue est définie de façon équivalente par $A \in \mathcal{L}(E, F')$ et

$$Ax(y) = a(x, y) \quad \forall x \in E,\ \forall y \in F.$$

Comme $\ell^{(n)} = Ax^{(n)}$ converge fortement vers $\ell = Ax$ lorsque $n \to \infty$, la propriété (2.9) est une conséquence immédiate de (2.8). □

3. Compacité faible

Nous en venons maintenant à la raison principale de l'introduction de la convergence faible. Elle permet de retrouver la propriété de pouvoir extraire de toute suite bornée une sous-suite convergente. Ce résultat s'exprime en disant que la boule unité d'un espace de Hilbert est faiblement séquentiellement compacte. Cette propriété, et les préparatifs qui servent à l'établir, requièrent des démonstrations à un niveau un peu plus avancé que celui auquel se situe ce cours. La recommandation est donc de se concentrer sur la compréhension de l'énoncé du théorème 5.5 qui donne le résultat principal de cette section.

3.1. Convergence d'une suite d'opérateurs uniformément bornés. La proposition suivante établit qu'une suite d'opérateurs d'un espace normé à valeurs dans un espace de Banach uniformément bornés, converge dès qu'elle converge pour les éléments d'un sous-ensemble $\mathcal{B}$ dont les combinaisons linéaires forment $\mathcal{V}$ un sous-espace dense de E.

PROPOSITION 5.6. *Soit E un espace normé et F un espace de Banach. Soit $\mathcal{B}$ une famille de la boule unité de E telle que $\mathcal{V}$ le sous-espace engendré par $\mathcal{B}$ est dense dans E. Alors, si $\{T_n\}_{n\geq 0} \subset \mathcal{L}(E, F)$ vérifie*

$$\|T_n v\|_F \leq M \|v\|_E \quad \forall v \in E,\ \forall n \geq 0, \tag{3.1}$$

$$\lim_{n\to\infty} T_n v = Tv \quad \forall v \in \mathcal{B} \tag{3.2}$$

Alors, T peut être prolongé sur E en un opérateur dans $\mathcal{L}(E, F)$ vérifiant

$$\|Tv\|_F \leq M \|v\|_E \quad \forall v \in E. \tag{3.3}$$

DÉMONSTRATION. Observons que la condition (3.2) permet de définir T sur $\mathcal{V}$. Suite à (3.1) et (3.2), il vérifie

$$\|Tv\|_F \leq M \|v\|_E \quad \forall v \in \mathcal{V}. \tag{3.4}$$

Soit maintenant $v \in E$. Il existe donc $\left\{v^{(n)}\right\}_{n\geq 0} \subset \mathcal{V}$ telle que $\lim_{n\to\infty} v^{(n)} = v$. Il résulte de (3.4) que $\left\|Tv^{(n+p)} - Tv^{(n)}\right\|_F \leq M \left\|v^{(n+p)} - v^{(n)}\right\|_E$ et donc que $\left\{Tv^{(n)}\right\}_{n\geq 0}$ est une suite de Cauchy convergente donc puisque F est de Banach. De même si $\left\{w^{(n)}\right\}_{n\geq 0} \subset \mathcal{V}$ est une autre suite convergeant vers v, $\left\|Tw^{(n)} - Tv^{(n)}\right\|_F \leq M \left\|w^{(n)} - v^{(n)}\right\|_E$ montre que la limite $\lim_{n\to\infty} Tv^{(n)} = Tv$ ne dépend pas de $\left\{v^{(n)}\right\}_{n\geq 0} \subset \mathcal{V}$ et définit un prolongement de $T \in \mathcal{L}(E, F)$ vérifiant (3.3). □

3.2. Compacité faible de la boule unité d'un espace de Banach réflexif. Nous aurons besoin du procédé diagonal qu'on rappelle ci-dessous.

PROPOSITION 5.7 (Procédé diagonal). *Soit U un ensemble non vide et $\left\{x^{(n)}\right\}_{n\geq 0}$ une suite d'éléments de U. Soit la famille de suites extraites $\left\{x^{(m,n)}\right\}_{n\geq 0}$ pour $m \geq 0$ construite de la façon suivante*

$$\left\{x^{(m+1,n)}\right\}_{n\geq 0} \text{ est une suite extraite de } \left\{x^{(m,n)}\right\}_{n\geq 0}.$$

Alors, la suite $\left\{x^{(n,n)}\right\}_{n\geq 0}$ *est une suite extraite de* $\left\{x^{(n)}\right\}_{n\geq 0}$ *telle que* $\left\{x^{(n,n)}\right\}_{n\geq k}$ *est une suite extraite de* $\left\{x^{(k,n)}\right\}_{n\geq 0}$.

DÉMONSTRATION. C'est une simple conséquence de la définition des suites extraites. □

Nous passons maintenant au résultat de compacité qui a justifié les développements précédents. Il est clair que la compacité séquentielle faible de la boule unité est équivalente à celle de n'importe quelle boule centrée à l'origine par normalisation.

THÉORÈME 5.5. *Soit* X *un espace de Hilbert. De toute suite bornée* $\left\{x^{(n)}\right\}_{n\geq 0} \subset X$, *on peut extraire une sous-suite faiblement convergente.*

DÉMONSTRATION. On traduit de façon plus précise l'hypothèse du théorème par

$$\text{il existe } M : \left\|x^{(n)}\right\|_X \leq M, \quad \text{pour tout } n \geq 0.$$

Notons par $\mathcal{V}$ le sous-espace de X engendré les éléments de la suite $\left\{x^{(n)}\right\}_{n\geq 0}$. Si $\mathcal{V}$ est de dimension finie, on sait qu'il coïncide avec son adhérence V. La suite $\left\{x^{(n)}\right\}_{n\geq 0}$ étant donc une suite contenue dans un espace normé de dimension finie, on peut y extraire une sous-suite fortement, et donc faiblement, convergente.

On peut supposer donc que $\mathcal{V}$ est de dimension infinie.

Notons par $\mathcal{B}$ les éléments de $\left\{x^{(n)}\right\}_{n\geq 0}$ qui ne peuvent pas s'écrire comme une combinaison linéaire des autres éléments de $\left\{x^{(n)}\right\}_{n\geq 0}$. On note par $\left\{w^{(m)}\right\}_{m\geq 0}$ les éléments de $\mathcal{B}$.

La construction suivante utilise le procédé diagonal. On a

$$\left|(x^{(n)}, w^{(0)})_X\right| \leq \left\|x^{(n)}\right\|_X \leq M.$$

On peut donc extraire de $\left\{x^{(n)}\right\}_{n\geq 0}$ une sous-suite $\left\{x^{(0,n)})\right\}_{n\geq 0}$ telle que la suite numérique $\left\{(x^{(0,n)}, w^{(0)})_X\right\}_{n\geq 0}$ converge vers $\alpha^{(0)}$. De même, on a

$$\left|(x^{(0,n)}, w^{(1)})_X\right| \leq \left\|x^{(0,n))}\right\|_X \leq M.$$

On peut de même extraire de $\left\{x^{(0,n)}\right\}_{n\geq 0}$ une sous-suite $\left\{x^{(1,n)}\right\}_{n\geq 0}$ telle que $\left\{(x^{(1,n)}, w^{(1)})\right\}_{n\geq 0}$ converge vers $\alpha^{(1)}$. Continuant ainsi, on définit donc pour tout $k \geq 1$ une suite $\left\{x^{(k,n)}\right\}_{n\geq 0}$ extraite de $\left\{x^{(k-1,n)}\right\}_{n\geq 0}$ telle que $\left\{(x^{(k,n)}, w^{(k)})_X\right\}_{n\geq 0}$ converge vers $\alpha^{(k)}$.

On considère alors la suite extraite de $\left\{x^{(n)}\right\}_{n\geq 0}$ définie par (procédé diagonal)

$$\left\{x^{(n,n)}\right\}_{n\geq 0}.$$

On a donc $\lim_{n\to\infty}(x^{(n,n)}, w^{(j)})_X = \alpha^{(j)}$ pour tout $j \geq 0$.

Notons $\ell^{(n)} := J_V x^{(n,n)}$. La proposition 5.6 montre que la limite de $\left\{\ell^{(n)}\right\}_{n\geq 0}$ est définie comme élément de V', c'est-à-dire

$$\lim_{n\to\infty}(x^{(n,n)}, v)_X = \lim_{n\to\infty}\ell^{(n)}(v) = \ell(v) = (x, v)_X \quad \forall v \in V$$

avec $x := J_V^{-1}\ell$. Comme pour $v \in V^{\perp}$, le sous-espace orthogonal à V, on a $(x^{(n,n)}, v)_X = 0$ et $(x, v)_X$, on a en fait

$$\lim_{n\to\infty}(x^{(n,n)}, v)_X = (x, v)_X \quad \forall v \in X.$$

Ceci exprime exactement que

$$\text{w-}\lim_{n\to\infty} x^{(n,n)} = x.$$

□

4. Opérateurs compacts

L'intérêt de la convergence faible vient du fait que, dans certaines conditions, on est capable de déduire une convergence forte d'une convergence faible. Avant d'examiner ce point, démontrons d'abord le résultat suivant.

PROPOSITION 5.8. *Soient X et Y deux espaces de Hilbert et $T \in \mathcal{L}(X,Y)$. Alors,*

$$w\text{-}\lim_{n\to\infty} x^{(n)} = x \text{ dans } X \text{ entraine } w\text{-}\lim_{n\to\infty} Tx^{(n)} = Tx \text{ dans } Y. \tag{4.1}$$

On exprime ce résultat en disant que tout opérateur linéaire continu est faiblement séquentiellement continu.

DÉMONSTRATION. Soit $v \in F$. On a $(Tx^{(n)}, v)_Y = (x^{(n)}, T^*v)_X$ par définition de l'adjoint. Si w-$\lim_{n\to\infty} x^{(n)} = x$ dans X, il vient

$$\lim_{n\to\infty} (Tx^{(n)}, v)_Y = \lim_{n\to\infty} (x^{(n)}, T^*v)_X = (x, T^*v)_X = (Tx, v)_Y.$$

Ceci exprime exactement la propriété (4.1). □

On dit qu'un **opérateur linéaire** T de E dans F est **compact**, s'**il transforme toute suite faiblement convergente en une suite fortement convergente.**

On a tout d'abord.

PROPOSITION 5.9. *Tout opérateur compact d'un espace de Hilbert X dans un espace normé F est continu.*

DÉMONSTRATION. Il suffit de montrer que

$$\sup_{\|v\|_X \le 1} \|Tv\|_X < +\infty.$$

Soit $\left\{v^{(n)}\right\}_{n\ge 0}$ une suite maximisant $\|Tv\|_X$ sur la boule unité, i.e. telle que

$$\lim_{n\to\infty} \left\|Tv^{(n)}\right\|_F = \sup_{\|v\|_X \le 1} \|Tv\|_F .$$

Quitte à passer à une suite extraite, on peut supposer que

$$\text{w-}\lim_{n\to\infty} v^{(n)} = z$$

avec $\|z\|_X \le 1$. Comme T est compact, on a

$$\lim_{n\to\infty} Tv^{(n)} = Tz,$$

ce qui implique que $\sup_{\|v\|_X \le 1} \|Tv\| = \|Tz\|$ et démontre la proposition. □

On dit que l'espace normé E s'injecte continument dans un espace normé F si E s'identifie à un sous-espace de F avec une norme plus fine, c'est-à-dire

$$\|v\|_F \le \|v\|_E \quad \forall v \in E,$$

ou encore

$$\lim_{n\to\infty} v^{(n)} = v \text{ dans } E \text{ entraine } \lim_{n\to\infty} v^{(n)} = v \text{ dans } F.$$

EXEMPLE 5.1. *L'espace $H^1(\Omega)$ s'injecte continument dans $L^2(\Omega)$.*

La proposition précédente montre que si l'espace de Hilbert X s'injecte continument dans l'espace de Hilbert Y, alors toute suite faiblement continue dans X est faiblement continue dans Y. On dira que l'injection de X dans Y est **compacte** si l'injection précédente est compacte, autrement dit si toute suite faiblement convergente dans X est fortement convergente dans Y.

On a le résultat fondamental suivant.

THÉORÈME 5.6 (Théorème de compacité de Rellich). *Soit Ω un domaine de $\mathbb{R}^N$ **borné**. Sous de larges hypothèses sur la régularité de la frontière de Ω (toujours satisfaites en pratique), l'injection de $H^1(\Omega)$ dans $L^2(\Omega)$ est compacte.*

DÉMONSTRATION. Admise. □

5. Applications

THÉORÈME 5.7 (Théorème de Peetre-Tartar). *Soient X, Y et G trois espaces de Hilbert, $T \in \mathcal{L}(X,Y)$ et S compact de X dans G tels que il existe une constante C_0*

$$\|v\|_X \leq C_0 \left\{\|Tv\|_Y + \|Sv\|_G\right\} \quad \forall v \in X. \tag{5.1}$$

Alors,

(1) $\ker T := \{v \in X \,:\, Tv = 0\}$ *est de dimension finie,*

(2) *il existe une constante C telle que*

$$\inf_{p \in \ker T} \|v + p\|_E \leq C \|Tv\|_F . \tag{5.2}$$

DÉMONSTRATION. Il est clair que $\ker T$ est un sous-espace fermé de X. C'est donc un espace de Hilbert pour le produit scalaire induit par celui de X. Soit $\left\{x^{(n)}\right\}_{n\geq 0}$ une suite dans la boule unité de $\ker T$. Quitte à passer à une suite extraite, on peut supposer que

$$\text{w-}\lim_{n\to\infty} x^{(n)} = x \text{ dans } \ker T.$$

Il vient ainsi

$$\left\|x^{(n)} - x\right\|_X \leq C_0 \left\|Sx^{(n)} - Sx\right\|_G$$

et comme S est compact $\lim_{n\to\infty} Sx^{(n)} = Sx$ dans G entrainant

$$\lim_{n\to\infty} x^{(n)} = x \text{ dans } \ker T.$$

La boule unité de $\ker T$ est ainsi compacte. Le théorème de Riesz assure donc que $\ker T$ est de dimension finie.

L'existence et l'unicité de la projection sur un sous-espace fermé d'un espace de Hilbert assurent que pour tout $v \in X$, il existe un et un seul $p_v \in \ker T$ tel que

$$\|v + p_v\|_X = \inf_{q \in \ker T} \|v + q\|_X \leq \|v + q\|_X \quad \forall q \in \ker T.$$

Supposons que (5.2) soit fausse. Il existe donc une suite $\left\{v^{(n)}\right\}_{n\geq 0} \subset X$ et une suite $\left\{p^{(n)}\right\}_{n\geq 0} \subset \ker T$ telles que

$$\left\|v^{(n)} + p^{(n)}\right\|_X = 1 \leq \left\|v^{(n)} + q\right\|_X \quad \forall q \in \ker T, \tag{5.3}$$

$$\lim_{n\to\infty} \left\|Tv^{(n)}\right\|_Y = 0. \tag{5.4}$$

Posons $w^{(n)} = v^{(n)} + p^{(n)}$. Comme $q - p^{(n)}$ est quelconque dans $\ker T$ si q est quelconque dans $\ker T$ et que $Tv^{(n)} = Tw^{(n)}$, les propriétés (5.3) et (5.4) peuvent être réécrites

$$\left\|w^{(n)}\right\|_X = 1 \leq \left\|w^{(n)} + q\right\|_X \quad \forall q \in \ker T, \tag{5.5}$$

$$\lim_{n\to\infty} \left\|Tw^{(n)}\right\|_Y = 0. \tag{5.6}$$

Comme X est un espace de Hilbert, quitte à passer à une sous-suite, on peut supposer que

$$\text{w-}\lim_{n\to\infty} w^{(n)} = w.$$

Comme S est compact, on a donc

$$\lim_{n\to\infty} Sw^{(n)} = Sw. \tag{5.7}$$

Mais l'hypothèse (5.1) entraine

$$\left\|w^{(n+p)} - w^{(n)}\right\|_E \leq C_0 \left\{\left\|Tw^{(n+p)} - Tw^{(n)}\right\|_F + \left\|Sw^{(n+p)} - Sw^{(n)}\right\|_G\right\}$$

et montre ainsi que la suite $\left\{w^{(n)}\right\}_{n\geq 0}$ est de Cauchy. Elle converge donc et, par suite de l'unicité de la limite faible, vers w. En prenant $q = -w$ dans (5.5), on obtient

$$1 \leq \left\|w^{(n)} - w\right\|_E.$$

ce qui est en contradiction avec

$$\lim_{n\to\infty} \left\|w^{(n)} - w\right\|_E = 0.$$

□

On va appliquer le résultat précédent pour démontrer l'inégalité de Poincaré suivante dans le cas d'un domaine Ω borné assez régulier de $\mathbb{R}^N$ $(N = 2, 3)$. La régularité, étant celle requise par le théorème d'injection compacte de Rellich, est assez large pour inclure tous les cas pouvant se présenter en pratique : frontière présentant des angles, des points de rebroussement, domaines avec des fissures, etc.

PROPOSITION 5.10. *Soit un domaine Ω comme ci-dessus et Γ_N et Γ_D une parttion sans recouvrement de sa frontière Γ avec la mesure $|\Gamma_D|$ de Γ_D non nulle. Notons par*

$$V := \{v \in \Omega\,;\ v|_{\Gamma_D} = 0\}.$$

Alors, il existe C telle que

$$\|v\|_{1,\Omega} \leq C\,|v|_{1,\Omega} \quad \forall v \in \Omega. \tag{5.8}$$

DÉMONSTRATION. Posons

$$Y = \underbrace{L^2(\Omega) \times \cdots \times L^2(\Omega)}_{N \text{ fois}}$$

On considère alors l'opérateur $T \in \mathcal{L}(V, F)$

$$Tv := (\partial_1 v, \cdots, \partial_N v)$$

et S l'injection de V dans $L^2(\Omega)$. Comme $\|Tv\|_F = |v|_{1,\Omega}$, on a ainsi par définition de la norme dans $H^1(\Omega)$

$$\begin{aligned}\|v\|_V := \|v\|_{1,\Omega} &= \left\{\|Tv\|_Y^2 + \|v\|_{L^2(\Omega)}^2\right\}^{1/2} \\ &\leq \left\{\|Tv\|_Y^2 + \|v\|_{L^2(\Omega)}^2 + 2\,\|Tv\|_Y\,\|v\|_{L^2(\Omega)}\right\}^{1/2} = \|Tv\|_Y + \|v\|_{L^2(\Omega)}.\end{aligned}$$

Le théorème d'injection compacte de Rellich assure S est compacte. Soit $p \in \ker T$, i.e.

$$\partial_j p = 0.$$

Ceci donne $p = \alpha$ avec α une constante. Mais comme $\alpha|_{\Gamma_D} = 0$ et que $|\Gamma_D| > 0$ ceci entraine que $\alpha = 0$. Ceci donne donc que T est injectif. Le lemme de Peetre-Tartar assure alors que l'estimation (5.8) est vérifiée. □

CHAPITRE 6

Systèmes et problèmes d'ordre supérieur

Nous examinons dans ce chapitre comment les méthodes d'éléments finis du chapitre précédent peuvent être étendues au cas des systèmes comportant plusieurs inconnues et équations et aux problèmes posés avec des dérivées d'ordre 4.

1. Systèmes de l'élasticité

1.1. Le problème variationnel. Pour rester simple, nous examinons le cas des systèmes de l'élasticité bidimensionnelle : problème en contraintes planes ou en déformations planes. Le système de l'élasticité tridimensionnelle peut être obtenu de façon analogue.

Pour décrire dans un seul cadre les systèmes en déformations planes et en contraintes planes, nous écrirons les équations telles qu'elles sont fournies par le principe des travaux virtuels.

Soit Ω un domaine du plan $\mathbb{R}^2$. On note par Γ sa frontière et $\mathbf{n}$ la normale unitaire à Γ orientée vers l'extérieur de Ω. Notons $L^2(\Omega;\mathbb{R}^2)$ l'espace des vecteurs

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \quad v_1, v_2 \in L^2(\Omega),$$

qu'on peut identifier à $L^2(\Omega) \times L^2(\Omega)$ et

$$H^1(\Omega;\mathbb{R}^2) := \left\{ \mathbf{v} \in L^2(\Omega;\mathbb{R}^2)\,;\ \partial_{x_j} v_i,\ i,j = 1,2 \right\}$$

qu'on peut de même identifier à $H^1(\Omega) \times H^1(\Omega)$. On se donne une partition sans recouvrement de Γ en Γ_D et Γ_N. On note alors

$$\mathbf{V} := \left\{ \mathbf{v} \in H^1(\Omega;\mathbb{R}^2)\,;\ \mathbf{v}|_{\Gamma_D} = 0 \right\}$$

le sous-espace des déplacements $\mathbf{v} \in H^1(\Omega;\mathbb{R}^2)$ vérifiant une condition d'encastrement sur Γ_D.

Le problème à résoudre s'énonce alors

$$\left\{ \begin{array}{l} \mathbf{u} \in \mathbf{V},\ \forall \mathbf{v} \in \mathbf{V}, \\ a(\mathbf{u},\mathbf{v}) = L\mathbf{v}, \end{array} \right. \tag{1.1}$$

où

$$a(\mathbf{u},\mathbf{v}) = \int_\Omega \left\{ 2\mu \textstyle\sum_{i,j=1}^2 \varepsilon_{ij}(\mathbf{u})\varepsilon_{ij}(\mathbf{v}) + \lambda \sum_{i=1}^2 \varepsilon_{ii}(\mathbf{u}) \sum_{i=1}^2 \varepsilon_{ii}(\mathbf{u}) \right\} d\Omega \tag{1.2}$$

$$L\mathbf{v} := \int_\Omega \mathbf{f}\cdot\mathbf{v}\, d\Omega + \int_{\Gamma_N} \mathbf{h}\cdot\mathbf{v}\, d\Gamma, \tag{1.3}$$

avec $\mathbf{f}$ donné dans $L^2(\Omega;\mathbb{R}^2)$, représentant la densité des forces extérieures appliquées à l'intérieur, $\mathbf{h}$ celle des forces appliquées sur le bord. On a noté de façon usuelle par $\mathbf{f}\cdot\mathbf{v} = f_1 v_1 + f_2 v_2$ et $\mathbf{h}\cdot\mathbf{v} = h_1 v_1 + h_2 v_2$ les produits scalaires respectifs de $\mathbf{f}$ et $\mathbf{v}$ et de $\mathbf{h}$ et $\mathbf{v}$. Les coefficients μ et λ dépendent du problème étudié. Ils sont donnés à l'aide du module d'Young E et du coefficient de Poisson ν selon le tableau suivant :

	2μ	λ
Contraintes planes (d épaisseur)	$Ed/(1+\nu)$	$Ed\nu/(1-\nu^2)$
Déformations planes	$E(1-\nu)/(1+\nu)(1-2\nu)$	$E\nu/(1+\nu)(1-2\nu)$

Les constantes μ et λ peuvent dépendre de $x \in \Omega$. Elles sont dans tous les cas dans $L^\infty(\Omega)$ et vérifient

$$\exists \mu_0 > 0 : \mu(x) \geq \mu_0, \quad \text{p.p. tout } x \in \Omega, \tag{1.4}$$

$$\lambda(x) \geq 0, \quad \text{p.p. tout } x \in \Omega. \tag{1.5}$$

1.2. Existence-unicité de la solution du problème variationnel. L'existence-unicité d'une solution pour le problème variationnel (1.1) n'est pas aussi simple que pour les problèmes scalaires. Elle repose sur une estimation difficile, fondamentale en élasticité : l'inégalité de Korn.

LEMME 6.1 (Inégalité de Korn). *Il existe une constante c ne dépendant que de Ω telle que*

$$\|\mathbf{v}\|_{1,\Omega} \leq c \left\{ \|\varepsilon(\mathbf{v})\|_{0,\Omega} + |\mathbf{v}|_{0,\Omega} \right\}, \quad \forall \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in H^1(\Omega;\mathbb{R}^2) \tag{1.6}$$

avec

$$\|\mathbf{v}\|_{1,\Omega} = \left\{ |\mathbf{v}|^2_{1,\Omega} + |\mathbf{v}|^2_{0,\Omega} \right\}^{1/2}$$

$$|\mathbf{v}|_{j,\Omega} = \left\{ \textstyle\sum_{i=1}^2 |v_i|^2_{j,\Omega} \right\}^{1/2}, \quad j = 1, 0,$$

$$\|\varepsilon(\mathbf{v})\|_{0,\Omega} = \left\{ \textstyle\sum_{i,j=1}^2 |\varepsilon_{ij}(\mathbf{v})|^2_{0,\Omega} \right\}^{1/2}, \quad \varepsilon_{ij}(\mathbf{v}) = \left(\partial_{x_j} v_i + \partial_{x_i} v_j \right)/2.$$

DÉMONSTRATION. Démonstration admise. Disons simplement que la difficulté vient du fait qu'il faut contrôler toutes les dérivées $\partial_j v_i$ à l'aide seulement des combinaisons suivantes

$$\partial_{x_j} v_i + \partial_{x_i} v_j.$$

□

L'autre ingrédient est de caractériser les déplacements linéarisés rigides, c'est à dire "ne donnant lieu à aucun travail pour les forces internes d'élasticité", définis par

$$\varepsilon(\mathbf{v}) = 0. \tag{1.7}$$

LEMME 6.2 (Caractérisation des déplacements linéarisés rigides). *Soit $\mathbf{v} \in H^1(\Omega;\mathbb{R}^2)$ vérifiant (1.7). Alors, il existe $\mathbf{a} \in \mathbb{R}^2$ et $b \in \mathbb{R}$ tels que*

$$\mathbf{v}(x) = \mathbf{a} + b \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix}. \tag{1.8}$$

DÉMONSTRATION. On écrit d'abord

$$\partial_{x_i}\partial_{x_j} v_\ell = \partial_{x_i} \left(\partial_{x_j} v_\ell + \partial_{x_\ell} v_j \right) - \partial_{x_\ell}\partial_{x_i} v_j,$$
$$\partial_{x_j}\partial_{x_i} v_\ell = \partial_{x_j} \left(\partial_{x_i} v_\ell + \partial_{x_\ell} v_i \right) - \partial_{x_\ell}\partial_{x_j} v_i.$$

En prenant la moyenne des deux relations suivantes, on obtient

$$\partial_{x_i}\partial_{x_j} v_\ell = \partial_{x_i}(\varepsilon_{j\ell}(\mathbf{v})) + \partial_{x_j}(\varepsilon_{i\ell}(\mathbf{v})) - \partial_{x_\ell}(\varepsilon_{ij}(\mathbf{v})).$$

D'où si $\mathbf{v}$ vérifie (1.7)

$$\partial_{x_i}\partial_{x_j} v_\ell \ (i,j,\ell = 1,2).$$

Il en résulte que

$$v_\ell(x) = v_\ell(0) + x_1\partial_{x_1}v_\ell(0) + x_2\partial_{x_2}v_\ell(0)\ (\ell = 1, 2).$$

La condition (1.7) donne en particulier

$$\begin{array}{ll} \partial_{x_1}v_1(0) = 0 & \partial_{x_2}v_1(0) = \partial_{x_2}v_1(0) - \partial_{x_1}v_2(0) \\ \partial_{x_1}v_2(0) = \partial_{x_1}v_2(0) - \partial_{x_2}v_1(0) & \partial_{x_2}v_2(0) = 0 \end{array}$$

Soit, en notant

$$\mathbf{a} := \begin{bmatrix} v_1(0) \\ v_2(0) \end{bmatrix}, \qquad b := \partial_{x_2}v_1(0) - \partial_{x_1}v_2(0),$$

$$\mathbf{v}(x) = \mathbf{a} + b \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix}.$$

□

LEMME 6.3. *Soit* $\mathbf{v}$ *un déplacement rigide et* p *et* q *deux points distincts du plan tels que*

$$\mathbf{v}(p) = \mathbf{v}(q) = 0. \tag{1.9}$$

Alors, $\mathbf{v}$ *est indentiquement nul.*

DÉMONSTRATION. Soit D une droite du plan passant par le point z et de cosinus directeurs le vecteur unitaire $\mathbf{t}$,

$$D := \left\{x \in \mathbb{R}^2\,;\ x = z + s\mathbf{t},\ \mathbf{t} \in \mathbb{R}\right\},$$

i.e. $x \in D$ si et seulement si $x - z$ est colinéaire à $\mathbf{t}$, ou encore

$$\det \begin{bmatrix} t_1 & x_1 - z_1 \\ t_2 & x_2 - z_2 \end{bmatrix} = \mathbf{t} \cdot \begin{bmatrix} x_2 - z_2 \\ -x_1 + z_1 \end{bmatrix} = 0.$$

Il résulte de cette propriété que si $\mathbf{v}$ est un déplacement rigide alors

$$\begin{aligned} \mathbf{v}(x) \cdot \mathbf{t} &= \mathbf{a} \cdot \mathbf{t} + b\mathbf{t} \cdot \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix} = \mathbf{a} \cdot \mathbf{t} + b\mathbf{t} \cdot \begin{bmatrix} z_2 \\ -z_1 \end{bmatrix} \\ &= \mathbf{v}(z) \cdot \mathbf{t} \end{aligned}$$

pour tout $x \in D$.

Soit maintenant r un point du plan tel que p, q et r forment un vrai triangle. Notons par

$$\mathbf{t}^{(pq)} := (q - p)/\left\|q - p\right\|, \quad \mathbf{t}^{(qr)} := (r - q)/\left\|r - q\right\|, \quad \mathbf{t}^{(rp)} := (p - r)/\left\|p - r\right\|$$

les vecteurs directeurs des droites supportant les arêtes du triangle de sommets p, q, r. Si $\mathbf{v}$ est le déplacement rigide de l'énoncé, il vient

$$\mathbf{v}(r) \cdot \mathbf{t}^{(qr)} = 0 \text{ et } \mathbf{v}(r) \cdot \mathbf{t}^{(rp)} = 0.$$

Comme $\mathbf{t}^{(qr)}$ et $\mathbf{t}^{(rp)}$ sont non colinéaires, il forment une base des vecteurs du plan. Il s'ensuit que

$$\mathbf{v}(r) = \alpha\mathbf{t}^{(qr)} + \beta\mathbf{t}^{(rp)}$$

et donc $\mathbf{v}(r) \cdot \mathbf{v}(r) = \left\|\mathbf{v}(r)\right\|^2 = 0$. On en déduit que $\mathbf{v}(r) = 0$. On termine la démonstration du lemme en observant que chaque composante de $\mathbf{v}$ est un polynôme de degré ≤ 1 qui s'annule aux trois sommets d'un triangle et s'annule donc identiquement sur $\mathbb{R}^2$. □

On a alors le théorème suivant.

THÉORÈME 6.1. *Si la longueur de Γ_D est non nulle, alors le problème (1.1) possède une solution et une seule.*

DÉMONSTRATION. Clairement, il suffit d'établir la coercivité. On a tout d'abord

$$a(\mathbf{v},\mathbf{v}) \geq 2\mu_0 \left\|\varepsilon(\mathbf{v})\right\|_{0,\Omega}.$$

L'inégalité de Korn assure l'existence d'une constante C_0 telle que

$$\left\|\mathbf{v}\right\|_{1,\Omega} \leq C_0 \left\{ \left\|\varepsilon(\mathbf{v})\right\|_{0,\Omega} + \left\|\mathbf{v}\right\|_{0,\Omega} \right\}.$$

Cette inégalité s'exprime aussi de la façon suivante. Notons par T l'opérateur de $\mathbf{V}$ dans $\left\{L^2(\Omega)\right\}^4$ qui à $\mathbf{v}$ associe $T\mathbf{v} := (\varepsilon_{11}(\mathbf{v}), \varepsilon_{12}(\mathbf{v}), \varepsilon_{21}(\mathbf{v}), \varepsilon_{22}(\mathbf{v}))$ et par S l'injection de $\mathbf{V}$ dans $L^2(\Omega;\mathbb{R}^2)$. Le théorème de Rellich montre que S est compact et, puisque la longueur de Γ_D est non nulle, le lemme précédent que $\ker T = \{0\}$. Le lemme de Peetre-Tartar assure donc qu'il existe $C > 0$ telle que

$$\left\|\varepsilon(\mathbf{v})\right\|_{0,\Omega} \geq (1/C) \left\|\mathbf{v}\right\|_{1,\Omega} \quad \forall \mathbf{v} \in \mathbf{V}.$$

Ce qui avec l'inégalité sur $a(\mathbf{v},\mathbf{v})$ ci-dessus assure la coercivité. □

1.3. Discrétisation par éléments finis. On utilise l'approximation par éléments finis $\mathbb{P}_1$-continus pour approcher l'espace $H^1(\Omega;\mathbb{R}^2)$ introduite au chapitre II. Le sous-espace de dimension finie devient maintenant

$$\mathbf{X}^h := \left\{ \mathbf{v}^h \in \mathcal{C}^0(\overline{\Omega};\mathbb{R}^2)\,;\ v_i^h|_T \in \mathbb{P}_1,\ i=1,2,\ \forall T \in \mathcal{T}^h \right\}.$$

Tout $\mathbf{v}^h \in \mathbf{X}^h$ est ainsi décrit par les valeurs nodales vectorielles

$$\mathbf{v}_j^h := \mathbf{v}^h(a_j) := \begin{bmatrix} v_{1,j}^h \\ v_{2,j}^h \end{bmatrix}, \quad j = 1, \ldots, N_S.$$

REMARQUE 6.1. *Par rapport au cas scalaire, on a deux valeurs nodales par noeud. Pour un système plus général avec M inconnues et M équations, on aurait eu M valeurs nodales par sommet. Par exemple pour un problème en thermoélasticité en contraintes planes, il y a en plus du vecteur déplacement la température qui serait aussi à déterminer. On aurait dans ce cas 3 valeurs nodales par sommet : 2 pour les deux compsantes du déplacement et une pour la température.*

La matrice raideur élémentaire est définie par la relation

$$\begin{bmatrix} \mathbf{v}_1^{\mathrm{T}} & \mathbf{v}_2^{\mathrm{T}} & \mathbf{v}_3^{\mathrm{T}} \end{bmatrix} K^{[e]} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{bmatrix} =$$

$$\int_{T^{[e]}} \left\{ 2\mu^{[e]} \textstyle\sum_{i,j=1}^2 \varepsilon_{ij}(\mathbf{u})\varepsilon_{ij}(\mathbf{v}) + \lambda^{[e]} \sum_{i=1}^2 \varepsilon_{ii}(\mathbf{u}) \sum_{i=1}^2 \varepsilon_{ii}(\mathbf{u}) \right\} dT^{[e]}$$

avec

$$\mathbf{u} = \begin{bmatrix} u_1 \in \mathbb{P}_1 \\ u_2 \in \mathbb{P}_1 \end{bmatrix}, \quad \mathbf{u}_j = \begin{bmatrix} u_{1,j} = u_1(a_j^{[e]}) \\ u_{2,j} = u_2(a_j^{[e]}) \end{bmatrix}, \quad a_j^{[e]} \text{ sommet de } T^{[e]}.$$

Si $j = \mathrm{IDL}(\ell)$ est la numérotation des sommets permettant de repérer les sommets où il y a encastrement (sommets $a_\ell \in \overline{\Gamma_D}$), on peut numéroter les degrés de liberté inconnus par

$$x_{2j-1} := u_{1,\ell}, \quad x_{2j} := u_{2,\ell} \quad \text{si } j > 0,$$

i.e. par un numéro impair les premières composantes et un numéro pair les secondes.

On peut alors vérifier aisément que si N est le nombre de degrés de liberté non bloqués pour le problème scalaire et ℓ_B la largeur de bande, on a ici (avec $M = 2$ le nombre d'inconnues du système d'EDP)

– **Ordre du système à résoudre :** MN
– **Largeur de bande :** $M\ell_B$.

2. Problèmes du quatrième ordre

Nous introduisons ci-dessous les problèmes les plus usuels en ingénierie où intervient une équation aux dérivées partielles d'ordre 4 : le système de Stokes qui décrit le modèle le plus simple d'un fluide visqueux incompressible et le modèle de Kirchoff-Love d'une plaque en flexion.

2.1. Formulation du système de Stokes en courant - tourbillon. Nous considérons le modèle simple suivant décrivant l'écoulement stationnaire d'un fluide incompressible confiné dans un domaine Ω de $\mathbb{R}^2$ de bord Γ

$$\left\{ \begin{array}{l} -\mu\Delta\mathbf{u} + \nabla p = \mathbf{f} \text{ dans } \Omega \\ \nabla \cdot \mathbf{u} = 0 \text{ dans } \Omega \\ \mathbf{u} = 0 \text{ sur } \Gamma \end{array} \right. \tag{2.1}$$

Dans ce système, les données sont la viscosité du fluide $\mu > 0$ et les forces volumiques $\mathbf{f}$ produisant l'écoulement ; les inconnues sont la vitesse de l'écoulement $\mathbf{u}$, qui est un vecteur à deux composantes qu'on note commodément sous forme d'un vecteur colonne

$$\mathbf{u}(x,y) = \left[\begin{array}{c} u_1(x,y) \\ u_2(x,y) \end{array} \right]$$

et la pression p qui est une inconnue scalaire. C'est pourquoi, le problème (2.1) est appelé formulation en vitesse-pression du système de Stokes. Rappelons que le laplacien vectoriel est défini par

$$\Delta\mathbf{u}(x,y) = \left[\begin{array}{c} \Delta u_1(x,y) \\ \Delta u_2(x,y) \end{array} \right]$$

et que la divergence $\nabla \cdot \mathbf{u}$ est donnée par

$$\nabla \cdot \mathbf{u}(x,y) = \partial_x u_1(x,y) + \partial_y u_2(x,y).$$

La résolution du système précédent à l'aide d'une méthode d'éléments finis n'est pas triviale car on ne peut pas discrétiser la vitesse $\mathbf{u}$ et la pression p de façon indépendante. Ce type de discrétisation par éléments finis est à un second niveau de difficulté que celui considéré dans ce cours. On utilise ce qu'on appelle une méthode d'éléments finis mixtes.

Nous allons utiliser ici une formulation qui évite la difficulté précédente. Introduisons pour cela le tourbillon ω qui est donné par le rotationnel scalaire de la vitesse

$$\omega = \operatorname{rot} \mathbf{u} := \partial_x u_2 - \partial_y u_1 = \nabla \cdot \left[\begin{array}{c} u_2 \\ -u_1 \end{array} \right].$$

Comme $\operatorname{rot} \nabla p = \partial_x \partial_y p - \partial_y \partial_x p = 0$, en prenant le rotationnel scalaire des deux membres de la première équation dans (2.1), on obtient l'équation suivante sur le tourbillon

$$-\mu \operatorname{rot} \Delta\mathbf{u} = -\mu\boldsymbol{\Delta} \operatorname{rot} \mathbf{u} = -\mu\Delta\omega = \operatorname{rot} \mathbf{f}. \tag{2.2}$$

On arrive ainsi à une équation scalaire posée avec le laplacien. La difficulté provient du fait ici qu'on ne dispose pas de condition aux limites sur le tourbillon ω. C'est pourquoi, on va plutôt utiliser une formulation basée sur la fonction courant.

En supposant que le domaine Ω est simplement connexe (i.e., sans trou), la condition d'incompressibilité $\nabla \cdot \mathbf{u} = 0$ assure qu'il une fonction dite fonction courant w telle que

$$\mathbf{u} = \nabla \times w := \begin{bmatrix} \partial_y w \\ -\partial_x w \end{bmatrix}. \tag{2.3}$$

Le rotationnel vecteur $\nabla \times w$ peut s'interpéter de diverses façons : c'est ou bien le rotationnel usuel dans $\mathbb{R}^3$ du champ de vecteurs colinéaires à l'axe des z

$$\mathbf{W}(x,y) = \begin{bmatrix} 0 \\ 0 \\ w(x,y) \end{bmatrix}$$

ou bien encore le gradient ∇w de w tourné de $\pi/2$ dans le sens contraire au sens trigonométrique. Observons que la fonction courant est définie à une constante prés. Pour fixer la constante, on impose à w d'être à moyenne nulle sur Γ

$$\int_\Gamma w \, ds = 0. \tag{2.4}$$

La relation entre la fonction courant w et le tourbillon est simplement donnée par

$$\omega = \operatorname{rot} \begin{bmatrix} \partial_y w \\ -\partial_x w \end{bmatrix} = -\partial^2_{xx} w - \partial^2_{yy} w = -\Delta w.$$

L'équation (2.2) se réécrit alors avec le bilaplacien $\mu \Delta^2 w = \operatorname{rot} \mathbf{f}$.

Il reste maintenant à trouver les conditions aux limites vérifiées par w. La condition de non glissement $\mathbf{u} = 0$ sur Γ se décompose de façon équivalent en les deux conditions suivantes sur la composante normale $\mathbf{u} \cdot \mathbf{n} = 0$ et la composante tangentielle $\mathbf{u} \cdot \boldsymbol{\tau} = 0$ où $\boldsymbol{\tau}$ est le vecteur tangent unitaire à Γ obtenu par rotation de $\pi/2$ dans le sens trigonométrique de la normale $\mathbf{n}$ à Γ sortante de Ω. Notons par s l'abcisse curviligne sur Γ croissante dans le sens de τ. On a

$$\mathbf{u} \cdot \mathbf{n} = \nabla \times w \cdot \mathbf{n} = \nabla w \cdot \boldsymbol{\tau} = \partial_s (w|_\Gamma) = 0.$$

Cette relation entraine que $w|_\Gamma$ est une constante. Comme w est de moyenne nulle sur Γ, on obtient donc la première condition aux limites vérifiée par w :

$$w = 0 \text{ sur } \Gamma.$$

La condition $\mathbf{u} \cdot \tau = 0$ donne simplement

$$\partial_{\mathbf{n}} w = 0 \text{ sur } \Gamma.$$

On obtient alors le problème aux limites du $4^{\text{ème}}$ ordre permettant de déterminer la fonction courant

$$\begin{cases} \Delta^2 w = \mu^{-1} \operatorname{rot} \mathbf{f} \text{ dans } \Omega \\ w = 0, \ \partial_{\mathbf{n}} w = 0 \text{ sur } \Gamma \end{cases} \tag{2.5}$$

Nous montrerons dans la suite que ce problème est un cas particulier d'un problème de plaque en flexion. On pourra en déduire ainsi non seulement un résultat d'existence-unicité pour sa solution mais aussi une façon de le résoudre avec une méthode d'éléments finis.

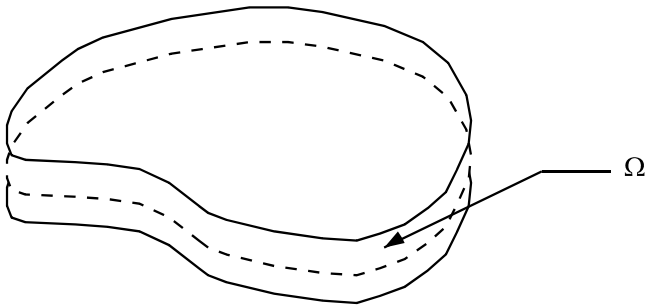


FIG. 1. Représentation de la plaque

2.2. Les équations variationnelles d'une plaque en flexion. On note par Ω un domaine borné de $\mathbb{R}^2$ représentant la surface moyenne de la plaque (voir Fig. 1).

On note par Γ la frontière de Ω.

Le cadre fonctionnel pour poser le problème de l'équilibre de la plaque en flexion est l'espace de Sobolev suivant

$$H^2(\Omega) := \left\{ v \in L^2(\Omega)\,;\ \partial^\alpha v \in L^2(\Omega),\ \forall \alpha,\ |\alpha| \leq 2 \right\}.$$

On peut définir $H^2(\Omega)$ de façon équivalente par

$$H^2(\Omega) := \left\{ v \in H^1(\Omega)\,;\ \partial^2_{xx} v \in L^2(\Omega),\ \partial^2_{yy} v \in L^2(\Omega),\ \partial^2_{xy} v \in L^2(\Omega) \right\}.$$

On montre exactement comme pour $H^1(\Omega)$ que $H^2(\Omega)$ est un espace de Hilbert pour la norme

$$\|v\|_{2,\Omega} := \left\{ \|v\|^2_{1,\Omega} + |v|_{2,\Omega} \right\}^{1/2} \tag{2.6}$$

$$|v|_{2,\Omega} := \left\{ \left|\partial^2_{xx} v\right|^2_{0,\Omega} + \left|\partial^2_{yy} v\right|^2_{0,\Omega} + \left|\partial^2_{xy} v\right|^2_{0,\Omega} \right\}^{1/2}$$

et que

$$H^2(\Omega) \subset \mathcal{C}^0(\overline{\Omega}).$$

Pour décrire les différents type de conditions aux limites, on se donne partition sans recouvrement Γ_D, Γ_N et Γ_L de Γ. On a alors les différentes situations cinématiques sur le déplacement inconnu u :

– Encastrement sur Γ_D

$$u|_{\Gamma_D} = 0 \text{ et } \partial_{\mathbf{n}} u|_{\Gamma_D} = 0 \tag{2.7}$$

– Appui simple sur Γ_L

$$u|_{\Gamma_L} = 0 \tag{2.8}$$

– Bord libre sur Γ_N : pris en compte directement par la formulation variationnelle.

Si la plaque est en équilibre sous l'action de la densité de forces normales $f \in L^2(\Omega)$, le déplacement u (flèche) est solution du problème variationnel :

$$\left\{ \begin{array}{l} u \in V, \quad \forall v \in V, \\ a(u,v) = Lv, \end{array} \right. \tag{2.9}$$

avec

$$a(u,v) := \int_\Omega D\left\{\partial^2_{xx}u\partial^2_{xx}v + \nu(\partial^2_{xx}u\partial^2_{yy}v + \partial^2_{yy}u\partial^2_{xx}v) + \partial^2_{yy}u\partial^2_{yy}v + 2(1-\nu)\partial^2_{xy}u\partial^2_{xy}v\right\} d\Omega$$

où $D = Eh^3/12(1-\nu^2)$ est le module de rigidité à la flexion, h étant la hauteur de la plaque, et

$$V := \left\{v \in H^2(\Omega)\,;\ v|_{\Gamma_D\cup\Gamma_L} = 0,\ \partial_{\mathbf{n}}v|_{\Gamma_D} = 0\right\}$$

et L une forme linéaire continue sur V.

2.3. Existence-unicité de la solution. L'existence-unicité de la solution du problème (2.9) est assurée par le théorème suivant.

THÉORÈME 6.2. *Si le coefficient de Poisson v vérifie*

$$0 \le \nu(x) \le \nu_0 < 1, \quad \textit{p.p. tout } x \in \Omega, \tag{2.10}$$

le module de rigidité à la flexion D vérifie

$$D(x) \ge D_0 > 0 \quad \textit{p.p. tout } x \in \Omega, \tag{2.11}$$

et le seul p dans $\mathbb{P}_1$ qui est dans V est le polynôme 0, i.e.

$$V \cap \mathbb{P}_1 = \{0\} \tag{2.12}$$

alors, le problème (2.9) possède une solution et une seule.

DÉMONSTRATION. La seule condition, non triviale à vérifier pour appliquer le théorème de Lax-Milgram, est la coercivité. On observe tout d'abord que $a(v,v)$ peut être écrit

$$\begin{aligned}
a(v,v) &= \int_\Omega D\left\{\left|\partial^2_{xx}v\right|^2 + 2\nu\partial^2_{xx}v\partial^2_{yy}v + \left|\partial^2_{yy}v\right|^2 + 2(1-\nu)\left|\partial^2_{xy}v\right|^2\right\} d\Omega \\
&= \int_\Omega D\left\{\nu\left|\partial^2_{xx}v\right|^2 + 2\nu\partial^2_{xx}v\partial^2_{yy}v + \nu\left|\partial^2_{yy}v\right|^2\right\} d\Omega \\
&\qquad + \int_\Omega D(1-\nu)\left(\left|\partial^2_{xx}v\right|^2 + \left|\partial^2_{yy}v\right|^2 + 2\left|\partial^2_{xy}v\right|^2\right) d\Omega \\
&= \int_\Omega D\nu\left|\Delta v\right|^2 d\Omega + \int_\Omega D(1-\nu)\left(\left|\partial^2_{xx}v\right|^2 + \left|\partial^2_{yy}v\right|^2 + 2\left|\partial^2_{xy}v\right|^2\right) d\Omega.
\end{aligned}$$

Il vient donc

$$a(v,v) \ge D_0(1-\nu_0)\left|v\right|^2_{2,\Omega}.$$

Comme l'injection de $H^2(\Omega)$ dans $H^1(\Omega)$ est compacte, le lemme de Peetre-Tartar permet à partir de (2.6) de se ramener à s'assurer que si $v \in V$ est tel que $|v|_{2,\Omega} = 0$, alors $v = 0$. Mais comme $|v|_{2,\Omega} = 0$ équivaut à $v \in \mathbb{P}_1$, le résultat suit de l'hypothèse (2.12). □

2.4. Eléments sur l'approximation numérique. La résolution par une méthode d'éléments finis repose sur la construction d'un sous-espace de $H^2(\Omega)$. Ceci impose de raccorder de les fonctions et leurs dérivées premières aux interfaces. Sur un maillage en triangles, cela impose d'utiliser l'élément d'Argyris (voir document sur `getfem` annexé) dont l'espace des fonctions de forme est $\mathbb{P}_5$ (et donc de dimension 21) et de **raccorder aussi** les dérivées secondes aux sommets du maillage. Cela a les deux inconvénients suivants par rapport par exemple à la résolution du problème de la poutre en flexion :

- on n'a plus d'interprétation simple en langage mécanique du torseur des forces concentrées liées aux matrices élémentaires,

– on impose un raccord trop fort sur les dérivées secondes aux sommets du maillage qui peut donner une solution fausse si le sommet est sur une discontinuité des données.

C'est pour cette raison qu'on peut être amené à utiliser des éléments dits composites où les fonctions de forme sont polynomiales sur une partition du triangle en trois triangles de sommets le centre de gravité et les trois sommets.

2.5. Liens des problèmes de Stokes et de la plaque en flexion. Supposons que la plaque est encastrée sur toute sa frontière Γ. En notant $w := u$, c'est exactement sous cette forme que sont écrites les conditions aux limites dans (2.5). Supposons en outre que la plaque est homogène, i.e. D est constant (on peut ainsi supposer $D = 1$) et ν est constant, avec donc $0 < \nu < 1$. Prenons Lv sous la forme

$$Lv = \int_\Omega \mu^{-1} \operatorname{rot} \mathbf{f}\, v \, d\Omega.$$

Montrons que la solution du problème (1.3), , est celle du système de Stokes en formulation fonction courant (2.5). En effet la dérivation au sens des distribution donne pour $\varphi \in \mathcal{D}(\Omega)$

$$\begin{aligned} a(w,\varphi) = & \\ & \int_\Omega \left\{ \partial^2_{xx} w \partial^2_{xx} \varphi + \nu(\partial^2_{xx} w \partial^2_{yy} v + \partial^2_{yy} w \partial^2_{xx} \varphi) + \partial^2_{yy} w \partial^2_{yy} \varphi + 2(1-\nu)\partial^2_{xy} w \partial^2_{xy} \varphi \right\} d\Omega \\ & = \left\langle (\partial_x)^4 w + 2\nu \partial^2_{xx} \partial^2_{yy} w + (\partial_y)^4 w + 2(1-\nu)\partial^2_{xx}\partial^2_{yy} w, \varphi \right\rangle_{\mathcal{D}'(\Omega),\mathcal{D}(\Omega)} \\ & = \left\langle \Delta^2 w, \varphi \right\rangle_{\mathcal{D}'(\Omega),\mathcal{D}(\Omega)} \end{aligned}$$

On obtient bien que w est la solution du problème (2.5).

REMARQUE 6.2. *On aurait pu écrire la formulation en fonction courant à l'aide de la forme bilinéaire*

$$a^{(2)}(w,v) := \int_\Omega \Delta w \, \Delta v \, d\Omega$$

qui est a priori plus simple. On s'aperçoit dans beaucoup d'expériences numériques que l'utilisation de la forme blinéaire $a(\cdot,\cdot)$ avec un coefficient de Poisson $0 < \nu < 1$ artificiel donne une meilleure précision.

CHAPITRE 7

Problèmes dépendant du temps

Nous nous limitons dans ce chapitre à introduire les principes à la base des procédés de résolution numérique des problèmes modélisant des phénomènes évoluant au cours du temps. Ce type de problème est dit d'évolution.

1. Problèmes modèles d'évolution

Nous donnons ci-dessous deux exemples typiques de problèmes dépendant du temps. La plupart des situations rencontrées en ingénierie ont les caractéristiques de l'un ou l'autre de ces deux exemples.

1.1. Les problèmes modèles. On reprend les données et notations du chapitre I de (2.1) à (2.6). On suppose pour simplifier que $g = 0$ et $h = 0$ et on fait l'hypothèse fondamentale que la matrice $A(x)$ est symétrique (et donc définie positive) pour presque tout $x \in \Omega$. On suppose en outre que la donnée f dépend non seulement de x mais aussi de $t \in [0,T]$ (i.e. on se donne $(x,t) \to f(x,t)$ définie dans $\Omega \times [0,T]$). On se donne en plus une fonction ϱ dans $L^\infty(\Omega)$ vérifiant

$$0 < \varrho_m \leq \varrho(x) \leq \varrho_M, \quad \text{p.p. en } x \in \Omega. \tag{1.1}$$

On pose de cette façon les deux exemples type de problèmes dépendant du temps t

(1) **Problème parabolique**

$$\begin{cases} (x,t) \to u(x,t), \\ \varrho(x)\partial_t u(x,t) - \sum_{i=1}^N \sum_{j=1}^N \partial_{x_i} a_{ij}(x) \partial_{x_j} u(x,t) + a_0(x) u(x,t) = f(x,t), \\ \qquad\qquad (x,t) \in \Omega \times]0,T[\,, \\ u(x,t) = 0, \quad (x,t) \in \Gamma_D \times]0,T[\,, \\ \sum_{i=1}^N \sum_{j=1}^N n_i(x) a_{ij}(x) \partial_{x_j} u(x,t) = 0, \quad (x,t) \in \Gamma_N \times]0,T[\,, \\ u(x,0) = u^{(0)}(x), \ x \in \Omega \end{cases} \tag{1.2}$$

où la fonction $u^{(0)}$ constitue la donnée initiale du problème,

(2) **Problème hyperbolique**

$$\begin{cases} (x,t) \to u(x,t), \\ \varrho(x)\partial_t^2 u(x,t) - \sum_{i=1}^N \sum_{j=1}^N \partial_{x_i} a_{ij}(x) \partial_{x_j} u(x,t) + a_0(x) u(x,t) = f(x,t), \\ \qquad\qquad (x,t) \in \Omega \times]0,T[\,, \\ u(x,t) = 0, \quad (x,t) \in \Gamma_D \times]0,T[\,, \\ \sum_{i=1}^N \sum_{j=1}^N n_i(x) a_{ij}(x) \partial_{x_j} u(x,t) = 0, \quad (x,t) \in \Gamma_N \times]0,T[\,, \\ u(x,0) = u^{(0)}(x), \ \partial_t u(x,0) = u^{(1)}(x), \quad x \in \Omega, \end{cases} \tag{1.3}$$

où maintenant le problème nécessite deux données initiales $u^{(0)}$ et $u^{(1)}$.

1.2. Formulation variationnelle. En multipliant l'EDP par une fonction test v, indépendante de t, dans

$$V := \left\{ v \in H^1(\Omega)\,;\; v|_{\Gamma_D} = 0 \right\},$$

on met les problèmes (1.2) et (1.3) sous la forme

$$\text{(1.4)} \qquad \begin{cases} t \to u(\cdot,t) \text{ définie dans } [0,T] \text{ à valeurs dans } V, \\ \partial_t^m b(u(\cdot,t),v) + a(u(\cdot,t),v) = \displaystyle\int_\Omega f(x,t)v(t)dx, \quad \forall v \in V, \\ \partial_t^\ell u(\cdot,t) = u^{(\ell)}, \quad 0 \le \ell \le m-1, \end{cases}$$

avec $m = 1$ pour le problème parabolique et $m = 2$ pour le problème hyperbolique. La forme bilinéaire $a(\cdot,\cdot)$ a été définie en (2.9) au chapitre I et la forme $b(\cdot,\cdot)$ est donnée par

$$\text{(1.5)} \qquad b(w,v) := \int_\Omega \varrho w v \, dx.$$

2. Semi-discrétisation en espace

2.1. Méthode d'éléments finis. On utilise un maillage $\mathcal{T}^h$ de Ω :
– en segments si Ω est un segment,
– en triangles si Ω est un domaine polygonal du plan,
– en tétraèdres si Ω est un domine polyèdrique de l'espace
compatible avec les discontinuités des données et Γ_D. On utilise alors la méthode d'éléments finis $\mathbb{P}_1$-continus introduite au chapitre III pour approcher respectivement $H^1(\Omega)$ et V par X^h et V^h.

Les fonctions test ne dépendent pas du temps t. Elles sont approchées, comme au chapitre III, par des fonctions v^h complètement déterminées à l'aide de leurs valeurs nodales : v_i $(i = 1, \ldots, N_S)$ où N_S est le nombre de sommets du maillage —nous supprimons l'exposant h pour alléger la notation—. On note par $\left[v^\#\right]$ le vecteur colonne formé par les valeurs nodales de v^h

$$\text{(2.1)} \qquad \left[v^\#\right] := \begin{bmatrix} v_1 \\ \vdots \\ v_N \\ v_{N+1} \\ \vdots \\ v_{N_S} \end{bmatrix}$$

Si on suppose que la numérotation des sommets est telle que $a_i \in \overline{\Gamma_D}$ est numéroté par i variant de $N+1$ à N_S, le vecteur des valeurs nodales non fixées à zéros de $v^h \in V^h$ est donné par

$$\text{(2.2)} \qquad [v] := \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$$

L'approximation de la fonction inconnue u est effectué de la façon suivante : en tout point $t \in [0,T]$, $u(\cdot,t)$ est approchée par une fonction $u^h(\cdot,t) \in V^h$. En particulier, la fonction u^h est complètement déterminée par N_S fonctions de $t \in [0,T]$: $t \to u_j^h(t)$ donnant les **valeurs nodales** de $u^h(\cdot,t)$. On note de même la fonction de t à valeurs dans

$\mathbb{R}^{N_S}$

$$\left[u^{\#}\right](t) := \begin{bmatrix} u_1(t) \\ \vdots \\ u_N(t) \\ u_{N+1}(t) \\ \vdots \\ u_{N_S}(t) \end{bmatrix} \tag{2.3}$$

De même, l'appartenance de $u^h(\cdot,t)$ à V^h est caractérisée par les conditions

$$u_{N+1}(t) = \cdots = u_{N_S}(t) = 0 \tag{2.4}$$

de sorte que les seules fonctions inconnues à déterminer sont les composantes du vecteur

$$[u](t) := \begin{bmatrix} u_1(t) \\ \vdots \\ u_N(t) \end{bmatrix} \tag{2.5}$$

En supposant que les données initiales $u^{(\ell)}$ $(\ell \leq m-1)$ sont suffisamment régulières (en fait dans $\mathcal{C}^0(\overline{\Omega})$), on peut les approcher à partir de leurs valeurs aux noeuds $u^{(\ell)}(a_i)$ par un vecteur $\left[u^{\#}\right]^{(\ell)}$ qu'on suppose vérifier $u_i^{(\ell)} = 0$ $(i = N+1, \ldots, N_S)$: condition de compatibilité sur les données initiales. Ces valeurs nodales déterminent des fonctions $\left(u^{(\ell)}\right)^h \in V^h$.

Pour approcher u, on est ainsi conduit de façon naturelle à résoudre le problème variationnel

$$\begin{cases} t \longrightarrow u^h(\cdot,t) \text{ définie dans } [0,T] \text{ à valeurs dans } V^h, \\ \partial_t^m b(u^h(\cdot,t), v^h) + a(u^h(\cdot,t), v) = \displaystyle\int_\Omega f(x,t) v^h(x) dx, \quad \forall v^h \in V^h, \\ \partial_t^\ell u^h(\cdot,t) = (u^{(\ell)})^h, \quad 0 \leq \ell \leq m-1, \end{cases} \tag{2.6}$$

dont les inconnues sont N fonctions inconnues $[0,T] \ni t \longrightarrow u_j(t)$ $(j = 1, \ldots, N)$.

2.2. Réduction à un système différentiel. En plus de la matrice raideur totale définie par la relation

$$\left[v^{\#}\right] K^{\#} \left[w^{\#}\right] = a(w^h, v^h), \quad w^h, v^h \in X^h, \tag{2.7}$$

on forme la matrice masse totale $M^{\#}$ définie par

$$\left[v^{\#}\right] M^{\#} \left[w^{\#}\right] = b(w^h, v^h), \quad w^h, v^h \in X^h, \tag{2.8}$$

où $\left[w^{\#}\right]$ est comme ci-dessus pour $\left[v^{\#}\right]$ le vecteur colonne formé des valeurs nodales de w^h.

On définit aussi pour tout $t \in [0,T]$ le vecteur chargement total dont les composantes dépendent du temps t sont données par la relation

$$\left[v^{\#}\right] L^{\#}(t) = \int_\Omega f(x,t) v^h(x)\, d\Omega, \quad v^h \in X^h. \tag{2.9}$$

Le problème (2.6) s'écrit donc à l'aide de ces notations

$$\begin{cases} \left[v^{\#}\right]^{\mathrm{T}} M^{\#} \left[\partial_t^\ell u^{\#}\right](t) + \left[v^{\#}\right]^{\mathrm{T}} K \left[u^{\#}\right](t) = \left[v^{\#}\right]^{\mathrm{T}} L^{\#}(t), \quad t \in [0,T], \\ \left[\partial_t^\ell u^{\#}\right](0) = \left[u^{\#}\right]^{(\ell)}, \ \ell \leq m-1, \end{cases} \tag{2.10}$$

où $\left[\partial_t^\ell u^\#\right](t)$ est le vecteur de composantes $\partial_t^\ell u_j(t)$ où $u_j(t)$ vérifie les conditions (2.4) et v_i les conditions analogues

$$v_{N+1} = \cdots = v_{N_S} = 0. \tag{2.11}$$

Exactement comme pour le cas des problèmes ne dépendant pas de t, en supprimant les lignes et les colonnes des matrices raideur, masse et chargement correspondant à des degrés de liberté bloqués à zéro, le système variationnel matriciel précédent se réécrit

$$\begin{cases} M\left[\partial_t^\ell u\right](t) + K\left[u\right](t) = L(t), \quad t \in [0,T], \\ \left[\partial_t^\ell u\right](0) = \left[u^{(\ell)}\right], \quad \ell \leq m-1. \end{cases} \tag{2.12}$$

Dans ce système, K, M et $L(t)$ désignent respectivement les matrices raideur, masse et chargement finales.

2.3. Propriétés des matrices du système différentiel. Nous avons supposé que la matrice A des coefficients de l'EDP était symétrique. Celà entraine que la forme bilinéaire $a(\cdot,\cdot)$ est symétrique et par là-même que la matrice raideur K est symétrique

$$K^{\mathrm{T}} = K. \tag{2.13}$$

La matrice finale M est aussi symétrique car la forme $b(\cdot,\cdot)$ vérifie

$$b(w,v) = \int_\Omega \varrho w v \, d\Omega = \int_\Omega \varrho v w \, d\Omega = b(v,w). \tag{2.14}$$

On a alors le résultat fondamental pour la résolution suivant.

PROPOSITION 7.1. *Quelque soit sa détermination, exacte ou condensée, la matrice masse M est définie positive.*

DÉMONSTRATION. On a d'abord, pour $\left[v^\#\right]$ satisfaisant les conditions (2.11),

$$\left[v\right]^{\mathrm{T}} M \left[v\right] = \left[v^\#\right]^{\mathrm{T}} M^\# \left[v^\#\right].$$

On a donc

$$\left[v\right]^{\mathrm{T}} M \left[v\right] = \begin{cases} \displaystyle\int_\Omega \varrho v^2 \, d\Omega \geq 0, & \text{si } M \text{ est exact,} \\ \displaystyle\sum_{e=1}^{N_E} \frac{|T^{[e]}|}{N_e} \sum_{i=1}^{N_e} \varrho_i^{[e]} \left(v_i^{[e]}\right)^2 \geq 0, & \text{si } M \text{ est condensée,} \end{cases}$$

où $\varrho_i^{[e]}$ et $v_i^{[e]}$ les valeurs nodales respectives de $\varrho^{[e]} = \varrho|_{T^{[e]}}$ et $v^h|_{T^{[e]}}$ aux sommets de l'élément $T^{[e]}$ et N_e est le nombre de sommets de $T^{[e]}$. La condition $\varrho \geq \varrho_0 > 0$ donne alors directement que $v^{\mathrm{T}} M v > 0$ dès que $v^h \neq 0$. □

REMARQUE 7.1.

– *La matrice masse condensée est une matrice diagonale*

$$M = \begin{bmatrix} M_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & M_N \end{bmatrix}$$

avec $M_i > 0$ $(i = 1, \ldots, N)$.

– *Si Γ_D est vide, c'est à dire si on a une condition de Neumann sur toute la frontière Γ, la forme $a(\cdot,\cdot)$ est généralement positive mais non définie. Dans ce cas la matrice K est symétrique et positive mais non définie. C'est ce qu'on supposera dans toute la suite.*

3. Schémas en temps

On est ainsi ramené à résoudre le système différentiel (2.12) qu'on peut effectuer à l'aide d'une méthode numérique de détermination de la solution aux points d'une grille de l'intervalle $[0,T]$

$$t_0 < t_1 < \cdots < t_m < t_{m+1} < t_M = T, \tag{3.1}$$

ou encore aux points

$$t_0 = 0,\ t_{m+1} = t_m + \Delta t_m\,(m = 0, \ldots M-1). \tag{3.2}$$

Le système (2.12) est un système différentiel dont la constante de Lipschitz est en $1/h^2$. Il est donc raide. On ne peut pas le résoudre par n'importe quel schéma. Nous décrivons ci-dessous rapidement une famille de schémas pour les problèmes paraboliques, les θ-schémas, et un schéma pour les équations hyperboliques, le schéma de Neumark.

3.1. θ-schéma pour les équations paraboliques. Ce sont des schémas **à un pas** d'intégration du sytème différentiel (2.12). Rappelons que la construction de ce type de schéma revient à utiliser une formule d'intégration numérique. En effet, si $[u]$ est solution du système différentiel

$$\left\{\begin{array}{l} M\left[\dot{u}\right](t) + K\left[u\right](t) = L(t), \quad t \in [0,T]\,, \\ \left[u\right](0) = \left[u^{(0)}\right], \end{array}\right. \tag{3.3}$$

où $\left[\dot{u}\right](t)$ désigne le vecteur dont les composantes sont les dérivées $\dot{u}_j(t)$ des composantes de $[u]$, on a

$$\int_{t_m}^{t_m+\Delta t_m} M\left[\dot{u}\right](t)dt = \int_{t_m}^{t_{m+1}} \left(L(t) - K\left[u\right](t)\right) dt$$

ou encore

$$M\left[u\right](t_{m+1}) - M\left[u\right](t_m) = \int_{t_m}^{t_{m+1}} \left(L(t) - K\left[u\right](t)\right) dt.$$

On approche alors l'intégrale du second membre par la formule d'intégration numérique

$$\int_{t_m}^{t_{m+1}} \varphi(t)\,dt \approx \Delta T\left((1-\theta)\varphi(t_m) + \theta\varphi(t_{m+1})\right), \quad 0 \le \theta \le 1.$$

On obtient alors le schéma

$$\left\{\begin{array}{l} \dfrac{1}{\Delta t_m}\left(M\left[u\right]_{m+1} - M\left[u\right]_m\right) = (1-\theta)\left\{\left[L\right]_m - K\left[u\right]_m\right\} + \theta(\left[L\right]_{m+1} - K\left[u\right]_{m+1}), \\ \left[u\right]_0 = \left[u^{(0)}\right], \end{array}\right.$$

où

$$\left[u\right]_m = \left[\begin{array}{c} u_{1,m} \\ \vdots \\ u_{N,m} \end{array}\right]$$

est le vecteur dont la composante $u_{j,m}$ donne l'approximation de la valeur nodale $u_j(t)$ à l'instant t_m et avec des notations analogues pour $\left[L\right]_m$.

On réécrit ce schéma en mettant tous les termes connus après m pas d'intégration dans le second membre

$$\left\{\begin{array}{l} (\dfrac{1}{\Delta t_m}M + \theta K)\left[u\right]_{m+1} = (1-\theta)\left[L\right]_m + \theta\left[L\right]_{m+1} + (\dfrac{1}{\Delta t_m}M - (1-\theta)K)\left[u\right]_m \\ \left[u\right]_0 = \left[u^{(0)}\right] \end{array}\right. \tag{3.4}$$

Les schémas pour $\theta < 1/2$ sont **conditionnellement stables**. Ceci les rend difficiles à utiliser, ou même quasiment inutisables, en pratique. En fait, le seul schéma intéressant de ce type est **le schéma d'Euler explicite** obtenu pour $\theta = 0$

$$\begin{cases} \dfrac{1}{\Delta t_m} M \left[u\right]_{m+1} = \left[L\right]_m + (\dfrac{1}{\Delta t_m} M - K) \left[u\right]_m \\ \left[u\right]_0 = \left[u^{(0)}\right] \end{cases}$$

Lorsque la matrice masse M est condensée, le calcul de $\left[u\right]_{m+1}$ se fait par résolution d'un système diagonal.

Les schémas pour $\theta > 1/2$ sont des schémas **fortement stables**. Le schéma avec la meilleure propriété de stabilité est le **schéma d'Euler implicite** obtenu pour $\theta = 1$

$$\begin{cases} (\dfrac{1}{\Delta t_m} M + K) \left[u\right]_{m+1} = \left[L\right]_{m+1} + \dfrac{1}{\Delta t_m} M \left[u\right]_m \\ \left[u\right]_0 = \left[u^{(0)}\right] \end{cases}$$

Pour $\theta = 1/2$, on obtient le **schéma de Crank-Nicholson**. C'est un schéma faiblement stable. Mais c'est le seul qui est un schéma d'ordre 2.

Observons que pour tout schéma avec $\theta > 0$, on doit résoudre à chaque pas d'intégration un problème de type éléments finis

$$Ax = b$$

avec

$$\begin{aligned} A &= (\frac{1}{\Delta t_m} M + \theta K) \\ x &= \left[u\right]_{m+1} \\ b &= (1-\theta) \left[L\right]_m + \theta \left[L\right]_{m+1} + (\frac{1}{\Delta t_m} M - (1-\theta) K) \left[u\right]_m \end{aligned}$$

Il est clair que les méthodes directes, si la taille du système n'est pas trop grande, sont les plus indiquées. On effectue la décomposition de A à l'aide de l'algorithme de Gauss sous la forme

$$A = LU.$$

A chaque pas d'intégration, on est amené ainsi à résoudre deux systèmes triangulaires.

3.2. Schéma de Neumark. Il s'agit maintenant de résoudre le système d'ordre 2 en temps

$$\begin{cases} M \left[\ddot{u}\right](t) + K \left[u\right](t) = L(t), \quad t \in \left[0, T\right], \\ \left[u\right](0) = \left[u^{(0)}\right], \quad \left[\dot{u}\right](0) = \left[u^{(1)}\right]. \end{cases} \tag{3.5}$$

Nous nous limitons au schéma de Neumark le plus utilisé. Il est utilisé sur une grille uniforme

$$\Delta t_m = \Delta t, \quad i = 0, \ldots, M-1.$$

Son obtention est effectuée par les deux approximations suivantes :

(1) **une approximation à l'ordre deux** de la dérivée seconde par une dérivation au sens des différences finies centrées

$$\begin{aligned} \left[\ddot{u}\right](t_m) &\approx \frac{\left[u\right](t_{m+1}) - 2\left[u\right](t_m) + \left[u\right](t_{m-1})}{\Delta t^2} \\ M \frac{\left[u\right](t_{m+1}) - 2\left[u\right](t_m) + \left[u\right](t_{m-1})}{\Delta t^2} &\approx L(t_m) - K\left[u\right](t_m) \end{aligned}$$

(2) **une approximation à l'ordre deux** pour le second membre pour améliorer les propriétés de stabilité

$$L(t_m) - K[u](t_m) \approx \frac{1}{4}\left((L(t_{m-1}) - K[u](t_{m-1})) + 2(L(t_m) - K[u](t_m)) + (L(t_{m+1}) - K[u](t_{m+1}))\right)$$

On obtient ainsi le schéma

$$\left(\frac{1}{\Delta t^2}M + \frac{1}{4}K\right)[u]_{m+1} = b_{m+1}$$

avec

$$b_{m+1} = \frac{1}{4}\left([L]_{m-1} + 2[L]_m + [L]_{m+1}\right) + \frac{1}{\Delta t^2}M\left(2[u]_m - [u]_{m-1}\right) - \frac{1}{4}K(2[u]_m + [u]_{m-1}).$$

On a besoin de deux données initiales : $[u]_0$ et $[u]_1$. On a

$$[u]_0 = \left[u^{(0)}\right].$$

On calcule $[u]_1$ par un développement de Taylor à l'ordre 2

$$[u]_1 \approx [u](\Delta t) \approx [u](0) + \Delta t[\dot{u}](0) + (\Delta t^2/2)[\ddot{u}](0)$$

On a

$$[\dot{u}](0) = \left[u^{(1)}\right]$$

En revenant à l'équation (3.5), on peut calculer

$$[\ddot{u}](0) = M^{-1}([L]_0 - K[u]_0).$$

4. Analyse modale

4.1. Résolution d'un système différentiel par diagonalisation. Rappelons le principe de résolution d'un système différentiel

$$\left\{\begin{array}{l} \frac{d^m}{dt^m}w(t) + Aw(t) = f(t) \\ \frac{d^\ell}{dt^\ell}w(0) = w^{(\ell)}, \quad 0 \le \ell \le m-1 \end{array}\right. \tag{4.1}$$

par diagonalisation. Supposons que A soit diagonalisable, i.e. qu'il existe une matrice inversible V et une matrice diagonale $\Lambda := \mathrm{diag}[\lambda_1, \cdots, \lambda_N]$ telles que

$$V^{-1}AV = \Lambda \tag{4.2}$$

Ceci revient en fait à supposer qu'il existe une base (de $\mathbb{R}^N$ ou de $\mathbb{C}^N$ selon que l'une au moins des valeurs propres soit complexe ou non) qui est en fait la famille de vecteurs $\{V_{*,j}\}_{j=1}^{j=N}$ formée des vecteurs colonne de V qui sont les vecteurs propres correspondant aux valeurs propres $\{\lambda_j\}_{j=1}^{j=N}$.

Multiplions à gauche le système (4.1) par V^{-1} et faisons le changement de fonction inconnue

$$v(t) = V^{-1}w(t). \tag{4.3}$$

On est ramené au système

$$\left\{\begin{array}{l} \frac{d^m}{dt^m}v(t) + \Lambda v(t) = F(t) \\ \frac{d^\ell}{dt^\ell}v(0) = v^{(\ell)}, \quad 0 \le \ell \le m-1 \end{array}\right. \tag{4.4}$$

avec

$$F(t) = V^{-1}f(t), \quad v^{(\ell)} = V^{-1}w^{(\ell)},\ 0 \le \ell \le m-1. \tag{4.5}$$

Ce système est découplé : il équivaut à la résolution de N équations différentielles découplées

$$\begin{cases} \frac{d^m}{dt^m} v_j(t) + \lambda_j v_j(t) = F_j(t) \\ \frac{d^\ell}{dt^\ell} v_j(0) = v_j^{(\ell)}, \quad 0 \leq \ell \leq m-1 \end{cases} \tag{4.6}$$

On le résout directement. On suppose pour simplifier la discussion et parce que cela sera vérifié pour le cas qui nous intéresse par la suite que les valeurs propres sont réelles et positives. On les ordonne de la façon suivante

$$0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N \tag{4.7}$$

(1) **Système d'ordre 1**

$$v_j(t) = v_j^{(0)} \exp(-\lambda_j t) + \int_0^t \exp(-\lambda_j (t-s)) F_j(s)\, ds, \quad j = 1, \ldots, N. \tag{4.8}$$

(2) **Système d'ordre 2**

(4.9)

$$v_j(t) = v_j^{(0)} \cos(t\sqrt{\lambda_j}) + v_j^{(1)} \frac{\sin(t\sqrt{\lambda_j})}{\sqrt{\lambda_j}} + \int_0^t \frac{\sin(\sqrt{\lambda_j}(t-s))}{\sqrt{\lambda_j}} F_j(s)\, ds, \quad j = 1, \ldots, N,$$

avec

$$v_j(t) = v_j^{(1)} t + \int_0^t (t-s) F_j(s)\, ds \tag{4.10}$$

si $\lambda_j = 0$.

On revient à la solution w par la transformation inverse de (4.3)

$$w(t) = V v(t) \tag{4.11}$$

4.2. Problème aux valeurs propres généralisé. Cependant, le sytème (2.12) n'est pas sous la forme (4.1). Comme M est symétrique définie positive, on peut en fait s'y ramener. On décompose M par l'algorithme de Cholesky

$$M = R^{\mathrm{T}} R \tag{4.12}$$

où R est triangulaire supérieure et inversible. Multiplions le système (2.12) par $(R^{-1})^{\mathrm{T}}$ et par R les conditions initiales. On note $u(t) = [u]\,(t)$ pour simplifier :

$$\begin{cases} \frac{d^m}{dt^m} Ru(t) + (R^{-1})^{\mathrm{T}} Ku(t) = (R^{-1})^{\mathrm{T}} L(t) \\ \frac{d^\ell}{dt^\ell} Ru(0) = Ru^{(\ell)}, \quad 0 \leq \ell \leq m-1 \end{cases} \tag{4.13}$$

Posons alors

$$w(t) := Ru(t) \tag{4.14}$$

et $(R^{-1})^{\mathrm{T}} L(t) = f(t)$. On est ainsi ramené au cas précédent avec

$$A := \left(R^{-1}\right)^{\mathrm{T}} K R^{-1}. \tag{4.15}$$

Cependant, comme K est symétrique positive, on a des résultats plus précis concernant ses valeurs propres et ses vecteurs propres.

PROPOSITION 7.2. *Il existe une matrice $V \in \mathbb{R}^{N,N}$ orthogonale, i.e.*

$$V^{-1} = V^T$$

telle que

$$V^T A V = \Lambda \tag{4.16}$$

avec

$$\Lambda = \mathrm{diag}[\lambda_1, \ldots, \lambda_N] \tag{4.17}$$

et

$$0 \leq \lambda_1 \leq \cdots \leq \lambda_N \tag{4.18}$$

DÉMONSTRATION. Il suffit de s'assurer que A est symétrique et positive. On a d'abord

$$A^{\mathrm{T}} = \left(\left(R^{-1}\right)^{\mathrm{T}} K R^{-1}\right)^{\mathrm{T}} = \left(R^{-1}\right)^{\mathrm{T}} K^{\mathrm{T}} R^{-1} = A$$

car $K^{\mathrm{T}} = K$. On a de même

$$v^{\mathrm{T}} A v = v^{\mathrm{T}} \left(R^{-1}\right)^{\mathrm{T}} K R^{-1} v = z^{\mathrm{T}} K z$$

avec $z = R^{-1}v$. Il vient donc

$$v^{\mathrm{T}} A v = z^{\mathrm{T}} K z \geq 0.$$

Ceci termine la démonstration. □

On déduit de cette proposition le théorème suivant.

THÉORÈME 7.1. *Il existe une matrice $U \in \mathbb{R}^{N,N}$ telle que*

$$U^T K U = \Lambda := \mathrm{diag}[\lambda_1, \ldots, \lambda_N] \tag{4.19}$$

$$U^T M U = \mathbb{I}_N \tag{4.20}$$

avec $\{\lambda_j\}_{j=1}^{j=N}$ vérifiant (4.18) et $\mathbb{I}_N$ la matrice identité de $\mathbb{R}^{N,N}$.

DÉMONSTRATION. On a en utilisant la proposition précédente

$$V^{\mathrm{T}} A V = \Lambda.$$

Posons alors

$$U = R^{-1} V.$$

On a ainsi (4.19). On a ensuite

$$U^{\mathrm{T}} M U = U^{\mathrm{T}} R^{\mathrm{T}} R U = V^T V = \mathbb{I}_N.$$

Ceci démontre le théorème. □

REMARQUE 7.2. *Il résulte de (4.19)*

$$K U = \left(U^{-1}\right)^T \Lambda$$

et de (4.20)

$$M U = \left(U^{-1}\right)^T$$

Il s'ensuit

$$K U = M U \Lambda.$$

Multiplions les deux membres de l'égalité précédente par le j-ème vecteur colonne de la matrice $\mathbb{I}_N$. Les vecteurs colonne $U_{*,j}$ de la matrice U constituent en fait une base de vecteurs vérifiantI

$$K U_{*,j} = \lambda_j M U_{*,j}. \tag{4.21}$$

Le couple $(\lambda_j, U_{*,j})$ forme une valeur propre et un vecteur propre pour le problème de valeurs propres généralisé

$$K w = \lambda M w. \tag{4.22}$$

4.3. Résolution modale. A partir des matrices intervenant dans les relations (4.19) et (4.20), on écrit le système (2.12) sous la forme

$$\begin{cases} \frac{d^m}{dt^m} U^{\mathrm{T}} M u(t) + U^{\mathrm{T}} K u(t) = F(t) \\ \frac{d^\ell}{dt^\ell} U u(0) = U u^{(\ell)}, \quad 0 \leq \ell \leq m-1 \end{cases}$$

On pose alors

$$u(t) = U v(t) \tag{4.23}$$

On se ramène à un système découplé

$$\begin{cases} \frac{d^m}{dt^m} v(t) + \Lambda v(t) = F(t) \\ \frac{d^\ell}{dt^\ell} v(0) = U u^{(\ell)}, \quad 0 \leq \ell \leq m-1 \end{cases} \tag{4.24}$$

qu'on résout par les formules (4.8) ou (4.9) ou (4.10).

La relation (4.23) s'écrit aussi

$$u(t) = \sum_{j=1}^{N} v_j(t) U_{*,j} \tag{4.25}$$

chaque $v_j(t)$ est solution de l'équation différentielle

$$\begin{cases} \frac{d^m}{dt^m} v_j(t) + \lambda_j v_j(t) = F_j(t) \\ \frac{d^\ell}{dt^\ell} v_j(0) = v_j^{(\ell)}, \quad 0 \leq \ell \leq m-1 \end{cases} \tag{4.26}$$

Chaque terme $v_j(t) U_{*,j}$ est appelé un mode pour le problème (2.12). Il nécessite pour être calculé, la détermination de la valeur propre λ_j et d'un vecteur propre $U_{*,j}$ normalisé par la condition

$$U_{*,j}^{\mathrm{T}} M U_{*j} = 1.$$

En fait, on n'a pas besoin de calculer tous les termes de la somme (4.25) pour obtenir une bonne approximation de $u(t)$. Il suffit de disposer des premières valeurs propres et vecteurs propres

$$u(t) \approx \sum_{j=1}^{J} v_j(t) U_{*,j}, \quad \text{avec } j \ll N$$

pour obtenir une approximationde la solution avec une bonne précision.

a Generic Finite Element library in C++

Documentation, part 3

DESCRIPTION OF FINITE ELEMENT AND INTEGRATION METHODS

Yves RENARD[1], JULIEN POMMIER[2]

16th January 2006

Introduction

This documentation describes the different finite element methods and cubature formulas available in GETFEM++.

[1] *MIP, INSAT, Complexe scientifique de Rangueil, 31077 Toulouse, France, Yves.Renard@insa-toulouse.fr*
[2] *MIP, INSAT, Complexe scientifique de Rangueil, 31077 Toulouse, France, Julien.Pommier@insa-toulouse.fr*

Contents

1 Finite element methods

All finite element methods defined in GETFEM++ are interfaced in the file `getfem_fem.h`. A descriptor on a finite element method is available thanks to the function

```
getfem::pfem pf = getfem::fem_descriptor("name of method");
```

where `"name of method"` is a string to be choosen among the existing methods.

1.1 Finite element methods description

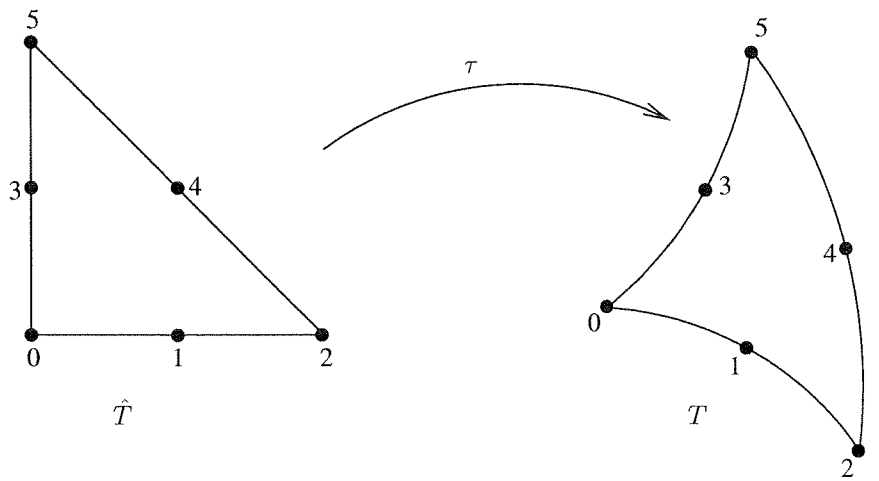


Figure 1: *Example of geometric transformation for a triangle.*

A finite element method is defined on a reference element $\hat{T} \subset \mathbb{R}^P$ by a set of n_d nodes a^i and corresponding base functions

$$\hat{\varphi}^i : \hat{T} \subset \mathbb{R}^P \longrightarrow \mathbb{R}^Q.$$

Each base function corresponds to a degree of freedom (d.o.f). Most finite element methods are scalar, which means that $Q = 1$, but GETFEM++ support also intrinsic vectorial elements. The map between the reference element and the real element is called the geometric transformation and is denoted by

$$\tau : \hat{T} \longrightarrow T,$$

and is supposed to be polynomial (see [1] or [2]). The base functions $\hat{\varphi}^i$ defined on the reference element define a set of base function on the real element defined by

$$\tilde{\varphi}^i(x) = \hat{\varphi}^i(\hat{x}) = \hat{\varphi}^i(\tau^{-1}(x)),$$

If the element is said to be equivalent throught the geometric transformation τ (or τ-equivalent) then base functions on the real element are just defined by

$$\varphi^i(x) = \tilde{\varphi}^i(x).$$

This is generally the case for Lagrange element, but not for Hermite elements (when some dof represent the gradient of the unkown). When the element is not equivalent throught the geometric transformation then GETFEM++ allows to define a square matrix $\tilde{M}$ depending on the real element (i.e. on the geometric transformation) such that base functions on the real element are defined by

$$\varphi^i(x) = \sum_{j=0}^{n_d-1} \tilde{M}_{ij}\tilde{\varphi}^j(x).$$

We denote by

$$[\hat{\varphi}(\hat{x})] = \begin{pmatrix} \hat{\varphi}^0(\hat{x}) \\ \hat{\varphi}^1(\hat{x}) \\ \cdots \\ \hat{\varphi}^{n_d-1}(\hat{x}) \end{pmatrix},$$

the $n_d \times Q$ matrix, such that when a function is defined by

$$f(x) = \sum_{i=0}^{n_d-1} \alpha_i \varphi^i(x),$$

one has

$$f(\tau(\hat{x})) = \alpha^T \tilde{M}[\hat{\varphi}(\hat{x})],$$

where α is the vector of components α_i.

1.1.1 Different types of d.o.f.

To each base function of a finite element method corresponds a degree of freedom (d.o.f) which is a linear form on this function. The following table gives the most significant types of d.o.f.

type	expression	commentary
Lagrange type	$\phi(a_i)$	Value of ϕ on the node a_i. The most simple d.o.f. Allows the Lagrange interpolation.
Hierarchical Lagrange type	$\phi(a_i) - \ldots$	Difference between the value of ϕ on the node a_i and the value of some other base functions. This is generally the bubble functions type of d.o.f .
mean type	$\frac{1}{\lvert T\rvert}\int_T \phi(x)dx$	Value of the mean value of ϕ on the element. Exists also for the restriction on a face.
derivative type	$\frac{\partial}{\partial x_i}\phi(a_i)$ or $\frac{\partial}{\partial n}\phi(a_i)$	Value of a derivative of ϕ on the node a_i. This kind of d.o.f. makes the element not to be τ-equivalent. $\frac{\partial}{\partial n}\phi(a_i)$ denotes the normal derivative with respect to a face.
second derivative type	$\frac{\partial^2}{\partial x_i \partial x_j}\phi(a_i)$	Value of a second derivative of ϕ on the node a_i. This kind of d.o.f. makes also the element not to be τ-equivalent.

1.1.2 Graphical codification of d.o.f.

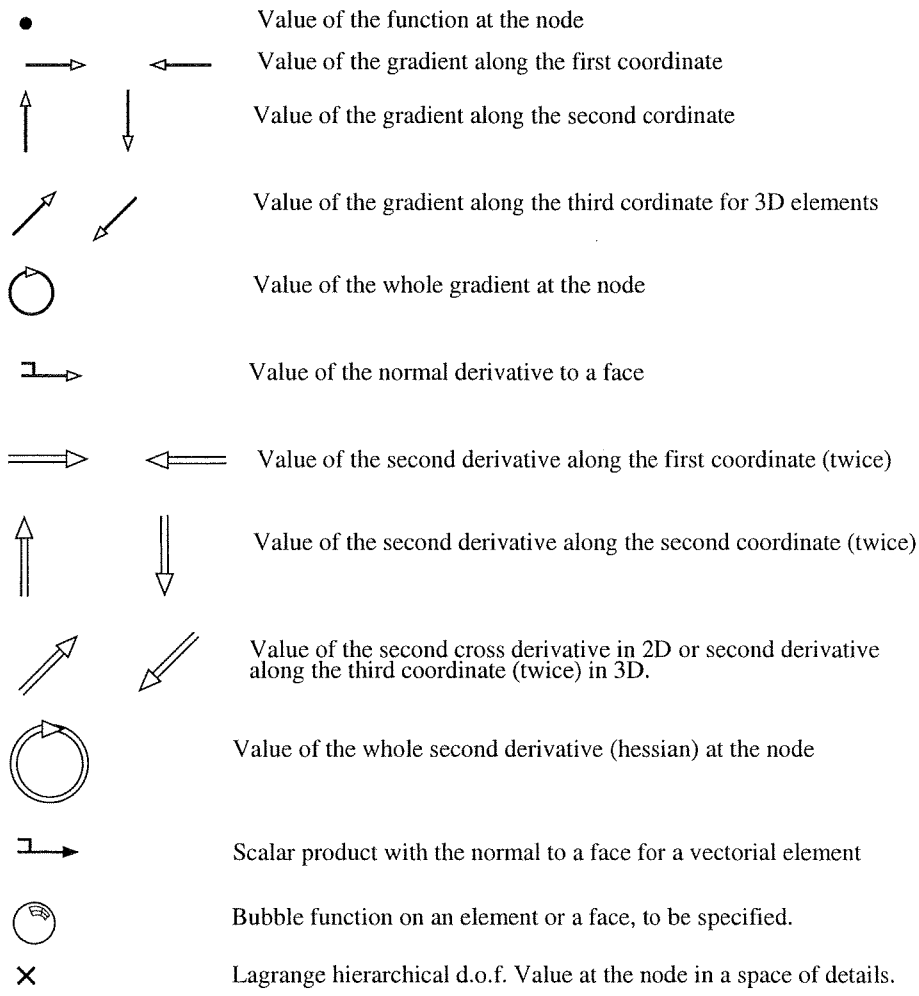


Figure 2: *Symbols representing degree of freedom types*

1.2 Classical "P_K" Lagrange elements on simplices

It is possible to define a classical "P_K" Lagrange element of arbitrary dimension and arbitrary degree. This element has only degrees of freedom which corresponds to the value of the function on a node. The grid of node is the so-called Lagrange grid. Figures 3, 4 and 5 show examples of dimension 1, 2 and 3.

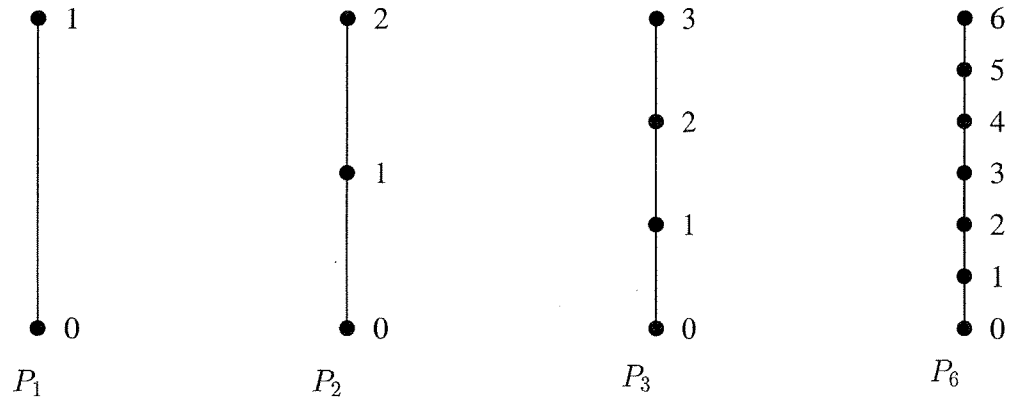


Figure 3: *Examples of classical P_K Lagrange elements on a segment.*

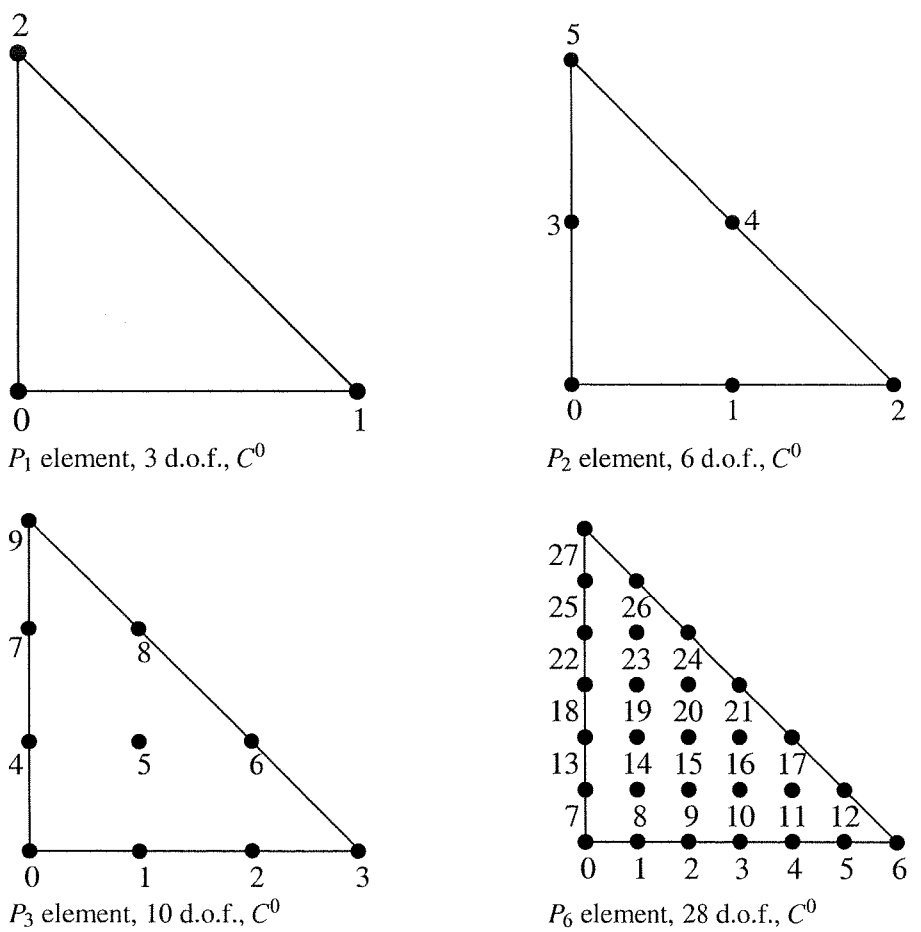


Figure 4: *Examples of classical P_K Lagrange elements on a triangle.*

The number of degree of freedom for a classical "P_K" Lagrange element of dimension P and degree K is $\frac{(P+K)!}{P!K!}$. For instance, in dimension 2 ($P=2$), this value is $\frac{(P+1)(P+2)}{2}$, in dimension 3 ($P=3$), this value is

$\frac{(P+1)(P+2)(P+3)}{6}$ ...

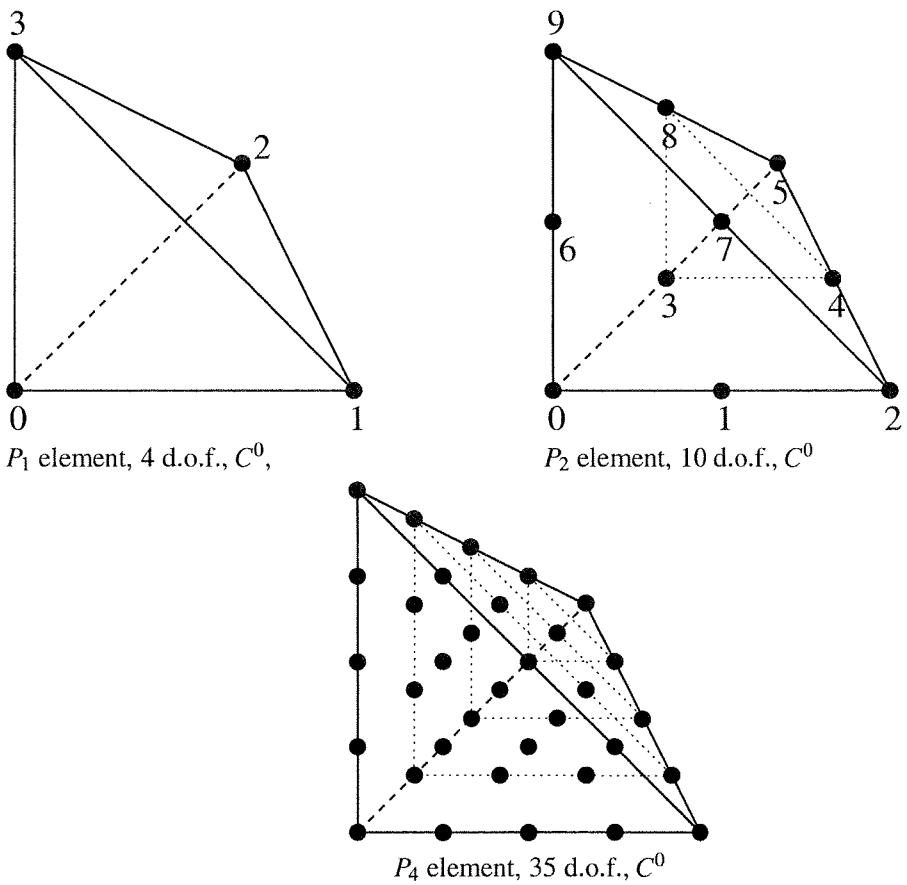


Figure 5: *Examples of classical P_K Lagrange elements on a tetrahedron.*

The particular way selected in GETFEM++ to numerate the nodes are also shown in figures 3, 4 and 5. Using another numeration, let

$$i_0, i_1, ...i_P,$$

be somme indexes such that

$$0 \leq i_0, i_1, ...i_P \leq K, \text{ and } \sum_{n=0}^{P} i_n = K.$$

Then, the coordinate of a node can be computed as

$$a_{i_0,i_1,...i_P} = \sum_{n=0}^{P} \frac{i_n}{K} S_n, \quad \text{for } K \neq 0,$$

where $S_0, S_1, ...S_N$ are the vertexes of the simplex (for $K = 0$ the particular choice $a_{0,0,...0} = \sum_{n=0}^{P} \frac{1}{P+1} S_n$ has been chosen). Then each base function, corresponding of each node $a_{i_0,i_1,...i_P}$ is defined by

$$\phi_{i_0,i_1,...i_P} = \prod_{n=0}^{P} \prod_{j=0}^{i_n-1} \left(\frac{K\lambda_n - j}{j+1} \right).$$

where λ_n are the barycentric coordinates, i.e. the polynomials of degree 1 whose value is 1 on the vertex S_n and whose value is 0 on other vertices. On the reference element, one has

$$\lambda_n = x_n, \ 0 \leq n < P,$$

$$\lambda_P = 1 - x_0 - x_1 - \ldots - x_{P-1}.$$

When between two elements of the same degrees (even with different dimensions), the d.o.f. of a common face are linked, the element is of class C^0. This means that the global polynomial is continuous. If you try to link elements of different degrees, you will get some trouble with the unlinked d.o.f. This is not automatically supported by GETFEM++, so you will have to support it (add constraints on these d.o.f.).

For some applications (computation of a gradient for instance) one does not want the d.o.f. of a common face to be linked. This is why there are two versions of the classical "P_K" Lagrange element.

Classical "P_K" Lagrange element `"FEM_PK(P, K)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K, $0 \leq K \leq 255$	P, $1 \leq P \leq 255$	$\frac{(K+P)!}{K!P!}$	C^0	No ($Q=1$)	Yes ($\tilde{M} = Id$)	Yes

Discontinuous "P_K" Lagrange element `"FEM_PK_DISCONTINUOUS(P, K)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K, $0 \leq K \leq 255$	P, $1 \leq P \leq 255$	$\frac{(K+P)!}{K!P!}$	discon-tinuous	No ($Q=1$)	Yes ($\tilde{M} = Id$)	Yes

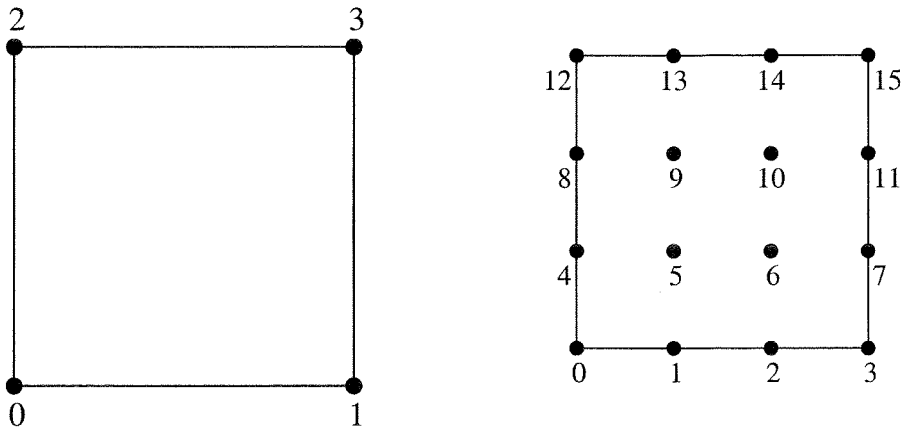
Even thought Lagrange elements are defined for arbitrary degrees, to choose a hight degree can be problematic for a large number of applications due to the "noisy" caracteristic of the lagrange basis. Those element are recommended for the basic interpolation but for p.d.e. applications elements with hierarchical basis are preferable (see the corresponding section).

1.3 Classical Lagrange elements on other geometries

Classical Lagrange elements on parallelepipeds or prisms are obtained as tensorial product of Lagrange elements on simplices. When two element are defined, one on a dimension P_1 and the other in dimension P_2, one obtains the base functions of the tensorial product (on the reference element) as

$$\hat{\phi}_{ij}(x,y) = \hat{\phi}^1_i(x)\hat{\phi}^2_j(y), \ x \in \mathbb{R}^{P_1}, y \in \mathbb{R}^{P_2},$$

where $\hat{\phi}^1_i$ and $\hat{\phi}^2_i$ are respectively the base functions of the first and second element.

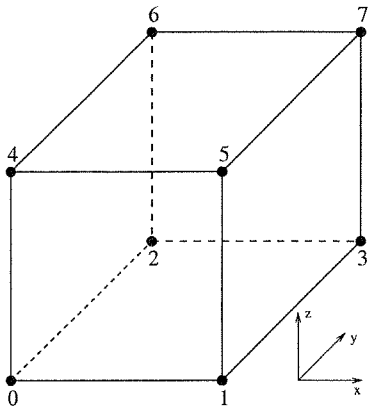


Q_1 element, 4 d.o.f., C^0 Q_3 element, 16 d.o.f., C^0

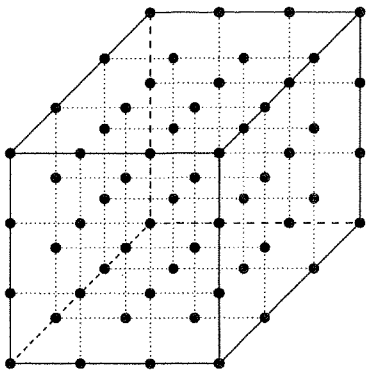
Figure 6: *Examples of classical Lagrange elements in dimension 2*

The Q_K element on a parallelepiped of dimension P is obtained as the tensorial product of P classical P_K element on the segment. Examples in dimension 2 are shown in figure 6 and in dimension 3 in figure 7.

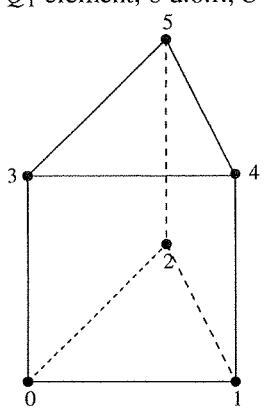
A prism in dimension $P > 1$ is the direct product of a simplex of dimension $P-1$ with a segment. The $P_K \otimes P_K$ element on this prism is the tensorial product of the classical P_K element on a simplex of dimension $P-1$ with the classical P_K element on a segment. For $P = 2$ this coincide with a parallelepiped. Examples in dimension 3 are shown in figure 7. This is also possible not to have the same degree on each dimension. An example is shown on figure 8.



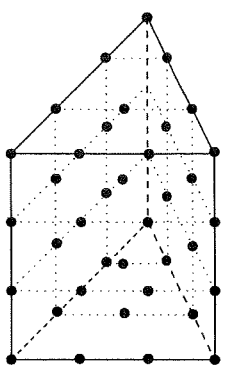
Q_1 element, 8 d.o.f., C^0



Q_3 element, 64 d.o.f., C^0



$P_1 \otimes P_1$ element, 6 d.o.f., C^0



$P_3 \otimes P_3$ element, 40 d.o.f., C^0

Figure 7: *Examples of classical Lagrange elements in dimension 3*

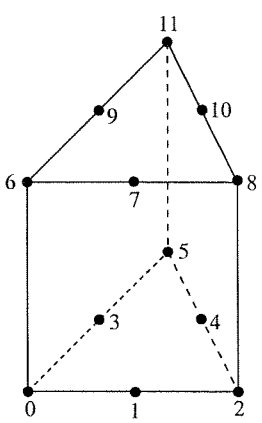
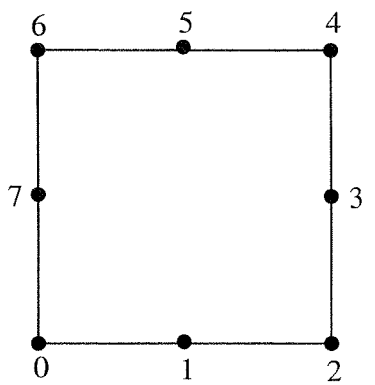


Figure 8: $P_2 \otimes P_1$ *Lagrange element on a prism, 12 d.o.f.,* C^0

Q_K **Lagrange element on parallelepipeds** `"FEM_QK(P, K)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
KP, $0 \leq K \leq 255$	P, $2 \leq P \leq 255$	$(K+1)^P$	C^0	No ($Q=1$)	Yes ($\tilde{M}=Id$)	Yes

$P_K \otimes P_K$ **Lagrange element on prisms** "FEM_PK_PRISM(P, K)"						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
$2K$, $0 \leq K \leq 255$	P, $2 \leq P \leq 255$	$(K+1)$ $\times \frac{(K+P-1)!}{K!(P-1)!}$	C^0	No $(Q=1)$	Yes $(\tilde{M}=Id)$	Yes

$P_{K_1} \otimes P_{K_2}$ **Lagrange element on prisms** "FEM_PRODUCT(FEM_PK(P-1, K_1), FEM_PK(1, K_2))"						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K_1+K_2, $0 \leq K_1, K_2 \leq 255$	P, $2 \leq P \leq 255$	(K_2+1) $\times \frac{(K_1+P-1)!}{K_1!(P-1)!}$	C^0	No $(Q=1)$	Yes $(\tilde{M}=Id)$	Yes



Incomplete Q_2 element, 8 d.o.f., C^0

Incomplete Q_2 Lagrange element on quadrilateral (Quad 8 serendipity element) "FEM_INCOMPLETE_Q2"						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
3	2	8	C^0	No $(Q=1)$	Yes $(\tilde{M}=Id)$	Yes

1.4 Elements with hierarchical basis

The idea behind hierarchical basis is the desciption of the solution at different level : a rought level, a more refined level ... In the same discretisation some degrees of freedom represent the rought description, some other the more rafined and so on. This correspond to imbricated spaces of discretisation. The hierarchical basis contains a basis of each of these spaces (this is not the case in classical Lagrange elements when the mesh is refined).

Among the advantages, the condition number of rigidity matrices can be greatly improved, it allows local raffinement and a resolution with a multigrid approach.

1.4.1 Hiercarchical elements with respect to the degree

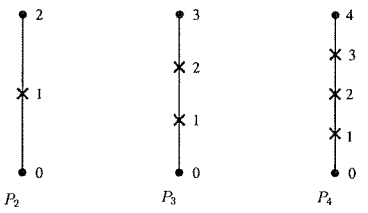


Figure 9: *P_K Hierarchical element on a segment, C^0*

P_K Classical Lagrange element on simplices but with a hierarchical basis with respect to the degree `"FEM_PK_HIERARCHICAL(P,K)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K, $0 \leq K \leq 255$	P, $1 \leq P \leq 255$	$\frac{(K_1+P)!}{K_1!P!}$	C^0	No $(Q=1)$	Yes $(\tilde{M}=Id)$	Yes

Q_K Classical Lagrange element on parallelepipeds but with a hierarchical basis with respect to the degree `"FEM_QK_HIERARCHICAL(P,K)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K, $0 \leq K \leq 255$	P, $2 \leq P \leq 255$	$(K+1)^P$	C^0	No $(Q=1)$	Yes $(\tilde{M}=Id)$	Yes

P_K Classical Lagrange element on prisms but with a hierarchical basis with respect to the degree `"FEM_PK_PRISM_HIERARCHICAL(P,K)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K, $0 \leq K \leq 255$	P, $2 \leq P \leq 255$	$(K+1) \times \frac{(K+P-1)!}{K!(P-1)!}$	C^0	No $(Q=1)$	Yes $(\tilde{M}=Id)$	Yes

some particular choices : P_4 will be build with the basis of the P_1, the additional basis of the P_2 then the additionnal basis of the P_4.

P_6 will be build with the basis of the P_1, the additional basis of the P_2 then the additionnal basis of the P_6 (not with the basis of the P_1, the additional basis of the P_3 then the additionnal basis of the P_6, this is possible to build the latter with `"FEM_GEN_HIERARCHICAL(a,b))`

1.4.2 Composite elements

The principal interest of the composite elements is to build hierarchical elements. But this tool can also be used to build piecewise polynomial elements.

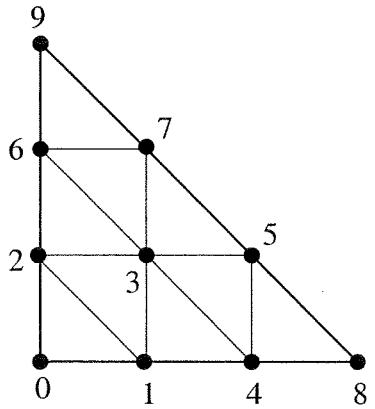


Figure 10: *composite element* `"FEM_STRUCTURED_COMPOSITE(FEM_PK(2,1), 3)"`

composition of a finite element method on a element with `S` **subdivisions** `"FEM_STRUCTURED_COMPOSITE(FEM1, S)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
degree of FEM1	dimension of FEM1	variable	variable	No ($Q = 1$)	If `FEM1` is	piecewise

It is important to use a corresponding composite integration method.

1.4.3 Hierarchical composite elements

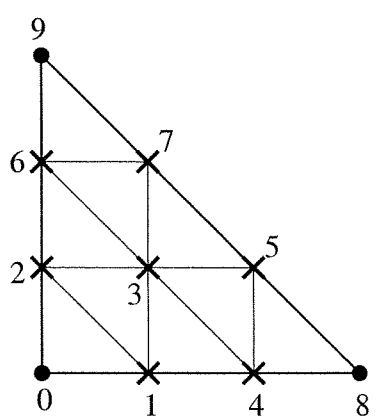


Figure 11: *hierarchical composite element* `"FEM_PK_HIERARCHICAL_COMPOSITE(2,1,3)"`

hierarchical composition of a P_K finite element method on a simplex with `S` **subdivisions** `"FEM_PK_HIERARCHICAL_COMPOSITE(P,K,S)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K	P	$\frac{(SK+P)!}{(SK)!P!}$	variable	No ($Q = 1$)	If `FEM1` is	piecewise

hierarchical composition of a hierarchical P_K finite element method on a simplex with `S` **subdivisions** `"FEM_PK_FULL_HIERARCHICAL_COMPOSITE(P,K,S)"`

Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K	P	$\frac{(SK+P)!}{(SK)!P!}$	variable	No ($Q=1$)	If FEM1 is	piecewise

Other constructions are possible thanks to "FEM_GEN_HIERARCHICAL(FEM1, FEM2)" and "FEM_STRUCTURED_COMPOSITE(FEM1, S)"

It is important to use a corresponding composite integration method.

1.5 Specific elements in dimension 1

1.5.1 GaussLobatto element

The 1D GaussLobatto P_K element is similar to the classical P_K fem on the segment, but the nodes are given by the Gauss-Lobatto-Legendre quadrature rule of order $2K-1$. This FEM is known to lead to better conditioned linear systems, and can be used with the correspounding quadrature to perform mass-lumping (on segments or parallelepipeds).

The polynomials coefficients have been pre-computed with Maple (they require the inversion of an ill-conditionned system), hence they are only available for the following values
of K: $1,2,3,4,5,6,7,8,9,10,11,12,13,14,16,24,32$. Note that for $K=1$ and $K=2$, this is the classical $P1$ and $P2$ fem.

GaussLobatto P_K element on the segment "FEM_PK_GAUSSLOBATTO1D(K)"						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
K	1	$K+1$	C^0	No ($Q=1$)	Yes	Yes

1.5.2 Hermite element

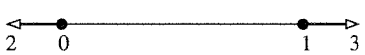


Figure 12: *P_3 Hermite element on a segment, 4 d.o.f., C^1*

Base functions on the reference element

$$\hat{\varphi}_0 = (2x+1)(x-1)^2, \qquad \hat{\varphi}_1 = x^2(3-2x),$$
$$\hat{\varphi}_2 = x(x-1)^2, \qquad \hat{\varphi}_3 = x^2(x-1).$$

This element is close to be τ-equivalent but it is not. On the real element the value of the gradient on vertices will be multiplied by the gradient of the geometric transformation. The matrix $\tilde{M}$ is not equal to identity but is still diagonal.

Hermite element on the segment "FEM_HERMITE(1)"						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
3	1	4	C^1	No ($Q=1$)	No	Yes

1.5.3 Lagrange element with an additional bubble function

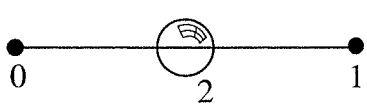
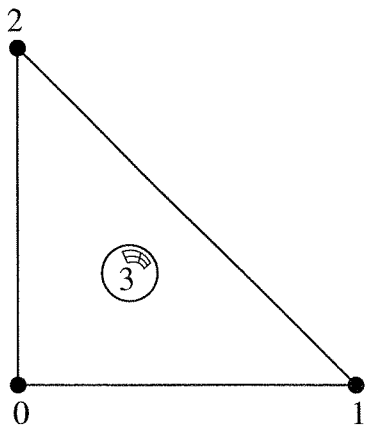


Figure 13: *P_1 Lagrange element on a segment with additional internal bubble function, 3 d.o.f., C^0*

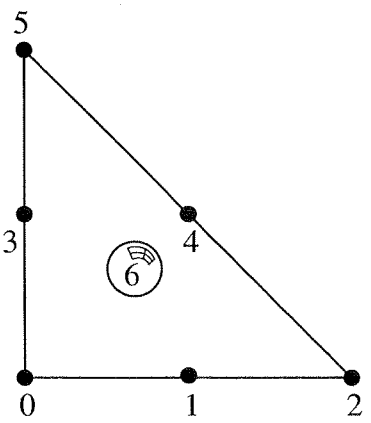
Lagrange P_1 element with an additional internal bubble function `"FEM_PK_WITH_CUBIC_BUBBLE(1, 1)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
2	1	3	C^0	No ($Q = 1$)	Yes	Yes

1.6 Specific elements in dimension 2

1.6.1 Elements with additional bubble functions



P_1 with additional bubble function, 4 d.o.f., C^0



P_2 with additional bubble function, 7 d.o.f., C^0

Figure 14: *Lagrange element on a triangle with additional internal bubble function*

Lagrange P_1 or P_2 element with an additional internal bubble function `"FEM_PK_WITH_CUBIC_BUBBLE(2, K)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
3	2	4 or 7	C^0	No ($Q = 1$)	Yes	Yes

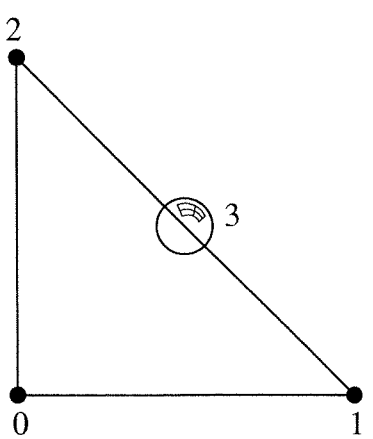


Figure 15: P_1 *Lagrange element on a triangle with additional bubble function on face 0, 4 d.o.f.,* C^0

Lagrange P_1 element with an additional bubble function on face 0 `"FEM_P1_BUBBLE_FACE(2)"`

Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
2	2	4	C^0	No ($Q=1$)	Yes	Yes

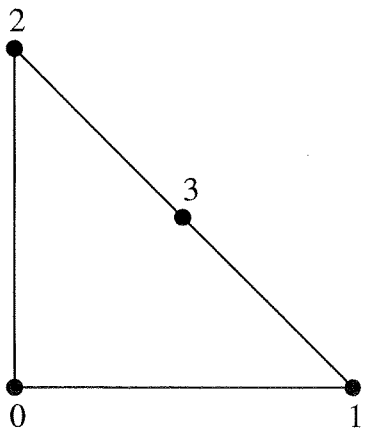


Figure 16: *P_1 Lagrange element on a triangle with additional d.o.f on face 0, 4 d.o.f., C^0*

P_1 Lagrange element on a triangle with additional d.o.f on face 0 `"FEM_P1_BUBBLE_FACE_LAG"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
2	2	4	C^0	No ($Q=1$)	Yes	Yes

1.6.2 Non-conforming P_1 element

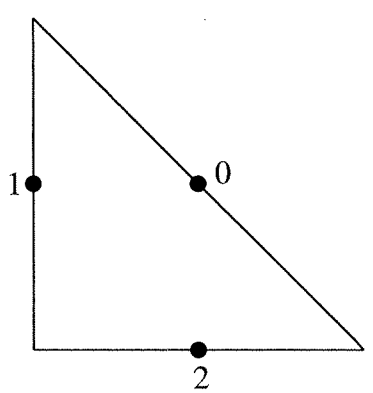


Figure 17: *P_1 non-conforming element on a triangle, 3 d.o.f., discontinuous*

P_1 non-conforming element on a triangle `"FEM_P1_NONCONFORMING"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
1	2	3	discon-tinuous	No ($Q=1$)	Yes	Yes

1.6.3 Hermite element

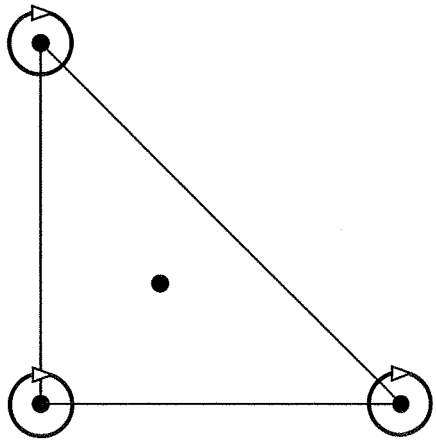


Figure 18: *Hermite element on a triangle, P_3, 10 d.o.f., C^0*

Base functions on the reference element:

$$
\begin{aligned}
\hat{\varphi}_0 &= (1-x-y)(1+x+y-2x^2-2y^2-11xy), & (\hat{\varphi}_0(0,0)=1),\\
\hat{\varphi}_1 &= -2x^3+7x^2y+7xy^2+3x^2-7xy, & (\hat{\varphi}_1(1,0)=1),\\
\hat{\varphi}_2 &= 7x^2y+7xy^2-2y^3+3y^2-7xy, & (\hat{\varphi}_2(0,1)=1),\\
\hat{\varphi}_3 &= 27xy(1-x-y), & (\hat{\varphi}_3(1/3,1/3)=1),\\
\hat{\varphi}_4 &= x(1-x-y)(1-x-2y), & (\partial_x\hat{\varphi}_4(0,0)=1),\\
\hat{\varphi}_5 &= x^3-2x^2y-2xy^2-x^2+2xy, & (\partial_x\hat{\varphi}_5(1,0)=1),\\
\hat{\varphi}_6 &= xy(x+2y-1), & (\partial_x\hat{\varphi}_6(0,1)=1),\\
\hat{\varphi}_7 &= y(1-x-y)(1-2x-y), & (\partial_y\hat{\varphi}_7(0,0)=1),\\
\hat{\varphi}_8 &= xy(y+2x-1), & (\partial_y\hat{\varphi}_8(1,0)=1),\\
\hat{\varphi}_9 &= y^3-2x^2y-2xy^2-y^2+2xy, & (\partial_y\hat{\varphi}_9(0,1)=1),
\end{aligned}
$$

This element is not τ-equivalent (The matrix $\tilde{M}$ is not equal to identity). On the real element linear combinaisons of $\hat{\varphi}_4$ and $\hat{\varphi}_7$ are used to match the gradient on the corresponding vertex. Idem for the two couples ($\hat{\varphi}_5$, $\hat{\varphi}_8$) and ($\hat{\varphi}_6$, $\hat{\varphi}_9$) for the two other vertices.

Hermite element on a triangle "FEM_HERMITE(2)"						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
3	2	10	C^0	No ($Q=1$)	No	Yes

1.6.4 Argyris element

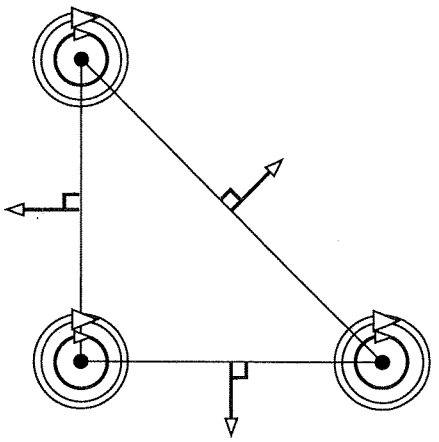


Figure 19: *Argyris element, P_5, 21 d.o.f., C^1*

The base functions on the reference element are:

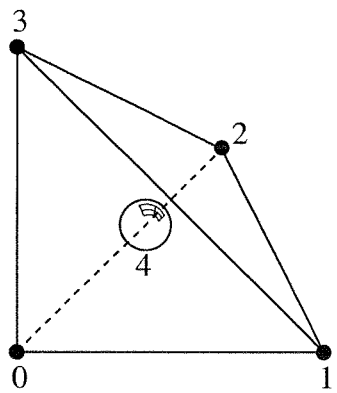
$$
\begin{aligned}
\hat{\varphi}_0(x,y) &= 1-10x^3-10y^3+15x^4-30x^2y^2+15y^4-6x^5+30x^3y^2+30x^2y^3-6y^5, && (\hat{\varphi}_0(0,0)=1),\\
\hat{\varphi}_1(x,y) &= 10x^3-15x^4+15x^2y^2+6x^5-15x^3y^2-15x^2y^3, && (\hat{\varphi}_1(1,0)=1),\\
\hat{\varphi}_2(x,y) &= 10y^3+15x^2y^2-15y^4-15x^3y^2-15x^2y^3+6y^5, && (\hat{\varphi}_2(0,1)=1),\\
\hat{\varphi}_3(x,y) &= x-6x^3-11xy^2+8x^4+10x^2y^2+18xy^3-3x^5+x^3y^2-10x^2y^3-8xy^4, && (\partial_x\hat{\varphi}_3(0,0)=1),\\
\hat{\varphi}_4(x,y) &= -4x^3+7x^4-3.5x^2y^2-3x^5+3.5x^3y^2+3.5x^2y^3, && (\partial_x\hat{\varphi}_4(1,0)=1),\\
\hat{\varphi}_5(x,y) &= -5xy^2+18.5x^2y^2+14xy^3-13.5x^3y^2-18.5x^2y^3-8xy^4, && (\partial_x\hat{\varphi}_5(0,1)=1),\\
\hat{\varphi}_6(x,y) &= y-11x^2y-6y^3+18x^3y+10x^2y^2+8y^4-8x^4y-10x^3y^2+x^2y^3-3y^5, && (\partial_y\hat{\varphi}_6(0,0)=1),\\
\hat{\varphi}_7(x,y) &= -5x^2y+14x^3y+18.5x^2y^2-8x^4y-18.5x^3y^2-13.5x^2y^3, && (\partial_y\hat{\varphi}_7(1,0)=1),\\
\hat{\varphi}_8(x,y) &= -4y^3-3.5x^2y^2+7y^4+3.5x^3y^2+3.5x^2y^3-3y^5, && (\partial_y\hat{\varphi}_8(0,0)=1),\\
\hat{\varphi}_9(x,y) &= 0.5x^2-1.5x^3+1.5x^4-1.5x^2y^2-0.5x^5+1.5x^3y^2+x^2y^3, && (\partial^2_{xx}\hat{\varphi}_9(0,0)=1),\\
\hat{\varphi}_{10}(x,y) &= 0.5x^3-x^4+0.25x^2y^2+0.5x^5-0.25x^3y^2-0.25x^2y^3, && (\partial^2_{xx}\hat{\varphi}_{10}(1,0)=1),\\
\hat{\varphi}_{11}(x,y) &= 1.25x^2y^2-1.25x^3y^2-0.75x^2y^3, && (\partial^2_{xx}\hat{\varphi}_{11}(0,1)=1),\\
\hat{\varphi}_{12}(x,y) &= xy-4x^2y-4xy^2+5x^3y+10x^2y^2+5xy^3-2x^4y-6x^3y^2-6x^2y^3-2xy^4, && (\partial^2_{xy}\hat{\varphi}_{12}(0,0)=1),\\
\hat{\varphi}_{13}(x,y) &= x^2y-3x^3y-3.5x^2y^2+2x^4y+3.5x^3y^2+2.5x^2y^3, && (\partial^2_{xy}\hat{\varphi}_{13}(1,0)=1),\\
\hat{\varphi}_{14}(x,y) &= xy^2-3.5x^2y^2-3xy^3+2.5x^3y^2+3.5x^2y^3+2xy^4, && (\partial^2_{xy}\hat{\varphi}_{14}(0,1)=1),\\
\hat{\varphi}_{15}(x,y) &= 0.5y^2-1.5y^3-1.5x^2y^2+1.5y^4+x^3y^2+1.5x^2y^3-0.5y^5, && (\partial^2_{yy}\hat{\varphi}_{15}(0,0)=1),\\
\hat{\varphi}_{16}(x,y) &= 1.25x^2y^2-0.75x^3y^2-1.25x^2y^3, && (\partial^2_{yy}\hat{\varphi}_{16}(1,0)=1),\\
\hat{\varphi}_{17}(x,y) &= 0.5y^3+0.25x^2y^2-y^4-0.25x^3y^2-0.25x^2y^3+0.5y^5, && (\partial^2_{yy}\hat{\varphi}_{17}(0,1)=1),\\
\hat{\varphi}_{18}(x,y) &= \sqrt{2}(-8x^2y^2+8x^3y^2+8x^2y^3), && (\sqrt{0.5}(\partial_x\hat{\varphi}_0(0.5,0.5)+\partial_y\hat{\varphi}_{18}(0.5,0.5))=1),\\
\hat{\varphi}_{19}(x,y) &= -16xy^2+32x^2y^2+32xy^3-16x^3y^2-32x^2y^3-16xy^4, && (-\partial_x\hat{\varphi}_{19}(0,0.5)=1),\\
\hat{\varphi}_{20}(x,y) &= -16x^2y+32x^3y+32x^2y^2-16x^4y-32x^3y^2-16x^2y^3, && (-\partial_y\hat{\varphi}_{20}(0.5,0)=1),
\end{aligned}
$$

This element is not τ-equivalent (The matrix $\tilde{M}$ is not equal to identity). On the real element linear combinaisons of the transformed base functions $\hat{\varphi}_i$ are used to match the gradient, the second derivatives and the normal derivatives on the faces. Note that the use of the matrix $\tilde{M}$ (see also the documentation on the finite element kernel [2]) allows to define Argyris element even with nonlinear geometric transformations (for instance to treat curved boundaries).

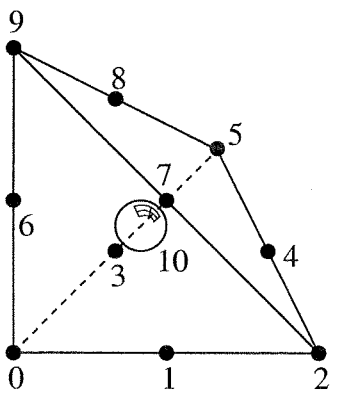
Argyris element on a triangle "FEM_ARGYRIS"						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
5	2	21	C^1	No ($Q=1$)	No	Yes

1.7 Specific elements in dimension 3

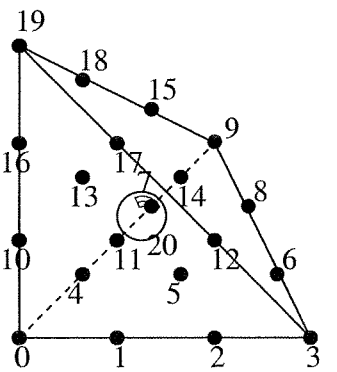
1.7.1 Elements with additional bubble functions



P_1 with additional bubble function, 5 d.o.f., C^0



P_2 with additional bubble function, 11 d.o.f., C^0



P_3 with additional bubble function, 21 d.o.f., C^0

Figure 20: *Lagrange element on a tetrahedron with additional internal bubble function.*

P_K Lagrange element with an additional internal bubble function `"FEM_PK_WITH_CUBIC_BUBBLE(3, K)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
4	3	5, 11 or 21	C^0	No ($Q = 1$)	Yes	Yes

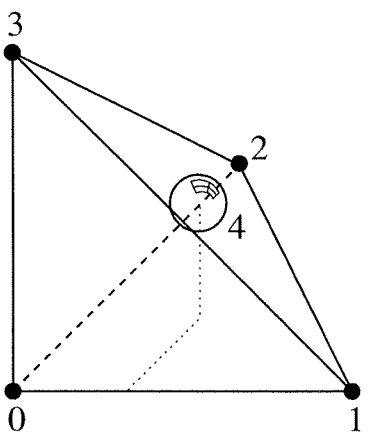


Figure 21: *P_1 Lagrange element on a tetrahedron with additional bubble function on face 0, 5 d.o.f., C^0*

Lagrange P_1 element with an additional bubble function on face 0 `"FEM_P1_BUBBLE_FACE(3)"`						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
3	3	5	C^0	No ($Q = 1$)	Yes	Yes

1.7.2 Hermite element

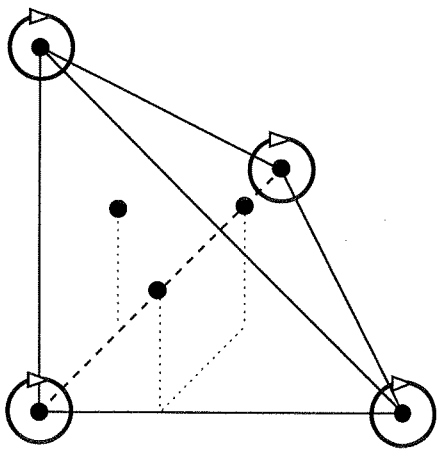


Figure 22: *Hermite element on a tetrahedron, P_3, 20 d.o.f., C^0*

Base functions on the reference element:

$$
\begin{aligned}
\hat{\varphi}_0(x,y) &= 1-3x^2-13xy-13xz-3y^2-13yz-3z^2+2x^3+13x^2y+13x^2z \\
&\quad +13xy^2+33xyz+13xz^2+2y^3+13y^2z+13yz^2+2z^3, && (\hat{\varphi}_0(0,0,0)=1),\\
\hat{\varphi}_1(x,y) &= 3x^2-7xy-7xz-2x^3+7x^2y+7x^2z+7xy^2+7xyz+7xz^2, && (\hat{\varphi}_0(1,0,0)=1),\\
\hat{\varphi}_2(x,y) &= -7xy+3y^2-7yz+7x^2y+7xy^2+7xyz-2y^3+7y^2z+7yz^2, && (\hat{\varphi}_0(0,1,0)=1),\\
\hat{\varphi}_3(x,y) &= -7xz-7yz+3z^2+7x^2z+7xyz+7xz^2+7y^2z+7yz^2-2z^3, && (\hat{\varphi}_0(0,0,1)=1),\\
\hat{\varphi}_4(x,y) &= 27xyz, && (\hat{\varphi}_0(1/3,1/3,1/3)=1),\\
\hat{\varphi}_5(x,y) &= 27yz-27xyz-27y^2z-27yz^2, && (\hat{\varphi}_0(0,1/3,1/3)=1),\\
\hat{\varphi}_6(x,y) &= 27xz-27x^2z-27xyz-27xz^2, && (\hat{\varphi}_0(1/3,0,1/3)=1),\\
\hat{\varphi}_7(x,y) &= 27xy-27x^2y-27xy^2-27xyz, && (\hat{\varphi}_0(1/3,1/3,0)=1),\\
\hat{\varphi}_8(x,y) &= x-2x^2-3xy-3xz+x^3+3x^2y+3x^2z+2xy^2+4xyz+2xz^2, && (\hat{\partial_x\varphi}_0(0,0,0)=1),\\
\hat{\varphi}_9(x,y) &= -x^2+2xy+2xz+x^3-2x^2y-2x^2z-2xy^2-2xyz-2xz^2, && (\hat{\partial_x\varphi}_0(1,0,0)=1),\\
\hat{\varphi}_{10}(x,y) &= -xy+x^2y+2xy^2, && (\hat{\partial_x\varphi}_0(0,1,0)=1),\\
\hat{\varphi}_{11}(x,y) &= -xz+x^2z+2xz^2, && (\hat{\partial_x\varphi}_0(0,0,1)=1),\\
\hat{\varphi}_{12}(x,y) &= y-3xy-2y^2-3yz+2x^2y+3xy^2+4xyz+y^3+3y^2z+2yz^2, && (\hat{\partial_y\varphi}_0(0,0,0)=1),\\
\hat{\varphi}_{13}(x,y) &= -xy+2x^2y+xy^2, && (\hat{\partial_y\varphi}_0(1,0,0)=1),\\
\hat{\varphi}_{14}(x,y) &= 2xy-y^2+2yz-2x^2y-2xy^2-2xyz+y^3-2y^2z-2yz^2, && (\hat{\partial_y\varphi}_0(0,1,0)=1),\\
\hat{\varphi}_{15}(x,y) &= -yz+y^2z+2yz^2, && (\hat{\partial_y\varphi}_0(0,0,1)=1),\\
\hat{\varphi}_{16}(x,y) &= z-3xz-3yz-2z^2+2x^2z+4xyz+3xz^2+2y^2z+3yz^2+z^3, && (\hat{\partial_z\varphi}_0(0,0,0)=1),\\
\hat{\varphi}_{17}(x,y) &= -xz+2x^2z+xz^2, && (\hat{\partial_z\varphi}_0(1,0,0)=1),\\
\hat{\varphi}_{18}(x,y) &= -yz+2y^2z+yz^2, && (\hat{\partial_z\varphi}_0(0,1,0)=1),\\
\hat{\varphi}_{19}(x,y) &= 2xz+2yz-z^2-2x^2z-2xyz-2xz^2-2y^2z-2yz^2+z^3, && (\hat{\partial_z\varphi}_0(0,0,1)=1),
\end{aligned}
$$

This element is not τ-equivalent (The matrix $\tilde{M}$ is not equal to identity). On the real element linear combinaisons of $\hat{\varphi}_8$, $\hat{\varphi}_{12}$ and $\hat{\varphi}_{16}$ are used to match the gradient on the corresponding vertex. Idem on the orther vertices.

Hermite element on a tetrahedron "FEM_HERMITE(3)"						
Degree	dimension	d.o.f. number	class	vectorial	τ-equivalent	Polynomial
3	3	20	C^0	No ($Q=1$)	No	Yes

1.8 Interpolation of elements on different meshes

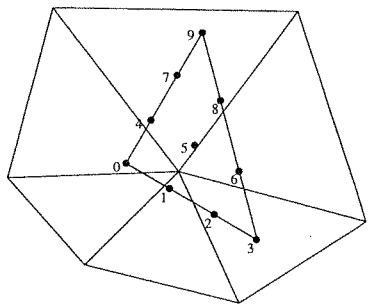


Figure 23: *Element which intepolates a finite element method defined on another mesh. The element has as many d.o.f. as the union of d.o.f. of elements of the other mesh having an intersection with it. The interpolation is made on Gauss points of the integration method.*

To increase the precision, it is not necessary to raise the order of the integration method. It is recommended to keep the normal order and use composite integration methods (see below).

Element which interpolates an element defined on another mesh

```
getfem::virtual_link_fem(getfem::mesh_fem mf1, getfem::mesh_fem mf2,
                         getfem::pintegration_method pim)
          getfem::virtual_link_fem_with_gradient(getfem::mesh_fem mf1,
          getfem::mesh_fem mf2, getfem::pintegration_method pim)
```

2 Integration methods

2.1 Integration methods description

The integration methods are of two kinds. Exact integrations of polynomials and approximated integrations (cubature formulas) of any function. The exact integration can only be used if all the elements are polynomial and if the geometric transformation is linear.

A descriptor on an integration method is available thanks to the function

```
ppi = getfem::int_method_descriptor("name of method");
```

where `"name of method"` is a string to be choosen among the existing methods.

The program `integration` located in the `tests` directory lists and checks the degree of each integration method.

2.2 Exact Integration methods

The list of available Exact integration methods is the following

`"IM_NONE()"`	Dummy integration method (new in getfem++-1.7).
`"IM_EXACT_SIMPLEX(n)"`	Description of the exact integration of polynomials on the simplex of reference of dimension `n`.
`"IM_PRODUCT(a, b)"`	Description of the exact integration on the convex which is the direct product of the convex in `a` and in `b`.

"IM_EXACT_PARALLELEPIPED(n)"	Description of the exact integration of polynomials on the parallelepiped of reference of dimension n
"IM_EXACT_PRISM(n)"	Description of the exact integration of polynomials on the prism of reference of dimension n

Even though a description of exact integration method exists on parallelepipeds or prisms, most of the time the geometric transformations on such elements are not linear and the exact integration cannot be used.

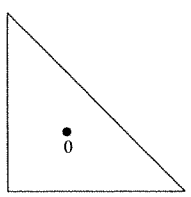
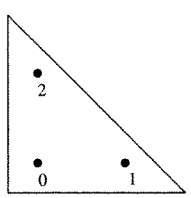
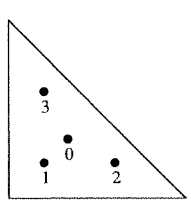
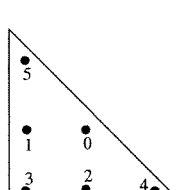
2.3 Newton cotes Integration methods

use "IM_NC(N,K)", "IM_NC_PARALLELEPIPED(N,K)" and "IM_NC_PRISM(N,K)" to have the Newton cotes integration of order K respectively on simplices, parallelepipeds and prisms.

2.4 Gauss Integration methods on dimension 1

use "IM_GAUSS1D(K)" to have the Gauss-Legendre integration on the segment of order K (with K/2 + 1 points), and "IM_GAUSSLOBATTO1D(K)" to have the Gauss-Lobatto-Legendre integration on the segment of order K (with K/2 + 1 points). The latter integration method is only available for odd values of K. The Gauss-Lobatto integration method can be used in conjunction with "FEM_PK_GAUSSLOBATTO1D(K/2)" to perform mass-lumping.

2.5 Gauss Integration methods on dimension 2

graphic	coordinates x	y	weights	function to call / order
	1/3	1/3	1/2	"IM_TRIANGLE(1)" 1 point, order 1.
	1/6 2/3 1/6	1/6 1/6 2/3	1/6 1/6 1/6	"IM_TRIANGLE(2)" 3 points, order 2.
	1/3 1/5 3/5 1/5	1/3 1/5 1/5 3/5	-27/96 25/96 25/96 25/96	"IM_TRIANGLE(3)" 4 points, order 3.
	a $1-2a$ a b $1-2b$ b	a a $1-2a$ b b $1-2b$	c c c d d d	"IM_TRIANGLE(4)" 6 points, order 4, a = 0.445948490915965, b = 0.091576213509771, c = 0.111690794839005, d = 0.054975871827661.

of the time for composite element, it is preferable to choose the basic method `IM1` with no points on the boundary (because the gradient coulb be not defined on the boundary of sub-elements).

References

[1] G. DHATT, AND G. TOUZOT *The Finite Element Method Displayed*, J. Wiley & Sons, New York, 1984. 3

[2] Y. RENARD, *Elementary Computations in* GETFEM++, 2002. 3, 18

[3] Y. RENARD, J. POMMIER, *Short User Documentation of* GETFEM++, 2003.

[4] J.-C. NEDELEC, *Notions sur les techniques d'éléments finis*, Ellipses, SMAI, Mathématiques & Applications n^{o}7, 1991.

[5] R. COOLS *An Encyclopaedia of Cubature Formulas*, J. Complexity, `http://www.cs.kuleuven.ac.be/ñines/research/ecf/ecf.html` 23

[6] P. SOLIN, K. SEGETH, I. DOLEZEL, *Higher-Order Finite Element Methods*, Chapman and Hall/CRC, Studies in advanced mathematics, 2004.

Université Ahmed ZABANA Relizane

Département de Physique

Master 1ere Année

METHODES DES ELEMENTS FINIS

Cours et exercices