Analyse numérique des équations aux dérivées partielles

Thierry Gallouët, Raphaèle Herbin

Analyse numérique des équations aux dérivées partielles

Thierry Gallouët et Raphaèle Herbin

4 août 2023

Table des matières

Introduction

Analyse numérique des équations aux dérivées partielles et calcul scientifique

Pour aborder le calcul numérique (à l'aide d'un outil informatique) des solutions d'un problème "réel", on passe par les étapes suivantes :

Description qualitative des phénomènes physiques Cette étape, effectuée par des spécialistes des phénomènes que l'on veut quantifier (ingénieurs, chimistes, biologistes etc...) consiste à répertorier tous les mécanismes qui entrent en jeu dans le problème qu'on étudie.

Modélisation Il s'agit, à partir de la description qualitative précédente, d'écrire un modèle mathématique. On supposera ici que ce modèle amène à un système d'équations aux dérivées partielles (EDP). Selon les hypothèses effectuées, la modélisation peut aboutir à plusieurs modèles, plus ou moins complexes. Dans la plupart des cas, on ne saura pas calculer une solution analytique, explicite, du modèle ; on devra faire appel à des techniques de résolution approchée.

Analyse mathématique du modèle mathématique Même si l'on ne sait pas trouver une solution explicite du modèle, il est important d'en étudier les propriétés mathématiques, dans la mesure du possible. Il est bon de se poser les questions suivantes :

- Le problème est-il bien posé, c'est-à-dire y a-t-il existence et unicité de la solution ?
- Les propriétés physiques auxquelles on s'attend sont elles satisfaites par les solutions du modèle mathématique ? Si l'inconnue est une concentration, par exemple, peut-on prouver que la solution du modèle censé représenter le modèle physique est toujours positive ?
- Y a-t-il continuité de la solution par rapport aux données ?

Discrétisation et résolution numérique Un problème posé sur un domaine continu (espace–temps) n'est pas résoluble tel quel par un ordinateur, qui ne peut traiter qu'un nombre fini d'inconnues. Pour se ramener à un problème en dimension finie, on discrétise l'espace et/ou le temps. Si le problème original est linéaire on obtient un système linéaire. Si le problème original est non linéaire (par exemple s'il s'agit de la minimisation d'une fonction) on aura un système non linéaire à résoudre par une méthode *ad hoc* (méthode de Newton...)

Analyse numérique Il s'agit maintenant de l'analyse mathématique du schéma numérique. En effet, une fois le problème discret obtenu, il est raisonnable de se demander si la solution de ce problème est proche et en quel sens, du problème continu. De même, si on doit mettre en œuvre une méthode itérative pour le traitement des non-linéarités, il faut étudier la convergence de la méthode itérative proposée.

Mise en œuvre, programmation et analyse des résultats La partie mise en œuvre est une grosse consommatrice de temps. Actuellement, de nombreux codes commerciaux de type "boîte noire" existent, qui permettent en théorie de résoudre "tous" les problèmes. Il faut cependant procéder à une analyse critique des résultats obtenus par ces codes, qui ne sont pas toujours compatibles avec les propriétés physiques attendues...

Principales méthodes de discrétisation

Méthodes de différences finies et volumes finis

On considère un domaine physique $\Omega \subset \mathbb{R}^d$, où d est la dimension de l'espace. Le principe des méthodes de différences finies (DF) consiste à se donner un certain nombre de points du domaine, qu'on notera $(x_1 \dots x_N) \subset (\mathbb{R}^d)^N$. On approche l'opérateur différentiel en espace en chacun des x_i

par des quotients différentiels. Il faut alors discrétiser la dérivée en temps : on pourra par exemple considérer un schéma d'Euler [1]. Les méthodes de volumes finis (VF) sont adaptées aux équations de conservation et utilisées en mécanique des fluides depuis plusieurs décennies. Le principe consiste à découper le domaine Ω en des *volumes de contrôle* ; on intègre ensuite l'équation de conservation sur les volumes de contrôle ; on approche alors les flux sur les bords du volume de contrôle par exemple par une technique de différences finies.

Méthodes variationnelles, méthodes d'éléments finis

On met le problème d'équations aux dérivées partielles sous la forme dite *variationnelle* :

$$\begin{cases} a(u,v) = (f,v)_H \quad \forall v \in H, \\ u \in H, \end{cases}$$

où H est un espace de Hilbert [2] bien choisi (par exemple parce qu'il y a existence et unicité de la solution dans cet espace), $(\,\cdot\,,\,\cdot\,)_H$ le produit scalaire sur H et a une forme bilinéaire sur H. Dans un tel cadre fonctionnel, la discrétisation consiste à remplacer H par un sous espace de dimension finie H_k, construit par exemple à l'aide de fonctions de base éléments finis qu'on introduira plus loin :

$$\begin{cases} a(u_k,v_k) = (f,v_k)_H \quad \forall v \in H_k, \\ u_k \in H_k, \end{cases}$$

Méthodes spectrales

L'idée de ces méthodes est de chercher une solution approchée sous forme d'un développement sur une certaine famille de fonctions. On peut par exemple écrire la solution approchée sous la forme $u = \sum_{i=1}^{n} \alpha_i(u) p_i$, où les p_i sont des fonctions polynomiales. On choisit la base p_i de manière à ce que les dérivées de α_i et p_i soient faciles à calculer. Ces dernières méthodes sont réputées coûteuses, mais précises. Elles sont d'ailleurs plus souvent utilisées comme aide à la compréhension des phénomènes physiques sur des problèmes modèles que dans des applications industrielles.

Quelques exemples d'équations aux dérivées partielles

Équation de Poisson [3] Soit d la dimension de l'espace (en général, $d = 1, 2$ ou 3), l'équation de Poisson s'écrit :

$$-\operatorname{div}(\kappa \nabla u) = f,$$

où :

— le symbole div représente l'opérateur divergence, qu'on définit de la manière suivante : pour $\boldsymbol{w} : \boldsymbol{x} = (x_1, \ldots, x_d)^t \in \mathbb{R}^d \mapsto \boldsymbol{w}(x) = (w_1(x_1, \ldots, x_d), \ldots, w_d(x_1, \ldots, x_d) \in \mathbb{R}^3$, on définit

$$\operatorname{div} \boldsymbol{w} = \sum_{i=1}^{d} \partial_i w_i, \tag{1}$$

la notation ∂_i désignant la dérivée partielle par rapport à x_i ;

1. Leonhard Paul Euler, né le 15 avril 1707 à Bâle et mort le 18 septembre 1783 à Saint-Pétersbourg, est un mathématicien et physicien suisse, qui passa la plus grande partie de sa vie en Russie et en Allemagne. Euler fit d'importantes découvertes dans des domaines aussi variés que le calcul infinitésimal et la théorie des graphes. Il introduisit également une grande partie de la terminologie et de la notation des mathématiques modernes, en particulier pour l'analyse mathématique, comme pour la notion d'une fonction mathématique. Il est également connu pour ses travaux en mécanique, en dynamique des fluides, en optique et en astronomie.

2. David Hilbert (1862–1943) est un très grand mathématicien allemand du XX–ème siècle. Il est en particulier connu pour les vingt-trois problèmes qu'il a énoncés comme défis aux mathématiciens. Certains de ces problèmes sont à ce jour non résolus. Un *espace de Hilbert* H ou espace hilbertien est un espace vectoriel normé complet dont la norme, notée $\|\cdot\|$, découle d'un produit scalaire $(\,\cdot\,,\,\cdot\,)_H$: pour tout $u \in H$, $\|u\|_H^2 = (u,u)_H$.

— le symbole nabla ∇ (parfois aussi appelé del) représente le gradient, défini, pour une fonction scalaire $u : \mathbb{R}^d \to \mathbb{R}$, par :

$$\nabla u = \begin{pmatrix} \partial_1 u \\ \vdots \\ \partial_d u \end{pmatrix} \tag{2}$$

En une dimension d'espace, l'équation de Poisson s'écrit donc :

$$(-(\kappa u')' = f$$

Le réel κ est un coefficient de diffusion et le terme $-\kappa \nabla u$ est un *flux de diffusion*. Il peut s'agir de la diffusion d'une espèce chimique, dans ce cas, κ est le coefficient de diffusion (souvent noté D, en $\mathrm{m/s^2}$) de la loi de Fick qui donne le flux de diffusion J (quantité de matière par unité de surface et de temps) en fonction de la concentration u :

$$J = -\kappa u' \text{ en dimension 1 qui devient } J = -\kappa \nabla u \text{ en dimension supérieure}$$

Il peut s'agir d'une diffusion thermique, et on parle dans ce cas plus souvent de conduction. Le coefficient de diffusion est souvent noté λ et on l'appelle la *conductivité thermique* (en $\mathrm{W\ m^{-1}\ K^{-1}}$). La loi de Fourier (1807) donne la densité de flux de chaleur j (en $\mathrm{W\ m^{-2}}$) en fonction de la température u (en Kelvin) :

$$J = -\lambda u' \text{ en dimension 1 qui devient } J = -\lambda \nabla u \text{ en dimension supérieure}$$

Dans le cas d'une conduction électrique, le coefficient de diffusion est dans ce cas noté σ et on l'appelle la conductivité électrique (en $\mathrm{W\ m^{-1}\ K^{-1}}$). La loi d'Ohm donne la densité de courant électrique j (en $\mathrm{m/s^2}$) en fonction du potentiel électrique u :

$$J = -\sigma u' \text{ en dimension 1 qui devient } j = -\sigma \nabla u \text{ en dimension supérieure}$$

Avec un coefficient constant $\kappa = 1$, l'équation de Poisson s'écrit en une dimension d'espace $-u'' = f$. En deux dimensions d'espace, elle s'écrit $-\Delta u = f$ avec $\Delta u = \partial^2_{xx} u + \partial^2_{yy} u$, où $\partial^2_{xx} u$ (respectivement $\partial^2_{yy} u$) désigne la dérivée partielle seconde de u par rapport à x (respectivement y). Dans le cas $f = 0$, on obtient l'équation de Laplace $-\Delta u = 0$.

Équation de la chaleur Elle s'écrit :

$$\partial_t u - \Delta y = 0$$

où $\partial_t u$ désigne la dérivée partielle de u par rapport au temps t : la fonction u est ici une fonction du temps et de l'espace. C'est la version “instationnaire" de l'équation de Laplace.

Équation de transport En une dimension d'espace, elle s'écrit :

$$\partial_t u + c\partial_x u = 0$$

où c est un réel (la vitesse de transport) et $\partial_x u$ désigne la dérivée partielle de u par rapport à la variable d'espace x. Si on se donne comme condition initiale $u(x, 0) = u_0(x)$, la solution de l'équation au temps t est $u(x, t) = u_0(x - ct)$ (ceci est facile à vérifier au moins dans le cas régulier). En dimension supérieure, cette équation devient :

$$\partial_t u + c\nabla u = 0$$

Équation des ondes Elle s'écrit :

$$u_{tt} - \Delta u = 0$$

Classification Considérons maintenant une équation aux dérivées partielles linéaire, de degré 2, de la forme :

$$A\partial^2_{xx} u + B\partial^2_{xy} u + C\partial^2_{yy} u = 0$$

L'appellation *elliptique*, *parabolique* ou *hyperbolique* d'une équation aux dérivées partielles de cette forme correspond à la nature de la conique décrite par l'équation caractéristique correspondante, c'est-à-dire :

$$Ax^2 + Bxy + Cy^2 = 0.$$

— Si $B^2 - 4AC < 0$, l'équation est dite elliptique ;

— Si $B^2 - 4AC = 0$, elle est dite parabolique ;

— si $B^2 - 4AC > 0$, elle est dite hyperbolique.

Vérifiez que l'équation de Laplace est elliptique alors que l'équation des ondes est hyperbolique.

Équations non linéaires Notons que tous les exemples que nous avons présentés sont des équations aux dérivées partielles linéaires, c'est-à-dire des équations qui ne font pas intervenir que des termes linéaires en u et ses dérivées.

L'appellation *elliptique*, *parabolique* ou *hyperbolique* s'étend à des équations qui ne sont pas forcément linéaires ni de degré 2. On dira par exemple que l'équation du p-laplacien, qui s'écrit $\mathrm{div}(|\nabla u|^{p-2}\nabla u) = 0$, avec $p \geqslant 2$ est une équation elliptique non linéaire (c'est l'équation de Laplace dans le cas $p = 2$).

Bien sûr, de nombreux modèles comportent des équations non linéaires comme par exemple l'équation hyperbolique non linéaire de Bürgers qui s'écrit :

$$\partial_t u + (u^2)_x = 0 \quad t \in \mathbb{R}+ \quad x \in \mathbb{R}$$

avec la condition initiale $u(x, 0) = u_0(x)$. Une telle équation est dite *hyperbolique non linéaire*.

1

Différences finies et volumes finis pour les problèmes de diffusion stationnaires

1.1 Principe des deux méthodes

1.1.1 Cas de la dimension 1

On considère le problème unidimensionnel

$$\begin{cases} -u''(x) = f(x) \quad \forall x \in]0\,;1[& (1.1) \\ u(0) = u(1) = 0 & (1.2) \end{cases}$$

où $f \in \mathcal{C}([0\,;1])$. Les conditions aux limites (1.2) considérées ici sont dites de type Dirichlet [1] homogène (le terme homogène désigne les conditions nulles). Cette équation modélise par exemple la diffusion de la chaleur dans un barreau conducteur chauffé (terme source f) dont les deux extrémités sont plongées dans de la glace.

Méthode de différences finies

Sur la figure 1.1, on se donne une subdivision de $[0\,;1]$, c'est-à-dire une suite de points $(x_k)_{k=0,\ldots,N+1}$ tels que $0 = x_0 < x_1 < x_2 < \ldots < x_N < x_{N+1} = 1$. Pour $i = 0, \ldots, N$, on

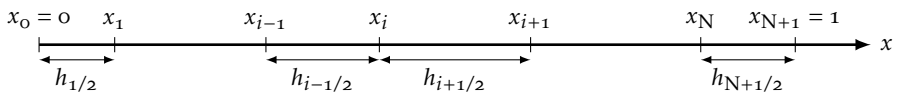


FIGURE 1.1 – Maillage unidimensionnel en différences finies

note $h_{i+1/2} = x_{i+1} - x_i$ et on définit le *pas* du maillage par :

$$h = \max_{i=0,\ldots,N} h_{i+1/2} \tag{1.3}$$

Pour simplifier l'exposé, on se limitera dans un premier temps à un pas constant :

$$h_{i+1/2} = h \qquad \forall i = 0, \ldots, N.$$

Le principe de la méthode des différences finies consiste à écrire l'équation aux dérivées partielles (1.1) aux points de discrétisation x_i :

$$-u''(x_i) = f(x_i) \qquad \forall i = 1, \ldots, N$$

1. Johann Peter Gustav Dirichlet, mathématicien allemand, né à Düren en 1805 et mort à Göttingen en 1859. Il a effectué ses études supérieures à Paris où il a cotoyé les plus grands mathématiciens français de l'époque, dont Legendre, Laplace, Poisson et Fourier. Il retourne ensuite en 1825 en Allemagne où il travaille en particulier avec son ami Jacobi et avec Gauss, dont il reprendra en la chaire à l'Université de Göttingen. Il eut entre autres comme élève Riemann et Kronecker (qui a donné son nom au fameux symbole). Les travaux de Dirichlet ont surtout porté sur les séries de Fourier et l'arithmétique.

puis à approcher l'opérateur différentiel (ici $-u''$) par un quotient différentiel, de manière à en déduire un système d'équations en fonction d'inconnues discrètes censées représenter des approximations de u aux points de discrétisation. Voici comment on procède pour l'équation de Poisson unidimensionnelle. Effectuons d'abord un développement de Taylor en x_i, en supposant que $u \in \mathcal{C}^4([0\,;1])$:

$$u(x_{i+1}) = u(x_i) + hu'(x_i) + \frac{h^2}{2}u''(x_i) + \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\zeta_i) \tag{1.4}$$

$$u(x_{i-1}) = u(x_i) - hu'(x_i) + \frac{h^2}{2}u''(x_i) - \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\eta_i) \tag{1.5}$$

avec $\zeta_i \in [x_i\,;x_{i+1}]$ et $\eta_i \in [x_{i-1}\,;x_i]$. En additionnant, on obtient :

$$u(x_{i+1}) + u(x_{i-1}) = 2u(x_i) + h^2u''(x_i) + \mathcal{O}(h^2)$$

Il semble donc raisonnable d'approcher la dérivée seconde $-u''(x_i)$ par le *quotient différentiel* :

$$\frac{2u(x_i) - u(x_{i-1}) - u(x_{i+1})}{h^2}$$

Sous des hypothèses de régularité sur u, on peut montrer [lemme 1.14] que cette approximation est d'ordre 2 au sens :

$$R_i = u''(x_i) + \frac{2u(x_i) - u(x_{i-1}) - u(x_{i+1})}{h^2} = \mathcal{O}(h^2)$$

On appelle *erreur de consistance* au point x_i la quantité R_i. L'approximation de $u''(x_i)$ par un quotient différentiel suggère de considérer les équations discrètes suivantes :

$$\frac{2u_i - u_{i-1} - u_{i+1}}{h^2} = f(x_i) \qquad i = 1, \ldots, N$$

dont les inconnues discrètes sont les u_i, $i = 1, \ldots, N$. Notons que la première équation fait intervenir u_0 tandis que la dernière fait intervenir u_{N+1}. Ces valeurs ne sont pas à proprement parler des inconnues, puisqu'elles sont données par les conditions aux limites (1.2). On pose donc $u_0 = 0$ et $u_{N+1} = 0$. Le système complet d'équations s'écrit donc

$$\begin{cases} \dfrac{2u_i - u_{i-1} - u_{i+1}}{h^2} = f(x_i) \qquad i = 1, \ldots, N & \text{(1.6a)} \\ u_0 = 0 & \text{(1.6b)} \\ u_{N+1} = 0 & \text{(1.6c)} \end{cases}$$

Remarque 1.1 — Inconnues discrètes et solution exacte. Attention à ne pas confondre u_i et $u(x_i)$: les équations discrètes (1.6) font intervenir les *inconnues* discrètes u_i, $i = 1, \ldots, N$ et non pas les *valeurs* $u(x_i)$, $i = 1, \ldots, N$ de la solution exacte. En général, ces valeurs ne sont pas les mêmes. Si la discrétisation a été effectuée correctement (comme c'est le cas ici et comme nous le démontrerons mathématiquement plus loin), la résolution du système discret nous permettra d'obtenir des valeurs u_i, $i = 1, \ldots, N$ des inconnues discrètes qui seront des bonnes approximations des valeurs $u(x_i)$, $i = 1, \ldots, N$ de la solution exacte que nous ne pouvons pas, dans le cas général, calculer explicitement.

Méthode des volumes finis

Sur la figure 1.2, on se donne non plus des points mais des volumes de contrôle K_i, $i = 1, \ldots, N$ avec $K_i =]x_{i-1/2}\,;x_{i+1/2}[$ et on note $h_i = x_{i+1/2} - x_{i-1/2}$. Pour chaque volume de contrôle K_i, on se donne un point $x_i \in K_i$. On pourra considérer par exemple (mais ce n'est pas le seul point possible) $x_i = 1/2(x_{i+1/2} + x_{i-1/2})$. On intègre l'équation $-u'' = f$ sur K_i :

$$\int_{x_{i-1/2}}^{x_{i+1/2}} -u''(x)\,\mathrm{d}x = \int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,\mathrm{d}x$$

et on pose :

$$f_i = \frac{1}{h_i}\int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,\mathrm{d}x$$

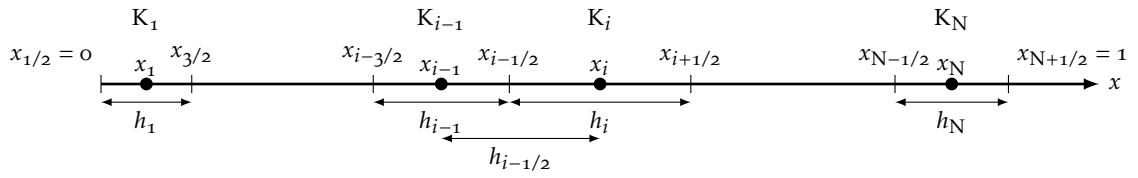


Figure 1.2 – Maillage unidimensionnel en volumes finis

On obtient :

$$-u'(x_{i+1/2}) + u'(x_{i-1/2}) = h_i f_i \qquad i = 1, \dots, N \tag{1.7}$$

Cette équation est un bilan de flux. La quantité $\overline{F}_{i+1/2} = -u'(x_{i+1/2})$ est le flux de diffusion en $x_{i+1/2}$. Pour la première maille ($i = 1$), on obtient plus particulièrement :

$$-u'(x_{3/2}) + u'(0) = h_1 f_1 \tag{1.8}$$

et pour la dernière ($i = N$) :

$$-u'(1) + u'(x_{N-1/2}) = h_N f_N \tag{1.9}$$

On cherche donc à approcher les flux $-u'(x_{i+1/2})$ aux interfaces $x_{i+1/2}$ des mailles et les flux $u'(0)$ et $u'(1)$ au bord. Notons que l'opérateur à approcher est ici d'ordre 1, alors qu'il était d'ordre 2 en différences finies pour la même équation. On se donne une inconnue par maille (ou volume de contrôle i) que l'on note u_i et on espère approcher ainsi la valeur $u(x_i)$ (ou $\frac{1}{h_i}\int_{K_i} u$). En supposant u suffisamment régulière, on peut effectuer deux développements de Taylor à l'ordre 2 de u entre x_{i+1} et $x_{i+1/2}$ et entre x_i et $x_{i+1/2}$; en soustrayant ces développements de Taylor l'un de l'autre, on se rend compte qu'il est "raisonnable" [exercice 13] d'approcher le terme $u'(x_{i+1/2})$ dans l'équation (1.7) par le quotient différentiel :

$$\frac{u(x_{i+1}) - u(x_i)}{h_{i+1/2}}, \text{ avec } h_{i+1/2} = \frac{1}{2}(h_i + h_{i+1}),$$

au sens où l'erreur de consistance sur les flux, définie par :

$$R_{i+1/2} = u'(x_{i+1/2}) - \frac{u(x_{i+1}) - u(x_i)}{h_{i+1/2}}$$

est d'ordre 1 si $u \in \mathcal{C}^2([0\,;1], \mathbb{R})$ [lemme 1.28]. Le schéma numérique s'écrit donc :

$$-\frac{u_{i+1} - u_i}{h_{i+1/2}} + \frac{u_i - u_{i-1}}{h_{i-1/2}} = h_i f_i \qquad i = 2, \dots, N-1 \tag{1.10}$$

Pour les première et N–ème équations, on tient compte des conditions aux limites de Dirichlet homogènes (1.2) et on approche $u'(0)$ dans l'équation (1.8) (respectivement $u'(1)$ dans l'équation (1.9)) par $(u(x_1) - 0)/h_{1/2}$ (respectivement $(0 - u(x_N))/h_{N+1/2}$), ce qui donne comme première et dernière équations du schéma numérique :

$$\begin{cases} -\dfrac{u_2 - u_1}{h_{3/2}} + \dfrac{u_1}{h_{1/2}} = h_1 f_1 & (1.11) \\ \dfrac{u_N}{h_{N+1/2}} + \dfrac{u_N - u_{N-1}}{h_{N-1/2}} = h_N f_N & (1.12) \end{cases}$$

Là encore, comme dans le cas des différences finies, il faut faire attention parce que les équations discrètes (1.10)-(1.12) font intervenir les *inconnues* discrètes u_i, $i = 1, \dots, N$ et non pas les *valeurs* $u(x_i)$, $i = 1, \dots, N$ de la solution exacte. En général, ces valeurs ne sont pas les mêmes.

Remarque 1.2 — Comparaison des deux schémas. Si le pas du maillage est constant $h_i = h$, $\forall i = 1, \dots, N$ (on dit aussi que le maillage est uniforme), on peut montrer [exercice 1] que les équations des schémas volumes finis et différences finies coïncident aux conditions de bord et au second membre près. Si le maillage n'est pas uniforme, les deux schémas diffèrent.

Autres conditions limites

Conditions de Dirichlet non homogènes Supposons que les conditions aux limites en 0 et en 1 soit maintenant de type Dirichlet non homogènes, c'est-à-dire :

$$\begin{aligned} u(0) &= a \\ u(1) &= b \end{aligned} \tag{1.13}$$

avec a et b pas forcément nuls. Dans ce cas :

1. les équations discrètes du *schéma aux différences finies* (1.6) restent identiques mais les valeurs u_0 et u_{N+1} sont maintenant données par $u_0 = a$ et $u_{N+1} = b$;
2. les équations discrètes du *schéma de volumes finis* (1.10) associés aux nœuds internes restent identiques mais les valeurs $u'(0)$ et $u'(1)$ sont maintenant approchées par $(u(x_1) - a)/h_{1/2}$ et $(b - u(x_N))/h_{N+1/2}$, ce qui donne comme première et dernière équations du schéma numérique :

$$\begin{cases} -\dfrac{u_2 - u_1}{h_{3/2}} + \dfrac{u_1 - a}{h_{1/2}} = h_1 f_1, & (1.14) \\ -\dfrac{b - u_N}{h_{N+1/2}} + \dfrac{u_N - u_{N-1}}{h_{N-1/2}} = h_N f_N. & (1.15) \end{cases}$$

Conditions de Neumann et Fourier On appelle *condition de Neumann*[2] une condition qui impose une valeur de la dérivée, par exemple :

$$u'(0) = a \tag{1.16}$$

On appelle *condition de Fourier*[3] ou *condition de Robin*[4] une condition [11] qui impose une relation entre la valeur de la dérivée et la valeur de la solution, par exemple,

$$u'(1) + \alpha u(1) = b \tag{1.17}$$

avec $\alpha > 0$. Cette condition est donc un mélange des conditions de Dirichlet et de Neumann et est souvent utilisée pour exprimer une condition de transfert, thermique par exemple, entre un milieu et l'extérieur.

Enfin, on dit que les conditions aux limites sont *mixtes* si elles sont de type différent sur des portions de frontière du domaine : on a des conditions mixtes dans le cas unidimensionnel si, par exemple, on a une condition de Dirichlet en 0 et une condition de Neumann en 1.

Prenons par exemple le cas de conditions mixtes, en considérant l'équation (1.1) en $x = 0$ avec les conditions (1.16) et (1.17) en $x = 1$. Voyons comment tenir compte de ces nouvelles conditions limites avec les méthodes différences finies et volumes finis :

Schéma aux différences finies, maillage uniforme Pour approcher la condition de Neumann en 0, on effectue un développement de Taylor à l'ordre 1 en 0 :

$$u'(0) = \frac{u(x_1) - u(0)}{h} + \varepsilon(h) = a$$

Ceci suggère l'équation discrète suivante pour u_0 en écrivant que :

$$\frac{u_1 - u_0}{h} = a \;\Leftrightarrow\; u_0 = u_1 - ah$$

2. Karl Gottried Neumann est un mathématicien allemand, né en 1832 à Königsberg et mort 1925 à Leipzig. Il fut l'un des pionniers de la théorie des équations intégrales. Il a laissé son nom aux conditions aux limites mentionnées ici.

3. Joseph Fourier est un mathématicien et physicien français, né en 1768 à Auxerre et mort en 1830 à Paris. Il est connu pour ses travaux sur la décomposition de fonctions périodiques en séries trigonométriques convergentes appelées séries de Fourier et leur application au problème de la propagation de la chaleur. Il a participé à la révolution, a échappé de peu à la guillotine et a été nommé préfet de l'Isère par Napoléon. Il a fait construire la route entre Grenoble et Briançon et fondé en 1810 l'Université Royale de Grenoble, dont il fut le recteur. L'université scientifique de Grenoble et l'un des laboratoires de mathématiques de cette université portent son nom.

4. Victor Gustave Robin est un mathématicien français né en 1855 et mort en 1897 qui a travaillé en particulier en thermodynamique et sur la théorie du potentiel.

Rappelons que pour la condition de Dirichlet homogène, la valeur de u_0 était simplement prise comme $u_0 = 0$. De la même manière, on écrit un développement limité pour la dérivée dans la condition de Fourier (1.17) :

$$\frac{u(1) - u(x_N)}{h} + \varepsilon(h) + \alpha u(1) = b$$

ce qui suggère l'approximation suivante :

$$\frac{u_{N+1} - u_N}{h} + \alpha u_{N+1} = b \;\Leftrightarrow\; u_{N+1} = \frac{u_N + bh}{1 + \alpha h}$$

Schéma de volumes finis La condition de Neumann est particulièrement simple à prendre en compte, puisque le schéma de volumes finis fait intervenir l'approximation du flux $u'(0)$ dans l'équation (1.8), que l'on discrétise donc par :

$$a + \frac{u_1 - u_2}{h_{3/2}} = h_1 f_1 \tag{1.18}$$

On tient compte ensuite de la condition de Fourier (1.17) pour approcher le terme $u'(1)$ dans l'équation (1.8) : on peut par exemple [5] approcher $u'(1)$ par $b - \alpha u_N$ ce qui nous donne comme N–ème équation discrète :

$$F_{N+1/2} - F_{N-1/2} = h_N f_N \quad \text{avec} \quad F_{N+1/2} = \alpha u_N - b \quad \text{et} \quad F_{N-1/2} = -\frac{u_N - u_{N-1}}{h_{N-1/2}} \tag{1.19}$$

1.1.2 Cas de la dimension 2 ou 3

On considère maintenant le problème de Laplace en dimension 2 ou 3 sur un ouvert borné Ω de $\mathbb{R}^d$, $d = 2$ ou 3, avec conditions aux limites de Dirichlet homogènes :

$$\begin{cases} -\Delta u = f \;\text{ sur }\; \Omega & \text{(1.20a)} \\ u(x) = 0 \;\text{ sur }\; \partial\Omega & \text{(1.20b)} \end{cases}$$

où f est une fonction de Ω dans $\mathbb{R}$.

Méthode de différences finies

Supposons pour simplifier que le domaine Ω soit un carré (c'est-à-dire $d = 2$, le cas rectangulaire se traite tout aussi facilement). On se donne un pas de maillage constant h et des points $x_{i,j} = (ih, jh)$, $i = 1, \ldots, N$, $j = 1, \ldots, N$. En effectuant les développements limités de Taylor comme au paragraphe 1.1.1 dans les deux directions [exercice 21], on approche $-\partial_i^2 u(x_{i,j})$ (respectivement $-\partial_j^2 u(x_{i,j})$) par :

$$\frac{2u(x_{i,j}) - u(x_{i+1,j}) - u(x_{i-1,j})}{h^2} \quad \left(\text{respectivement par } \frac{2u(x_{i,j}) - u(x_{i,j+1}) - u(x_{i,j-1})}{h^2}\right)$$

Ce type d'approche est limité à des géométries simples. Pour mailler des géométries compliquées, on utilise souvent des triangles (tétraèdres en dimension 3), auquel cas la méthode des différences finies est plus difficile à généraliser car on ne peut pas approcher la dérivée seconde comme en maillages cartésiens.

Méthode de volumes finis

On suppose maintenant que Ω est un ouvert polygonal de $\mathbb{R}^2$ et on se donne un maillage $\mathcal{T}$ de Ω, c'est-à-dire, en gros, un découpage de Ω en volumes de contrôle polygonaux K. En intégrant l'équation (1.20a) sur K, on obtient :

$$\int_K -\Delta u \, \mathrm{d}x = \int_K f \, \mathrm{d}x$$

5. Ce n'est pas la seule possibilité [exercice 12].

Par la formule de Stokes, on peut réécrire cette équation :

$$-\int_{\partial K} \nabla u(x) \cdot \boldsymbol{n}_K(x)\ \mathrm{d}\gamma(x) = \int_K f(x)\,\mathrm{d}x$$

où $\mathrm{d}\gamma(x)$ désigne l'intégrale par rapport à la mesure uni-dimensionnelle sur le bord de l'ouvert Ω et où $\boldsymbol{n}_K$ désigne le vecteur normal unitaire à ∂K extérieur à K. Comme K est polygonal, on peut décomposer ∂K en arêtes σ qui sont des segments de droite et en appelant $\mathcal{E}_K$ l'ensemble des arêtes de ∂K (trois arêtes dans le cas d'un triangle), on a :

$$-\sum_{\sigma\in\mathcal{E}_K}\int_\sigma \nabla u \cdot \boldsymbol{n}_{K,\sigma}\mathrm{d}\gamma(x) = \int_K f(x)\,\mathrm{d}x$$

où $\boldsymbol{n}_{K,\sigma}$ désigne le vecteur normal unitaire à σ extérieur à K (noter que ce vecteur est constant sur σ). On cherche donc maintenant à approcher la dérivée normale $\nabla u \cdot \boldsymbol{n}_{K,\sigma}$ de manière consistante sur chaque arête σ. On se donne des inconnues discrètes notées $(u_K)_{K\in\mathcal{T}}$ qui, on l'espère, vont s'avérer être des approximations de $u(x_K)$. Pour une arête $\sigma = K|L$ séparant les volumes de contrôle K et L, il est tentant d'approcher la dérivée normale $\nabla u \cdot \boldsymbol{n}_{K,\sigma}$ par le quotient différentiel :

$$\frac{u(x_L) - u(x_K)}{d_{K,L}}$$

où $d_{K,L}$ est la distance entre les points x_K et x_L. Cependant, cette approximation ne pourra être justifiée que si la direction du vecteur défini par les deux points x_K et x_L est la même que celle de la normale $\boldsymbol{n}_{K,\sigma}$, c'est-à-dire si le segment de droite $x_K x_L$ est orthogonal à l'arête $K|L$. Pour un maillage triangulaire à angles strictement inférieurs à $\pi/2$, ceci est facile à obtenir en choisissant les points x_K comme intersection des médiatrices du triangle K [6] à l'image de la figure 1.3. Supposons

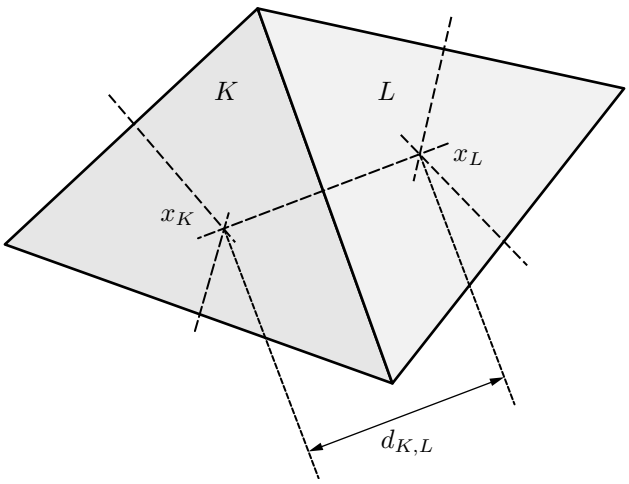


Figure 1.3 – Exemple de volumes de contrôle pour la méthode des volumes finis en deux dimensions d'espace

que cette hypothèse, dite d'orthogonalité du maillage, soit satisfaite ; on approche donc $\nabla u \cdot \boldsymbol{n}_K|_\sigma$ par $\dfrac{u(x_L) - u(x_K)}{d_{K,L}}$ et en notant $|\sigma|$ la longueur de l'arête σ, on approche :

$$\int_\sigma \nabla u \cdot \boldsymbol{n}_K \mathrm{d}\gamma \ \text{ par } \ F_{K,\sigma} = |\sigma|\frac{u_L - u_K}{d_{K,L}}$$

pour tout $\sigma \in \mathcal{E}_K$ et pour tout $K \in \mathcal{T}$. Le schéma volumes finis s'écrit donc :

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} = |K| f_K \quad \text{où} \quad f_K = \frac{1}{|K|}\int_K f(x)\,\mathrm{d}x \tag{1.21}$$

6. Les médiatrices d'un triangle se coupent en un point qui est le centre du cercle circonscrit au triangle, alors que les médianes se coupent au barycentre, qui est le centre du cercle inscrit dans le triangle ; ces deux points coïncident dans le cas d'un triangle équilatéral.

$|K|$ étant la mesure de K et où les flux numériques $F_{K,\sigma}$ sont définis (en tenant compte des conditions limites pour les arêtes du bord) par :

$$F_{K,\sigma} = \begin{cases} -|\sigma|\dfrac{u_L - u_K}{d_{K,L}} & \text{si } \sigma = K|L \\ -|\sigma|\dfrac{u_K}{d_{K,\sigma}} & \text{si } \sigma \subset \partial\Omega \text{ et } \sigma \in \mathcal{E}_K \end{cases}$$

où $d_{K,\sigma}$ est la distance entre le point x_K et l'arête σ.

Comparaison des méthodes

Cette introduction aux différences finies et volumes finis nous permet de remarquer que les différences finies sont particulièrement bien adaptées dans le cas de domaines rectangulaires ou parallèlepipédiques, pour lesquels on peut facilement définir des maillages structurés (cartésiens dans le cas présent) c'est-à-dire dont on peut indexer les mailles par un ordre (i, j) naturel.

Dans le cas de domaines plus complexes, on maille souvent à l'aide de triangles (ou tétraèdres) et dans ce cas la méthode des différences finies ne se généralise pas facilement. On a alors recours soit aux volumes finis, dont on vient de donner le principe, soit aux éléments finis, que nous aborderons ultérieurement.

1.1.3 Questions d'analyse numérique

Voici un certain nombre de questions du domaine de l'analyse numérique, auxquelles nous tenterons de répondre dans la suite :

1. Le problème obtenu en dimension finie (avec des inconnues localisées aux nœuds du maillage dans le cas de la méthode des différences finies et dans les mailles dans le cas de la méthode des volumes finis) admet-il une (unique) solution ?
2. La solution du problème discret satisfait-elle les propriétés physiques qui sont vérifiées par la solution du modèle mathématique ?
3. La solution du problème discret converge-t-elle vers la solution du problème continu lorsque le pas du maillage h tend vers 0 ? Dans le cas des différences finies en une dimension d'espace, le pas du maillage est défini par :

$$h = \sup_{i=1,\ldots,N} |x_{i+1} - x_i| \tag{1.22}$$

Dans le cas des volumes finis en une dimension d'espace, il est défini par :

$$h = \sup_{i=1,\ldots,N} |x_{i+1/2} - x_{i-1/2}|. \tag{1.23}$$

en deux dimensions d'espace, le pas h est défini par :

$$h = \sup_{K\in\mathcal{T}} \operatorname{diam}(K) \quad \text{avec} \quad \operatorname{diam}(K) = \sup_{x,y\in K} \mathrm{d}(x, y)$$

où le maillage $\mathcal{T}$ est l'ensemble des volumes de contrôle K. Notons que la réponse à cette question de convergence n'est pas évidente *a priori*. La solution discrète peut converger vers la solution continue, elle peut aussi converger vers autre chose que la solution du problème continu et, enfin, elle pourrait ne pas converger du tout.

1.2 Analyse de la méthode des différences finies

On cherche à discrétiser le problème aux limites suivant :

$$\begin{cases} -u''(x) + c(x)u(x) = f(x) \quad 0 < x < 1 \\ u(0) = 0 \\ u(1) = 0 \end{cases} \tag{1.24}$$

où $c \in \mathcal{C}([0\,;1],\mathbb{R}_+)$ et $f \in \mathcal{C}([0\,;1],\mathbb{R})$, qui peut modéliser par exemple un phénomène de diffusion-réaction d'une espèce chimique. On se donne un pas du maillage constant $h = 1/(N+1)$ et une subdivision de $]0\,;1[$, notée $(x_k)_{k=0,\ldots,N+1}$ avec $x_0 = 0 < x_1 < x_2 < \ldots < x_N < x_{N+1} = 1$. Soit u_i l'inconnue discrète associée au nœud i, $i = 1,\ldots,N$. On pose $u_0 = u_{N+1} = 0$. On obtient les équations discrètes en approchant $u''(x_i)$ par quotient différentiel par développement de Taylor, comme au paragraphe 1.1.1. On obtient le système suivant :

$$\begin{cases} \dfrac{2u_i - u_{i-1} - u_{i+1}}{h^2} + c_i u_i = f_i \quad i = 1,\ldots,N \\ u_0 = 0 \\ u_{N+1} = 0 \end{cases} \tag{1.25}$$

avec $c_i = c(x_i)$ et $f_i = f(x_i)$. On peut écrire ces équations sous forme matricielle :

$$A_h U_h = b_h \text{ avec}$$
$$A_h = \frac{1}{h^2}\begin{bmatrix} 2+c_1h^2 & -1 & 0 & \ldots & 0 \\ -1 & 2+c_2h^2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2+c_{N-1}h^2 & -1 \\ 0 & \ldots & 0 & -1 & 2+c_Nh^2 \end{bmatrix}, U_h = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix} \text{ et } b_h = \begin{bmatrix} f_1 \\ \vdots \\ f_N \end{bmatrix} \tag{1.26}$$

Remarque 1.3 — Notations pour les vecteurs et matrices. Un vecteur : $\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}$ sera aussi noté, par souci de simplicité, $\boldsymbol{u} = (u_1,\ldots,u_N)$ (ces deux égalités signifiant que les composantes de u dans la base canonique de $\mathbb{R}^N$ sont u_1, u_2, …, u_N. Attention toutefois à ne pas confondre cette notation avec $\boldsymbol{u}^t = (u_1 \ \ldots \ u_N)$, qui est une matrice $1 \times N$; c'est la matrice transposée de $\boldsymbol{u}$ vu comme une matrice $N \times 1$. On peut écrire le produit scalaire de deux vecteurs $\boldsymbol{u}$ et $\mathbf{v}$ de $\mathbb{R}^N$ avec ces notations :

$$\boldsymbol{u} \cdot \mathbf{v} = \sum_{i=1}^{N} u_i v_i = \boldsymbol{u}^t \mathbf{v} = \mathbf{v}^t \boldsymbol{u}.$$

Les questions suivantes surgissent alors naturellement :

1. Le système (1.26) admet-il une unique solution ?
2. A-t-on convergence de U_h vers u et en quel sens ?

Nous allons répondre par l'affirmative à ces deux questions. Commençons par la première.

Proposition 1.4 Soit $c = (c_1,\ldots,c_N) \in \mathbb{R}^N$ tel que $c_i \geqslant 0$ pour $i = 1,\ldots,N$; alors la matrice A_h définie dans (1.26) est symétrique définie positive et donc inversible.

Démonstration. La matrice A_h est évidemment symétrique. Montrons qu'elle est définie positive. Soit $\mathbf{v} = (v_1 \ldots v_N)$, on pose $v_0 = v_{N+1} = 0$. Calculons le produit scalaire $A_h\mathbf{v} \cdot \mathbf{v} = \mathbf{v}^t A_h \mathbf{v}$. On a :

$$A_h v \cdot v = \frac{1}{h^2}\begin{bmatrix} v_1 & v_2 & \cdots & v_N \end{bmatrix}\begin{bmatrix} 2+c_1h^2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & & -1 \\ 0 & & -1 & 2+c_Nh^2 \end{bmatrix}\begin{bmatrix} v_1 \\ \vdots \\ \vdots \\ v_N \end{bmatrix}$$

c'est-à-dire :

$$\begin{aligned} A_h\mathbf{v} \cdot \mathbf{v} &= \frac{1}{h^2}\sum_{i=1}^{N} v_i\left(-v_{i-1} + (2+c_ih^2)v_i - v_{i+1}\right) \\ &= \sum_{i=1}^{N} c_i v_i^2 + \frac{1}{h^2}\sum_{i=1}^{N} v_i\left(-v_{i-1} + v_i + v_i - v_{i+1}\right) \end{aligned}$$

On a donc, par changement d'indice, et puisqu'on a posé $v_0 = 0$ et $v_{N+1} = 0$:

$$\begin{aligned} A_h\mathbf{v}\cdot\mathbf{v} &= \sum_{i=1}^{N} c_i v_i^2 + \frac{1}{h^2}\sum_{i=1}^{N} v_i\,(v_i - v_{i-1}) - \sum_{j=2}^{N} v_{j-1}\,(v_j - v_{j-1}) \\ &= \sum_{i=1}^{N} c_i v_i^2 + \frac{1}{h^2}\sum_{i=1}^{N} (v_i - v_{i-1})^2 + v_N^2 \\ &= \sum_{i=1}^{N} c_i v_i^2 + \frac{1}{h^2}\sum_{i=1}^{N+1} (v_i - v_{i-1})^2 \\ &\geqslant 0 \qquad \forall \mathbf{v} = (v_1,\dots,v_N) \in \mathbb{R}^N \end{aligned}$$

Si on suppose $A_h\mathbf{v}\cdot\mathbf{v} = 0$, on a alors :

$$\sum_{i=1}^{N} c_i h^2 v_i^2 = 0 \text{ et } v_i - v_{i-1} = 0 \qquad \forall i = 1,\dots,N+1.$$

On a donc $v_1 = v_2 = \ldots = v_N = v_0 = v_{N+1} = 0$. Remarquons que ces égalités sont vérifiées même si les c_i sont nuls. Ceci démontre que la matrice A_h est bien définie. •

Remarque 1.5 — Existence et unicité de la solution. On a montré ci-dessus que A_h est symétrique définie positive, donc inversible, ce qui entraîne l'existence et l'unicité de la solution de (1.26). On aurait pu aussi démontrer l'existence et l'unicité de la solution de (1.26) directement, en montrant que le noyau de A_h est réduit à $\{0\}$ [exercice 7]. On rappelle qu'en dimension finie, toute application linéaire injective ou surjective est bijective.

Remarque 1.6 — Caractère défini et conditions limites. Dans la démonstration de la proposition 1.4, si $c_i > 0$ pour tout $i = 1,\dots,N$ le terme $\sum_{i=1}^{N} c_i h^2 v_i^2 = 0$ permet de conclure que $v_i = 0$ pour tout $i = 1,\dots,N$. Par contre, si on n'a que $c_i \geqslant 0$ (ou, bien même $c_i = 0$ pour tout $i = 1,\dots,N$), on peut encore montrer que $v_i = 0$ pour tout $i = 1,\dots,N$ grâce aux conditions aux limites de Dirichlet homogènes (représentées par le fait qu'on pose $v_0 = 0$ et $v_{N+1} = 0$ qui permet d'écrire alors les équations 1 et N sous la même forme que l'équation i) ; on a en effet $v_i = v_{i-1}$ pour tout $i = 1,\dots,N$ et $v_0 = 0$. En particulier, la matrice de discrétisation de $-u''$ par différences finies avec conditions aux limites de Neumann homogènes :

$$\begin{cases} -u'' = f \\ u'(0) = 0 \\ u'(1) = 0 \end{cases} \tag{1.27}$$

donne une matrice A_h qui est symétrique et semi définie positive, mais non définie [exercice 16]. De fait la solution du problème continu (1.27) n'est pas unique, puisque les fonctions constantes sur $[0\,;1]$ sont solutions de (1.27).

Nous allons maintenant nous préoccuper de la question de la convergence.

Définition 1.7 — Matrice d'inverse à coefficients positifs, ou IP–matrice, ou matrice monotone. Soit $A \in \mathcal{M}_N(\mathbb{R})$ de coefficients $a_{i,j}$, $i,j = 1,\dots,N$. On dit que A est à coefficients positifs (ou $A \geqslant 0$) si $a_{i,j} \geqslant 0$, $\forall i,j = 1,\dots,N$. On dit que A est d'inverse à coefficients positifs (ou en abrégé IP-matrice), ou que A est monotone (c'est la terminologie classique, mais pas la plus claire...) si A est inversible et $A^{-1} \geqslant 0$, c.à.d. que tous les coefficients de A^{-1} sont positifs ou nuls. [7].

L'avantage des schémas dont la matrice est une IP-matrice est de satisfaire la propriété de conservation de la positivité qui peut être cruciale dans les applications physiques :

Proposition 1.8 — Caractérisation des matrices d'inverse à coefficients positifs. Soit $A \in \mathcal{M}_N(\mathbb{R})$ de coefficients $a_{i,j}$, $i,j = 1,\dots,N$; alors A est une IP-matrice (ou matrice monotone) si et seulement si

$$\forall v \in \mathbb{R}^N, \qquad [Av \geqslant 0] \Longrightarrow v \geqslant 0. \tag{1.28}$$

7. Voir les exercices sur les matrices d'inverse à coefficients positifs du polycopié de L3 à l'adresse : `https://www.i2m.univ-amu.fr/perso/raphaele.herbin/PUBLI/anamat.pdf`.

(Les inégalités s'entendent composante par composante.)

Démonstration. Supposons d'abord que A vérifie (1.28) et montrons que A inversible et que A^{-1} a des coefficients positifs ou nuls. Si x est tel que $Ax = 0$, alors $Ax \geqslant 0$ et donc, par hypothèse, $x \geqslant 0$. Mais on a aussi $Ax \leqslant 0$, soit $A(-x) \geqslant 0$ et donc par hypothèse, $x \leqslant 0$. On en déduit $x = 0$, ce qui prouve que A est inversible. La propriété (1.28) donne alors que $y \geqslant 0 \Rightarrow A^{-1}y \geqslant 0$. En prenant $y = e_1$ on obtient que les coefficients de la première colonne de A^{-1} sont positifs, puis en prenant $y = e_i$ on obtient que les coefficients de la i-ième colonne de A^{-1} sont positifs pour $i = 2, \ldots, N$. Par conséquent, A^{-1} a tous ses coefficients positifs.

Réciproquement, supposons maintenant que A est inversible et que A^{-1} a des coefficients positifs. Soit $x \in \mathbb{R}^N$ tel que $Ax = y \geqslant 0$, alors $x = A^{-1}y \geqslant 0$. Donc A conserve la positivité. •

Remarque 1.9 — Principe du maximum. On appelle principe du maximum continu le fait que si $f \geqslant 0$ alors le minimum de la fonction u solution du problème 1.24 est atteint sur les bords. Cette propriété mathématique correspond à l'intuition physique qu'on peut avoir du phénomène : si on chauffe un barreau tout en maintenant ses deux extrémités à une température fixe, la température aux points intérieurs du barreau sera supérieure à celle des extrémités. Il est donc souhaitable que la solution approchée satisfasse la même propriété [exercice 3].

Lemme 1.10 — Monotonie de la matrice de discrétisation. Soit $c = (c_1, \ldots, c_N) \in \mathbb{R}^N$ et $A_h \in \mathcal{M}_N(\mathbb{R})$ définie en (1.26). Si $c_i \geqslant 0$, $\forall i = 1, \ldots, N$ alors A_h est une IP-matrice.

Démonstration. On va montrer que si $v \in \mathbb{R}^N$, $A_h v \geqslant 0$ alors $v \geqslant 0$. On peut alors utiliser la proposition 1.8 pour conclure. Soit $v = (v_1, \ldots, v_N) \in \mathbb{R}^N$. Posons $v_0 = v_{N+1} = 0$. Supposons que $A_h v \geqslant 0$, on a donc

$$-\frac{v_{i-1}}{h^2} + \left(\frac{2}{h^2} + c_i\right)v_i - \frac{v_{i+1}}{h^2} \geqslant 0 \qquad i = 1, \ldots, N \tag{1.29}$$

Soit $p = \min\left\{i \in \{1, \ldots, N\}; v_i = \min_{j=1,\ldots,N} v_j\right\}$. Supposons que $p \geqslant 1$, on a alors :

$$\frac{v_p - v_{p-1}}{h^2} + c_p v_p + \frac{v_p - v_{p+1}}{h^2} \geqslant 0 \text{ soit encore } c_p v_p \geqslant \frac{v_{p-1} - v_p}{h^2} + \frac{v_{p+1} - v_p}{h^2} \geqslant 0.$$

Par hypothèse, $c_p \geqslant 0$.

- Si $c_p > 0$, on a donc $v_p \geqslant 0$ et donc $v_i \geqslant 0$, $\forall i = 1, \ldots, N$.
- Si $c_p = 0$, on doit alors avoir $v_{p-1} = v_p = v_{p+1}$ ce qui est impossible car p est le plus petit indice i tel que $v_i = \min_{j=1,\ldots,N} v_j$. Donc dans ce cas le minimum ne peut pas être atteint pour $p > 1$ et on a donc une contradiction.

On a ainsi finalement montré que $\min_{i\in\{1,\ldots,N\}} v_i \geqslant 0$, on a donc $v \geqslant 0$. •

Définition 1.11 — Erreur de consistance. On appelle erreur de consistance la quantité obtenue en remplaçant l'inconnue par la solution exacte dans le schéma numérique. Dans le cas du schéma (1.25), l'erreur de consistance au point x_i est donc définie par :

$$R_i = \frac{2u(x_i) - u(x_{i-1}) - u(x_{i+1})}{h^2} + c(x_i)u(x_i) - f(x_i) \tag{1.30}$$

L'erreur de consistance R_i est donc l'erreur qu'on commet en remplaçant l'opérateur $-u''$ par le quotient différentiel

$$\frac{2u(x_i) - u(x_{i-1}) - u(x_{i+1})}{h^2}$$

Cette erreur peut être évaluée si u est suffisamment régulière en effectuant des développements de Taylor.

Définition 1.12 — Ordre du schéma. On dit qu'un schéma de discrétisation à N points est d'ordre p s'il existe $C \in \mathbb{R}$, ne dépendant que de la solution exacte, tel que l'erreur de consistance satisfait :

$$\max_{i=1,\ldots,N}(R_i) \leqslant Ch^p,$$

où h est le le pas du maillage défini par (1.3).

Définition 1.13 — Consistance du schéma. On dit qu'un schéma de discrétisation est consistant si

$$\max_{i=1,\dots N}(R_i) \to 0 \text{ lorsque } h \to 0$$

où N est le nombre de points de discrétisation.

Lemme 1.14 Si la solution de (1.24) vérifie $u \in \mathcal{C}^4([0\,;1])$, le schéma (1.25) est consistant d'ordre 2 et on a plus précisément :

$$|R_i| \leqslant \frac{h^2}{12} \sup_{[0\,;1]} |u^{(4)}| \qquad \forall i = 1, \dots, N \tag{1.31}$$

Démonstration. Par développement de Taylor, on a :

$$u(x_{i+1}) = u(x_i) + hu'(x_i) + \frac{h^2}{2}u''(x_i) + \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\xi_i) \tag{1.32}$$

$$u(x_{i-1}) = u(x_i) - hu'(x_i) + \frac{h^2}{2}u''(x_i) - \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\eta_i) \tag{1.33}$$

En additionnant ces deux égalités, on obtient que :

$$\frac{u(x_{i+1}) + u(x_i) - 2u(x_i)}{h^2} = u''(x_i) + \frac{h^2}{24}(u^{(4)}(\xi_i) + u^{(4)}(\eta_i))$$

ce qui entraîne que :

$$|R_i| \leqslant \frac{h^2}{12} \sup_{[0\,;1]} |u^{(4)}| \tag{1.34}$$

•

Remarque 1.15 — Sur l'erreur de consistance.

1. Si on note $\bar{U}_h : (u(x_i))_{i=1\dots N}$ le vecteur dont les composantes sont les valeurs exactes de la solution de (1.24) et $U_h = (u_1 \dots u_N)$ la solution de (1.25), on a :

   $$R = A_h(U_h - \bar{U}_h) \tag{1.35}$$

   où $R \in \mathbb{R}^N$ est le vecteur de composantes R_i, $i = 1, \dots, N$, erreur de consistance en x_i définie en (1.30).

2. On peut remarquer que si $u^{(4)} = 0$, les développements de Taylor se résument à :

   $$-u''(x_i) = \frac{2u(x_i) - u(x_{i-1}) - u(x_{i+1})}{h^2}$$

   et on a donc $R_i = 0$, pour tout $i = 1, \dots, N$ et donc $u_i = u(x_i)$ pour tout $i = 1 \dots N$. Dans ce cas (rare !), le schéma de discrétisation donne la valeur exacte de la solution en x_i, pour tout $i = 1, \dots, N$. Cette remarque est bien utile lors de la phase de validation de méthodes numériques et/ou programmes informatiques pour la résolution de l'équation (1.24). En effet, si on choisit f telle que la solution soit un polynôme de degré inférieur ou égal à 3, alors on doit avoir une erreur entre solution exacte et approchée inférieure à l'erreur machine.

La preuve de convergence du schéma utilise la notion de consistance, ainsi qu'une notion de stabilité, qui consiste en une borne sur la solution approchée, indépendante du maillage.

Proposition 1.16 — Stabilité du schéma. Le schéma (1.25) est *stable*, au sens où la norme infinie de la solution approchée est bornée par un nombre ne dépendant que de f.

Plus précisément, la matrice de discrétisation A_h satisfait :

$$\|A_h^{-1}\|_\infty \leqslant \frac{1}{8} \tag{1.36}$$

inégalité qui peut aussi s'écrire comme une estimation sur les solutions du système (1.26) :

$$\|U_h\|_\infty \leqslant \frac{1}{8}\|f\|_\infty \tag{1.37}$$

Démonstration. On rappelle que par définition, si $M \in \mathcal{M}_N(\mathbb{R})$,

$$\|M\|_\infty = \sup_{v\in\mathbb{R}^N v\neq 0} \frac{\|Mv\|_\infty}{\|v\|_\infty} \text{ avec } \|v\|_\infty = \sup_{i=1,\dots,N} |v_i|.$$

Pour montrer que $\|A_h^{-1}\|_\infty \leqslant 1/8$, on décompose la matrice A_h sous la forme $A_h = A_{0h} + \mathrm{diag}(c_i)$ où A_{0h} est la matrice de discrétisation de l'opérateur $-u''$ avec conditions aux limites de Dirichlet homogènes et

$$A_{0h} = \frac{1}{h^2}\begin{bmatrix} 2 & -1 & & 0 \\ -1 & \ddots & & \\ & & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix} \text{ et } \mathrm{diag}(c_i) = \begin{bmatrix} c_1 & 0 & & 0 \\ 0 & c_2 & & \\ & & \ddots & 0 \\ 0 & & 0 & c_N \end{bmatrix}. \tag{1.38}$$

Les matrices A_{0h} et A_h sont inversibles et on a :

$$A_{0h}^{-1} - A_h^{-1} = A_{0h}^{-1}A_hA_h^{-1} - A_{0h}^{-1}A_{0h}A_h^{-1} = A_{0h}^{-1}(A_h - A_{0h})A_h^{-1}.$$

Comme $\mathrm{diag}(c_i) \geqslant 0$, on a $A_h \geqslant A_{0h}$ et comme A_{0h} et A_h sont des IP-matrices, on en déduit que $0 \leqslant A_h^{-1} \leqslant A_{0h}^{-1}$ composante par composante. On peut maintenant remarquer que si $B \in \mathcal{M}_N(\mathbb{R})$ et si $B \geqslant 0$ (c'est-à-dire $B_{ij} \geqslant 0$ pour tout i et j), on a

$$\|B\|_\infty = \sup_{\substack{v\in\mathbb{R}^N \\ \|v\|=1}} \sup_{i=1,\dots,N} |(Bv)_i| = \sup_{\substack{v\in\mathbb{R}^N \\ \|v\|=1}} \sup_{i=1,\dots,N} \left|\sum_{j=1}^N B_{ij}v_j\right| \text{ c'est-à-dire } \|B\|_\infty = \sup_{i=1,\dots,N}\sum_{j=1}^N B_{ij}.$$

On a donc

$$\|A_h^{-1}\| = \sup_{i=1,\dots,N}\sum_{j=1}^N (A_h^{-1})_{ij} \leqslant \sup_{i=1,\dots,N}\sum_{j=1}^N (A_{0h}^{-1})_{ij}$$

car $A_h^{-1} \leqslant A_{0h}^{-1}$ d'où on déduit que $\|A_h^{-1}\|_\infty \leqslant \|A_{0h}^{-1}\|_\infty$. Il ne reste plus qu'à estimer $\|A_{0h}^{-1}\|_\infty$. Comme $A_{0h}^{-1} \geqslant 0$, on a $\|A_{0h}^{-1}\|_\infty = \|A_{0h}^{-1}e\|_\infty$ avec $e = (1,\dots,1)$. Soit $d = A_{0h}^{-1}e \in \mathbb{R}^N$. On veut calculer $\|d\|_\infty$, où d vérifie $A_{0h}d = e$. Or le système linéaire $A_{0h}d = e$ n'est autre que la discrétisation par différences finies du problème

$$\begin{cases} -u'' = 1 \\ u(0) = 0 \\ u(1) = 0 \end{cases} \tag{1.39}$$

dont la solution exacte est $u_0(x) = x(1-x)/2$ qui vérifie $u_0^{(4)}(x) = 0$. On en conclut, par la remarque 1.15, que $u_0(x_i) = d_i$, $\forall i = 1\dots N$. Donc $\|d\|_\infty = \sup_{i=1,N} \frac{ih(ih-1)}{2}$ où $h = 1/(N+1)$ est le pas de discrétisation. Ceci entraîne que

$$\|d\|_\infty \leqslant \sup_{[0\,;1]} \left|\frac{x(x-1)}{2}\right| = \frac{1}{8}$$

et donc que $\|A_h^{-1}\|_\infty \leqslant 1/8$. •

Remarque 1.17 — Sur la stabilité. L'inégalité (1.37) donne une estimation sur les solutions approchées indépendantes du pas de maillage. C'est ce type d'estimation que l'on recherchera par la suite comme garant de la stabilité d'un schéma numérique.

Définition 1.18 — Erreur de discrétisation. On appelle erreur de discrétisation en x_i, la différence entre la solution exacte en x_i et la i-ème composante de la solution donnée par le schéma numérique

$$e_i = u(x_i) - u_i \qquad \forall i = 1,\dots,N \tag{1.40}$$

Théorème 1.1 Soit u la solution exacte de

$$\begin{cases} -u'' + cu = f \\ u(0) = 0 \\ u(1) = 0 \end{cases}$$

On suppose $u \in \mathcal{C}^4([0\,;1])$. Soit u_h la solution de (1.25). Alors l'erreur de discrétisation définie par (1.40) satisfait

$$\max_{i=1,\dots,N} |e_i| \leqslant \frac{1}{96}\|u^{(4)}\|_\infty h^2.$$

Le schéma est donc convergent d'ordre 2.

Démonstration. Soit $U_h = (U_1, \ldots, U_n)$ et $\bar{U}_h = (u(x_1), \ldots, u(x_N))$, on cherche à majorer $\|\bar{U}_h - U_h\|_\infty$. On a $A(\bar{U}_h - U_h) = R$ où R est l'erreur de consistance [remarque 1.15]. On a donc

$$\|\bar{U}_h - U_h\|_\infty \leqslant \|A_h^{-1}\|_\infty \|R\|_\infty \leqslant \frac{1}{8} \times \frac{1}{12}\|u^{(4)}\|_\infty = \frac{1}{96}\|u^{(4)}\|_\infty$$

•

Remarque 1.19 — Sur la convergence. On peut remarquer que la preuve de la convergence s'appuie sur la stabilité (elle-même déduite de la conservation de la positivité) et sur la consistance. Dans certains livres d'analyse numérique, vous trouverez la "formule" : stabilité + consistance $\Rightarrow$ convergence. Il faut toutefois prendre garde au fait que ces notions de stabilité et consistance peuvent être variables d'un type de méthode à un autre comme nous le verrons en étudiant la méthode des volumes finis, par exemple.

Remarque 1.20 — Contrôle des erreurs d'arrondi. On cherche à calculer la solution approchée de $-u'' = f$. Le second membre f est donc une donnée du problème. Supposons que des erreurs soient commises sur cette donnée (par exemple des erreurs d'arrondi, ou des erreurs de mesure). On obtient alors un nouveau système, qui s'écrit $A_h\widetilde{U}_h = b_h + \varepsilon_h$ où ε_h représente la discrétisation des erreurs commises sur le second membre. Si on résout $A_h\widetilde{U}_h = b_h + \varepsilon_h$ au lieu de $A_hU_h = b_h$, l'erreur commise sur la solution du système s'écrit :

$$E_h = \widetilde{U}_h - U_h = A_h^{-1}\varepsilon_h$$

On en déduit que

$$\|E_h\|_\infty \leqslant \frac{1}{8}\|\varepsilon_h\|_\infty$$

On a donc une borne d'erreur sur l'erreur qu'on obtient sur la solution du système par rapport à l'erreur commise sur le second membre. Le problème des erreurs relatives est beaucoup plus subtil [exercice 7].

1.3 Volumes finis pour un problème elliptique unidimensionnel

1.3.1 Écriture du schéma

On va étudier la discrétisation par volumes finis du problème (1.1)-(1.2). On utilise ici la notation u'' plutôt que $\partial^2_{xx}u$ puisque l'inconnue u ne dépend que de la variable x.

Définition 1.21 — Maillage volumes finis. On appelle maillage volumes finis de l'intervalle $[0\,;1]$, un ensemble de N mailles $(K_i)_{i=1,\ldots,N}$ tel que $K_i =]x_{i-1/2}\,;x_{i+1/2}[$ et on note $h_i = x_{i+1/2} - x_{i-1/2}$. On se donne également N points $(x_i)_{i=1,\ldots,N}$ situés dans les mailles K_i. On a donc :

$$0 = x_{1/2} < x_1 < x_{3/2} < \ldots < x_{i-1/2} < x_i < x_{i+1/2} < \ldots < x_{N+1/2} = 1$$

On notera $h_{i+1/2} = x_{i+1} - x_i$ et $h = \max_{i=1,\ldots,N} h_i$. Pour des questions de notations, on posera également $x_0 = 0$ et $x_{N+1} = 1$.

On rappelle que pour obtenir un schéma volumes finis, on part de la forme intégrale (bilan des flux) obtenue en intégrant l'équation (1.1) sur K_i :

$$-u'(x_i + 1/2) + u'(x_i - 1/2) = \int_{K_i} f(x)\,\mathrm{d}x. \tag{1.41}$$

On pose

$$f_i = \frac{1}{h_i}\int_{K_i} f(x)\,\mathrm{d}x$$

et on introduit les inconnues discrètes $(u_i)_{i=1\ldots N}$ (une par maille) et les équations discrètes du schéma numérique :

$$F_{i+1/2} - F_{i-1/2} = h_i f_i \qquad i = 1, \ldots, N, \tag{1.42}$$

où $F_{i+1/2}$ est le flux numérique en $x_{i+1/2}$ qui devrait être une approximation raisonnable de $-u'(x_{i+1/2})$. On pose alors :

$$\begin{cases} F_{i+1/2} = -\dfrac{u_{i+1}-u_i}{h_{i+1/2}}, & i = 1,\dots,N, \\ F_{1/2} = -\dfrac{u_1}{h_{1/2}}, \\ F_{N+1/2} = \dfrac{u_N}{h_{N+1/2}}. \end{cases}$$

Notons que les flux $F_{1/2}$ et $F_{N+1/2}$ tiennent compte des conditions aux limites de Dirichlet homogènes $u(0) = u(1) = 0$. On peut aussi écrire :

$$\begin{cases} F_{i+1/2} = -\dfrac{u_{i+1}-u_i}{h_{i+1/2}}, & i = 0,\dots,N, \quad (1.43) \\ u_0 = u_{N+1} = 0. \quad (1.44) \end{cases}$$

On peut écrire le système linéaire obtenu sur $(u_1,\dots,u_N)$ sous la forme

$$A_h U_h = b_h \quad \text{avec} \quad (A_h U_h)_i = \frac{1}{h_i}\Big(\frac{u_i - u_{i+1}}{h_{i+1/2}} + \frac{u_i - u_{i-1}}{h_{i-1/2}}\Big) \quad \text{et} \quad (b_h)_i = f_i \tag{1.45}$$

Remarque 1.22 — Non consistance au sens des différences finies. L'approximation de $-u''(x_i)$ par

$$\frac{1}{h_i}\Big(\frac{u_i - u_{i+1}}{h_{i+1/2}} + \frac{u_i - u_{i-1}}{h_{i-1/2}}\Big)$$

n'est pas consistante dans le cas général [exercice 11].

On peut montrer que les deux schémas différences finies et volumes sont identiques *au bord près* dans le cas d'un maillage uniforme lorsque x_i est supposé être le centre de la maille (voir l'exercice 1).

1.3.2 Analyse mathématique du schéma

On va démontrer ici qu'il existe une unique solution $(u_1,\dots u_N)$ au schéma (1.42)-(1.44) et que cette solution converge, en un certain sens, vers la solution de problème continu (1.1)-(1.2) lorsque le pas du maillage tend vers 0.

Proposition 1.23 — Existence de la solution du schéma volumes finis. Soit $f \in \mathcal{C}([0\,;1])$ et $u \in \mathcal{C}^2([0\,;1])$ solution du problème (1.1)-(1.2). Soit $(K_i)_{i=1,\dots,N}$ le maillage de la définition 1.21. Alors il existe une unique solution $u_h = (u_1,\dots,u_N)$ de (1.42)-(1.44).

Démonstration. Ce résultat se déduit facilement de la proposition suivante, qui donne la stabilité du shéma, c'est-à-dire une estimation *a priori* sur les solutions approchées. Si $f_i = 0$ pour tout $i = 1,\dots,N$, la proposition 1.25 nous donne que $\|D_\mathcal{T} u_\mathcal{T}\| = 0$ où $D_\mathcal{T} u_\mathcal{T}$ est la *dérivée discrète* et donc $u_i - u_{i-1} = 0$ pour tout $i = 1 \dots N$; mais comme $u_0 = 0$, on en déduit que $u_i = 0$ pour tout $i = 1 \dots N$. Ceci démontre l'unicité de $(u_i)_{i=1\dots N}$ solution de (1.42)-(1.44) et donc son existence, puisque le système (1.42)-(1.44) est un système linéaire carré d'ordre N. (On rappelle qu'une matrice carrée d'ordre N est inversible si et seulement si son noyau est réduit à $\{0\}$). •

Nous allons maintenant prouver la stabilité, sous forme d'une estimation dite *a priori*, car on effectue une majoration sur une fonction dont on n'a pas forcément prouvé l'existence : on établit l'estimation *a priori* en premier et on en déduit l'existence.

Pour démontrer cette propriété, on commence par introduire une *dérivée discrète* des fonctions constantes par mailles, qui nous servira dans la suite des démonstrations.

Définition 1.24 — Dérivée discrète. On considère le maillage $(K_i)_{i=1,\dots N}$ de la définition 1.21. Soit v une fonction constante par mailles sur les mailles K_i, qui représente une approximation

d'une fonction définie sur $[0\,;1]$ et nulle en 0 et en 1. En posant $v_0 = v_{N+1} = 0$, on peut définir une dérivée discrète de v par les pentes

$$p_{i+1/2} = \frac{v_{i+1} - v_i}{h_{i+1/2}} \qquad i = 0, \dots, N.$$

On peut alors définir une fonction $D_{\mathcal{T}} v$, constante par intervalle et égale à $p_{i+1/2}$ sur l'intervalle $K_{i+1/2} =]x_i\,; x_{i+1}[$. La norme L^2 de $D_{\mathcal{T}} v$ est définie par :

$$\|D_{\mathcal{T}} v\|^2_{L^2(]0\,;1[)} = \sum_{i=0}^{N} \int_{K_{i+1/2}} |D_{\mathcal{T}} v|^2 = \sum_{i=0}^{N} h_{i+1/2}\; p^2_{i+1/2} = \sum_{i=0}^{N} \frac{(v_{i+1} - v_i)^2}{h_{i+1/2}}.$$

(On rappelle que $h_{i+1/2} = \dfrac{h_i + h_{i+1}}{2}$.)

Proposition 1.25 — Stabilité : estimation *a priori* sur les solutions approchées. Soit $f \in L^2([0\,;1])$. On considère le maillage $(K_i)_{i=1,\dots N}$ de la définition 1.21 ; pour $i = 1, \dots, N$, on note f_i la valeur moyenne de f sur la maille K_i. Si $u_{\mathcal{T}}$ est la fonction constante par maille dont les valeurs sur les mailles sont des valeurs $(u_1, \dots, u_N)$ qui vérifient le schéma volumes finis (1.42)-(1.44), alors

$$\|D_{\mathcal{T}} u_{\mathcal{T}}\|_{L^2} \leqslant \|f\|_{L^2} \tag{1.46}$$

Démonstration. La preuve de cette proposition est calquée sur l'estimation *a priori* qu'on peut faire sur les solutions du problème continu : en effet, si u est une solution qu'on supposera aussi régulière que l'on veut [8] du problème (1.1)-(1.2), alors

$$\|u'\|_{L^2(]0\,;1[)} \leqslant \|f\|_{L^2((0,1))} \tag{1.47}$$

Nous allons donc mener les preuves de (1.47) et (1.46) en parallèle. Soit $u \in \mathcal{C}^2(]0\,;1[)$ solution de (1.1)-(1.2) et soit $(u_1, \dots, u_N)$ solution de (1.42)-(1.44).

Estimation continue. On multiplie (1.1) par u et on intègre entre 0 et 1 :

$$-\int_0^1 u''(x)u(x)\,\mathrm{d}x = \int_0^1 f(x)u(x)\,\mathrm{d}x$$

On intègre par parties et on utilise les conditions aux limites (1.2) :

$$\int_0^1 (u'(x))^2\,\mathrm{d}x = \int_0^1 f(x)u(x)\,\mathrm{d}x$$

En utilisant l'inégalité de Cauchy-Schwarz pour le membre de droite, on obtient

$$\int_0^1 (u'(x))^2\,\mathrm{d}x \leqslant \|f\|_{L^2(]0\,;1[)}\|u\|_{L^2(]0\,;1[)}$$

On utilise alors l'inégalité de Poincaré qui s'écrit (voir Proposition 1.26 plus bas) $\|u\|_{L^2(]0\,;1[)} \leqslant \|u'\|_{L^2(]0\,;1[)}$. On en déduit (1.47).

Estimation discrète. On multiplie (1.42) par u_i et on somme sur i :

$$\sum_{i=1}^{N} \left(-\frac{u_{i+1} - u_i}{h_{i+1/2}} + \frac{u_i - u_{i-1}}{h_{i-1/2}} \right) u_i = \sum_{i=1}^{N} h_i f_i u_i$$

On somme par parties en utilisant $u_0 = u_{N+1} = 0$. On a :

$$\sum_{i=1}^{N} -\frac{u_{i+1} - u_i}{h_{i+1/2}} u_i + \sum_{i=1}^{N} \frac{u_i - u_{i-1}}{h_{i-1/2}} u_i = \sum_{i=1}^{N} h_i f_i u_i.$$

En effectuant un changement d'indice sur la deuxième somme, on obtient :

$$\sum_{i=1}^{N} -\frac{u_{i+1} - u_i}{h_{i+1/2}} u_i + \sum_{i=0}^{N-1} \frac{u_{i+1} - u_i}{h_{i+1/2}} u_{i+1} = \sum_{i=1}^{N} h_i f_i u_i;$$

en regroupant les sommes, on a donc :

$$\sum_{i=0}^{N} \frac{(u_{i+1} - u_i)^2}{h_{i+1/2}} = \sum_{i=1}^{N} h_i f_i u_i$$

8. En fait la solution unique de (1.1)-(1.2) appartient à l'espace $H^1(]0\,;1[)$, comme on le verra plus tard.

Par définition de $D_{\mathcal{T}}u$, on a

$$\|D_{\mathcal{T}}u_{\mathcal{T}}\|_{L^2}^2 = \sum_{i=0}^{N} \int_{K_{i+1/2}} \frac{(u_{i+1}-u_i)^2}{h_{i+1/2}^2} = \sum_{i=0}^{N} \frac{(u_{i+1}-u_i)^2}{h_{i+1/2}}.$$

On a donc

$$\|D_{\mathcal{T}}u_{\mathcal{T}}\|_{L^2}^2 \leqslant \sum_{i=1}^{N} h_i f_i u_i$$

On utilise alors l'inégalité de Cauchy-Schwarz pour le membre de droite :

$$\sum_{i=1}^{N} h_i f_i u_i \leqslant \left(\sum_{i=1}^{N} h_i f_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{N} h_i u_i^2\right)^{\frac{1}{2}}$$

Or, toujours par l'inégalité de Cauchy-Schwarz,

$$\sum_{i=1}^{N} h_i f_i^2 = \sum_{i=1}^{N} h_i \frac{1}{h_i^2} \left(\int_{K_i} f(x)\,\mathrm{d}x\right)^2 \leqslant \sum_{i=1}^{N} \int_{K_i} f(x)^2\,\mathrm{d}x = \|f\|_{L^2}^2.$$

On obtient donc alors $\|(D_{\mathcal{T}}u_{\mathcal{T}})^2\|_{L^2(]0\,;1[)} \leqslant \|f\|_{L^2(]0,1[)}\|u_{\mathcal{T}}\|_{L^2(]0,1[)}$. On utilise maintenant l'inégalité de Poincaré discrète (voir Proposition 1.26 plus bas), qui s'écrit $\|u_{\mathcal{T}}\|_{L^2(]0\,;1[)} \leqslant \|D_{\mathcal{T}}u_{\mathcal{T}}\|_{L^2(]0\,;1[)}$, pour en déduire (1.46).

•

Il nous faut maintenant donner les démonstrations des inégalités de Poincaré continue et discrète que nous avons utilisées dans la preuve ci-dessus.

Proposition 1.26 — Inégalité de Poincaré, cas continu et discret.

Inégalité de Poincaré pour une fonction régulière Soit $u \in \mathcal{C}^1([0\,;1])$ telle que $u(0) = 0$. Alors

$$\|u\|_{L^2(]0\,;1[)} \leqslant \|u'\|_{L^2(]0\,;1[)} \tag{1.48}$$

$$\|u\|_{L^\infty(]0\,;1[)} \leqslant \|u'\|_{L^2(]0\,;1[)} \tag{1.49}$$

Inégalité de Poincaré dans le cadre discret On considère le maillage $(K_i)_{i=1,\dots N}$ de la définition 1.21. Soit v une fonction constante par mailles sur les mailles K_i et soit $D_{\mathcal{T}}v$ sa "dérivée discrète" au sens de la définition 1.24. Alors :

$$\|v\|_{L^2(]0\,;1[} \leqslant \|D_{\mathcal{T}}v\|_{L^2(]0\,;1[)} \tag{1.50}$$

$$\|v\|_{\infty} \leqslant \|D_{\mathcal{T}}v\|_{L^2(]0\,;1[)} \tag{1.51}$$

Démonstration. Là encore, on va effectuer les démonstrations en parallèle, vu que la démonstration de l'inégalité "discrète" copie la démonstration de l'inégalité continue.

Le cas continu.

On écrit que u est l'intégrale de sa dérivée, en utilisant le fait que $u(0) = 0$:

$$u(x) = \int_0^x u'(s)\,\mathrm{d}s$$

On a donc :

$$|u(x)| \leqslant \int_0^x |u'(s)|\,\mathrm{d}s \leqslant \int_0^1 |u'(s)|\,\mathrm{d}s$$

En utilisant l'inégalité de Cauchy-Schwarz pour le membre de droite, on obtient que

$$|u(x)|^2 \leqslant \|u'\|_{L^2(]0\,;1[}^2,\ \forall x \in]0,1[, \tag{1.52}$$

ce qui donne immédiatement (1.49). On intègre ensuite l'inégalité (1.52) entre 0 et 1 pour aboutir à (1.48).

Inégalité de Poincaré discrète. En tenant compte du fait que $v_0 = 0$, on écrit que :

$$v_i = \sum_{k=1}^{i} (v_k - v_{k-1}).$$

On majore :

$$|v_i| \leqslant \sum_{k=1}^{i} |v_k - v_{k-1}| \leqslant \sum_{k=1}^{N+1} h_{k+1/2} \frac{|v_k - v_{k-1}|}{h_{k+1/2}} = \int_0^1 |D_{\mathcal{T}} v(s)\, \mathrm{d}s$$

On utilise l'inégalité de Cauchy-Schwarz à droite $|v_i|^2 \leqslant \|D_{\mathcal{T}} v\|^2_{L^2(]0\,;1[)}$, ce qui donne tout de suite (1.51) et en intégrant entre 0 et 1, on aboutit au résultat $\|v\|_{L^2(]0\,;1[} \leqslant \|D_{\mathcal{T}} v\|_{L^2(]0\,;1[)}$. Notons que dans les deux démonstrations, on obtient que $\|u\|_\infty \leqslant \|u\|_{L^2(]0\,;1[)}$ et que $\|u\|_\infty \leqslant \|u'\|_{L^2(]0\,;1[)}$. •

Définition 1.27 — Erreur de consistance sur le flux. Soit $u : [0\,;1] \to \mathbb{R}$ solution du problème (1.1)-(1.2). On se donne une subdivision de $[0\,;1]$. On appelle $\bar{F}_{i+1/2} = -u'(x_{i+1/2})$ le flux exact en $x_{i+1/2}$ et $F^*_{i+1/2} = -(u(x_{i+1}) - u(x_i))/h_{i+1/2}$, l'approximation du flux exact utilisée pour construire le flux numérique $F_{i+1/2} = -(u_{i+1} - u_i)/h_{i+1/2}$. On dit que le flux numérique est consistant d'ordre p s'il existe $C \in \mathbb{R}_+$ ne dépendant que de u telle que l'erreur de consistance sur le flux, définie par :

$$R_{i+1/2} = \bar{F}_{i+1/2} - F^*_{i+1/2}, \tag{1.53}$$

vérifie

$$|R_{i+1/2}| \leqslant Ch^p. \tag{1.54}$$

Lemme 1.28 — Consistance du flux de diffusion. Soit $u \in \mathcal{C}^2([0\,;1])$ solution du problème (1.1)-(1.2). Le flux numérique $F_{i+1/2} = -\frac{u_{i+1} - u_i}{h_{i+1/2}}$ est consistant d'ordre 1. Plus précisément il existe C ne dépendant que de $\|u''\|_\infty$ tel que l'erreur de consistance sur les flux définie par (1.53) vérifie :

$$|R_{i+1/2}| = \left| -u'(x_{i+1/2}) + \frac{u(x_{i+1}) - u(x_i)}{h_{i+1/2}} \right| \leqslant Ch \tag{1.55}$$

Démonstration. La démonstration de ce résultat s'effectue facilement à l'aide de développements de Taylor [exercice 13] où l'on montre aussi que si $x_{i+1/2}$ est au centre de l'intervalle $[x_i\,; x_{i+1}]$, l'erreur de consistance sur les flux est d'ordre 2, *i.e.* il existe $C \in \mathbb{R}_+$ ne dépendant que de u telle que $R_{i+1/2} \leqslant Ch^2$. Notez que cette propriété de consistance est vraie sur les flux et non pas sur l'opérateur $-u''$ [remarque 1.22 et exercice 11]. •

Définition 1.29 — Conservativité. On dit que le schéma volumes finis (1.42)-(1.44) est conservatif, au sens où, lorsqu'on considère une interface $x_{i+1/2}$ entre deux mailles K_i et K_{i+1}, le flux numérique entrant dans une maille est égal à celui sortant de l'autre.

C'est grâce à la conservativité et à la consistance des flux qu'on va montrer la convergence du schéma volumes finis.

Théorème 1.2 — Convergence du schéma volumes finis. On suppose que la solution u du problème (1.1)-(1.2) vérifie $u \in \mathcal{C}^2([0\,;1])$. On pose $e_i = u(x_i) - u_i$ pour $i = 1, \ldots, N$ et $e_0 = e_{N+1} = 0$. On note $e_{\mathcal{T}}$ la fonction constante par mailles et égale à e_i sur la maille i et $D_{\mathcal{T}} e_{\mathcal{T}}$ sa dérivée discrète, au sens de la définition 1.24. Il existe $C \geqslant 0$ ne dépendant que de u tel que (on rappelle que $h = \sup_{i=1\ldots N} h_i$) :

$$\|D_{\mathcal{T}} e_{\mathcal{T}}\|_{L^2(]0\,;1[)} = \Big(\sum_{i=0}^{N} \frac{(e_{i+1} - e_i)^2}{h} \Big)^{1/2} \leqslant Ch, \tag{1.56}$$

$$\|e_{\mathcal{T}}\|_{L^2(]0\,;1[)} = \Big(\sum_{i=1}^{N} h e_i^2 \Big)^{1/2} \leqslant Ch \tag{1.57}$$

$$\|e_{\mathcal{T}}\|_\infty = \max_{i=1,\ldots,N} |e_i| \leqslant Ch \tag{1.58}$$

Démonstration. écrivons le schéma volumes finis (1.42) ainsi que l'équation exacte (1.41) intégrée sur la maille K_i : $\bar{F}_{i+1/2} - \bar{F}_{i-1/2} = h_i f_i$ où $\bar{F}_{i+1/2}$ est défini dans le lemme 1.28 et soustrayons : $\bar{F}_{i+1/2} -$

$F_{i+1/2} - \bar{F}_{i-1/2} + F_{i-1/2} = 0$. En introduisant $R_{i+1/2} = \overline{F}_{i+1/2} - F^*_{i+1/2}$, on obtient :

$$F^*_{i+1/2} - F_{i+1/2} - F^*_{i-1/2} + F_{i-1/2} = -R_{i+1/2} + R_{i-1/2}$$

ce qui s'écrit encore, au vu de la définition de e_i,

$$-\frac{e_{i+1} - e_i}{h_{i+1/2}} + \frac{e_i - e_{i-1}}{h_{i-1/2}} = -R_{i+1/2} + R_{i-1/2}$$

On multiplie cette dernière égalité par e_i et on somme de 1 à N :

$$\sum_{i=1}^{N} -\frac{e_{i+1} - e_i}{h_{i+1/2}} e_i + \sum_{i=1}^{N} \frac{e_i - e_{i-1}}{h_{i-1/2}} e_i = \sum_{i=1}^{N} -R_{i+1/2} e_i + \sum_{i=1}^{N} R_{i-1/2} e_i,$$

ce qui s'écrit encore :

$$\sum_{i=1}^{N} -\frac{e_{i+1} - e_i}{h_{i+1/2}} e_i + \sum_{i=0}^{N-1} \frac{e_{i+1} - e_i}{h_{i+1/2}} e_{i+1} = \sum_{i=1}^{N} -R_{i+1/2} e_i + \sum_{i=0}^{N-1} R_{i+1/2} e_{i+1}$$

En réordonnant les termes, on obtient, en remarquant que $e_0 = 0$ et $e_{N+1} = 0$:

$$\sum_{i=0}^{N} \frac{(e_{i+1} - e_i)^2}{h_{i+1/2}} = \sum_{i=0}^{N} R_{i+1/2}(e_{i+1} - e_i)$$

Mais

$$\sum_{i=0}^{N} \frac{(e_{i+1} - e_i)^2}{h_{i+1/2}} = \sum_{i=0}^{N} h_{i+1/2} \Big(\frac{e_{i+1} - e_i}{h_{i+1/2}}\Big)^2 = \int_0^1 (D_{\mathcal{T}} e_{\mathcal{T}}(s))^2 \,\mathrm{d}s = \|D_{\mathcal{T}} e_{\mathcal{T}}\|^2_{L^2(]0\,;1[)}.$$

De plus, $R_{i+1/2} \leqslant Ch$ (par le lemme 1.28). On a donc

$$\|D_{\mathcal{T}} e_{\mathcal{T}}\|^2_{L^2(]0\,;1[)} \leqslant Ch \sum_{i=0}^{N} |e_{i+1} - e_i| = Ch \sum_{i=0}^{N} h_{i+1/2} \frac{|e_{i+1} - e_i|}{h_{i+1/2}} = Ch \int_0^1 |D_{\mathcal{T}} e_{\mathcal{T}}| \,\mathrm{d}t$$

et, par l'inégalité de Cauchy-Schwarz, on a $\|D_{\mathcal{T}} e_{\mathcal{T}}\|_{L^2(]0\,;1[)} \leqslant Ch$. On a ainsi démontré (1.56). On obtient (1.57) et (1.58) par l'inégalité de Poincaré discrète [proposition 1.30]. •

Rappels : espaces fonctionnels et normes discrètes On rappelle qu'une fonction u de l'espace de Lebesgue[9] [10] $L^2(]0\,;1[)$ admet une dérivée faible dans $L^2(]0\,;1[)$ s'il existe $v \in L^2(]0\,;1[)$ telle que

$$\int_0^1 u(x)\varphi'(x) \,\mathrm{d}x = -\int_0^1 v(x)\varphi(x) \,\mathrm{d}x \tag{1.59}$$

pour toute fonction $\varphi \in \mathcal{C}^1_c(]0\,;1[)$ où $\mathcal{C}^1_c(]0\,;1[)$ désigne l'espace des fonctions de classe $\mathcal{C}^1$ à support compact dans $]0\,;1[$. On peut montrer que v est unique [1]. On notera $v = Du$. On peut remarquer que si $u \in \mathcal{C}^1([0\,;1])$ et si on note u' sa dérivée classique, alors $Du = u'$ *presque partout*. On note $H^1(]0\,;1[)$ l'ensemble des fonctions de $L^2(]0\,;1[)$ qui admettent une dérivée faible dans $L^2(]0\,;1[)$:

$$H^1(]0\,;1[) = \{u \in L^2(]0\,;1[) \ ; \ Du \in L^2(]0\,;1[)\}.$$

9. Henri-Léon Lebesgue (1875–1941) est un mathématicien français. Il est reconnu pour sa théorie d'intégration publiée initialement dans son mémoire *Intégrale, longueur, aire* à l'université de Nancy en 1902. Il fut l'un des grands mathématiciens français de la première moitié du XX-ème siècle.

10. Une fonction f de $]0\,;1[$ dans $\mathbb{R}$ est intégrable au sens de Lebesgue si f est mesurable et

$$\int_0^1 |f| \,\mathrm{d}t < +\infty.$$

L'espace $L^2(]0\,;1[)$ désigne l'ensemble des classes d'équivalence des fonctions de carré intégrable au sens de Lebesgue, pour la relation d'équivalence *égale presque partout*, ce qui permet de définir une norme sur $L^2(]0\,;1[)$ par

$$\|u\|_{L^2(]0\,;1[)} = \Big(\int_0^1 u^2 \,\mathrm{d}t\Big)^{1/2}$$

qui en fait un espace complet. Cette norme est associée au produit scalaire

$$(u, v)_{L^2(]0\,;1[)} = \int_0^1 uv \,\mathrm{d}t$$

qui en fait un espace de Hilbert [8].

C'est un espace de Hilbert pour le produit scalaire :

$$(u,v)_{H^1(]0\,;1[)} = \int_{(]0\,;1[)} \big(u(x)v(x)\,\mathrm{d}x + Du(x)Dv(x)\big)\,\mathrm{d}x.$$

Tout élément de $H^1(]0\,;1[)$ (au sens classe d'équivalence de fonctions) admet un représentant continu et on peut donc définir les valeurs en 0 et en 1 d'une "fonction" de $H^1(]0\,;1[)$:

$$H^1_0(]0\,;1[) = \{u \in H^1(]0\,;1[)\ ; u(0) = u(1) = 0\}.$$

Pour $u \in H^1(]0\,;1[)$, on note :

$$\|u\|_{H^1_0} = \left(\int_0^1 (Du(x))^2\,\mathrm{d}x\right)^{1/2}.$$

C'est une norme sur H^1_0 qui est équivalente à la norme $\|\cdot\|_{H^1}$ définie par

$$\|u\|_{H^1} = \left(\int u^2(x)\,\mathrm{d}x + \int (Du)^2(x)\,\mathrm{d}x\right)^{1/2}$$

ce qui se démontre grâce à l'inégalité de Poincaré [11] qu'on rappelle :

Proposition 1.30 — Inégalité de Poincaré, cas uni-dimensionnel.

$$\|u\|_{L^2(]0\,;1[)} \leqslant \|Du\|_{L^2(]0\,;1[)} \text{ pour tout } u \in H^1_0(]0\,;1[). \tag{1.60}$$

La démonstration de cette proposition a été faite dans le cadre de la preuve de la stabilité du schéma en une dimension [proposition 1.25] pour des fonctions régulières mais la même preuve s'applique dans le cas d'une fonction H^1.

Norme H^1 discrète Soit maintenant $\mathcal{T}$ un maillage volumes finis de $[0\,;1]$ [définition 1.21], on note $X(\mathcal{T})$ l'ensemble des fonctions de $[0\,;1]$ dans $\mathbb{R}$, constantes par maille de ce maillage. Pour $v \in X(\mathcal{T})$, on note v_i la valeur de v sur la maille i ; on peut écrire les normes L^2 et L^∞ de v :

$$\|v\|^2_{L^2(]0\,;1[)} = \sum_{i=1}^N h_i v_i^2 \quad \text{et} \quad \|v\|_{L^\infty(]0\,;1[)} = \max_{i=1,\dots,N} |v_i|$$

Par contre, la fonction v étant constante par maille, elle n'est pas dérivable au sens classique, ni même au sens faible. On peut toutefois définir une norme H^1 discrète de v comme suit :

$$|v|_{1,\mathcal{T}} = \left(\sum_{i=0}^N h_{i+1/2}\Big(\frac{v_{i+1}-v_i}{h_{i+1/2}}\Big)^2\right)^{1/2}$$

ce qui est la norme L^2 de la dérivée discrète $D_{\mathcal{T}}v$ [définition 1.24]. On peut montrer [exercice 18] que si $u_{\mathcal{T}} :]0\,;1[\to \mathbb{R}$ est définie par $u_{\mathcal{T}}(x) = u_i$, $\forall x \in K_i$ où $(u_i)_{i=1,\dots,N}$ solution de (1.42)-(1.44) alors $|u_{\mathcal{T}}|_{1,\mathcal{T}}$ converge dans $L^2(]0\,;1[)$ lorsque h tend vers 0, vers $\|Du\|_{L^2(]0\,;1[)}$, où u est la solution du problème (1.1)-(1.2).

Dimensions supérieures

En une dimension d'espace, on a obtenu une estimation d'erreur en norme H^1_0 discrète et en norme L^∞. En dimension supérieure ou égale à 2, on aura une estimation en h, en norme H^1_0 discrète, en norme L^2, mais pas en norme L^∞. Ceci tient au fait que l'injection de Sobolev $H^1(]0\,;1[) \subset \mathcal{C}(]0\,;1[)$ n'est vraie qu'en dimension 1. La démonstration de l'estimation d'erreur en norme L^2 (1.57) se prouve alors directement à partir de l'estimation en norme H^1_0 discrète, grâce à une "inégalité de Poincaré discrète", équivalent discret de la célèbre inégalité de Poincaré continue, que l'on rappelle (voir (1.60) pour la dimension 1).

Lemme 1.31 — Inégalité de Poincaré. Soit Ω un ouvert borné de $\mathbb{R}^N$ et $u \in H^1_0(\Omega)$, alors $\|u\|_{L^2(\Omega)} \leqslant \mathrm{diam}(\Omega)\|Du\|_{L^2(\Omega)}\}$.

11. Henri Poincaré (1854–1912) est un mathématicien, physicien et philosophe français. C'est probablement l'un des plus grands hommes de science de cette époque.

1.3.3 Prise en compte de discontinuités

On considère ici un barreau conducteur constitué de deux matériaux de conductivités λ_1 et λ_2 différentes et dont les extrémités sont plongées dans de la glace. On suppose que le barreau est de longueur 1, que le matériau de conductivité λ_1 (resp. λ_2) occupe le domaine $\Omega_1 =]0\,;1/2[$ (resp. $\Omega_2 =]1/2, 1[$). Le problème de conduction de la chaleur s'écrit alors :

$$\begin{cases} (-\lambda_1(x)u')' = f(x) & x \in]0\,;1/2[\\ (-\lambda_2(x)u')' = f(x) & x \in]1/2\,;1[\\ u(0) = u(1) = 0, \\ -(\lambda_1 u')(1/2) = -(\lambda_2 u')(1/2) \end{cases} \tag{1.61}$$

Remarque 1.32 La dernière égalité traduit la conservation du flux de chaleur à l'interface $x = 1/2$. On peut noter que comme λ est discontinu en ce point, la dérivée u' le sera forcément elle aussi.

On choisit de discrétiser le problème par volumes finis. On se donne un maillage volumes finis comme défini par la définition 1.21, en choisissant les mailles telles que la discontinuité de λ soit située sur un interface de deux mailles qu'on note K_k et K_{k+1}. On a donc, avec les notations du paragraphe (1.1.1) $x_{k+1/2} = 0.5$. La discrétisation par volumes finis s'écrit alors

$$F_{i+1/2} - F_{i-1/2} = h_i f_i \quad i = 1, \ldots, N,$$

où les flux numériques $F_{i+1/2}$ sont donnés par

$$F_{i+1/2} = -\lambda_* \frac{u_{i+1} - u_i}{h_{i+1/2}} \text{ avec } \lambda_* = \begin{cases} \lambda_1 & \text{si } x_{i+1/2} > 1/2, \\ \lambda_2 & \text{si } x_{i+1/2} < 1/2. \end{cases}$$

Il ne reste donc plus qu'à calculer le flux $F_{k+1/2}$, approximation de $(\lambda u')(x_{k+1/2})$ (avec $x_{k+1/2} = 1/2$). On introduit pour cela une inconnue auxiliaire $u_{k+1/2}$ que l'on pourra éliminer plus tard et on écrit une discrétisation du flux de part et d'autre de l'interface :

$$F_{k+1/2} = -\lambda_1 \frac{u_{k+1/2} - u_k}{\frac{h_k}{2}}, \tag{1.62}$$

$$F_{k+1/2} = -\lambda_2 \frac{u_{k+1} - u_{k+1/2}}{\frac{h_{k+1}}{2}}. \tag{1.63}$$

L'élimination (et le calcul) de l'inconnue se fait en écrivant la conservation du flux numérique :

$$-\lambda_1 \frac{u_{k+1/2} - u_k}{h_k} = -\lambda_2 \frac{u_{k+1} - u_{k+1/2}}{h_{k+1}}$$

On en déduit la valeur de $u_{k+1/2}$

$$u_{k+1/2} = \frac{\frac{\lambda_1}{h_k} u_k + \frac{\lambda_2}{h_{k+1}} u_{k+1}}{\frac{\lambda_1}{h_k} + \frac{\lambda_2}{h_{k+1}}}$$

On remplace $u_{k+1/2}$ par cette valeur dans l'expression du flux $F_{k+1/2}$ et on obtient :

$$F_{k+1/2} = -\frac{2\lambda_1\ \lambda_2}{h_k\lambda_2 + h_{k+1}\lambda_1}(u_{k+1} - u_k).$$

Si le maillage est uniforme, on obtient

$$F_{k+1/2} = -\frac{2\lambda_1\lambda_2}{\lambda_1 + \lambda_2}\left(\frac{u_{k+1} - u_k}{h}\right).$$

Le flux est donc calculé en faisant intervenir la moyenne harmonique des conductivités λ_1 et λ_2. Notons que lorsque $\lambda_1 = \lambda_2$, on retrouve la formule habituelle du flux.

1.4 Diffusion bidimensionnelle

1.4.1 Différences finies

On considère maintenant le problème de diffusion dans un ouvert Ω de $\mathbb{R}^2$:

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \tag{1.64}$$

Le problème est bien posé au sens où si $f \in \mathcal{C}^1(\overline{\Omega})$, alors il existe une unique solution $u \in \mathcal{C}(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$, solution de (1.64). Si $f \in L^2(\Omega)$et si Ω est convexe (ou à bord régulier) alors il existe une unique fonction $u \in H^2(\Omega)$ au sens faible [12] de (1.64), c'est-à-dire qui vérifie :

$$\begin{cases} u \in H_0^1(\Omega), \\ \displaystyle\int_\Omega \nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x \qquad \forall v \in H_0^1(\Omega). \end{cases}$$

On peut montrer que si $u \in \mathcal{C}^2(\overline{\Omega})$, alors u est solution de (1.64) si et seulement si u est solution faible de (1.64). Pour discrétiser le problème, on se donne un certain nombre de points, alignés dans les directions x et y, comme représentés sur la figure 1.4 (on prend un pas de maillage uniforme et égal à h). Certains de ces points sont à l'intérieur du domaine Ω, d'autres sont situés sur la frontière $\partial\Omega$. Comme en une dimension d'espace, les inconnues discrètes sont associées aux nœuds

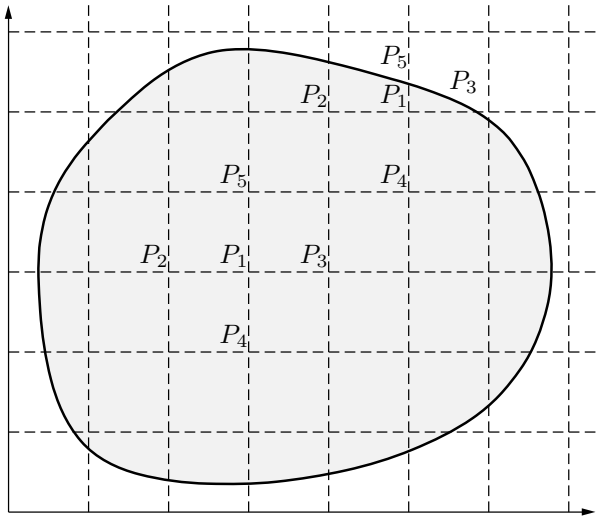


Figure 1.4 – Discrétisation par différences finies en dimension 2 d'espace

du maillage. On note $\{P_i, i \in I\}$ les points de discrétisation et on écrit l'équation aux dérivées partielles en ces points :

$$-\Delta u(P_i) - \frac{\partial^2 u}{\partial x^2}(P_i) - \frac{\partial^2 u}{\partial y^2}(P_i) = f(P_i).$$

Dans le cas de points points "vraiment intérieurs", tel que le point P_1 sur la figure 1.4, *i.e.* dont tous les points voisins sont situés à l'intérieur de Ω, les quotients différentiels

$$\frac{2u(P_1) - u(P_2) - u(P_3)}{h^2} \quad \text{et} \quad \frac{2u(P_1) - u(P_5) - u(P_4)}{h^2}$$

sont des approximations consistantes à l'ordre 2 de $-\partial_1^2 u(P_1)$ et $-\partial_2^2 u(P_1)$. Par contre, pour un point "proche" du bord tel que le point $\widetilde{P}_1$, les mêmes approximations (avec les points $\widetilde{P}_2$, $\widetilde{P}_3$, $\widetilde{P}_4$ et $\widetilde{P}_5$) ne seront que d'ordre 1 en raison des différences de distance entre les points (faire les développements de Taylor pour s'en convaincre).

12. Par définition, $H^2(\Omega)$ est l'ensemble des fonctions de $L^2(\Omega)$ qui admet des dérivées faibles jusqu'à l'ordre 2 dans $L^2(\Omega)$.

Une telle discrétisation amène à un système linéaire $A_h U_h = b_h$ où la structure de A_h, en particulier sa *largeur de bande*, c'est-à-dire le nombre de diagonales non nulles, dépend de la numérotation des nœuds. On peut montrer que la matrice A_h est une IP-matrice (au sens de la définition 1.7 et que le schéma est stable. De la consistance et la stabilité, on déduit, comme en une dimension d'espace, la convergence du schéma.

1.4.2 Volumes finis

Problème modèle

On considère le problème modèle suivant (qui peut modéliser par exemple la conduction de la chaleur) :

$$-\operatorname{div}(\lambda_i \nabla u(x)) = f(x) \qquad x \in \Omega_i \quad i = 1, 2 \tag{1.65}$$

où l'opérateur divergence div est défini par (1), $\lambda_1 > 0$ et $\lambda_2 > 0$ sont les conductivités thermiques dans les domaines Ω_1 et Ω_2 avec $\Omega_1 =]0\,;1[\times]0\,;1[$ et $\Omega_2 =]0\,;1[\times]1\,;2[$. On appelle $\Gamma_1 =]0\,;1[\times\{0\}$, $\Gamma_2 = \{1\}\times]0\,;2[$, $\Gamma_3 =]0\,;1[\times\{2\}$ et $\Gamma_4 = \{0\}\times]0\,;2[$ les frontières extérieures de Ω et on note $I =]0\,;1[\times\{1\}$ l'interface entre Ω_1 et Ω_2 (voir figure 1.5). Dans la suite, on notera λ la conductivité thermique sur Ω avec $\lambda|_{\Omega_i} = \lambda_i$, $i = 1, 2$. On va considérer plusieurs types de conditions aux limites

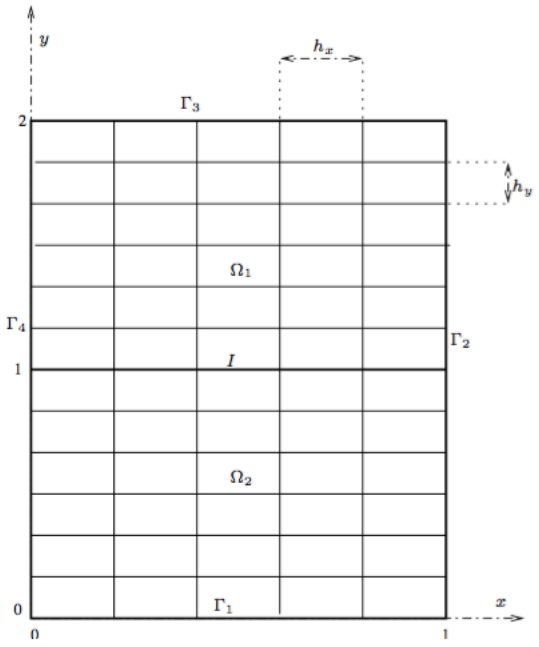


Figure 1.5 – Domaine d'étude

en essayant d'expliquer leur sens physique. On rappelle que le flux de chaleur par diffusion $\mathbf{q}$ est donné par la loi de Fourier $\mathbf{q} = -\lambda \nabla u \cdot \boldsymbol{n}$ où $\boldsymbol{n}$ est le vecteur normal unitaire à la surface à travers laquelle on calcule le flux.

Conditions aux limites de type Fourier sur $\Gamma_1 \cup \Gamma_3$ On suppose qu'il existe un transfert thermique entre les parois Γ_1 et Γ_3 et l'extérieur. Ce transfert est décrit par la condition de

Fourier [13] qui exprime que le flux transféré est proportionnel à la différence de température entre l'extérieur et l'intérieur :

$$-\lambda \nabla u \cdot \boldsymbol{n}(x) = \alpha(u(x) - u_{\text{ext}}), \qquad \forall x \in \Gamma_1 \cup \Gamma_3. \tag{1.66}$$

où $\alpha > 0$ est le coefficient de transfert thermique, $\boldsymbol{n}$, le vecteur unitaire normal à $\partial\Omega$ extérieur à Ω et u_{ext} est la température extérieure (donnée).

Conditions aux limites de type Neumann sur Γ_2 On suppose que la paroi Γ_2 est parfaitement isolée et que le flux de chaleur à travers cette paroi est donc nul. Ceci se traduit par une condition dite de *Neumann homogène* :

$$-\lambda \nabla u \cdot \boldsymbol{n} = 0, \qquad \forall x \in \Gamma_2. \tag{1.67}$$

Conditions aux limites de type Dirichlet sur Γ_4 Sur la paroi Γ_4, on suppose que la température est fixée. Ceci est une condition assez difficile à obtenir expérimentalement pour un problème de type chaleur, mais qu'on peut rencontrer dans d'autres problèmes pratiques.

$$u(x) = g(x), \qquad \forall x \in \Gamma_4. \tag{1.68}$$

Conditions sur l'interface I On suppose que l'interface I est par exemple le siège d'une réaction chimique surfacique θ qui provoque un dégagement de chaleur surfacique. On a donc un saut du flux de chaleur au travers de l'interface I. Ceci se traduit par la condition de saut suivante :

$$-\lambda_1 \nabla u_1(x) \cdot \boldsymbol{n}_1 - \lambda_2 \nabla u_2(x) \cdot \boldsymbol{n}_2 = \theta(x) \qquad x \in I \tag{1.69}$$

où $\boldsymbol{n}_i$ désigne le vecteur unitaire normal à I et extérieur à Ω_i et θ est une fonction donnée.

Discrétisation par volumes finis

On se donne un maillage *admissible* $\mathcal{T}$ de Ω

$$\bar{\Omega} = \bigcup_{K \in \mathcal{T}} \bar{K}.$$

Par *admissible*, on entend un maillage tel qu'il existe des points $(x_K)_{K\in\mathcal{T}}$ situés dans les mailles, tels que chaque segment $x_K x_L$ soit orthogonal à l'arête $K|L$ séparant la maille K de la maille L, comme visible sur la figure 1.6. Cette condition permet d'obtenir une approximation consistante du

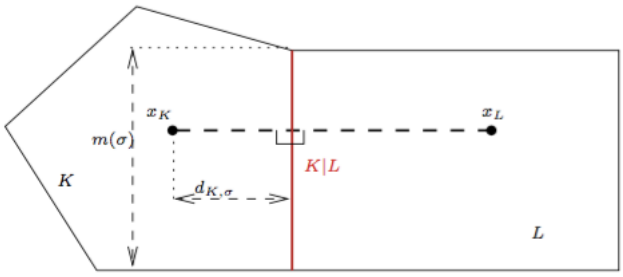


Figure 1.6 – Condition d'orthogonalité pour un maillage volumes finis

13. Robin dans la littérature anglo-saxonne

flux de diffusion (c'est-à-dire de la dérivée normale sur l'arête $K|L$) avec deux inconnues discrètes (voir la proposition 1.33 ci-dessous). Dans le cas présent, le domaine representé sur la figure 1.5 étant rectangulaire, cette condition est particulièrement facile à vérifier en prenant un maillage rectangulaire. Par souci de simplicité, on prendra ce maillage uniforme et on notera $h_x = 1/n$ le pas de discrétisation dans la direction x et $h_y = 1/p$ le pas de discrétisation dans la direction y. Le maillage est donc choisi de telle sorte que l'interface I coïncide avec un ensemble d'arêtes du maillage qu'on notera $\mathcal{E}_I$. On a donc

$$\bar{I} = \bigcup_{\sigma \in \mathcal{E}_I} \bar{\sigma},$$

où le signe $\bar{}$ désigne l'adhérence de l'ensemble. On se donne ensuite des inconnues discrètes $(u_K)_{K\in\mathcal{T}}$ associées aux mailles et $(u_\sigma)_{\sigma\in\mathcal{E}}$ associées aux arêtes.

Pour obtenir le schéma volumes finis, on commence par établir les bilans par maille en intégrant l'équation sur chaque maille K (notons que ceci est faisable en raison du fait que l'équation est sous forme conservative, c'est-à-dire sous la forme $-\operatorname{div}(\text{flux}) = f$). On obtient donc :

$$\int_K -\operatorname{div}(\lambda_i \nabla u(x))\,\mathrm{d}x = \int_K f(x)\,\mathrm{d}x,$$

soit encore, par la formule de Stokes,

$$\int_{\partial K} -\lambda_i \nabla u(x) \cdot \boldsymbol{n}(x) \mathrm{d}\gamma(x) = m(K) f_K,$$

où $\boldsymbol{n}$ est le vecteur unitaire normal à $\partial\Omega$ extérieur à Ω et $\mathrm{d}\gamma$ désigne le symbole d'intégration sur la frontière. On décompose ensuite le bord de chaque maille K en arêtes du maillage :

$$\partial K = \bigcup_{\sigma \in \mathcal{E}_K} \bar{\sigma},$$

où $\mathcal{E}_K$ représente l'ensemble des arêtes de K. On obtient alors :

$$\sum_{\sigma\in\mathcal{E}_K} \int_\sigma -\lambda_i \nabla u \cdot \boldsymbol{n}_{K,\sigma}\ \mathrm{d}\gamma(x) = m(K) f_K$$

où $\boldsymbol{n}_{K,\sigma}$ est le vecteur unitaire normal à σ extérieur à K. On écrit alors le schéma numérique :

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} = m(K) f_K,$$

où $F_{K,\sigma}$ est le flux numérique à travers σ, qui approche le flux exact

$$\bar{F}_{K,\sigma} = \int_\sigma -\lambda_i \nabla u \cdot \boldsymbol{n}_{K,\sigma}\ \mathrm{d}\gamma(x). \tag{1.70}$$

Pour obtenir le schéma numérique, il nous reste à exprimer le flux numérique $F_{K,\sigma}$ en fonction des inconnues discrètes $(u_K)_{K\in\mathcal{T}}$ associées aux mailles et $(u_\sigma)_{\sigma\in\mathcal{E}}$ associées aux arêtes (ces dernières seront ensuite éliminées) :

$$F_{K,\sigma} = -\lambda_i \frac{u_\sigma - u_K}{d_{K,\sigma}} m(\sigma) \tag{1.71}$$

où $d_{K,\sigma}$ est la distance du point x_K à l'arête σ et $m(\sigma)$ est la longueur de l'arête σ (voir figure 1.6. L'équation associée à l'inconnue u_K est donc :

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} = m(K) f_K$$

On a ainsi obtenu autant d'équations que de mailles. Il nous reste maintenant à écrire une équation pour chaque arête, afin d'obtenir autant d'équations que d'inconnues.

En ce qui concerne les arêtes intérieures, on écrit la conservativité du flux, ce qui nous permettra d'éliminer les inconnues associées aux arêtes internes. Soit $\sigma = K|L \subset \Omega_i$ On a alors :

$$F_{K,\sigma} = -F_{L,\sigma} \tag{1.72}$$

On vérifiera par le calcul (voir l'exercice 24) que, après élimination de u_σ, ceci donne

$$F_{K,\sigma} = -F_{L,\sigma} = \lambda_i \frac{m(\sigma)}{d_\sigma}(u_K - u_L) \tag{1.73}$$

où $d_\sigma = d(x_K, x_L)$, et $\mathrm{d}(\,\cdot\,,\,\cdot\,)$ désigne la distance euclidienne.

Définition 1.33 — Consistance du flux. On appelle erreur de consistance associée au flux (1.71) l'expression :

$$R_{K,\sigma} = \bar{F}_{K,\sigma} - F^*_{K,\sigma}$$

où $\bar{F}_{K,\sigma}$ est le flux exact défini par (1.70) et

$$F^*_{K,\sigma} = -\lambda_i \frac{u(x_\sigma) - u(x_K)}{d_{K,\sigma}} m(\sigma),$$

où x_σ est l'intersection de σ avec l'arête $K|L$, u la solution exacte. On dit que le flux numérique donné par l'expression (1.71) est *consistant* si

$$\lim_{h(\mathcal{T})\to 0} \max_{\substack{K\in\mathcal{T}\\ \sigma\in\mathcal{E}(K)}} |R_{K,\sigma}| = 0$$

où $h(\mathcal{T})$ est le pas du maillage, *i.e.* $h(\mathcal{T}) = \max_{K\in\mathcal{T}} \mathrm{diam}(K)$ avec $\mathrm{diam}(K) = \sup_{(x,y)\in K^2} \mathrm{d}(x,y)$. On vérifie facilement que si u est suffisamment régulière et si le segment $x_K x_L$ est colinéaire au vecteur normal $\boldsymbol{n}$ alors le flux numérique est consistant. Cette propriété, alliée a la propriété de conservativité des flux, permet de démontrer la convergence du schéma, comme on l'a fait dans le cas unidimensionnel.

Remarque 1.34 — Cas du maillage cartésien de la figure 1.5. Dans le cas du maillage carésien considéré pour notre problème, il est naturel de choisir les points x_K comme les centres de gravité des mailles. Comme le maillage est uniforme, on a donc $d_{K,\sigma} = h_x/2$ (respectivement $h_y/2$) et $|\sigma| = h_y$ (respectivement $|\sigma| = h_x$) pour une arête σ verticale (respectivement horizontale).

Écrivons maintenant la discrétisation des conditions aux limites et interface :

Condition de Neumann sur Γ_2 Sur Γ_2, on a la condition de Neumann (1.67) $\lambda_i \nabla u \cdot \boldsymbol{n} = 0$, qu'on discrétise par $\sigma \in \mathcal{E}_K$ et $\sigma \subset \Gamma_2, F_{K,\sigma} = 0$.

Condition de Dirichlet sur Γ_4 La discrétisation de la condition de Dirichlet (1.68) peut s'effectuer de la manière suivante :

$$u_\sigma = \frac{1}{m(\sigma)} \int_\sigma g(y)\ \mathrm{d}\gamma(y).$$

L'expression du flux numérique est alors :

$$F_{K,\sigma} = -\lambda_i \frac{u_\sigma - u_K}{d_{K,\sigma}} m(\sigma)$$

Condition de Fourier sur $\Gamma_1 \cup \Gamma_3$ Sur $\Gamma_1 \cup \Gamma_3$ on a la condition de Fourier (1.66) :

$$-\lambda_i \nabla u \cdot \boldsymbol{n} = \alpha(u(x) - u_{\mathrm{ext}}) \qquad \forall x \in \Gamma_1 \cup \Gamma_3$$

qu'on discrétise par

$$F_{K,\sigma} = -m(\sigma)\lambda_i \frac{u_\sigma - u_K}{d_{K,\sigma}} = m(\sigma)\alpha(u_\sigma - u_{\mathrm{ext}}) \text{ pour } \sigma \subset \Gamma_1 \cup \Gamma_3.$$

Après élimination de u_σ [exercice 24], on obtient :

$$F_{K,\sigma} = \frac{\alpha\lambda_i m(\sigma)}{\lambda_i + \alpha d_{K,\sigma}}(u_K - u_{\mathrm{ext}}) \tag{1.74}$$

Condition de saut pour le flux sur I Si $\sigma = K|L \in \mathcal{E}_I$, la discrétisation de la condition de saut (1.69) se discrétise facilement en écrivant :

$$F_{K,\sigma} + F_{L,\sigma} = \theta_\sigma \quad \text{avec} \quad \theta_\sigma = \frac{1}{|\sigma|}\int_\sigma \theta(x)\ \mathrm{d}\gamma(x). \tag{1.75}$$

Après élimination de l'inconnue u_σ [exercice 24], on obtient

$$F_{K,\sigma} = \frac{\lambda_1 m(\sigma)}{\lambda_1 d_{L,\sigma} + \lambda_2 d_{K,\sigma}}\left[\lambda_2(u_K - u_L) + d_{L,\sigma}\theta_\sigma\right] \tag{1.76}$$

On a ainsi éliminé toutes les inconnues u_σ, ce qui permet d'obtenir un système linéaire dont les inconnues sont les valeurs $(u_K)_{K\in\mathcal{T}}$.

Remarque 1.35 — Implantation informatique de la méthode. Lors de l'implantation informatique, la matrice du système linéaire est construite *par arête* (contrairement à une matrice éléments finis, dont nous verrons plus tard la construction *par élément*), c'est-à-dire que pour chaque arête, on additionne la contribution du flux au coefficient de la matrice correspondant à l'équation et à l'inconnue concernées.

1.5 Exercices

1.5.1 Énoncés

Exercice 1 — Différences et volumes finis avec conditions de Dirichlet non homogènes.
corrigé p.48
On considère le problème :

$$\begin{cases} -u''(x) + \sin(u(x)) = f(x) & x \in]0\,;1[\\ u(0) = a \\ u(1) = b \end{cases} \tag{1.77}$$

1. Écrire les schémas de différences finies et volumes finis avec pas constant pour le problème (1.77). Pour le schéma volumes finis, on utilisera l'approximation

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \sin(u(x))dx \simeq (x_{i+1/2} - x_{i-1/2})\sin(u(x_i))$$

2. Comparer les schémas ainsi obtenus.

Exercice 2 — Différences et volumes finis avec conditions mixtes. On considère le problème :

$$\begin{cases} -u''(x) = f(x) & x \in]0\,;1[\\ u(0) - u'(0) = a \\ u'(1) = b \end{cases} \tag{1.78}$$

Écrire les schémas de différences finies et volumes finis avec pas constant pour le problème (1.78), et comparer les schémas ainsi obtenus.

Exercice 3 — Différences finies et principe du maximum.
corrigé p.48
On considère le problème :

$$-u''(x) + c(x)u(x) = f(x) \qquad 0 < x < 1, \tag{1.79a}$$
$$u(0) = a, \tag{1.79b}$$
$$u(1) = b, \tag{1.79c}$$

où $c \in C([0\,;1], \mathbb{R}_+)$, $f \in C([0\,;1], \mathbb{R})$ et $(a, b) \in \mathbb{R}^2$.

1. Donner la discrétisation par différences finies de ce problème. On appelle U_h la solution approchée c'est-à-dire $U_h = (u_1, \ldots, u_N)$ où u_i est l'inconnue discrète en $x_i = ih$ où $h = \frac{1}{N+1}$ le pas du maillage.

2. On suppose ici que $c = 0$ et $f \geqslant 0$. Montrer que $u_i \geqslant \min(a, b)$ pour tout $i = 1, \ldots, N$.

Exercice 4 — Équation de diffusion réaction. On s'intéresse au problème elliptique unidimensionnel suivant :

$$\begin{cases} -u''(x) + 2u(x) = x & x \in]0\,;1[\\ u(0) = 1 \\ u'(1) + u(1) = 0 \end{cases} \tag{1.80}$$

1. Écrire une discrétisation de (1.80) par différences finies pour un maillage uniforme. écrire le système linéaire obtenu.

2. Écrire une discrétisation de (1.80) par volumes finis de (1.80) pour un maillage uniforme. Écrire le système linéaire obtenu.

Remarque 1.36 — Forme conservative et non conservative. Les deux exercices suivants concernent la discrétisation d'une équation de transport-diffusion sous forme non-conservative (exercice 5) puis conservative (exercice 6). On a déjà vu dans le cours qu'en une dimension d'espace, le terme de diffusion unidimensionnel est de la forme $-u''$ (tout du moins dans le cas d'un matériau homogène de conductivité constante). On appelle terme de transport un terme de la forme $v(x)u'(x)$, dite *non conservative*, ou $(v(x)u(x))'$, dite *conservative*, où v est la *vitesse de transport* donnée et u l'inconnue, qui est la quantité transportée (une concentration de polluant, par exemple). Remarquez d'abord que si la vitesse v est constante, les deux formes sont identiques, puisque $(v(x)u(x))' = v'(x)u(x) + v(x)u'(x) = v(x)u'(x)$. La deuxième forme est dite conservative car elle est obtenue à partir de l'écriture de la conservation de la masse (par exemple) sur un petit élement $x + \delta x$, en passant à la limite lorsque δx tend vers 0. La première forme, non conservative, apparaît dans des modèles de mécanique de fluides (écoulements compressibles polyphasiques, par exemple).

corrigé p.49

Exercice 5 — Équation de transport-diffusion sous forme non-conservative.
Soient $v \in C([0\,;1], \mathbb{R}_+)$ et $a, b \in \mathbb{R}$. On considère le problème suivant :

$$\begin{cases} -u''(x) + v(x)\partial_x u(x) = 0 \quad x \in]0\,;1[\\ u(0) = a \\ u(1) = b. \end{cases} \tag{1.81}$$

On admettra qu'il existe une unique solution $u \in C([0\,;1], \mathbb{R}) \cap \mathcal{C}^2(]0\,;1[, \mathbb{R})$ à ce problème. On cherche à approcher cette solution par une méthode de différences finies. On se donne un pas de maillage $h = 1/N + 1$ uniforme, des inconnues discrètes $u_1, \dots, u_N$ censées approcher les valeurs $u(x_1), \dots, u(x_N)$. On considère le schéma aux différences finies suivant :

$$\begin{cases} \dfrac{2u_i - u_{i+1} - u_{i-1}}{h^2} + \dfrac{v_i(u_i - u_{i-1})}{h} = 0 \qquad i = 1, \dots, N \\ u_0 = a, \\ u_{N+1} = b, \end{cases} \tag{1.82}$$

où $v_i = v(x_i)$ pour $i = 1, \dots, N$. Noter que le terme de transport $v(x_i)u'(x_i)$ est approché par $v(x_i)(u(x_{i+1/2}) - u(x_{i-1/2}))/h$. Comme la vitesse v_i est positive ou nulle, on choisit d'approcher $u(x_{i+1/2})$ par la valeur *amont*, c'est-à-dire $u(x_i)$, d'où le schéma (1.82).

1. Montrer que le système (1.82) s'écrit sous la forme $MU = b$ avec $U = (u_1, \dots, u_N)$, $b \in \mathbb{R}^N$ et M est une matrice telle que :

- **(a)** $MU \geqslant 0 \Rightarrow U \geqslant 0$ (les inégalités s'entendent composante par composante) ;
- **(b)** M est inversible ;
- **(c)** Si U est solution de $MU = b$ alors $\min(a, b) \leqslant u_i \leqslant \max(a, b)$.

2. Montrer que M est une M-matrice, c'est-à-dire que M vérifie :

- **(a)** $m_{i,i} > 0$ pour $i = 1, \dots, n$;
- **(b)** $m_{i,j} \leqslant 0$ pour $i, j = 1, \dots, n$, $i \neq j$;
- **(c)** M est inversible ;
- **(d)** $M^{-1} \geqslant 0$.

Exercice 6 — Équation de transport-diffusion sous forme conservative. Il est conseillé d'étudier l'exercice 5 avant celui-ci.

Soit $v \in C([0\,;1],\mathbb{R}_+) \cap C^1(]0\,;1[,\mathbb{R})$, et on considère le problème :

$$\begin{cases} -u''(x) + (vu)'(x) = 0, \quad x \in]0\,;1[, \\ u(0) = a, \\ u(1) = b. \end{cases} \tag{1.83}$$

On admettra qu'il existe une unique solution $u \in C([0\,;1],\mathbb{R}) \cap \mathcal{C}^2(]0\,;1[,\mathbb{R})$ à ce problème. On cherche ici encore à approcher cette solution par une méthode de différences finies. On se donne un pas de maillage $h = 1/(N+1)$ uniforme, des inconnues discrètes $u_1, \ldots, u_N$ censées approcher les valeurs $u(x_1), \ldots, u(x_N)$. On considère le schéma aux différences finies suivant :

$$\begin{cases} \dfrac{2u_i - u_{i+1} - u_{i-1}}{h^2} + \dfrac{v_{i+\frac{1}{2}}u_i - v_{i-\frac{1}{2}}u_{i-1}}{h} = 0 \quad i = 1, \ldots, N \\ u_0 = a \\ u_{N+1} = b \end{cases} \tag{1.84}$$

où $v_{i+1/2} = v((x_i + x_{i+1})/2)$, pour $i = 0, \ldots, N$. Noter que le terme de convection $(vu)'(x_i)$ peut être approché par :

$$\frac{v(x_{i+1/2})u(x_{i+1/2}) - v(x_{i-1/2})u(x_{i-1/2})}{h}.$$

Comme $v(x_{i+1/2}) \geqslant 0$, on choisit d'approcher $u(x_{i+1/2})$ par la valeur *amont*, c'est-à-dire $u(x_i)$. C'est une valeur amont dans le sens où elle est choisie en amont de l'écoulement, si l'on suppose que v est la vitesse de l'écoulement. On dit que le schéma est *décentré amont*.

1. Montrer que le système (1.84) s'écrit sous la forme $MU = B$ avec $U = (u_1, \ldots, u_N)$, , $B \in \mathbb{R}^N$;

2. Pour $U = (u_1, \ldots, u_N)$ et $W = (w_1, \ldots, w_N) \in \mathbb{R}^N$, calculer $MU \cdot W$, et en déduire l'expression de $(M^tW)_i$, pour $i = 1, \ldots, N$ (on distinguera les cas $i = 2, \ldots, N-1$, $i = 1$ et $i = N$;

3. Soit $W \in \mathbb{R}^N$:

(a) montrer que si $M^tW \geqslant 0$ alors $W \geqslant 0$;

(b) en déduire que si $U \in \mathbb{R}^N$ est tel que $MU \geqslant 0$ alors $U \geqslant 0$.

4. Montrer que M est une M-matrice ;

5. Montrer que U solution de (1.84) peut ne pas vérifier $\min(a,b) \leqslant u_i \leqslant \max(a,b)$.

corrigé p.50

Exercice 7 — Conditionnement efficace. Soit $f \in C([0\,;1])$. Soit $N \in \mathbb{N}^\star$, N impair. On pose $h = 1/(N+1)$. Soit A la matrice définie par (1.38), issue d'une discrétisation par différences finies (vue en cours) du problème (1.1)-(1.2). Pour $u \in \mathbb{R}^N$, on note $u_1, \ldots, u_N$ les composantes de u. Pour $u \in \mathbb{R}^N$, on dit que $u \geqslant 0$ si $u_i \geqslant 0$ pour tout $i \in \{1, \ldots, N\}$. Pour u, $v \in \mathbb{R}^N$, on note $u \cdot v = \sum_{i=1}^N u_i v_i$. On munit $\mathbb{R}^N$ de la norme suivante : pour $u \in \mathbb{R}^N$, $\|u\| = \max\{|u_i|,\ i \in \{1, \ldots, N\}\}$. On munit alors $\mathcal{M}_N(\mathbb{R})$ de la norme induite, également notée $\|\cdot\|$, c'est-à-dire $\|B\| = \max\{\|Bu\|,\ u \in \mathbb{R}^N \text{ tel que } \|u\| = 1\}$, pour tout $B \in \mathcal{M}_N(\mathbb{R})$.

Partie I Conditionnement de la matrice et borne sur l'erreur relative

1. Existence et positivité de A^{-1}. Soient $b \in \mathbb{R}^N$ et $u \in \mathbb{R}^N$ tel que $Au = b$. Remarquer que $Au = b$ peut s'écrire :

$$\begin{cases} \dfrac{u_i - u_{i-1}}{h^2} + \dfrac{u_i - u_{i+1}}{h^2} = b_i \quad \forall i \in \{1, \ldots, N\} \\ u_0 = 0 \\ u_{N+1} = 0 \end{cases} \tag{1.85}$$

Montrer que $b \geqslant 0 \Rightarrow u \geqslant 0$. [On pourra considérer $p \in \{0, \ldots, N+1\}$ tel que $u_p = \min\{u_j,\ j \in \{0, \ldots, N+1\}\}$.] En déduire que A est inversible.

2. On considère la fonction $\varphi \in C([0\,;1],\mathbb{R})$ définie par $\varphi(x) = (1/2)x(1-x)$ pour tout $x \in [0\,;1]$. On définit alors $\phi \in \mathbb{R}^N$ par $\phi_i = \phi(ih)$ pour tout $i \in \{1,\dots,N\}$. Montrer que $(A\phi)_i = 1$ pour tout $i \in \{1,\dots,N\}$.

3. Calcul de $\|A^{-1}\|$ — Soient $b \in \mathbb{R}^N$ et $u \in \mathbb{R}^N$ tels que $Au = b$. Montrer que $\|u\| \leqslant (1/8)\|b\|$ [Calculer $A(u \pm \|b\|\phi)$ avec ϕ défini à la question 2 et utiliser la question 1]. En déduire que $\|A^{-1}\| \leqslant 1/8$ puis montrer que $\|A^{-1}\| = 1/8$.

4. Calcul de $\|A\|$ — Montrer que $\|A\| = \frac{4}{h^2}$.

5. Conditionnement pour la norme $\|\cdot\|$ — Calculer $\|A^{-1}\|\|A\|$. Soient b, $\delta_b \in \mathbb{R}^N$. Soient u, $\delta_u \in \mathbb{R}^N$ tel que $Au = b$ et $A(u+\delta_u) = b + \delta_b$. Montrer que

$$\frac{\|\delta_u\|}{\|u\|} \leqslant \|A^{-1}\|\|A\|\frac{\|\delta_b\|}{\|b\|}$$

Montrer qu'un choix convenable de b et δ_b donne l'égalité dans l'inégalité précédente.

Partie II Borne réaliste sur l'erreur relative : Conditionnement efficace On se donne maintenant $f \in C([0\,;1],\mathbb{R})$ et on suppose (pour simplifier…) que $f(x) > 0$ pour tout $x \in]0\,;1[$. On prend alors, dans cette partie, $b_i = f(ih)$ pour tout $i \in \{1,\dots,N\}$. On considère aussi le vecteur φ défini à la question 2 de la partie I.

1. Montrer que $h\sum_{i=1}^N b_i\varphi_i \to \int_0^1 f(x)\phi(x)dx$ quand $N \to \infty$ et que $\sum_{i=1}^N b_i\varphi_i > 0$ pour tout N. En déduire qu'il existe $\alpha > 0$, ne dépendant que de f tel que $h\sum_{i=1}^N b_i\varphi_i \geqslant \beta$ pour tout $N \in \mathbb{N}^\star$.

2. Soit $u \in \mathbb{R}^N$ tel que $Au = b$. Montrer que $N\|u\| \geqslant \sum_{i=1}^N u_i = u \cdot A\varphi \geqslant \beta/h$ (avec β donné à la question 1). Soit $\delta_b \in \mathbb{R}^N$ et $\delta_u \in \mathbb{R}^N$ tel que $A(u+\delta_u) = b + \delta_b$. Montrer que

$$\frac{\|\delta_u\|}{\|u\|} \leqslant \frac{\|f\|_{L^\infty(]0\,;1[)}}{8\beta}\frac{\|\delta_b\|}{\|b\|}$$

3. Comparer $\|A^{-1}\|\,\|A\|$ et $\|f\|_{L^\infty(]0\,;1[)}/8\beta$ quand $N \to \infty$.

Exercice 8 — Conditionnement en réaction diffusion 1D. On s'intéresse au conditionnement pour la norme euclidienne de la matrice issue d'une discrétisation par différences finies du problème aux limites suivant :

$$\begin{cases} -u''(x) + u(x) = f(x) \quad x \in]0\,;1[\\ u(0) = 0 \\ u(1) = 0 \end{cases} \tag{1.86}$$

Soit $N \in \mathbb{N}^\star$. On note $U = (u_j)_{j=1\dots N}$ une valeur approchée de la solution u du problème (1.86) aux points $(j/(N+1))_{j=1\dots N}$.

1. Montrer que la discrétisation par différences finies de ce problème sur maillage uniforme de pas $h = 1/(N+1)$ consiste à chercher U comme solution du système linéaire $= AU = (f(j/(N+1)))_{j=1\dots N}$ où la matrice $A \in M_N(\mathbb{R})$ est définie par $A = (N+1)^2B + \mathrm{Id}$ où Id désigne la matrice identité et

$$B = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix}$$

2. Valeurs propres de la matrice B. — On rappelle que le problème aux valeurs propres

$$\begin{cases} -u''(x) = \lambda u(x) \quad x \in]0\,;1[\\ u(0) = 0 \\ u(1) = 0 \end{cases} \tag{1.87}$$

admet la famille $(\lambda_k, u_k)_{k\in\mathbb{N}^*}$, $\lambda_k = (k\pi)^2$ et $u_k(x) = \sin(k\pi x)$ comme solution. Montrer que les vecteurs $U_k = (u_k(j/(N+1)))_{j=1\dots N}$ sont des vecteurs propres de la matrice B. En déduire toutes les valeurs propres de la matrice B.

3. En déduire les valeurs propres de la matrice A.

4. En déduire le conditionnement pour la norme euclidienne de la matrice A.

Exercice 9 — Erreur de consistance. On considère la discrétisation à pas constant par le schéma aux différences finies symétrique à trois points du problème (1.1)-(1.2) avec $f \in C([0\,;1])$. Soit $N \in \mathbb{N}^\star$, N impair. On pose $h = 1/(N+1)$. On note u la solution exacte, $x_i = ih$, pour $i = 1, \dots, N$ les points de discrétisation, et $(u_i)_{i=1,\dots,N}$ la solution du système discrétisé.

1. Écrire le système linéaire obtenu.

2. Montrer que si f est constante, alors

$$\max_{1\leqslant i\leqslant N} |u_i - u(x_i)| = 0.$$

3. Soit N fixé, et $\max_{1\leqslant i\leqslant N} |u_i - u(x_i)| = 0$. A-t-on forcément f est constante sur $[0\,;1]$?

Exercice 10 — Problème elliptique 1D et discrétisation par différences finies. Soit $f \in \mathcal{C}^2([0\,;1])$. On s'intéresse au problème suivant [5] :

$$\begin{cases} -u''(x) + \dfrac{u'(x)}{1+x} = f(x) \quad x \in]0\,;1[\\ u(0) = a \\ u(1) = b \end{cases} \tag{1.88}$$

On admet que ce problème admet une et une seule solution u et on suppose que $u \in \mathcal{C}^4(]0\,;1[)$. On cherche une solution approchée de (1.88) par la méthode des différences finies. Soit $N \in \mathbb{N}^*$, et $h = 1/(N+1)$. On note u_i la valeur approchée recherchée de u au point ih, pour $i = 0, \cdots, N+1$. On utilise les approximations centrées les plus simples de u' et u'' aux points ih, $i = 1, \dots, N$. On pose $u_h = (u_1, \cdots, u_N)$.

1. Montrer que u_h est solution d'un système linéaire de la forme $A_h u_h = b_h$; donner A_h et b_h.

2. Montrer que le schéma numérique obtenu est consistant et donner une majoration de l'erreur de consistance (on rappelle que l'on a supposé $u \in \mathcal{C}^4$).

3. Soit $v \in \mathbb{R}^N$, montrer que $A_h v \geqslant 0 \Rightarrow v \geqslant 0$ (ceci s'entend composante par composante). Cette propriété s'appelle conservation de la positivité. En déduire que A_h est une IP-matrice.

4. On définit θ par

$$\theta(x) = -\frac{1}{2}(1+x)^2 \ln(1+x) + \frac{2}{3}(x^2+2x)\ln 2 \qquad x \in [0\,;1].$$

(a) Montrer qu'il existe $C \geqslant 0$, indépendante de h, tel que

$$\max_{1\leqslant i\leqslant N} \left| \frac{-\theta(x_{i-1}) + 2\theta(x_i) - \theta(x_{i+1})}{h^2} + \frac{\theta(x_{i+1}) - \theta(x_{i-1})}{2h(1+ih)} - 1 \right| \leqslant Ch^2.$$

(b) On pose $\theta_h = (\theta_1, \cdots, \theta_N)$. Montrer que $(A_h\theta_h)_i \geqslant 1 - Ch^2$, pour $i = 1, \dots, N$.

(c) Montrer qu'il existe $M \geqslant 0$ ne dépendant pas de h tel que $\|A_h^{-1}\|_\infty \leqslant M$.

5. Montrer la convergence, en un sens à définir, de u_h vers u.

6. Que peut on dire si $u \notin \mathcal{C}^4$ mais seulement $u \in \mathcal{C}^2$ ou $\mathcal{C}^3$?

7. On remplace dans (1.88) $\frac{1}{1+x}$ par $\alpha u'(x)$ avec α donné (par exemple $\alpha = 100$). On utilise pour approcher (1.88) le même principe que précédemment (approximations centrées de u' et u''. Que peut on dire sur la consistance, la stabilité, la convergence du schéma numérique ?

corrigé p.52 **Exercice 11 — Non consistance des volumes finis.**

1. Montrer que la discrétisation de l'opérateur $-u''$ par le schéma volumes finis n'est pas toujours consistante au sens des différences finies (voir remarque 1.22), *i.e.* que l'erreur de consistance définie par
$$R_i = \frac{1}{h_i}\Big(\frac{-u(x_{i+1})+u(x_i)}{h_{i+1/2}} + \frac{u(x_i)-u(x_{i-1})}{h_{i-1/2}}\Big) + u''(x_i)$$
ne tend pas toujours vers 0 lorsque h tend vers 0.
2. Ecrire un schéma de différences finies consistant aux points x_i, en supposant que x_i est le centre de la maille $K_i =]x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}$, pour $i = 1, \ldots, N$.

Exercice 12 — Condition de Fourier. On reprend ici la discrétisation du flux sortant en 1 pour le problème (1.1) comme la condition de Fourier (1.17) en 1 et la condition de Neumann (1.16) en 0. On reprend les notations du paragraphe 1.1.1, et pour simplifier on note $h = h_N$ et on suppose que x_N est au milieu de la N–ème maille, de sorte que $h_{N+1}/2 = h/2$. On modifie l'approximation de l'équation (1.8) : au lieu d'approcher $u'(1)$ par $b - \alpha u_N$, on introduit une onconnue auxiliaire $u_{N+1/2}$ censée approcher la valeur $u(1)$, on approche ensuite $u'(1)$ par $\frac{u_{N+1/2}-u_{N-1}}{h_{N-1/2}}$.

1. Montrer que par cette méthode, en éliminant l'inconnue auxiliaire $u_{N+1/2}$, on obtient comme N–ème équation discrète non plus (1.19) mais l'équation suivante :
$$F_{N+1/2} - F_{N-1/2} = h_N f_N \text{ avec } F_{N+1/2} = \frac{1}{1+\alpha h/2}(\alpha u_N - b) \text{ et } F_{N-1/2} = -\frac{u_N - u_{N-1}}{h_{N-1/2}} \tag{1.89}$$

2. Calculer l'erreur de consistance sur le flux approché $F_{N+1/2}$ en 1 dans le cas des discrétisations (1.19) et (1.89). Montrer qu'elle est d'ordre 1 dans le premier cas, et d'ordre 2 dans le second.

corrigé p.52

Exercice 13 — Consistance des flux.

1. Montrer que si $u \in \mathcal{C}^2([0\,;1])$, le flux défini par (1.43) est consistant d'ordre 1 dans le cas du maillage général décrit dans la figure 1.2 ;

2. Montrer que si $u \in \mathcal{C}^3([0\,;1])$, le flux défini par (1.43) est d'ordre 2 si $x_{i+1/2} = (x_{i+1} + x_i)/2$.

Exercice 14 — Convergence des VF par une méthode DF. *Cet exercice est inspiré de l'article de Forsyth, P.A. and P.H. Sammon, paru en 1988 (Appl. Num. Math. 4, 377-394) et intitulé : "Quadratic Convergence for Cell-Centered Grids."*

Soit $f \in C([0,1],\mathbb{R})$, on rappelle qu'il existe une unique solution $u \in C^2([0,1],\mathbb{R})$ du problème suivant :
$$-u''(x) = f(x), \quad x \in (0,1), \tag{1.90}$$
$$u(0) = 0, \tag{1.91}$$
$$u(1) = 0. \tag{1.92}$$

Dans la suite, cette solution exacte sera notée u. Afin de calculer une approximation numérique de u, on se donne un maillage, noté $\mathcal{T}$, de l'intervalle $]0,1[$, constitué de N cellules (ou volumes de contrôle), notées K_i, $i = 1, \ldots, N$, et des familles de points de $]0,1[$, notés x_i, $i = 0, \ldots, N+1$ et $x_{i+\frac{i}{2}}$, $i = 0, \ldots, N$ qui vérifient :
$$x_0 = x_{\frac{1}{2}} = 0 < x_1 < x_{\frac{3}{2}} < \cdots < x_{i-\frac{1}{2}} < x_i < x_{i+\frac{1}{2}} < \cdots < x_N < x_{N+\frac{1}{2}} = x_{N+1} = 1,$$
$$K_i =]x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[,$$
et on note
$$h_i = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}},\ i = 1, \ldots, N,\ h_{i+\frac{1}{2}} = x_{i+1} - x_i,\ i = 0, \ldots, N,$$
$$h = \max\{h_i,\ i = 1, \ldots, N\}.$$

Dans la suite, on va s'intéresser à la convergence d'un schéma numérique lorsque $h \to 0$, et on considère une famille de tels maillages tels qu'il existe $\alpha > 0$ tel que $\min\{h_i,\ i = 1, \ldots, N\} \geqslant \alpha h$ (on dit que les maillages sont quasi-uniformes).

1. *Schéma volumes finis*
   En notant $(u_i)_{i=1,\dots,N}$ les inconnues discrètes, expliquer rapidement comment une discrétisation par volumes finis pour l'approximation du problème (1.90)–(1.92) amène au schéma suivant

$$F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}} = \int_{K_i} f(x)\ \mathrm{d}x,\ i = 1,\dots,N \tag{1.93}$$

$$F_{i+\frac{1}{2}} = -\frac{u_{i+1}-u_i}{h_{i+\frac{1}{2}}},\ i = 0,\dots,N, \tag{1.94}$$

$$u_0 = u_{N+1} = 0, \tag{1.95}$$

2. *Consistance des flux*
   Montrer que le flux numérique $F_{i+\frac{1}{2}}$ défini par (1.94) approche de manière consistante (au sens des différences finies) le flux continu $-u'(x_{i+\frac{1}{2}})$, dans les deux cas suivants :
   (a) on considère que u_i est une approximation de $u(x_i)$, où x_i est un point quelconque de K_i ;
   (b) on considère que u_i est une approximation de la moyenne de u sur la maille i.

3. *Ecriture matricielle et monotonie de la matrice*
   (a) Montrer que le schéma volumes finis vu comme une approximation de $-u''(x_i)$ s'écrit sous la forme matricielle suivante.

$$AU = b, \tag{1.96}$$

   où $U = (u_1,\dots,u_N)^t$, $b = (b_1,\dots,b_N)^t$, avec (1.95) et avec A et b définis par

$$(AU)_i = \frac{1}{h_i}\Big(-\frac{u_{i+1}-u_i}{h_{i+\frac{1}{2}}} + \frac{u_i-u_{i-1}}{h_{i-\frac{1}{2}}}\Big),\ i = 1,\dots,N, \tag{1.97}$$

$$u_0 = u_{N+1} = 0, \tag{1.98}$$

$$b_i = \frac{1}{h_i}\int_{K_i} f(x)\ \mathrm{d}x,\ i = 1,\dots,N. \tag{1.99}$$

   (b) Montrer que la matrice A est une IP-matrice (ou matrice monotone), c.à.d. que pour tout $V \in \mathbb{R}^N$, $AV \geqslant 0 \Longrightarrow V \geqslant 0$ (au sens composante par composante). (On pourra introduire $i_0 = \min\{i \in \{0,\dots,N+1\}\ ;\ v_i = a\}$ avec $a = \min\{v_i, i = 0,\dots,N+1\}$). En déduire que le schéma (1.93)-(1.95) admet une solution unique.

On suppose dans toute la suite que x_i est le centre de la maille K_i. Soit $\bar{U} = (u(x_1),\dots,u(x_N))$. On pose

$$r = A(\bar{U} - U),\ \text{et, pour } i = 0,\dots,N,\ R_{i+\frac{1}{2}} = -\frac{u(x_{i+1})-u(x_i)}{h_{i+\frac{1}{2}}} + u'(x_{i+\frac{1}{2}}).$$

On définit $V = (v_1,\dots,v_N)^t \in \mathbb{R}^N$ par $v_i = \frac{h_i^2 u''(x_i)}{8}$, $i = 1,\dots,N$.

4. *Succédané de consistance*
   (a) Montrer que

$$R_{i+\frac{1}{2}} = -\frac{1}{4}(h_{i+1}-h_i)u''(x_{i+\frac{1}{2}}) + h\varepsilon_{i+\frac{1}{2}}(h),\ i = 1,\dots,N-1,$$

$$R_{\frac{1}{2}} = -\frac{1}{4}h_1 u''(0) + h\varepsilon_{\frac{1}{2}}(h),\ R_{N+\frac{1}{2}} = \frac{1}{4}h_N u''(1) + h\varepsilon_{N+\frac{1}{2}}(h),$$

   avec $\max_{i=1,\dots,N}\varepsilon_{i+\frac{1}{2}}(h) \to 0$ lorsque $h \to 0$, pour tout $i = 1,\dots,N$.

(b) Montrer que

$$-\frac{v_{i+1}-v_i}{h_{i+\frac{1}{2}}} = R_{i+\frac{1}{2}} + h\eta_{i+\frac{1}{2}}(h),\ i=1,\dots,N-1, \tag{1.100}$$

$$-\frac{2v_1}{h_1} = R_{\frac{1}{2}} + h\eta_{\frac{1}{2}}(h), \tag{1.101}$$

$$\frac{2v_N}{h_N} = R_{N+\frac{1}{2}} + h\eta_{N+\frac{1}{2}}(h). \tag{1.102}$$

avec $\max_{i=1,\dots,N}\eta_{i+\frac{1}{2}}(h)\to 0$ lorsque $h\to 0$.

En déduire que $r_i = (AV)_i + \eta_i(h)$ avec $\eta_i(h)\to 0$ lorsque $h\to 0$.

Soient φ la fonction définie sur $[0,1]$ par $\varphi(x)=\frac{1}{2}x(1-x)$, et $\Phi=(\phi_1,\dots,\phi_N)\in\mathbb{R}^N$ défini par $\phi_i=\varphi(x_i)$.

5. *Stabilité*

   (a) Montrer qu'il existe $W\in\mathbb{R}^N$ tel que $A(\Phi-W)=\mathbf{1}$ où $\mathbf{1}$ est le vecteur de $\mathbb{R}^N$ dont toutes les composantes sont égales à 1 et donner l'expression de W.

   (b) En déduire que

$$A(U-\|b\|_\infty(\Phi-W))\leqslant 0 \text{ et } A(U+\|b\|_\infty(\Phi-W))\geqslant 0. \tag{1.103}$$

   (c) Montrer que $\|A^{-1}\|_\infty\leqslant\frac{1}{4}$.

6. *Convergence*

   Déduire des questions 5. et 6. que $\|U-\bar{U}\|_\infty\to 0$ lorsque $h\to 0$.

7. *Estimation d'erreur*

   On suppose maintenant que $f\in C^1([0,1],\mathbb{R})$, et donc $u\in C^3([0,1],\mathbb{R})$. Quel est l'ordre du schéma (1.93)–(1.95) ?

Exercice 15 — Conditions aux limites de Neumann. On considère ici l'équation le problème de diffusion réaction avec conditions aux limites de Neumann homogènes (correspondant à une condition physique de flux nul sur le bord) :

$$\begin{cases} -u''(x)+cu(x)=f(x) & x\in\,]0\,;1[\\ u'(0)=0\\ u'(1)=0 \end{cases} \tag{1.104}$$

avec $c\in\mathbb{R}_+^*$ et $f\in\mathcal{C}([0\,;1])$. Donner la discrétisation de ce problème par :

1. différences finies ;

2. volumes finis.

Montrer que les matrices obtenues ne sont pas inversibles. Proposer une manière de faire en sorte que le problème soit bien posé, compatible avec ce qu'on connaît du problème continu.

corrigé p.53

Exercice 16 — Conditions aux limites de Fourier (ou Robin). On considère le problème :

$$\begin{cases} -u''(x)+cu(x)=f(x) & x\in\,]0\,;1[\\ u'(0)-\alpha(u-\widetilde{u})=0\\ u'(1)+\alpha(u-\widetilde{u})=0 \end{cases} \tag{1.105}$$

avec $c\in\mathbb{R}_+$, $f\in C([0\,;1])$, $\alpha\in\mathbb{R}_+^*$ et $\widetilde{u}\in\mathbb{R}$. Donner la discrétisation de ce problème par :

1. différences finies ;

2. volumes finis.

Dans les deux cas, écrire le schéma sous la forme d'un système linéaire de N équations à N inconnues, en explicitant matrice et second membre (N est le nombre de noeuds internes en différences finies, de mailles en volumes finis).

Exercice 17 — Différences finies et volumes finis pour les conditions de Fourier. On cherche à calculer une approximation de la fonction $u : [0\,;1] \to \mathbb{R}$ vérifiant

$$\begin{cases} -u'' = f \quad \text{dans }]0\,;1[& \text{(1.106a)} \\ u(0) - u'(0) = c & \text{(1.106b)} \\ u(1) + u'(1) = d & \text{(1.106c)} \end{cases}$$

où f est une fonction continue de $[0\,;1]$ dans $\mathbb{R}$. On va étudier pour cela un schéma de différences finies (partie A) et de volumes finis (partie B). Les deux parties sont indépendantes.

Partie A – Approximation par différences finies

On considère le schéma de différences finies à pas constant $h = 1/N$ pour le problème (1.106) :

$$\begin{cases} u_0 - \dfrac{u_1 - u_0}{h} = c & \text{(1.107a)} \\ \dfrac{2u_i - u_{i-1} - u_{i+1}}{h^2} = f(x_i) \qquad i = 1, \dots, N, & \text{(1.107b)} \\ u_{N+1} + \dfrac{u_{N+1} - u_N}{h} = d & \text{(1.107c)} \end{cases}$$

1. En respectant l'ordre des équations, écrire le schéma de différences finies (1.107) sous la forme $AU = b$ où A est une matrice et U et b des vecteurs qu'on explicitera.

2. Soit $U = (u_0, u_1, \dots, u_N, u_{N+1}) \in \mathbb{R}^{N+2}$; on pose $V = (v_0, v_1, \dots, v_N, v_{N+1})$, avec $v_0 = u_0/h$, $v_{N+1} = u_{N+1}/h$, et $v_i = u_i$ pour $i = 1, \dots, N$. Montrer que $AU \cdot V > 0$ si $U \neq 0$. En déduire qu'il existe une unique solution au système linéaire (1.107).

3. Peut on mettre le système sous une forme équivalente $\widetilde{A}U = \widetilde{b}$ où $\widetilde{A}$ est une matrice symétrique définie positive ?

4. Soit A la matrice définie à la question 1. Montrer que si $V = (v_0, v_1, \dots, v_N, v_{N+1}) \in \mathbb{R}^{N+2}$ est tel que $AV \geqslant 0$ (composante par composante, c'est-à-dire $(AV)_i \geqslant 0$ pour tout $i = 0, \dots, N+1$), alors $v_i \geqslant 0$ pour $i = 0, \dots, N+1$.

5. On définit l'erreur de consistance en $x_i = ih$ par $R_i = (A\overline{U} - b)_i$ (où A est la matrice définie par la question 1), avec $\overline{U} = (u_0, \dots, u_{N+1})$ et $\overline{u}_i = u(ih)$, $i = 0, \dots, N+1$ où u est la solution exacte de (1.106). Donner l'ordre de l'erreur de consistance du schéma en $x_i = ih$ pour $i = 0, \dots, N+1$ (en supposant u suffisamment régulière).

6. On définit la fonction φ de $[0\,;1]$ dans $\mathbb{R}$ par $\varphi(x) = -x^2 + 4$, et $\phi \in \mathbb{R}^{N+2}$ par $\phi_i = \varphi(x_i)$ pour $i = 0, \dots, N+1$.

(a) Montrer que $A\phi = S$ avec $S = (s_0, s_1, \dots, s_N, s_{N+1})$ et $s_0 = 4 + h$, $s_i = 2$ pour $i = 1, \dots, N$ et $s_{N+1} = 1 + h$.

(b) Soit U la solution du système $AU = b$ et soit

$$\mu = \max\Big(\max_{i=1,\dots,N}\Big(\frac{|b_i|}{2}\Big), \frac{|b_0|}{4}, |b_{N+1}|\Big)$$

Montrer que $A(U + \mu\phi) \geqslant 0$ et que $A(U - \mu\phi) \leqslant 0$. En déduire que $\|U\|_\infty \leqslant 4\mu$ et que $\|A^{-1}\|_\infty \leqslant 4$. On rappelle que pour une matrice carrée M à n lignes et n colonnes :

$$\|M\|_\infty = \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\|Mx\|_\infty}{\|x\|_\infty}$$

7. En déduire que le schéma est convergent et donner son ordre de convergence.

Partie B – Approximation par volumes finis

On se donne maintenant un ensemble de N mailles $(K_i)_{i=1,\dots,N}$, telles que $K_i =]x_{i-1/2}\,;x_{i+1/2}[$ avec $x_{1/2} = 0 < x_{3/2} < x_{i-1/2} < x_{i+1/2} < \ldots < x_{N+1/2} = 1$ et on note $h_i = x_{i+1/2} - x_{i-1/2}$. On se donne également N points $(x_i)_{i=1,\dots,N}$ situés dans les mailles K_i. On a donc :

$$0 = x_{1/2} < x_1 < x_{3/2} < \ldots < x_{i-1/2} < x_i < x_{i+1/2} < \ldots < x_{N+1/2} = 1$$

On notera $x_0 = 0$ et $x_{N+1} = 1$, $h_{i+1/2} = x_{i+1} - x_i$ pour $i = 0,\dots,N$ et $h = \max_{i=1,\dots,N} h_i$ [figure 1.2].

1. Écrire un schéma de volumes finis associé à ce maillage pour le problème (1.106).

2. Écrire le schéma sous la forme $AU = b$ où A est une matrice U et b des vecteurs qu'on explicitera.

3. Montrer qu'il existe une unique solution au système linéaire $AU = b$ défini à la question précédente.

4. Soit A la matrice définie à la question 2. Montrer que si $V = (v_0, v_1, \dots, v_N, v_{N+1}) \in \mathbb{R}^{N+2}$ est tel que $AV \geqslant 0$, alors $v_i \geqslant 0$ pour $i = 0,\dots,N+1$.

5. Donner l'ordre de consistance sur les flux.

6. Donner une estimation d'erreur pour ce schéma.

Exercice 18 — Convergence de la norme H^1 discrète. Montrer que si $u_{\mathcal{T}} :]0\,;1[\longrightarrow \mathbb{R}$ est définie par $u_{\mathcal{T}}(x) = u_i$, $\forall x \in K_i$ où $(u_i)_{i=1,\dots,N}$ solution de (1.42)-(1.44), alors $|u_{\mathcal{T}}|_{1,\mathcal{T}}$ converge dans $L^2(]0\,;1[)$ lorsque h tend vers 0, vers $\|Du\|_{L^2(]0\,;1[)}$, où u est la solution du problème (1.1)-(1.2).

corrigé p.54

Exercice 19 — Problème elliptique 1d, discrétisation par volumes finis. Soient a, $b \geqslant 0$, c, $d \in \mathbb{R}$ et $f \in C([0\,;1],\mathbb{R})$; on cherche à approcher la solution u du problème suivant :

$$\begin{cases} -u''(x) + au'(x) + b(u(x) - f(x)) = 0 \quad x \in [0\,;1] & (1.108)\\ u(0) = c & (1.109)\\ u(1) = d & (1.110) \end{cases}$$

On suppose que (1.108)-(1.109)-(1.110) admet une solution unique $u \in \mathcal{C}^2([0\,;1],\mathbb{R})$. Soient $N \in \mathbb{N}^\star$ et $h_1,\dots,h_N > 0$ tels que $\sum_{i=1}^N h_i = 1$. On pose $x_{1/2} = 0$ de sorte que $x_{N+1/2} = 1$ et :

$$\begin{cases} x_{i+1/2} = x_{i-1/2} + h_i & i = 1,\dots,N\\ h_{i+1/2} = \dfrac{h_{i+1} + h_i}{2} & i = 1,\dots,N-1\\ f_i = \dfrac{1}{h_i}\displaystyle\int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,\mathrm{d}x & i = 1,\dots,N \end{cases}$$

Pour approcher la solution u de (1.108)-(1.109)-(1.110), on propose le schéma numérique suivant :

$$F_{i+1/2} - F_{i-1/2} + bh_i u_i = bh_i f_i \quad i \in \{1,\dots,N\}, \tag{1.111}$$

avec $(F_{i+1/2})_{i\in\{0,\dots,N\}}$ donné par les expressions suivantes :

$$\begin{cases} F_{i+1/2} = -\dfrac{u_{i+1} - u_i}{h_{i+1/2}} + au_i \qquad i = 1,\dots,N-1 & (1.112)\\ F_{1/2} = -\dfrac{u_1 - c}{h_1/2} + ac \qquad F_{N+1/2} = -\dfrac{d - u_N}{h_N/2} + au_N & (1.113) \end{cases}$$

En tenant compte des expressions (1.112) et (1.113), le schéma numérique (1.111) donne donc un système de N équations à N inconnues $u_1,\dots,u_N$:

1. Expliquer comment, à partir de (1.108), (1.109) et (1.110), on obtient ce schéma numérique.

2. Existence de la solution approchée.

(a) On suppose ici que $c = d = 0$ et $f_i = 0$ pour tout $i \in \{1,\dots,N\}$. Montrer qu'il existe un unique vecteur $U = (u_1,\dots,u_N) \in \mathbb{R}^N$ solution de (1.111). Ce vecteur est obtenu en prenant $u_i = 0$, pour tout $i \in \{1,\dots,N\}$. (On rappelle que dans (1.111) les termes $F_{i+1/2}$ et $F_{i-1/2}$ sont donnés par (1.112) et (1.113))

(b) On revient maintenant au cas général (c'est-à-dire $c, d \in \mathbb{R}$ et $f \in C([0\,;1], \mathbb{R})$. Montrer qu'il existe un unique vecteur $U = (u_1, \ldots, u_N) \in \mathbb{R}^N$ solution de (1.111). (On rappelle, encore une fois, que dans (1.111) les termes $F_{i+1/2}$ et $F_{i-1/2}$ sont donnés par (1.112) et (1.113))

Soient $\alpha, \beta > 0$. On suppose, dans tout la suite de l'exercice, qu'il existe $h > 0$ tel que $\alpha h \leqslant h_i \leqslant \beta h$, pour tout $i \in \{1, \ldots, N\}$. On note $\overline{u}_i = \dfrac{1}{h_i} \int_{x_{i-1/2}}^{x_{i+1/2}} u(x)\,\mathrm{d}x$, pour $i = 1, \ldots, N$. On rappelle que u est la solution exacte de (1.108)-(1.109)-(1.110).

3. Non consistance du schéma au sens des différences finies

(a) Montrer que le système peut se mettre sous la forme $AU = B$, où B est définie par

$$\begin{cases} B_1 = bf_1 + \dfrac{2c}{h_1^2} + \dfrac{ac}{h_1} & (1.114) \\ B_i = bf_i \qquad i = 2, \ldots, N-1 & (1.115) \\ B_N = bf_n + \dfrac{2d}{h_N^2} & (1.116) \end{cases}$$

(b) On pose $\bar{R} = A\bar{U} - B$ avec $\bar{U} = (\bar{u}_1, \ldots, \bar{u}_N)$. Vérifier que pour tout $i \in \{1, \ldots, N\}$, $\bar{R}_i$ peut se mettre sous la forme :

$$\bar{R}_i = \bar{R}_i^1 + \bar{R}_i^2$$

où $\sup_{i=1,\ldots,N} \mid \bar{R}_i^1 \mid \leqslant C_1$ et $\sup_{i=1,\ldots,N} \mid \bar{R}_i^2 \mid \leqslant C_2 h$.

(c) On se restreint dans cette question au cas où $a = 0$, $b > 0$, $f = 0$, $c = 1$, $d = e^{\sqrt{b}}$, $N = 2q$, $h_i = h$ si i est pair et $h_i = h/2$ si i est impair avec $h = 2/3N$. Montrer que $\|\bar{R}\|_\infty$ ne tend pas vers 0 avec h.

4. Consistance des flux. En choisissant convenablement $(\bar{F}_{i+1/2})_{i\in\{0,\ldots,N\}}$, montrer que :

$$\bar{F}_{i+1/2} - \bar{F}_{i-1/2} + bh_i\overline{u}_i = bh_i f_i,\ i \in \{1, \ldots, N\}, \tag{1.117}$$

et que $(\bar{F}_{i+1/2})_{i\in\{0,\ldots,N\}}$ vérifie les égalités suivantes :

$$\begin{cases} \bar{F}_{i+1/2} = -\dfrac{\overline{u}_{i+1} - \overline{u}_i}{h_{i+1/2}} + a\overline{u}_i + R_{i+1/2} \qquad i \in \{1, \ldots, N-1\}, & (1.118) \\ \bar{F}_{1/2} = -\dfrac{\overline{u}_1 - c}{h_1/2} + ac + R_{1/2} \qquad \bar{F}_{N+1/2} = -\dfrac{d - \overline{u}_N}{h_N/2} + au_N + R_{N+1/2} & (1.119) \end{cases}$$

avec :

$$|R_{i+1/2}| \leqslant C_1 h \quad i \in \{0, \ldots, N\}, \tag{1.120}$$

où $C_1 \in \mathbb{R}$ et C_1 ne dépend que de α, β et u.

5. Estimation d'erreur. On pose $e_i = \overline{u}_i - u_i$, pour $i \in \{1, \ldots, N\}$ et $E = (e_1, \ldots, e_N)$.

(a) Montrer que E est solution du système (de N équations) suivant :

$$G_{i+1/2} - G_{i-1/2} + bh_i e_i = 0,\ i \in \{1, \ldots, N\}, \tag{1.121}$$

avec $(G_{i+1/2})_{i\in\{0,\ldots,N\}}$ donné par les expressions suivantes :

$$\begin{cases} G_{i+1/2} = -\dfrac{e_{i+1} - e_i}{h_{i+1/2}} + ae_i + R_{i+1/2} \quad i \in \{1, \ldots, N-1\} & (1.122) \\ G_{1/2} = -\dfrac{e_1}{h_1/2} + R_{1/2} \qquad G_{N+1/2} = -\dfrac{-e_N}{h_N/2} + ae_N + R_{N+1/2} & (1.123) \end{cases}$$

(b) En multipliant (1.121) par e_i et en sommant sur $i = 1, \ldots, N$, montrer qu'il existe $C_2 \in \mathbb{R}$, ne dépendant que de α, β et u tel que :

$$\sum_{i=0}^{N} (e_{i+1} - e_i)^2 \leqslant C_2 h^3 \tag{1.124}$$

avec $e_0 = e_{N+1} = 0$.

(c) Montrer qu'il existe $C_3 \in \mathbb{R}$, ne dépendant que de α, β, et u tel que :

$$|e_i| \leqslant C_3 h, \text{ pour tout } i \in \{1, \dots, N\}. \tag{1.125}$$

6. (Principe du maximum.) On suppose, dans cette question, que $f(x) \leqslant d \leqslant c$, pour tout $x \in [0\,;1]$. Montrer que $u_i \leqslant c$, pour tout $i \in \{1, \dots, N\}$. (On peut aussi montrer que $u(x) \leqslant c$, pour tout $x \in [0\,;1]$.)

7. On remplace, dans cette question, (1.112) et (1.113) par :

$$\begin{cases} F_{i+1/2} = -\dfrac{u_{i+1} - u_i}{h_{i+1/2}} + a u_{i+1} \qquad i \in \{1, \dots, N-1\} & (1.126) \\ F_{1/2} = -\dfrac{u_1 - c}{h_1/2} + a u_1 \qquad F_{N+1/2} = -\dfrac{d - u_N}{h_N/2} + a d & (1.127) \end{cases}$$

Analyser brièvement le nouveau schéma obtenu (existence de la solution approchée, consistance des flux, estimation d'erreur, principe du maximum).

Exercice 20 — Un problème de rayonnement. On cherche à résoudre un modèle de diffusion thermique avec rayonnement volumique (dans un matériau comme le verre, par exemple) avec une discrétisation par différences finies et une méthode de monotonie.

Le modèle (simplifié) consiste à chercher la fonction u de $[0, 1]$ à valeurs dans $\mathbb{R}$ solution du problème suivant :

$$-\kappa u''(x) + \int_0^1 \frac{1}{\sqrt{|x-y|}}(u^4(x) - u^4(y))dy = 0 \text{ pour } x \in]0, 1[, \tag{1.128}$$

$$u(0) = c_1,\ u(1) = c_2. \tag{1.129}$$

Pour $n \geqslant 1$, on pose $h = 1/(n+1)$.

1. Montrer que la discrétisation de (1.128)-(1.129) par différences finies avec un pas uniforme $h = \frac{1}{n}$ consiste à chercher le vecteur u de $\mathbb{R}^n$ solution de

   $$Au + R(u) = b, \tag{1.130}$$

   où A est une matrice $n \times n$ et b un vecteur de $\mathbb{R}^n$ qu'on explicitera et où R est une application de $\mathbb{R}^n$ dans $\mathbb{R}^n$, dont les composantes, calculées par une méthode d'intégration numérique que l'on précisera, sont données par :

$$R_i(u) = \sum_{j \in \{1,\dots,n\},\, j \neq i} \frac{\sqrt{h}}{\sqrt{|i-j|}}(u_i^4 - u_j^4) + \frac{\sqrt{h}}{2\sqrt{i}}(u_i^4 - 1) + \frac{\sqrt{h}}{2\sqrt{n+1-i}}(u_i^4 - 16), \qquad i \in \{1, \dots, n\} \tag{1.131}$$

Pour trouver une solution de (1.130), on se donne $\beta \geqslant 0$ et on utilise la méthode itérative suivante :

$$\textbf{Initialisation } u^{(0)} \in \mathbb{R}^n,\ u_i^{(0)} = 1 \text{ pour tout } i \in \{1, \dots, n\}. \tag{1.132}$$

Itérations $Au^{(k+1)} + \beta u^{(k+1)} = -R(u^{(k)}) + \beta u^{(k)} + b$.

Dans toute la suite, on prendra $\kappa = 10$, $c_1 = 1$, $c_2 = 2$, et on note m l'élément de $\mathbb{R}^n$ dont toutes les composantes sont égales à 1 (c'est-à-dire $m = u^{(0)}$) et M l'élément de $\mathbb{R}^n$ dont toutes les composantes sont égales à 2.

2. (Propriété de A) Soient $\beta \geqslant 0$ et $u \in \mathbb{R}^n$. Montrer que $Au + \beta u \geqslant 0 \Rightarrow u \geqslant 0$.
3. (Propriété de R)
   (a) Soient $u \in \mathbb{R}^n$, $m \leqslant u \leqslant M$, et $i, j \in \{1, \dots, n\}$.
   Montrer que $\partial_j R_i(u) \leqslant 0$ pour $i \neq j$ et $\partial_i R_i(u) \leqslant 128$.
   (b) Soient $u, v \in \mathbb{R}^n$, $m \leqslant u \leqslant v \leqslant M$, et $\beta \geqslant 128$. Montrer que $\beta u - R(u) \leqslant \beta v - R(v)$.

4. (Sous et sur solutions) Montrer que $Am + R(m) \leqslant b$ et $AM + R(M) \geqslant b$.
5. On choisit $\beta \geqslant 128$. Montrer, par récurrence sur k, que pour tout $k \geqslant 0$,

   $$m \leqslant u^{(k)} \leqslant u^{(k+1)} \leqslant M.$$

   En déduire qu'il existe $u \in \mathbb{R}^n$ tel que $u^{(k)} \to u$ quand $k \to +\infty$ et que $Au + R(u) = b$, $m \leqslant u \leqslant M$.
6. On initialise maintenant la méthode (1.132) par $u_i^{(0)} = 2$ au lieu de $u_i^0 = 1$ (pour tout $i = 1, \dots, n$). La méthode converge-t'elle ?
7. On discrétise maintenant le problème (1.128)-(1.129) par la méthode des volumes finis, avec un pas uniforme $h = \frac{1}{n}$.
   (a) Montrer que le problème discrétisé consiste encore à chercher un vecteur u de $\mathbb{R}^n$ solution d'un système non linéaire de la forme (1.130), et donner la matrice $\widetilde{A}$ et les vecteurs $\widetilde{b}$ et $\widetilde{R}(u)$ correspondants.
   (b) Construire un algorithme de la forme (1.132) adapté à cette nouvelle discrétisation et qui converge.

Exercice 21 — Discrétisation 2D par différences finies. Écrire le système linéaire obtenu lorsqu'on discrétise le problème

$$\begin{cases} -\Delta u = f & \text{dans } \Omega =]0\,;1[\times]0\,;1[\\ u = 0 & \text{sur } \partial\Omega \end{cases}$$

par différences finis avec un pas uniforme $h = 1/N$ dans les deux directions d'espace. Montrer l'existence et l'unicité de la solution du système linéaire obtenu.

corrigé p.56

Exercice 22 — Problème elliptique 2d et différences finies. Soit $\Omega =]0\,;1[^2 \subset \mathbb{R}^2$. On se propose d'étudier deux schémas numériques pour le problème suivant :

$$\begin{cases} -\Delta u(x,y) + k\dfrac{\partial u}{\partial x}(x,y) = f(x,y) & (x,y) \in \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases} \tag{1.133}$$

où $k > 0$ est un réel donné et $f \in C(\bar{\Omega})$ est donnée. On note u la solution exacte de (1.133) et on suppose que $u \in \mathcal{C}^4(\bar{\Omega})$.

1. Principe du maximum — Montrer que pour tout $\varphi \in \mathcal{C}^1(\bar{\Omega})$ tel que $\varphi = 0$ sur $\partial\Omega$, on a :

$$\int_\Omega \nabla u(x) \cdot \nabla\varphi(x)\,\mathrm{d}x + \int_\Omega k\frac{\partial u}{\partial x}(x)\varphi(x) \qquad\qquad dx = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x$$

En déduire que si $f \leqslant 0$ sur $\bar{\Omega}$, on a alors $u \leqslant 0$ sur $\bar{\Omega}$. Soit $N \in \mathbb{N}$; on pose $h = 1/(N+1)$ et $u_{i,j}$ est la valeur approchée recherchée de $u(ih, jh)$, $(i,j) \in \{0, \dots, N+1\}^2$. On pose $f_{i,j} = f(ih, jh)$, pour tout $(i,j) \in \{1, \dots, N\}^2$. On s'intéresse à deux schémas de la forme :

$$\begin{cases} a_0 u_{i,j} - a_1 u_{i-1,j} - a_2 u_{i+1,j} - a_3 u_{i,j-1} - a_4 u_{i,j+1} = f_{i,j} & \forall (i,j) \in \{1, \dots, N\}^2 \\ u_{i,j} = 0 & (i,j) \in \gamma \end{cases} \tag{1.134}$$

où a_0, a_1, a_2, a_3 et a_4 sont données (ce sont des fonctions données de h) et $\gamma = \{(i,j)\ ;\ (ih, jh) \in \partial\Omega\}$ (γ dépend aussi de h). Le premier schéma, centré pour la discrétisation de la dérivée $\partial_x u$, noté schéma [I], correspond au choix suivant des a_i :

$$a_0 = \frac{4}{h^2} \qquad a_1 = \frac{1}{h^2} + \frac{k}{2h} \qquad a_2 = \frac{1}{h^2} + \frac{k}{2h} \qquad a_3 = a_4 = \frac{1}{h^2}.$$

Le deuxième schéma, décentré pour la discrétisation de la dérivée $\partial_x u$, noté schéma [II], correspond au choix suivant des a_i :

$$a_0 = \frac{4}{h^2} + \frac{k}{h} \qquad a_1 = \frac{1}{h^2} + \frac{k}{h} \qquad a_2 = a_3 = a_4 = \frac{1}{h^2}$$

2. Consistance — Donner une majoration de l'erreur de consistance en fonction de k, h et des dérivées de u, pour les schémas [I] et [II]. Donner l'ordre des schémas [I] et [II].

3. Principe du maximum discret — Dans le cas du schéma [II] montrer que si $(w_{i,j})$ vérifie :

$$a_0 w_{i,j} - a_1 w_{i-1,j} - a_2 w_{i+1,j} - a_3 w_{i,j-1} - a_4 w_{i,j+1} \leqslant 0 \qquad \forall (i,j) \in \{1,\dots,N\}^2$$

on a alors

$$w_{i,j} \leqslant \max_{(n,m)\in\gamma} (w_{n,m}) \qquad \forall (i,j) \in \{1,\dots,N\}^2$$

Montrer que ceci est aussi vrai dans le cas du schéma [I] si h vérifie une condition à déterminer.

4. Stabilité — Montrer que le schéma [II] et le schéma [I] sous la condition trouvée en 3. sont stables (au sens $||U||_\infty \leqslant C||f||_\infty$, avec une constante C à déterminer explicitement, où $U = \{u_{i,j}\}_{(i,j)\in\{0,\dots,N+1\}^2}$ est solution de (1.134). [On pourra utiliser la fonction $\phi(x,y) = y^2/2$].
En déduire que dans le cas du schéma [II] et du schéma [I] sous la condition trouvée en 3. le problème (1.134) admet, pour tout f, une et une seule solution.

5. Convergence — Les schémas [I] et [II] sont-ils convergents ? (au sens $\max_{(i,j)\in\{0,\dots,N+1\}^2}(|u_{i,j} - u(ih,jh)|) \to 0$ quand $h \to 0$). Quel est l'ordre de convergence de chacun des schémas ?

6. Commentaires — Quels sont, à votre avis, les avantages respectifs des schémas [I] et [II] ?

corrigé p.59

Exercice 23 — Implantation de la méthode des volumes finis. On considère le problème de conduction du courant électrique

$$-\operatorname{div}(\mu_i \nabla\phi(x)) = 0 \qquad x \in \Omega_i \quad i = 1,2 \tag{1.135}$$

où ϕ représente le potentiel électrique, $j = -\mu\nabla\phi(x)$ est donc le courant électrique, $\mu_1 > 0$, $\mu_2 > 0$ sont les conductivités thermiques dans les domaines Ω_1 et avec Ω_2, avec $\Omega_1 =]0\,;1[\times]0\,;1[$ et $\Omega_2 =]0\,;1[\times]1\,;2[$. On appelle $\Gamma_1 =]0\,;1[\times \{0\}$, $\Gamma_2 = \{1\} \times]0\,;2[$, $\Gamma_3 =]0\,;1[\times \{2\}$ et $\Gamma_4 = \{0\} \times]0\,;2[$ les frontières extérieures de Ω et on note $I =]0\,;1[\times \{0\}$, l'interface entre Ω_1 et Ω_2 (voir Figure 1.5). Dans la suite, on notera μ la conductivité électrique sur Ω avec $\mu|_{\Omega_i} = \mu_i$, $i = 1,2$.

On suppose que les frontières Γ_2 et Γ_4 sont parfaitement isolées. Le potentiel électrique étant défini à une constante près, on impose que sa moyenne soit nulle sur le domaine, pour que le problème soit bien posé.

La conservation du courant électrique impose que

$$\int_{\Gamma_1} j \cdot \boldsymbol{n} + \int_{\Gamma_3} j \cdot \boldsymbol{n} = 0$$

où $\boldsymbol{n}$ désigne le vecteur unitaire normal à la frontière $\partial\Omega$ et extérieure à Ω.

Enfin, on suppose que l'interface I est le siège d'une réaction électrochimique qui induit un saut de potentiel. On a donc pour tout point de l'interface I :

$$\phi_2(x) - \phi_1(x) = \psi(x) \qquad \forall x \in I$$

où ϕ_i désigne la restriction de ϕ au sous domaine i. La fonction ϕ est donc discontinue sur l'interface I. Notons que, par contre, le courant électrique est conservé et on a donc

$$(-\mu\nabla\phi \cdot \boldsymbol{n})|_2(x) + (-\mu\nabla\phi \cdot \boldsymbol{n})|_1(x) = 0 \qquad \forall x \in I$$

1. Écrire le problème complet, avec conditions aux limites.

2. Discrétiser le problème par la méthode des volumes finis, avec un maillage rectangulaire uniforme, (considérer deux inconnues discrètes pour chaque arête de l'interface) et écrire le système linéaire obtenu sur les inconnues discrètes.

corrigé p.60

Exercice 24 — Élimination des inconnues d'arêtes. On se place ici dans le cadre des hypothèses et notations du paragraphe 1.4.2 :

1. Pour chaque arête interne $\sigma = K|L$, calculer la valeur u_σ en fonction de u_K et u_L et en déduire que les flux numériques $F_{K,\sigma}$ et $F_{L,\sigma}$ vérifient bien (1.73) ;

2. Pour chaque arête $\sigma \subset \Gamma_1 \cup \Gamma_3$, telle que $\sigma \in \mathcal{E}_K$, calculer u_σ en fonction de u_K et montrer que $F_{K,\sigma}$ vérifie bien (1.74) ;

3. Pour chaque arête $\sigma \in \mathcal{E}_I$, avec $\sigma = K|L$, $K \in \Omega_1$, calculer la valeur u_σ en fonction de u_K et u_L et en déduire que les flux numériques $F_{K,\sigma}$ et $F_{L,\sigma}$ vérifient bien (1.76) ;

4. Écrire le système linéaire que satisfont les inconnues $(u_K)_{K\in\mathcal{T}}$.

1.5.2 Corrigés

Exercice 1 — Différences et volumes finis avec conditions de Dirichlet non homogènes.

1. Le schéma différences finies pour l'équation (1.77) s'écrit :

$$\begin{cases} \dfrac{1}{h^2}\left(2u_i - u_{i-1} - u_{i+1}\right) + \sin u_i = f_i \quad i = 1, \dots, N \\ u_0 = a \\ u_{N+1} = b \end{cases}$$

ce qui s'écrit encore

$$\begin{cases} \dfrac{2u_i - u_{i+1} - u_{i-1}}{h^2} + \sin u_i = f_i \quad i = 2, \dots, N-1 \\ \dfrac{2u_1 - u_2 - a}{h^2} + \sin u_1 = f_1 \\ \dfrac{2u_N - u_{N-1} - b}{h^2} + \sin u_N = f_N \end{cases}$$

Le schéma volumes finis pour la même équation s'écrit :

$$F_{i+1/2} - F_{i-1/2} + h \sin u_i = h f_i \qquad i = 1, \dots, N \tag{1.136}$$

avec :

$$\begin{cases} F_{i+1/2} = -\dfrac{u_{i+1} - u_i}{h} \qquad i = 1, \dots, N-1 \\ F_{1/2} = -\dfrac{u_1 - a}{h/2} \qquad F_{N+1/2} = -\dfrac{b - u_N}{h/2} \end{cases}$$

En remplaçant les expressions des flux dans l'équation (1.136). On obtient :

$$\begin{cases} \dfrac{2u_i - u_{i+1} - u_{i-1}}{h^2} + \sin u_i = f_i \qquad i = 2, \dots, N-1 \\ \dfrac{3u_1 - u_2 - 2a}{h^2} + \sin u_1 = 2f_1 \\ \dfrac{3u_N - u_{N-1} - 2b}{h^2} + \sin u_N = 2f_N \end{cases}$$

2. La différence entre les deux schémas réside uniquement dans les conditions aux limites.

Exercice 3 — Différences finies et principe du maximum.

1. On se donne N points de discrétisation et on écrit l'équation(1.79a) en chaque point x_i, $i = 1, \dots, N$. Avec la discrétisation donnée dans le cours pour $u''(x)$, on obtient, en tenant compte des conditions limites (1.79b)-(1.79c)

$$\begin{aligned} &\frac{1}{h^2}(2u_i - u_{i+1} - u_{i-1}) + c(x_i)u_i = f(x_i), \qquad i = 1, \dots, N, \\ &u_0 = a \\ &u_{N+1} = b. \end{aligned}$$

ce qui donne le schéma suivant :

$$\frac{1}{h^2}(2u_i - u_{i+1} - u_{i-1}) + c(x_i)u_i = f(x_i), \qquad i = 2, \dots, N-1,$$
$$\frac{1}{h^2}(2u_1 - u_2) + c(h)u_1 = \frac{1}{h^2}a + f(h)$$
$$\frac{1}{h^2}(2u_N - u_{N-1}) + c(1-h)u_N = \frac{1}{h^2}b + f(1-h).$$

2. Supposons $a \leqslant b$ (l'autre cas se traite de manière similaire). On note ici encore $u_0 = a$ et $u_{N+1} = b$. Soit $i_0 = \min\{i \ ; \ 0 \leqslant i \leqslant N+1 \text{ et } u_i = \min_{j=1,N+1} u_j\}$. C'est donc le plus petit des indices i pour lesquels u_i est minimum. Si $i_0 > 0$, on a

$$\frac{1}{h^2}(u_{i_0} - u_{i_0+1}) + \frac{1}{h^2}(u_{i_0} - u_{i_0-1}) \geqslant 0,$$

ce qui entraine $u_{i_0} - u_{i_0+1} = 0$ et $u_{i_0} - u_{i_0-1} = 0$. Or la deuxième égalité est impossible par définition de i_0. On a donc $i_0 = 0$ et $u_i \geqslant a$ pour tout $i = 1, \dots, N$.

Exercice 5 — Équation de transport-diffusion sous forme non-conservative.

1. La matrice M et le second membre B sont donnés par :

$$\begin{cases} (MU)_i = \dfrac{2u_i - u_{i+1} - u_{i-1}}{h^2} + \dfrac{v_i(u_i - u_{i-1})}{h}, \text{ pour } i = 2, \dots, N \\ (MU)_1 = \dfrac{2u_1 - u_2}{h^2} + \dfrac{v_1 u_1}{h} \\ (MU)_N = \dfrac{2u_N - u_{N-1}}{h^2} + \dfrac{v_N(u_N - u_{N-1})}{h} \end{cases} \quad \text{et } B = \begin{pmatrix} \left(\frac{1}{h^2} + \frac{v_1}{h}\right)a \\ 0 \\ \vdots \\ 0 \\ \frac{1}{h^2}b \end{pmatrix}$$

(a) Supposons $MU \geqslant 0$. Soit $i_0 = \min\{i; u_i = \min_j u_j\}$.

i. Si $i_0 = 1$, comme $\frac{2u_1 - u_2}{h^2} + \frac{v_1 u_1}{h} \geqslant 0$, on a

$$\left(\frac{1}{h^2} + \frac{v_1}{h}\right)u_1 + \frac{u_1 - u_2}{h^2} \geqslant 0$$

et comme $u_1 - u_2 \leqslant 0$, ceci entraine $u_1 \geqslant 0$.

ii. Si $2 \leqslant i_0 \leqslant N-1$, on a :

$$\frac{u_{i_0} - u_{i_0+1}}{h^2} + \frac{u_{i_0} - u_{i_0-1}}{h^2} + \frac{v_i(u_{i_0} - u_{i_0-1})}{h} \geqslant 0$$

Mais par définition de i_0, on a $u_{i_0} - u_{i_0+1} \leqslant 0$ et , et $u_{i_0} - u_{i_0-1} < 0$ donc ce cas est impossible.

iii. Si $i_0 = N$, on a :

$$\frac{1}{h^2}u_N + \frac{1}{h^2}(u_N - u_{N-1}) + \frac{1}{h}v_i(u_N - u_{N-1}) \geqslant 0.$$

Or par définition de $i_0 = N$, on a $u_N < u_{N-1}$, et donc $u_N > 0$.

On a donc montré que si $MU \geqslant 0$ alors $U \geqslant 0$.

(b) Comme $MU \geqslant 0 \Rightarrow U \geqslant 0$, on a donc (en prenant U puis $-U$) $MU = 0 \Rightarrow U = 0$, ce qui prouve que M est inversible,

(c) Soit U solution de $MU = b$. Posons $\widetilde{U} = (u_0, u_1, \dots, u_N, u_{N+1})$, avec $u_0 = a_0$ et $u_{N+1} = a_1$. Remarquons d'abord que le minimum et le maximum des composantes u_i de $\widetilde{U}$ ne peuvent être atteints pour $i = 1, \dots, N$ que si les u_i sont tous égaux (auquel cas $u_i = a_0 = a_1$ pour tout $i = 0, \dots, N+1$). En effet, pour $i = 1, \dots, N$, on a :

$$\frac{1}{h^2}(u_i - u_{i+1}) + \frac{1}{h^2}(u_i - u_{i-1}) + \frac{1}{h}v_i(u_i - u_{i-1}) = 0 \tag{1.137}$$

Soit $i_0 = \min\{i; u_i = \min_j u_j\}$ et $i_1 = \min\{i; u_i = \max_j u_j\}$. Par définition de i_0, on a $u_{i_0} - u_{i_0+1} \leqslant 0$ et $u_{i_0} - u_{i_0-1} < 0$ donc (1.137) est impossible si $1 < i_0 < N$. Donc $i_0 = 1$

ou N. Soit $i_1 = \min\{i; u_i = \max_j u_j\}$, on a $u_{i_1} - u_{i_1+1} \geqslant 0$ et $u_{i_1} - u_{i_1-1} > 0$ et (1.137) est encore impossible si $1 < i_1 < N$.

On en déduit que $i_0 = 0$ ou $N+1$ et $i_1 = 0$ ou $N+1$, ce qui prouve que $\min(a,b) \leqslant u_i \leqslant \max(a,b)$.

2. (a) Par définition, $m_{i,i} = \frac{1}{h^2} + \frac{1}{h}v_i$ avec $v_i \geqslant 0$, ce qui prouve le résultat.

(b) Par définition, $m_{i,j}$ est soit nul, soit égal à $-\frac{1}{h^2} + \frac{1}{h}v_i$ si $j = i-1$, soit égal à $-\frac{1}{h^2}$ si $j = i+1$, ce qui prouve le résultat.

(c) On a montré que M est inversible à la question précédente.

(d) D'après la question 1.2, on sait que si $MU \geqslant 0$, alors $U \geqslant 0$. Soit e_i le i–ème vecteur de la base canonique. On a $e_i = M(M^{-1})e_i \geqslant 0$, et donc $M^{-1}e_i \geqslant 0$, ce qui montre que tous les coefficients de M^{-1} doivent être positifs.

Exercice 7 — Conditionnement efficace.
Partie I

1. Soit $u = (u_1, \ldots, u_N)$. On a

$$Au = b \Leftrightarrow \begin{cases} \dfrac{1}{h^2}(u_i - u_{i-1}) + \dfrac{1}{h^2}(u_i - u_{i+1}) = b_i \quad \forall i = 1, \ldots N \\ u_0 = u_{N+1} = 0 \end{cases}$$

Supposons $b_i \geqslant 0, \forall i = 1, \ldots, N$, et soit $p \in \{0, \ldots, N+1\}$ tel que $u_p = \min(u_i,\ i = 0, \ldots, N+1)$.

— Si $p = 0$ ou $N+1$, alors $u_i \geqslant 0\ \forall i = 0, N+1$ et donc $u \geqslant 0$.

— Si $p \in \{1, \ldots, N\}$, alors

$$\frac{1}{h^2}(u_p - u_{p-1}) + \frac{1}{h^2}(u_p - u_{p+1}) \geqslant 0$$

et comme $u_p - u_{p-1} < 0$ et $u_p - u_{p+1} \leqslant 0$, on aboutit à une contradiction.

Montrons maintenant que A est inversible. On vient de montrer que si $Au \geqslant 0$ alors $u \geqslant 0$. On en déduit par linéarité que si $Au \leqslant 0$ alors $u \leqslant 0$, et donc que si $Au = 0$ alors $u = 0$. Ceci démontre que l'application linéaire représentée par la matrice A est injective donc bijective (car on est en dimension finie).

2. Soit $\varphi \in C([0\,;1], \mathbb{R})$ tel que $\varphi(x) = 1/2x(1-x)$ et $\phi_i = \varphi(x_i)$, $i = 1, N$, où $x_i = ih$. $(A\phi)_i$ est le développement de Taylor à l'ordre 2 de $\varphi''(x_i)$, et comme φ est un polynôme de degré 2, ce développement est exact. Donc $(A\phi)_i = \varphi''(x_i) = 1$.

3. Soient $b \in \mathbb{R}^N$ et $u \in \mathbb{R}^N$ tels que $Au = b$. On a :

$$(A(u \pm \|b\|\varphi))_i = (Au)_i \pm \|b\|(A\phi)_i = b_i \pm \|b\|.$$

Prenons d'abord $\widetilde{b}_i = b_i + \|b\| \geqslant 0$, alors par la question (1),

$$u_i + \|b\|\phi_i \geqslant 0 \qquad \forall i = 1, \ldots, N.$$

Si maintenant on prend $\bar{b}_i = b_i - \|b\| \leqslant 0$, alors

$$u_i - \|b\|\phi_i \leqslant 0 \qquad \forall i = 1, \ldots, N.$$

On a donc $-\|b\|\phi_i \leqslant \|b\|\phi_i$.

On en déduit que $\|u\|_\infty \leqslant \|b\|\ \|\phi\|_\infty$; or $\|\phi\|_\infty = 1/8$, d'où $\|u\|_\infty \leqslant \|b\|/8$. On peut alors écrire que pour tout $b \in \mathbb{R}^N$,

$$\|A^{-1}b\|_\infty \leqslant \frac{1}{8}\|b\|, \text{ donc } \frac{\|A^{-1}b\|_\infty}{\|b\|_\infty} \leqslant \frac{1}{8}, \text{ d'où } \|A^{-1}\| \leqslant \frac{1}{8}$$

On montre que $\|A^{-1}\| = 1/8$ en prenant le vecteur b défini par $b(x_i) = 1$, $\forall i = 1, \ldots, N$. On a en effet $A^{-1}b = \phi$, et comme N est impair, $\exists i \in \{1, \ldots, N\}$ tel que $x_i = 1/2$; or $\|\varphi\|_\infty = \varphi(1/2) = 1/8$.

4. Par définition, on a $\|A\| = \sup_{\|x\|_\infty = 1} \|Ax\|$, et donc $\|A\| = \max_{i=1,N} \sum_{j=1,N} |a_{i,j}|$, d'où le résultat.

5. Grâce aux questions 3 et 4, on a, par définition du conditionnement pour la norme $\|\cdot\|$, $\mathrm{cond}(A) = \|A\|\|A^{-1}\| = \frac{1}{2h^2}$. Comme $A\delta_u = \delta_b$, on a :

$$\|\delta_u\| \leqslant \|A^{-1}\|\|\delta_b\|\frac{\|b\|}{\|b\|} \leqslant \|A^{-1}\|\|\delta_b\|\frac{\|A\|\|u\|}{\|b\|},$$

d'où le résultat.

Pour obtenir l'égalité, il suffit de prendre $b = Au$ où u est tel que $\|u\| = 1$ et $\|Au\| = \|A\|$, et δ_b tel que $\|\delta_b\| = 1$ et $\|A^{-1}\delta_b\| = \|A^{-1}\|$. On obtient alors

$$\frac{\|\delta_b\|}{\|b\|} = \frac{1}{\|A\|} \text{ et } \frac{\|\delta u\|}{\|u\|} = \|A^{-1}\|.$$

D'où l'égalité.

Partie 2

1. Soient $\varphi^{(h)}$ et $f^{(h)}$ les fonctions constantes par morceaux définies par

$$\varphi^h(x) = \begin{cases} \varphi(ih) = \phi_i & \text{si } x \in]x_i - h/2\,;x_i + h/2[\qquad i = 1,\dots,N \\ 0 & \text{si } x \in [0\,;h/2] \quad \text{ou} \quad x \in]1-h/2\,;1] \end{cases}$$

$$f^{(h)}(x) = \begin{cases} f(ih) = b_i & \text{si } x \in]x_i - h/2\,;x_i + h/2[\qquad i = 1,\dots,N \\ f(ih) = 0 & \text{si } x \in [0\,;h/2] \quad \text{ou} \quad x \in]1-h/2\,;1] \end{cases}$$

Comme $f \in C([0\,;1],\mathbb{R})$ et $\varphi \in C^2([0\,;1],\mathbb{R})$, la fonction f_h (resp. φ_h) converge uniformément vers f (resp. φ) lorsque $h \to 0$. On a donc

$$h\sum_{i=1}^{N} b_i\varphi_i = \int_0^1 f^{(h)}(x)\varphi^{(h)}(x)\,\mathrm{d}x \to \int_0^1 f(x)\varphi(x)\,\mathrm{d}x \text{ lorsque } h \to 0.$$

Comme $b_i > 0$ et $f_i > 0$, $\forall i = 1,\dots,N$, on a évidemment :

$$S_N = \sum_{i=1}^{N} b_i\varphi_i > 0 \text{ et } S_N \to \int_0^1 f(x)\varphi(x)\,\mathrm{d}x = \beta > 0 \text{ lorsque } h \to 0$$

Donc il existe $N_0 \in \mathbb{N}$ tel que si $N \geqslant N_0$, $S_N \geqslant \beta/2$, et donc $S_N \geqslant \alpha = \min(S_0, S_1 \dots S_{N_0}, \beta/2) > 0$.

2. On a $N\|u\| = N\sup_{i=1,N}|u_i| \geqslant \sum_{i=1}^N u_i$. D'autre part, $A\varphi = \begin{pmatrix}1 & \dots & 1\end{pmatrix}^t$ donc $u \cdot A\varphi = \sum_{i=1}^N u_i$; or $u \cdot A\varphi = A^t u \cdot \varphi = Au \cdot \varphi$ car A est symétrique donc :

$$u \cdot A\varphi = \sum_{i=1}^{N} b_i\varphi_i \geqslant \frac{\alpha}{h}$$

d'après la question 1. Comme $\delta_u = A^{-1}\delta_b$, on a donc $\|\delta_u\| \leqslant \|A^{-1}\|\ \|\delta_b\|$ et comme $N\|u\| \geqslant \frac{\alpha}{h}$, on obtient :

$$\frac{\|\delta_u\|}{\|u\|} \leqslant \frac{1}{8}\frac{hN}{\alpha}\|\delta_b\|\frac{\|f\|_\infty}{\|b\|}$$

Or $hN = 1$ et on a donc bien :

$$\frac{\|\delta_u\|}{\|u\|} \leqslant \frac{\|f\|_\infty}{8\alpha}\frac{\|\delta_b\|}{\|b\|}$$

3. Le conditionnement $\mathrm{cond}(A)$ calculé dans la partie 1 est d'ordre $1/h^2$, et donc tend vers l'infini lorsque le pas du maillage tend vers 0, alors qu'on vient de montrer dans la partie 2 que la variation relative $\|\delta_u\|/\|u\|$ est inférieure à une constante multipliée par la variation relative de $\|\delta_b\|/\|b\|$. Cette dernière information est nettement plus utile et réjouissante pour la résolution effective du système linéaire.

Exercice 9 — Erreur de consistance.

1. Si f est constante, alors $-u''$ est constante, et donc les dérivées d'ordre supérieur de u sont nulles. Donc par l'estimation (1.31) sur l'erreur de consistance, on a $R_i = 0$ pour tout $i = 1, \ldots, N$. Si on appelle U le vecteur de composantes u_i et $\bar{U}$ le vecteur de $\mathbb{R}^N$ de composantes $u(x_i)$, on peut remarquer facilement que $U - \bar{U} = A^{-1}R$, où R est le vecteur de composantes R_i. On a donc $U - \bar{U} = 0$, c.q.f.d.

2. Il est facile de voir que f n'est pas forcément constante, en prenant $f(x) = \sin 2\pi x$, et $h = 1/2$, on n'a donc qu'une seule inconnue u_1 qui vérifie $u_1 = 0$, et on a également $u(1/2) = \sin \pi = 0$.

Exercice 11 — Non consistance des volumes finis. 1. Par développement de Taylor, pour $i = 1, \ldots, N$, il existe $\xi_i \in [x_i, x_{i+1}]$ tel que :

$$u(x_{i+1}) = u(x_i) + h_{i+1/2}u'(x_i) + 1/2h_{i+1/2}^2u''(x_i) + \frac{h_{i+1/2}^3}{6}u'''(\xi_i),$$

et donc

$$R_i = -\frac{1}{h_i}\frac{h_{i+1/2} + h_{i-1/2}}{2}u''(x_i) + u''(x_i) + \rho_i \qquad i = 1, \ldots, N,$$

où $|\rho_i| \leqslant Ch$, C ne dépendant que de la dérivée troisième de u. Il est facile de voir que, en général, R_i, ne tend pas vers 0 lorsque h tend vers 0 (sauf dans des cas particuliers). En effet, prenons par exemple $f \equiv 1$, $h_i = h$ pour i pair, $h_i = h/2$ pour i impair et $x_i = (x_{i+1/2} + x_{i-1/2})/2$ pour $i = 1, \ldots, N$. On a dans ce cas $u'' \equiv -1$, $u''' \equiv 0$ et donc :

$$R_i = \begin{cases} -1/4 & \text{si } i \text{ est pair} \\ +1/2 & \text{si } i \text{ est impair.} \end{cases}$$

On en conclut que $\sup\{|R_i|, i = 1, \ldots, N\} \not\to 0$ lorsque $h \to 0$.
Notons cependant que si on définit la fonction R constante par morceaux et égale à R_i sur chaque maille tend vers 0 faiblement dans L^1 lorsque $h \to 0$.

2. Par développement de Taylor,

$$u(x_{i+1}) = u(x_i) + h_{i+\frac{1}{2}}u'(x_i) + \frac{1}{2}h_{i+\frac{1}{2}}^2u''(x_i) + h^2\varepsilon_i(h),$$
$$u(x_{i-1}) = u(x_i) - h_{i-\frac{1}{2}}u'(x_i) + \frac{1}{2}h_{i-\frac{1}{2}}^2u''(x_i) + h^2\varepsilon_{i-1}(h).$$

avec, pour $k = i-1, i$, $\varepsilon_k(h) \to 0$ lorsque $h \to 0$. En multipliant la première équation par $h_{i-\frac{1}{2}}$ et la deuxième par $h_{i+\frac{1}{2}}$ et en additionnant, il vient

$$h_{i-\frac{1}{2}}(u(x_{i+1}) - u(x_i)) + h_{i+\frac{1}{2}}(u(x_{i-1}) - u(x_i)) = \frac{1}{2}h_{i+\frac{1}{2}}h_{i-\frac{1}{2}}(h_{i+\frac{1}{2}} + h_{i-\frac{1}{2}}))u''(x_i) + h^3\varepsilon(h),$$

avec $\varepsilon(h) = \varepsilon_{i-1}(h) + \varepsilon_{i-1}(h) \to 0$ lorsque $h \to 0$. On a donc

$$-u''(x_i) = -\frac{2}{h_{i+\frac{1}{2}} + h_{i-\frac{1}{2}}}\left(\frac{u(x_{i+1}) - u(x_i)}{h_{i+\frac{1}{2}}} + \frac{u(x_{i-1}) - u(x_i)}{h_{i-\frac{1}{2}}}\right) + \varepsilon(h),$$

ce qui suggère le schéma suivant

$$\frac{2}{h_{i+\frac{1}{2}} + h_{i-\frac{1}{2}}}\Big[-\frac{u_{i+1} - u_i}{h_{i+\frac{1}{2}}} + \frac{u_i - u_{i-1}}{h_{i-\frac{1}{2}}}\Big] = f(x_i),\ i = 2, \ldots, N-1, \tag{1.138}$$

Exercice 13 — Volumes finis 1D, consistance des flux.

1. Si $u \in C^2([0,1])$, par développement de Taylor, pour $i = 1, \ldots, N$:

$$u(x_{i+1}) = u(x_i) + h_{i+1/2}u'(x_i) + \frac{1}{2}h_{i+1/2}^2u''(\xi_{i+1/2}).$$

Donc

$$|u'(x_i) - \frac{u(x_{i+1}) - u(x_i)}{h_{i+1/2}}| \leqslant h||u''||_\infty.$$

De plus,

$$|u'(x_{i+1/2}) - u'(x_i)| \leqslant h||u''||_\infty.$$

Par inégalité triangulaire, on obtient donc que

$$|u'(x_{i+1/2}) - \frac{u(x_{i+1}) - u(x_i)}{h_{i+1/2}}| \leqslant 2h||u''||_\infty,$$

ce qui prouve que le flux est consistant d'ordre 1.

2. Dans le cas où $u \in \mathcal{C}^3([0\,;1])$ et $x_{i+1/2} = \dfrac{(x_{i+1} + x_i)}{2}$, on a

$$x_{i+1} - x_{i+1/2} = x_{i+1/2} - x_i = \frac{h_{i+1/2}}{2}.$$

Un développement de Taylor entre x_{i+1} et $x_{i+1/2}$ puis entre $x_{i+1/2}$ et x_i donne alors

$$u(x_{i+1}) = u(x_{i+1/2}) + \frac{1}{2}h_{i+1/2}u'(x_{i+1/2}) + \frac{1}{8}h^2_{i+1/2}u''(x_{i+1/2}) + \frac{1}{24}h^3_{i+1/2}u'''(\eta_{i+1/2})$$
$$u(x_i) = u(x_{i+1/2}) - \frac{1}{2}h_{i+1/2}u'(x_{i+1/2}) + \frac{1}{8}h^2_{i+1/2}u''(x_{i+1/2}) - \frac{1}{24}h^3_{i+1/2}u'''(\zeta_{i+1/2}$$

avec $\eta_{i+1/2}, \in [x_{i+1/2}, x_{i+1}]$ et $\zeta_{i+1/2}, \in [x_i, x_{i+1/2}]$. Par soustraction, on obtient

$$\frac{1}{h_{i+1/2}}\big(u(x_{i+1}) - u(x_i)\big) - u'(x_{i+1/2}) = \frac{1}{24}h^2_{i+1/2}(u''(\eta_{i+1/2}) + u''(\zeta_{i+1/2})),$$

de sorte que

$$|\frac{u(x_{i+1}) - u(x_i)}{h_{i+1/2}} - u'(x_{i+1/2})| \leqslant \frac{1}{12}h^2||u''||_\infty.$$

Le flux est donc consistant d'ordre 2.

Exercice 16 — Conditions aux limites de Fourier (ou Robin).

1. La discrétisation par différences finies donne comme i–ème équation (voir par exemple exercice 15 :

$$\frac{1}{h^2}(2u_i - u_{i-1} - u_{i+1}) + c_i u_i = f_i \qquad i = 1, \ldots, N.$$

Il reste donc à déterminer les inconnues u_0 et u_{N+1} à l'aide de la discrétisation des conditions aux limites, qu'on approche par :

$$\frac{u_1 - u_0}{h} + \alpha(u_0 - \widetilde{u}) = 0,$$
$$\frac{u_{N+1} - u_N}{h} + \alpha(u_{N+1} - \widetilde{u}) = 0$$

où u_0 et u_{N+1} sont les valeurs approchées en x_0 et x_{N+1}, on a donc par élimination :

$$u_0 = \frac{1}{\alpha - 1/h}\left(\alpha\widetilde{u} - \frac{u_1}{h}\right) \text{ et } u_{N+1} = \frac{1}{\alpha + 1/h}\left(\alpha\widetilde{u} + \frac{u_N}{h}\right).$$

Ce qui termine la définition du schéma.

2. Par volumes finis, la discrétisation de l'équation s'écrit

$$F_{i+1/2} - F_{i-1/2} = h_i f_i, i = 1, \ldots, N,$$

et les seuls flux "nouveaux" sont encore $F_{1/2}$ et $F_{N+1/2}$, qu'on obtient à partir de la discrétisation des conditions aux limites. Ceci peut se faire de plusieurs manières. On peut, par exemple, discrétiser la condition aux limites en 0 par :

$$F_{1/2} + \alpha(u_0 - \widetilde{u}) = 0, \text{ avec } F_{1/2} = \frac{u_1 - u_0}{\frac{h_1}{2}}.$$

On a ddans ce cas $-\alpha(u_0-\widetilde{u})\times\frac{h_1}{2}=-u_1+u_0$ d'où on déduit que $u_0=\frac{\alpha\widetilde{u}h_1+2u_1}{\alpha h_1+2}$ et qui conduit à l'expression :

$$F_{1/2}=\frac{\alpha}{\alpha h_1+2}\left(2(u_1-\widetilde{u})-\alpha h_1\widetilde{u}\right).$$

Le calcul est semblable pour $F_{N+1/2}$

Exercice 19 — Problème elliptique 1D et discrétisation par volumes finis.

1. On intègre (1.108) sur une maille $[x_{i-1/2},x_{i+1/2}]$ et on obtient :

$$-u'(x_{i+1/2})+u'(x_{i-1/2})+a[u(x_{i+1/2})-u(x_{i-1/2})]+b\int_{x_{i-1/2}}^{x_{i+1/2}}u(x)dx=bh_if_i. \quad (1.139)$$

Pour justifier le schéma numérique proposé on remarque que :

$$u(x_{i+1})=u(x_{i+1/2})+(x_{i+1}-x_{i+1/2})u'(x_{i+1/2})+1/2(x_{i+1}-x_{i+1/2})^2u''(\xi_i)$$

avec $\xi_i\in[x_{i+1/2},x_{i+1}]$ et de même

$$u(x_i)=u(x_{i+1/2})+(x_i-x_{i+1/2})u'(x_{i+1/2})+1/2(x_i-x_{i+1/2})^2u''(\gamma_i)$$

avec $\gamma_i\in[x_{i-1/2},x_i]$, dont on déduit :

$$u(x_{i+1})-u(x_i)=h_{i+1/2}u'(x_{i+1/2})+\frac{1}{8}(h_{i+1}^2u''(\xi_i)-h_i^2u''(\gamma_i)).$$

De plus en utilisant le fait que x_i est le milieu de $[x_{i-1/2},x_{i+1/2}]$ on a (voir démonstration plus loin)

$$\int_{x_{i-1/2}}^{x_{i+1/2}}udx=u(x_i)h_i+\frac{1}{24}u''(\alpha_i)h_i^3 \quad (1.140)$$

D'où le schéma numérique.

Démontrons la formule (1.140). Pour cela il suffit (par changement de variable) de démontrer que si $u\in\mathcal{C}^2(\mathbb{R})$, alors pour tout $\alpha\geqslant 0$, on a :

$$\int_{-\alpha}^{\alpha}udx=2\alpha u(0)+\frac{1}{3}u''(\alpha_i)\alpha^3. \quad (1.141)$$

Pour cela, on utilise une formule de type Taylor avec reste intégral, qu'on obtient en remarquant que si on pose $\varphi(t)=u(tx)$, alors $\varphi'(t)=xu(tx)$, et $\varphi''(t)=x^2u''(tx)$. Or $\varphi(1)-\varphi(0)=\int_0^1\varphi'(t)dt$, et par inégration par parties, on obtient donc :

$$\varphi(1)=\varphi(0)+\varphi'(0)+\int_0^1\varphi''(t)(1-t)ds.$$

On en déduit alors que

$$u(x)=u(0)+xu'(0)+\int_0^1x^2u''(tx)(1-t)dt.$$

En intégrant entre $-\alpha$ et α, on obtient alors :

$$\int_{-\alpha}^{\alpha}u(x)\,\mathrm{d}x=2\alpha u(0)+A,\ \text{avec}\ A=\int_0^1x^2u''(tx)(1-t)\,\mathrm{d}t\,\mathrm{d}x.$$

Comme la fonction u'' est continue elle est minorée et majorée sur $[-\alpha,\alpha]$. Soient donc $m=\min_{[-\alpha,\alpha]}u''$ et $M=\min_{[-\alpha,\alpha]}u''$. Ces deux valeurs sont atteintes par u'' puisqu'elle est continue. On a donc $u''([-\alpha,\alpha])=[m,M]$. De plus, la fonction $(x,t)\mapsto x^2(1-t)$ est positive ou nulle sur $[-\alpha,\alpha]\times[0\,;1]$. On peut donc minorer et majorer A de la manière suivante

$$m\int_0^1x^2(1-t)\,\mathrm{d}t\ \,\mathrm{d}x\leqslant A\leqslant M\int_0^1x^2(1-t)\,\mathrm{d}t\,\mathrm{d}x.$$

Or

$$\int_0^1 x^2(1-t)\,\mathrm{d}t\ \ \mathrm{d}x = \frac{1}{3}\alpha^3$$

On en déduit que $\frac{1}{3}\alpha^3 m \leqslant A \leqslant \frac{1}{3}\alpha^3 M$, et donc que

$$A = \frac{1}{3}\alpha^3\gamma$$

avec $\gamma \in [m, M]$; mais comme u'' est continue, elle prend toutes les valeurs entre m et M, il existe donc $\beta \in [-\alpha, \alpha]$ tel que $\gamma = u''(\beta)$, ce qui termine la preuve de (1.141).

2. (a) On multiplie (1.111) par u_i et on somme pour $i = 1, \ldots, N$. On obtient après changement d'indice que

$$\sum_{i=0}^{i=N} \frac{(u_{i+1}-u_i)^2}{h_{i+1/2}} + \frac{a}{2}\sum_{i=0}^{i=N}(u_{i+1}-u_i)^2 + b\sum_{i=0}^{i=N} u_i^2 h_i = 0.$$

On a donc $u_i = 0$ pour tout $i = 1 \ldots N$, d'où en mettant le schéma sous la forme matricielle $AU = B$ on déduit que l'application linéaire représentée par la matrice A est injective donc bijective (grâce au fait qu'on est en dimension finie) et donc que (1.111) admet une unique solution.

3. (a) Evident.

(b) On pose $\bar{R} = A\bar{U} - B$. On a donc $R_i = R_i^{(1)} + R_i^{(2)}$, avec

$$R_i^{(1)} = -\frac{1}{h_i}\left[\left(\frac{\bar{u}_{i+1}-\bar{u}_i}{h_{i+1/2}} - u'(x_{i+1/2})\right) - \left(\frac{\bar{u}_i - \bar{u}_{i-1}}{h_{i-1/2}} - u'(x_{i-1/2})\right)\right],$$

$$R_i^{(2)} = \frac{a}{h_i}[(\bar{u}_i - u(x_{i+1/2})) - (\bar{u}_{i-1} - u(x_{i+1/2}))].$$

De plus on remarque que

$$\bar{u}_i = \frac{1}{h_i}\int_{x_{i-1/2}}^{x_{i+1/2}} u\,dx = u(x_{i+1/2}) - \frac{h_i}{2}u'(x_{i+1/2}) + \frac{h_i^2}{6}u''(x_{i+1/2}) - \frac{h_i^3}{24}u^{(3)}(d_i) \tag{1.142}$$

$$\bar{u}_{i+1} = \frac{1}{h_{i+1}}\int_{x_{i+1/2}}^{x_{i+3/2}} u\,dx = u(x_{i+1/2}) + \frac{h_{i+1}}{2}u'(x_{i+1/2}) + \frac{h_{i+1}^2}{6}u''(x_{i+1/2}) - \frac{h_{i+1}^3}{24}u^{(3)}(\delta_i) \tag{1.143}$$

avec $d_i \in [x_{i-1/2}, x_{i+1/2}]$ et $\delta_i \in [x_{i+1/2}, x_{i+3/2}]$. Ce qui implique que :

$$\frac{\bar{u}_{i+1}-\bar{u}_i}{h_{i+1/2}} = u'(x_{i+1/2}) + \frac{1}{3}u''(x_{i+1/2})(h_{i+1}-h_i) + \frac{1}{24}\ \frac{1}{h_{i+1/2}}\left[u^{(3)}(\delta_i)h_{i+1}^3 + u^{(3)}(d_i)h_i^3\right]$$

et donc

$$\frac{1}{h_i}\left[\frac{\bar{u}_{i+1}-\bar{u}_i}{h_{i+1/2}} - u'(x_{i+1/2})\right] = S_i + K_i,$$

avec

$$|S_i| = |\frac{1}{3}u''(x_{i+1/2})\left(\frac{h_{i+1}}{h_i} - 1\right) + \frac{1}{24}| \leqslant Ch$$

et

$$|K_i| = |\frac{1}{h_i h_{i+1/2}}\left[u^{(3)}(\delta_i)h_{i+1}^3 + u^{(3)}(d_i)h_i^3\right]| \leqslant Ch,$$

où C ne dépend que de u. De plus si on pose $L_i = (\bar{u}_i - u(x_{i+1/2}))/h_i$ par développement de Taylor, il existe $\widetilde{C}$ ne dépendant que de u telle que $|L_i| \leqslant Ch$. Finalement on conclut que $|R_i^{(1)}| = |-S_i + S_{i+1}| \leqslant C_1$ et $|R_i^{(2)}| = |-K_i + K_{i-1} + a(L_i - L_{i-1})| \leqslant C_2 h$.

4. Reprendre les résultats précédents. Pour $|R_{i+1/2}| \leqslant Ch$ reprendre calcul du 3 $|R_{i+1/2}| = |h_i(-S_i - K_i + L_i)|$.

5. **(a)** On pose $e_i = \bar{u}_i - u_i$. Cette définition implique que e_i est solution du système (1.121)-(1.123).

(b) Un calcul similaire à celui de la question 2. donne que

$$b\sum_1^N h_i e_i + \sum_{i=0}^{i=N} \frac{(e_{i+1}-e_i)^2}{h_{i+1/2}} + \frac{a}{2}\sum_{i=0}^{i=N}(e_{i+1}-e_i)^2 = \sum_{i=0}^{i=N} R_{i+1/2}(e_{i+1}-e_i)$$

D'où en utilisant le fait que $\alpha h \leqslant h_i \leqslant \beta h$ et l'inégalité de Cauchy-Schwarz on déduit que

$$\frac{1}{\beta h}\sum_{i=0}^{i=N}(e_{i+1}-e_i)^2 \leqslant \left(\sum_{i=0}^{i=N}(e_{i+1}-e_i)^2\right)^{1/2}\left(\sum_{i=0}^{i=N} R_{i+1/2}^2\right)^{1/2}$$

et en utilisant (1.120), et le fait que $\sum_{i=0}^{i=N} h_i = 1$ entraine $N \leqslant \frac{1}{\alpha h}$, on déduit :

$$\sum_{i=0}^{i=N}(e_{i+1}-e_i)^2 \leqslant C_1\frac{\beta}{\alpha}h^3.$$

En remarquant que $e_i = \sum_{j=0}^{j=i-1}(e_{j+1}-e_j)$ on a pour tout $0 < i \leqslant N$ que

$$|e_i| \leqslant \left(\sum_{j=0}^{j=i}(e_{j+1}-e_j)^2\right)^{1/2} i^{1/2} \leqslant \left(C_1\frac{\beta}{\alpha}h^3\right)^{1/2} N^{1/2}$$

et donc $|e_i| \leqslant \frac{\sqrt{C_1\beta}}{\alpha}h$, pour tout $0 < i \leqslant N$.

Exercice 22 — Problème elliptique 2d et différences finies. On note (x, y) les coordonnées d'un point de $\mathbb{R}^2$.

1. En multipliant la première équation de (1.133) par φ et en intégrant par parties, on trouve, pour tout $\varphi \in C^1(\bar{\Omega})$ tel que $\varphi = 0$ sur $\partial\Omega$:

$$\int_\Omega \nabla u(x,y)\cdot\nabla\varphi(x,y)\ \mathrm{d}x\,\mathrm{d}y + \int_\Omega k\frac{\partial u(x,y)}{\partial x}\varphi(x,y)\ \mathrm{d}x = \int_\Omega f(x,y)\varphi(x,y)\,\mathrm{d}x\,\mathrm{d}y. \quad (1.144)$$

On suppose maintenant que $f \leqslant 0$ sur Ω. On se donne une fonction $\psi \in C^1(\mathbb{R},\mathbb{R})$ tel que $\psi(s) = 0$ si $s \leqslant 0$ et $\psi(s) > 0$ si $s > 0$. (On peut choisir, par exemple, $\psi(s) = s^2$ pour $s > 0$ et $\psi(s) = 0$ pour $s \leqslant 0$) et on prend dans (1.144) $\varphi = \psi \circ u$. On obtient ainsi :

$$\int_\Omega \psi'(u(x,y))\,|\nabla u(x,y)|^2\ \mathrm{d}x\,\mathrm{d}y + \int_\Omega k\frac{\partial u}{\partial x}(x,y)\psi(u(x,y))\,\mathrm{d}x = \int_\Omega f(x,y)\psi(u(x,y))\ \mathrm{d}x\,\mathrm{d}y \leqslant 0. \quad (1.145)$$

En notant G la primitive de ψ s'annulant en 0, on a :

$$\frac{\partial}{\partial x}G(u(x,y)) = \psi(u(x,y))\frac{\partial u}{\partial x}(x,y)$$

Comme $u = 0$ sur $\partial\Omega$, on obtient donc :

$$\int_\Omega k\frac{\partial u}{\partial x}(x,y)\psi(u(x,y))\ \mathrm{d}x\,\mathrm{d}y = \int_\Omega k\frac{\partial}{\partial x}G(u(x,y))\ \mathrm{d}x\,\mathrm{d}y = \int_{\partial\Omega} kG(u(x,y))n_x\ d\gamma(x,y) = 0,$$

où n_x désigne la première composante du vecteur normal $\boldsymbol{n}$ à $\partial\Omega$ extérieur à Ω, et $d\gamma(x,y)$ le symbole d'intégration par rapport à la mesure de Lebesgue unidimensionnelle sur $\partial\Omega$. De (1.145) on déduit alors :

$$\int_\Omega \psi'(u(x,y))\,|\nabla u(x,y)|^2\ \mathrm{d}x\,\mathrm{d}y \leqslant 0,$$

et donc, comme $\psi' \geqslant 0$ et que la fonction $(x,y) \mapsto \psi'(u(x,y))|\nabla u(x,y)|^2$ est continue :

$$\psi'(u(x,y))|\nabla u(x,y)|^2 = 0, \forall (x,y) \in \bar{\Omega}.$$

Ceci donne aussi

$$\nabla\psi(u(x,y)) = 0, \forall (x,y) \in \bar{\Omega}.$$

La fonction $\psi \circ u$ est donc constante sur $\bar{\Omega}$, comme elle est nulle sur $\partial\Omega$, elle est nulle sur $\bar{\Omega}$, ce qui donne $u \leqslant 0$ sur $\bar{\Omega}$.

2. On s'intéresse ici à la consistance au sens des différences finies. On pose donc $\bar{u}_{i,j} = u(ih, jh)$ pour $i, j \in \{0, \dots, N+1\}^2$. On a bien $\bar{u}_{i,j} = 0$ pour $(i,j) \in \gamma$, et pour $(i,j) \in \{1, \dots, N\}^2$, on pose $R_{ij} = a_0\bar{u}_{i,j} - a_1\bar{u}_{i-1,j} - a_2\bar{u}_{i+1,j} - a_3\bar{u}_{i,j-1} - a_4\bar{u}_{i,j+1} - f_{i,j}$.

3. On rappelle que u est solution de (1.133), R_{ij} est donc l'erreur de consistance du schéma. Dans le cas du schéma [I] on a :

$$R_{ij} = \frac{2\bar{u}_{ij} - \bar{u}_{i+1,j} - \bar{u}_{i-1,j}}{h^2} + \frac{2\bar{u}_{ij} - \bar{u}_{i,j+1} - \bar{u}_{i,j-1}}{h^2} + k\frac{\bar{u}_{i+1,j} - \bar{u}_{i-1,j}}{2h} - f_{ij}.$$

Comme $u \in \mathcal{C}^4(\bar{\Omega})$, il existe $\xi_{ij} \in]0\,;1[$ et $\eta_{ij} \in]0\,;1[$ tel que :

$$\bar{u}_{i+1,j} = \bar{u}_{i,j} + h\partial_x u(ih, jh) + \frac{h^2}{2}\partial_x^2 u(ih, jh) + \frac{h^3}{6}\partial_x^3 u(ih, jh) + \frac{h^4}{24}\partial_x^4 u(ih + \xi_{ij}h, jh) \tag{1.146}$$

$$\bar{u}_{i-1,j} = \bar{u}_{i,j} - h\partial_x u(ih, jh) + \frac{h^2}{2}\partial_x^2 (ih, jh) - \frac{h^3}{6}\partial_x^3 u(ih, jh) + \frac{h^4}{24}\partial_x^4 u(ih + \eta_{ij}h, jh). \tag{1.147}$$

On obtient des formules analogues pour $\bar{u}_{i,j+1}$ et $\bar{u}_{i,j-1}$; de plus, comme u est solution de (1.133), on a

$$f_{ij} = -\partial_x^2 u(ih, jh) - \partial_y^2 u(ih, jh) + k\partial_x u(ih, jh).$$

On en déduit

$$|R_{i,j}| \leqslant \frac{h^2}{12}\|\partial_x^4 u\|_\infty + \frac{h^2}{12}\|\partial_y^4 u\|_\infty + k\frac{h^2}{6}\|\partial_x^3 u\|_\infty$$

où $\|\partial_x^4 u\|_\infty$ désigne la norme uniforme sur $\bar{\Omega}$ de la dérivée quatrième de u par rapport à x (notations analogues pour $\|\partial_y^4 u\|_\infty$ et $\|\partial_x^3 u\|_\infty$). On obtient finalement $|R_{ij}| \leqslant C_1 h^2$ où C_1 ne dépend que de u et k. Comme pour h petit, on a $h^2 \leqslant h$, on en déduit que schéma [I] est donc d'ordre 2.

Pour le schéma [II], on a :

$$R_{ij} = \frac{2\bar{u}_{ij} - \bar{u}_{i+1,j} - \bar{u}_{i-1,j}}{h^2} + \frac{2\bar{u}_{ij} - \bar{u}_{i,j+1} - \bar{u}_{i,j-1}}{h^2} + k\frac{\bar{u}_{ij} - \bar{u}_{i-1,j}}{h} - f_{ij}.$$

D'où l'on déduit

$$|R_{ij}| \leqslant \frac{h^2}{12}\|\partial_{xxxx}^4 u\|_\infty + \frac{h^2}{12}\|\partial_{yyyy}^4 u\|_\infty + \frac{kh}{2}\|\partial_{xx}^2 u\|_\infty,$$

et donc $|R_{ij}| \leqslant C_2 h$ où C_2 ne dépend que de u et k. Le schéma [II] est donc d'ordre 1.

4. Dans le cas du schéma [II], la famille $\{w_{ij}, i, j \in \{0, \dots, N+1\}^2\}$ vérifie $\forall i, j \in \{1, \dots, N\}$:

$$\frac{1}{h^2}(w_{ij} - w_{i+1,j}) + \left(\frac{1}{h^2} + \frac{k}{h}\right)(w_{ij} - w_{i-1,j}) + \frac{1}{h^2}(w_{ij} - w_{i,j+1}) + \frac{1}{h^2}(w_{ij} - w_{i,j-1}) \leqslant 0$$

On pose $M = \max\{w_{i,j}, (i,j) \in \{0, \dots, N+1\}^2\}$ et $m = \max\{w_{i,j}, (i,j) \in \gamma\}$. Noter que $\gamma = \{0, \dots, N+1\}^2 \setminus \{1, \dots, N\}^2$. On a bien sûr $m \leqslant M$ et il reste donc à montrer que $M \leqslant m$. Soit $A = \{(i,j) \in \{0, \dots, N+1\}^2, w_{i,j} = M\}$ et soit $(\bar{ß}, \bar{æ}) \in A$ tel que $\bar{i} = \max\{i\ ;\ (i,j) \in A\}$ et $\bar{j} = \max\{j\ ;\ (i,j) \in A\}$. On distingue deux cas :

— Si $\bar{i} \in \{0, N+1\}$ ou $\bar{j} \in \{0, \dots, N+1\}$, on a alors $(\bar{ß}, \bar{j}) \in \gamma$ et donc $M = w_{\bar{ß},\bar{j}} \leqslant m$.

— Sinon, on a $\bar{i} \notin \{0, N+1\}$ et $\bar{j} \notin \{0, N+1\}$, et donc $(\bar{i}, \bar{j}) \in \{1, \dots, N\}^2$. On en déduit que :

$$\frac{1}{h^2}\left(w_{\bar{i},\bar{j}} - w_{\bar{i}+1,\bar{j}}\right) + \left(\frac{1}{h^2} + \frac{k}{h}\right)\left(w_{\bar{i},\bar{j}} - w_{\bar{i}-1,\bar{j}}\right) + \frac{1}{h^2}\left(w_{\bar{i},\bar{j}} - w_{\bar{i},\bar{j}+1}\right) + \frac{1}{h^2}\left(w_{\bar{i},\bar{j}} - w_{\bar{i},\bar{j}-1}\right) \leqslant 0, \quad (1.148)$$

ce qui est impossible car $w_{\bar{i},\bar{j}} = M$ et donc

$$\begin{aligned} w_{\bar{i},\bar{j}} - w_{\bar{i},\bar{j}-1} &\geqslant 0, \\ w_{\bar{i},\bar{j}} - w_{\bar{i},\bar{j}+1} &\geqslant 0, \\ w_{\bar{i},\bar{j}} - w_{\bar{i}-1,\bar{j}} &\geqslant 0, \\ w_{\bar{i},\bar{j}} - w_{\bar{i}+1,\bar{j}} &> 0, \end{aligned}$$

noter que la dernière inégalité est bien stricte car $(\bar{i}+1, \bar{j}) \notin A$ (c'est l'intérêt du choix de $\bar{i}$). On a donc bien montré que $M \leqslant m$.

Dans le cas du schéma [II], si on a $\bar{i} \notin \{0, N+1\}$ et $\bar{j} \notin \{0, N+1\}$, et donc $(\bar{i}, \bar{j}) \in \{1, \dots, N\}^2$, le même raisonnement que celui du schéma 1 donne :

$$\left(\frac{1}{h^2} - \frac{k}{2h}\right)\left(u_{\bar{i}\bar{j}} - u_{\bar{i}+1,\bar{j}}\right) + \left(\frac{1}{h^2} + \frac{h}{2h}\right)\left(u_{\bar{i},\bar{j}} - u_{\bar{i}-1,\bar{j}}\right) + \frac{1}{h^2}\left(u_{\bar{i}\bar{j}} - u_{\bar{i},\bar{j}+1}\right) + \frac{1}{h^2}\left(u_{\bar{i}\bar{j}} - u_{\bar{i},\bar{j}-1}\right) \leqslant 0. \quad (1.149)$$

On ne peut conclure à une contradiction que si $\frac{1}{h^2} - \frac{k}{2h} \geqslant 0$. Le schéma [II] vérifie

$$w_{i,j} \leqslant \max_{(k,\ell)\in\gamma}(w_{k,\ell}) \qquad \forall i, j \in \{1, \dots, N\}^2$$

lorsque h vérifie la condition (dite condition de CFL (pour Courant-Levy-Friedriechs, voir section 5.3.1) :

$$h \leqslant \frac{2}{k} \quad (1.150)$$

5. La fonction ϕ vérifie $\phi_x = 0$, $\phi_y = y$ et $\phi_{yy} = 1$ et donc $-\Delta\phi + k\frac{\partial\phi}{\partial x} = -1$. On pose maintenant $\phi_{i,j} = \phi(ih, jh)$ pour $i, j \in \{a, \dots, N+1\}^2$ (Noter que ϕ ne vérifie pas la condition $\phi_{i,j} = 0$ si $(i,j) \in \gamma$). Comme $\phi_{xx} = \phi_{xxx} = \phi_{xxxx} = \phi_{yyyy} = 0$, les calculs de la question 2 montrent que pour les schémas [I] et [II] :

$$a_0\phi_{i,j} - a_1\phi_{i-1,j} - a_2\phi_{i+1,j} - a_3\phi_{i,j-1} - a_4\phi_{i,j+1} = -1$$

pour $i, j \in \{1, \dots, N\}^2$. En posant $w_{i,j} = u_{i,j} + C\phi_{i,j}$ pour $(1,j) \in \{0, \dots, N+1\}^2$ (et U solution de (1.134)) on a donc :

$$a_0 w_{i,j} - a_1 w_{i-1,j} - a_2 w_{i+1,j} - a_3 w_{i,j-1} - a_4 w_{i,j+1} = f_{ij} - C \quad \forall i, j \in \{1, \dots, N\}$$

On prend $C = \|f\|_\infty$, de sorte que $f_{i,j} - C \leqslant 0$ pour tout (i,j) pour le schéma [II] et pour le schéma [I] avec $h \leqslant 2/k$, la question 3 donne alors pour $(i,j) \in \{1, \dots, N\}^2$,

$$w_{ij} \leqslant \max\{w_{k\ell}, (k\ell) \in \gamma\} \leqslant \frac{C}{2}$$

car $u_{i,j} = 0$ si $(i,j) \in \gamma$ et $-\max_\Omega \phi = 1/2$. On en déduit pour $(i,j) \in \{1, \dots, N\}^2$, $w_{ij} \leqslant C/2 = 1/2\|f\|_\infty$. Pour montrer que $-w_{ij} \leqslant 1/2\|f\|_\infty$, on prend maintenant $w_{ij} = C\phi_{ij} - u_{i,j}$ pour $(i,j) \in \{0, \dots, N+1\}^2$, avec $C = \|f\|_\infty$. On a donc

$$a_0 w_{i,j} - a_1 w_{i-1,j} - a_2 w_{i+1,j} - a_3 w_{i,j-1} - a_4 w_{i,j+1} = -C - f_{i,j} \leqslant 0, \forall i, j \in \{1, \dots, N\}.$$

Ici encore, pour le schéma [II] ou le schéma [I] avec la condition $h \leqslant 2/k$, la question 3 donne

$$w_{ij} \leqslant \max\{w_{k\ell}, (k\ell) \in \gamma\} = \frac{C}{2}$$

donc $u_{ij} \geqslant -C/2 = -\|f\|_\infty/2$ pour tout$(i,j) \in \{1,\dots,N\}^2$. Pour le schéma [II] ou le schéma [I] avec la condition $h \leqslant 2/k$, on a donc :

$$\|U\|_\infty \leqslant 1/2\|f\|_\infty. \tag{1.151}$$

Le système (1.134) peut être vu comme un système linéaire de N^2 équation, à N^2 inconnues (qui sont les $u_{i,j}$ pour $(i,j) \in \{1,\dots,N\}^2$). Si le second membre de ce système linéaire est nul, l'inégalité (1.151)(I) prouve que la solution est nulle. Le système (1.134) admet donc, pour tout f, au plus une solution. Ceci est suffisant pour affirmer qu'il admet, pour tout f, une et une seule solution.

6. Pour $(i,j) \in \{0,\dots,N+1\}^2$ on pose $e_{ij} = u(ih,jh) - u_{i,j}$. On a donc, pour les schémas [I] et [II], avec les notations de la question 2 :

$$a_0 e_{ij} - a_1 e_{i-1,j} - a_2 e_{i+1,j} - a_3 e_{i,j-1} - a_4 e_{i,j+1} = R_{ij}, \quad \forall i,j \in \{1,\dots,N\}^2.$$

avec les questions 2 et 4, on a donc, pour le schéma [I], si $h \leqslant 2/k$:

$$\max\{|e_{ij}|, (i,j) \in \{1,\dots,N\}^2\} \leqslant 1/2C_1 h^2,$$

où C_1 et C_2 ne dépendent que de u et k (et sont données à la question 2). Les deux schémas sont convergents. Le schéma [I] converge en "h^2" et le schéma [II] en "h".

7. Le schéma [1] converge plus vite mais a une condition de stabilité $k \leqslant 2/h$. Le schéma [II] est inconditionnellement stable.

Exercice 23 — Implantation de la méthode des volumes finis.

1. Le problème complet s'écrit :

$$\begin{cases} -\operatorname{div}(\mu_i \nabla\phi)(x) = 0 & x \in \Omega_i \qquad i = 1,2 \\ \nabla\phi(x) \cdot \boldsymbol{n}(x) = 0 & x \in \Gamma_2 \cup \Gamma_4 \\ \displaystyle\int_{\Gamma_1} \mu_1 \nabla\phi(x) \cdot \boldsymbol{n}(x) d\gamma(x) + \int_{\Gamma_3} \mu_2 \nabla\phi(x) \cdot \boldsymbol{n}(x) d\gamma(x) = 0 & \\ \phi_2(x) - \phi_1(x) = 0 & \forall x \in I \\ -(\mu\nabla\phi \cdot \boldsymbol{n})|_2(x) - (\mu\nabla\phi \cdot \boldsymbol{n})|_1(x) = 0 & \forall x \in I \end{cases} \tag{1.152}$$

2. On se donne le même maillage rectangulaire uniforme qu'à l'exercice précédent. On note ϕ_K l'inconnue associée à la maille K ou ϕ_k si on référence la maille K par son numéro $k = n(j-1)+i$ où $i \in \{1,\dots,n\}$ et $j \in \{1,\dots,2p\}$. Pour une maille intérieure, l'équation obtenue est la même que (1.154) en remplaçant λ_m par μ_m.

Etudions maintenant le cas d'une maille proche de l'interface. Comme indiqué, on va considérer deux inconnues discrètes par arête de l'interface. Soient K et L ayant en commun l'arête $\widetilde{\sigma} \subset I$, K est située au dessous de L. Les équations associées à K et L s'écrivent alors

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} = 0 \quad \text{et} \quad \sum_{\sigma\in\mathcal{E}_L} F_{L,\sigma} = 0.$$

Pour les arêtes $\sigma \in \mathcal{E}_K$ autres que $\widetilde{\sigma}$, le flux s'écrit de manière habituelle

$$F_{K,\sigma} = \mu_1 \frac{\phi_K - \phi_M}{d_\sigma}$$

avec $\sigma = K|M$. Pour l'arête $\sigma = \widetilde{\sigma}$, on a $F_{K,\sigma} = \mu_1 \frac{\phi_K - \phi_\sigma}{d_\sigma} m(\sigma)$ et $F_{L,\sigma} = \mu_2 \frac{\phi_L - \phi_\sigma^+}{d_\sigma} m(\sigma)$, où les deux inconnues discrètes ϕ_σ^+ et ϕ_σ^- sont reliées par les relations :

$$\phi_\sigma^+ - \phi_\sigma^- = \psi_\sigma \left(= \frac{1}{m(\sigma)} \int_\sigma \psi(x) d\gamma(x) \right)$$
$$F_{K,\sigma} + F_{L,\sigma} = 0.$$

On peut alors éliminer ϕ_σ^+ et ϕ_σ^- ; en utilisant par exemple $\phi_\sigma^+ = \psi_\sigma + \phi_\sigma^-$ et en remplaçant dans la deuxième équation, on obtient :

$$-\mu_1 \frac{\phi_\sigma^- - \phi_K}{d_{K,\sigma}} + \mu_2 \frac{\phi_\sigma^- + \psi_\sigma - \phi_L}{d_{L,\sigma}} = 0,$$

ce qui donne :

$$\phi_{\sigma^-} = \frac{1}{\frac{\mu_1}{d_{K,\sigma}} + \frac{\mu_2}{d_{L,\sigma}}} \Big(\frac{\mu_1}{d_{K,\sigma}} \phi_K + \frac{\mu_2}{d_{L,\sigma}} \phi_L - \frac{\mu_2}{d_{L,\sigma}} \psi_{\sigma.} \Big).$$

En remplaçant cette expression dans les flux, on obtient :

$$F_{K,\sigma} = -F_{L,\sigma} = m(\sigma) \frac{\mu_1 \mu_2}{\mu_1 d_{L,\sigma} + \mu_2 d_{K,\sigma}} (\phi_K - \phi_L + \psi_\sigma)$$

On peut alors écrire l'équation discrète associée à une maille de numéro k située sous l'interface (avec $k = n(p-1) + i$, $i = 2, \ldots, n-1$). On pose :

$$\frac{\mu_1 \mu_2}{\mu_1 d_{L,\sigma} + \mu_2 d_{K,\sigma}} = \frac{\mu_I}{d_\sigma}$$

(μ_I est donc la moyenne harmonique pondérée entre μ et μ_2). Notons que pour une arête de I, $d_\sigma = h_y$, et $m(\sigma) = h_x$. L'équation associée à la maille k s'écrit donc :

$$\Big(2\mu_1 \frac{h_y}{h_x} + \mu_1 \frac{h_x}{h_y} + \frac{\mu_I h_x}{h_y} \Big) u_k - \mu_1 \frac{h_y}{h_x} (u_{k-1} + u_{k+1}) - \mu_1 \frac{h_x}{h_y} u_{k-n} - \mu_I \frac{h_x}{h_y} u_{k+n} = -\mu_I \frac{h_x}{h_y} \psi_i$$

où ψ_i est le saut de potentiel à travers l'arête σ_i de l'interface considérée ici. De même, l'équation associée à une maille k avec $k = np + i$, $i = 2, \ldots, n-1$ située au dessus de l'interface s'écrit :

$$\Big(2\mu_1 \frac{h_y}{h_x} + \mu_1 \frac{h_x}{h_y} + \mu_I \frac{h_x}{h_y} \Big) u_k - \mu_1 \frac{h_y}{h_x} (u_{k-1} + u_{k+1}) - \mu_1 \frac{h_x}{h_y} u_{k+n} - \mu_I \frac{h_x}{h_y} u_{k-n} = +\mu_I \frac{h_x}{h_y} \psi_i.$$

La discrétisation des conditions aux limites de Neumann sur Γ_2 et Γ_4 est effectuée de la même manière que pour le cas du problème thermique, voir exercice 24.

Il ne reste plus qu'à discrétiser la troisième équation du problème (1.152), qui relie les flux sur la frontière Γ_1 avec les flux sur la frontière Γ_3. En écrivant la même condition avec les flux discrets, on obtient :

$$\mu_1 \sum_{i=1}^{n} \frac{2h_x}{h_y} (u_i - u_{B,i}) + \mu_2 \sum_{i=1}^{n} \frac{2h_x}{h_y} (u_{H,i} - u_{k(i)}) = 0,$$

où $\mu_{B,i}$ représente l'inconnue discrète sur la i–ème arête de Γ_1 et $\mu_{H,i}$ l'inconnue discrète sur la i–ème arête de Γ_3 et $k(i) = n(p-1) + i$ est le numéro de la maille jouxtant la i–ème arête de Γ_3.

Remarquons que tel qu'il est posé, le système n'est pas inversible : on n'a pas assez d'équations pour éliminer les inconnues $u_{B,i}$ et $u_{H,i}, i = 1 \ldots N$. On peut par exemple pour les obtenir considérer une différence de potentiel fixée entre Γ_1 et Γ_3, et se donner un potentiel fixé sur Γ_1.

Exercice 24 — Élimination des inconnues d'arêtes.

1. On a vu au paragraphe 1.4.2 que si σ est une arête du volume de contrôle K, alors le flux numérique $F_{K,\sigma}$ s'écrit :

$$F_{K,\sigma} = \lambda_i \frac{u_\sigma - u_K}{d_{K,\sigma}} m(\sigma). \tag{1.153}$$

On cherche à éliminer les inconnues auxiliaires u_σ. Pour cela, si σ est une arête interne, $\sigma = K|L$, on écrit la conservativité du flux numérique $F_{K,\sigma} = -F_{L,\sigma}$, ce qui entraîne, si σ n'est pas une arête de l'interface I, que :

$$-\lambda_i \frac{u_\sigma - u_K}{d_{K,\sigma}} m(\sigma) = \lambda_i \frac{u_\sigma - u_L}{d_{L,\sigma}} m(\sigma)$$

On en déduit que

$$u_\sigma \Big(\frac{1}{d_{K,\sigma}} + \frac{1}{d_{L,\sigma}} \Big) = \frac{u_K}{d_{K,\sigma}} + \frac{u_L}{d_{L,\sigma}},$$

soit encore que

$$u_\sigma = \frac{d_{K,\sigma} d_{L,\sigma}}{d_\sigma} \Big(\frac{u_K}{d_{K,\sigma}} + \frac{u_L}{d_{L,\sigma}} \Big).$$

Remplaçons alors dans (1.153). On obtient :

$$F_{K,\sigma} = \lambda_i\Big(\frac{d_{L,\sigma}}{d_\sigma}\Big(\frac{u_K}{d_{K,\sigma}} + \frac{u_L}{d_{L,\sigma}}\Big) - \frac{u_K}{d_{K,\sigma}}\Big) = -\frac{\lambda_i}{d_\sigma}\Big(\frac{d_{L,\sigma}}{d_{K,\sigma}}u_K + u_L - u_K - \frac{d_{L,\sigma}}{d_{K,\sigma}}u_K\Big)$$

On obtient donc finalement bien la formule (1.73).

2. Considérons maintenant le cas d'une arête $\sigma \subset \Gamma_1 \cup \Gamma_3$, où l'on a une condition de Fourier, qu'on a discrétisée par :

$$F_{K,\sigma} = -m(\sigma)\lambda_i\frac{u_\sigma - u_K}{d_{K,\sigma}} = m(\sigma)\alpha(u_\sigma - u_{ext}).$$

On a donc

$$u_\sigma = \frac{1}{\alpha + \frac{\lambda_i}{d_{K,\sigma}}}\Big(\frac{\lambda_i u_K}{d_{K,\sigma}} + \alpha u_{ext}\Big)$$

On remplace cette expression dans l'égalité précédente. Il vient :

$$F_{K,\sigma} = \frac{m(\sigma)\alpha}{\alpha + \frac{\lambda_i}{d_{K,\sigma}}}\Big(\frac{\lambda_i}{d_{K,\sigma}}u_K + \alpha u_{ext} - \alpha u_{ext} - \frac{\lambda_i}{d_{K,\sigma}}u_{ext}\Big),$$

Ce qui, après simplifications, donne exactement (1.74).

3. Considérons maintenant une arête $\sigma = K|L$ appartenant à l'interface I. La discrétisation de la condition de saut de flux sur I. S'écrit :

$$F_{K,\sigma} + F_{L,\sigma} = \int_\sigma \theta(x)d\gamma(x) = m(\sigma)\theta_\sigma$$

Supposons que K (resp. L) soit situé dans le milieu de conductivité (resp. λ_2). En remplaçant $F_{K,\sigma}$ et $F_{L,\sigma}$ par leurs expressions, on obtient :

$$-\lambda_i m(\sigma)\frac{u_\sigma - u_K}{d_{K,\sigma}} - \lambda_2 m(\sigma)\frac{u_\sigma - u_L}{d_{L,\sigma}} = m(\sigma)\theta_\sigma.$$

On en déduit que

$$u_\sigma\Big(\frac{\lambda_1}{d_{K,\sigma}} + \frac{\lambda_2}{d_{L,\sigma}}\Big) = \Big(\frac{\lambda_1 u_K}{d_{K,\sigma}} + \frac{\lambda_2 u_L}{d_{L,\sigma}} - \theta_\sigma\Big).$$

En remplaçant u_σ dans l'expression de $F_{K,\sigma}$, on obtient :

$$F_{K,\sigma} = -\frac{m(\sigma)}{d_{K,\sigma}}\lambda_1\frac{1}{\frac{\lambda_1}{d_{K,\sigma}} + \frac{\lambda_2}{d_{L,\sigma}}}\Big(\frac{\lambda_1 u_K}{d_{K,\sigma}} + \frac{\lambda_2 u_L}{d_{L,\sigma}} - \theta_\sigma - \frac{\lambda_1 u_K}{d_{K,\sigma}} - \frac{\lambda_2 u_K}{d_{L,\sigma}}\Big).$$

En simplifiant, on obtient :

$$F_{K\sigma} = -\frac{m(\sigma)\lambda_1}{\lambda_1 d_{L,\sigma} + \lambda_2 d_{K,\sigma}}\Big(\lambda_2 u_L - \lambda_2 u_K - d_{L,\sigma}\theta_\sigma\Big),$$

ce qui est exactement (1.76). On obtient alors l'expression $F_{L,\sigma} = m(\sigma)\theta_\sigma - F_{K,\sigma}$ ce qui donne, après simplifications :

$$F_{L,\sigma} = \frac{\lambda_2 m(\sigma)}{\lambda_1 d_{L,\sigma} + \lambda_2 d_{K,\sigma}}\left[\lambda_1(u_L - u_K) + d_{K,\sigma}\theta_\sigma\right].$$

On vérifie bien que $F_{K\sigma} + F_{L,\sigma} = m(\sigma)\theta_\sigma$.

4. Le système linéaire que satisfont les inconnues $(u_K)_{K\in\mathcal{M}}$ s'écrit $AU = b$ avec $U = (u_K)_{K\in\mathcal{T}}$. Pour construire les matrices A et b, il faut se donner une numérotation des mailles. On suppose qu'on a $n \times 2p$ mailles ; on considère un maillage uniforme du type de celui décrit sur la figure 1.5 et on note $h_x = 1/n$ (resp. $h_y = 1/p$) la longueur de la maille dans la direction x (resp. y). Comme le maillage est cartésien, il est facile de numéroter les mailles dans l'ordre "lexicographique" ; c'est-à-dire que la k–ème maille a comme centre le point $x_{i,j} = ((i-1/2)h_x, (j-1/2)h_y)$, avec

$k = n(j-1)+i$. On peut donc déterminer le numéro de la maille (et de l'inconnue associée) k à partir de la numérotation cartésienne (i,j) de la maille.

$$k = n(j-1)+i$$

Remarquons que, comme on a choisi un maillage uniforme, on a pour tout $K \in \mathcal{T}$: $m(K) = h_x h_y$, pour toute arête intérieure verticale σ : $d_\sigma = h_x m(\sigma) = h_y$ et pour toute arête intérieure horizontale, $d_\sigma = h_y$ et $m(\sigma) = h_x$. Pour chaque numéro de maille, nous allons maintenant construire l'équation correspondante.

— Mailles internes — $i = 2,\dots,n-1$; $j = 2,\dots,p-1,p+1,\dots,2p-1$. L'équation associée à une maille interne K s'écrit

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} = m(K) f_K.$$

Avec l'expression de $F_{K,\sigma}$ donnée par (1.73), ceci amène à :

$$2\lambda_m\left(\frac{h_x}{h_y}+\frac{h_y}{h_x}\right)u_k - \lambda_m\frac{h_x}{h_y}(u_{k-n}+u_{k+n}) - \lambda_m\frac{h_y}{h_x}(u_{k+1}+u_{k-1}) = h_x h_y f_k, \quad (1.154)$$

avec $m = 1$ si $j \leqslant p-1$ et $m = 2$ si $j \geqslant p+1$.

— Mailles du bord Γ_2 — Les mailles du bord Γ_2 sontrepérées par les indices $(n,j), j = 2$ à $p-1$, $j = p+1$ à $2p-1$, (on exclut pour l'instant les coins). L'équation des flux est la même que pour les mailles internes, mais le flux sur la frontière Γ_2 est nul. Ceci donne :

$$\lambda_m\left(2\frac{h_x}{h_y}+\frac{h_y}{h_x}\right)u_k - \lambda_m\frac{h_x}{h_y}(u_{k-n}+u_{k+n}) - \lambda_m\frac{h_y}{h_x}u_{k-1} = h_x h_y f_k,$$

avec $k = n(j-1)+n, j = 2$ à $p-1, j = p+1$ à $2p-1$ et $m = 1$ si $j \leqslant p-1, m = 2$ si $j \geqslant p+1$.

— Mailles de bord Γ_4 — Les mailles du bord Γ_4 sont repérées par les indices $(1,j), j = 2$ à $p-1, j = p+1$ à $2p-1$. Pour ces mailles, il faut tenir compte du fait que sur une arête de Γ_4, le flux $F_{K,\sigma}$ est donné par :

$$F_{K,\sigma} = -\lambda_m\frac{g_\sigma - u_K}{d_{K,\sigma}}m(\sigma)$$

avec $g_\sigma = \frac{1}{m(\sigma)}\int g(y)d\gamma(y)$. On en tire l'équation relative à la maille $k = n(j-1)+1, j = 2,\dots,p-1,p+1,\dots,2p-1$:

$$\lambda_m\left(2\frac{h_x}{h_y}+3\frac{h_y}{h_x}\right)u_k - \lambda_m\frac{h_x}{h_y}(u_{k-n}+u_{k+n}) - \lambda_m\frac{h_y}{h_x}u_{k+1} = h_x h_y f_k + 2\frac{h_y}{h_x}\lambda_m g_j,$$

avec $g_j = g_{\sigma_j}$ et $m = 1$ si $j \leqslant p-1, m = 2$ si $j \geqslant p+1$.

— Mailles du bord $\Gamma_1 \cup \Gamma_3$ — Pour $j = 1$, où $j = 2p, i = 2\dots n-1$. On tient compte ici de la condition de Fourier sur la maille qui appartient au bord, pour laquelle l'expression du flux est :

$$F_{K,\sigma} = \frac{\alpha\lambda_m m(\sigma)}{\lambda_m + \alpha d_{K,\sigma}}(u_K - u_{ext}).$$

Pour une arête σ horizontale, on note : $C_{F,\sigma} = \frac{\alpha m(\sigma)}{\lambda_m+\alpha d_{K,\sigma}}$. Comme le maillage est uniforme, $C_{F,\sigma}$ est égal à

$$C_F = \frac{2\alpha h_x}{2\lambda_m + \alpha h_y},$$

et ne dépend donc pas de σ. Les équations s'écrivent donc :

$$\lambda_1\left(2\frac{h_y}{h_x}+\frac{h_x}{h_y}+C_F\right)u_k - \lambda_1\frac{h_x}{h_y}u_{k+n} - \lambda_1\frac{h_y}{h_x}(u_{k+1}+u_{k-1}) = h_x h_y f_k + \lambda_1 C_F u_{ext},$$
$$k = 2,\dots,n-1,$$

$$\lambda_2\left(\frac{2h_y}{h_x}+\frac{h_x}{h_y}+C_F\right)u_k-\lambda_2\frac{h_x}{h_y}u_{k-n}-\lambda_2\frac{h_y}{h_x}(u_{k+1}+u_{k-1})=h_xh_yf_k+\lambda_2C_Fu_{ext},$$
$$k=n(2p-1)+2,\dots,2np-1,$$

— Mailles des coins extérieurs — Il suffit de synthétiser les calculs déjà faits :

— coin sud-est : $i=1,j=1,k=1$; un bord Dirichlet, un bord Fourier :

$$\lambda_1\left(\frac{3h_y}{h_x}+\frac{h_x}{h_y}+C_F\right)u_1-\lambda_1\frac{h_y}{h_x}u_2-\lambda_1\frac{h_x}{h_y}u_{n+1}=h_xh_yf_1+\lambda_1C_Fu_{ext}+\frac{2h_y}{h_x}\lambda_1g_1$$

— coin sud-ouest : $i=1n,j=1,k=n$; un bord Fourier, un bord Neumann :

$$\lambda_1\left(\frac{h_y}{h_x}+\frac{h_x}{h_y}+C_F\right)u_1-\lambda_1\frac{h_y}{h_x}u_{n-1}-\lambda_1\frac{h_x}{h_y}u_{2n}=h_xh_yf_n+\lambda_1C_Fu_{ext}$$

— coin nord-ouest : $i=2n,j=2p,k=2np$. On a encore un bord Fourier, un bord Neumann, et l'équation s'écrit :

$$\lambda_2\left(\frac{h_y}{h_x}+\frac{h_x}{h_y}+C_F\right)u_{2np}-\lambda_2\frac{h_y}{h_x}u_{2np-1}-\lambda_2\frac{h_x}{h_y}u_{2n(p-1)}=h_xh_yf_{2np}+\lambda_2C_Fu_{ext}$$

— coin nord-est : $i=1,j=2p\quad k=n(2p-1)+1$ un bord Dirichlet, un bord Fourier :

$$\lambda_2\left(\frac{3h_y}{h_x}+\frac{h_x}{h_y}+C_F\right)u_k-\lambda_2\frac{h_y}{h_x}u_{k+1}-\lambda_2\frac{h_x}{h_y}(u_{k-n}=h_xh_yf_k+\lambda_2C_Fu_{ext},+\frac{2h_y}{h_x}\lambda_2g_k.$$

— Interface — L'expression du flux sur une arête de l'interface est donnée par (1.76). On pose, pour chaque arête σ de l'interface,

$$s_{I,\sigma}=\frac{m(\sigma)}{\lambda_1d_{L,\sigma}+\lambda_2d_{K,\sigma}}.$$

Notons que dans le cas du maillage uniforme considéré, ce coefficient est égal à :

$$s_I=\frac{2h_x}{(\lambda_1+\lambda_2)h_y},$$

et qu'il est indépendant de l'arête σ. Tenant compte de ce flux, on obtient, pour $k=n(p-1)+i,i=2,\dots,N-1$

$$\lambda_1\left(\frac{2h_y}{h_x}+\frac{h_x}{h_y}+\lambda_2S_I\right)u_k-\lambda_1\frac{h_y}{h_x}u_{k+1}-\lambda_1\frac{h_y}{h_x}u_{k-1}-\lambda_1\frac{h_x}{h_y}u_{k-n}-\lambda_1S_Iu_{k+n}=h_xh_yf_k+\lambda_1S_I\frac{h_y}{2}\theta_i,$$

avec

$$\theta_i=\int_{\sigma_i}\theta(x)d\gamma(x).$$

Et de même, pour $k=np+i,i=2,\dots,N-1$,

$$\lambda_1\left(\frac{2h_y}{h_x}+\frac{h_x}{h_y}+\lambda_1S_I\right)u_k-\lambda_2\frac{h_y}{h_x}u_{k+1}-\lambda_2\frac{h_y}{h_x}u_{k-1}-\lambda_2\frac{h_x}{h_y}u_{k+n})-\lambda_2S_Iu_{k-n}=h_xh_yf_k+\lambda_2S_I\frac{h_y}{2}\theta_i.$$

Il ne reste plus qu'à traiter les coins des interfaces.

— $i=1,j=p\quad k=n(p-1)+1$. Dirichlet sous l'interface

$$\lambda_1\left(\frac{3h_y}{h_x}+\frac{h_x}{h_y}+\lambda_2s_I\right)u_k-\lambda_1\frac{h_y}{h_x}u_{k+1}-\lambda_1\frac{h_x}{h_y}u_{k+n}-\lambda_1s_Iu_{k+n}=h_xh_yf_k+\lambda_1S_I\frac{h_y}{2}\theta_i+\frac{2h_y}{h_x}\lambda_1g_j$$

— $i=1,j=p+1\quad k=np+1$, Dirichlet, dessus de l'interface

$$\lambda_2\left(\frac{3h_y}{h_x}+\frac{h_x}{h_y}+\lambda_1s_I\right)u_k-\lambda_2\frac{h_y}{h_x}u_{k+1}-\lambda_2\frac{h_x}{h_y}u_{k+n}-\lambda_2s_Iu_{k-n}=h_xh_yf_k+\lambda_2S_I\frac{h_y}{2}\theta_i+\frac{2h_y}{h_x}\lambda_2g_j$$

— $i=n,j=p,k=n(p-1)+n$. Neumann, sous l'interface.

$$\lambda_1\left(\frac{h_y}{h_x}+\frac{h_x}{h_y}+\lambda_2s_I\right)u_k-\lambda_1\frac{h_y}{h_x}u_{k-1}-\lambda_1\frac{h_x}{h_y}u_{k-n}-\lambda_1s_Iu_{k+n}=h_xh_yf_k+\lambda_1S_I\frac{h_y}{2}\theta_i$$

— $i = n, j = p + 1, k = np + n$, Neuman, dessus de l'interface

$$\lambda_2 \left(\frac{h_y}{h_x} + \frac{h_x}{h_y} + \lambda_1 s_I \right) u_k - \lambda_2 \frac{h_y}{h_x} u_{k-1} - \lambda_2 \frac{h_x}{h_y} u_{k+n} - \lambda_2 s_I u_{k-n} = h_x h_y f_k + \lambda_2 S_I \frac{h_y}{2} \theta_i.$$

On a ainsi obtenu $2np$ équations à $2np$ inconnues. Notons que chaque équation fait intervenir au plus 5 inconnues.

2

Discrétisation en temps des problèmes paraboliques

Dans l'introduction, on a vu comme exemple type de problème parabolique l'équation de la chaleur instationnaire :

$$\partial_t u - \Delta u = f,$$

qui fait intervenir la dérivée en temps d'ordre 1 de la fonction u, qu'on note $\partial_t u$, ainsi que le laplacien de u, Δu, qui est un opérateur différentiel d'ordre 2 en espace défini (en deux dimensions d'espace) par $\Delta u = \partial^2_{xx} u + \partial^2_{yy} u$, où $\partial^2_{xx} u$ désigne la dérivée partielle d'ordre 2 de u par rapport à x. Pour que ce problème soit bien posé, il faut spécifier des conditions aux limites sur la frontière de Ω, et une condition initiale en $t = 0$.

Dans la suite de ce chapitre, on va considérer ce problème en une dimension d'espace seulement, car c'est la discrétisation en temps qui nous importe ici. Au temps $t = 0$, on se donne une condition initiale u_0, et on considère des conditions aux limites de type Dirichlet homogène. Le problème unidimensionnel s'écrit :

$$\begin{cases} \partial_t u - \partial^2_{xx} u = 0, & \forall x \in]0\,;1[, \quad \forall t \in]0, T[, \\ u(x,0) = u_0(x), & \forall x \in]0\,;1[, \\ u(0,t) = u(1,t) = 0, & \forall t \in]0, T[, \end{cases} \tag{2.1}$$

où $u(x,t)$ représente la température au point x et au temps t.

2.1 Propriétés du problème continu et discrétisation espace-temps

On admettra le théorème d'existence et unicité suivant :

Théorème 2.1 — Résultat d'existence et unicité. Si $u_0 \in \mathcal{C}(]0\,;1[, \mathbb{R})$ alors il existe une unique fonction $u \in \mathcal{C}^2(]0\,;1[\times]0, T[, \mathbb{R}) \cap \mathcal{C}([0\,;1] \times [0\,;T], \mathbb{R})$ qui vérifie (2.1).

On a même $u \in \mathcal{C}^\infty(]0\,;1[\times]0, T[, \mathbb{R})$: c'est l'*effet régularisant* de l'équation de la chaleur.

Proposition 2.1 — Principe du maximum. Sous les hypothèses du théorème 2.1, soit u la solution du problème (2.1) ;

1. si $u_0(x) \geqslant 0$ pour tout $x \in [0\,;1]$, alors $u(x,t) \geqslant 0$, pour tout $t \geqslant 0$, pour tout $x \in]0\,;1[$;
2. $\|u\|_{L^\infty([0\,;1[\times]0,T[)} \leqslant \|u_0\|_{L^\infty(]0\,;1[)}$.

Ces dernières propriétés peuvent être importantes par rapport au modèle physique ; supposons par exemple que u représente une fraction massique. Par définition de la fraction massique, celle-ci est toujours comprise entre 0 et 1. La proposition 2.1 nous dit que la quantité u donnée par le modèle mathématique supposé représenter la fraction massique d'une espèce qui diffuse dans un milieu, par exemple, est aussi comprise entre 0 et 1, dès que la fraction massique initiale u_0 est dans l'intervalle $[0\,;1]$ ce qui est plutôt une bonne nouvelle : le modèle mathématique respecte les bornes de la physique. Mais on ne peut pas en général calculer la solution de (2.1) de manière analytique. On a recours à la discrétisation en temps et en espace pour se ramener à un système d'équations de dimension finie. Il est souhaitable pour la validité du calcul que la solution approchée obtenue par la résolution de ce système, qui est supposée approcher une fraction massique soit aussi comprise

à tout instant entre 0 et 1. On dit souvent d'une méthode de discrétisation (ou d'un schéma de discrétisation) qu'elle (ou il) est *robuste* ou *stable* s'il préserve les bornes imposées par la physique (0 et 1 dans le cas de la fraction massique évoquée ci-dessus).

Pour calculer une solution approchée, on choisit pour l'instant de discrétiser par différences finies en temps et en espace. La discrétisation consiste donc à se donner un ensemble de points t_n, $n = 1, \dots, M$ de l'intervalle $]0, T[$, et un ensemble de points x_i, $i = 1, \dots, N$. Pour simplifier, on considère un pas constant en temps et en espace. Soit $h = 1/(N+1)$ le pas de discrétisation en espace, et $k = T/M$, le pas de discrétisation en temps. On pose alors $t_n = nk$ pour $n = 0, \dots, M$ et $x_i = ih$ pour $i = 0, \dots, N+1$. Comme on a des conditions de Dirichlet homogènes, les valeurs de u en $x_0 = 0$ et $x_{N+1} = 1$ ne sont pas des inconnues. On cherche à calculer une solution approchée du problème (2.1) en calculant des valeurs $u_i^{(n)}$, $i = 1, \dots, N$ et $n = 1, \dots, M$, censées être des approximations des valeurs $u(x_i, t_n)$, $i = 1, \dots, N$, et $n = 1, \dots, M$.

2.2 Discrétisation par un schéma d'Euler explicite en temps

L'approximation en temps par la méthode d'Euler explicite consiste à écrire la première équation de (2.1) en chaque point x_i et temps t_n, à approcher la dérivée en temps $\partial_t u(x_i, t_n)$ par le quotient différentiel :

$$\frac{u(x_i, t_{n+1}) - u(x_i, t_n)}{k},$$

et la dérivée en espace $-\partial^2_{xx} u(x_i, t_n)$ par le quotient différentiel :

$$\frac{1}{h^2}(2u(x_i, t_n) - u(x_{i-1}, t_n) - u(x_{i+1}, t_n)).$$

Remarque 2.2 — Autres discrétisations en espace. On a choisi ici une discrétisation en espace de type différences finies en espace mais on aurait aussi bien pu prendre un schéma de volumes finis ou d'éléments finis.

On obtient le schéma suivant :

$$\frac{u_i^{(n+1)} - u_i^{(n)}}{k} + \frac{1}{h^2}(2u_i^{(n)} - u_{i-1}^{(n)} - u_{i+1}^{(n)}) = 0, \qquad i = 1, \dots N, \quad n = 1, \dots, M, \tag{2.2a}$$

$$u_i^0 = u_0(x_i), \qquad i = 1, \dots, N, \tag{2.2b}$$

$$u_0^{(n)} = u_{N+1}^{(n)} = 0, \qquad \forall n = 1, \dots, M. \tag{2.2c}$$

Le schéma est dit explicite car la formule (2.2a) donne $u_i^{(n+1)}$ de manière explicite en fonction des $(u_i^{(n)})_{i=1,\dots,N}$. De fait, on peut réécrire la formule (2.2a) de la manière suivante :

$$u_i^{(n+1)} = u_i^{(n)} - \lambda(2u_i^{(n)} - u_{i-1}^{(n)} - u_{i+1}^{(n)}) \quad \text{avec} \quad \lambda = \frac{k}{h^2}.$$

2.2.1 Consistance

Définition 2.3 — Erreur de consistance, Euler explicite. Avec les notations du paragraphe 2.1, on pose $\bar{u}_i^{(n)} = u(x_i, t_n)$ la valeur exacte de la solution en x_i et t_n du problème (2.1) ; l'erreur de consistance du schéma (2.2) en (x_i, t_n), notée $R_i^{(n)}$, est définie comme la somme des erreurs de consistance en temps et en espace :

$$R_i^{(n)} = \widetilde{R}_i^{(n)} + \widehat{R}_i^{(n)}, \text{ avec}$$

$$\widetilde{R}_i^{(n)} = \frac{\bar{u}_i^{(n+1)} - \bar{u}_i^{(n)}}{k} - \partial_t u(x_i, t_n) \quad \text{et} \quad \widehat{R}_i^{(n)} = -\partial^2_{xx} u(x_i, t_n) - \frac{2\bar{u}_i^{(n)} - \bar{u}_{i-1}^{(n)} - \bar{u}_{i+1}^{(n)}}{h^2}. \tag{2.3}$$

Proposition 2.4 — Consistance, Euler explicite. Le schéma (2.2) est consistant d'ordre 1 en temps et d'ordre 2 en espace, c'est-à-dire qu'il existe $C \in \mathbb{R}_+$ ne dépendant que de u tel que l'erreur de consistance $R^{(n)} = (R_1^{(n)}, \dots, R_N^{(n)})^t$ définie par (2.3) vérifie :

$$\|R^{(n)}\| = \max_{i=1,\dots,N} |R_i^{(n)}| \leqslant C(k + h^2). \tag{2.4}$$

Démonstration. On a vu lors de l'étude des problèmes elliptiques que l'erreur de consistance en espace $\widetilde{R}_i^{(n)}$ est d'ordre 2 [formule (1.31)]. Un développement de Taylor en temps donne facilement que $\widetilde{R}_i^{(n)}$ est d'ordre 1 en temps. •

2.2.2 Stabilité

Par la proposition 2.1, la solution exacte vérifie

$$\|u\|_{L^\infty(]0\,;1[\times]0,T[)} \leqslant \|u_0\|_{L^\infty(]0\,;1[)}$$

Nous allons voir que si l'on choisit correctement les pas de temps et d'espace, on peut avoir l'équivalent discret sur la solution approchée.

Définition 2.5 — Stabilité d'un schéma de discrétisation temps-espace. On dit qu'un schéma est L^∞-stable si la solution approchée qu'il donne est bornée en norme infinie, indépendamment du pas du maillage.

Proposition 2.6 — Stabilité, Euler explicite. Si la condition de stabilité

$$\lambda = \frac{k}{h^2} \leqslant \frac{1}{2}, \tag{2.5}$$

est vérifiée, alors le schéma (2.2) est L^∞–stable au sens où :

$$\max_{\substack{i=1,\dots,N\\ n=1,\dots,M}} |u_i^{(n)}| \leqslant \|u_0\|_\infty.$$

Démonstration. On peut écrire le schéma sous la forme

$$u_i^{(n+1)} = u_i^{(n)} - \lambda(2u_i^{(n)} - u_{i-1}^{(n)} - u_{i+1}^{(n)})$$

soit encore :

$$u_i^{(n+1)} = (1-2\lambda)u_i^{(n)} + \lambda u_{i-1}^{(n)} + \lambda u_{i+1}^{(n)}.$$

Si $0 \leqslant \lambda \leqslant 1/2$, on a $\lambda \geqslant 0$ et $1 - 2\lambda \geqslant 0$, et la quantité $u_i^{(n+1)}$ est donc combinaison convexe de $u_i^{(n)}$, $u_{i-1}^{(n)}$ et $u_{i+1}^{(n)}$. Soit $\mu^{(n)} = \max_{i=1,\dots,N} u_i^{(n)}$, on a alors :

$$u_i^{(n+1)} \leqslant (1-2\lambda)\mu^{(n)} + \lambda\mu^{(n)} + \lambda\mu^{(n)}, \quad \forall i = 1, \dots, N,$$

et donc $u_i^{(n+1)} \leqslant \mu^{(n)}$. On a donc, en passant au maximum sur i que $\mu^{(n+1)} \leqslant \mu^{(n)}$. On montre de la même manière que

$$\min_{i=1,\dots,N} u_i^{(n+1)} \geqslant \min_{i=1,\dots,N} u_i^{(n)}$$

On en déduit $\max_{i=1,\dots,N}(u_i^{(n+1)}) \leqslant \max u_i^0$ et $\min_{i=1,\dots,N}(u_i^{(n+1)}) \geqslant \min u_i^0$ d'où le résultat. •

Notons que la condition de stabilité (2.5) est très contraignante sur le pas de temps : en effet, pour obtenir une bonne précision en espace, on veut pouvoir choisir un pas d'espace petit…, mais du coup on doit prendre un pas de temps "tout petit", puisque le pas de temps doit être de l'ordre du carré du pas d'espace. Ceci engendre trop de temps calcul, de sorte que le schéma d'Euler explicite est rarement utilisé pour les problèmes paraboliques dans les codes industriels.

2.2.3 Convergence

Définition 2.7 — Erreur de discrétisation. Soit u la solution excate du problème (2.1) et $(u_i^{(n)})_{i=1,\dots N}^{n=1,\dots,M}$ la solution du schéma Euler explicite (2.2). Pour $i = 1, \dots N$ et $n = 1, \dots, M$, on

note $\bar{u}_i^{(n)} = u(x_i, t_n)$ et on appelle *erreur de discrétisation au point* (x_i, t_n) la quantité

$$e_i^n = \bar{u}_i^{(n)} - u_i^{(n)}.$$

L'erreur de discrétisation associée au schéma (2.2) est alors

$$\|e^{(n)}\|_\infty = \max_{i=1,\dots,N} |e_i^{(n)}|.$$

Théorème 2.2 — Convergence du schéma d'Euler explicite. Sous les hypothèses du théorème 2.1 et sous la condition de stabilité (2.5), il existe $C \in \mathbb{R}_+$ ne dépendant que de u tel que

$$\|e^{(n+1)}\|_\infty \leqslant \|e^{(0)}\|_\infty + TC(k+h^2) \qquad \forall i = 1,\dots,N, \quad n = 0,\dots,M-1.$$

Ainsi, si $\|e_i^{(0)}\|_\infty = 0$ alors $\max_{i=1,\dots N} \|e_i^{(n)}\|$ tend vers 0 lorsque k et h tendent vers 0 pour tout $n = 1,\dots M$. Le schéma (2.2) est donc convergent.

Démonstration. On a donc, par définition (2.3) de l'erreur de consistance,

$$\frac{\bar{u}_i^{(n+1)} - \bar{u}_i^{(n)}}{k} + \frac{2\bar{u}_i^{(n)} - \bar{u}_{i-1}^{(n)} - \bar{u}_{i+1}^{(n)}}{h^2} = R_i^{(n)}.$$

D'autre part, le schéma numérique s'écrit :

$$\frac{u_i^{(n+1)} - u_i^{(n)}}{k} + \frac{2u_i^{(n)} - u_{i-1}^{(n)} - u_{i+1}^{(n)}}{h^2} = 0.$$

La différence des deux équations précédentes donne :

$$\frac{e_i^{(n+1)} - e_i^{(n)}}{k} + \frac{2e_i^{(n)} - e_{i+1}^{(n)} - e_{i-1}^{(n)}}{h^2} = R_i^{(n)},$$

soit encore $e_i^{(n+1)} = (1-2\lambda)e_i^{(n)} + \lambda e_{i-1}^{(n)} + \lambda e_{i+1}^{(n)} + kR_i^{(n)}$; or $(1-2\lambda)e_i^{(n)} + \lambda e_{i-1}^{(n)} + \lambda e_{i+1}^{(n)} \leqslant \|e^{(n)}\|_\infty$ car $\lambda \leqslant 1/2$ et donc, grâce à la consistance du schéma ('inégalité (2.4)) on a :

$$|e_i^{(n+1)}| \leqslant \|e^{(n)}\|_\infty + kC(k+h^2).$$

On a donc par récurrence $\|e^{(n+1)}\|_\infty \leqslant \|e^{(0)}\|_\infty + MkC(k+h^2)$ ce qui démontre le théorème. •

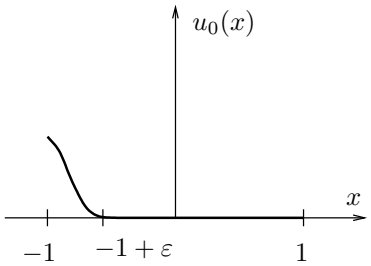
2.2.4 Exemple de non convergence

Montrons que si la condition de stabilité (2.5), *i.e.* $\lambda \leqslant 1/2$, n'est pas respectée, on peut construire une condition initiale pour laquelle le schéma n'est pas stable.

Soit $u_0 \in \mathcal{C}([-1\,;1], \mathbb{R})$ qui vérifie :

$$\begin{cases} u_0(x) \geqslant 0 \\ u_0(x) \neq 0 & \text{si } x \in]-1\,;-1+\varepsilon[\\ u_0(x) = 0 & \text{si } x > -1+\varepsilon \end{cases}$$

Un exemple de fonction qui vérifie ces conditions est représenté sur la figure ci-contre.



On considère le problème :

$$\begin{cases} \partial_t u - \partial_{xx}^2 u = 0 & \forall x \in]-1\,;1[\quad \forall t > 0 \\ u(x,0) = u_0(x) & \forall x \in]-1\,;1[\\ u(1,t) = u(-1,t) = 0 & \forall t > 0 \end{cases} \tag{2.6}$$

On peut montrer que la solution exacte u de (2.6) vérifie $u(x,t) > 0, \forall x \in]-1\,;1[, \forall t > 0$. En particulier, pour un temps $T > 0$ donné, on a $u(0,T) > 0$. Soit $M \in \mathbb{N}$ et $k = T/M$. Soit $u_i^{(n)}$ la solution approchée par (2.2), censée approcher $u(x_i, t_n)$ $(i \in \{-N,\dots,N\}, n \in N)$. On va montrer que $u_0^M = 0$ pour des pas de temps k et h ne respectant pas la condition (2.5) ; ceci montre que le schéma ne peut pas converger. Calculons u_0^M :

$$u_0^M = (1-2\lambda)u_0^{M-1} + \lambda u_{-1}^{M-1} + \lambda u_1^{M-1}$$

Donc u_0^M dépend de :

$$\begin{array}{ll} u^{(M-1)} & \text{sur } [-h\,;h] \\ u^{(M-2)} & \text{sur } [-2h\,;2h] \\ \vdots & \\ u^{(0)} & \text{sur } [-Mh\,;Mh] = [-Th/k\,;Th/k] \end{array}$$

Par exemple, si on prend $h/k = 1/(2T)$ on obtient $[-Th/k\,;Th/k] = [-1/2\,;1/2]$, et donc, si $\varepsilon < 1/2$, on a $u_0^M = 0$. On peut donc remarquer que si $h/k = 1/(2T)$, même si $h \to 0$ et $k \to 0$, $u_0^M = 0$ ne tend pas vers $u(0,T) > 0$ lorsque $m \to +\infty$. Le schéma ne converge pas ; notons que ceci n'est pas en contradiction avec le résultat de convergence 2.2, puisqu'ici, la condition $k/h^2 \leqslant 1/2$ n'est pas satisfaite.

2.2.5 Stabilité au sens des erreurs d'arrondi

On considère le schéma d'Euler explicite pour l'équation (2.1). On rappelle que $\bar{u}_i^{(n)}$ désigne la solution exacte du problème continu (2.1) en (x_i, t_n). Soit $(u_i^{(n)})_{i=1,\dots N}^{n=1,\dots,M}$ la solution du schéma Euler explicite (2.2). On désigne enfin par $(v_i^{(n)})_{i=1,\dots N}^{n=1,\dots,M}$ la solution effectivement calculée par l'ordinateur, qui est différente de $(u_i^{(n)})_{i=1,\dots N}^{n=1,\dots,M}$ en raison des erreurs d'arrondi de l'ordinateur utilisé pour les calculs. On peut écrire :

$$\bar{u}_i^{(n)} - v_i^{(n)} = \bar{u}_i^{(n)} - u_i^{(n)} + u_i^{(n)} - v_i^{(n)}.$$

On a vu (théorème 2.2) que l'erreur de discrétisation $e_i^{(n)} = \bar{u}_i^{(n)} - u_i^{(n)}$ tend vers 0 lorsque h et k tendent vers 0, sous condition de stabilité (2.5), c'est-à-dire $\lambda \leqslant 1/2$. Pour contrôler l'erreur entre la solution $(u_i^{(n)})_{i=1,\dots N}^{n=1,\dots,M}$ du schéma (2.2) et la solution numérique obtenue $(v_i^{(n)})_{i=1,\dots N}^{n=1,\dots,M}$, on cherche à estimer l'amplification de l'erreur commise sur la donnée initiale en raison des erreurs arrondis. Le schéma (2.2) peut s'écrire sous la forme $U^{(n+1)} = A_N U^{(n)}$, où A_N est une matrice réelle carrée d'ordre N, souvent appelée matrice d'amplification, qui est définie par .

$$A_N = \begin{bmatrix} 1-2\lambda & \lambda & 0 & \dots & 0 \\ \lambda & 1-2\lambda & \lambda & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \lambda & 1-2\lambda & \lambda \\ 0 & \dots & 0 & \lambda & 1-2\lambda \end{bmatrix} \tag{2.7}$$

Définition 2.8 — Stabilité au sens des erreurs d'arrondi. Supposons que l'on commette une erreur ε^0 sur la condition initiale. La nouvelle condition initiale $\widetilde{u}^0$, s'écrit donc $\widetilde{u}^0 = u^0 + \varepsilon^0$. À cette nouvelle condition initiale correspond une nouvelle solution calculée $\widetilde{u}^{(n)} = u^{(n)} + \varepsilon^{(n)}$. On dit que le schéma est stable au sens des erreurs d'arrondi s'il existe $C > 0$ indépendant de n tel que $\varepsilon^{(n)} \leqslant C\varepsilon^{(0)}$.

On peut trouver une condition suffisante pour que le schéma (2.2) soit stable au sens des erreurs d'arrondi. Pour cela, on aura besoin du petit lemme d'algèbre linéaire suivant.

Lemme 2.9 — Valeurs propres du laplacien discret unidimensionnel. L'ensemble $\mathcal{VP}(K_N)$ des valeurs propres de la matrice K_N définie par

$$K_N = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix} \tag{2.8}$$

est donné par :

$$\mathcal{VP}(K_N) = \{4\sin^2\frac{j\pi}{2(N+1)}, j = 1,\dots,N\}$$

Démonstration. Les valeurs propres de K_N peuvent se calculer à partir des valeurs propres de l'opérateur continu ; commençons donc par chercher u solution du problème au valeur propre continu :

$$\begin{cases} -u'' + \alpha u = 0, \\ u(0) = u(1) = 0. \end{cases}$$

Cherchons $u(x)$ sous la forme :

$$u(x) = a\cos\sqrt{\alpha}x + b\sin\sqrt{\alpha}x$$

Comme $u(0) = 0$, on a $a = 0$. De même, $u(1) = B\sin\sqrt{\alpha} = 0$ et donc $\sqrt{\alpha} = k\pi$. Les valeurs propres et vecteurs propres associés de l'opérateur continu sont donc $\left(k^2\pi^2, \sin k\pi x\right)$ $k \in \mathbb{N}^*$. Pour $k = 1,\dots,N$, soit $v^{(k)} \in \mathbb{R}^N$ tel que $v_i^{(k)} = \sin k\pi ih$. Calculons $K_N v^{(k)}$:

$$(K_N v^{(k)})_i = -v_{i-1}^{(k)} + 2v_i^{(k)} - v_{i+1}^{(k)}$$

et donc

$$(K_N v^{(k)})_i = -\sin k\pi(i-1)h + 2\sin k\pi ih - \sin k\pi(i+1)h.$$

En développant, on obtient

$$(K_N v^{(k)})_i = -\sin k\pi ih\cos(-k\pi h) - \cos k\pi ih\sin(-k\pi h) + 2\sin k\pi ih - \sin k\pi ih\cos k\pi h - \cos k\pi ih\sin k\pi h.$$

Après simplification, il vient $(K_N v^{(k)})_i = -2\sin k\pi ih(-1+\cos k\pi h)$. Or, $\cos k\pi h = 1 - 2\sin^2\frac{k\pi h}{2}$. On a donc $(K_N v^{(k)})_i = 2\sin k\pi ih \times (2\sin^2\frac{k\pi h}{2}) = 4\sin^2\frac{k\pi h}{2}(v^{(k)})_i$, $\forall k = 1\dots N$. On a $h = \frac{1}{N+1}$, et donc les valeurs propres de K_N s'écrivent $\mu_k = 4\sin^2\frac{k\pi}{2(N+1)}$, $\quad k = 1,\dots,N$. •

On peut maintenant démontrer le résultat suivant :

Proposition 2.10 — Stabilité au sens des erreurs d'arrondi, Euler explicite. On suppose que $\lambda = k/h^2 < 1/2$. Alors le schéma (2.2) est stable au sens des erreurs d'arrondi.

Démonstration. Soit donc une condition initiale perturbée $\widetilde{u}^{(0)} = u^{(0)} + \varepsilon^{(0)}$ à laquelle on associe une nouvelle solution calculée $\widetilde{u}^{(n)} = u^{(n)} + \varepsilon^{(n)}$. On a $\varepsilon^{(n)} = A_N^n\varepsilon^{(0)}$, où A_N est la matrice d'amplification définie par (2.7). Comme A_N est symétrique, A_N est diagonalisable dans $\mathbb{R}$. Soient $\mu_1,\dots,\mu_N$ les valeurs propres de A_N, et $e_1,\dots,e_N$ les vecteurs propres associés, c'est-à-dire tels que $A_N e_i = \mu_i e_i, \forall i = 1,\dots N$. Décomposons la perturbation $\varepsilon^{(0)}$ sur la base des vecteurs propres :

$$\varepsilon^{(0)} = \sum_{i=1}^N a_i e_i, \text{ et donc } A_N^n\varepsilon^{(0)} = \sum_{i=1}^N a_i\mu_i^n e_i = \varepsilon^{(n)}.$$

Si on prend par exemple : $\varepsilon^{(0)} = a_i e_i$, on obtient $\varepsilon^{(n)} = a_i\mu_i^n e_i$. Il y a diminution de l'erreur d'arrondi sur $\varepsilon^{(0)}$ si

$$\sup_{i=1\dots N} |\mu_i| < 1.$$

c'est-à-dire si $\rho(A_N) < 1$ où $\rho(A_N)$ désigne le rayon spectral de A. Calculons $\rho(A_N)$. On écrit $A_N = I - \lambda K_N$ où K_N est la matrice symétrique définie positive, définie par (2.8) : Soit $\mathcal{VP}(A_N)$ l'ensemble de valeurs propres de A. Alors $\mathcal{VP}(A_N) = \{1 - \lambda\mu, \mu \in \mathcal{VP}(K_N))\}$. Or d'après le lemme 2.9, $\mathcal{VP}(K_N) = \{4\sin^2\frac{j\pi}{2(N+1)}, j = 1,\dots,N\}$. Pour que $\varepsilon^{(n)} < \varepsilon^0$, il faut donc que :

$$\sup_{j=1,\dots,N} |1 - 4\lambda\sin^2\frac{j\pi}{2(N+1)}| < 1 \quad \Leftrightarrow \quad \lambda\sin^2\frac{j\pi}{2(N+1)} < \frac{1}{2}$$

Une condition suffisante pour avoir une diminution de l'erreur est donc que $\lambda < 1/2$. •

2.2.6 Stabilité au sens de Von Neumann

L'analyse de stabilité au sens de Von Neumann [1] consiste à étudier l'impact du schéma sur un mode de Fourier isolé. Pour que le mode de Fourier en question soit solution du problème continu, on remplace les conditions de Dirichlet homogènes du problème (2.1) par des conditions périodiques, et pour alléger les notations, on considère l'intervalle $]0\,;2\pi[$ comme intervalle d'étude en espace plutôt que l'intervalle $]0\,;1[$. .

1. John von Neumann (1903-1957), mathématicien et physicien américain d'origine hongroise, a apporté d'importantes contributions tant en mécanique quantique, qu'en analyse fonctionnelle, en théorie des ensembles, en informatique, en sciences économiques ainsi que dans beaucoup d'autres domaines des mathématiques et de la physique. Il a de plus participé aux programmes militaires américains.

Remarque 2.11 — Attention, complexes.... Ici et dans tout le paragraphe 2.2.6, on travaille avec des nombres complexes car on utilise l'analyse de Fourier. On va donc utiliser des fonctions de $\mathbb{R}$ à valeurs dans $\mathbb{C}$; en particulier, la notation i ne peut plus être utilisée comme indice des inconnues de discrétisation, car i désigne dans toute cette section le complexe imaginaire pur tel que $i^2 = -1$; on notera donc j l'indice de discrétisation en espace.

Problème continu avec conditions aux limites périodiques On considère le problème avec conditions aux limites périodiques et donnée initiale $u_0 \in \mathcal{C}([0\,;2\pi],\mathbb{C})$ (attention : u_0 est donc à valeurs dans $\mathbb{C}$) :

$$\partial_t u - \partial^2_{xx} u = 0, \qquad \forall t \in]0,T[, \qquad \forall x \in]0\,;2\pi[, \tag{2.9a}$$
$$u(0,t) = u(2\pi,t), \qquad \forall t \in]0,T[, \tag{2.9b}$$
$$u(x,0) = u_0(x). \tag{2.9c}$$

Le problème (2.9) est bien posé, au sens où, commeon suppose que $u_0 \in \mathcal{C}([0\,;2\pi],\mathbb{C})$, il existe une unique $u \in \mathcal{C}^2(]0\,;2\pi[\times]0,T[,\mathbb{C})$ solution de (2.9). De plus $u_0 \in L^2_{\mathbb{C}}(]0\,;2\pi[)$. On rappelle que l'espace $L^2_{\mathbb{C}}$ des fonctions mesurables de $\mathbb{R}$ à valeurs dans $\mathbb{C}$ est un espace de Hilbert ; une base hilbertienne de $L^2(]0\,;2\pi[)$ [2] est la famille $\{\psi_p,\ p \in \mathbb{Z}\}$, où ψ_p est le p-ième mode de Fourier défini par

$$\psi_p : \quad \begin{aligned} &\mathbb{R} \to \mathbb{C}, \\ &x \mapsto \mathrm{e}^{ipx} \end{aligned}$$

On décompose donc la condition initiale dans cette base hilbertienne : $u_0 = \sum_{p\in\mathbb{Z}} c_p(0)\psi_p$ (au sens de la convergence dans L^2). Dans un premier temps, calculons formellement les solutions de (2.9) sous la forme d'un développement dans la base hilbertienne : $u(x,t) = \sum_{p\in\mathbb{Z}} c_p(t)\psi_p(x)$. En supposant qu'on ait le droit de dériver terme à terme, on a alors :

$$\partial_t u(x,t) = \sum_{p\in\mathbb{Z}} c_p'(t)\,\mathrm{e}^{ipx} \text{ et } \partial^2_{xx} u(x,t) = \sum_{p\in\mathbb{Z}} c_p(t)p^2\,\mathrm{e}^{ipx}\,.$$

En remplaçant dans l'équation (2.9a), on obtient : $c_p'(t) = p^2 c_p(t)$, c'est-à-dire, en tenant compte de la condition initiale, $c_p(t) = c_p(0)\,\mathrm{e}^{p^2 t}$. On a donc finalement :

$$u(x,t) = \sum_{p\in\mathbb{Z}} c_p(0)\,\mathrm{e}^{p^2 t}\,\mathrm{e}^{ipx}\,. \tag{2.10}$$

Justifions maintenant ce calcul formel. On a : $\sum_{p\in\mathbb{Z}} |c_p(0)|^2 = \|u^0\|^2_{L^2} < +\infty$. De plus, en dérivant (2.10) terme à terme, on obtient : $\partial_t u - \partial^2_{xx} u = 0$. La condition de périodicité est bien vérifiée par u donnée par (2.10). Enfin on a bien $u(x,t) \to u_0(x)$ lorsque $t \to 0$, donc la condition initiale est vérifiée. On peut remarquer qu'il y a *amortissement* des coefficients de Fourier $c_p(0)$ lorsque t croît, c'est-à-dire qu'on a $c_p(t) < c_p(0)$, $\forall t > 0$.

Discrétisation du problème Si on utilise le schéma d'Euler explicite, pour la discrétisation de (2.9) avec comme condition initiale le p-ième mode de Fourier : $u^0(x) = \psi_p(x) = \mathrm{e}^{ipx}$ pour $p \in \mathbb{Z}$ fixé, on obtient :

$$u_j^{(n+1)} = (1-2\lambda)u_j^{(n)} + \lambda u_{j-1}^{(n)} + \lambda u_{j+1}^{(n)}, \text{ avec } \lambda = k/h^2, \tag{2.11a}$$
$$u_j^0(x) = \mathrm{e}^{ipjh} \text{ pour } j = 1,\dots,N \text{ avec } h = 2\pi/(N+1) \tag{2.11b}$$

2. Soit H un espace de Hilbert, $(\psi_p)_{p\in\mathbb{Z}}$ est une base hilbertienne de H si $(\psi_p)_{p\in\mathbb{Z}}$ est une famille orthonormée telle que $\forall v \in H$, $\exists (v_p)_{i\in\mathbb{Z}} \subset \mathbb{R}$; $v = \sum_{p\in\mathbb{Z}} v_p\psi_p$ au sens de la convergence dans H, avec $v_p = (x,\psi_p)$ où $(\,\cdot\,,\,\cdot\,)$ désigne le produit scalaire sur H.

On a bien $u_0^0 = u_{N+1}^0 = 0$, et $u_j^{(1)} = (1-2\lambda)\,\mathrm{e}^{ipjh} + \lambda\,\mathrm{e}^{ip(j-1)h} + \lambda\,\mathrm{e}^{ip(j+1)h}$, donc $u_j^{(1)} = \mathrm{e}^{ipjh}\,\xi_p$ avec

$$\begin{aligned}\xi_p &= 1 - 2\lambda + \lambda\,\mathrm{e}^{-iph} + \lambda\,\mathrm{e}^{iph}\\ &= 1 - 2\lambda + 2\lambda\cos ph\\ &= 1 - 2\lambda + 2\lambda\Big(1 - 2\sin^2\frac{ph}{2}\Big)\\ &= 1 - 4\lambda\sin^2\Big(\frac{p\pi}{N+1}\Big).\end{aligned} \tag{2.12}$$

Le coefficient ξ_p s'appelle facteur d'amplification associé au p-ième mode de Fourier ψ_p. Il est facile alors de calculer $u_j^{(n)}$:

$$u_j^{(1)} = \xi_p u_j^{(0)}, u_j^{(2)} = (\xi_p)^2 u_j^{(0)}, \text{ et par récurrence sur } n, u_j^{(n)} = (\xi_p)^n u_j^{(0)}$$

On dit que le schéma est *stable au sens de Von Neumann* s'il conserve la propriété d'amortissement des modes de Fourier du cas continu, vu plus haut, c'est-à-dire si $|\xi_p| < 1$ pour tout $p \in \mathbb{Z}$. Pour cela, il faut et il suffit que

$$-1 < 1 - 4\lambda\sin^2\Big(\frac{p\pi}{N+1}\Big) < 1, \text{ c'est-à-dire } \lambda < \frac{1}{2}.$$

Une condition suffisante pour que le schéma soit stable au sens de Von Neumann est donc que $\lambda < 1/2$ (on rappelle que $\lambda = k/h^2$) ; c'est la même condition que pour la stabilité au sens des erreurs d'arrondis et au sens de la norme L^∞.

Convergence du schéma avec la technique de Von Neumann Soit $u \in \mathcal{C}^2(]0\,;2\pi[\times]0,T[,\mathbb{C})$ la solution exacte de (2.9) ; on a donc

$$u(jh,nk) = \sum_{p\in\mathbb{Z}} c_p(0) e^{-p^2 nk} e^{ipjh}$$

où $h = 2\pi/(N+1)$ est le pas de discrétisation en espace et $k = T/M$ le pas de discrétisation en temps. Soit $u_j^{(n)}$ l'approximation de $u(jh,nk)$ obtenue par le schéma d'Euler explicite (2.11) ; on a $u_j^{(n)} = \sum_{p\in\mathbb{Z}} c_p(0)(\xi_p)^n\,\mathrm{e}^{ipjh}$, et on cherche à montrer la convergence de la solution approchée vers la solution exacte au sens suivant :

Proposition 2.12 — Convergence par la technique Von Neumann. Soient $u_0 = \sum_{p\in\mathbb{Z}} c_n(0)\psi_p$, u la solution du problème (2.9), et $(u_j^{(n)}, j = 1,\dots,N, n = 1,\dots,M)$ la solution approchée obtenue par le schéma d'Euler explicite (2.11). On suppose que

$$\sum_{p\in\mathbb{Z}} |c_p(0)| < +\infty. \tag{2.13}$$

On alors la majoration d'erreur suivante :

$$\forall\varepsilon > 0,\ \exists\eta \geqslant 0 \text{ tel que si } k \leqslant \eta \text{ et } \frac{k}{h^2} \leqslant \frac{1}{2}, \text{ alors } |u(jh,nk) - u_j^{(n)}| \leqslant \varepsilon, \forall j = 1,\dots,N, \forall n = 1,\dots,M.$$

Démonstration. Pour alléger les notations, on va montrer le résultat pour $n = M$. Le résultat pour tout $n \leqslant M$ se déduit facilement en remplaçant dans la démonstration M par n et $T = kM$ par $t_n = kn$. On pose $e_j^{(M)} = u(jh,kM) - u_j^{(M)}$. Par l'hypothèse (2.13), pour tout $\varepsilon \in \mathbb{R}+$, il existe $A \in \mathbb{R}$ tel que $2\sum_{|p|\geqslant A} |c_p(0)| \leqslant \varepsilon$. On écrit alors :

$$e_j^{(M)} = \sum_{|p|\leqslant A} c_p(0)(e^{-p^2T} - \xi_p^M)e^{ipjh} + \sum_{|p|\geqslant A} c_p(0)(e^{-p^2T} - \xi_p^M)e^{ipjh}.$$

On a donc :

$$|e_j^{(M)}| \leqslant X^{(M)} + 2\sum_{|p|\geqslant A} |c_p(0)| \leqslant X^{(M)} + 2\varepsilon, \text{ avec } X^{(M)} = \sum_{|p|\leqslant A} |c_p(0)||e^{-p^2T} - \xi_p^M|$$

Montrons maintenant que $X^{(M)} \to 0$ lorsque $h \to 0$. On a vu que $\xi_p = 1 - 4\lambda \sin^2 \frac{ph}{2}$. Or, $\sin^2 \frac{ph}{2} = \frac{p^2h^2}{4} + O(h^4)$ et $\lambda = \frac{k}{h^2}$. On a donc $4\lambda \sin^2 \frac{ph}{2} = p^2 k + O(kh^2)$, et on déduit que

$$(\xi_p)^M = \left(1 - 4\lambda \sin^2 \frac{ph}{2}\right)^{T/k} \text{ et donc } \ln \xi_p^M = \frac{T}{k} \ln\left(1 - 4\lambda \sin^2 \frac{ph}{2}\right) = -Tp^2 + O(h^2)$$

Il s'ensuit que $\xi_p^M \to e^{-p^2 T}$ lorsque $h \to 0$. Tous les termes de X tendent vers 0, et X est une somme finie ; on a donc ainsi montré que $e_j^{(M)}$ tend vers 0 lorsque h tend vers 0. •

Remarque 2.13 On peut adapter la technique de Von Neumann au cas Dirichlet homogène sur $[0\,;1]$, en effectuant le développement de u par rapport aux fonctions propres de l'opérateur u'' avec conditions aux limites de Dirichlet : $u(x,t) = \sum c_n(t) \sin(n\pi x)$. L'avantage du développement en série de Fourier est qu'il fonctionne pour n'importe quel opérateur linéaire, sans avoir besoin de la connaissance de ses fonctions propres, à condition d'avoir pris des conditions aux limites périodiques.

2.3 Schéma implicite et schéma de Crank-Nicolson

2.3.1 Le θ-schéma

Commençons par un petit rappel sur les équations différentielles (voir aussi polycopié d'analyse numérique de licence). On considère le problème de Cauchy :

$$\begin{cases} y'(t) = f(y(t)) \quad t > 0 \\ y(0) = y_0 \end{cases}$$

Soit k un pas (constant) de discrétisation, on rappelle que les schémas d'Euler explicite et implicite pour la discrétisatision de ce problème s'écrivent respectivement :

$$\text{Euler explicite : } \frac{y^{(n+1)} - y^{(n)}}{k} = f(y^{(n)}), \quad n \geqslant 0, \tag{2.14}$$

$$\text{Euler implicite : } \frac{y^{(n+1)} - y^{(n)}}{k} = f(y^{(n+1)}), \quad n \geqslant 0, \tag{2.15}$$

avec $y^{(0)} = y_0$. On rappelle également que le θ-schéma, où θ est un paramètre de l'intervalle $[0\,;1]$ s'écrit :

$$y^{(n+1)} = y^{(n)} + k\theta f(y^{(n+1)}) + k(1-\theta) f(y^{(n)}).$$

Remarquons que pour $\theta = 0$ on retrouve le schéma explicite (2.14) et pour $\theta = 1$ le schéma implicite (2.15). On peut facilement adapter le θ-schéma à la résolution des équations paraboliques. Par exemple, le θ-schéma pour la discrétisation en temps du problème (2.1), avec une discrétisation par différences finies en espace s'écrit :

$$\begin{aligned} &\frac{u_i^{(n+1)} - u_i^{(n)}}{k} = \frac{\theta(-2u_i^{(n+1)} + u_{i-1}^{(n+1)} + u_{i+1}^{(n+1)})}{h^2} + \frac{(1-\theta)(-2u_i^{(n)} + u_{i-1}^{(n)} + u_{i+1}^{(n)})}{h^2}, \ n \geqslant 0, \ i = 1, \ldots, N \\ &u_i^{(0)} = u_0(x_i) \quad i = 1, \ldots, N, \\ &u_0^{(n)} = u_{N+1}^{(n)} = 0, \forall n \geqslant 0. \end{aligned} \tag{2.16}$$

Si $\theta = 0$, on retrouve le schéma d'Euler explicite ; si $\theta = 1$, celui d'Euler implicite. Dans ce cas où $\theta = 1/2$ ce schéma s'appelle schéma de Crank [3]-Nicolson [4]. Notons que dès que $\theta > 0$, le schéma est implicite, au sens où on n'a pas d'expression explicite de $u_i^{(n+1)}$ en fonction des $u_j^{(n)}$.

3. John Crank, 6 février 1916–3 octobre 2006, mathématicien britannique
4. Phyllis Nicolson, 21 Septembre 1917–6 Octobre 1968, mathématicienne britannique

2.3.2 Consistance et stabilité

Proposition 2.14 — Consistance du θ-schéma. Le θ schéma (2.16) pour la discrétisation du problème (2.1) est d'ordre 2 en espace. Il est d'ordre 2 en temps si $\theta = \frac{1}{2}$, et d'ordre 1 sinon.

Démonstration. On pose $\bar{u}_j^n = u(x_j, t_n), h = \frac{1}{N+1}$. Comme $\partial_t u(x_j, t_n) + \partial^2_{xx} u(x_j, t_n) = 0$, on a par définition de l'erreur de consistance :

$$R_j^{(n)} = \frac{1}{k}(\bar{u}_j^{(n+1)} - \bar{u}_j^{(n)}) + \frac{\theta}{h^2}\left(2\bar{u}_j^{(n+1)} - \bar{u}_{j-1}^{(n+1)} - \bar{u}_{j+1}^{(n+1)}\right) + \frac{1-\theta}{h^2}\left(-2\bar{u}_j^{(n)} + \bar{u}_{j-1}^{(n)} + \bar{u}_{j+1}^{(n)}\right)$$

On va montrer, en effectuant des dévelopements limités, que :

$$\left|R_j^{(n)}\right| \leqslant C(k+h^2) \text{ si } \theta \neq \tfrac{1}{2}, \text{ et } \left|R_j^{(n)}\right| \leqslant C(k^2+h^2) \text{ si } \theta = \tfrac{1}{2}.$$

Dans ce but, décomposons $R_j^{(n)}$:

$$R_j^{(n)} = T_j^{(n)} + \theta X_j^{(n+1)} + (1-\theta) X_j^{(n)}, \text{ avec}$$

$$T_j^{(n)} = \frac{\bar{u}_j^{(n+1)} - \bar{u}_j^{(n)}}{k},$$

$$X_j^{(n+1)} = \frac{1}{h^2}\left(-2\bar{u}_i^{(n+1)} + \bar{u}_{i-1}^{(n+1)} + \bar{u}_{i+1}^{(n+1)}\right),$$

$$X_j^{(n)} = \frac{1}{h^2}\left(-2\bar{u}_i^{(n)} + \bar{u}_{i-1}^{(n)} + \bar{u}_{i+1}^{(n)}\right).$$

Effectuons un développement limité pour calculer $T_j^{(n)}$:

$$T_j^{(n)} = (\partial_t u)(x_j, t_n) + \frac{k}{2}(\partial^2_{tt} u)(x_j, t_n) + R_1 \quad \text{avec } |R_1| \leqslant Ck^2.$$

Faisons de même pour $X_j^{(n+1)}$;

$$X_j^{(n+1)} = \partial^2_{xx} u(x_j, t_{n+1}) + R_2 \text{ avec } |R_2| \leqslant Ch^2.$$

Or

$$\partial^2_{xx} u(x_j, t_{n+1}) = \partial^2_{xx} u(x_j, t_n) + k\partial^3_{xxt} u(x_j, t_n) + R_3 \text{ avec } |R_3| \leqslant Ck^2,$$

et donc

$$X_j^{(n+1)} = (\partial^2_{xx} u(x_j, t_n) + + k\partial^3_{xxt} u(x_j, t_n) + R_4) \text{ avec } |R_4| \leqslant C(h^2+k^2).$$

De même pour $X_j^{(n)}$, on a :

$$X_j^{(n)} = \partial_{xx} u(x_j, t_n) + R_5, \text{ avec } |R_5| \leqslant Ck^2.$$

En regroupant, on obtient que

$$R_j^{(n)} = \partial_t u(x_j, t_n) + \frac{k}{2}\partial^2_{tt} u(x_j, t_n) + \theta(\partial^2_{xx} u(x_j, t_n) + k\partial^3_{xxt})u(x_j, t_n) + (1-\theta)\partial^2_{xx} u(x_j, t_n) + R,$$

avec $R = R_1 + \theta R_4 + (1-\theta)R_5$ et $R \leqslant C(k^2+h^2)$. Mais $\partial_t u(x_j, t_n) + \partial^2_{xx} u(x_j, t_n) = 0$ et donc $\partial^3_{xxt} u(x_j, t_n) = -\partial^2_{tt} u(x_j, t_n)$. On en déduit que

$$R_j^{(n)} = \partial_t u(x_j, t_n) + k(\frac{1}{2} - \theta)\partial^2_{tt} u(x_j, t_n) + \partial^2_{xx} u(x_j, t_n) + R,$$

Le schéma est donc d'ordre 2 en temps et en espace si $\theta = 1/2$. Si $\theta \neq 1/2$ on a un schéma d'ordre 2 en espace et d'ordre 1 en temps. •

Proposition 2.15 — Stabilité au sens de Von Neumann.

- Si $\theta \geqslant 1/2$, le θ-schéma est inconditionnellement stable. En particulier, les schémas d'Euler implicite et de Crank-Nicolson sont inconditionnellement stables.
- Si $\theta < 1/2$, le schéma est stable à condition que : $\lambda \leqslant \frac{1}{2(1-2\theta)}$.

On retrouve en particulier que le schéma d'Euler explicite n'est stable que si $\lambda \leqslant 1/2$.

Démonstration. On remplace les conditions aux limites de Dirichlet sur $[0\,;1]$ par des conditions périodiques sur $[0\,;2\pi]$. La solution exacte sécrit alors $u(x,t) = \sum_{p\in\mathbb{Z}} c_p(0)\,\mathrm{e}^{-p^2 t}\,\mathrm{e}^{ipx}$. Prenons comme condition initiale le p-ième mode de Fourier : $u_0 = \psi_p$, c'est-à-dire $u_0(x) = \mathrm{e}^{ipx}$. On a :

$$u_j^{(n+1)} - u_j^{(n)} = \frac{k}{h^2}\left(-\theta(2u_j^{(n+1)} - u_{j-1}^{(n+1)} - u_{j+1}^{(n+1)}) - (1-\theta)(2u_j^{(n)} - u_{j-1}^{(n)} - u_{j+1}^{(n)})\right)$$

ce qui s'écrit encore, avec $\lambda = k/h^2$:

$$(1+2\lambda)u_j^{(n+1)} - \lambda\theta u_{j-1}^{(n+1)} - \lambda\theta u_{j+1}^{(n+1)} = (1-2\lambda(1-\theta))u_j^{(n)} + \lambda(1-\theta)u_{j+1}^{(n)} + \lambda(1-\theta)u_{j-1}^{(n)}.$$

En discrétisant la condition initiale u_0, on obtient $u_j^{(0)} = \mathrm{e}^{ipjh}$ et on cherche le facteur d'amplification ξ_p tel que $u_j^{(1)} = \xi_p u_j^{(0)} = \xi_p \mathrm{e}^{ipjh}$; en appliquant le schéma ci-dessus pour $n = 0$, on obtient :

$$(1 + 2\lambda\theta)\xi_p - \lambda\theta\xi_p[e^{-iph} + e^{iph}] = [1 - 2\lambda(1 - \theta)] + \lambda(1 - \theta)[e^{-iph} + e^{iph}]$$

et donc :

$$\xi_p = \frac{1 - 2\lambda(1-\theta) + 2\lambda(1-\theta)\cos ph}{(1 + 2\lambda\theta) - 2\lambda\cos ph} = \frac{1 - 4\lambda(1-\theta)\sin^2 ph/2}{1 + 4\lambda\theta\sin^2 \frac{ph}{2}}.$$

Pour que le schéma soit stable au sens de Von Neumann, il faut que $|\xi_p| < 1$ pour tout p ; comme $1 + 4\lambda\theta\sin^2 \frac{ph}{2} > 0$, il suffit que les deux conditions suivantes soient vérifiées :

$$1 - 4\lambda(1-\theta)\sin^2\frac{ph}{2} < 1 + 4\lambda\theta\sin^2\frac{ph}{2}, \tag{2.17}$$

$$4\lambda(1-\theta)\sin^2\frac{ph}{2} - 1 < 1 + 4\lambda\theta\sin^2\frac{ph}{2}. \tag{2.18}$$

L'inégalité (2.17) est toujours vérifiée. En ce qui concerne l'inégalité (2.18), on distingue deux cas :

1. Si $\theta \geqslant 1/2$ alors $0 \leqslant 1 - \theta \leqslant \theta$ et dans ce cas (2.18) est toujours vraie.
2. Si $\theta < 1/2$, on veut que :

$$4\lambda\left((1-\theta)\sin^2\frac{ph}{2} - \theta\sin^2\frac{ph}{2}\right) < 2, \text{ soit encore } \lambda < \frac{1}{2}\left((1-2\theta)\sin^2\frac{ph}{2}\right)^{-1}.$$

Une condition suffisante est donc : $\lambda \leqslant \dfrac{1}{2(1-2\theta)}$ si $\theta < \dfrac{1}{2}$.

•

2.3.3 Convergence du schéma d'Euler implicite

Prenons $\theta = 1$ dans le θ-schéma : on obtient le schéma d'Euler implicite :

$$(1 + 2\lambda)u_j^{(n+1)} - \lambda u_{j-1}^{(n+1)} - \lambda u_{j+1}^{(n+1)} = u_j^{(n)} \text{ avec } \lambda = k/h^2. \tag{2.19}$$

On rappelle que ce schéma est inconditionnellement stable au sens de Von Neumann. On va montrer de plus qu'il est inconditionnellement L^∞-stable :

Proposition 2.16 — Stabilité L^∞ pour Euler implicite. Si $(u_j^{(n)})_{j=1,\dots,N}$ est solution du schéma (2.19), alors :

$$\max_{j=1,\dots,N} u_j^{(n+1)} \leqslant \max_{j=1,\dots,N} u_j^{(n)} \leqslant \max_{j=1,\dots,N} u_j^{(0)} \tag{2.20}$$

de même :

$$\min_{j=1,\dots,N} u_j^{(n+1)} \geqslant \min_{j=1,\dots,N} u_j^{(n)} \geqslant \min_{j=1,\dots,N} u_j^{(0)} \tag{2.21}$$

Le schéma (2.19) est donc L^∞ stable.

Démonstration. Prouvons l'estimation (2.20), la preuve de (2.21) est similaire. Soit j_0 tel que $u_{j_0}^{(n+1)} = \max_{j=1,\dots,N} u_j^{(n+1)}$. Par définition du schéma d'Euler implicite (2.19), on a :

$$u_{j_0}^{(n)} = (1+2\lambda)u_{j_0}^{(n+1)} - \lambda u_{j_0-1}^{(n+1)} - \lambda u_{j_0+1}^{(n+1)} = u_{j_0}^{(n+1)} + \lambda(u_{j_0}^{(n+1)} - u_{j_0-1}^{(n+1)}) + \lambda(u_{j_0}^{(n+1)} - u_{j_0+1}^{(n+1)}) \geqslant u_{j_0}^{(n+1)}.$$

On en déduit $u_{j_0}^{(n+1)} \leqslant \max_{j=1,\dots,N} u_j^{(n)}$, ce qui prouve que $\max_{j=1,\dots,N} u_j^{(n+1)} \leqslant \max_{j=1,\dots,N} u_j^{(n)}$. Donc le schéma (2.19) est L^∞ stable. •

Théorème 2.3 — Convergence, Euler implicite. Soit $e^{(n)} = (e_1^{(n)}, \dots, e_N^{(n)})$ l'erreur de discrétisation, définie par $e_j^{(n)} = u(x_j, t_n) - u_j^{(n)}$ pour $j = 1, \dots, N$, avec u solution de (2.1) et $(u_j^{(n)})_{\substack{j=1,\dots,N \\ n=1,\dots,M}}$ solution du schéma (2.19). Alors

$$\|e^{(n+1)}\|_\infty \leqslant \|e^{(0)}\|_\infty + TC(k + h^2)$$

Si $\|e^{(0)}\|_\infty = 0$, le schéma est donc convergent d'ordre 1 en temps et 2 en espace.

Démonstration. En utilisant la définition de l'erreur de consistance, on obtient :

$$(1+2\lambda)e_j^{(n+1)} - \lambda e_{j-1}^{(n)} - \lambda e_{j+1}^{(n)} = e_j^{(n)} + kR_j^{(n)},$$

et donc $\|e^{(n+1)}\|_\infty \leqslant \|e^{(n)}\|_\infty + kC(k+h^2)$. On en déduit, par récurrence sur n, que

$$\|e^{(n+1)}\|_\infty \leqslant \|e^{(0)}\|_\infty + TC(k+h^2),$$

ce qui montre la convergence du schéma. •

On peut montrer que le schéma saute-mouton

$$\frac{u_j^{(n+1)} - u_j^{(n-1)}}{2k} = \frac{u_{j-1}^{(n)} - 2u_j^{(n)} + u_{j+1}^{(n)}}{h^2}$$

est d'ordre 2 en espace et en temps [exercice 33]. Malheureusement il est aussi inconditionnellement instable. On peut le modifier pour le rendre stable, en introduisant le schéma de Dufort-Frankel, qui s'écrit :

$$\frac{u_j^{(n+1)} - u_j^{(n-1)}}{2k} = \frac{u_{j-1}^{(n)} - (u_j^{(n+1)} + u_j^{(n-1)}) + u_{j+1}^{(n)}}{h^2}$$

Ce schéma est consistant et inconditionnellement stable [exercice 33].

Remarque 2.17 — Cas de la dimension 2. Soit Ω un ouvert borné de $\mathbb{R}^2$, on considère le problème suivant :

$$\begin{cases} \partial_t u - \Delta u = 0 & x \in \Omega \quad t \in]0,T[\\ u(x,0) = u_0(x) & x \in \Omega \\ u(x,t) = g(t) & x \in \partial\Omega \quad \forall t \in]0,T[\end{cases}$$

Si le domaine est rectangulaire, ce problème se discrétise facilement à l'aide de θ schéma en temps et de différences finies en espace, en prenant un maillage rectangulaire. On peut montrer, comme dans le cas 1D, la consistance, la stabilité, la L^∞ stabilité, la stabilité au sens de Von Neumann.

2.4 Exercices

2.4.1 Énoncés

corrigé p.85

Exercice 25 — Existence de solutions "presque classiques". Soit $u_0 \in L^2(]0\,;1[)$. On s'intéresse au problème :

$$\begin{cases} \partial_t u(x,t) - \partial_{xx}^2 u(x,t) = 0, & x \in]0\,;1[, \quad t \in \mathbb{R}_+^\star, \\ u(0,t) = u(1,t) = 0, & t \in \mathbb{R}_+^\star, \\ u(x,0) = u_0(x), & x \in]0\,;1[. \end{cases} \tag{2.22}$$

1. On définit $u : [0\,;1] \times \mathbb{R}_+^\star \to \mathbb{R}$ par :

$$u(x,t) = \sum_{n\in\mathbb{N}^\star} e^{-n^2\pi^2 t} a_n \sin(n\pi x) \quad x \in [0\,;1] \quad t \in \mathbb{R}_+^\star,$$

avec :

$$a_n = \frac{\displaystyle\int_0^1 u_0(x)\sin(n\pi x)dx}{\displaystyle\int_0^1 \sin^2(n\pi x)dx}$$

Montrer que u est bien définie de $[0\,;1] \times \mathbb{R}_+^\star$ dans $\mathbb{R}$ et est solution de (2.22) au sens suivant :

$$\begin{cases} u \in \mathcal{C}^\infty([0\,;1] \times \mathbb{R}_+^\star, \mathbb{R}) \\ \partial_t u(x,t) - \partial_{xx}^2 u(x,t) = 0 & \forall x \in [0\,;1] \quad \forall t \in \mathbb{R}_+^\star \\ u(0,t) = u(1,t) = 0 & \forall t \in \mathbb{R}_+^\star \\ \lim\limits_{t\to 0} \|u(\,\cdot\,,t) - u_0\|_{L^2(]0\,;1[)} \to 0 \end{cases} \tag{2.23}$$

2. Montrer qu'il existe une unique fonction u solution de (2.23).

Exercice 26 — Discrétisation par différences finies. Soit $u_0 \in \mathcal{C}(]0\,;1[)$. On s'intéresse au problème :

$$\begin{cases} \partial_t u(x,t) + \partial_x u(x,t) - \partial^2_{xx} u(x,t) = 0 & x \in]0\,;1[\quad t \in]0,T[\\ u(0,t) = a \qquad u'(1,t) = b & t \in \mathbb{R}_+^\star \\ u(x,0) = u_0(x) & x \in]0\,;1[\end{cases} \tag{2.24}$$

avec $T > 0$, a et $b \in \mathbb{R}$ donnés.

Écrire une discrétisation espace-temps du problème (2.24) avec le schéma d'Euler explicite en temps et par différences finies avec un maillage uniforme en espace, en utilisant un schéma décentré amont pour le terme d'ordre 1 $\partial_x u(x,t)$.

corrigé p.86

Exercice 27 — Exemple de schéma non convergent. Soit $u_0 \in L^2(]-4\,;4[)$. On note u l'unique solution (au sens vu en cours ou en un sens inspiré de l'exercice précédent) du problème suivant :

suggestions p.84

$$\begin{cases} \partial_t u(x,t) - \partial^2_{xx} u(x,t) = 0 & x \in]-4\,;4[\quad t \in]0\,;1[\\ u(-4,t) = u(4,t) = 0 & t \in]0\,;1[\\ u(x,0) = u_0(x) & x \in]-4\,;4[. \end{cases} \tag{2.25}$$

On sait que la solution de (2.25) est de classe $\mathcal{C}^\infty$ sur $[-4,4]\times]0,1]$ (voir l'exercice précédent). On admettra que si $u_0 \geqslant 0$ p.p. sur $]-4\,;4[$ et $u_0 \neq 0$ (dans $L^2(]-4\,;4[)$) alors $u(x,t) > 0$ pour tout $x \in]-4\,;4[$ et tout $t \in]0,1]$.

On suppose maintenant que $u_0 \in \mathcal{C}([-4,4],\mathbb{R})$, $u_0(-4) = u_0(4) = 0$, $u_0 \geqslant 0$ sur $]-4\,;4[$, u_0 nulle sur $[-3,4]$ et qu'il existe $a \in]-4,-3[$ t.q. $u_0(a) > 0$. On a donc $u(x,t) > 0$ pour tout $x \in]-4\,;4[$.

Avec les notations du cours, on considère la solution de (2.25) donnée par le schéma d'Euler explicite (2.2) avec le pas de temps $k = 1/(M+1)$ et le pas d'espace $h = 8/(N+1)$ ($M, N \in \mathbb{N}^\star$, N impair). La solution approchée est définie par les valeurs u_i^n pour $i \in \{-(N+1)/2, \dots, (N+1)/2\}$ et $n \in \{0, \dots, M+1\}$. La valeur u_i^n est censée être une valeur approchée de $\overline{u}_i^n = u(ih, nk)$.

1. Donner les équations permettant de calculer u_i^n pour $i \in \{-(N+1)/2, \dots, (N+1)/2\}$ et $n \in \{0, \dots, M+1\}$.

2. On suppose maintenant que $k = h$. Montrer que $u_i^n = 0$ pour $i \geqslant 0$ et $n \in \{0, \dots, M+1\}$. En déduire que $\max\{|u_i^{M+1} - \overline{u}_i^{M+1}|,\ i \in \{-(N+1)/2, \dots, (N+1)/2\}$ ne tend pas vers 0 quand $h \to 0$ (c'est-à-dire quand $N \to \infty$).

corrigé p.87

Exercice 28 — Schémas explicites centré et décentré. Soient $\alpha > 0$, $\mu > 0$, $T > 0$ et $u_0 : \mathbb{R} \to \mathbb{R}$. On s'intéresse au problème suivant :

$$\begin{cases} \partial_t u(x,t) + \alpha \partial_x u(x,t) - \mu \partial^2_{xx} u(x,t) = 0 & x \in]0\,;1[\quad t \in]0,T[\\ u(0,t) = u(1,t) = 0 & t \in]0,T[\\ u(x,0) = u_0(x) & t \in]0\,;1[\end{cases} \tag{2.26}$$

On rappelle que $\partial_t u = \partial u/\partial t$, $\partial_x u = \partial u/\partial x$ et $\partial^2_{xx} u = \partial^2 u/\partial x^2$. On suppose qu'il existe $u \in C^4([0\,;1]\times[0\,;T])$ solution (classique) de (2.26) (noter que ceci implique $u_0(0) = u_0(1) = 0$). On pose $A = \min\{u_0(x),\ x \in [0\,;1]\}$ et $B = \max\{u_0(x),\ x \in [0\,;1]\}$ (noter que $A \leqslant 0 \leqslant B$).

On discrétise le problème (2.26). On reprend les notations du cours. Soient $h = 1/(N+1)$ et $k = T/M$ (N, $M \in \mathbb{N}^\star$).

1. Propriétés de la solution de (2.26).

(a) Montrer que $A \leqslant u(x,t) \leqslant B$ pour tout $(x,t) \in [0\,;1]\times[0\,;T]$ [Pour montrer $u(x,t) \leqslant B$, on pourra, par exemple, multiplier la première équation de (2.26) par $\varphi(u)$ et integrer sur $[0\,;1]\times[0\,;T]$, avec $\varphi \in C^1(\mathbb{R},\mathbb{R})$, $\varphi(s) = 0$ si $s \leqslant B$ et $\varphi'(s) > 0$ si $s > B$]. En déduire que $\|u(.,t)\|_{L^\infty(]0\,;1[)} \leqslant \|u_0\|_{L^\infty(]0\,;1[)}$ pour tout $t \in [0\,;T]$.

(b) Montrer que $\|u(.,t)\|_{L^2(]0\,;1[)} \leqslant \|u_0\|_{L^2(]0\,;1[)}$ pour tout $t \in [0\,;T]$.

2. Schéma explicite décentré. Pour approcher la solution u de (2.26), on considère le schéma suivant :

$$\begin{cases} \dfrac{u_i^{n+1}-u_i^n}{k}+\dfrac{\alpha(u_i^n-u_{i-1}^n)}{h}-\dfrac{\mu(u_{i+1}^n-2u_i^n+u_{i-1}^n)}{h^2}=0 & i=1,\dots,N \qquad n=0,\dots,M-1 \\ u_0^n=u_{N+1}^n=0 & n=1,\dots,M \\ u_i^0=u_0(ih) & i=0,\dots,N+1. \end{cases} \tag{2.27}$$

On pose $\overline{u}_i^n = u(ih, nk)$ pour $i = 0, \dots, N+1$ et $n = 0, \dots, M$.

(a) Consistance. Montrer que l'erreur de consistance du schéma (2.27) est majorée par $C_1(k+h)$, où C_1 ne dépend que de u, T, α et μ.

(b) Stabilité. Sous quelle condition sur k et h (cette condition peut dépendre de α et μ) a-t-on $A \leqslant u_i^n \leqslant B$ pour tout $i \in \{0, \dots, N+1\}$ et tout $n \in \{0, \dots, M\}$? Sous cette condition, en déduire $\|u^n\|_\infty \leqslant \|u_0\|_{L^\infty(]0\,;1[)}$ pour tout $n \in \{0, \dots, M\}$ (avec $\|u^n\|_\infty = \max\{|u_i^n|, i \in \{0, \dots, N+1\}\}$)

(c) Estimation d'erreur. On pose $e_i^n = \overline{u}_i^n - u_i^n$. Sous la condition sur k et h trouvée précédemment, montrer que $|e_i^n| \leqslant C_2(k+h)$ pour tout $i \in \{0, \dots, N+1\}$ et tout $n \in \{0, \dots, M\}$ avec C_2 ne dépendant que de u, T, α et μ.

3. Schéma explicite centré. On change dans le schéma (2.27) la quantité $(\alpha/h)(u_i^n - u_{i-1}^n)$ par $(\alpha/2h)(u_{i+1}^n - u_{i-1}^n)$.

(a) Consistance. Montrer que l'erreur de consistance est maintenant majorée par $C_3(k+h^2)$, où C_3 ne dépend que de u, T, α et μ.

(b) Reprendre les questions de stabilité et d'estimation d'erreur du schéma (2.27).

Exercice 29 — Schéma implicite et principe du maximum. Cet exercice est le pendant instationnaire des exercices 5 et 6, il est donc utile d'avoir effectué ces deux exercices avant celui-ci.

1. Forme non conservative – Soient $T > 0$, $v \in C^1([0\,;1], \mathbb{R}_+)$, $a, b \in \mathbb{R}$ et $u_0 \in C([0\,;1])$. On considère le problème d'évolution suivant :

$$\begin{cases} \partial_t u(x,t) - \partial^2_{xx} u(x,t) + v(x)\partial_x u(x,t) = 0, \quad x \in]0\,;1[, \quad t \in]0, T[, \\ u(0,t) = a, \qquad u(1,t) = b, \ t \in]0, T[, \\ u(x,0) = u_0(x), \end{cases} \tag{2.28}$$

dont on cherche à approcher la solution par différences finies. On choisit pour cela le schéma de l'exercice 5 pour la discrétisation en espace, et on discrétise par le schéma d'Euler implicite en temps avec un pas de temps uniforme $k = \frac{T}{P}$ où $P \geqslant 1$.

(a) Écrire le schéma ainsi obtenu et montrer qu'il admet une solution qu'on notera $U = (u_i^{(p)})_{i=1,\dots N, p=1,\dots,P}$, où $u_i^{(p)}$ est censé être une approximation de $u(x_i, t_p)$, où $t_p = pk, p = 0, \dots, P$.

(b) Montrer que

$$\min(\min_{[0\,;1]} u_0, \min(a,b)) \leqslant u_i^{(p)} \leqslant \max(\max_{[0\,;1]} u_0, \max(a,b)) \qquad i = 1, \dots, N \quad p = 1, \dots, P \tag{2.29}$$

2. Forme conservative – Soit $T > 0$, et $u_0 \in \mathcal{C}([0\,;1])$. On considère maintenant le problème d'évolution suivant :

$$\begin{cases} \partial_t u(x,t) - \partial^2_{xx} u(x,t) + \partial_x(vu)(x,t) = 0, \quad x \in]0\,;1[, \quad t \in]0, T[, \\ u(0) = a, \qquad u(1) = b, \\ u(x,0) = u_0(x). \end{cases}$$

dont on cherche à approcher la solution par différences finies. On choisit pour cela le schéma de la question 2 de l'exercice 6 pour la discrétisation en espace, et on discrétise par le schéma d'Euler implicite en temps avec un pas de temps uniforme $k = \frac{T}{P}$ où $P \geqslant 1$.

(a) Écrire le schéma ainsi obtenu et montrer qu'il admet une solution qu'on notera $U = (u_i^{(p)})_{i=1,\dots N, p=1,\dots,P}$, où $u_i^{(p)}$ est censé être une approximation de $u(x_i, t_p)$, où $t_p = pk$, $p = 0, \dots, P$.

(b) Montrer que si $a \geqslant 0$, $b \geqslant 0$ et $u_0 \geqslant 0$, alors on a $u_i^{(p)} \geqslant 0$, $i = 1, \dots, N$ et $p = 1, \dots, P$.

3. Le cas bi-dimensionnel – On considère maintenant $\Omega =]0\,;1[^2$; soient $v \in \mathcal{C}^\infty(\Omega, (\mathbb{R}_+)^2)$ ($v(x)$ est donc un vecteur de $\mathbb{R}^2$), $a \in \mathcal{C}(\partial\Omega, \mathbb{R})$ et $u_0 \in \mathcal{C}(\Omega, \mathbb{R}_+)$. En s'inspirant des schémas étudiés aux questions précédentes, donner une discrétisation en espace et en temps des deux problèmes suivants (avec pas uniforme) :

$$\begin{cases} \partial_t u - \Delta u + v \cdot \nabla u = 0, \\ u(x,t) = a, x \in \partial\Omega, t \in]0, T[, \\ u(\,\cdot\,, 0) = u_0. \end{cases}$$

$$\begin{cases} \partial_t u - \Delta u + \operatorname{div}(vu) = 0, \\ u(x,t) = a, x \in \partial\Omega, t \in]0, T[, \\ u(\,\cdot\,, 0) = u_0. \end{cases}$$

Exercice 30 — Discrétisation d'un problème parabolique. On s'intéresse à des schémas numériques pour le problème parabolique :

suggestions p.84

$$\begin{cases} \partial_t u + \partial_x u - \varepsilon \partial^2_{xx} u = 0 & (x,t) \in \mathbb{R}^+ \times]0, T[\\ u(1,t) = u(0,t) = 0 & t \in]0, T[\\ u(x,0) = u_0(x) & x \in]0\,;1[\end{cases} \tag{2.30}$$

où $u_0 \in C([0\,;1])$ et $\varepsilon > 0$ sont donnés. On admettra que qu'il existe une unique solution $u \in C^4(\mathbb{R}, \mathbb{R})$ de (2.30).

1. Euler explicite

(a) Ecrire le schéma d'approximation de (2.30) par différences finies à pas constant (noté h, et tel que $Nh = 1$), centré en espace (c'est-à-dire en approchant $u'(ih)$ par $\frac{1}{2h}(u((i+1)h) - u((i-1)h))$ et $u''(ih)$ par $\frac{1}{h^2}(u((i+1)h) + u((i-1)h) - 2u(ih))$), et avec le schéma d'Euler explicite à pas constant (noté k, avec $T = Mk$) en temps.

(b) Montrer que l'erreur de consistance est majorée par $C(k + h^2)$ où C ne dépend que de la solution exacte de (2.30).

(c) Sous quelle(s) condition(s) sur k et h a-t-on le résultat de stabilité $\|u^n\|_\infty \leqslant \|u^0\|_\infty, \forall n \leqslant M$ où u^n désigne la solution approchée au temps $t_n = nk$?

(d) Donner un résultat de convergence pour ce schéma.

2. Euler implicite — Mêmes questions qu'en 1. en remplaçant Euler explicite par Euler implicite.

3. Crank Nicolson

(a) En s'inspirant du schéma de Crank-Nicolson (vu en cours) construire un schéma d'ordre 2 (espace et temps).

(b) Sous quelle(s) condition(s) sur k et h a-t'on $\|u^n\|_2 \leqslant \|u^0\|_2, \forall n \leqslant M$?

(c) Donner un résultat de convergence pour ce schéma.

4. Approximation décentrée amont — Dans les schémas trouvés aux questions 1., 2. et 3. on remplace l'approximation de $\partial_x u$ par une approximation décentrée amont (c'est-à-dire qu'on approche $u'(ih)$ par $\frac{u(ih) - u((i-1)h)}{h}$.

(a) Quel est l'ordre des schémas obtenus ?

(b) Sous quelle(s) condition(s) sur k et h a-t-on $\|u^n\|_\infty \leqslant \|u^0\|_\infty$ ou $\|u\|_2 \leqslant \|u^0\|_2, \forall n \leqslant M$?

(c) Donner un résultat de convergence pour ces schémas.

corrigé p.89

Exercice 31 — Équation de diffusion-réaction. Soit u_0 une fonction donnée de $[0\,;1]$ dans $\mathbb{R}$. On s'intéresse ici à la discrétisation du problème suivant :

$$\partial_t u(t,x) - \partial^2_{xx} u(t,x) - u(t,x) = 0 \quad t \in \mathbb{R}^+ \quad x \in [0\,;1] \tag{2.31}$$

$$u(t,0) = u(t,1) = 0,\ t \in \mathbb{R}^\star_+ \quad u(0,x) = u_0(x) \quad x \in [0\,;1]. \tag{2.32}$$

On note u la solution de (2.31), (2.32), et on suppose que u est la restriction à $\mathbb{R}_+ \times [0\,;1]$ d'une fonction de classe $\mathcal{C}^\infty$ de $\mathbb{R}^2$ dans $\mathbb{R}$.

Pour $h = \frac{1}{N+1}$ $(N \in \mathbb{N}^\star)$ et $k > 0$, on pose $x_i = ih, i \in \{0,\dots,N+1\}$, $t_n = nk, n \in \mathbb{N}$, $\overline{u}^n_i = u(x_i,t_n)$, et on note u^n_i la valeur approchée recherchée de $\overline{u}^n_i$.

On considère deux schémas numériques, (2.33)–(2.35) et (2.34)–(2.35) définis par les équations suivantes :

$$\frac{u^{n+1}_i - u^n_i}{k} - \frac{(u^{n+1}_{i+1} + u^{n+1}_{i-1} - 2u^{n+1}_i)}{h^2} - u^{n+1}_i = 0,\ n \in \mathbb{N},\ i \in \{1,\dots,N\}, \tag{2.33}$$

$$\frac{u^{n+1}_i - u^n_i}{k} - \frac{(u^{n+1}_{i+1} + u^{n+1}_{i-1} - 2u^{n+1}_i)}{h^2} - u^n_i = 0,\ n \in \mathbb{N},\ i \in \{1,\dots,N\}, \tag{2.34}$$

$$u^{n+1}_0 = u^{n+1}_{N+1} = 0,\ n \in \mathbb{N};\ u^0_i = u_0(x_i), =\ i \in \{0,\dots,N+1\}. \tag{2.35}$$

Pour $n \in N$, on note $u^n = (u^n_1,\dots,u^n_N) \in \mathbb{R}^N$.

1. Consistance. Soit $T > 0$. Pour $n \in \mathbb{N}$, et $i \in \{1,\dots,N\}$, on note R^n_i l'erreur de consistance (définie en cours) du schéma numérique (2.33), (2.35) [resp. du schéma numérique (2.34), (2.35)]. Montrer qu'il existe $C \in \mathbb{R}$, ne dépendant que de u et T, t. q. $|R^n_i| \leqslant C(k+h^2)$, pour tout $i \in \{1,\dots,N\}$ et tout $n \in \mathbb{N}$, t.q. $kn \leqslant T$.

2. Montrer que le schéma (2.33), (2.35) [resp. (2.34), (2.35)] demande, à chaque pas de temps, la résolution du système linéaire $Au^{n+1} = a$ [resp. $Bu^{n+1} = b$] avec $A, B \in \mathbb{R}^{N,N}$ et $a, b \in \mathbb{R}^N$ à déterminer.

Montrer que B est inversible (et même s.d.p.) pour tout $h > 0$ et $k > 0$. Montrer que A est inversible (et même s.d.p.) pour tout $h > 0$ et $k \in]0\,;1[$.

3. (Stabilité) Pour $n \in \mathbb{N}$, on pose $\|u^n\|_\infty = \sup_{i\in\{1,\dots,N\}} |u^n_i|$. Soit $T > 0$. On considère le schéma (2.34), (2.35). Montrer qu'il existe $C_1(T) \in \mathbb{R}$, ne dépendant que de T, t.q. $\|u^n\|_\infty \leqslant C_1(T)\|u_0\|_\infty$, pour tout $h > 0$, $k > 0$, et $n \in \mathbb{N}$ tel que $kn \leqslant T$.

Soit $\alpha \in [0\,;1]$. On considère le schéma (2.33), (2.35). Montrer qu'il existe $C_2(T,\alpha) \in \mathbb{R}$, ne dépendant que de T et de α, t.q. $\|u^n\|_\infty \leqslant C_2(T,\alpha)\|u_0\|_\infty$, pour tout $h > 0$, $k \in]0,\alpha[$, et $n \in \mathbb{N}$ tel que $kn \leqslant T$.

4. (Estimation d'erreur) Pour $n \in \mathbb{N}$ et $i \in \{1,\dots,N\}$, on pose $e^n_i = \overline{u}^n_i - u^n_i$. Soit $T > 0$. Donner, pour $kn \leqslant T$, des majorations de $\|e^n\|_\infty$ en fonction de T, C, $C_1(T)$, $C_2(T,\alpha)$ (définis dans les questions précédentes), k et h pour les deux schémas étudiés.

Exercice 32 — Discrétisation par volumes finis. Écrire une discrétisation espace-temps du problème (2.24) de l'exercice 26 avec le schéma de Crank-Nicolson en temps, et par volumes finis avec un maillage uniforme en espace, en utilisant un schéma décentré amont pour le terme d'ordre 1 $\partial_x u(x,t)$.

corrigé p.91

Exercice 33 — Schémas "saute-mouton" et Dufort-Frankel. On considère le problème suivant :

suggestions p.84

$$\begin{cases} \partial_t u(x,t) - \partial^2_{xx} u(x,t) = 0 & x \in]0\,;1[\quad t \in]0,T[, \\ u(0,t) = u(1,t) = 0 & t \in]0,T[, \\ u(x,0) = u_0(x) & x \in]0\,;1[. \end{cases} \tag{2.36}$$

Pour trouver une solution approchée de ((2.36)), on considère le schéma "saute-mouton" :

$$\begin{cases} \dfrac{u^{n+1}_j - u^{(n-1)}_j}{2k} = \dfrac{u^n_{j-1} - 2u^n_j + u^n_{j+1}}{h^2} & j = 1,\dots,N-1, \quad n = 1,\dots,M-1, \\ u^{n+1}_0 = u^{n+1}_{N+1} = 0 & n = 1,\dots,M-1, \end{cases} \tag{2.37}$$

où $(u^0_j)_{j=1,\dots,N}$ et $(u^1_j)_{j=1,\dots,N}$ sont supposés connus, $h = 1/N$, $k = T/M$.

1. Montrer que le schéma (2.37) est consistant. Quel est son ordre ?

2. Montrer que le schéma (2.37) est inconditionnellement instable au sens de Von Neumann.
On modifie “légèrement" le schéma (2.37) en prenant

$$\begin{cases} \dfrac{u_j^{n+1}-u_j^{(n-1)}}{2k} = \dfrac{u_{j-1}^n-(u_j^{n+1}+u_j^{(n-1)})+u_{j+1}^n}{h^2} & j=1,\dots,N \qquad n=1,\dots,M-1, \\ u_0^{n+1}=u_{N+1}^{n+1}=0 & n=1,\dots,M-1, \end{cases} \tag{2.38}$$

(schéma de Dufort-Frankel).

3. Montrer que le schéma (2.38) est consistant avec (2.36) quand $h,k\to 0$ sous la condition $\frac{k}{h}\to 0$.

4. Montrer que (2.38) est inconditionnellement stable.

Exercice 34 — Schéma de Gear. On considère le problème suivant :

$$\begin{cases} \partial_t u-\partial_{xx}^2 u=0 & \forall x\in\,]0\,;1[\quad \forall t\in]0,T[\\ u(x,0)=u_0(x) & \forall x\in\,]0\,;1[\\ u(0,t)=u(1,t)=0 & \forall t\in]0,T[\end{cases} \tag{2.39}$$

On suppose que $u_0\in\mathcal{C}(]0\,;1[,\mathbb{R})$. On rappelle que dans ce cas, il existe une unique fonction $u\in\mathcal{C}^2(]0\,;1[\times]0,T[,\mathbb{R})\cap\mathcal{C}([0\,;1]\times[0\,;T],\mathbb{R})$ qui vérifie (2.39). On cherche une approximation de la solution de ce problème, par une discrétisation par différences finies en espace et en temps. On se donne un ensemble de points $\{t_n\}$, $n=1,\dots,M$ de l'intervalle $]0,T[$, et un ensemble de points $\{x_i\}$, $i=1,\dots,N$. Pour simplifier, on considère un pas constant en temps et en espace. Soit $h=1/(N+1)$ le pas de discrétisation en espace, et $k=\frac{T}{M}$, le pas de discrétisation en temps. On pose alors $t_n=nk$ pour $n=0,\dots,M$ et $x_i=ih$ pour $i=0,\dots,N+1$. On cherche à calculer une solution approchée du problème (2.39) ; plus précisement, on cherche à déterminer des approximations $u_i^{(n)}$ de $u(x_i,t_n)$ pour $i=1,\dots,N$, et $n=1,\dots,M$.

On considère le schéma suivant :

$$\begin{cases} \dfrac{3u_i^{(n+1)}-4u_i^{(n)}+u_i^{(n-1)}}{2k}+\dfrac{2u_i^{(n+1)}-u_{i-1}^{(n+1)}-u_{i+1}^{(n+1)}}{h^2}=0 \qquad i=1,\dots N \quad n=1,\dots,M \\ u_i^0=u_0(x_i) \qquad i=1,\dots,N, \\ u_i^1=u_1(x_i) \qquad i=1,\dots,N, \\ u_0^{(n)}=u_{N+1}^{(n)}=0 \qquad \forall n=1,\dots,M, \end{cases}$$

où $u_1(x_i)=u(x_i,k)$ est supposée connue.

1. Montrer que ce schéma est consistant d'ordre 2 en temps et en espace.

2. Montrer que le schéma s'écrit sous forme matricielle :

$$U^{n+1}=BW^n$$

où

$$U^{n+1}=\begin{pmatrix} u_1^{n+1} \\ \vdots \\ u_N^{n+1} \end{pmatrix}, \quad B=(3\,\mathrm{Id}+\frac{2k}{h^2}A)^{-1}, \quad A=\begin{bmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 2 \end{bmatrix}$$

et W^n ne dépend que de $U^{n-1}=\begin{pmatrix} u_1^{n-1} & \cdots & u_N^{n-1} \end{pmatrix}^t$ et $U^n=\begin{pmatrix} u_1^n & \cdots & u_N^n \end{pmatrix}^t$. Donner l'expression de W^n en fonction de U^{n-1} et U^n.

3. En posant $V^n=\begin{pmatrix} U^n & U^{n-1} \end{pmatrix}^t\in\mathbb{R}^{2M}$, mettre le schéma sous la forme $V^{n+1}=MV^n$. Donner la matrice M en fonction de A.

4. Montrer que μ est valeur propre de M si et seulement si $\mu^2 - 4\beta\mu + \beta = 0$ où β est une valeur propre de la matrice B.

5. Montrer que les valeurs propres de la matrice M sont toutes de module strictement inférieur à 1.

6. Montrer qu'il existe $C \in \mathbb{R}$, qui ne dépend pas de n, tel que $|U^n|_2 \leqslant C$, où $|\cdot|_2$ désigne la norme euclidienne dans $\mathbb{R}^N$.

corrigé p.93

Exercice 35 — Problème parabolique non linéaire. On se propose, dans cet exercice, de montrer l'existence d'une solution faible au problème (2.40)-(2.42), à partir de l'existence de la solution approchée donnée par un schéma numérique. L'inconnue de ce problème est la fonction u de $[0\,;1] \times [0\,;T]$ dans $\mathbb{R}$, elle doit être solution des équations suivantes :

$$\partial_x u(x,t) - \partial^2_{xx}(\varphi(u))(x,t) = v(x,t) \quad x \in]0\,;1[\quad t \in]0,T[, \tag{2.40}$$

$$\partial_x(\varphi(u))(0,t) = \partial_x(\varphi(u))(1,t) = 0 \quad t \in]0,T[, \tag{2.41}$$

$$u(x,0) = u_0(x) \quad x \in]0\,;1[, \tag{2.42}$$

où φ, v, T, u_0 sont donnés et sont t.q.

- $T > 0$, $v \in L^\infty(]0\,;1[\times]0,T[)$,
- φ croissante, lipschitzienne de $\mathbb{R}$ dans $\mathbb{R}$,
- $u_0 \in L^\infty(]0\,;1[)$ et $\varphi(u_0)$ lipschitzienne de $[0\,;1]$ dans $\mathbb{R}$.

Un exemple important est donné par $\varphi(s) = \alpha_1 s$ si $s \leqslant 0$, $\varphi(s) = 0$ si $0 \leqslant s \leqslant L$ et$\varphi(s) = \alpha_2(s-L)$ si $s \geqslant L$, avec α_1, α_2 et L donnés dans $\mathbb{R}_+^\star$. Noter pour cet exemple que $\varphi' = 0$ sur $]0,L[$. Les ensembles $]0\,;1[$ et $D =]0\,;1[\times]0,T[$ sont munis de leur tribu borélienne et de la mesure de Lebesgue sur cette tribu.

On appelle "solution faible" de (2.40)-(2.42) une fonction u vérifiant :

$$u \in L^\infty(]0\,;1[\times]0,T[), \tag{2.43}$$

$$\int_D \big(u(x,t)\partial_t\psi(x,t) + \varphi(u(x,t))\partial^2_{xx}\psi(x,t) + v(x,t)\psi(x,t)\big)\,\mathrm{d}x\,\mathrm{d}t + \int_0^1 u_0(x)\psi(x,0)dx = 0, \forall \psi \in C_T^{2,1}(\mathbb{R}^2), \tag{2.44}$$

où $\psi \in C_T^{2,1}(\mathbb{R}^2)$ signifie que ψ est une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ deux fois continûment dérivable par rapport à x, une fois continûment dérivable par rapport à t et telle que

$$\partial_x\psi(0,t) = \partial_x\psi(1,t) = 0 \quad \forall t \in [0\,;T] \qquad \text{et} \qquad \psi(x,T) = 0 \quad \forall x \in [0\,;1] \tag{2.45}$$

1. Solution classique versus solution faible — On suppose, dans cette question seulement, que φ est de classe $\mathcal{C}^2$, v est continue sur $[0\,;1]\times[0\,;T]$ et u_0 est continue sur $[0\,;1]$. Soit $u \in \mathcal{C}^2(\mathbb{R}^2,\mathbb{R})$; on note encore u la restriction de u à$]0\,;1[\times]0,T[$. Montrer que u est solution de (2.43)-(2.44) si et seulement si u vérifie (2.40)-(2.42) au sens classique (c'est-à-dire pour tout $(x,t) \in [0\,;1]\times[0\,;T]$).

On cherche maintenant une solution approchée de (2.40)-(2.42).

Soient $N, M \in \mathbb{N}^\star$. On pose $h = \frac{1}{N}$ et $k = \frac{T}{M}$. On va construire une solution approchée de (2.40)-(2.42) à partir de la famille $\{u_i^n,\ i = 1,\dots,N,\ n = 0,\dots,M\}$ (dont on va prouver l'existence et l'unicité) vérifiant les équations suivantes :

$$u_i^0 = \frac{1}{h}\int_{(i-1)h}^{ih} u_0(x)dx,\ i = 1,\dots,N, \tag{2.46}$$

$$\frac{u_i^{n+1} - u_i^n}{k} - \frac{\varphi(u_{i-1}^{n+1}) - 2\varphi(u_i^{n+1}) + \varphi(u_{i+1}^{n+1})}{h^2} = v_i^n,\ i = 1,\dots,N,\ n = 0,\dots,M-1, \tag{2.47}$$

avec $u_0^{n+1} = u_1^{n+1}$, $u_{N+1}^{n+1} = u_N^{n+1}$, pour tout $n = 0,\dots,M-1$ et, pour tout $i = 1,\dots,N$, pour tout $n = 0,\dots,M$,

$$v_i^n = \frac{1}{kh}\int_{nk}^{(n+1)k}\int_{(i-1)h}^{ih} v(x,t)\,\mathrm{d}x\,\mathrm{d}t.$$

2. Existence et unicité de la solution approchée. Soit $n \in \{0, \dots, M-1\}$. On suppose connue la famille $\{u_i^n,\ i = 1, \dots, N\}$. On va prouver dans cette question l'existence et l'unicité de la famille $\{u_i^{n+1},\ i = 1, \dots, N\}$ vérifiant (2.47) (avec $u_0^{n+1} = u_1^{n+1}$, $u_{N+1}^{n+1} = u_N^{n+1}$).

(a) Soit $a > 0$, pour $s \in \mathbb{R}$, on pose $g_a(s) = s + a\varphi(s)$. Montrer que g_a est une application strictement croissante bijective de $\mathbb{R}$ dans $\mathbb{R}$.

(b) Soit $\overline{w} = (\overline{w}_i)_{i=1,\dots,N} \in \mathbb{R}^N$; on pose $\overline{w}_0 = \overline{w}_1$ et $\overline{w}_{N+1} = \overline{w}_N$. Montrer qu'il existe un et un seul couple $(u, w) \in \mathbb{R}^N \times \mathbb{R}^N$, $u = (u_i)_{i=1,\dots,N}$, $w = (w_i)_{i=1,\dots,N}$, tel que :

$$\varphi(u_i) = w_i, \text{ pour tout } i \in \{1, \dots, N\}, \tag{2.48}$$

$$u_i + \frac{2k}{h^2} w_i = \frac{k}{h^2}(\overline{w}_{i-1} + \overline{w}_{i+1}) + u_i^n + k v_i^n, \text{ pour tout } i = 1, \dots, N. \tag{2.49}$$

On peut donc définir une application F de $\mathbb{R}^N$ dans $\mathbb{R}^N$ par $\overline{w} \mapsto F(\overline{w}) = w$ où w est solution de (2.48)–(2.49).

(c) On munit $\mathbb{R}^N$ de la norme usuelle $\| \cdot \|_\infty$. Montrer que l'application F est strictement contractante. [On pourra utiliser la monotonie de φ et remarquer que, si $a = \varphi(\alpha)$ et $b = \varphi(\beta)$, on a $|\alpha - \beta| \geqslant (1/L)|a - b|$, où L ne dépend que de φ.]

(d) Soit $\{u_i^{n+1},\ i = 1, \dots, N\}$ solution de (2.47). On pose $w = (w_i)_{i=1,\dots,N}$, avec $w_i = \varphi(u_i^{n+1})$ pour $i \in \{1, \dots, N\}$. Montrer que $w = F(w)$.

(e) Soit $w = (w_i)_{i=1,\dots,N}$ t.q. $w = F(w)$. Montrer que pour tout $i \in \{1, \dots, N\}$ il existe $u_i^{n+1} \in \mathbb{R}$ tel que $w_i = \varphi(u_i^{n+1})$. Montrer que $\{u_i^{n+1},\ i = 1, \dots, N\}$ est solution de (2.47).

(f) Montrer qu'il existe une unique famille $\{u_i^{n+1},\ i = 1, \dots, N\}$ solution de (2.47).

3. Estimation $L^\infty(]0\,;1[\times]0, T[)$ sur u. On pose $A = \|u_0\|_{L^\infty(]0\,;1[)}$ et $B = \|v\|_{L^\infty(]0\,;1[\times]0,T[)}$. Montrer, par récurrence sur n, que $u_i^n \in [-A - nkB, A + nkB]$ pour tout $i = 1, \dots, N$ et tout $n = 0, \dots, M$. [On pourra, par exemple, considérer (2.47) avec i t.q. $u_i^{n+1} = \min\{u_j^{n+1},\ j = 1, \dots, N\}$.]

En déduire qu'il existe $c_{u_0,v,T} \in \mathbb{R}_+$ t.q. $\|u^n\|_{L^\infty(]0\,;1[)} \leqslant c_{u_0,v,T}$.

4. Estimation de la dérivée par rapport à x de $\varphi(u)$. Montrer qu'il existe C_1 (ne dépendant que de T, φ, v et u_0) t.q., pour tout $n = 0, \dots, M-1$,

$$\sum_{n=0}^{M-1} \sum_{i=1}^{N-1} (\varphi(u_{i+1}^{n+1}) - \varphi(u_i^{n+1}))^2 \leqslant C_1 \frac{h}{k}. \tag{2.50}$$

[Multiplier (2.47) par u_i^{n+1} et sommer sur i et sur n et utiliser l'inégalité $a^2 - ab \geqslant \frac{a^2}{2} - \frac{b^2}{2}$.]

5. Estimation de la dérivée par rapport à t de $\varphi(u)$. Montrer qu'il existe C_2 (ne dépendant que de T, φ, v et u_0) t.q.

$$\sum_{n=0}^{M-1} h \sum_{i=0}^{N+1} (\varphi(u_{i+1}^n) - \varphi(u_{i+1}^{n+1}))^2 \leqslant C_2 k. \tag{2.51}$$

et

$$\sum_{i=0}^{N+1} (\varphi(u_i^{n+1}) - \varphi(u_{i+1}^{n+1}))^2 \leqslant C_2 h, \text{ pour tout } n \in \{0, \dots, M\}. \tag{2.52}$$

[indication : multiplier (2.47) par $\varphi(u_i^{n+1}) - \varphi(u_i^n)$ et sommer sur i et n]

Dans la suite de l'exercice, il s'agit de passer à la limite (quand $N, M \to \infty$) pour trouver une solution de (2.40)-(2.42).

Pour $M \in \mathbb{N}^\star$ donné, on prend $N = M^2$ (et donc h et k sont donnés et $k = T\sqrt{h}$), on définit (avec les u_i^n trouvés dans les questions précédentes) une fonction, u_h, sur $[0\,;1] \times [0\,;T]$ en posant

$$u_h(x,t) = \frac{t - nk}{k} u_h^{(n+1)}(x) + \frac{(n+1)k - t}{k} u_h^{(n)}(x), \text{ si } t \in [nk, (n+1)k]$$

et

$$u_h^{(n)}(x) = u_i^n, \text{ si } x \in](i-1)h, ih[,\ i = 1, \dots, N,\ n = 0, \dots, M.$$

Enfin, on définit $\varphi(u_h)$ par $\varphi(u_h)(x,t) = \varphi(u_h(x,t))$.

6. Montrer que les suites $(u_h)_{M\in\mathbb{N}^\star}$ et $(\varphi(u_h))_{M\in\mathbb{N}^\star}$ sont bornées dans $L^\infty(]0\,;1[\times]0,T[)$ (on rappelle que h est donné par M).

7. Montrer qu'il existe C (ne dépendant que de T, φ, v et u_0) tel que l'on ait, pour tout $M \in \mathbb{N}^\star$:

(a) Pour tout $t \in [0\,;T]$,

$$\int_{\mathbb{R}} |\varphi(u_h)(x+\eta,t) - \varphi(u_h)(x,t)|^2 \,\mathrm{d}x \leqslant C\eta,$$

pour tout $\eta \in \mathbb{R}_+^\star$, avec $\varphi(u_h)(\,\cdot\,,t)$ prolongée par 0 hors de $[0\,;1]$.

(b) $\|\varphi(u_h)(\,\cdot\,,t) - \varphi(u_h)(\,\cdot\,,s)\|_{\mathcal{L}^2(]0\,;1[)} \leqslant C|t-s|$, pour tout $t, s \in [0\,;T]$.

Une conséquence des questions 6 et 7 (que l'on admet ici est que l'on peut trouver une suite $(h_n)_{n\in\mathbb{N}}$ et $u \in L^\infty(]0\,;1[\times]0,T[)$ telle que, en posant $u_n = u_{h_n}$ (on rappelle que $k_n = T\sqrt{h_n}$), l'on ait, quand $n \to \infty$,

(a) $h_n \to 0$ et $k_n \to 0$,

(b) $u_n \to u$ dans $L^\infty(]0\,;1[\times]0,T[)$ pour la topologie faible-$\star$,

(c) $\varphi(u_n) \to \varphi(u)$ dans $L^p(]0\,;1[\times]0,T[)$, pour tout $p \in [1,\infty[$.

8. Montrer que la fonction u ainsi trouvée est solution de (2.43),(2.44).

Remarque : On peut aussi montrer l'unicité de la solution de (2.43),(2.44).

2.4.2 Suggestions

Exercice 27 — Exemple de schéma non convergent.

1. Ecrire le schéma d'Euler explicite.
2. Démontrer par récurrence que

$$\text{Si } n \in \{0,\dots,M+1\},\ i \in \left\{-\frac{N+1}{2},\dots,\frac{N+1}{2}\right\} \text{ et } \quad i \geqslant -\frac{N+1}{4} + n \text{ alors } u_i^n = 0.$$

En déduire que $u_i^n = 0$ pour $n \in \{0,\dots,M+1\}$ et $i \in \left\{0,\dots,\frac{N+1}{2}\right\}$ et conclure.

Exercice 30 — Discrétisation d'un problème parabolique.

1. Calculer l'erreur de consistance et la majorer par des développements de Taylor.
Chercher ensuite les conditions pour que :

$$\|u^n\|_\infty \leqslant \|u^0\|_\infty.$$

Pour étudier la convergence du schéma, majorer l'erreur de discrétisation : $e_j^n = \bar{u}_j^n - u_j^n$ où u_j^n est solution du schéma, et $\bar{u}_j^n$ est la solution du problème (2.30) en $x_j = jh$ et $t_n = nk$.

Même chose pour les questions suivantes...

Exercice 31 — Équation de diffusion-réaction.

1. Effectuer des développements de Taylor...

3. Montrer par récurrence que $\max_{j=1,\dots,N} u_j^n \leqslant (1+k)^n \max_{j=1,\dots,N} u_j^0$. et que $\min_{j=1,\dots,N} u_j^{(n)} \geqslant (1+k)^n \min_{j=1,\dots,N} u_j^{(0)}$.

4. Utiliser l'équation, le schéma, et l'erreur de consistance.

Exercice 33 — Schémas "saute-mouton" et Dufort-Frankel.

1. Effectuer des développements de Taylor pour majorer l'erreur de consistance.

2. Montrer que le facteur d'amplification ξ_n obtenu par l'analyse de stabilité de Von Neumann satisfait :

$$\xi_{n+1} - \alpha\xi_n - \xi_{n-1} = 0,\ n \geqslant 2.$$

Etudier ensuite les racines de l'équation $r^2 - \alpha r - 1 = 0$ et montrer que l'une de ses racines est, en module, supérieure à 1.

4. Reprendre la méthode développée à la question 2, en montrant que l'équation caractéristique pour ξ est maintenant :

$$p(r) = ar^2 + br + c = 0,$$

avec

$$a = \frac{1}{2k} + \frac{1}{h^2},\ b = -\frac{2\cos(ph)}{h^2} \text{ et } c = \frac{1}{h^2} - \frac{1}{2h}.$$

Etudier ensuite les racines de cette équation.

2.4.3 Corrigés

Exercice 25 — Existence de solutions "presque classiques". On note $\|\cdot\|_2 = \|\cdot\|_{L^2(]0\,;1[)}$.

1. Pour $n \in \mathbb{N}^*$, on a

$$\int_0^1 \sin^2(n\pi x)\,\mathrm{d}x = \int_0^1 \frac{1-\cos(2n\pi x)}{2}\,\mathrm{d}x = \frac{1}{2},$$

et

$$\int_0^1 |u_0(x)\sin(n\pi x)|\,\mathrm{d}x \leqslant \|u_0\|_2 \left(\int_0^1 \sin^2(n\pi x)dx\right)^{1/2} = \frac{\sqrt{2}}{2}\|u_0\|_2.$$

La quantité a_n est donc bien définie et $|a_n| \leqslant \sqrt{2}\|u_0\|_2$. Pour tout $t > 0$ et $x \in [0,1]$, on a

$$|e^{-n^2\pi^2t^2} a_n \sin(n\pi x)| \leqslant \sqrt{2}\|u_0\|_2 e^{-n^2\pi^2t^2} \qquad \forall n \in \mathbb{N}^*.$$

Ceci montre que la série

$$\sum_{n>0} e^{-n^2\pi^2t^2} a_n \sin(n\pi x)$$

est absolument convergente et donc que u est bien définie pour tout $t > 0$ et tout $x \in [0\,;1]$ et même pour tout $x \in \mathbb{R}$. On remarque ensuite que u est de classe $\mathcal{C}^\infty$ sur $\mathbb{R} \times \mathbb{R}_+^*$, en appliquant les théorèmes classiques de dérivation terme à terme d'une série. En effet, soit $\varepsilon > 0$, pour tout $x \in \mathbb{R}$ et $t > \varepsilon$ on a

$$\left|e^{-n^2\pi^2t^2} a_n \sin(n\pi x)\right| \leqslant \sqrt{2}\|u_0\|_2 e^{-n^2\pi^2\varepsilon^2}, \forall n \in \mathbb{N}^*$$

Comme $(x,t) \to e^{-n^2\pi^2t^2} a_n \sin(n\pi t)$ est continue (pour tout $n \in \mathbb{N}^*$), on en déduit que u est continue sur $\mathbb{R}\times]\varepsilon,\infty[$, et finalement sur $\mathbb{R}\times]0,\infty[$ car $\varepsilon > 0$ est arbitraire. Pour dériver terme à terme la série définissant u, il suffit également d'obtenir sur $]\varepsilon,\infty[\times\mathbb{R}$ (pour tout $\varepsilon > 0$) une majoration du terme général de la série des dérivées par le terme général d'une série convergente (indépendant de $(x,t) \in \mathbb{R}\times]\varepsilon,\infty[$. On obtient cette majoration en remarquant que, pour $(x,t) \in \mathbb{R}\times]\varepsilon,\infty[$,

$$|-n^2\pi^2 e^{-n^2\pi^2t^2} a_n \sin(n\pi x)| \leqslant n^2\pi^2 e^{-n^2\pi^2\varepsilon^2}\sqrt{2}\|u_0\|_2$$

On montre ainsi finalement que u est de classe C^1 par rapport à t et que

$$\partial_t u(x,t) = \sum_{n>0} -n^2\pi^2 e^{-n^2\pi^2t^2} a_n \sin(n\pi x), x \in \mathbb{R}, t > 0.$$

En itérant ce raisonnement on montre que u est de classe $\mathcal{C}^\infty$ par rapport à t sur $\mathbb{R}\times\mathbb{R}_+^*$. Un raisonnement similaire montre que u est de classe $\mathcal{C}^\infty$ par rapport à x sur $\mathbb{R}\times\mathbb{R}_+^*$ et que l'on peut dériver terme à terme la série définissant u. On obtient donc aussi

$$\partial_{xx}^2 u(xt) = \sum_{n>0} -n^2\pi^2 e^{-n^2\pi^2t^2} a_n \sin(n\pi x), x\in\mathbb{R}, t>0,$$

et ceci donne $\partial_t u = \partial_{xx}^2 u$ sur $\mathbb{R}\times\mathbb{R}_t^*$ et donc aussi un $[0\,;1]\times\mathbb{R}_+^*$. Le fait que $u(0,t)=u(1,t)$ pour tout $t>0$ est immédiat car $\sin n\pi t = \sin 0 = 0$, pour tout $n\in\mathbb{N}^*$. Il reste à montrer que $u(\,\cdot\,,t)\to u_0$ dans $L^2(]0\,;1[)$ quand $t\to 0$. On définit $e_n\in L^2(]0\,;1[)$ par $e_n(x)=\sqrt{2}\sin(n\pi x)$. La famille $\{e_n, n\in\mathbb{N}^*\}$ est une base hilbertienne de $L^2(]0\,;1[)$. On a donc :

$$\lim_{N\to\infty}\left(\sum_{n=1}^N a_n \sin n\pi x\right) = u_0 \text{ dans } L^2(]0\,;1[) \quad \text{et} \quad \sum_{n=1}^\infty a_n^2 = 2\|u_0\|_2^2.$$

On remarque maintenant que

$$u(x,t)-u_0(x) = u(x,t)-u^{(N)}(x,t)+u^{(N)}(xt)-u_0^{(N)}(x)-u_0^{(N)}(x)-u_0(x),$$

avec

$$u^{(N)}(x,t)=\sum_{n=1}^N a_n e^{-n^2\pi^2t^2}\sin(n\pi x) \quad \text{et} \quad u_0^{(N)}(x)=\sum_{n=1}^N a_n\sin(n\pi x).$$

Il est clair que, pour tout $N\in\mathbb{N}^*$, on a $u^{(N)}(\,\cdot\,,t)\to u_0^{(N)}$ uniformément sur $\mathbb{R}$, quand $t\to 0$, et donc $u^{(N)}(\,\cdot\,,t)\to u_0^{(N)}$ dans $L^2(]0\,;1[)$ quand $t\to 0$. Comme la famille $(\sin n\pi x)_{n\in\mathbb{N}^*}$ est une base hilbertienne de $L^2(]0,1[)$, on a $\int_0^1 \sin n\pi x\ \sin m\pi x\ \mathrm{d}x = 0$ si $n\neq m$; on rappelle que $\int_0^1 \sin^2 n\pi x\,\mathrm{d}x = \frac{1}{2}$, on a donc

$$\|u(\,\cdot\,,t)-u^{(N)}(\,\cdot\,,t)\|_2^2 = \frac{1}{2}\sum_{n=N+1}^\infty a_n^2 e^{-2n^2\pi^2t^2} \leqslant \frac{1}{2}\sum_{n=N+1}^\infty a_n^2 = \|u_0^{(N)}-u_0\|_2^2 \to 0 \text{ lorsque } N\to\infty,$$

on en déduit que $u(\,\cdot\,,t)\to u_0$ quand $t\to 0$ dans $L^2(]0\,;1[)$.

2. On note w la différence de deux solutions de (2.23). On a donc

$$\begin{array}{ll} w\in\mathcal{C}^\infty([0\,;1]\times\mathbb{R}_+^*,\mathbb{R}), & \\ w_t - w_{xx} = 0 & \text{sur } [0\,;1]\times\mathbb{R}_+^*, \\ w(0,t)=w(1,t)=0 & \text{pour } t>0, \\ w(.,t)\to 0 & \text{dans } L^2(]0\,;1[) \quad \text{quand } t\to 0. \end{array}$$

Soit $0<\varepsilon<T<\infty$. On intègre l'équation $ww_t - ww_{xx}=0$ sur $]0\,;1[\times]\varepsilon,T[$. En utilisant une intégration par parties (noter que $w\in\mathcal{C}^\infty([0\,;1]\times[\varepsilon,T])$, on obtient :

$$\frac{1}{2}\int_0^1 w^2(x,T)\,\mathrm{d}x - \frac{1}{2}\int_0^1 w^2(x,\varepsilon)\,\mathrm{d}x + \int_0^1\int_\varepsilon^T w_x^2(x,t)\,\mathrm{d}x\mathrm{d}t = 0.$$

D'où l'on déduit $\|w(.,T)\|_2 \leqslant \|w(.,\varepsilon)\|_2$. Comme $w(.,t)\to 0$ dans $L^2(]0\,;1[)$ quand $t\to 0$, on a $\|w(.,\varepsilon)\|_2\to 0$ lorsque $\varepsilon\to 0$ et donc $\|w(.,T)\|_2=0$. Comme $T>0$ est arbitraire, on a finalement $w(x,t)=0$, $\forall t\in[0\,;1]$, ce qui montre bien l'unicité de la solution de (2.23).

Exercice 27 — Exemple de schéma non convergent.

1. La formule pour calculer u_i^0 est :

$$u_1^0 = u_0(ih,0), \quad i=-\frac{N+1}{2},\dots,\frac{N+1}{2}$$

Soit maintenant $n\in\{0,\dots,M\}$. On a :

$$\begin{array}{ll} u_i^{n+1}=0 & i=-\dfrac{N+1}{2} \qquad i=\dfrac{N+1}{2} \\ u_i^{n+1}=u_i^n+\dfrac{k}{h^2}\left(u_{i+1}^n+u_{i-1}^n-2u_i^n\right), & i=-\dfrac{N+1}{2}+1,\dots,\dfrac{N+1}{2}-1. \end{array}$$

2. On va montrer, par récurrence (finie) sur n, que :

$$\text{Si } n \in \{0,\dots,M+1\},\ i \in \left\{-\frac{N+1}{2},\dots,\frac{N+1}{2}\right\} \text{ et } \quad i \geqslant -\frac{N+1}{4}+n \text{ alors } u_i^n = 0. \quad (2.53)$$

Pour initialiser la récurrence, on suppose que $n = 0$ et $i \geqslant -\dfrac{N+1}{4}$. On a alors

$$ih \geqslant -\frac{N+1}{4} \quad \frac{8}{N+1} = -2 > -3$$

et donc $u_i^0 = 0$. Soit maintenant $n \in \{0,\dots,M\}$. On suppose que l'hypothèse de récurrence est vérifiée jusqu'au rang n, et on démontre la propriété au rang $n+1$. Soit donc $i \in \left\{-\frac{N+1}{2},\dots,\frac{N+1}{2}\right\}$ tel que $i \geqslant -\frac{N+1}{4} + (n+1)$. Alors :

— Si $i = \frac{N+1}{2}$ on a bien $u_i^{N+1} = 0$.

— Si $i < \frac{N+1}{2}$, les indices $i-1$, i et $i+1$ sont tous supérieurs ou égaux à $-\frac{N+1}{4} + n$, et donc par hypothèse de récurrence,

$$u_i^{n+1} = u_i^n\left(1 - \frac{2k}{h^2}\right) + \frac{k}{h^2}u_{i+1}^n + \frac{k}{h^2}u_{i-1}^n = 0.$$

On a donc bien démontré (2.53). On utilise maintenant l'hypothèse $k = h$, c'est–à–dire $\frac{1}{M+1} = \frac{8}{N+1}$. On a alors

$$-\frac{N+1}{4} + M + 1 = -2(M+1) + M + 1 = -(M+1) < 0.$$

On en déduit que si $n \in \{0,\dots,M+1\}$ et $i \geqslant 0$, alors $i \geqslant -\frac{N+1}{4} + n$. On en déduit que $u_i^n = 0$ pour $n \in \{0,\dots,M+1\}$ et $i \in \left\{0,\dots,\frac{N+1}{2}\right\}$. On remarque alors que

$$\begin{aligned}\max\left\{|u_i^{M+1} - \bar{u}_i^{M+1}|, i \in \left\{-\frac{N+1}{2},\dots,\frac{N+1}{2}\right\}\right\} &\geqslant \max\left\{|\bar{u}_i^{M+1}|, i \in \left\{0,\dots,\frac{N+1}{2}\right\}\right\} \\ &\geqslant \inf_{[0,4]} u(x,1) > 0,\end{aligned}$$

et donc ne tend pas vers 0 quand $h \to 0$.

Exercice 28 — Schémas explicites centré et décentré.

1. Schéma explicite décentré

(a) Propriétés de la solution de (2.26).

i. On multiplie la première équation de (2.26) par $\varphi(u)$ et on intègre sur $[0\,;1]\times[0\,;T]$, avec $\varphi \in C^1(\mathbb{R},\mathbb{R})$, $\varphi(s) = 0$ si $s \leqslant B$ et $\varphi'(s) > 0$ si $s > B$]. On en déduit que $A \leqslant u(x,t) \leqslant B$ pour tout $(x,t) \in [0\,;1]\times[0\,;T]$ et que que $\|u(.,t)\|_{L^\infty(]0\,;1[)} \leqslant \|u_0\|_{L^\infty(]0\,;1[)}$ pour tout $t \in [0\,;T]$.

ii. Il faut montrer que $\|u(.,t)\|_{L^2(]0\,;1[)} \leqslant \|u_0\|_{L^2(]0\,;1[)}$ pour tout $t \in [0\,;T]$.

(b) Par définition, l'erreur de consistance en (x_i,t_n) s'écrit : On s'intéresse ici à l'ordre du schéma au sens des différences finies. On suppose que $u \in C^4([0\,;1]\times[0\,;T])$ est solution de (2.26) et on pose

$$\bar{u}_i^n = u(ih,nk),\ i = 0,\dots,N,\ k = 0,\dots,M.$$

Pour $i = 1,\dots,N-1$ et $k = 1,\dots,M-1$, l'erreur de consistance en (x_i,t_k) est définie par :

$$R_i^n = \frac{1}{k}(\bar{u}_i^{n+1} - \bar{u}_i^n) - \frac{\alpha}{h}(\bar{u}_i^n - \bar{u}_{i-1}^n) - \frac{\mu}{h^2}(\bar{u}_{i-1}^n - 2\bar{u}_i^n + \bar{u}_{i+1}^n). \quad (2.54)$$

Soit $i \in \{1,\dots,N-1\}$, $k \in \{1,\dots,M-1\}$. On cherche une majoration de R_i^n en utilisant des développements de Taylor. En utilisant ces développements, on obtient qu'il existe

$(\xi_\ell, t_\ell) \in [0\,;1] \times [0\,;T]$, $\ell = 1, \ldots, 4$, t.q. :

$$\bar{u}_i^{n+1} = \bar{u}_i^n + k\partial_t u(ih, nk) + \frac{k^2}{2} u_{tt}(\xi_1, t_1), \tag{2.55}$$

$$\bar{u}_{i-1}^n = \bar{u}_i^n - h\partial_x u(ih, nk) + \frac{h^2}{2}\partial_{xx}^2 u(\xi_2, t_2), \tag{2.56}$$

$$\bar{u}_{i-1}^n = \bar{u}_i^n - h\partial_x u(ih, nk) + \frac{h^2}{2}\partial_{xx}^2 u(ih, nk) - \frac{h^3}{6}\partial_{xxx} u(ih, nk) - \frac{h^4}{24}\partial_{xxxx} u(\xi_3, t_3), \tag{2.57}$$

$$\bar{u}_{i+1}^n = \bar{u}_i^n + h\partial_x u(ih, nk) + \frac{h^2}{2}\partial_{xx}^2 u(ih, nk) + \frac{h^3}{6}\partial_{xxx} u(ih, nk) + \frac{h^4}{24}\partial_{xxxx} u(\xi_4, t_4). \tag{2.58}$$

On en déduit :

$$\begin{aligned} R_i^n =& \partial_t u(ih, nk) + \frac{k}{2} u_{tt}(\xi_1, t_1) + \alpha\partial_x u(ih, nk) + \alpha\frac{h}{2}\partial_{xx}^2 u(\xi_2, t_2) \\ & - \mu\partial_{xx}^2 u(ih, nk) - \mu\frac{h^2}{24}\left(\partial_{xxxx} u(\xi_3, t_3) + \mu\partial_{xxxx} u(\xi_4, t_4)\right) \end{aligned}$$

et donc, comme u est solution de (2.26), pour h assez petit, on a $|R_i^n| \leqslant C_1(h + k)$ où C_1 ne dépend que de u. Le schéma (2.27) est donc consistant d'ordre 1 en temps et en espace.

(c) Cherchons les conditions pour que u_i^{n+1} s'écrive comme combinaison convexe de u_i^n, u_{i-1}^n et u_{i+1}^n. On peut réécrire le schéma (2.27) :

$$u_i^{n+1} = au_i^n + bu_{i+1}^n + cu_{i-1}^n, \text{ avec } a = 1 - \frac{\alpha k}{h} - \frac{2\mu k}{h^2}, b = \frac{\mu k}{h^2} \text{ et } c = \frac{\alpha k}{h} + \frac{\mu k}{h^2}.$$

Il est facile de voir que $a + b + c = 1$, et que $b \geqslant 0$, $c \geqslant 0$. Il reste à vérifier que $a \geqslant 0$; pour cela, il faut et il suffit que $\frac{\alpha k}{h} + \frac{2\mu k}{h^2} \leqslant 1$. Cette condition sécrit encore :

$$k \leqslant \frac{h^2}{\alpha h + 2\mu}. \tag{2.59}$$

Si h et k vérifient la condition (2.59), on pose : $M^n = \max_{i=1\ldots N} u_i^n$ (resp. $m^n = \min_{i=1\ldots N} u_i^n$. Comme u_i^{n+1} est une combinaison convexe de u_i^n, u_{i-1}^n et u_{i+1}^n, on a alors : $u_i^{n+1} \leqslant M^n$, $i = 1, \ldots, N$ (resp. $u_i^{n+1} \geqslant m^n$, $i = 1, \ldots, N$) et donc : $M^{n+1} \leqslant M^n$ (resp. $m^{n+1} \geqslant m^n$). On a ainsi montré que :

$$\|u^{n+1}\|_\infty \leqslant \|u^n\|_\infty.$$

On a de même :

$$\|u^n\|_\infty \leqslant \|u^{n-1}\|_\infty.$$

$$\vdots$$

$$\|u^1\|_\infty \leqslant \|u^0\|_\infty.$$

En sommant ces inégalités, on obtient :

$$\|u^n\|_\infty \leqslant \|u^0\|_\infty.$$

Donc, sous la condition (2.59), on a $\|u^{n+1}\|_\infty \leqslant \|u^n\|_\infty$ et donc $\|u^n\|_\infty \leqslant \|u^0\|_\infty$, pour tout $n = 1, \ldots, N$.

(d) En retranchant l'égalité (2.54) au schéma (2.27), on obtient l'équation suivante sur e_i^n :

$$\frac{1}{k}(e_i^{n+1} - e_i^n) + \frac{\alpha}{h}(e_i^n - e_{i-1}^n) - \frac{\mu}{h^2}(e_{i-1}^n - 2e_i^n + e_{i+1}^n) = R_i^n.$$

ce qu'on peut encore écrire :

$$e_i^{n+1} = (1 - \frac{k\alpha}{h} - 2\frac{k\mu}{h^2})e_i^n + e_{i-1}^n\frac{k\mu}{h^2} + kR_i^n.$$

Sous la condition de stabilité (2.59), on obtient donc :

$$\begin{array}{rcccl} |e_i^{n+1}| & \leqslant & \|e^{n+1}\|_\infty & + & C_1(k+h)k, \\ |e_i^{n}| & \leqslant & \|e^{n-1}\|_\infty & + & C_1(k+h)k, \\ \vdots & \leqslant & \vdots & + & \vdots \\ |e_i^{1}| & \leqslant & \|e^{0}\|_\infty & + & C_1(k+h)k, \end{array}$$

Si à $t=0$, on a $\|e^0\|=0$, alors on éduit des inégalités précédentes que $|e_i^n| \leqslant C1T(k+h)$ pour tout $n \in \mathbb{N}$. Le schéma est donc convergent d'ordre 1.

2. Schéma explicite centré.

(a) Consistance. En utilisant les développements de Taylor (2.55) (2.57) et (2.58), et les développements suivants :

$$\bar{u}_{i-1}^n = \bar{u}_i^n - h\partial_x u(ih,nk) + \frac{h^2}{2}\partial_{xx}^2 u(ih,nk) - \frac{h^3}{6}\partial_{xxx}u(\xi_5,t_5),$$

$$\bar{u}_{i+1}^n = \bar{u}_i^n + h\partial_x u(ih,nk) + \frac{h^2}{2}\partial_{xx}^2 u(ih,nk) + \frac{h^3}{6}\partial_{xxx}u(\xi_6,t_6),$$

on obtient maintenant :

$$\begin{aligned} R_i^n &= \partial_t u(ih,nk) + \frac{k}{2}u_{tt}(\xi_1,t_1) + \alpha\partial_x u(ih,nk) + \alpha\frac{h^2}{12}\left(\partial_{xxx}u(\xi_5,t_5) + \mu\partial_{xxx}u(\xi_6,t_6)\right) \\ &\quad -\mu\partial_{xx}^2 u(ih,nk) - \mu\frac{h^2}{24}\left(\partial_{xxxx}u(\xi_3,t_3) + \mu\partial_{xxxx}u(\xi_4,t_4)\right), \end{aligned}$$

On en déduit que

$$|R_i^n| \leqslant C_3(k+h^2),$$

où $C_3 = \max(\frac{1}{2}\|u_{tt}\|_\infty, \frac{1}{6}\|\partial_{xxx}u\|_\infty, \frac{1}{12}\|\partial_{xxxx}u\|_\infty)$.

(b) Le schéma s'écrit maintenant :

$$u_i^{n+1} = \widetilde{a}u_i^n + \widetilde{b}u_{i+1}^n + \widetilde{c}u_{i-1}^n, \text{ avec } \widetilde{a} = 1 - \frac{2\mu k}{h^2}, \widetilde{b} = \frac{\mu k}{h^2} - \frac{\alpha k}{h} \text{ et } \widetilde{c} = \frac{\mu k}{h^2} + \frac{\alpha k}{h}.$$

Remarquons que l'on a bien : $\widetilde{a}+\widetilde{b}+\widetilde{c}=1$. Pour que u_i^{n+1} soit combinaison convexe de u_i^n, u_{i+1}^n et u_{i-1}^n, il faut et il suffit donc que $\widetilde{a} \geqslant 0$, $\widetilde{b} \geqslant 0$, et $\widetilde{c} \leqslant 0$. L'inégalité $\widetilde{c} \geqslant 0$ est toujours vérifiée. Les deux conditions qui doivent être vérifiées par h et k s'écrivent donc :

i. $\widetilde{a} \geqslant 0$, i.e. $1 - \frac{2\mu k}{h^2} \geqslant 0$, soit encore

$$k \leqslant \frac{h^2}{2\mu}.$$

ii. $\widetilde{b} \geqslant 0$ i.e. $\frac{\mu k}{h^2} - \frac{\alpha k}{h} \geqslant 0$, soit encore

$$h \leqslant \frac{\mu}{2\alpha}.$$

Le schéma centré est donc stable sous les deux conditions suivantes :

$$h \leqslant \frac{\mu}{2\alpha} \text{ et } k \leqslant \frac{1}{2\mu}h^2. \tag{2.60}$$

Pour obtenir une borne d'erreur, on procède comme pour le schéma (2.27) : on soustrait la définition de l'erreur de consistance au schéma numérique, et on obtient :

$$e_i^{n+1} = \widetilde{a}e_i^n + \widetilde{b}e_{i+1}^n + \widetilde{c}e_{i-1}^n + kR_i^n.$$

Par le même raisonnement que pour le schéma décentré, on obtient donc que si $e_i^0 = 0$, on a $|e_i^n| \leqslant C_4(k+h^2)$, avec $C_4 = TC_3$.

Exercice 31 — Équation de diffusion-réaction.

1. Notons $R_i^{(n)}$ l'erreur de consistance en (x_i, t_n). Pour le schéma (2.33), on a donc par définition :

$$R_i^{(n)} = \frac{\bar{u}_i^{(n+1)} - \bar{u}_i^{(n)}}{k} + \frac{1}{h^2}(2\bar{u}_i^{(n+1)} - \bar{u}_{i-1}^{(n+1)} - \bar{u}_{i+1}^{(n+1)}) - \bar{u}_i^{n+1} = \widetilde{R}_i^{(n)} + \hat{R}_i^n$$

où

$$\widetilde{R}_i^n = \frac{\bar{u}_i^{n+1} - \bar{u}_i^n}{k} - \partial_t u(x_i, t_n)$$

est l'erreur de consistance en temps et

$$\hat{R}_i^n = \frac{1}{h^2}(2\bar{u}_i^{n+1} - \bar{u}_{i-1}^{n+1} - \bar{u}_{i+1}^{n+1}) - (\partial_{xx}^2 u(x_i, t_n))$$

est l'erreur de consistance en espace. On a vu (voir (1.31)) que

$$\left|\hat{R}_i^n\right| \leqslant \frac{h^2}{12} \sup_{[0\,;1]} \left|\frac{\partial^4 u}{\partial x^4}(\,\cdot\,, t_n)\right|, \forall i \in \{1, \ldots, N\}$$

Effectuons maintenant un développement de Taylor en fonction du temps d'ordre 2 :

$$u(x_i, t_{n+1}) = u(x_i, t_n) + k\partial_t u + \frac{k^2}{2} u_{tt}(x_i, \xi_n)$$

avec $\xi_n \in [t_n, t_{n+1}]$. Donc

$$\frac{u(x_i, t_{n+1}) - u(x_i, t_n)}{k} - \partial_t u = \frac{k}{2} u_{tt}(x_i, \xi_n)$$

Comme $\xi_n \in [0\,;T]$, et u_{tt} admet un maximum (à x_i fixé) dans $[0\,;T]$ (qui est compact), on a donc

$$\left|\widetilde{R}_i^n\right| \leqslant \frac{k}{2} \max_{[0\,;T]} |u_{tt}(x_i, \,\cdot\,)|\,.$$

Par conséquent,

$$|R_i^n| = \left|\widetilde{R}_i^n + \hat{R}_i^n\right| \leqslant \left|\widetilde{R}_i^n\right| + \left|\hat{R}_i^n\right| \leqslant \frac{k}{2} \max_{[0\,;T]} |u_{tt}(x_i, \,\cdot\,)| + \frac{h^2}{12} \max_{[0\,;1]} \left|\frac{\partial^4 u}{\partial x^4}(\,\cdot\,, t_{n+1})\right|.$$

Donc $|R_i^n| \leqslant C(k + h^2)$ avec

$$C = \frac{1}{2} \max\left(\|u_{tt}\|_{L^\infty([0\,;1]\times[0\,;T]}; \frac{1}{6}\|\frac{\partial^4 u}{\partial x^4}\|_{L^\infty([0\,;1]\times[0\,;T]}\right).$$

Le calcul de l'erreur de consistance pour le schéma (2.34) s'effectue de manière semblable.

2. Le schéma (2.33) est complètement implicite alors que le schéma (2.34) ne l'est que partiellement, puisque le terme de réaction est pris à l'instant n. Le schéma (2.33) s'écrit $AU^{n+1} = U^n$ avec $U^{n+1} = (U_1^{n+1}, \ldots U_N^{n+1})$, $U^n = (U_1^n, \ldots, U_N^n)$ et

$$A = \begin{bmatrix} 1+2\lambda-k & -\lambda & 0 & \ldots & 0 \\ -\lambda & 1+2\lambda-k & -\lambda & \ddots & 0 \\ 0 & & & & \\ & & & & \\ \vdots & & & & 0 \\ 0 & 0 & -\lambda & & 1+2\lambda-k \end{bmatrix}$$

où $\lambda = k/h^2$. Notons que par définition, A est symétrique. De même, le schéma (2.34) s'écrit $BU^{n+1} = U^n$ avec

$$B = \frac{1}{1+k} \begin{bmatrix} 1+2\lambda & -1 & 0 & \ldots & 0 \\ -1 & 1+2\lambda & & \ddots & 0 \\ 0 & & & & -1 \\ & & & & \\ 0 & 0 & -1 & & 1+2\lambda \end{bmatrix}$$

On a donc $A = \lambda A_h$, où A_h est définie en (1.26), avec $c_i = \frac{1-k}{\lambda}$, et $B = \frac{\lambda}{k+1} \quad A_h$ avec $c_i = \frac{1}{\lambda}$. Dans les deux cas, les matrices sont donc s.d.p. en vertu de la proposition 1.4. Notons que l'hypothèse $k \in]0\,;1[$ est nécessaire dans le cas du premier schéma, pour assurer la positivité de c_i.

3. Le schéma (2.34) s'écrit $(1+k)u_i^n = u_i^{n+1} + \lambda(2u_i^{n+1} - u_{i+1}^{n+1} - u_{i-1}^{n+1})$. On montre facilement par récurrence que $\max_{j=1,\dots,N} u_j^n \leqslant (1+k)^n \max_{j=1,\dots,N} u_j^0$, (voir preuve de la stabilité L^∞ d'Euler implicite page 75) et que $\min_{j=1,\dots,N} u_j^{(n)} \geqslant (1+k)^n \min_{j=1,\dots,N} u_j^{(0)}$. On en déduit que $\|u^{(n)}\|_\infty \leqslant (1+k)^n \|u_0\|_\infty$. Or $(1+k)^n \leqslant (1+k)^{T/k}$ car $kn \leqslant T$. Or

$$(1+k)^{T/k} = \exp\left(\frac{T}{k}\ell n(1+k)\right) \leqslant \exp(\frac{T}{k}k) = \mathrm{e}^T\,.$$

On en déduit le résultat, avec $C_1(T) = \mathrm{e}^T$. De même, pour le schéma (2.33), on montre par récurrence que :

$$\|u^{(n)}\|_\infty \leqslant \frac{1}{(1-k)^n}\|u^{(0)}\|.$$

Mais pour $k \in]0, \alpha[$, avec $\alpha \in]0\,;1[$, on a :

$$\frac{1}{(1-k)} \leqslant 1 + \beta k \quad \text{avec} \quad \beta = \frac{1}{(1-\alpha)}.$$

On en déduit par un calcul similaire au précédent que $(1-k)^{T/k} \leqslant e^{\beta T}$ d'où le résultat avec $C_2(T,\alpha) = e^{\beta T}$.

4. Par définition de l'erreur de consistance, on a pour le schéma (2.33)

$$\frac{\bar{u}_j^{(n+1)} - \bar{u}_j^n}{k} - \frac{\bar{u}_{j+1}^{(n+1)} + \bar{u}_{j-1}^{n+1} - 2\bar{u}_j^{(n+1)}}{h^2} - \bar{u}_j^{n+1} = R_i^{(n,1)}$$

et donc, en notant $e_j^{(n)} = \bar{u}_j^n - u_j^{(n)}$ l'erreur de discrétisation en (x_j, t_n), on a :

$$e_j^{(n+1)}(1 + 2\lambda - k) - \lambda e_{j-1}^{(n+1)} - \lambda e_{j+1}^{(n+1)} = e_j^{(n)} + kR_j^{(n,1)}$$

On obtient donc, de manière similaire à la question 3 (en considérant $e_{j_0}^{(n+1)} = \max e_j^{(n+1)}$ puis $e_{j_0}^{(n+1)} = \min e_j^{(n+1)}$)

$$\frac{1}{1-k}\|e^{(n+1)}\| \leqslant \|e^{(n)}\| + kC(k+h^2).$$

Par récurrence sur n, on obtient alors

$$\|e^{(n)}\|_\infty \leqslant \left(\frac{1}{1-k}\right)^n [kC(k+h^2) + \|e^0\|_\infty]$$

d'où $\|e^{(n)}\|_\infty \leqslant C_2(T,\alpha)(TC(k+h^2) + \|e^0\|_\infty)$. De même, pour le schéma (2.34), on écrit l'erreur de consistance :

$$\frac{\bar{u}_j^{(n+1)} - \bar{u}_j^n}{k} - \frac{u_{j+1}^{(n+1)} + \bar{u}_{j-1}^{n+1} - 2\bar{u}_j^{(n+1)}}{h^2} - \bar{u}_j^n = R_j^{(n,2)}$$

et donc :

$$e_j^{(n+1)}(1 + 2\lambda) - \lambda e_{j-1}^{(n+1)} - \lambda e_{j+1}^{(n+1)} = e_j^{(n)}(1+k) + kR_j^{(n,2)}.$$

Par des raisonnements similaires à ceux de la question 3 on obtient alors :

$$\|e_j^{(n)}\| \leqslant (1+k)^n(\|e^{(0)}\| + kC(k+h^2))$$

d'où

$$\|e^{(n)}\|_\infty \leqslant C_1(T)(\|e^{(0)}\| + kC(k+h^2)).$$

Exercice 33 — Schémas "saute-mouton" et Dufort-Frankel.

1. On s'intéresse ici à l'ordre du schéma au sens des différences finies. On suppose que $u \in C^4([0\,;1]\times[0\,;T])$ est solution de (2.36) et on pose

$$\bar{u}_j^n = u(jh, nk),\ j = 0,\dots,N,\ k = 0,\dots,M.$$

L'erreur de consistance est définie par :

$$R_j^n = \frac{\bar{u}_j^{n+1} - \bar{u}_j^{n-1}}{2k} - \frac{\bar{u}_{j-1}^n - 2\bar{u}_j^n + \bar{u}_{j+1}^n}{h^2}, j = 1,\dots,N-1,\ k = 1,\dots,M-1.$$

On cherche une majoration de R_j^n en utilisant des développement de Taylor. Soit $j \in \{1,\dots,\ N-1\}$, $k \in \{1,\dots,\ M-1\}$. Il existe $(\xi_i, t_i) \in [0\,;1]\times[0\,;T]$, $i = 1,\dots,4$, t.q. :

$$\begin{aligned}
\bar{u}_j^{n+1} &= \bar{u}_j^n + k\partial_t u(jh,nk) + \frac{k^2}{2}u_{tt}(jh,nk) + \frac{k^3}{6}u_{ttt}(\xi_1,t_1),\\
\bar{u}_j^{n-1} &= \bar{u}_j^n - k\partial_t u(jh,nk) + \frac{k^2}{2}u_{tt}(jh,nk) - \frac{k^3}{6}u_{ttt}(\xi_2,t_2),\\
\bar{u}_{j-1}^n &= \bar{u}_j^n - h\partial_x u(jh,nk) + \frac{h^2}{2}\partial_{xx}^2 u(jh,nk) - \frac{h^3}{6}\partial_{xxx}u(jh,nk) - \frac{h^4}{24}\partial_{xxxx}u(\xi_3,t_3),\\
\bar{u}_{j+1}^n &= \bar{u}_j^n + h\partial_x u(jh,nk) + \frac{h^2}{2}\partial_{xx}^2 u(jh,nk) + \frac{h^3}{6}\partial_{xxx}u(jh,nk) + \frac{h^4}{24}\partial_{xxxx}u(\xi_4,t_4).
\end{aligned}$$

On en déduit :

$$R_j^n = \partial_t u(jh,nk) + \frac{k^2}{12}\left(u_{ttt}(\xi_1,t_1) + u_{ttt}(\xi_2,t_2)\right) - \partial_{xx}^2 u(jh,nk) - \frac{h^2}{24}\left(\partial_{xxxx}u(\xi_3,t_3) + \partial_{xxxx}u(\xi_4,t_4)\right),$$

et donc, comme u est solution de (2.36),$|R_j^n| \leqslant C_1(k^2+h^2)$, où C_1 ne dépend que de u. Le schéma (2.37) est donc consistant d'ordre 2.

2. Pour étudier la stabilité au sens de Von Neumann, on "oublie" les conditions aux limites dans (2.36). Plus précisément, on s'intéresse à (2.36)avec $x \in \mathbb{R}$ (au lieu de $x \in]0\,;1[$) et on remplace les conditions aux limites par des conditions de périodicité (exactement comme on l'a vu au paragraphe2.2.6). Enfin, on prend une condition initiale de type "mode de Fourier", avec $p \in \mathbb{R}$ arbitraire, et u_0 défini par $u_0(x) = \mathrm{e}^{ipx}$, $x \in \mathbb{R}$. La solution exacte est alors $u(x,t) = \mathrm{e}^{-p^2t}\,\mathrm{e}^{ipx}$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$, c'est-à-dire $u(\,\cdot\,,t) = \mathrm{e}^{-p^2t}\,u_0$, $t \in \mathbb{R}_+$. Le facteur d'amplification est donc, pour tout $t \in \mathbb{R}_+$, le nombre e^{-p^2t}. Ce facteur est toujours, en module, inférieur à 1. On va maintenant chercher la solution du schéma numérique sous la forme :

$$u_j^n = \xi_n\,\mathrm{e}^{ipjh}, j \in \mathbb{Z}, n \in \mathbb{N}, \tag{2.61}$$

où ξ_0 et $\xi_1 \in \mathbb{R}$ sont donnés (ils donnent u_j^0 et u_j^1 pour tout $j \in \mathbb{Z}$) et $\xi_n \in \mathbb{R}$ est à déterminer de manière à ce que la première équation de (2.37)) soit satisfaite. Ce facteur ξ_n va dépendre de k, h et p. Pour k et h donnés, le schéma est stable au sens de Von Neumann si, pour tout $p \in \mathbb{R}$, la suite $(\xi_n)_{n\in\mathbb{N}}$ est bornée. Dans le cas contraire, le schéma est (pour ces valeurs de k et h) dit instable au sens de Von Neumann. Un calcul immédiat donne que la famille des u_j^n, définie par (2.61), est solution de la première équation si et seulement si la suite $(\xi_n)_{n\in\mathbb{N}}$ vérifie (on rappelle que ξ_0 et ξ_1 sont donnés) :

$$\frac{\xi_{n+1} - \xi_{n-1}}{2k} = \frac{2}{h^2}(\cos ph - 1)\xi_n,\ n \geqslant 2,$$

ou encore, en posant $\alpha = 4k/h^2(\cos ph - 1) \leqslant 0$:

$$\xi_{n+1} - \alpha\xi_n - \xi_{n-1} = 0,\ n \geqslant 2 \tag{2.62}$$

Le polynôme associé est

$$r^2 - \alpha r - 1 = 0 \tag{2.63}$$

En excluant le cas $\alpha = -2$ (qui correspond à une racine double), ce polynôme admet deux racines distinctes r_1, r_2 réelles, et comme $r_1 r_2 = 1$, l'un de ces nombres est, en module, supérieur à 1. La solution de (2.62) est donc

$$\xi_n = Ar_1^n + Br_2^n,\ \forall n \geqslant 0 \tag{2.64}$$

où A et B sont déterminés par ξ_0 et ξ_1 (de sorte que $\xi_0 = A + B, \xi_1 = Ar_1 + Br_2$). Ceci montre que $(\xi_n)_n$ est une suite non bornée (sauf pour des choix particuliers de ξ_0 et ξ_1, ceux pour lesquels $\xi_1 = \xi_0 r_2$ et donc $A = 0$, où r_2 est la racine de (2.63) de module inférieur à 1). Ce schéma est donc instable au sens de Von Neumann, pour tout $k > 0$ et $h > 0$.

3. On reprend les notations de la question 1. On s'intéresse maintenant à la quantité S_j^n (qui est toujours l'erreur de consistance) :

$$S_j^n = \frac{\bar{u}_j^{n+1} - u_j^{n-1}}{2k} - \frac{\bar{u}_{j-1}^n - (\bar{u}_j^{n+1} + \bar{u}_j^{n-1}) + \bar{u}_{j+1}}{h^2}, j = 1, \ldots, N-1, \quad k = 0, \ldots, M-1.$$

En reprenant la technique de la question 1, il existe (ξ_i, t_i), $i = 1, \ldots, 6$ $\quad t.q.$

$$S_j^n = \frac{h^2}{12}\left(u_{ttt}(\xi_1, t_1) + u_{ttt}(\xi_2, t_2)\right) - \frac{h^2}{24}\left(\partial_{xxxx} u(\xi_3, t_3) - \partial_{xxxx} u(\xi_4, t_4)\right) + \frac{k^2}{2h^2} h_{tt}(\xi_5, t_5) + \frac{k^2}{2h^2} u_{tt}(\xi_6, t_6).$$

Ce qui donne, avec C_2 ne dépendant que de u,

$$|S_j^n| \leqslant C_2 \left(h^2 + k^2 + \frac{k^2}{h^2}\right), j = 1, \ldots, N-1, \quad k = 0, \ldots, M-1.$$

Le schéma est donc consistant quand $h \to 0$ avec $k/h \to 0$.

4. On reprend la méthode développée à la question 2, la suite $(\xi_n)_n$ doit maintenant vérifier la relation suivante (avec ξ_0, ξ_1 donnés).

$$\frac{\xi_{n+1} - \xi_{n-1}}{2k} = \frac{2\cos(ph)}{h^2}\xi_n - \frac{\xi_{n-1} + \xi_{n+1}}{h^2}, \; n \geqslant 2$$

c'est à dire :

$$\xi_{n+1}\left(\frac{1}{2k} + \frac{1}{h^2}\right) - \frac{2\cos(ph)}{h^2}\xi_n + \xi_{n-1}\left(\frac{1}{h^2} - \frac{1}{2k}\right) = 0, \; n \geqslant 2.$$

L'équation caractéristique est maintenant :

$$p(r) = ar^2 + br + c = 0 \quad \text{avec} \quad a = \frac{1}{2k} + \frac{1}{h^2} \quad b = -\frac{2\cos(ph)}{h^2} \quad \text{et} \quad c = \frac{1}{h^2} - \frac{1}{2h}.$$

Pour montrer la stabilité au sens de Von Neumann, il suffit d'après (2.64) de montrer que les deux racines du polynôme p sont de module inférieur ou égal à 1. On note r_1 et r_2 ces deux racines (qui peuvent être confondues) et on distingue 2 cas :

(a) Les racines de p ne sont pas réelles. Dans ce cas, on a $|r_1| = |r_2| = \gamma$ et $\gamma = |c/a| < 1$ car $k > 0$.

(b) Les racines de p sont réelles. Dans ce cas, on remarque que $r_1 r_2 = c/a < 1$ et l'une des racines, au moins, est donc entre -1 et 1 (strictement). De plus on a

$$p(1) = \frac{2}{h^2} - \frac{2\cos ph}{h^2} \geqslant 0 \quad \text{et} \quad p(-1) = \frac{2}{h^2} + \frac{2\cos ph}{h^2} \geqslant 0,$$

et l'autre racine est donc aussi entre -1 et 1 (au sens large).

On en déduit que le schéma (2.38) est stable au sens de Von Neumann.

Exercice 35— Discrétisation d'un problème parabolique non linéaire.

1. Solution classique versus solution faible — Soit $u \in C(\mathbb{R}^2, \mathbb{R})$; notons u sa restriction à $D =]0,1[\times]0,T[$; notons que l'on a bien $u \in L^\infty(]0\,;1[\times]0,T[)$. Supposons que u satisfait (2.40)-(2.42), et montrons qu'alors u vérifie (2.44). Soit $\psi \in C_T^{2,1}(\mathbb{R}^2)$. En multipliant (2.40) par ψ et en intégrant sur D, on obtient :

$$\int_D \partial_t u(x,t)\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t - \int_D \partial_{xx}^2(\varphi(u))(x,t)\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t = \int_D v(x,t)\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t. \quad (2.65)$$

Par intégration par parties, il vient :

$$\int_D \partial_t u(x,t)\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t = \int_0^1 u(x,T)\psi(x,T)\,\mathrm{d}x - \int_0^1 u(x,0)\psi(x,0)\,\mathrm{d}x - \int_D u(x,t)\partial_t\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t.$$

Comme $\psi \in C_T^{2,1}(\mathbb{R}^2)$ on a donc $\psi(x,T)=0$ pour tout $x \in [0\,;1]$ et comme u vérifie (2.42), on a $u(x,0)=u_0(x)$. On en déduit que

$$\int_D \partial_t u(x,t)\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t = -\int_0^1 u_0(x)\psi(x,0)\,\mathrm{d}x - \int_D \partial_t\psi(x,t)u(x,t)\,\mathrm{d}x\,\mathrm{d}t. \tag{2.66}$$

Intégrons par parties le deuxième terme de (2.65) :

$$\int_D \partial^2_{xx}(\varphi(u))(x,t)\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T [\partial_x(\varphi(u))(1,t)\psi(1,t) - \partial_x(\varphi(u))(0,t)\psi(0,t)]\,\mathrm{d}t - \int_D \partial_x(\varphi(u))(x,t)\partial_x\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t \tag{2.67}$$

et comme u vérifie (2.41), on a

$$\partial_x(\varphi(u))(0,t) = \partial_x(\varphi(u))(1,t) = 0 \quad t \in]0,T[.$$

En tenant compte de ces relations et en ré–intégrant par parties, on obtient :

$$\int_D \partial^2_{xx}(\varphi(u))(x,t)\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t = -\int_D \varphi(u)(x,t)\partial^2_{xx}\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t. \tag{2.68}$$

En remplaçant dans (2.66) et (2.68) dans (2.65), on obtient (2.40). Réciproquement, supposons que u satisfait (2.44), et soit ψ continûment différentiable à support compact dans D. En intégrant (2.44) par parties et en tenant compte que ψ et toutes ses dérivées sont nulles au bord de D, on obtient :

$$\int_D \big[-\partial_t u(x,t) + \partial^2_{xx}(\varphi(u))(x,t) - v(x,t)\big]\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t = 0, \forall \psi \in C_c^\infty(D).$$

Comme u est régulière, ceci entraîne que l'équation (2.40) est satisfaite par u. On prend ensuite $\psi \in C_T^{2,1}(\mathbb{R}^2)$, et on intègre (2.44) par parties ; en tenant compte de (2.45), on obtient :

$$-\int_0^1 u(x,0)\psi(x,0)dx - \int_D \partial_t u(x,t)\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t + \int_0^T \partial_x(\varphi(u))(1,t)\psi(1,t)\,\mathrm{d}t - \int_0^T \partial_x(\varphi(u))(0,t)\psi(0,t)\,\mathrm{d}t + \int_D \partial^2_{xx}(\varphi(u))\psi(x,t)\,\mathrm{d}x\,\mathrm{d}t + \int_0^1 u_0(x)\psi(x,0)\,\mathrm{d}x = 0. \tag{2.69}$$

En regroupant et en utilisant le fait que u satisfait (2.40), on obtient :

$$\int_0^1 (u_0(x)-u(x,0))\psi(x,0)\,\mathrm{d}x + \int_0^T \partial_x(\varphi(u))(1,t)\psi(1,t)\,\mathrm{d}t - \int_0^T \partial_x(\varphi(u))(0,t)\psi(0,t)\,\mathrm{d}t = 0.$$

En choisissant successivement une fonction ψ nulle en $x=0$ et $x=1$ puis nulle en $x=1$ et $t=T$ et enfin nulle en $x=0$ et $t=T$, on obtient que u satisfait la condition initiale (2.42) et les conditions aux limites (2.41), ce qui conclut la question.

2. Existence et unicité de la solution approchée — Soit $n \in \{0,\dots,M-1\}$. On suppose connue la famille $\{u_i^n, i=1,\dots,N\}$, et on veut prouver dans cette question l'existence et l'unicité de la famille $\{u_i^{n+1}, i=1,\dots,N\}$ vérifiant (2.47) avec $u_0^{n+1}=u_1^{n+1}$ et $u_{N+1}^{n+1}=u_N^{n+1}$.

(a) L'application $s \mapsto s$ est strictement croissante, et par hypothèse sur φ, l'application $s \mapsto a\varphi(s)$ est croissante ; on en déduit que g_a est strictement croissante, comme somme d'une fonction strictement croissante et d'une fonction croissante. D'autre part, comme φ est croissante, on a $\varphi(s) \leqslant \varphi(0)$, $\forall s \leqslant 0$, et donc $\lim_{s\to-\infty} g_a(s) = -\infty$. De même, $\varphi(s) \geqslant \varphi(0)$, $\forall s \geqslant 0$, et donc $\lim_{s\to+\infty} g_a(s) = +\infty$. La fonction g_a est continue et prend donc toutes les valeurs de l'intervalle $]-\infty,+\infty[$; comme elle est strictement croissante, elle est bijective.

(b) Posons $a = \frac{k}{h^2}$; l'équation (2.49) s'écrit alors :

$$g_a(u_i) = \frac{k}{h^2}(\overline{w}_{i-1} + \overline{w}_{i+1}) + u_i^n + kv_i^n, \text{ pour tout } i = 1,\dots,N.$$

Par la question précédente, il existe donc un unique u_i qui vérifie cette équation ; il suffit alors de poser $\varphi(u_i) = w_i$ pour déterminer de manière unique la solution de (2.48)-(2.49). On peut donc définir une application F de $\mathbb{R}^N$ dans $\mathbb{R}^N$ par $\overline{w} \mapsto F(\overline{w}) = w$ où w est solution de (2.48)–(2.49).

(c) Soit $\overline{w}^1$ et $\overline{w}^2 \in \mathbb{R}^N$ et soit $w^1 = F(\overline{w}^1)$ et $w^2 = F(\overline{w}^2)$. Par définition de F, on a :

$$u_i^1 - u_i^2 + \frac{2k}{h^2}(w_i^1 - w_i^2) = \frac{k}{h^2}\big((\overline{w}_{i-1}^1 + \overline{w}_{i+1}^1) - (\overline{w}_{i-1}^2 + \overline{w}_{i+1}^2)\big) \quad \forall i = 1, \dots, N. \tag{2.70}$$

Comme φ est monotone, le signe de $w_i^1 - w_i^2 = \varphi(u_i^1) - \varphi(u_i^2)$ est le même que celui de $u_i^1 - u_i^2$, et donc

$$|u_i^1 - u_i^2 + \frac{2k}{h^2}(w_i^1 - w_i^2)| = |u_i^1 - u_i^2| + \frac{2k}{h^2}|w_i^1 - w_i^2|. \tag{2.71}$$

Et comme φ est lipschitzienne de rapport L, on a

$$|w_i^1 - w_i^2| = |\varphi(u_i^1) - \varphi(u_i^2)| \leqslant L|u_i^1 - u_i^2|,$$

d'où :

$$|u_i^1 - u_i^2| \geqslant \frac{1}{L}|w_i^1 - w_i^2|. \tag{2.72}$$

On déduit donc de (2.70),(2.71) et(2.72) que

$$\frac{1}{L}|w_i^1 - w_i^2| + \frac{2k}{h^2}|w_i^1 - w_i^2| \leqslant \frac{k}{h^2}\big(|\overline{w}_{i-1}^1 - \overline{w}_{i-1}^2| + |\overline{w}_{i+1}^1 - \overline{w}_{i+1}^2|\big) \quad \forall i = 1, \dots, N.$$

On a donc

$$|w_i^1 - w_i^2| \leqslant \frac{1}{1 + \frac{h^2}{2kL}} \max_{i=1,\dots,N} |\overline{w}_i^1 - \overline{w}_i^2| \quad \forall i = 1, \dots, N.$$

d'où on déduit que $\|w^1 - w^2\|_\infty \leqslant C\|\overline{w}^1 - \overline{w}^2\|_\infty$ avec $C = \frac{1}{1+\frac{h^2}{2kL}} < 1$. L'application F est donc bien strictement contractante.

(d) Si $\{u_i^{n+1},\ i = 1, \dots, N\}$ est solution de (2.47) et $w_i = \varphi(u_i^{n+1})$ pour $i \in \{1, \dots, N\}$, alors on remarque que $(u_i^{n+1})_{i=1,\dots,N}$ et $(w_i)_{i=1,\dots,N}$ vérifient (2.48)–(2.49) avec $\overline{w}_i = w_i$ pour $i = 1, \dots, N$. On en déduit que $w = F(w)$.

(e) Par définition de F, on a $F(w) = \widetilde{w}$ avec $(\widetilde{u}, \widetilde{w}) \in \mathbb{R}^N \times \mathbb{R}^N$, $\widetilde{u} = (\widetilde{u}_i)_{i=1,\dots,N}$, $\widetilde{w} = (\widetilde{w}_i)_{i=1,\dots,N}$, t.q. $\varphi(\widetilde{u}_i) = \widetilde{w}_i$, $\forall i = 1, \dots, N$ et

$$\widetilde{u}_i + \frac{2k}{h^2}\widetilde{w}_i = \frac{k}{h^2}(w_{i-1} + w_{i+1}) + u_i^n + kv_i^n \quad \forall i = 1, \dots, N. \tag{2.73}$$

Comme $F(w) = w$, on a donc $\widetilde{w}_i = w_i$ et on obtient l'existence de $u_i^{n+1} = \widetilde{u}_i$ tel que $w_i = \varphi(u_i^{n+1})$ pour $i = 1, \dots, N$. Il suffit alors de remplacer w_i et $\widetilde{w}_i$ par $\varphi(u_i^{n+1}$ dans (2.73) pour conclure que $\{u_i^{n+1},\ i = 1, \dots, N\}$ est solution de (2.47).

(f) On vient de montrer dans les questions précédentes que $\{u_i^{n+1},\ i = 1, \dots, N\}$ est solution de (2.47) si et seulement si w défini par $w_i = \varphi(u_i^{n+1})$ est solution de $w = F(w)$, où F est définie par $(2.48) - (2.49)$. Comme F est une application strictement croissante, il existe un unique point fixe $w = F(w)$; par définition de F, il existe donc une unique famille $\{u_i^{n+1}\}$, $i = 1, \dots, N$ solution de (2.47).

3. Estimation $L^\infty(]0\,;1[\times]0, T[)$ sur u — La relation à démontrer par récurrence est clairement vérifiée au rang $n = 0$, par définition de A. Supposons qu'elle soit vraie jusqu'au rang n, et démontrons–la au rang $n + 1$. La relation (2.47) s'écrit encore :

$$u_i^{n+1} = u_i^n + \frac{k}{h^2}(\varphi(u_{i-1}^{n+1}) - \varphi(u_i^{n+1})) + \frac{k}{h^2}(\varphi(u_{i+1}^{n+1} - \varphi(u_i^{n+1})) + kv_i^n$$

pour $i = 1, \dots, N$ et $n = 0, \dots, M - 1$. Supposons que i est tel que $u_i^{n+1} = \min_{j=1,\dots,N} u_j^{n+1}$. Comme φ est croissante, on a dans ce cas :

$$\varphi(u_{i-1}^{n+1}) - \varphi(u_i^{n+1}) \geqslant 0 \quad \text{et} \quad \varphi(u_{i+1}^{n+1}) - \varphi(u_i^{n+1}) \geqslant 0,$$

et on en déduit que

$$\min_{j=1,\dots,N} u_j^{n+1} \geqslant u_i^n - kB$$

d'où, par hypothèse de récurrence

$$\min_{j=1,\dots,N} u_j^{n+1} \geqslant -A - nkB - kB$$

Un raisonnement similaire en considérant maintenant i tel que $u_i^{n+1} = \max_{j=1,\dots,N} u_j^{n+1}$ conduit à :

$$\max_{j=1,\dots,N} u_j^{n+1} \leqslant u_i^n + kB \leqslant A + nkB + kB$$

On a donc bien :

$$-A - (n+1)kB \leqslant u_i^{n+1} \leqslant A + (n+1)kB \quad \forall i = 1,\dots,N, \quad \forall n = 0,\dots,M$$

On en déduit alors que $\|u^n\|_{L^\infty(]0\,;1[)} \leqslant c_{u_0,v,T}$, avec $c_{u_0,v,T} = A + BT$.

4. Estimation de la dérivée par rapport à x de $\varphi(u)$ — En multipliant (2.47) par u_i^{n+1} et en sommant sur i, on obtient $A_n + B_n = C_n$, avec

$$A_n = \sum_{i=1}^N \frac{u_i^{n+1} - u_i^n}{k} u_i^{n+1},\ B_n = -\sum_{i=1}^N \frac{\varphi(u_{i-1}^{n+1}) - 2\varphi(u_i^{n+1}) + \varphi(u_{i+1}^{n+1})}{h^2} u_i^{n+1} \text{ et } C_n = \sum_{i=1}^N v_i^n u_i^{n+1}.$$

En utilisant l'inégalité $a^2 - ab = \frac{a^2}{2} - \frac{b^2}{2}$, on obtient :

$$A_n \geqslant \alpha_{n+1} - \alpha_n, \text{ avec } \alpha_n = \frac{1}{2k}\sum_{i=1}^N (u_i^n)^2.$$

En développant B_n, on obtient :

$$B_n = -\frac{1}{h^2}\Big(\sum_{i=1}^N (\varphi(u_{i-1}^{n+1}) - \varphi(u_i^{n+1}))u_i^{n+1} + \sum_{i=1}^N (-\varphi(u_i^{n+1}) + \varphi(u_{i+1}^{n+1}))u_i^{n+1})\Big).$$

Par un changement d'indice sur les sommes, on obtient alors :

$$B_n = -\frac{1}{h^2}\Big(\sum_{i=0}^{N-1} (\varphi(u_i^{n+1}) - \varphi(u_{i+1}^{n+1}))u_{i+1}^{n+1} - \sum_{i=1}^N (-\varphi(u_i^{n+1}) + \varphi(u_{i+1}^{n+1}))u_i^{n+1})\Big).$$

En tenant compte du fait que $u_0^{n+1} = u_1^{n+1}$, $u_{N+1}^{n+1} = u_N^{n+1}$, pour tout $n = 0,\dots,M-1$, on obtient alors que :

$$B_n = \frac{1}{h^2}\Big(\sum_{i=1}^{N-1} (\varphi(u_{i+1}^{n+1}) - \varphi(u_i^{n+1}))(u_{i+1}^{n+1} - u_i^{n+1})\Big).$$

En utilisant le caractère lipschitzien de φ, on obtient la minoration suivante :

$$B_n \geqslant \frac{1}{Lh^2}\sum_{i=1}^{N-1} (\varphi(u_{i+1}^{n+1}) - \varphi(u_i^{n+1}))^2.$$

Enfin, on majore C_n :

$$C_n \leqslant \frac{Bc_{u_0,v,T}}{h}.$$

L'égalité $A_n + B_n = C_n$ entraîne donc :

$$\alpha_{n+1} - \alpha_n + \frac{1}{Lh^2}\sum_{i=1}^{N-1} (\varphi(u_{i+1}^{n+1}) - \varphi(u_i^{n+1}))^2 \leqslant \frac{Bc_{u_0,v,T}}{h}.$$

En sommant pour $n = 0$ à $M-1$, et en notant que $\alpha_M \geqslant 0$,on obtient alors :

$$\frac{1}{Lh^2}\sum_{n=0}^{M-1}\sum_{i=1}^{N-1}(\varphi(u_{i+1}^{n+1})-\varphi(u_i^{n+1}))^2 \leqslant \frac{Bc_{u_0,v,T}}{h}+\alpha_0.$$

Il reste à remarquer que $\alpha_0 \leqslant \frac{h}{2k}c_{u_0,v,T}^2$ pour conclure que :

$$\sum_{n=0}^{M-1}\sum_{i=1}^{N-1}(\varphi(u_{i+1}^{n+1})-\varphi(u_i^{n+1}))^2 \leqslant C_1\frac{h}{k},\ \text{avec } C_1 = Lc_{u_0,v,T}(B+\frac{1}{2}c_{u_0,v,T}).$$

5. Estimation de la dérivée par rapport à t de $\varphi(u)$ — Multiplions (2.47) par $\varphi(u_i^{n+1})-\varphi(u_i^n)$ et sommons pour $i = 1,\dots,N$. On obtient :

$$A_n + B_n = C_n, \tag{2.74}$$

avec

$$A_n = \sum_{i=1}^{N}\frac{u_i^{n+1}-u_i^n}{k}(\varphi(u_i^{n+1})-\varphi(u_i^n))$$

$$B_n = -\sum_{i=1}^{N}\frac{\varphi(u_{i-1}^{n+1})-2\varphi(u_i^{n+1})+\varphi(u_{i+1}^{n+1})}{h^2}(\varphi(u_i^{n+1})-\varphi(u_i^n))$$

$$C_n = \sum_{i=1}^{N}v_i^n(\varphi(u_i^{n+1})-\varphi(u_i^n))$$

En utilisant le caractère lipschitzien de φ, on obtient la minoration suivante :

$$A_n \geqslant \frac{1}{Lk}\sum_{i=1}^{N-1}(\varphi(u_i^{n+1})-\varphi(u_i^n))^2. \tag{2.75}$$

En développant B_n, on obtient :

$$B_n = -\frac{1}{h^2}\Big(\sum_{i=1}^{N}(\varphi(u_{i-1}^{n+1})-\varphi(u_i^{n+1}))(\varphi(u_i^{n+1})-\varphi(u_i^n)) + \sum_{i=1}^{N}(-\varphi(u_i^{n+1})+\varphi(u_{i+1}^{n+1}))(\varphi(u_i^{n+1})-\varphi(u_i^n))\Big). \tag{2.76}$$

Par un changement d'indice sur les sommes, on obtient alors :

$$B_n = -\frac{1}{h^2}\Big(\sum_{i=0}^{N-1}(\varphi(u_i^{n+1})-\varphi(u_{i+1}^{n+1}))(\varphi(u_{i+1}^{n+1})-\varphi(u_{i+1}^n)) + \sum_{i=1}^{N}(-\varphi(u_i^{n+1})+\varphi(u_{i+1}^{n+1}))(\varphi(u_i^{n+1})-\varphi(u_i^n))\Big). \tag{2.77}$$

En tenant compte du fait que $u_0^{n+1} = u_1^{n+1}$ et $u_{N+1}^{n+1} = u_N^{n+1}$ pour tout $n = 0,\dots,M-1$, on obtient alors :

$$B_n = \frac{1}{h^2}\big(\sum_{i=1}^{N-1}(\varphi(u_{i+1}^{n+1})-\varphi(u_i^{n+1}))((\varphi(u_{i+1}^{n+1})-\varphi(u_i^{n+1}))-(\varphi(u_{i+1}^n)-\varphi(u_i^n)))\big).$$

En utilisant à nouveau la relation $a(a-b) \geqslant \frac{a^2}{2}-\frac{b^2}{2}$, on obtient :

$$B_n \geqslant \beta_{n+1}-\beta_n,\ \text{avec } \beta_n = \frac{1}{2h^2}\sum_{i=1}^{N-1}(\varphi(u_{i+1}^n)-\varphi(u_i^n))^2 \tag{2.78}$$

Enfin, on majore C_n par :

$$C_n \leqslant \frac{1}{2Lk}\sum_{i=1}^{N}(\varphi(u_i^{n+1})-\varphi(u_i^n))^2 + C\sum_{i=1}^{N} k \geqslant C\frac{k}{h}. \tag{2.79}$$

En utilisant (2.74), (2.75), (2.78) et (2.79), on obtient :

$$\frac{1}{2Lk}\sum_{i=1}^{N}(\varphi(u_i^{n+1})-\varphi(u_i^n))^2 + \beta_{n+1} - \beta_n \leqslant C\frac{k}{h}. \tag{2.80}$$

En sommant sur n, on obtient d'une part, en utilisant le fait que $\beta_n \geqslant 0$:

$$\sum_{n=0}^{M-1}\sum_{i=1}^{N}(\varphi(u_i^{n+1})-\varphi(u_i^n))^2 \leqslant 2LC\frac{k}{h} + 2L\beta_0 k. \tag{2.81}$$

d'autre part, en utilisant que le fait que le premier terme est positif, on obtient par (2.80) une majoration sur β_M, et donc sur β_n pour tout $n \leqslant M$:

$$\beta_n \leqslant \frac{C}{h} + \beta_0. \tag{2.82}$$

Il ne reste donc plus qu'à majorer β_0 pour obtenir (2.51) et (2.52). Par définition, on a

$$\beta_0 = \sum_{i=1}^{N-1}\frac{\varphi(u_i^0)-\varphi(u_{i+1}^0)}{2h^2}.$$

En utilisant le fait que φ est lipschitzienne et que la différence entre u_i^0 et u_{i+1}^0 est en h, on obtient (2.52) à partir de (2.81) et (2.51) à partir de (2.82).

6. Par définition de la fonction u_h, et grace au résultat de la question 3, on a :

$$\sup_{\substack{x\in](i-1)h,ih[\\ t\in[nk,(n+1)k]}} u_h(x,t) \leqslant \frac{t-nk}{k}\|u_h^{(n+1)}\|_\infty + \frac{(n+1)k-t}{k}\|u_h^{(n)}\|_\infty \leqslant c_{u_0,v,T}.$$

ce qui prouve que la suite $(u_h)_{M\in\mathbb{N}^\star}$ est bornée dans $L^\infty(]0\,;1[\times]0,T[)$. Comme φ est continue, on en déduit immédiatement que $(\varphi(u_h))_{M\in\mathbb{N}^\star}$ est bornée dans $L^\infty(]0\,;1[\times]0,T[)$

3

Méthodes variationnelles

3.1 Exemples de problèmes variationnels

3.1.1 Problème de Dirichlet

Soit Ω un ouvert borné de $\mathbb{R}^d$, $d \geqslant 1$. On considère le problème suivant :

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

où $f \in \mathcal{C}(\bar{\Omega})$ et $\Delta u = \partial_1^2 u + \partial_2^2 u$ où l'on désigne par $\partial_i^2 u$ la dérivée partielle d'ordre 2 par rapport à la i-ème variable.

Définition 3.1 — Solution classique. On appelle solution classique de (3.1) une fonction $u \in \mathcal{C}^2(\bar{\Omega})$ qui vérifie (3.1).

Soit $u \in \mathcal{C}^2(\bar{\Omega})$ une solution classique de (3.1) et soit $\varphi \in \mathcal{C}_c^\infty(\Omega)$ où $\mathcal{C}_c^\infty(\Omega)$ désigne l'ensemble des fonctions de classe $\mathcal{C}^\infty$ à support compact dans Ω. On multiplie (3.1) par φ et on intègre sur Ω (on appellera par la suite *fonction test* une telle fonction φ) pour obtenir :

$$\int_\Omega -\Delta u(x)\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x.$$

Notons que les deux intégrales présentes dans cette équation sont bien définies, puisque $\Delta u \in \mathcal{C}(\Omega)$ et $f \in \mathcal{C}(\Omega)$. Par intégration par parties (formule de Green), on a :

$$\begin{aligned} \int_\Omega -\Delta u(x)\varphi(x)\,\mathrm{d}x &= -\sum_{i=1}^d \int_\Omega \partial_i^2 u(x)\varphi(x)\,\mathrm{d}x \\ &= \sum_{i=1}^d \int_\Omega \partial_i u(x)\partial_i\varphi(x)\,\mathrm{d}x + \sum_{i=1}^d \int_{\partial\Omega} \partial_i u\; n_i(s)\varphi(s)\;\mathrm{d}\gamma(s), \end{aligned}$$

où n_i désigne la i–ème composante du vecteur unitaire normal à la frontière $\partial\Omega$ de Ω extérieur à Ω et $\mathrm{d}\gamma$ désigne le symbole d'intégration pour la mesure de Lebesgue sur $\partial\Omega$. Comme φ est nulle sur $\partial\Omega$, on obtient :

$$\sum_{i=1}^d \int_\Omega \partial_i u(x)\partial_i\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x.$$

ce qui s'écrit encore :

$$\int_\Omega \nabla u(x)\,\cdot\,\nabla\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x. \tag{3.2}$$

Donc toute solution classique de (3.1) satisfait (3.2). Prenons maintenant comme fonction test φ, non plus une fonction de $\mathcal{C}_c^\infty(\Omega)$ mais une fonction de $H_0^1(\Omega)$. On rappelle que l'espace $H_0^1(\Omega)$ est défini comme l'adhérence de $\mathcal{C}_c^\infty(\Omega)$ dans $H^1(\Omega) = \{u \in L^2(\Omega); Du \in L^2(\Omega)\}$ où Du désigne la dérivée faible de u, au sens de la définition suivante :

Définition 3.2 — Dérivée par transposition, dérivée faible. Soit Ω un ouvert de $\mathbb{R}^N$, $N \geqslant 1$; soit $\mathcal{D}(\Omega) = \mathcal{D}(\Omega)$ et $\mathcal{D}^\star(\Omega)$ son dual algébrique, c'est-à-dire l'ensemble des formes linéaires sur $\mathcal{D}(\Omega)$;

- Soit $f \in \mathcal{L}^1_{\text{loc}}(\Omega)$, on appelle dérivée par transposition de f par rapport à sa i-ème variable la forme linéaire $D_i f$ sur $\mathcal{D}(\Omega)$ définie par :

$$\langle D_i f, \varphi \rangle_{\mathcal{D}^\star(\Omega),\mathcal{D}(\Omega)} = -\int_\Omega f \, \partial_i \varphi \, \mathrm{d}x.$$

  où $\partial_i \varphi$ désigne la dérivée partielle classique de φ par rapport à sa i-ème variable. Donc $D_i f$ est un élément de $\mathcal{D}^\star(\Omega)$. Noter que si $f \in C^1(\Omega)$, alors $D_i f$ n'est autre que $\partial_i f$ car on confond $\partial_i f$ et $T_{\partial_i f}$ (qui est l'élément de $\mathcal{D}^\star(\Omega)$ induit par $\partial_i f$). Il s'agit donc bien d'une généralisation de la notion de dérivée.

  Si la forme linéaire $D_i f$ peut être confondue avec une fonction localement intégrable, cette fonction est unique à un ensemble de mesure nulle près [7, Lemme 1.1], et on dit que cette fonction est la *dérivée faible de f dans la direction i*.

- Soit $T \in \mathcal{D}^\star(\Omega)$; on définit la dérivée par transposition $D_i T$ de T par :

$$\langle D_i T, \varphi \rangle_{\mathcal{D}^\star(\Omega),\mathcal{D}(\Omega)} = -\langle T, \partial_i \varphi \rangle_{\mathcal{D}^\star(\Omega),\mathcal{D}(\Omega)}, \ \forall \varphi \in \mathcal{D}(\Omega).$$

On rappelle que l'espace $H^1(\Omega)$ muni du produit scalaire

$$(u, v)_{H^1} = \int_\Omega u(x) v(x) \, \mathrm{d}x + \sum_{i=1}^{d} \int_\Omega D_i u(x) D_i v(x) \, \mathrm{d}x \tag{3.3}$$

est un espace de Hilbert. Les espaces $H^1(\Omega)$ et $H^1_0(\Omega)$ font partie des espaces dits *de Sobolev* [1, 7]. Si $\varphi \in H^1_0(\Omega)$, par définition, il existe $(\varphi_n)_{n\in\mathbb{N}} \subset \mathcal{C}^\infty_c(\Omega)$ telle que $\varphi_n \to \varphi$ dans H^1 lorsque $n \to +\infty$, c'est-à-dire qu'on a

$$\|\varphi_n - \varphi\|^2_{H^1} = \|\varphi_n - \varphi\|^2_{L^2} + \sum \|D_i \varphi_n - D_i \varphi\|^2_{L^2} \to 0 \text{ lorsque } n \to +\infty.$$

Pour chaque fonction $\varphi_n \in \mathcal{C}^\infty_c(\Omega)$ on a par (3.2) :

$$\sum_{i=1}^{N} \int_\Omega \partial_i u(x) \partial_i \varphi_n(x) \, \mathrm{d}x = \int_\Omega f(x) \varphi_n(x) \, \mathrm{d}x \qquad \forall n \in \mathbb{N}. \tag{3.4}$$

Or la i-ème dérivée partielle $\partial_i \varphi_n = \partial \varphi_n / \partial x_i$ converge vers $D_i \varphi$ dans L^2 donc dans L^2 faible lorsque n tend vers ∞ et φ_n tend vers φ dans $L^2(\Omega)$. On a donc :

$$\int_\Omega \partial_i u(x) \partial_i \varphi_n(x) \, \mathrm{d}x \to \int_\Omega \partial_i u(x) D_i \varphi(x) \, \mathrm{d}x \text{ lorsque } n \to +\infty$$
$$\int_\Omega f(x) \varphi_n(x) \, \mathrm{d}x \to \int_\Omega f(x) \varphi(x) \, \mathrm{d}x \text{ lorsque } n \to +\infty$$

L'égalité (3.4) est donc vérifiée pour toute fonction $\varphi \in H^1_0(\Omega)$. Montrons maintenant que si u est solution classique de (3.1) alors $u \in H^1_0(\Omega)$. En effet, si $u \in \mathcal{C}^2(\Omega)$, alors $u \in \mathcal{C}(\bar{\Omega})$ et donc $u \in L^2(\Omega)$; de plus $\partial_i u \in \mathcal{C}(\bar{\Omega})$ donc $\partial_i u \in L^2(\Omega)$. On a donc bien $u \in H^1(\Omega)$. Il reste à montrer que $u \in H^1_0(\Omega)$. Pour cela on rappelle les théorèmes de trace suivants (voir [7, chapitre 1]).

Théorème 3.1 — Existence de l'opérateur trace. Soit Ω un ouvert (borné ou non borné) de $\mathbb{R}^d$, $d \geqslant 1$, de frontière $\partial\Omega$ lipschitzienne, alors l'espace $\mathcal{C}^\infty_c(\bar{\Omega})$ des fonctions de classe $\mathcal{C}^\infty$ et à support compact dans $\bar{\Omega}$ est dense dans $H^1(\Omega)$. On peut donc définir par continuité l'application *trace* qui est linéaire continue de $H^1(\Omega)$ dans $L^2(\partial\Omega)$, définie par $\gamma(u) = u|_{\partial\Omega}$ si $u \in \mathcal{C}^\infty_c(\bar{\Omega})$ et par

$$\gamma(u) = \lim_{n\to+\infty} \gamma(u_n) \text{ si } u \in H^1(\Omega), \ \ u = \lim_{n\to+\infty} u_n,$$

où $(u_n)_{n\in\mathbb{N}} \subset \mathcal{C}_c^\infty(\bar{\Omega})$. Dire que l'application (linéaire) γ est continue est équivalent à dire qu'il existe $C \in \mathbb{R}_+$ tel que

$$\|\gamma(u)\|_{L^2(\partial\Omega)} \leqslant C\|u\|_{H^1(\Omega)} \text{ pour tout } u \in H^1(\Omega). \tag{3.5}$$

Notons que $\gamma(H^1(\Omega)) \subset L^2(\Omega)$ mais $\gamma(H^1(\Omega)) \neq L^2(\partial\Omega)$. On note $H^{1/2}(\Omega) = \gamma(H^1(\Omega))$.

Remarquons que si Ω est un ouvert borné, alors $\bar{\Omega}$ est compact et donc toutes les fonctions $\mathcal{C}^\infty$ sont à support compact dans $\bar{\Omega}$.

Théorème 3.2 — Noyau de l'opérateur trace. Soit Ω un ouvert borné de $\mathbb{R}^d$ de frontière $\partial\Omega$ lipschitzienne et γ l'opérateur trace défini par le théorème (3.1). Alors

$$\operatorname{Ker}\gamma = H_0^1(\Omega).$$

Si $u \in \mathcal{C}^2(\bar{\Omega})$ est une solution classique de (3.1), alors $\gamma(u) = u|_{\partial\Omega} = 0$ donc $u \in \operatorname{Ker}\gamma$ et par le théorème 3.2, ceci prouve que $u \in H_0^1(\Omega)$.

Nous avons ainsi montré que toute solution classique de (3.1) vérifie $u \in H_0^1(\Omega)$ et l'egalité (3.2). Cette remarque motive l'introduction de solutions plus générales qui permettent de s'affranchir de la régularité $\mathcal{C}^2$ et qu'on appellera *solutions faibles*.

Définition 3.3 — Formulation faible. Soit $f \in L^2(\Omega)$, on dit que u est solution faible de (3.1) si $u \in H_0^1(\Omega)$ est solution de

$$\sum_{i=1}^{N} \int_\Omega D_i u(x) D_i\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x \qquad \forall\varphi \in H_0^1(\Omega) \tag{3.6}$$

Définition 3.4 — Formulation variationnelle. Soit $f \in L^2(\Omega)$; on dit que u est solution variationnelle de (3.1) si $u \in H_0^1(\Omega)$ est solution du problème de minimisation :

$$J(u) \leqslant J(v) \quad \forall v \in H_0^1(\Omega) \quad \text{avec} \quad J(v) = \frac{1}{2}\int_\Omega \nabla v(x) \cdot \nabla v(x)\,\mathrm{d}x - \int_\Omega f(x)v(x)\,\mathrm{d}x \tag{3.7}$$

où on a noté :

$$\int_\Omega \nabla u(x) \cdot \nabla\varphi(x)\,\mathrm{d}x = \sum_{i=1}^{d} \int_\Omega D_i u(x) D_i\varphi(x)\,\mathrm{d}x.$$

On cherche à montrer l'existence et l'unicité de la solution de (3.6) et (3.7). Pour cela, on utilise le théorème de Lax [1]–Milgram [2] qu'on rappelle ici :

Théorème 3.3 — Lax-Milgram. Soit H un espace de Hilbert, soit a une forme bilinéaire continue coercive sur H et $T \in H'$. Il existe un unique élément $u \in H$ tel que

$$a(u, v) = T(v) \qquad \forall v \in H \tag{3.8}$$

De plus, si a est symétrique, $u \in H$ est l'unique solution du problème de minimisation suivant :

$$J(u) \leqslant J(v) \tag{3.9}$$

où J est définie de H dans $\mathbb{R}^N$ par :

$$J(v) = \frac{1}{2}a(v, v) - T(v). \tag{3.10}$$

Démonstration.

— Si a est symétrique l'existence et l'unicité de u est immédiate par le théorème de représentation de Riesz (car dans ce cas a est un produit scalaire et la forme linéaire définie par $\varphi \mapsto \int_\Omega f(x)\varphi(x)\,\mathrm{d}x$ est continue pour la norme associée à ce produit scalaire).

1. Peter Lax, mathématicien américain d'origine hongroise contemporain
2. Arthur Milgram (1912-1948), mathématicien américain

— Si a est non symétrique, on considère l'application de H dans H, qui à u associe Au, défini par $(Au, v) = a(u, v)$, $\forall v \in H$. L'application qui à u associe Au est linéaire continue et $(Au, v) \leqslant a(u, v) \leqslant M\|u\|\|v\|$ car a est continue. D'autre part, par le théorème de représentation de Riesz, on a existence et unicité de $\psi \in H$ tel que $T(v) = (\psi, v)$, pour tout $v \in H$. Donc u est solution de $a(u, v) = T(v)$, $\forall v \in H$ si et seulement si $Au = \psi$. Pour montrer l'existence et l'unicité de u, il faut donc montrer que A est bijectif.

— Montrons d'abord que A est injectif. On suppose que $Au = 0$. On a $(Au, u) \geqslant \alpha\|u\|^2$ par coercitivité de a et comme $\|Au\|\,\|v\| \geqslant (Au, v)$, on a donc $\|Au\| \geqslant \alpha\|u\|$. En conclusion, si $Au = 0 \Rightarrow u = 0$.

— Montrons maintenant que A est surjectif. On veut montrer que $AH = H$. Pour cela, on va montrer que AH est fermé et $AH^\top = \{0\}$. Soit $w \in \overline{AH}$; il existe alors une suite $(v_n)_{n\in\mathbb{N}} \subset H$ telle que $Av_n \to w$ dans H.

— Montrons que la suite $(v_n)_{n\in\mathbb{N}}$ converge dans H. On a :

$$\|Av_n - Av_m\| = \|A(v_n - v_m)\| \geqslant \alpha\|v_n - v_m\|_H$$

donc la suite $(v_n)_{n\in\mathbb{N}}$ est de Cauchy. On en déduit qu'elle converge vers un certain $v \in H$. Comme A est continue, on a donc : $Av_n \to Av$ dans H et donc $w = Av \in AH$.

— Montrons maintenant que $AH^\top = \{0\}$. Soit $v_0 \in AH^\top$, comme a est coercive, on a $\alpha\|v_0\|^2 \leqslant a(v_0, v_0) = (Av_0, v_0) = 0$, on en déduit que $v_0 = 0$, ce qui prouve que $AH^\top = \{0\}$.

Pour conclure la preuve du théorème, il reste à montrer que si a est symétrique, le problème de minimisation (3.9) est équivalent au problème (3.8) Soit $u \in H$ solution unique de (3.8) ; montrons que u est solution de (3.9). Soit $w \in H$, on va montrer que $J(u + w) \geqslant J(u)$.

$$\begin{aligned} J(u+w) &= \frac{1}{2}a(u+w, u+w) - T(u+w) \\ &= \frac{1}{2}a(u,u) + \frac{1}{2}\left[a(u,w) + a(w,u)\right] + \frac{1}{2}a(w,w) - T(u) - T(w) \\ &= \frac{1}{2}a(u,u) + \frac{1}{2}a(w,w) + a(u,w) - T(u) - T(w) \\ &= J(u) + \frac{1}{2}a(w,w) \geqslant J(u) + \frac{\alpha}{2}\|w\|^2 \end{aligned}$$

Donc $J(u + w) > J(u)$ sauf si $w = 0$.

Réciproquement, supposons maintenant que u est solution du problème de minimisation (3.9) et montrons que u est solution du problème (3.8). Soit $w \in H$ et $t > 0$; on a $J(u+tw) - J(u) \geqslant 0$ et $J(u-tw) - J(u) \geqslant 0$ car u minimise J. On en déduit que :

$$ta(u,w) - tT(w) + \frac{1}{2}t^2a(w,w) \geqslant 0 \quad \text{et} \quad -ta(u,w) + tT(w) + \frac{1}{2}t^2a(w,w) \geqslant 0$$

Comme t est strictement positif, on peut diviser ces deux inégalités par t :

$$a(u,w) - T(w) + \frac{1}{2}ta(w,w) \geqslant 0 \quad \text{et} \quad -a(u,w) + T(w) + \frac{1}{2}ta(w,w) \geqslant 0$$

On fait alors tendre t vers 0 et on obtient $a(u, w) = T(w)$ pour tout $w \in H$, ce qui montre que u est bien solution du problème (3.8). •

Montrons que l'on peut appliquer le théorème de Lax-Milgram pour les problèmes (3.6) et (3.7).

Proposition 3.5 — Existence et unicité de la solution de (3.1). Si $f \in L^2(\Omega)$, il existe un unique $u \in H_0^1(\Omega)$ solution de (3.6) et (3.7).

Démonstration. Montrons que les hypothèses du théorème de Lax–Milgram sont vérifiées. L'espace $H = H_0^1(\Omega)$ est un espace de Hilbert et :

$$a(u,v) = \int_\Omega \nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x \qquad T(v) = \int_\Omega f(x)v(x)\,\mathrm{d}x.$$

Montrons que $T \in H'$; en effet, la forme T est linéaire et on a $T(v) \leqslant \|f\|_{L^2}\|v\|_{L^2} \leqslant \|f\|_{L^2}\|v\|_{H^1}$. On en déduit que T est une forme linéaire continue sur $H_0^1(\Omega)$, ce qui est équivalent à dire que $T \in H^{-1}(\Omega)$ (dual topologique de $H_0^1(\Omega)$).

Montrons que a est bilinéaire, continue et symétrique. La continuité de a se démontre en écrivant que

$$a(u,v) = \int_\Omega \nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x \leqslant \|\nabla u\|_{L^2}\|\nabla v\|_{L^2} \leqslant \|u\|_{H^1}\|v\|_{H^1}$$

Les caractères bilinéaire et symétrique sont évidents. Montrons maintenant que a est coercitive. En effet :

$$a(v,v) = \int_\Omega \nabla v(x) \cdot \nabla v(x)\,\mathrm{d}x = \sum_{i=1}^N \int_\Omega D_i v(x) D_i v(x)\,\mathrm{d}x \geqslant \frac{1}{\mathrm{diam}(\Omega)^2 + 1}\|u\|_{H^1}^2$$

par l'inégalité de Poincaré. Comme $T \in H'$ et comme a est linéaire, continue, coercitive donc le théorème de Lax–Milgram s'applique : on en conclut qu'il existe une unique fonction $u \in H_0^1(\Omega)$ solution de (3.6) et comme a est symétrique, u est l'unique solution du problème de minimisation associée. •

Définition 3.6 — Solution forte dans H^2. Soit $f \in L^2(\Omega)$, on dit que u est solution forte de (3.1) dans H^2 si $u \in H^2(\Omega) \cap H_0^1(\Omega)$ vérifie $-\Delta u = f$ dans $L^2(\Omega)$.

Remarquons que si u est solution forte $\mathcal{C}^2$ de (3.1), alors u est solution forte H^2. De même, si u est solution forte H^2 de (3.1) alors u est solution faible de (3.1). Les réciproques sont fausses. On admettra le théorème de régularité (dont la preuve est difficile), qui s'énonce de la manière suivante :

Théorème 3.4 — Régularité. Soit Ω un ouvert borné de $\mathbb{R}^d$, $d \geqslant 1$ On suppose que Ω a une frontière de classe $\mathcal{C}^2$ ou que Ω est convexe à frontière lipschitzienne. Si $f \in L^2(\Omega)$ et si $u \in H_0^1(\Omega)$ est solution faible de (3.1), alors $u \in H^2(\Omega)$ et il existe $C_0 > 0$ ne dépendant que de f et Ω tel que $\|u\|_{H^2} \leqslant C_0 \|f\|_{L^2}$. De plus, si $f \in H^m(\Omega)$ alors $u \in H^{m+2}(\Omega)$, et il existe $C_m > 0$ ne dépendant que de f et Ω tel que $\|u\|_{H^{m+2}} \leqslant C_m \|f\|_{H^m}$.

Remarque 3.7 — Différences entre les méthodes de discrétisation. Lorsqu'on adopte une discrétisation par différences finies, on discrétise directement terme à terme le problème (3.1). Lorsqu'on adopte une méthode de volumes finis, on discrétise le *bilan* obtenu en intégrant (3.1) sur chaque maille. Lorsqu'on utilise une méthode variationnelle, on discrétise la formulation variationnelle (3.7) dans le cas de la méthode de Ritz, la formulation faible (3.6) dans le cas de la méthode de Galerkin [section 3.2].

Remarque 3.8 — Sur la prise en compte des conditions aux limites. Remarquons également que dans la formulation faible (3.6), les conditions aux limites de Dirichlet homogènes $u = 0$ sont prises en compte dans l'espace $u \in H_0^1(\Omega)$ et donc également dans l'espace d'approximation H_N. En revanche pour le problème de Neumann homongène, les conditions aux limites ne sont pas explicitées dans l'espace fonctionnel (voir exercice 47).

3.1.2 Problème de Dirichlet non homogène

On se place ici en dimension 1 d'espace, $d = 1$ et on considère le problème suivant :

$$\begin{cases} u'' = f & \text{sur }]0\,;1[\\ u(0) = a \\ u(1) = b \end{cases} \tag{3.11}$$

où a et b sont des réels donnés. Ces conditions aux limites sont dites de type Dirichlet non homogène ; comme a et b ne sont pas forcément nuls, on cherche une solution dans $H^1(\Omega)$ et non plus dans $H_0^1(\Omega)$. Cependant, pour se ramener à l'espace $H_0^1(\Omega)$ on va utiliser une technique dite de *relèvement* ; on va s'assurer en particulier que le problème est bien posé grâce au théorème de Lax–Milgram et à la coercivité de la forme bilinéaire

$$a(u,v) = \int_\Omega u'(x)\ v'(x)\,\mathrm{d}x \quad \text{sur } H_0^1(\Omega).$$

On pose $u = u_0 + \widetilde{u}$ où u_0 est définie par $u_0(x) = a + (b-a)x$. On a en particulier $u_0(0) = a$ et $u_0(1) = b$. On a alors $\widetilde{u}(0) = 0$ et $\widetilde{u}(1) = 0$. La fonction $\widetilde{u}$ vérifie donc le système :

$$\begin{cases} -\widetilde{u}'' = f, \\ \widetilde{u}(0) = 0, \\ \widetilde{u}(1) = 0, \end{cases}$$

dont on connait la formulation faible et dont on a vu qu'il est bien posé au paragraphe 3.1.1. Il existe donc une unique fonction $u \in H^1(\Omega)$ vérifiant $u = u_0 + \widetilde{u}$ où $\widetilde{u} \in H_0^1(\Omega)$ est l'unique solution du problème

$$\int_0^1 \widetilde{u}'(x)v'(x)\,\mathrm{d}x = \int_0^1 f(x)v(x)\,\mathrm{d}x \qquad \forall v \in H_0^1(]0\,;1[).$$

De manière plus générale, soit $u_1 \in H^1_{a,b}(]0\,;1[) = \{v \in H^1\,;\, v(0) = a \text{ et } v(1) = b\}$, et soit $\bar{u} \in H^1_0(]0\,;1[)$ l'unique solution faible du problème :

$$\begin{cases} -\bar{u}'' = u_1'' + f, \\ \bar{u}(0) = 0, \\ \bar{u}(1) = 0. \end{cases}$$

Alors $\bar{u} + u_1$ est l'unique solution faible de (3.11), c'est-à-dire la solution du problème

$$\begin{cases} u \in H^1_{a,b}(]0\,;1[) \\ \displaystyle\int_0^1 u'(x)v'(x)\,\mathrm{d}x = \int_0^1 f(x)v(x)\,\mathrm{d}x, \quad \forall v \in H^1_0(]0\,;1[). \end{cases}$$

Il est facile de montrer que u ne dépend pas du relèvement choisi, voir l'exercice 40.

Considérons maintenant le cas de la dimension d'espace $d = 2$. Soit Ω un ouvert borné de $\mathbb{R}^d$, on considère le problème :

$$\begin{cases} -\Delta u = f & \text{dans } \Omega, \\ u = g & \text{sur } \partial\Omega. \end{cases} \tag{3.12}$$

Pour se ramener au problème de Dirichlet homogène, on veut construire un relèvement, c'est-à-dire une fonction $u_0 \in H^1(\Omega)$ tel que $\gamma(u_0) = g$ où γ est l'application trace. On ne peut plus le faire de manière explicite comme en dimension 1. En particulier, on rappelle qu'en dimension 2, l'espace $H^1(\Omega)$ n'est pas inclus dans l'espace $\mathcal{C}(\bar{\Omega})$ des fonctions continues, contrairement au cas de la dimension 1. Mais par le théorème de trace (théorème 3.1), si $g \in H^{1/2}(\partial\Omega)$, il existe $u_0 \in H^1(\Omega)$ tel que $g = \gamma(u_0)$. On cherche donc u sous la forme $u = \widetilde{u} + u_0$ avec $\widetilde{u} \in H^1_0(\Omega)$ et $u_0 \in H^1(\Omega)$ telle que $\gamma(u_0) = g$. Soit $v \in H^1_0(\Omega)$; on multiplie (3.12) par v et on intègre sur Ω :

$$\int_\Omega -\Delta u(x)v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x,$$

c'est-à-dire :

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

Comme $u = u_0 + \widetilde{u}$, on a donc :

$$\begin{cases} \widetilde{u} \in H^1_0(\Omega), \\ \displaystyle\int_\Omega \nabla\widetilde{u}(x) \cdot \nabla v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x - \int \nabla u_0(x) \cdot \nabla v(x)\,\mathrm{d}x \quad \forall v \in H^1_0(\Omega) \end{cases} \tag{3.13}$$

Cependant, en dimension supérieure ou égale à 2, il souvent difficile de construire le relèvement u_0. Il est donc usuel, dans la mise en œuvre des méthodes d'approximation (par exemple par éléments finis), de servir de de la formulation suivante, qui est équivalente à la formulation (3.13) :

$$\begin{cases} u \in \{v \in H^1(\Omega); \gamma(v) = g \text{ sur } \partial\Omega\}, \\ \displaystyle\int_\Omega \nabla u(x)\nabla v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x \quad \forall v \in H^1_0(\Omega). \end{cases}$$

3.1.3 Problème avec conditions aux limites de Fourier

On considère ici le problème de diffusion avec conditions aux limites de type Fourier (ou Robin dans la littérature anglo-saxonne).

$$\begin{cases} -\Delta u = f & \text{dans } \Omega, \\ \nabla u \cdot \boldsymbol{n} + \lambda u = 0 & \text{sur } \partial\Omega, \end{cases} \tag{3.14}$$

où Ω est un ouvert borné de $\mathbb{R}^d$, $d = 1$, 2 ou 3 et $\partial\Omega$ sa frontière ; $f \in \mathcal{C}^2(\bar{\Omega})$; $\boldsymbol{n}$ est le vecteur unitaire normal à $\partial\Omega$, extérieur à Ω et $\lambda(x) > 0$, $\forall x \in \partial\Omega$, est un coefficient qui modélise par exemple un transfert thermique à la paroi. Supposons qu'il existe $u \in \mathcal{C}^2(\bar{\Omega})$ vérifiant (3.14). Soit $\varphi \in \mathcal{C}^\infty(\bar{\Omega})$ une *fonction test*. On multiplie formellement (3.14) par φ et on intègre sur Ω. On obtient :

$$-\int_\Omega \Delta u(x)\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x.$$

Par intégration par parties, on a alors

$$\int_\Omega \nabla u(x)\nabla\varphi(x)\,\mathrm{d}x - \int_{\partial\Omega} \nabla u(x)\cdot\boldsymbol{n}(x)\varphi(x)\,\mathrm{d}\gamma(x) = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x$$

Notons que la fonction φ n'est pas à support compact et que la condition aux limites

$$\nabla u\cdot\boldsymbol{n} = -\lambda u$$

va donc intervenir dans cette formulation. En remplaçant on obtient :

$$\int_\Omega \nabla u(x)\cdot\nabla\varphi(x)\,\mathrm{d}x + \int_{\partial\Omega}\lambda u(x)\varphi(x)d\gamma(x) = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x, \qquad \forall\,\varphi\in C^\infty(\bar{\Omega}).$$

Par densité de $\mathcal{C}^\infty(\bar{\Omega})$ dans $H^1(\Omega)$, on a donc également

$$\int_\Omega \nabla u(x)\cdot\nabla\varphi(x)\,\mathrm{d}x + \int_{\partial\Omega}\lambda u(x)\varphi(x) = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x \qquad \forall\,\varphi\in H^1(\Omega).$$

Définition 3.9 — Solution faible. On dit que u est solution faible de (3.14) si u est solution de :

$$\begin{cases} u\in H^1(\Omega) \\ \displaystyle\int_\Omega \nabla u(x)\cdot\nabla v(x) + \,\mathrm{d}x\int_{\partial\Omega}\lambda(x)u(x)v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x \quad \forall v\in H^1(\Omega) \end{cases} \tag{3.15}$$

On peut remarquer que sous les hypothèses $f \in L^2(\Omega)$ et $\lambda \in L^\infty(\partial\Omega)$, toutes les intégrales de (3.15) sont bien définies. On rappelle que si $\varphi \in L^2(\Omega)$ et $\psi \in L^2(\Omega)$, alors $\varphi\psi \in L^1(\Omega)$. Pour vérifier que le problème (3.15) est bien posé, on a envie d'appliquer le théorème de Lax–Milgram. Définissons pour cela la forme bilinéaire $a : H^1(\Omega)\times H^1(\Omega)\to\mathbb{R}$ par :

$$a(u,v) = \int_\Omega \nabla u(x)\cdot\nabla v(x)\,\mathrm{d}x + \int \lambda(x)u(x)v(x)\,\mathrm{d}x. \tag{3.16}$$

Il est facile de voir que a est une forme bilinéaire symétrique. On peut donc lui associer une forme quadratique définie par :

$$E(v) = \frac{1}{2}\int_\Omega \nabla v(x)\cdot\nabla v(x)\,\mathrm{d}x + \int_{\partial\Omega}\lambda(x)v^2(x)\,\mathrm{d}\gamma(x) - \int_\Omega f(x)v(x)\,\mathrm{d}x. \tag{3.17}$$

Définition 3.10 — Solution variationnelle. On dit que u est solution variationnelle de (3.14) si u vérifie :

$$\begin{cases} u\in H^1(\Omega), \\ E(u)\leqslant E(v) \quad \forall v\in H^1(\Omega), \end{cases} \tag{3.18}$$

où E est défini par (3.17).

Lemme 3.11 On suppose que $\lambda \in L^\infty(\partial\Omega)$. Alors la forme bilinéaire définie par (3.16) est continue sur $H^1(\Omega)\times H^1(\Omega)$.

Démonstration. On a :

$$\begin{aligned} a(u,v) &= \int_\Omega \nabla u(x)\nabla v(x)\,\mathrm{d}x + \int_{\partial\Omega}\lambda(x)u(x)v(x)\,\mathrm{d}\gamma(x) \\ &\leqslant \|\nabla u\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)} + \|\lambda\|_{L^\infty(\partial\Omega)}\|u\|_{L^2(\partial\Omega)}\|v\|_{L^2(\partial\Omega)} \end{aligned} \tag{3.19}$$

Or par le théorème de trace 3.1 et plus particulièrement grâce à la continuité de la trace (3.5), on a

$$\|u\|_{L^2(\partial\Omega)} \leqslant C\|u\|_{H^1(\Omega)}.$$

On en déduit que

$$a(u,v) \leqslant M\|u\|_{H^1}\|v\|_{H^1}$$

avec $M = 1 + \mathcal{C}^2\|\lambda\|_{L^\infty}(\partial\Omega)$. Donc a est bilinéaire continue. •

Lemme 3.12 Soit $\lambda \in L^\infty(\partial\Omega)$ tel qu'il existe $\underline{\lambda} > 0$ tel que $\lambda(x) \geqslant \underline{\lambda}$ p.p. sur $\partial\Omega$. Alors la forme bilinéaire a définie par (3.16) est coercitive.

Démonstration. Montrons qu'il existe $\alpha > 0$ tel que $a(v,v) \geqslant \alpha\|v\|^2$, pour tout $v \in H^1$ où

$$a(v,v) = \int_\Omega \nabla v(x) \cdot \nabla v(x)\,\mathrm{d}x + \int_\Omega \lambda(x)v^2(x)\,\mathrm{d}\gamma(x).$$

Attention, comme $v \in H^1(\Omega)$ et non $H^1_0(\Omega)$, on ne peut pas écrire l'inégalité de Poincaré, qui nous permettrait de minorer $\int_\Omega \nabla v(x) \cdot \nabla v(x)\,\mathrm{d}x$. On va montrer l'existence de α par l'absurde. On suppose que a n'est pas coercive, c'est-à-dire que :

$$\forall \alpha > 0, \quad \exists\, v \in H^1(\Omega); \quad a(v,v) < \alpha\|v\|^2.$$

On a donc en particulier, en prenant $\alpha = 1/n$:

$$\forall n \in \mathbb{N}, \quad \exists\, v_n \in H^1(\Omega); \quad a(v_n,v_n) < \frac{1}{n}\|v_n\|^2_{H^1}$$

Dans cette dernière assertion, on peut prendre v_n de norme 1, puisque l'inégalité est homogène de degré 2. On a donc :

$$\forall n \in \mathbb{N}, \quad \exists\, v_n \in H^1(\Omega); \quad \|v_n\|_{H^1(\Omega)} = 1; \quad a(v_n,v_n) < \frac{1}{n}$$

Or, par le théorème de Rellich, toute suite bornée $(v_n)_{n\in\mathbb{N}}$ de $H^1(\Omega)$, est relativement compacte dans $L^2(\Omega)$. Comme on a $\|v_n\|_{H^1(\Omega)} = 1$, il existe donc une sous-suite encore notée $(v_n)_{n\in\mathbb{N}} \subset H^1(\Omega)$ telle que v_n converge vers v dans $L^2(\Omega)$ lorsque n tend vers $+\infty$. De plus, comme :

$$a(v_n,v_n) = \int_\Omega \nabla v_n(x) \cdot \nabla v_n(x)\,\mathrm{d}x + \int_{\partial\Omega} \lambda v_n(x)v_n(x)\,\mathrm{d}x < \frac{1}{n} \to 0 \text{ lorsque } n \to +\infty$$

On en déduit que, chaque terme étant positif :

$$\int_\Omega \nabla v_n(x) \cdot \nabla v_n(x)\,\mathrm{d}x \to_{n\to+\infty} 0 \tag{3.20}$$

$$\int_{\partial\Omega} \lambda v_n(x)v_n(x)\,\mathrm{d}x \to_{n\to+\infty} 0 \tag{3.21}$$

On a donc : $\nabla v_n \to 0$ dans $L^2(\Omega)$ lorsque $n \to +\infty$. On en déduit que

$$\int_\Omega \partial_i v_n(x)\varphi\,\mathrm{d}x \to 0 \text{ lorsque } n \to +\infty, \text{ pour } i = 1,\dots,d.$$

Donc par définition de la dérivée faible (1.59), on a aussi

$$\int_\Omega v_n(x)\partial_i\varphi(x)\,\mathrm{d}x \to 0 \text{ lorsque } n \to +\infty.$$

Comme $v_n \to v$ dans $L^2(\Omega)$ lorsque $n \to +\infty$, on peut passer à la limite et écrire que $\int_\Omega v(x)\partial_i\varphi(x) = 0$. On en déduit que la dérivée faible $D_i v$ existe et est nulle dans Ω. La fonction v est donc constante par composante connexe. Mais par (3.21), on a $v = 0$ sur $\partial\Omega$ et la trace d'une fonction constante est la constante elle-même. On a donc

$$v = 0 \text{ dans } \Omega$$

On a ainsi montré que

$$D_i v_n \to D_i v \text{ et } v_n \to v = 0 \text{ dans } L^2(\Omega) \text{ lorsque } n \to +\infty.$$

Par conséquent, $v_n \to 0$ dans $H^1(\Omega)$ lorsque $n \to +\infty$, ce qui contredit le fait que $\|v_n\|_{H^1(\Omega)} = 1$. On a ainsi montré la coercivité de a. •

Proposition 3.13 Soit $f \in L^2(\Omega)$ et $\lambda \in L^\infty(\Omega)$ tel que $\lambda \geqslant \underline{\lambda}$ p.p. avec $\underline{\lambda} > 0$ alors il existe un unique u solution de (3.15) qui est aussi l'unique solution de (3.18).

3.1.4 Condition de Neumann

Considérons maintenant le problème (3.14) avec $\lambda = 0$, on obtient le problème :

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ \dfrac{\partial u}{\partial n} = 0 & \text{sur } \partial\Omega \end{cases} \tag{3.22}$$

qu'on appelle problème de Dirichlet avec conditions de Neumann homogènes. En intégrant la première équation du système, il est facile de voir qu'une condition nécessaire d'existence d'une solution de (3.22) est que :

$$\int_\Omega -\Delta u(x)\,\mathrm{d}x = \int_{\partial\Omega} \frac{\partial u}{\partial n}(x)\,\mathrm{d}x = \int_\Omega f(x)\,\mathrm{d}x = 0$$

Si la condition aux limites de Neumann est non-homogène, soit :

$$\frac{\partial u}{\partial n} = g$$

la condition de compatibilité devient

$$\int_\Omega f(x)\,\mathrm{d}x + \int_{\partial\Omega} g(x)\ \mathrm{d}\gamma(x) = 0.$$

Remarquons que si $\alpha = 0$, la forme bilinéaire est

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

et que celle-ci n'est pas coercive sur $H^1(\Omega)$. De fait, il est clair que la solution de (3.22) n'est pas unique, puisque si u est solution de (3.22) alors $u + c$ est aussi solution, pour tout $c \in \mathbb{R}$. Pour éviter ce problème on va chercher les solutions de (3.22) à moyenne nulle. On cherche donc à résoudre (3.22) dans l'espace

$$H = \Big\{ v \in H^1(\Omega); \int_\Omega v(x)\,\mathrm{d}x = 0 \Big\}$$

On admettra que a est coercive sur H (ceci est vrai grâce à l'inégalité de Poincaré-Wirtinger [3] [4]). Le problème

$$\begin{cases} u \in H, \\ a(u,v) = \displaystyle\int fv \quad \forall v \in H, \end{cases}$$

admet donc une unique solution.

3.1.5 Formulation faible et formulation variationnelle

Nous donnons ici un exemple de problème pour lequel on peut établir une formulation faible, mais pas variationnelle. On se place en une dimension l'espace $N = 1$ et on considère $\Omega =]0\,;1[$ et $f \in L^2(]0\,;1[)$. On s'intéresse au problème dit *d'advection diffusion* :

$$\begin{cases} -u'' + u' = f, & \text{dans }]0\,;1[\\ u(0) = u(1) = 0. \end{cases} \tag{3.23}$$

Cherchons une formulation faible. Par la même méthode qu'au paragraphe 3.1.1, on choisit $v \in H_0^1(\Omega)$, on multiplie (3.23) par v et on intègre par parties :

$$\int_\Omega u'(x)v'(x)\,\mathrm{d}x + \int_\Omega u'(x)v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x.$$

Il est donc naturel de poser :

$$a(u,v) = \int_\Omega u'(x)v'(x)\,\mathrm{d}x + \int_\Omega u'(x)v(x)\,\mathrm{d}x \quad \text{et} \quad T(v) = \int_\Omega f(x)v(x)\,\mathrm{d}x.$$

3. L'inégalité de Poincaré-Wirtinger s'énonce de la façon suivante : soit Ω un ouvert borné de $\mathbb{R}^d$ de frontière lipschitzienne, alors il existe $C \in \mathbb{R}_+$, ne dépendant que de Ω tel que pour tout $u \in H^1(\Omega)$, on a :

$$\|u\|^2_{L^2(\Omega)} \leqslant C|u|^2_{H^1(\Omega)} + 2(\mathrm{m}(\Omega))^{-1}\Big(\int_\Omega u(x)\,\mathrm{d}x\Big)^2.$$

4. Wilhelm Wirtinger, 1865-1945, mathématicien autrichien, dont les travaux portent sur plusieurs domaines des mathématiques.

Il est évident que T est une forme linéaire continue sur $H_0^1(\Omega)$ (c'est-à-dire $T \in H^{-1}(\Omega)$) et que la forme a est bilinéaire continue, mais pas symétrique. De plus elle est coercive. En effet, on a :

$$a(u,u) = \int_\Omega u'^2(x)\,\mathrm{d}x + \int_\Omega u'(x)u(x)\,\mathrm{d}x = \int_\Omega u'^2(x)\,\mathrm{d}x + \int_\Omega \frac{1}{2}(u^2)'(x)\,\mathrm{d}x.$$

Or, comme $u \in H_0^1(\Omega)$, on a $u = 0$ sur $\partial\Omega$ et donc

$$\int_\Omega (u^2)'(x)\,\mathrm{d}x = u^2(1) - u^2(0) = 0.$$

On en déduit que :

$$a(u,u) = \int_0^1 (u')^2\,\mathrm{d}x,$$

et par l'inégalité de Poincaré (1.31), on conclut que a est coercive sur $H_0^1(\Omega)$. On en déduit par le théorème de Lax-Milgram, l'existence et l'unicité de $u \in H_0^1(]0\,;1[)$ solution du problème :

$$\int_0^1 (u'(x)v'(x) + u'(x)v(x))\,\mathrm{d}x = \int_0^1 f(x)v(x)\,\mathrm{d}x.$$

3.2 Méthodes de Ritz et Galerkin

3.2.1 Principe général de la méthode de Ritz

On se place sous les hypothèses suivantes :

$$\begin{cases} H \text{ est un espace de Hilbert} \\ a \text{ est une forme bilinéaire continue, coercive et symétrique} \\ T \in H' \end{cases} \tag{3.24}$$

On cherche à calculer $u \in H$ telle que :

$$a(u,v) = T(v) \qquad \forall v \in H, \tag{3.25}$$

ce qui revient à calculer $u \in H$ solution du problème du problème de minimisation (3.9), avec J définie par (3.10). L'idée de la méthode de Ritz [5] est de remplacer H par un espace $H_N \subset H$ de dimension finie (où $\dim H_N = N$) et de calculer u_N solution de

$$\begin{cases} u_N \in H_N \\ J(u_N) \leqslant J(v) \quad \forall v \in H_N \end{cases} \tag{3.26}$$

en espérant que u_N soit "proche" (en un sens à définir) de u.

Théorème 3.5 Sous les hypothèses (3.24), si H_N est un sous-espace vectoriel de H et $\dim H_N < +\infty$ alors le problème (3.26) admet une unique solution.

Démonstration. Puisque H_N est un espace de dimension finie inclus dans H, c'est donc aussi un espace de Hilbert. On peut donc appliquer le théorème de Lax–Milgram et on en déduit l'existence et l'unicité de $u_N \in H_N$ solution de (3.26), qui est aussi solution de $a(u_N, v) = T(v)$, $\forall v \in H_N$. •

Nous allons maintenant exposer une autre méthode de démonstration du théorème 3.5, qui a l'avantage d'être constructive et qui nous permet d'introduire les idées principales des méthodes numériques envisagées plus loin. Comme l'espace H_N considéré dans le théorème 3.26 est de dimension N, il existe une base $(\phi_1, \ldots, \phi_N)$. Si $u \in H_N$, on peut donc développer $u = \sum_{i=1}^N u_i\phi_i$. On note : $U = (u_1, \ldots, u_N) \in \mathbb{R}^N$. L'application ξ qui à u associe U est une bijection de H_N dans $\mathbb{R}^N$. On peut donc définir

$$j = J \circ \xi^{-1}, \text{ c.à.d. } j(U) = J(u), \text{ pour } U = (u_1, \ldots, u_N) \in \mathbb{R}^N \text{ et } u = \sum_{i=1}^N u_i\phi_i. \tag{3.27}$$

5. Walter Ritz, né le 22 février 1878 à Sion et mort le 7 juillet 1909 à Göttingen, est un physicien suisse. Il a inventé la méthode dite *de Ritz* dans le cadre du calcul des valeurs propres de l'opérateur bi-harmonique

On a donc

$$J(u) = \frac{1}{2}a\Big(\sum_{i=1}^{N} u_i\phi_i, \sum_{i=1}^{N} u_i\phi_i\Big) - T\Big(\sum_{i=1}^{N} u_i\phi_i\Big) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} u_i u_j a(\phi_i,\phi_j) - \sum_{i=1}^{N} u_i T(\phi_i),$$

et $J(u)$ peut s'écrire sous la forme :

$$J(u) = \frac{1}{2}U^t\mathcal{K}U - U^t\mathcal{G} = j(U),$$

où $\mathcal{K} \in M^{N,N}(\mathbb{R})$ est définie par $\mathcal{K}_{ij} = a(\phi_i,\phi_j)$ et où $\mathcal{G}_i = T(\phi_i)$. Chercher u_N solution de (3.26) est donc équivalent à chercher $U \in \mathbb{R}^N$ solution de :

$$j(U) \leqslant j(V) \quad \forall V \in \mathbb{R}^N, \quad \text{avec} \quad j(V) = \frac{1}{2}V^t\mathcal{K}V - V^t\mathcal{G}. \tag{3.28}$$

Il est facile de vérifier que la matrice $\mathcal{K}$ est symétrique définie positive. Donc j est une fonctionnelle quadratique sur $\mathbb{R}^N$ et on a donc existence et unicité de $U \in \mathbb{R}^N$ tel que $j(U) \leqslant j(V)$, $\forall V \in \mathbb{R}^N$. La solution du problème de minimisation (3.28) est aussi la solution du système linéaire $\mathcal{K}U = \mathcal{G}$; on appelle souvent $\mathcal{K}$ la matrice de rigidité. Le résultat suivant est une conséquence immédiate du résultat général de minimisation dans $\mathbb{R}^N$ (voir par exemple [12, chapitre 3]).

Proposition 3.14 — Existence et unicité de la solution du problème de minimisation. Soit $j : \mathbb{R}^N \to \mathbb{R}$ définie par (3.27). Il existe un unique $u \in \mathbb{R}^N$ solution du problème de minimisation (3.28).

Résumé sur la technique de Ritz

1. On se donne $H_N \subset H$.
2. On trouve une base de H_N.
3. On calcule la matrice de rigidité $\mathcal{K}$ et le second membre $\mathcal{G}$. Les coefficients de $\mathcal{K}$ sont donnés par $\mathcal{K}_{ij} = a(\phi_i,\phi_j)$.
4. On minimise j par la résolution de $\mathcal{K}V = \mathcal{G}$.
5. On calcule la solution approchée $u^{(N)} = \sum_{i=1}^{N} u_i\phi_i$.

On appelle H_N l'espace d'approximation. Le choix de cet espace sera fondamental pour le développement de la méthode d'approximation. Le choix de H_N est formellement équivalent au choix de la base $(\phi_i)_{i=1,\dots,N}$. Pourtant, le choix de cette base est capital même si $u^{(N)}$ ne dépend que du choix de H_N et pas de la base.

Choix de la base Un premier choix consiste à choisir des bases de façon récursive, selon le schéma :

$$\{ \text{ base de } H_{N+1}\} = \{ \text{ base de } H_N\} \cup \{\phi_{N+1}\}.$$

Les bases sont donc emboitées les unes dans les autres. Considérons par exemple $H = H^1(]0\,;1[)$, et l'espace d'approximation $H_N = \text{Vect}\{\phi_1,\phi_2,\dots,\phi_{N-1}\}$, avec $\phi_i(x) = x^{i-1}$, $i = 1,\dots,N$. On peut remarquer la matrice de rigidité $\mathcal{K}$ résultante est une matrice pleine. Or, on veut justement éviter les matrices pleines car les systèmes linéaires associés sont coûteux (en temps et mémoire) à résoudre.

Le choix idéal serait de choisir une base (ϕ_i), $i = 1,\dots,N$ de telle sorte que

$$a(\phi_i,\phi_j) = \lambda_i\delta_{ij} \quad \text{où} \quad \delta_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{sinon.} \end{cases} \tag{3.29}$$

On a alors $\mathcal{K} = \text{diag}(\lambda_1,\dots,\lambda_N)$ et on a explicitement :

$$u^{(N)} = \sum_{i=1}^{N} \frac{T(\phi_i)}{a(\varphi_i,\varphi_i)}\phi_i.$$

Considérons par exemple le problème de Dirichlet (3.1). Si ϕ_i est la i–ème fonction propre de l'opérateur $-\Delta$ avec conditions aux limites de Dirichlet, associée à la valeur propre λ_i, on obtient bien la propriété souhaitée, car la matrice $\mathcal{K}$ est alors $\mathrm{diag}(\lambda_1, \ldots, \lambda_N)$. Malheureusement, il est rare que l'on puisse connaître explicitement les fonctions propres de l'opérateur concerné.

Un deuxième choix consiste à choisir des bases telles que la base de H_N n'est pas forcément incluse dans celle de H_{N+1}. La technique des éléments finis qu'on verra au chapitre suivant, est un exemple de ce choix. La matrice $\mathcal{K}$ obtenue n'est pas diagonale, mais elle est creuse, c'est-à-dire qu'un grand nombre de ses coefficients sont nuls. Par exemple, pour des éléments finis appliqués à un opérateur du second ordre, on peut avoir un nombre de coefficients non nuls de l'ordre de $0(N)$.

Convergence de l'approximation de Ritz Une fois qu'on a calculé u_N solution de (3.28), il faut se préocupper de savoir si $u^{(N)}$ est une bonne approximation de u solution de (3.25), c'est-à-dire de savoir si $u^{(N)} \to u$ lorsque $N \to +\infty$. Pour vérifier cette convergence, on va se servir de la notion de consistance.

Définition 3.15 — Consistance. Sous les hypothèses (3.24), on dit que l'approximation de Ritz définie par l'espace $H_N \subset H$ avec $\dim H_N = N < +\infty$ est consistante si $d(H, H_N)$ tend vers 0 lorsque $N \to +\infty$, c'est-à-dire $d(w, H_N) \to_{N\to+\infty} 0$, $\forall w \in H$ ou encore $\inf_{v\in H_N} \|w-v\| \to_{N\to+\infty} 0$, $\forall w \in H$.

L'autre notion fondamentale pour prouver la convergence est la stabilité, elle-même obtenue grâce à la propriété de coercivité de a. Par stabilité, on entend estimation a priori sur la solution approchée, c'est-à-dire une estimation sur la solution approchée avant même de savoir si elle existe ; la solution approchée $u^{(N)}$ est solution de (3.28) ou encore de :

$$\begin{cases} a(u^{(N)}, v) = T(v) \quad \forall v \in H_N \\ u^{(N)} \in H_N \end{cases} \tag{3.30}$$

On a l'estimation a priori suivante sur u_N :

Proposition 3.16 — Stabilité. Sous les hypothèses du théorème 3.24, on a :

$$\|u^{(N)}\|_H \leqslant \frac{\|T\|_{H'}}{\alpha}.$$

Démonstration. Le caractère coercif de a nous permet d'écrire $\alpha\|u^{(N)}\|^2 \leqslant a(u^{(N)}, u^{(N)})$. Or comme $u^{(N)}$ est solution de (3.30), on a $a(u^{(N)}, u^{(N)}) = T(u^{(N)})$. Comme T est linéaire continue, on obtient $T(u^{(N)}) \leqslant \|T\|_{H'}\|u^{(N)}\|_H$. •

Il est naturel de s'intéresser à l'erreur que l'on commet lorsqu'on remplace la résolution de la formulation faible (3.25) par la résolution de la formulation faible en dimension finie (3.30). Il nous faut pour cela pouvoir quantifier la norme de l'erreur commise $\|u - u^N\|_H$. Un premier pas dans cette direction est le lemme suivant, souvent appelé lemme de Céa [6] qui permet de contrôler l'erreur de discrétisation par l'erreur d'interpolation, définie comme la distance entre $u \in H$ solution de (3.25) et l'espace H_N d'approximation.

Définition 3.17 — Erreur d'interpolation, erreur de discrétisation. Soit H un espace de Hilbert réel et a une forme bilinéaire continue symétrique coercive. Soit T une forme linéaire continue et $T \in H'$, et soit $M > 0$ et $\alpha > 0$ tels que $a(u, v) \leqslant M\|u\|_H\|v\|_H$ et $a(u, u) \geqslant \alpha\|u\|_H^2$. Soit $u \in H$ l'unique solution du problème :

$$a(u, v) = T(v) \qquad \forall v \in H \tag{3.31}$$

On appelle *erreur d'interpolation* la distance $\mathrm{d}(u, H_N)$ entre la solution du problème continu (3.31) et l'espace d'approximation H_N, définie par $\mathrm{d}(u, H_N) = \inf\{\|u - v\|_H, v \in h_N\}$. Soit $H_N \subset H$ tel que $\dim H_N = N$ et soit $u^{(N)} \in H_N$ l'unique solution de

$$a(u^{(N)}, v) = T(v) \qquad \forall v \in H_N$$

6. Jean Céa, mathématicien français contemporain, voir `http://www.jean-cea.fr`

On appelle *erreur de discrétisation* la quantité $\|u - u^{(N)}\|_H$.

Lemme 3.18 — Lemme de Céa. Sous les hypothèses de la définition 3.17,

$$\|u - u^{(N)}\|_H \leqslant \sqrt{\frac{M}{\alpha}} d(u, H_N). \tag{3.32}$$

Démonstration. On procède en deux étapes.

1. On va montrer que $u^{(N)}$ est la projection de u sur H_N pour le produit scalaire $(\,\cdot\,,\,\cdot\,)_a$ induit par a, défini de $H \times H$ par $(u, v)_a = a(u, v)$. On note $\|u\|_a = \sqrt{a(u,u)}$, la norme induite par le produit scalaire a. La norme $\|\cdot\|_a$ est équivalente à la norme $\|\cdot\|_H$: en effet, grâce à la coercivité et la continuité de la forme bilinéaire a, on peut écrire $\alpha\|u\|_H^2 \leqslant \|u\|_a^2 \leqslant M\|u\|_H^2$. Donc $(H, \|\cdot\|_a)$ est un espace de Hilbert. Soit u la solution de (3.31) et soit $v = P_{H_N} u$ la projection orthogonale de u sur H_N relative au produit scalaire $a(\,\cdot\,,\,\cdot\,)$. Par définition de la projection orthogonale, on a donc $v - u \in H_N^\perp$. Soit encore $a(v - u, w) = 0$, $\forall w \in H_N$. On en déduit que $a(v, w) = a(u, w) = T(w)$, $\forall w \in H_N$ et donc que $v = u^{(N)}$. On a donc montré que $u^{(N)}$ est la projection orthogonale de v sur H_N, c'est-à-dire $u^{(N)} = P_{H_N} u$.
2. On va établir une estimation de la norme de la différence entre u et u_N ; par définition de P_{H_N}, on a :
$$\|u - P_{H_N} u\|_a^2 \leqslant \|u - v\|_a^2 \qquad \forall v \in H_N,$$
ce qui s'écrit endore (puisque $P_{H_N} u = u^{(N)}$) :
$$a(u - u^{(N)}, u - u^{(N)}) \leqslant a(u - v, u - v) \qquad \forall v \in H_N.$$
Par coercivité et continuité de la forme bilinéaire a, on a donc :
$$\alpha\|u - u^{(N)}\|_H^2 \leqslant a(u - u^{(N)}, u - u^{(N)}) \leqslant a(u - v, u - v) \leqslant M\|u - v\|_H^2 \qquad \forall\, v \in H_N.$$
On en déduit que :
$$\|u - u^{(N)}\|_H \leqslant \sqrt{\frac{M}{\alpha}} \|u - v\|_H \qquad \forall v \in H_N.$$
En passant à l'inf sur v, on obtient alors :
$$\|u - u^{(N)}\|_H \leqslant \sqrt{\frac{M}{\alpha}} \inf_{v \in H_N} \|u - v\|_H$$
Ce qui est exactement (3.32).

•

3.2.2 Méthode de Galerkin

On se place maintenant sous les hypothèses suivantes :

$$\begin{cases} H \text{ est un espace de Hilbert} \\ a \text{ est une forme bilinéaire continue et coercive} \\ T \in H' \end{cases} \tag{3.33}$$

Remarquons que maintenant, a n'est pas nécessairement symétrique, les hypothèses (3.33) sont donc plus générales que les hypothèses (3.24). On considère le problème

$$\begin{cases} u \in H \\ a(u, v) = T(v) \quad v \in H \end{cases} \tag{3.34}$$

Par le théorème de Lax–Milgram, il y a existence et unicité de $u \in H$ solution de (3.34). Le principe de la méthode de Galerkin [7] est similaire à celui de la méthode de Ritz. On se donne $H_N \subset H$, tel que $\dim H_N < +\infty$ et on cherche à résoudre le problème approché :

$$(P_N) \begin{cases} u^{(N)} \in H_N \\ a(u^{(N)}, v) = T(v) \quad \forall v \in H_N \end{cases} \tag{3.35}$$

7. Boris Grigoryevich Galerkin, né le 20 février 1871 à Polotsk en Biélorussie et mort le 12 juillet 1945, est un mathématicien et un ingénieur russe réputé pour ses contributions à l'étude des treillis de poutres et des plaques élastiques. Son nom reste lié à une méthode de résolution approchée des structures élastiques, qui est l'une des bases de la méthode des éléments finis.

Remarque 3.19 — Orthogonalité de l'erreur avec l'espace d'approximation. Comme $H_N \subset H$, on peut choisir $v \in H_N$ dans (3.34) ; en soustrayant (3.35) à cette équation, on obtient

$$a(u - u^{(N)}, v) = 0, \ \forall v \in H_N. \tag{3.36}$$

La différence entre u et son approximation $u^{(N)}$ est donc orthogonale à l'espace d'approximation.

Par le théorème de Lax–Milgram, on a immédiatement :

Théorème 3.6 — Existence et unicité, méthode de Galerkin. Sous les hypothèses $H_N \subset H$ et $\dim H_N = N$, il existe un unique $u^{(N)} \in H_N$ solution de (3.35).

Comme dans le cas de la méthode de Ritz, on va donner une autre méthode, constructive, de démonstration de l'existence et unicité de u_N qui permettra d'introduire la méthode de Galerkin. Comme $\dim H_N = N$, il existe une base $(\phi_1 \dots \phi_N)$ de H_N. Soit $v \in H_N$, on peut donc développer v sur cette base $: v = \sum_{i=1}^{N} v_i \phi_i$, et identifier v au vecteur $(v_1, \dots, v_N) \in \mathbb{R}^N$. En écrivant que u satisfait (3.35) pour tout $v = \phi_i$, $i = 1, \dots, N$:

$$a(u, \phi_i) = T(\phi_i) \qquad \forall i = 1, \dots, N,$$

et en développant u sur la base $(\phi_i)_{i=1,\dots,N}$, on obtient :

$$\sum_{j=1}^{N} a(\phi_j, \phi_i) u_j = T(\phi_i) \qquad \forall i = 1, \dots, N.$$

On peut écrire cette dernière égalité sous forme d'un système linéaire : $\mathcal{K}U = \mathcal{G}$,

$$\mathcal{K}_{ij} = a(\phi_j, \phi_i) \text{ et } \mathcal{G}_i = T(\phi_i), \text{ pour } i, j = 1, \dots, , N.$$

La matrice $\mathcal{K}$ n'est pas en général symétrique.

Proposition 3.20 — Existence et unicité de la solution du système linéaire. Sous les hypothèses du théorème 3.6 le système linéaire (3.2.2) admet une solution.

Démonstration. On va montrer que $\mathcal{K}$ est inversible en vérifiant que son noyau est réduit à $\{0\}$. Soit $w \in \mathbb{R}^N$ tel que $\mathcal{K}w = 0$. Décomposons w sur la base $(\phi_1, \dots, \phi_N)$ de H_N. On a donc $\sum_{j=1}^{N} a(\phi_j, \phi_i) w_j = 0$. Multiplions cette relation par w_i et sommons pour $i = 1$ à N, on obtient :

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a(\phi_j, \phi_i) w_j w_i = 0.$$

Soit encore $a(w, w) = 0$. Par coercitivité de a, ceci entraine que $w = 0$. On en déduit que $w_i = 0, \forall_i = 1, \dots, N$, ce qui achève la preuve. •

Remarque 3.21 Si a est symétrique, la méthode de Galerkin est équivalente à celle de Ritz.

En résumé, la méthode de Galerkin comporte les mêmes étapes que la méthode de Ritz, c'est-à-dire :

1. On se donne $H_N \subset H$
2. On trouve une base de H_N
3. On calcule $\mathcal{K}$ et $\mathcal{G}$
4. On résout $\mathcal{K}U = \mathcal{G}$
5. On écrit $u^{(N)} = \sum_{i=1}^{N} u_i \phi_i$.

La seule différence avec la méthode de Ritz est que l'étape 4 n'est pas issue d'un problème de minimisation. Comme pour la méthode de Ritz, il faut se poser la question du choix du sous-espace H_N et de sa base, ainsi que de la convergence de l'approximation de u solution de (3.34) par $u^{(N)}$ obtenue par la technique de Galerkin. En ce qui concerne le choix de la base $\{\phi_1, \dots, \phi_N\}$, les possibilités sont les mêmes que pour la méthode de Ritz (voir paragraphe 3.2.1). De même, la notion de consistance est identique à celle donnée pour la méthode de Ritz (définition 3.15) et la

démonstration de stabilité est identique à celles effectuée pour la méthode de Ritz (proposition 3.16).

Comme dans le cas de la méthode de Ritz, on a encore un contrôle de l'erreur de discrétisation $\|u - u^{(N)}\|_H$ par l'erreur d'interpolation $d(u, H_N)$, mais avec une constante plus grande que celle du lemme de Céa :

Lemme 3.22 Sous les hypothèses du théorème (3.6), si u est la solution de (3.34) et u_N la solution de (3.35), alors

$$\|u - u^{(N)}\|_H \leqslant \frac{M}{\alpha} d(u, H_N). \tag{3.37}$$

où M et α sont tels que : $\alpha\|v\|^2 \leqslant a(v,u) \leqslant M\|v\|^2$ pour tout v dans H (les réels M et α existent en vertu de la continuité et de la coercivité de a). Noter que la constante M/α est supérieure à la constante $\sqrt{M/\alpha}$ du lemme 3.18.

Démonstration. Comme la forme bilinéaire a est coercive de constante α, on a :

$$\alpha\|u - u^{(N)}\|_H^2 \leqslant a(u - u^{(N)}, u - u^{(N)})$$

On a donc, pour tout $v \in H$:

$$\alpha\|u - u^{(N)}\|_H^2 \leqslant a(u - u^{(N)}, u - v) + a(u - u^{(N)}, v - u^{(N)})$$

Or $a(u - u^{(N)}, v - u^{(N)}) = a(u, v - u^{(N)}) - a(u^{(N)}, v - u^{(N)})$ et par définition de u et $u^{(N)}$, on a :

$$a(u, v - u^{(N)}) = T(v - u^{(N)}) \text{ et } a(u^{(N)}, v - u^{(N)}) = T(v - u^{(N)}).$$

On en déduit que :

$$\alpha\|u - u^{(N)}\|_H^2 \leqslant a(u - u^{(N)}, u - v) \qquad \forall v \in H_N,$$

et donc, par continuité de la forme bilinéaire a :

$$\alpha\|u - u^{(N)}\|_H^2 \leqslant M\|u - u^{(N)}\|_H \|u - v\|_H.$$

On obtient donc :

$$\|u - u^{(N)}\|_H \leqslant \frac{M}{\alpha}\|u - v\|_H \qquad \forall v \in H_N,$$

ce qui entraine (3.37). •

Remarque 3.23 On peut remarquer que l'estimation (3.37) obtenue dans le cadre de la métode de Galerkin est moins bonne que l'estimation (3.32) obtenue dans le cadre de la méthode de Ritz. Ceci est moral, puisque la méthode de Ritz est un cas particulier de la méthode de Galerkin.

Grâce au théorème 3.22, on peut remarquer que $u^{(N)}$ converge vers u dans H lorsque N tend vers $+\infty$ dès que $d(u, H_N) \to 0$ lorsque $N \to +\infty$. C'est donc là encore une propriété de consistance dont nous avons besoin. La propriété de consistance n'est pas toujours facile à montrer directement. On utilise alors la caractérisation suivante :

Proposition 3.24 — Caractérisation de la consistance. Soit V un sous espace vectoriel de H dense dans H On suppose qu'il existe une fonction $r_N : V \to H_N$ telle que $\|v - r_N(v)\|_H \to 0$ lorque $N \to +\infty$; alors

$$d(u, H_N) \to 0 \text{ lorque } N \to +\infty.$$

Démonstration. Soient $v \in V$ et $w = r_N(v)$. Par définition, on a

$$d(u, H_N) \leqslant \|u - r_N(v)\|_H \leqslant \|u - v\|_H + \|v - r_N(v)\|_H$$

Comme V est dense dans H, pour tout $\varepsilon > 0$, il existe $v \in V$, tel que $\|u - v\|_H \leqslant \varepsilon$. Choisissons v qui vérifie cette dernière inégalité. Par hypothèse sur r_N,

$$\forall \varepsilon > 0,\ \exists N_0 \text{ tel que si } N \geqslant N_0 \text{ alors } \|v - r_N(v)\|_H \leqslant \varepsilon.$$

Donc si $N \geqslant N_0$, on a $d(u, H_N) \leqslant 2\varepsilon$. On en déduit que $d(u, H_N) \to 0$ quand $N \to +\infty$. •

3.2.3 Méthode de Petrov-Galerkin

La méthode de Petrov-Galerkin s'apparente à la méthode de Galerkin. On cherche toujours à résoudre :

$$\begin{cases} u \in H \\ a(u, v) = T(v) \quad \forall v \in H \end{cases} \tag{3.38}$$

Mais on choisit maintenant deux sous-espaces H_N et V_N de H, tous deux de même dimension finie $\dim H_N = \dim V_N = N$. On cherche une approximation de la solution du problème dans l'espace H_N et on choisit comme fonction test les fonctions de base de V_N. On obtient donc le système :

$$\begin{cases} u \in H_N \\ a(u,v) = T(v) \quad \forall v \in V_N \end{cases} \tag{3.39}$$

On appelle H_N l'espace d'approximation et V_N l'espace des fonctions test. Si $(\phi_1, \dots, \phi_N)$ est une base de H_N et $(\psi_1, \dots, \psi_N)$ une base de V_N, en développant $u^{(N)}$ sur la base de $(\phi_1, \dots, \phi_N)$. $u^{(N)} = \sum u_j \phi_j$ et en écrivant (3.39) pour $v = \psi_i$, on obtient :

$$\begin{cases} u \in H_N \\ a(u,\psi_i) = T(\psi_i) \qquad \forall i = 1, \dots, N \end{cases} \tag{3.40}$$

Le système à résoudre est donc :

$$\mathcal{K}U = \mathcal{G}$$

où $\mathcal{K}$ est une matrice carrée d'ordre N de coefficients $\mathcal{K}_{i,j} = a(\phi_j, \psi_i)$ et $\mathcal{G}_i = T(\psi_i)$, pour $i = 1, \dots, N$.

3.3 Méthode des éléments finis

La méthode des éléments finis est une façon particulière de choisir les bases des espaces d'approximation pour les méthodes de Ritz et Galerkin.

3.3.1 Principe

On se limitera dans le cadre de ce cours à des problèmes du second ordre. L'exemple type sera le problème de Dirichlet (3.1), qu'on rappelle ici :

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

et l'espace de Hilbert sera l'espace de Sobolev $H^1(\Omega)$ ou $H^1_0(\Omega)$.

On se limitera à un certain type d'éléments finis, dits *de Lagrange*. Donnons les principes généraux de la méthode.

Éléments finis de Lagrange Soient $\Omega \subset \mathbb{R}^2$ (ou $\mathbb{R}^3$) et H l'espace fonctionnel dans lequel on recherche la solution. Par exemple $H^1_0(\Omega)$ s'il s'agit du problème de Dirichlet (3.1). On cherche $H_N \subset H = H^1_0(\Omega)$ et les fonctions de base $\phi_1, \dots, \phi_N$. On va déterminer ces fonctions de base à partir d'un découpage de Ω en un nombre fini de cellules appelées *éléments*. la procédure est la suivante :

1. On construit un *maillage* $\mathcal{T}$ de Ω (en triangles ou rectangles) que l'on appelle *éléments* ; un élément générique sera noté K.
2. Dans chaque élément, on se donne des points que l'on appelle *nœuds*.
3. On définit H_N par $H_N = \{u : \Omega \to \mathbb{R} / u_{|K} \in \mathcal{P}_k, \forall K \in \mathcal{T}\} \cap H$ où $\mathcal{P}_k$ désigne l'ensemble des polynômes de degré inférieur ou égal à k. Le degré des polynômes est choisi de manière à ce que u soit entièrement déterminée par ses valeurs aux nœuds. Pour une méthode d'éléments finis de type Lagrange, les valeurs aux nœuds sont également les *degrés de liberté*, c'est-à-dire les valeurs qui déterminent entièrement la fonction recherchée.
4. On construit une base $\{\phi_i, \dots, \phi_N\}$ de H_N tel que le support de ϕ_i soit "le plus petit possible". Les fonctions ϕ_i sont aussi appelées fonctions de forme.

Remarque 3.25 — Éléments finis non conformes. Notons qu'on a introduit ici une méthode d'éléments finis dite *conforme*, au sens où l'espace d'approximation H_N est inclus dans l'espace H. Dans une méthode *non conforme*, on n'aura plus $H_N \subset H$ et par conséquence, on devra aussi

construire une forme bilinéaire approchée $a_{\mathcal{T}}$; on pourra voir à ce sujet l'exercice 50 où on exprime la méthode des volumes finis comme une méthode déléments finis non conformes.

Exemple en dimension 1 Soit $\Omega =]0\,;1[\subset \mathbb{R}$ et soit $H = H_0^1([0\,;1[)$; on cherche un espace H_N d'approximation de H. Pour cela, on divise l'intervalle $]0\,;1[$ en N intervalles de longueur $h = 1/(N+1)$. On pose $x_i = i$, $i = 0, N+1$. Les étapes 1. à 4. décrites précédemment donnent dans ce cas :

1. Construction des éléments : on construit $n+1$ éléments $K_i =]x_i\,;x_{i+1}[$, $i = 0, \ldots, N$.
2. Nœuds : on a deux nœuds par élément ; x_i et x_{i+1} sont les nœuds de K_i, $i = 0, \ldots, N$. Le fait que $H_N \subset H_0^1(]0\,;1[)$ impose que les fonctions de H_N soient nulles en $x_0 = 0$ et $x_{N+1} = 1$. On appelle $x_1, \ldots, x_N$ les nœuds libres et x_0, x_{N+1} les nœuds liés. Les degrés de liberté sont donc les valeurs de u en $x_1, \ldots, x_N$. Aux nœuds liés, on a $u(x_0) = u(x_{N+1}) = 0$.
3. Choix de l'espace : on choisit comme espace de polynômes $\mathcal{P}_1 = \{ax + b, a, b \in \mathbb{R}\}$ et on pose :

$$H_N = \{u : \Omega \to \mathbb{R} \,\big|\, u_{|K_i} \in \mathcal{P}_1 \,\forall i = 1 \ldots N,\ u \in C(\bar{\Omega}) = C([0\,;1]) \text{ et } u(0) = u(1) = 0\}.$$

   Il est facile de vérifier qu'on a bien $H_N \subset H = H_0^1(]0\,;1[)$. En effet, si $u \in H_N$ alors u est continue et bornée sur $]0\,;1[)$ et donc $u \in L^2(\Omega)$. De plus la dérivée faible Du de u est une fonction définie presque partout et constante par morceaux qui est intégrable. Enfin si $u \in H^N$, on a $u(0) = u(1) = 0$ ce qui prouve que $u \in H_0^1(]0\,;1[)$.
4. choix de la base de H_N : si on prend les fonctions de type 1 de la méthode de Ritz, on choisit les fonctions décrites sur la figure 3.1. On a donc $H_1 = \text{Vect}\{\phi_1\}$, $H_3 = \text{Vect}\{\phi_1, \phi_2, \phi_3\}$ et $H_7 = \text{Vect}\{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7\}$ où Vect désigne le sous espace engendré par la famille considérée. Avec ce choix, on a donc $H_1 \subset H_3 \subset H_7$. Si maintenant on choisit des fonctions

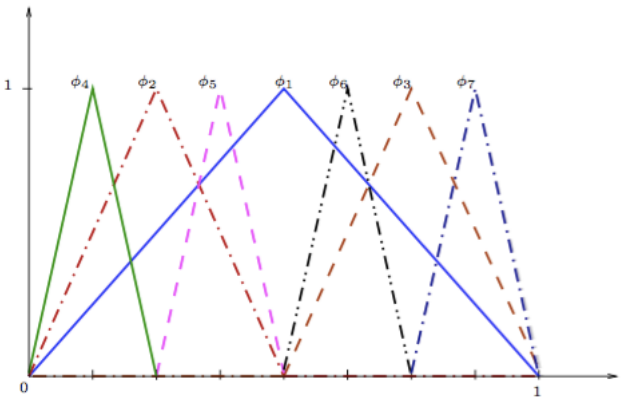


Figure 3.1 – Fonctions de forme de type 1 — espaces emboîtés

de forme de type 2, on peut définir ϕ_i pour $i = 1$ à N par :

$$\begin{cases} \phi_i : \text{continue et affine par morceau sur les intervalles } [x_{i-1}\,;x_{i+1}] \\ \text{supp}(\phi_i) = [x_{i-1}\,;x_{i+1}] \\ \phi_i(x_i) = 1 \\ \phi_i(x_{i-1}) = \phi_i(x_{i+1}) = 0 \end{cases} \tag{3.41}$$

Il est facile de voir que $\phi_i \in H_N$ et que $\{\phi_1, \ldots, \phi_N\}$ engendre H_N, c'est-à-dire que pour tout $u \in H_N$, il existe $(u_1, \ldots, u_N) \in \mathbb{R}^N$ tel que $u = \sum_{i=1}^N u_i \phi_i$. On a représenté sur la figure 3.2 les fonctions de base obtenue pour H_3 (à gauche) et H_7 (à droite). On peut remarquer que dans ce cas, les espaces d'approximation ne sont plus inclus les uns dans les autres.

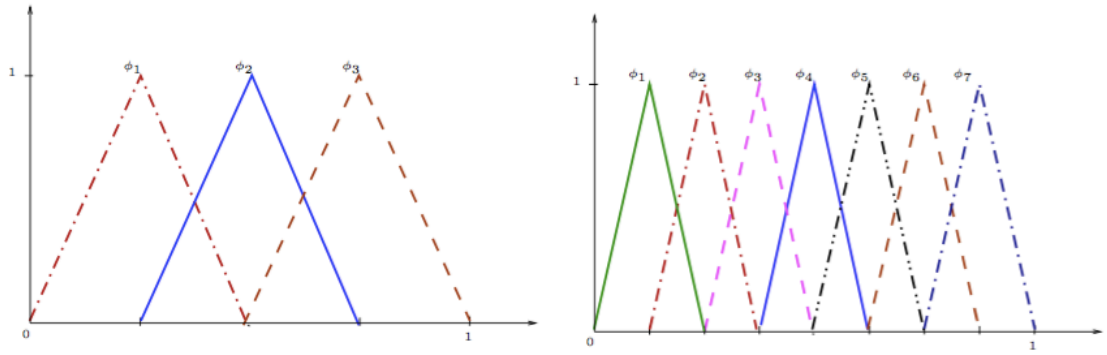


FIGURE 3.2 – Fonctions de forme de type 2 (fonction P1) en une dimension d'espace.

Exemple en dimension 2 Soit Ω un ouvert polygonal de $\mathbb{R}^2$ et $H = H_0^1(\Omega)$. Les étapes de construction de la méthode des éléments finis sont encore les mêmes.

1. éléments : on choisit des triangles.
2. nœuds : on les place aux sommets des triangles. Les nœuds $x_i \in \Omega$ (intérieurs à Ω) sont libres et les nœuds $x_i \in \partial\Omega$ (sur la frontière de Ω sont liés. On notera Σ l'ensemble des nœuds libres, Σ_F l'ensemble des nœuds liés et $\Sigma = \Sigma_I \cup \Sigma_F$.
3. espace d'approximation : l'espace des polynômes est l'ensemble des fonctions affines, noté $\mathcal{P}_1$. Une fonction $p \in \mathcal{P}_1$ est de la forme :

   $$p : \mathbb{R}^2 \to \mathbb{R}$$
   $$x = (x_1, x_2) \mapsto a_1 x_1 + a_2 x_2 + b$$

   avec $(a_1, a_2, b) \in \mathbb{R}^3$. L'espace d'approximation H_N est donc défini par :

   $$H_N : \{u \in C(\bar{\Omega}); u|_K \in \mathcal{P}_1, \forall K \text{ et } u(x_i) = 0, \forall x_i \in \Sigma_F\}$$

4. base de H_N : On choisit comme base de H_N la famille de fonctions $\{\phi_i\}$, $i = 1, \ldots, N$ où $N = \text{card}(\Sigma_I)$ et ϕ_i est définie, pour $i = 1$ à N, par :

   $$\begin{cases} \phi_i \text{ est affine par morceaux} \\ \phi_i(x_i) = 1 \\ \phi_i(x_j) = 0 \quad \forall j \neq 1 \end{cases} \tag{3.42}$$

   La fonction ϕ_i associée au nœud x_i a donc l'allure présentée sur la figure 3.3. Le support de chaque fonction ϕ_i, c'est-à-dire l'ensemble des points où ϕ_i est non nulle, est constitué de l'ensemble des triangles dont x_i est un sommet.

En résumé Les questions à se poser pour construire une méthode d'éléments finis sont donc :

1. la construction du maillage ;
2. un choix cohérent entre éléments, nœuds et espace des polynômes ;
3. la construction de l'espace d'approximation H_N et de sa base $\{\phi_i\}_{i=1\ldots N}$;
4. la construction de la matrice de rigidité $\mathcal{K}$ et du second membre $\mathcal{G}$;
5. l'évaluation de $d(u, H_N)$ en vue de l'analyse de convergence.

3.3.2 Maillage de l'espace H_N et de sa base ϕ_N

Construction des éléments

Soit $\Omega \in \mathbb{R}^2$ un ouvert borné polygonal. On construit un maillage de Ω en divisant $\bar{\Omega}$ en parties fermées $\{K_\ell\}_{\ell=1,\ldots,L}$ où L est le nombre d'éléments.

Les principes pour la construction du maillage sont :

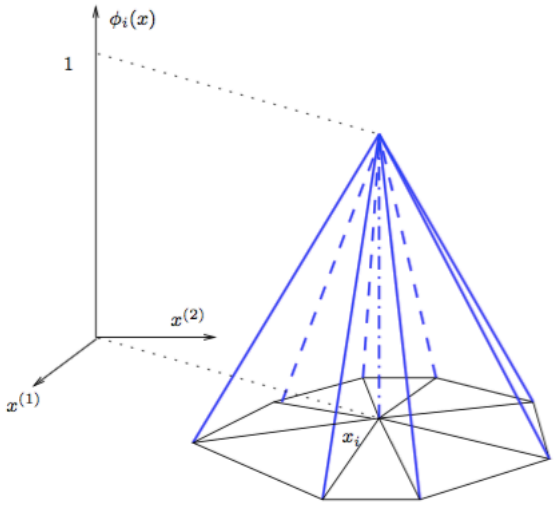


Figure 3.3 – Fonction de forme de type 2 (fonction de $\mathcal{P}_1$) en deux dimensions d'espace

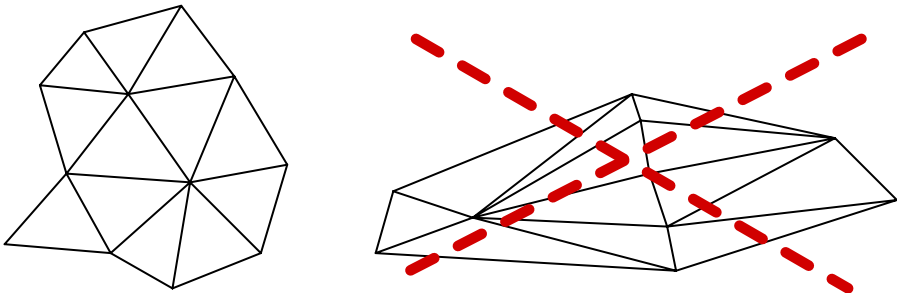


Figure 3.4 – Exemple de triangles *bons* (à gauche) et *mauvais* (à droite)

— Eviter les angles trop grands ou trop petits. On préfèrera par exemple les triangles de gauche plutôt que ceux de droite dans la figure 3.4.

— Mettre beaucoup d'éléments là où u varie rapidement (ceci ne peut se faire que si on connait a priori les zones de de variation rapide, ou si on a les moyens d'évaluer l'erreur entre la solution exacte du problème et la solution calculée et de remailler les zones ou celle-ci est jugée trop grande.

— On peut éventuellement mélanger des triangles et des rectangles, mais ceci n'est pas toujours facile.

Il existe un très grand nombre de logiciels de maillages en deux ou trois dimensions d'espace. On pourra pour s'en convaincre utiliser les moteurs de recherche sur internet avec les mots clés : "mesh 2D structured", "mesh 2D unstructured", "mesh 3D structured", "mesh 3D unstructured". Le mot "mesh" est le terme anglais pour maillage, les termes 2D et 3D réfèrent à la dimension de l'espace physique. Le terme "structured" (structuré en français) désigne des maillages que dont on peut numéroter les éléments de façon cartésienne, le terme "unstructured" (non structuré) désigne tous les autres maillages. L'avantage des maillages "structurés" est qu'ils nécessitent une base de données beaucoup plus simple que les maillages non structurés, car on peut connaitre tous les nœuds voisins à partir du numéro global d'un nœud d'un maillage structuré, ce qui n'est pas le cas dans un maillage non structuré (voir paragraphe suivant pour la numérotation des nœuds). La figure 3.5 montre un exemple de maillage de surface structuré ou non-structuré.

Choix des nœuds

On se donne une famille $\{S_i\}_{i=1,\dots,M}$ de M points de $\bar{\Omega}$ de composantes (x_i, y_i), pour $i = 1, \dots, M$.

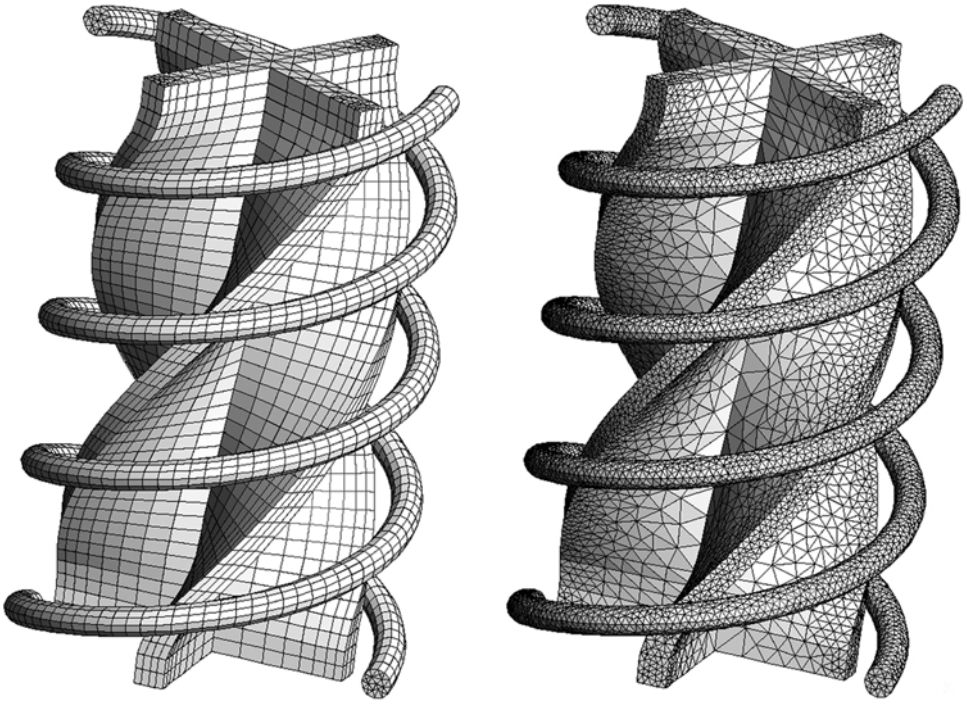


Figure 3.5 – Exemple de maillage structuré (à gauche) et non-structuré (à droite) d'une surface ; Gmsh : a three-dimensional finite element mesh generator with built-in pre-and post-processing facilities, dévelippé par C. Geuzaine and J.-F. Remacle — `http://www.geuz.org/gmsh/`

Le maillage éléments finis est défini par éléments $\{K_\ell\}_{\ell=1\dots L}$ et les nœuds $\{S_i\}_{i=1\dots M}$. Ces éléments et nœuds ne peuvent bien sûr pas être choisis indépendamment. Dans le cas général, on choisit tous les éléments de même type (par exemple, des triangles) et on se donne un nombre fixe de nœuds par élément, ce qui détermine le nombre total de nœuds. Chaque nœud appartient donc à plusieurs éléments. Dans le cas d'un maillage structuré tel que celui qu'on a décrit dans la figure 1.5, une numérotation globale des nœuds est suffisante pour retrouver les éléments dont font partie ce nœud, ainsi que tous les voisins du nœud. Par contre, dans le cas d'un maillage non structuré (un maillage en triangles, par exemple), on aura besoin d'une numérotation locale des nœuds c'est-à-dire une numérotation des nœuds de chaque élément, pour $k = 1, \dots, N_\ell$ où N_ℓ est le nombre de nœuds par élément ; on aura également besoin d'une numérotation globale des nœuds et d'une table de correspondance, l'une qui donne pour chaque élément, les numéros dans la numérotation globale des nœuds qui lui appartiennent :

$i_r^\ell =$ notation globale (ℓ, r) r-ème nœud de l'élément ℓ

Amélioration de la précision

On a vu aux paragraphes précédents que l'erreur entre la solution exacte u recherchée et la solution $u(N)$ obtenue par la méthode de Ritz ou de Galerkin est majorée par une constante fois la distance entre H et H_N. On a donc intérêt à ce que cette distance soit petite. Pour ce faire, il paraît raisonnable d'augmenter la dimension de l'espace H_N. Pour cela, on a deux possibilités :

— augmenter le nombre d'éléments : on augmente alors aussi le nombre global de nœuds, mais pas le nombre local.

— augmenter le degré des polynômes : on augmente alors le nombre de nœuds local, donc on augmente aussi le nombre global de nœuds, mais pas le nombre d'éléments. Ce deuxième choix (augmentation du degré des polynômes) ne peut se faire que si la solution est suffisamment régulière ; si la solution n'est pas régulière, on n'arrivera pas à diminuer $d(H, H_N)$ en augmentant le degré des polynômes.

3.4 Exercices

3.4.1 Énoncés

Exercice 36 — Dérivée par transposition, dérivée faible. Soit f la fonction définie par $f(x) = \dfrac{1}{x}$ sur $]0,1[$.

1. La fonction f est-elle dans un espace $L^p(]0,1[)$?
2. Montrer que f est la dérivée par transposition de la fonction g définie par $g(x) = -\ln x$. La fonction f est-elle une dérivée faible de g ?

Exercice 37 — Fonctions H^1 en une dimension. Montrer que si $u \in H^1(]0\,;1[)$ alors u est continue. En déduire que $H^2(]0\,;1[) \subset C^1([0\,;1])$.

corrigé p.124

Exercice 38 — Minimisation de la semi-norme. Soit Ω un ouvert borné de $\mathbb{R}^n$. On suppose que sa frontière est de classe C^1 par morceaux. Etant donné une fonction $u_0 \in H^1(\Omega)$, on désigne par $u_0 + H_0^1(\Omega)$ l'ensemble $\{u_0 + v, v \in H_0^1(\Omega)\}$.

suggestions p.123

1. Montrer qu'il existe une unique fonction $u \in u_0 + H_0^1(\Omega)$ tel que :

$$|u|_{1,\Omega} = \inf_{v \in u_0 + H_0^1(\Omega)} |v|_{1,\Omega}.$$

2. Caractériser u comme étant la solution d'un problème aux limites.

corrigé p.125

Exercice 39 — Formulation faible pour le problème de Dirichlet unidimensionnel. Soit $f \in L^2(]0\,;1[)$. On s'intéresse au problème suivant :

$$-u''(x) = f(x),\ x \in]0\,;1[, \tag{3.43}$$
$$u(0) = 0 \tag{3.44}$$
$$u(1) = 0. \tag{3.45}$$

Donner une formulation faible et une formulation variationnelle de (3.45).

corrigé p.126

Exercice 40 — Relèvement. Soient a et $b \in \mathbb{R}$, et $f \in C(\mathbb{R}, \mathbb{R})$.

1. Soient u_0 et u_1 définies de $[0\,;1]$ dans $\mathbb{R}$ par $u_0(x) = a + (b-a)x$ et $u_1(x) = a + (b-a)x^2$. Montrer qu'il existe un unique $\widetilde{u}$ (resp. $\overline{u}$) tel que $u = u_0 + \widetilde{u}$ (resp. $v = u_1 + \overline{u}$) soit solution de (3.11). Montrer que $u = v$.

2. Mêmes questions en supposant maintenant que u_0 et u_1 sont des fonctions de $C^2([0\,;1])$ telles que $u_0(0) = u_1(0) = a$ et $u_0(1) = u_1(1) = b$.

corrigé p.126

Exercice 41 — Lax-Milgram. Ecrire une formulation faible pour laquelle on puisse appliquer le théorème de Lax–Milgram, dans le cas du problème suivant :

suggestions p.123

$$\begin{cases} -u''(x) = f(x), \quad x \in [0\,;1], \\ u'(0) = 0, \\ u(1) = 1. \end{cases} \tag{3.46}$$

corrigé p.127

Exercice 42 — Conditions aux limites de Fourier et Neumann. Soit $f \in L^2(]0\,;1[)$. On s'intéresse au problème suivant :

$$\begin{aligned} &-u''(x) + u(x) = f(x) && x \in]0\,;1[\\ &u'(0) - u(0) = 0 \quad u'(1) = -1. \end{aligned} \tag{3.47}$$

Donner une formulation faible et une formulation variationnelle de (3.47) ; y-a-t-il existence et unicité des solutions faibles de (3.47) ?

Exercice 43 — Conditions aux limites de Fourier et Neumann, bis. Soit $f \in L^2(]0\,;1[)$. On s'intéresse au problème suivant :

$$\begin{cases} -u''(x) - u'(x) + u(x) = f(x),\ x \in]0\,;1[, \\ u(0) + u'(0) = 0,\ u(1) = 1 \end{cases} \tag{3.48}$$

1. Donner une formulation faible du problème de la forme

$$\begin{cases} \text{Trouver } u \in H^1(]0\,;1[); u(1) = 1, \\ a(u,v) = T(v),\ \forall v \in H. \end{cases} \tag{3.49}$$

où $H = \{v \in H^1(]0\,;1[); v(1) = 0\}$, a et T sont respectivement une forme bilinéaire sur $H^1(]0\,;1[)$ et une forme linéaire sur $H^1(]0\,;1[)$, à déterminer.

2. Y-a-t-il existence et unicité de solutions de cette formulation faible ?

Exercice 44 — Un problème non linéaire. Soit $\Omega =]0,1[$ et soit φ une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$ telle qu'il existe $M \geqslant 0$ avec $|\varphi(s)| \leqslant M$ pour tout $s \in \mathbb{R}$.

suggestions p.123

On s'intéresse à l'existence et l'approximation d'une fonction $u \in H^1_0(\Omega)$, solution faible du problème

$$-u'' + (\varphi(u))' = f,$$

où $f \in L^2(\Omega)$ est donnée. On définit le sens faible suivant :

$$u \in H^1_0(\Omega),\ \forall v \in H^1_0(\Omega),\ \int_\Omega \Big[u'(x)v'(x) - \varphi(u(x))\ v'(x)\Big]\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x. \tag{3.50}$$

Pour tout $n \in \mathbb{N}$, soit $V_n \subset H^1_0(\Omega)$ un sous-espace de dimension finie tel que

$$\forall u \in H^1_0(\Omega),\ \lim_{n\to\infty} \inf_{v \in V_n} \|u - v\|_{H^1(\Omega)} = 0.$$

1. Soit $n \in \mathbb{N}$ et $w \in V_n$. Montrer qu'il existe une et une seule fonction z solution de
$$z \in V_n,\ \forall v \in V_n,\ \int_\Omega z'(x)v'(x)\,\mathrm{d}x = \int_\Omega \Big[f(x)v(x) + \varphi(w(x))v'(x)\Big]\,\mathrm{d}x, \tag{3.51}$$
et qu'il existe $C > 0$ ne dépendant ni de w ni de n telle que l'on ait $\|z'\|_{L^2(\Omega)} \leqslant C$.
2. Montrer que, pour tout $n \in \mathbb{N}$, il existe au moins une fonction, notée $u_n \in V_n$, solution du problème
$$u_n \in V_n,\ \forall v \in V_n,\ \int_\Omega \Big[u_n'(x)v'(x) - \varphi(u_n(x))\ v'(x)\Big]\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x. \tag{3.52}$$
3. a) Montrer que, de la suite $(u_n)_{n\in\mathbb{N}}$, on peut extraire une sous-suite, notée $(u_{\psi(n)})_{n\in\mathbb{N}}$, telle qu'il existe une fonction $u \in H^1_0(\Omega)$, avec $u_{\psi(n)}$ tendant vers u dans $L^2(\Omega)$ et $u'_{\psi(n)}$ tendant vers u' faiblement dans $L^2(\Omega)$.
   b) Prouver que u est solution du problème (3.50).
   c) En déduire que $u'_{\psi(n)}$ tend vers u' dans $L^2(\Omega)$.

corrigé p.128

Exercice 45 — Conditions mixtes. Soit Ω un ouvert borné $\mathbb{R}^d$, $d = 1ou2$, de frontière $\partial\Omega = \Gamma_0 \cup \Gamma_1$, avec $\Gamma_0 \cap \Gamma_1 = \emptyset$; on suppose que la mesure $d-1$ dimensionnelle de Γ_0 est non nulle,

suggestions p.124

et soit $f \in L^2(\Omega)$. On s'intéresse ici au problème suivant :

$$\begin{cases} -\Delta u(x) = f(x),\ x \in \Omega, \\ u(x) = 0,\ x \in \Gamma_0, \\ \nabla u(x) \cdot \boldsymbol{n}(x) = 0,\ x \in \Gamma_1, \end{cases} \tag{3.53}$$

où $\boldsymbol{n}$ est la normale unitaire à $\partial\Omega$ extérieure à Ω.

Donner une formulation faible et une formulation variationnelle de (3.53) telle qu'on puisse appliquer le théorème de Lax-Milgram. (On rappelle que l'inégalité de Poincaré donnée en bas de

page 1.31 pour les fonctions de $H^1_0(\Omega)$ est encore valable pour les fonctions de $H^1(\Omega)$ dont la trace est nulle sur un sous-ensemble de $\partial\Omega$ de mesure $((d-1)$–dimensionnelle) non nulle.)

corrigé p.128

Exercice 46 — Problème elliptique pour un problème avec conditions mixtes. Soit Ω un ouvert borné $\mathbb{R}^d$, $d = 1$ ou 2, de frontière $\partial\Omega = \Gamma_0 \cup \Gamma_1$, avec $\Gamma_0 \cap \Gamma_1 = \emptyset$; on suppose que la mesure $d-1$ dimensionnelle de Γ_0 est non nulle. On s'intéresse ici au problème suivant :

$$\begin{cases} -\operatorname{div}(p(x)\nabla u(x)) + q(x)u(x) = f(x), x \in \Omega \\ \qquad\qquad\qquad\qquad u(x) = g_0(x), x \in \Gamma_0, \\ p(x)\nabla u(x) \cdot \boldsymbol{n}(x) + \sigma u(x) = g_1(x), x \in \Gamma_1, \end{cases} \tag{3.54}$$

où $f \in L^2(\Omega)$, $p \in L^\infty(\Omega)$, est telle qu'il existe $\alpha > 0$ t.q. $p(x) \geqslant \alpha$ p.p., $q \in L^\infty(\Omega)$, $q \geqslant 0$, $\sigma \in \mathbb{R}_+$, $g_0 \in L^2(\Gamma_0)$ est telle qu'il existe $\widetilde{g} \in H^1(\Omega)$ t.q. $\gamma(\widetilde{g})|_{\Gamma_0} = g_0$, $g_1 \in L^2(\Gamma_1)$ et $\boldsymbol{n}$ est la normale unitaire à $\partial\Omega$ extérieure à Ω.

1. Donner une formulation faible et une formulation variationnelle de (3.54) telle qu'on puisse appliquer le théorème de Lax-Milgram.

2. On suppose dans cette question que $p \in C^1(\overline{\Omega})$, $q \in C(\overline{\Omega})$ $g_0 \in C(\Gamma_0)$ et $g_1 \in C(\Gamma_1)$. Soit $u \in C^2(\overline{\Omega})$. Montrer que u est solution faible si et seulement si u est une solution classique de (3.54).

Exercice 47 — Problème de Neumann homogène. On considère le problème suivant :

suggestions p.124

$$\begin{aligned} -\Delta u + u &= f \quad \text{dans } \Omega, \\ \nabla u \cdot \boldsymbol{n} &= 0 \quad \text{sur } \Omega \end{aligned} \tag{3.55}$$

où Ω est un ouvert borné de $\mathbb{R}^d$ de frontière régulière de classe C^2 et de normale unitaire extérieure $\boldsymbol{n}$, et $f \in L^2(\Omega)$.

1. Montrer que si u est une solution régulière de (3.55) et $\varphi \in C^\infty(\mathbb{R}^d)$, alors

$$\int_\Omega (\nabla u \cdot \nabla\varphi + u\,\varphi)\,\mathrm{d}x = \int_\Omega f\varphi\,\mathrm{d}x$$

En déduire que u est solution de la formulation faible qui consiste à trouver $u \in H^1(\Omega)$ tel que

$$\int_\Omega (\nabla u \cdot \nabla v + uv)\,\mathrm{d}x = \int_\Omega fv\,\mathrm{d}x \qquad \forall v \in H^1(\Omega). \tag{3.56}$$

2. Montrer que si u est une solution régulière de classe C^2 de (3.56) et $\varphi \in C^\infty(\mathbb{R}^d)$, alors u est solution de (3.55).

corrigé p.130

Exercice 48 — Condition inf-sup. Soit V un espace de Hilbert réel de produit scalaire $(\cdot\,;\,\cdot)$ induisant une norme $||\cdot||$. On se donne $a(\cdot\,;\,\cdot)$ une forme bilinéaire continue sur $V \times V$, avec M comme constante de continuité. Soit L une forme linéaire continue sur V. On suppose de plus qu'il existe une solution $u \in V$ au problème suivant :

$$a(u, v) = L(v),\ \forall v \in V. \tag{3.57}$$

Soit V_h un sous-espace de V de dimension finie. On suppose qu'il existe $\beta_h \in R_+$ telle que :

$$\inf_{(v_h \in V_h, \|v_h\|=1)} \left(\sup_{(w_h \in V_h, \|w_h\|=1)} (a(v_h; w_h)) \right) \geqslant \beta_h \tag{3.58}$$

On cherche alors u_h solution de :

$$\begin{aligned} &u_h \in V_h, \\ &a(u_h, v_h) = L(v_h) \quad \forall v_h \in V_h \end{aligned} \tag{3.59}$$

1. Montrer que le problème (3.59) admet une unique solution.

2. Soit u la solution de (3.57) et u_h la solution de (3.59). Montrer que :

$$\|u - u_h\| \leqslant \left(1 + \frac{M}{\beta_h}\right) \inf_{v_h \in V_h} \|u - v_h\|. \tag{3.60}$$

Exercice 49 — Condition inf-sup pour un problème mixte. Soient V et Q deux espaces de Hilbert, on note $(\,\cdot\,,\,\cdot\,)_V$, $\|\cdot\|_V$ et $(\,\cdot\,,\,\cdot\,)_Q$, $\|\cdot\|_Q$ leurs produits scalaires et normes respectives, et on considère le problème suivant :

$$\begin{cases} \text{Trouver } u \in V,\, p \in Q, \text{ tels que} \\ a(u,v) + b(v,p) = (f,v)_H, \forall v \in V, \\ b(u,q) = (g,q)_Q, \forall q \in Q. \end{cases} \tag{3.61}$$

où a est une forme bilinéaire continue et coercive sur V et b est une application bilinéaire continue de $V \times Q$ dans $\mathbb{R}$. Pour (u,p) et (v,q) éléments de $V \times Q$, on pose :

$$\begin{aligned} B(u,p;v,q) &= a(u,v) + b(v,p) + b(u,q), \\ F(v,q) &= (f,v)_H + (g,q)_Q. \end{aligned}$$

et on munit $V \times Q$ d'une norme notée $\|(\,\cdot\,,\,\cdot\,)\|$, définie par $\|(v,q)\| = \|v\|_V + \|q\|_Q$ pour $(v,q) \in V \times Q$.

1. Montrer que B est une forme bilinéaire continue sur $V \times Q$.

2. Montrer que le problème (3.61) est équivalent au problème :

$$\begin{cases} \text{Trouver } (u,p) \in V \times Q, \text{ tels que} \\ B(u,p;v,q) = F(v,q), \forall (v,q) \in V \times Q. \end{cases} \tag{3.62}$$

On considère maintenant des espaces d'approximation (par exemple construits par éléments finis). Soient donc $(V_n)_{n\in\mathbb{N}}$ et $(Q_n)_{n\in\mathbb{N}}$ des espaces de Hilbert de dimension finie tels que $V_n \subset V$ et $Q_n \subset Q$, pour tout $n \in \mathbb{N}$.

3. On suppose dans cette question que la condition suivante (dite condition "inf-sup") est satisfaite :

$$\text{Il existe } \beta \in \mathbb{R}_+^* \text{ (indépendant de } n) \text{ tel que } \inf_{q\in Q_n} \sup_{\substack{w\in V_n, \\ \|w\|_V \neq 0}} \frac{b(w,q)}{\|w\|_V} \geqslant \beta \|q\|_Q. \tag{3.63}$$

(a) Montrer qu'il existe $\alpha \in \mathbb{R}_+^*$ tel que :

$$\text{Pour tout } q \in Q_n \text{ et } v \in V_n, B(v,q;v,-q) \geqslant \alpha \|v\|_V^2. \tag{3.64}$$

(b) Soit $(v,q) \in V_n \times Q_n$, montrer qu'il existe $w \in V_n$ tel que $\|w\|_V = \|q\|_Q$ et $b(w,q) \geqslant \beta \|q\|_Q^2$. Montrer que pour ce choix de w, on a :

$$B(v,q;w,0) \geqslant -M\|v\|_V\|w\|_V + \beta\|q\|_Q^2,$$

où M est la constante de continuité de a.

(c) En déduire qu'il existe des réels positifs C_1 et C_2 indépendants de n tels que

$$B(v,q;w,0) \geqslant -C_1\|v\|_V^2 + C_2\|q\|_Q^2.$$

(On pourra utiliser, en le démontrant, le fait que pour tout $a_1 \geqslant 0, a_2 \geqslant 0$ et $\varepsilon > 0$, on a $a_1 a_2 \leqslant \frac{1}{\varepsilon} a_1^2 + \varepsilon a_2^2$.)

(d) Soit $\gamma \in \mathbb{R}_+^*$. Montrer que si γ est suffisamment petit, on a :

$$B(v,q;v+\gamma w,-q) \geqslant C_3[\|v\|_V^2 + \|q\|_Q^2].$$

et

$$\|(v+\gamma w,-q)\| \leqslant C_4\|(v,q)\|,$$

où C_3 et C_4 sont deux réels positifs qui ne dépendent pas de n.

(e) En déduire que la condition suivante (dite de stabilité) est satisfaite :

$$\text{Il existe } \delta \in \mathbb{R}_+^* \text{ indépendant de } n, \text{ tel que pour tout } (u,p) \in V_n \times Q_n,$$
$$\sup_{\substack{(v,q)\in V_n\times Q_n, \\ \|(v,q)\|\neq 0}} \frac{B(u,p;v,q)}{\|(v,q)\|} \geqslant \delta\|(u,p)\|. \tag{3.65}$$

4. On suppose maintenant que la condition (3.65) est satisfaite.

(a) Montrer que pour tout $p \in Q$,

$$\sup_{\substack{(v,q)\in V_n\times Q_n,\\ \|(v,q)\|\neq 0}} \frac{b(v,p)}{\|(v,q)\|} \geqslant \delta\|p\|_Q.$$

(b) En déduire que pour tout $p \in Q$,

$$\sup_{\substack{(v,q)\in V_n\times Q_n,\\ \|(v,q)\|\neq 0}} \frac{b(v,p)}{\|v\|_V} \geqslant \delta\|p\|_Q.$$

5. Déduire des questions précédentes que la condition (3.63) est satisfaite si et seulement si la condition (3.65) est satisfaite.

corrigé p.131

Exercice 50 — Volumes finis vus comme des éléments finis non conformes. Soit un ouvert borné polygonal de $\mathbb{R}^2$ et $\mathcal{T}$ un maillage admissible au sens des volumes finis (voir page 1.4.2) de Ω.

1. Montrer que la discrétisation par volumes finis de (3.1) se ramène à chercher $(u_K)_{K\in\mathcal{T}}$, qui vérifie :

$$\sum_{\sigma\in\mathcal{E}_{\rm int},\ \sigma=K|L} \tau_\sigma(u_L - u_K) + \sum_{\sigma\in\mathcal{E}_{\rm ext},\ \sigma\in\mathcal{E}_K} \tau_\sigma\ u_K = m(K) f_K \tag{3.66}$$

où $\mathcal{E}_{\rm int}$ représente l'ensemble des arêtes internes (celles qui ne sont pas sur le bord) $\mathcal{E}_{\rm ext}$ l'ensemble des arêtes externes (celles qui sont sur le bord), et

$$\tau_\sigma = \begin{cases} \dfrac{m(\sigma)}{d_{K,\sigma} + d_{L,\sigma}} \text{ si } \sigma \in \mathcal{E}_{\rm int},\ \sigma = K|L, \\ \dfrac{m(\sigma)}{d_{K,\sigma}} \text{ si } \sigma \in \mathcal{E}_{\rm ext},\ \sigma \in \mathcal{E}_K, \end{cases} \tag{3.67}$$

(voir figure 1.6).

2. On note $H_\mathcal{T}(\Omega)$ le sous-espace de $L^2(\Omega)$ formé des fonctions constantes par maille (c.à.d. constantes sur chaque élément de $\mathcal{T}$). Pour $u \in H_\mathcal{T}(\Omega)$, on note u_K la valeur de u sur K. Montrer que $(u_K)_{K\in\mathcal{T}}$ est solution de (3.67) si et seulement si $u \in H_\mathcal{T}(\Omega)$ est solution de :

$$\begin{cases} u \in H_\mathcal{T}(\Omega), \\ a_\mathcal{T}(u,v) = T_\mathcal{T}(v),\ \forall v \in H_\mathcal{T}(\Omega), \end{cases} \tag{3.68}$$

où $a_\mathcal{T}$ est une forme bilinéaire sur $H_\mathcal{T}(\Omega)$ (à déterminer), et $T_\mathcal{T}$ est une forme linéaire sur $H_\mathcal{T}(\Omega)$ (à déterminer).

3.4.2 Suggestions pour les exercices

Exercice 38. Rappel ; par définition, l'ensemble $u_0 + H_0^1$ est égal à l'ensemble $\{v = u_0 + w, w \in H_0^1\}$.
1. Montrer que le problème s'écrit sous la forme : $J(u) \leqslant J(v), \forall v \in H$, ou H est un espace de Hilbert, avec $J(v) = a(v,v)$, où a est une forme bilinéaire symétrique définie positive.
2. Prendre une fonction test à support compact dans la formulation faible.

Exercice 41. Considérer les espaces $H^1_{1,1} = \{v \in H^1(]0,1[); v(1) = 1\}$ et $H^1_{1,0} = \{v \in H^1(]0,1[); v(1) = 0\}$.

Exercice 44 – Un problème non linéaire.
2. On pourra utiliser le théorème du point fixe de Brouwer : toute fonction continue d'une boule fermée B d'un espace vectoriel normé de dimension finie à valeurs dans la boule B admet au moins un point fixe.

Exercice 45. Considérer comme espace de Hilbert l'ensemble $\{u \in H^1(\Omega); u = 0 \text{ sur } \Gamma_0\}$.

Exercice 47. 1. On rappelle que l'espace des fonctions de $C^\infty(\mathbb{R}^d)$ restreintes à Ω est dense dans $H^1(\Omega)$.
2. On admettra que l'image de $H^1(\Omega)$ par l'application trace est dense dans $L^2(\partial\Omega)$.

Exercice 50. 1. Intégrer l'équation sur la maille et approcher les flux sur les arêtes par des quotients différentiels.
2. Pour montrer que (3.66) entraîne (3.68), multiplier par v_K, où $v \in H_{\mathcal{T}}(\Omega)$, et développer. Pour montrer la réciproque, écrire u comme combinaison linéaire des fonctions de base de $H_{\mathcal{T}}(\Omega)$, et prendre pour v la fonction caractéristique de la maille K. (3.66)

3.4.3 Corrigés

Exercice 38.

1. Par définition, on sait que

$$|u|_{1,\Omega} = (\int_\Omega \sum_{i=1}^N |\partial_i u(x)|^2 \, dx)^{\frac{1}{2}}$$

où $\partial_i u$ désigne la dérivée partielle de u par rapports à sa i–ème variable. Attention ceci $|\cdot|_{1,\Omega}$ définit une semi-norme et non une norme sur l'espace $H^1(\Omega)$. Cependant sur $H^1_0(\Omega)$ c'est bien une norme, grâce à l'inégalité de Poincaré. On rappelle que $H^1_0(\Omega) = \ker(\gamma) = \left\{u \in H^1(\Omega) \textit{ tel que } \gamma(u) = 0\right\}$, où γ est l'opérateur de trace linéaire et continu de $H^1(\Omega)$ dans $L^2(\Omega)$ (voir théorème 3.1). Le problème consiste à minimiser

$$\left(\int_\Omega \sum_{i=1}^N |\partial_i u|^2 \, dx\right)^{\frac{1}{2}}$$

sur $u_0 + H^1_0(\Omega)$. Tentons de nous ramener à minimiser une fonctionnelle sur $H^1_0(\Omega)$. Soit $v \in u_0 + H^1_0(\Omega)$. Alors $v = u_0 + w$ avec $w \in H^1_0(\Omega)$, et donc :

$$|v|^2_{1,\Omega} = |u_0 + w|^2_{1,\Omega} = \int_\Omega \sum_{i=1}^N |\partial_i(u_0 + w)|^2 \, \mathrm{d}x = \int_\Omega \sum_{i=1}^N \left|(\partial_i u_0)^2 + (\partial_i w)^2 + 2\,(\partial_i u_0)\,(\partial_i w)\right| \, \mathrm{d}x$$

$$= |u_0|^2_{1,\Omega} + |w|^2_{1,\Omega} + 2\int_\Omega \sum_{i=1}^N |\partial_i u_0 \partial_i w| \, \mathrm{d}x$$

Ainsi chercher à mimimiser $|v|_{1,\Omega}$ sur $u_0 + H^1_0(\Omega)$ revient à minimiser J sur $H^1_0(\Omega)$, où J est définie par :

$$J(w) = \inf_{H^1_0(\Omega)} 2\left(\int_\Omega \sum_{i=1}^N |\partial_i u_0 \partial_i w| \, dx + \frac{1}{2}\int_\Omega \sum_{i=1}^N (\partial_i w)^2\right).$$

Pour montrer l'existence et l'unicité du minimum de J, nous allons mettre ce problème sous une forme faible, puis utiliser le théorème de Lax–Milgram pour en déduire l'existence et l'unicité d'une solution faible, et finalement conclure que la fonctionnelle $J(w)$ admet un unique infimum. On pose :

$$a(w,v) = \int_\Omega \sum_{i=1}^N (\partial_i w \partial_i v) \, \mathrm{d}x, \ \forall w, v \in H^1_0(\Omega) \quad \text{et} \quad L(v) = -\int_\Omega \sum_{i=1}^N (\partial_i u_0 \partial_i v) \, \mathrm{d}x, \ \forall v \in H^1_0(\Omega)$$

Voyons si les hypothèses de Lax-Milgram sont vérifiées. La forme $a(w,v)$ est clairement symétrique, on peut changer l'ordre de w et de v dans l'expression sans changer la valeur de l'intégrale. La forme $a(w,v)$ est bilinéaire. En effet, elle est linéaire par rapport au premier argument, puisque : $\forall u, v, w \in H^1_0(\Omega)$ et $\forall \lambda, \mu \in \mathbb{R}$, on a : $a(\lambda w + \mu u, v) = \lambda a(w,v) + \mu a(u,v)$. Ainsi par symétrie, elle est aussi linéaire par rapport au second argument. Donc elle est bien bilinéaire. Pour montrer que la forme $a(w,v)$ est continue, on utilise la caractérisation de la continuité des applications

bilinéaires. On va donc montrer l'existence de $C \in \mathbb{R}_+$ tel que $|a(u,v)| \leqslant C\|u\|_{H^1}\|v\|_{H^1}$ pour tous u, $v \in H_0^1(\Omega)$. Or, par l'inégalité de Cauchy-Schwarz, on a :

$$|a(u,v)| = \left|\int_\Omega \sum_{i=1}^N (\partial_i u \partial_i v)\, \mathrm{d}x\right| \leqslant \left(\int_\Omega \sum_{i=1}^N (\partial_i u)^2\, dx\right)^{\frac{1}{2}} \left(\int_\Omega \sum_{i=1}^N (\partial_i v)^2\, dx\right)^{\frac{1}{2}} \leqslant \|u\|_{H^1}\|v\|_{H^1}.$$

La forme a est donc bien continue. Montrons alors qu'elle est coercive, c'est-à-dire qu'il existe $\alpha > 0$ tel que $a(v,v) \geqslant \alpha\|v\|_{H^1}^2$ pour tout $v \in H_0^1(\Omega)$.

$$a(v,v) = \int_\Omega \sum_{i=1}^N (\partial_i v(x))^2\, \mathrm{d}x = \int_\Omega \nabla v(x) \cdot \nabla v(x)\, \mathrm{d}x \geqslant \frac{1}{1+\mathrm{diam}(\Omega)^2}\|v\|_{H^1}^2,$$

grâce à l'inégalite de Poincaré, qu'on rappelle ici :

$$\|v\|_{L^2(\Omega)} \leqslant c(\Omega)\|\nabla v\|_{L^2(\Omega)},\ \forall v \in H_0^1(\Omega). \tag{3.69}$$

On aurait pu utiliser directement l'equivalence des normes de $\|v\|_{H^1}^2$ et de $|v|_{1,\Omega}$ aussi donnée par l'inégalité de Poincaré. En effet : $\|v\|_{H^1} = \|v\|_{L^2}^2 + \sum \|\frac{\partial v}{\partial xi}\|_{L^2}^2 = \int_\Omega \left(v^2 + \nabla v^2\right)$ L'inégalité de Poincaré peut s'ecrire aussi : $\|v\|_{L^2} \leqslant c(\Omega)\|\nabla u\|_{L^2} = c(\Omega)|v|_{1,\Omega}$
Donc a est bien une forme bilinéaire, symétrique, continue et coercive. Par le même genre de raisonnement, on montre facilement que L est linéaire et continue. Ainsi toutes les hypothèses de Lax-Milgram sont vérifiées, et donc le problème :

$$\text{Trouver } u \in H_0^1(\Omega) \text{ tel que } a(u,v) = L(v) \text{ pour tout } v \in H_0^1(\Omega)$$

admet une unique solution dans $H_0^1(\Omega)$. De plus, comme a est symétrique, la fonctionnelle J admet un unique minimum.

2. On va maintenant caractériser u comme étant la solution d'un problème aux limites. Soit $\varphi \in D(\Omega)$, donc φ est à support compact dans Ω. On a :

$$a(u,\varphi) = L(\varphi) \qquad \forall \varphi \in D(\Omega),$$

et donc :

$$\int_\Omega \nabla u(x)\nabla(x)\varphi\, \mathrm{d}x = -\int_\Omega \nabla u_0(x)\nabla(x)\varphi(x)dx.$$

Comme u et $u_0 \in H^1(\Omega)$, et comme φ est réguliére, on peut intégrer par parties ; en remarquant que φ est nulle sur $\partial\Omega$, on a donc :

$$-\int_\Omega \Delta u(x)\phi(x)dx = \int_\Omega \Delta u_0(x)\phi(x)dx.$$

On en déduit que $-\Delta u = \Delta u_0$. Comme $u \in H_0^1(\Omega)$, ceci revient à résoudre le problème aux limites $\widetilde{u} = u - u_0 \in H^1(\Omega)$, tel que $-\Delta\widetilde{u} = 0$ dans Ω et $\widetilde{u} = u_0$ sur $\partial\Omega$.

Exercice 39 — Formulation faible pour le problème de Dirichlet unidimensionnel. Soit $\varphi \in C_c^\infty([0\,;1])$, on multiplie la première équation de (3.45), on intègre par parties et on obtient :

$$\int_0^1 u'(x)\varphi'(x)\, \mathrm{d}x = \int_0^1 f(x)\varphi(x)\, \mathrm{d}x. \tag{3.70}$$

Pour trouver une formulation faible (ou variationnelle) il faut commencer par trouver un espace de Hilbert pour les fonctions duquel (3.70) ait un sens, et qui soit compatible avec les conditions aux limites. Comme $f \in L^2(]0\,;1[)$, le second membre de (3.70) est bien défini dès que $\varphi \in L^2(]0\,;1[)$. De même, le premier membre de (3.70) est bien défini dès que $u' \in L^2(]0\,;1[$ et $\varphi' \in L^2(]0\,;1[)$. Comme de plus, on doit avoir $u = 0$ en 0 et en 1, il est naturel de choisir par définition $H = H_0^1(]0\,;1[) = \{u \in L^2(]0\,;1[; Du \in L^2(]0\,;1[)$ et $u(0) = u(1) = 0\}$. (Rappelons qu'en une dimension d'espace $H^1(]0\,;1[) \subset C([0\,;1])$ et donc $u(0)$ et $u(1)$ sont bien définis). Une formulation faible naturelle est donc de trouver $u \in H = \{u \in H_0^1(\Omega); v(0) = v(1) = 0\}$ tel que :

$$a(u,v) = T(v), \quad \forall v \in H \quad \text{avec} \quad a(u,v) = \int_0^1 u'(x)v'(x)dx \quad \text{et} \quad T(v) = \int_0^1 f(x)v(x)dx$$

La formulation variationnelle associée (notons que a est clairement symétrique) consiste à trouver $u \in H$ tel que :

$$J(u) = \min_{v \in H} J(v) \quad \text{avec} \quad J(v) = \frac{1}{2}a(u,v) - T(v)$$

Le fait que a soit une forme bilinéaire continue symétrique et coercive etque $T \in H'$ a été prouvé (dans le cas plus général de la dimension quelconque) lors de la démonstration de la proposition 3.5.

Exercice 40. 1. Comme $f \in C(\mathbb{R},\mathbb{R})$, et comme $-u'' = f$, on a $u \in C^2(\mathbb{R},\mathbb{R})$. Or $u_0 \in C^2(\mathbb{R},\mathbb{R})$ et $u_0'' = 0$; de même, $u_1 \in C^2(\mathbb{R},\mathbb{R})$ et $u_1'' = 2(b-a)$. Les fonctions $\widetilde{u}$ et $\overline{u}$ doivent donc vérifier :

$$\begin{cases} -\widetilde{u}'' = f \\ \widetilde{u}(0) = 0 \\ \widetilde{u}(1) = 0. \end{cases} \quad \text{et} \quad \begin{cases} -\overline{u}'' = f + 2(b-a) \\ \overline{u}(0) = 0 \\ \overline{u}(1) = 0. \end{cases}$$

Donc $\widetilde{u}$ est l'unique solution du problème : trouver $\widetilde{u} \in H_0^1(\Omega)$ tel que

$$a(u,\varphi) = \widetilde{T}(\varphi) \qquad \forall \varphi \in H_0^1(\Omega)$$

avec

$$a(u,\varphi) = \int_0^1 u'(x)\varphi'(x)dx \quad \text{et} \quad \widetilde{T}(\varphi) = \int_0^1 f(x)\varphi(x)dx$$

et $\overline{u}$ est l'unique solution du problème $\overline{u} \in H_0^1(\Omega)$ tel que :

$$a(\overline{u},\varphi) = \overline{T}(\varphi) \qquad \forall \varphi \in H_0^1(\Omega) \quad \text{avec} \quad \overline{T}(\varphi) = \int_0^1 (f(x) + 2(b-a))\varphi(x)dx$$

Montrons maintenant que $u = v$. Remarquons que $w = u - v$ vérifie

$$\begin{cases} w" = 0 \\ w(0) = w(1) = 0 \end{cases}$$

ce qui prouve que $w \in H_0^1(\Omega)$ est solution de $a(w,\varphi) = 0$, $\forall \varphi \in H_0^1(\Omega)$ d'où $w = 0$.

2. Le même raisonnement s'applique pour u_0 et $u_1 \in C^2([0\,;1])$ tel que $u_0(0) = u_1(0) = a$ et $u_1(0) = u_1(1) = b$.

Exercice 41. On introduit les espaces :

$$H_{1,1}^1 = \{v \in H^1(]0\,;1[); v(1) = 1\} \tag{3.71}$$

$$H_{1,0}^1 = \{v \in H^1(]0\,;1[); v(1) = 0\} \tag{3.72}$$

Soit $u_0 :]0\,;1[\to \mathbb{R}$, définie par $u_0(x) = x$. On a bien $u_0(1) = 1$, et $u_0 \in H_{1,1}^1$. Cherchons alors u sous la forme $u = u_0 + \widetilde{u}$, avec $\widetilde{u} \in H_{1,0}^1$.

$$\int_0^1 u'(x)v'(x)\,\mathrm{d}x - u'(1)v(1) + u'(0)v(0) = \int_0^1 f(x)v\,\mathrm{d}x, \forall \in H_{1,0}^1.$$

Comme $v(1) = 0$ et $u'(0) = 0$, on obtient donc :

$$\int_0^1 u'(x)v'(x)\,\mathrm{d}x = \int_0^1 f(x)v(x)\,\mathrm{d}x,$$

ou encore :

$$\int_0^1 \widetilde{u}'(x)v'(x)\,\mathrm{d}x = \int_0^1 f(x)v(x)\,\mathrm{d}x - \int_0^1 u_0'(x)v'(x)\,\mathrm{d}x = \int_0^1 f(x)v(x)\,\mathrm{d}x - \int_0^1 v'(x)\,\mathrm{d}x.$$

car $u_0' = 1$.

Exercice 42. Soit $v \in C_c^\infty([0\,;1])$, on multiplie la première équation de (3.48) et intègre par parties :

$$\int_0^1 u'(x)v'(x)\,\mathrm{d}x - u'(1)v(1) + u'(0)v(0) + \int_0^1 u(x)v(x)\,\mathrm{d}x = \int_0^1 f(x)v(x)\,\mathrm{d}x.$$

En tenant compte des conditions aux limites sur u en 0 et en 1, on obtient :

$$\int_0^1 u'(x)v'(x)\,\mathrm{d}x + \int_0^1 u(x)v(x)\,\mathrm{d}x + u(0)v(0) = \int_0^1 f(x)\varphi(x)\,\mathrm{d}x - v(1). \tag{3.73}$$

Pour trouver une formulation faible (ou variationnelle) il faut commencer par trouver un espace de Hilbert pour les fonctions duquel (3.73) ait un sens, et qui soit compatible avec les conditions aux limites. Comme $f \in L^2(]0\,;1[)$, le second membre de (3.73) est bien défini dès que $v \in L^2(]0\,;1[)$. De même, le premier membre de (3.73) est bien défini dès que $u \in H^1(]0\,;1[$ et $v \in H^1(]0\,;1[) \overset{def}{=} \{u \in L^2(]0\,;1[; Du \in L^2(]0\,;1[)$. Il est donc naturel de choisir $H = H(]0\,;1[)$. On obtient ainsi la formulation faible suivante :

$$\begin{cases} u \in H = \{u \in H(\Omega)\}, \\ a(u,v) = T(v), \forall v \in H, \end{cases}$$

où $a(u,v) = \int_0^1 u'(x)v'(x)\,\mathrm{d}x + \int_0^1 u(x)v(x)dx + u(0)v(0)$ et $T(v) = \int_0^1 f(x)v(x)\,\mathrm{d}x - v(1)$.

La formulation variationnelle associée (notons que a est clairement symétrique), s'écrit :

$$\begin{cases} \text{Trouver } u \in H, \\ J(u) = \min\limits_{v\in H} J(v) \end{cases}$$

avec $J(v) = \frac{1}{2}a(u,v) - T(v)$

Pour montrer l'existence et l'unicité des solutions de (3.73), on cherche à appliquer le théorème de Lax–Milgram. On remarque d'abord que T est bien une forme linéaire sur H, et que de plus, par l'inégalité de Cauchy–Schwarz, :

$$|T(v)| = |\int_0^1 f(x)v(x)\,\mathrm{d}x| + |v(1)| \leqslant \|f\|_{L^2(]0\,;1[)}\|v\|_{L^2(]0\,;1[)} + |v(1)|. \tag{3.74}$$

Montrons maintenant que $|v(1)| \leqslant 2\|v\|_{H^1(]0\,;1[)}$. Ce résultat est une conséquence du théorème de trace, voir cours d'EDP. Dans le cas présent, comme l'espace est de dimension 1, la démonstration est assez simple en remarquant que comme $v \in H^1(]0\,;1[)$, on peut écrire que v est intégrale de sa dérivée. On a en particulier :

$$v(1) = v(x) + \int_x^1 v'(t)dt,$$

et donc par l'inégalité de Cauchy-Schwarz,

$$|v(1)| = |v(x)| + \int_x^1 |v'(t)|dt \leqslant |v(x)| + \|v'\|_{L^2(]0\,;1[)}.$$

En intégrant cette inégalité entre 0 et 1 on obtient :

$$|v(1)| \leqslant \|v(x)\|_{L^1(]0\,;1[)} + \|v'\|_{L^2(]0\,;1[)}.$$

Or $\|v\|_{L^1(]0\,;1[)} \leqslant \|v(x)\|_{L^2(]0\,;1[)}$. De plus

$$\|v\|_{L^2(]0\,;1[)} + \|v'\|_{L^2(]0\,;1[)} \leqslant 2\max(\|v(x)\|_{L^2(]0\,;1[)}, \|v'\|_{L^2(]0\,;1[)})$$

on a donc

$$\Big(\|v\|_{L^2(]0\,;1[)} + \|v'\|_{L^2(]0\,;1[)}\Big)^2 \leqslant 4\max(\|v(x)\|^2_{L^2(]0\,;1[)}, \|v'\|^2_{L^2(]0\,;1[)}) \leqslant 4(\|v\|^2_{L^2(]0\,;1[)} + \|v'\|^2_{L^2(]0\,;1[)}).$$

On en déduit que

$$|v(1)| \leqslant \|v\|_{L^2(]0\,;1[)} + \|v'\|_{L^2(]0\,;1[)} \leqslant 2\|v\|_{H^1(]0\,;1[)}.$$

En reportant dans (3.74), on obtient :

$$|T(v)| \leqslant (\|f\|_{L^2(]0\,;1[)} + 2)\|v\|_{H^1(]0\,;1[)}$$

ce qui montre que T est bien continue.

Remarquons que le raisonnement effectué ci–dessus pour montrer que $|v(1)| \leqslant 2\|v\|_{H^1(]0\,;1[)}$ s'applique de la même manière pour montrer que

$$|v(a)| \leqslant 2\|v\|_{H^1(]0\,;1[)} \text{ pour tout } a \in [0\,;1]. \tag{3.75}$$

Ceci est une conséquence du fait que $H^1(]0\,;1[)$ s'injecte continûment dans $C([0\,;1])$.

Il est clair que a est une forme bilinéaire symétrique (notons que le caractère symétrique n'est pas nécessaire pour l'application du théorème de Lax–Milgram). Montrons que a est continue. On a :

$$\begin{aligned}|a(u,v)| &\leqslant \int_0^1 |u'(x)v'(x)|\,\mathrm{d}x + \int_0^1 |u(x)||v(x)|\,\mathrm{d}x + |u(0)||v(0)| \\ &\leqslant \|u'\|_{L^2(]0\,;1[)}\|v'\|_{L^2(]0\,;1[)} + \|u\|_{L^2(]0\,;1[)}\|v\|_{L^2(]0\,;1[)} + |u(0)||v(0)|\end{aligned}$$

Grâce à (3.75), on en déduit que

$$\begin{aligned}|a(u,v)| &\leqslant \|u'\|_{L^2(]0\,;1[)}\|v'\|_{L^2(]0\,;1[)} + \|u\|_{L^2(]0\,;1[)}\|v\|_{L^2(]0\,;1[)} + 4\|u\|_{H^1(]0\,;1[)}\|v\|_{H^1(]0\,;1[)} \\ &\leqslant 6\|u\|_{H^1(]0\,;1[)}\|v\|_{H^1(]0\,;1[)}.\end{aligned}$$

On en déduit que a est continue. Soit $u \in H^1(]0\,;1[)$; par définition de a, on a :

$$a(u,u) = \int_0^1 (u'(x))^2\,\mathrm{d}x + \int_0^1 (u(x))^2\,\mathrm{d}x + u(0)^2 \geqslant \|u\|_{H^1(]0\,;1[)}.$$

Ceci prouve que la forme a est coercive. Par le théorème de Lax–Milgram, on en déduit l'existence et l'unicité des solutions faibles de (3.73).

Exercice 45. Soit $\varphi \in H = \{v \in H^1(\Omega) : u = 0 \text{ sur } \Gamma_0\}$.

Multiplions la première équation de (3.53) par $\varphi \in H$. On obtient :

$$\int_\Omega \nabla u(x) \cdot \nabla\varphi(x)\,\mathrm{d}x + \int_{\Gamma_0\cup\Gamma_1} \nabla u \cdot \boldsymbol{n}(x)\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)dx$$

et comme $\nabla u.n = 0$ sur Γ_1 et $\varphi = 0$ sur Γ_0, on obtient donc

$$\int_\Omega \nabla u(x) \cdot \nabla\varphi(x) = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x.$$

On obtient donc la formulation faible pour laquelle il faut trouver $u \in H$ tel que

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

Notons que cette formulation ne diffère de la formulation faible du problème (3.1) que par la donnée de la condition aux limites de Dirichlet sur Γ_0 et non $\partial\Omega$ dans l'espace H. La condition de Neumann homogène est implicitement prise en compte dans la formulation faible.

La démonstration du fait que cette formulation satisfait les hypothèses du théorème de Lax–Milgram est similaire à celle de la proposition 3.5 en utilisant, pour la coercivité, le fait que les fonctions à trace nulle sur une partie du bord de Ω (de mesure non nulle) vérifient encore l'inégalité de Poincaré.

Exercice 46.

1. Multiplions la première équation de (3.54) par $\varphi \in C^\infty(\Omega)$ et intégrons sur Ω. Par la formule de Green, on obtient :

$$\int_\Omega p(x)\nabla u(x) \cdot \nabla\varphi(x)\,\mathrm{d}x - \int_{\partial\Omega} p(x)\nabla u(x) \cdot \boldsymbol{n}(x)\varphi(x)\,\mathrm{d}x + \int_\Omega q(x)u(x)\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x.$$

En tenant compte des conditions aux limites sur u et en prenant φ nulle sur Γ_0, on obtient alors :

$$a(u,\varphi) = T(\varphi)$$

avec :

$$a(u,\varphi) = \int_\Omega (p(x)\nabla u(x).\nabla\varphi(x) + q(x)u(x)\varphi(x))\,\mathrm{d}x + \int_{\Gamma_1} \sigma(x)u(x)\varphi(x)d\gamma(x), \tag{3.76}$$

et

$$T(\varphi) = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x + \int_{\Gamma_1} g_1(x)\varphi(x)d\gamma(x). \tag{3.77}$$

Pour assurer la condition aux limites de type Dirichlet non homogène, on choisit donc $u \in H^1_{\Gamma_0,g_0}(\Omega) = \{u \in H^1(\Omega); u = g_0 \text{ sur } \Gamma_0\}$, qu'on peut aussi décomposer en : $u = \widetilde{u}+$ avec $\widetilde{u} \in H^1_{\Gamma_0,g_0}(\Omega)$ ("relèvement" de u) et $u_0 \in H^1_{\Gamma_0}(\Omega) = \{u \in H^1(\Omega); u = 0 \text{ sur } \Gamma_0\}$, Une formulation faible naturelle est alors :

$$\begin{cases} u \in H^1_{\Gamma_0,g_0}(\Omega) \\ a(u,v) = T(v), \forall v \in H^1_{\Gamma_0,0}(\Omega), \end{cases}$$

ou encore :

$$\begin{cases} u = u_0 + \widetilde{u} \\ \widetilde{u} \in H^1_{\Gamma_0,0}(\Omega) \\ a(\widetilde{u},v) = T(v) - T(u_0), \forall v \in H^1_{\Gamma_0,0}(\Omega), \end{cases} \tag{3.78}$$

L'espace $H = H^1_{\Gamma_0,0}(\Omega)$ muni de la norme H^1 est un espace de Hilbert. Il est facile de montrer que l'application a définie de $H \times H$ dans $\mathbb{R}$ est bilinéaire. Montrons qu'elle est continue ; soient $(u,v) \in H \times H$, alors

$$a(u,v) \leqslant \|p\|_{L^\infty(\Omega)}\|\nabla u\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)} + \|q\|_{L^\infty(\Omega)}\|u\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)} + \sigma\|\gamma(u)\|_{L^2(\Omega)}\|\gamma(v)\|_{L^2(\partial\Omega)}.$$

Par le théorème de trace, il existe C_Ω ne dépendant que de Ω tel que

$$\|\gamma(u)\|_{L^2(\partial\Omega)} \leqslant C_\Omega\|u\|_{H^1(\Omega)} \text{ et } \|\gamma(v)\|_{L^2(\partial\Omega)} \leqslant C_\Omega\|v\|_{H^1(\Omega)}.$$

On en déduit que

$$a(u,v) \leqslant \left(\|p\|_{L^\infty(\Omega)} + \|q\|_{L^\infty(\Omega)} + \sigma C_\Omega^2\right)\|u\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)},$$

ce qui montre que a est continue. La démonstration de la coercivité de a est similaire à la démonstration du lemme 3.12. Enfin, il est facile de voir que T définie par (3.77) est une forme linéaire. On en déduit que le théorème de Lax–Milgram s'applique.

2. On a déjà vu à la question précédente que si u est solution de (3.78), alors u est solution de (3.54). Il reste à démontrer la réciproque. Soit donc u solution de (3.78), et soit $\varphi \in C_c^\infty(\Omega)(\subset H)$. En utilisant la formule de Green, et en notant que φ est nulle sur $\partial\Omega$, on obtient :

$$\int_\Omega (-\operatorname{div}(p\nabla u)(x) + q(x)u(x) - f(x))\varphi(x)\,\mathrm{d}x = 0, \forall\varphi \in C_0^\infty(\Omega).$$

Comme $u \in C^2(\overline{\Omega})$, on en déduit que :

$$-div(p\nabla u)(x) + q(x)u(x) - f(x) = 0, \forall x \in \Omega.$$

Comme $u \in H^1_{\Gamma_0,g_0}$ et $u \in C^2(\overline{\Omega})$, on a aussi $u = g_0$ sur Γ_0. Prenons maintenant $\varphi \in H^1_{\Gamma_0,0}$ on a :

$$\int_\Omega p(x)\nabla u(x)\nabla\varphi(x)\,\mathrm{d}x + \int_\Omega q(x)u(x)\varphi(x)\,\mathrm{d}x + \int_{\Gamma_1}\sigma(x)u(x)\varphi(x)d\gamma(x) = \int_\Omega f(x)dx + \int_{\Gamma_1} g(x)\varphi(x)d\gamma(x).$$

Par intégration par parties, il vient donc :

$$\begin{aligned}\int_\Omega -\operatorname{div}(p(x)\nabla u(x))\varphi(x)\,\mathrm{d}x + \int_{\Gamma_1} p(x)\nabla u(x)\cdot \boldsymbol{n}(x)\varphi(x)\,\mathrm{d}x \\ + \int_{\Gamma_1}\sigma(x)u(x)\varphi(x)d\gamma(x) + \int_\Omega q(x)u(x)\varphi(x)\,\mathrm{d}x \\ = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x + \int_{\Gamma_1} g_1(x)\varphi(x)d\gamma(x).\end{aligned}$$

Or on a montré que $-\operatorname{div}(p\nabla u) + qu = 0$. On a donc :

$$\int_{\Gamma_1}(p(x)\nabla u(x)\cdot\boldsymbol{n}(x) + \sigma u(x) - g_1(x))\varphi(x)d\gamma(x) = 0, \quad \forall, \varphi \in H^1_{\Gamma_0,g_0}.$$

On en déduit que $p\nabla u\cdot\boldsymbol{n} + \sigma u - g_1 = 0$ sur Γ_1 et donc u vérifie bien (3.54).

Exercice 48.

1. Pour montrer que le problème (3.59) admet une unique solution, on aimerait utiliser le théorème de Lax–Milgram. Comme $V_h \subset V$ un Hilbert, que a une forme bilinéaire continue sur $V\times V$, et que L est une forme linéaire continue sur V, il ne reste qu'à montrer la coercivité de a sur V_h. Mais la condition (3.58) n'entraîne pas la coercivité de a sur V_h. Il suffit pour s'en convaincre de considérer la forme bilinéaire $a(u,v) = u_1u_2 - v_1v_2$ sur $V_h = \mathbb{R}^2$, et de vérifier que celle-ci vérifie la condition (3.58) sans être pour autant coercive. Il faut trouver autre chose...

On utilise le théorème représentation de F. Riesz, que l'on rappelle ici : Soit H un espace de Hilbert et T une une forme lineaire continue sur H, alors il existe un unique $\psi\in H$ tel que $T(v) = (\psi, v)\ \forall v\in H$. Soit A l'opérateur de V_h dans V_h défini par $a(u,v) = (Au, v)$ pour tout $v\in V_h$. Comme L est une forme linéaire continue sur $V_h\subset V$, par le théorème de Riesz, il existe un unique $\psi\in V_h$ tel que $L(v) = (\psi, v)$, pour tout $v\in V_h$. Le problème (3.59) s'écrit donc

$$\text{Trouver } u\in V_h \text{ tel que } (Au, v) = (\psi, v), \text{ pour tout } v\in V_h.$$

Si A est bijectif de V_h dans V_h, alors $u = A^{-1}\psi_u$ est donc la solution unique de (3.59). Comme V_h est de dimension finie, il suffit de montrer que A est injectif. Soit donc $w\in V_h$ tel que $Aw = 0$, on a dans ce cas $\|Aw\| = 0$ et donc

$$\sup_{(v\in V_h,,\|v\|=1)} a(w,v) = 0.$$

Or par la condition (3.58), on a

$$\inf_{w\in V_h, w\neq 0}\ \sup_{(v\in V_h,\|v\|=1)} a(w,v) \geqslant \beta_h > 0.$$

On en déduit que $w = 0$, donc que A est bijectif et que le problème (3.59) admet une unique solution. On peut remarquer de plus que si A est inversible,

$$\inf_{v\in V_h,\|v\|=1}\ \sup_{v\in V_h,\|v\|=1} a(w,v) = \|A^{-1}\|^{-1}, \tag{3.79}$$

et donc si (3.58) est vérifiée, alors

$$\|A^{-1}\| \leqslant \frac{1}{\beta_h} \tag{3.80}$$

En effet, par définition,

$$\begin{aligned}\|A^{-1}\|^{-1} &= \left(\sup_{v\in V_h, v\neq 0}\frac{\|A^{-1}v\|}{\|v\|}\right)^{-1} = \inf_{v\in V_h, v\neq 0}\frac{\|v\|}{\|A^{-1}v\|} \\ &= \inf_{f\in V_h, v\neq 0}\frac{\|Af\|}{\|f\|} = \inf_{f\in V_h,\|f\|=1}\|Af\| = \inf_{f\in V_h,\|f\|=1}\ \sup_{w\in V_h,\|w\|=1}(Af, w).\end{aligned}$$

2. Soit $v \in V_h$, $v \neq 0$; par l'inégalité triangulaire, on a :

$$\|u - u_h\| \leqslant \|u - v\| + \|v - u_h\|. \tag{3.81}$$

Mais grâce à (3.80), on a :

$$\begin{aligned}
\|v - u_h\| &= \|A^{-1}A(v - u_h)\| \\
&\leqslant \frac{1}{\beta_h}\|A(v - u_h)\| \\
&\leqslant \frac{1}{\beta_h} \sup_{w \in V_h, \|w\|=1)} a(v - u_h, w) \\
&\leqslant \frac{1}{\beta_h} \sup_{w \in V_h, \|w\|=1)} (a(v, w) - a(u_h, w)) \\
&\leqslant \frac{1}{\beta_h} \sup_{w \in V_h, \|w\|=1)} (a(v, w) - a(u, w)),
\end{aligned}$$

car $a(u_h, w) = L(w) = a(u, w)$. On a donc

$$\begin{aligned}
\|v - u_h\| &\leqslant \frac{1}{\beta_h} \sup_{w \in V_h, \|w\|=1)} a(v - u, w) \\
&\leqslant \frac{1}{\beta_h} \sup_{w \in V_h, \|w\|=1)} M\|v - u\|\|w\| \\
&\leqslant \frac{M}{\beta_h}\|v - u\|.
\end{aligned}$$

En reportant dans (3.81), il vient alors :

$$\|u - u_h\| \leqslant \|u - v\| + \frac{M}{\beta_h}\|v - u\|, \ \forall v \in V_h,$$

et donc

$$\|u - u_h\| \leqslant \left(1 + \frac{M}{\beta_h}\right) \inf_{v \in V_h} \|u - v\|.$$

Exercice 50

1. Soit K un volume de contrôle du maillage volumes finis. On intègre (3.1) sur K et en utilisant la formule de Stokes, on obtient :

$$\sum_{\sigma \in \mathcal{E}_K} \int_\sigma \nabla u(x) \cdot \boldsymbol{n}_{K,\sigma} d\gamma(x) = m(K) f_K,$$

avec les notations du paragraphe 1.1.2.

On approche cette équation par :

$$\sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma} = m(K) f_K,$$

où $F_{K,\sigma}$ est le flux numérique à travers σ, qu'on approche par :

$$F_{K,\sigma} = \begin{cases} \dfrac{m(\sigma)}{d_{K,\sigma} + d_{L,\sigma}}(u_K - u_L) \text{ si } \sigma \in \mathcal{E}_{int} \cap \mathcal{E}_K, \\ \\ \dfrac{m(\sigma)}{d_{K,\sigma}} u_K \text{ si } \sigma \in \mathcal{E}_{ext} \cap \mathcal{E}_K. \end{cases}$$

On obtient donc bien le schéma (3.66) - (3.67)

2. Soit $v = (v_K)_{K \in \mathcal{T}} \in H_{\mathcal{T}}(\Omega)$ une fonction constante par volumes de contrôle.

On multiplie l'équation (3.66) par V_K et on somme sur K. On obtient :

$$\sum_{K\in\mathcal{T}}\left(\sum_{\substack{\sigma\in\mathcal{E}_{int}\\ \sigma=K|L}}\tau_\sigma(u_K-u_L)v_K+\sum_{\sigma\in\mathcal{E}_{ext}\cap\mathcal{E}_K}\tau_\sigma u_K V_K\right)=\sum_K m(K)f_K v_K.$$

Remarquons maintenant que le premier membre de cette égalité est aussi égal, en sommant sur les arêtes du maillage à :

$$\sum_{\substack{\sigma\in\mathcal{E}\\ \sigma=K|L}}(\tau_\sigma(u_K-u_L)v_K)+\tau_\sigma(u_L-u_K)v_L)+\sum_{\tau\in\mathcal{E}_{ext}}\tau_\sigma u_{K_\sigma}v_{K_\sigma}$$

où K_σ désigne le volume de contrôle dont σ est une arête (du bord) dans la deuxième sommation. On obtient donc :

$$a_\mathcal{T}(u,v)=T_\mathcal{T}(V),\forall v\in H_\mathcal{T}(\Omega), \tag{3.82}$$

avec :

$$a_\mathcal{T}(u,v)=\sum_{\substack{\sigma\in\mathcal{E}\\ \sigma=K|L}}\tau_\sigma(u_K-u_L)(v_K-v_L)+\sum_{\sigma\in\mathcal{E}_{ext}}u_{K_\sigma}v_{K,\sigma}\text{et } T_\mathcal{T}(v)=\sum_K m(K)f_K v_K.$$

On a donc montré que si $u=(u_K)_{K\in\mathcal{T}}$ la solution de (3.66) - (3.67), alors u est solution de (3.82). Montrons maintenant la réciproque. Soit 1_K la solution caractéristique du volume de contrôle K, définie par

$$1_K(x)=\begin{cases}1 \text{ si } x\in K\\ 0 \text{ sinon },\end{cases}$$

Prenons $v=1_K$ dans (3.82), on obtient alors

$$\sum_{\sigma\in\mathcal{E}_{int}}\tau_\sigma(u_K-u_L)+\sum_{\sigma\in\mathcal{E}_{ext}\cap\mathcal{E}_K}\tau_\sigma u_K=m(K)f_K.$$

On retrouve donc bien (3.66).

Notons qu'en faisant ceci, on a introduit une discrétisation de la formulation faible (3.6) par une méthode de discrétisation non conforme, puisque $H_\mathcal{T}\not\subset H^1(\Omega)$.

4

Éléments finis de Lagrange

La construction d'une méthode d'éléments finis nécessite la donnée d'un maillage, de nœuds et d'un espace de polynômes, qui doivent être choisis de manière cohérente. Les éléments finis de type Lagrange font intervenir comme *degrés de liberté* (c'est-à-dire les valeurs qui permettent de déterminer entièrement une fonction) les valeurs de la fonction aux nœuds. Ils sont très largement utilisés dans les applications. Il existe d'autres familles d'éléments finis, comme les éléments finis de type Hermite qui font intervenir les valeurs des dérivées directionnelles.

Dans ce cours, nous n'aborderons que les éléments finis de type Lagrange et nous renvoyons aux ouvrages cités en introduction pour d'autres éléments.

4.1 Espace d'approximation

4.1.1 Cohérence *locale*

Soit $\mathcal{T}$ un maillage de Ω ; pour tout élément K de $\mathcal{T}$, on note Σ_K l'ensemble des nœuds de l'élément. On suppose que chaque élément a N_ℓ nœuds K : $\Sigma_K = \{a_1, \ldots, a_{N_\ell}\}$, qui ne sont pas forcément ses sommets.

On note $\mathcal{P}$ un espace de dimension finie constitué de polynômes, qui définit la méthode d'éléments finis choisie.

Définition 4.1 — Unisolvance, élément fini de Lagrange. Soit $\mathcal{T}$ un maillage de Ω ; soit K un élément de $\mathcal{T}$ et $\Sigma_K = (a_i)_{i=1,\ldots,N_\ell}$ un ensemble de nœuds de K. Soit $\mathcal{P}$ un espace de polynômes de dimension finie. On dit que le triplet $(K, \Sigma_K, \mathcal{P})$ est un élément fini de Lagrange si l'ensemble Σ_K est $\mathcal{P}$-unisolvant, c'est-à-dire si pour tout $(\alpha_1, \ldots, \alpha_{N_\ell}) \in \mathbb{R}^{N_\ell}$, il existe un unique élément $f \in \mathcal{P}$ tel que $f(a_i) = \alpha_i$, $\forall i = 1 \ldots N_\ell$. Pour $i = 1, \ldots, N_\ell$, on appelle degré de liberté la forme linéaire ζ_i définie par $\zeta_i(p) = p(a_i)$, $p \in \mathcal{P}$. La propriété d'unisolvance équivaut à dire que la famille $(\zeta_i)_{i=1,\ldots,N_\ell}$ forme une base de l'espace dual $\mathcal{P}'$ de $\mathcal{P}$.

La $\mathcal{P}$-unisolvance revient à dire que toute fonction de $\mathcal{P}$ est entièrement déterminée par ses valeurs aux nœuds de K.

Exemple 4.2 — Élément fini P1**.** Prenons en dimension 1 l'élément $K = [a_1, a_2]$ avec $\Sigma_K = \{a_1, a_2\}$ et $\mathcal{P} = \mathcal{P}_1$ (ensemble des polynômes de degré inférieur ou égal à 1). D'après la définition précédente, l'ensemble Σ_K est $\mathcal{P}$-unisolvant s'il existe une unique fonction f de $\mathcal{P}$ telle que :

$$\begin{cases} f(a_1) = \alpha_1, \\ f(a_2) = \alpha_2. \end{cases}$$

Or toute fonction f de $\mathcal{P}$ s'exprime sous la forme $f(x) = \lambda x + \mu$ et le système :

$$\begin{cases} \lambda a_1 + \mu = \alpha_1, \\ \lambda a_2 + \mu = \alpha_2, \end{cases}$$

détermine λ et μ de manière unique.

De même si on considère le cas $d = 2$. On prend comme élément K un triangle et comme nœuds les trois sommets a_1, a_2 et a_3 du triangle. Soit $\mathcal{P} = \mathcal{P}_1 = \{f : \mathbb{R} \to \mathbb{R}; f(x) = \lambda x_1 + \mu x_2 + \nu\}$ l'ensemble des fonctions affines. Alors le triplet $(K, \Sigma_K, \mathcal{P})$ est un élément fini de Lagrange car $f \in \mathcal{P}$ est entièrement déterminée par $f(a_1)$, $f(a_2)$ et $f(a_3)$.

Définition 4.3 — Fonctions de base locales. Si $(K, \Sigma_K, \mathcal{P})$ est un élément fini de Lagrange, alors toute fonction f de $\mathcal{P}$ peut s'écrire :

$$f = \sum_{i=1}^{N_\ell} f(a_i) f_i,$$

avec $f_i \in \mathcal{P}$ et $f_i(a_j) = \delta_{ij}$. Les fonctions f_i sont appelées fonctions de base locales.

Pour l'élément fini de Lagrange P1 en dimension 2, les fonctions de base locales sont décrites sur la figure 4.1.

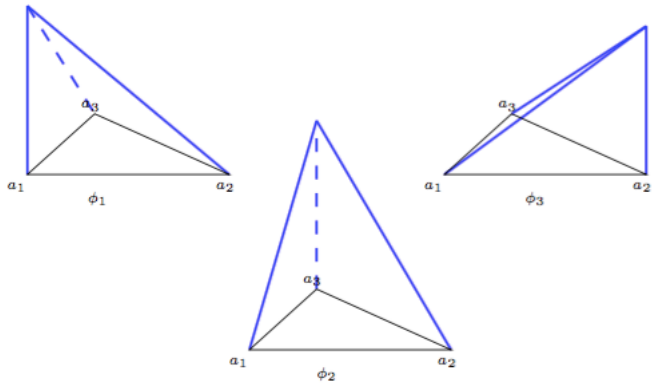


Figure 4.1 – Fonctions de base locales pour l'élément fini de Lagrange P1 en dimension 2

Remarque 4.4 — Condition nécessaire d'unisolvance. Pour que le triplet $(K, \Sigma_K, \mathcal{P})$ soit un élément fini de Lagrange, il faut, mais il ne suffit pas, que $\dim \mathcal{P} = \operatorname{card} \Sigma_K$. Par exemple si $\mathcal{P} = \mathcal{P}_1$ et qu'on prend comme nœuds du triangle deux sommets et le milieu de l'arête joignant les deux sommets, (voir figure 4.2), $(K, \Sigma_K, \mathcal{P})$ n'est pas un élément fini de Lagrange.

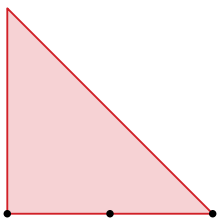


Figure 4.2 – Exemple de triangle à trois nœuds qui n'est pas un élément fini de Lagrange

Définition 4.5 — Interpolée. Soit $\mathcal{T}$ un maillage de Ω et $K \in \mathcal{T}$; on suppose que $(K, \Sigma_K, \mathcal{P})$ un élément fini de Lagrange et soit $v \in C(K, \mathbb{R})$. L'interpolée de v sur K est la fonction $\Pi_K v \in \mathcal{P}$ définie par :

$$\Pi_K v = \sum_{i=1}^{N_\ell} v(a_i) f_i,$$

où $(a_i)_{i=1,\dots,N_\ell}$ sont noeuds de K et $(f_i)_{i=1,\dots,N_\ell}$ les fonctions de base locales associées.

Soit maintenant $v \in C(\Omega, \mathbb{R})$, on suppose que pour tout $K \in \mathcal{T}$, $(K, \Sigma_K, \mathcal{P})$ est un élément fini de Lagrange ; alors l'interpolée de v sur Ω est la fonction continue Πv dont la restriction à chaque élément K est la fonction $\Pi_K v$.

On montre sur la figure 4.3 un exemple d'interpolée pour l'élément fini de Lagrange P1 en dimension 1. L'étude de $\|v - \Pi v\|$ va nous permettre d'établir une majoration de l'erreur de consistance

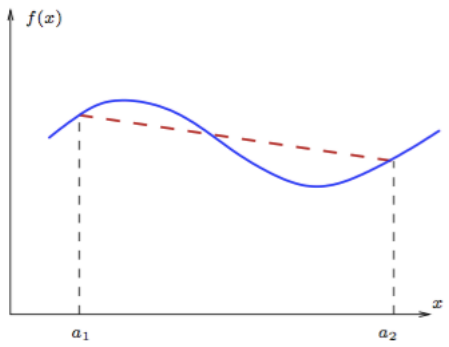


Figure 4.3 – Interpolée P1 sur $[a_1, a_2]$ (en trait pointillé) d'une fonction régulière (en trait continu)

$d(u, H_N)$.

Proposition 4.6 — Critère de détermination. Soit $(K, \Sigma, \mathcal{P})$ un triplet constitué d'un élément, d'un ensemble de nœuds et d'un espace de polynômes, tel que $\dim \mathcal{P} = \operatorname{card} \Sigma = N_\ell$. Si

$$\exists! f \in \mathcal{P}; f = 0 \text{ sur } \Sigma \tag{4.1}$$

ou si $\forall i \in \{1 \dots N_\ell\}$, $\exists f_i \in \mathcal{P}$, $f_i(a_j) = \delta_{ij}$ alors $(K, \Sigma, \mathcal{P})$ est un élément fini de Lagrange.

Démonstration. Soit :

$$\psi : \mathcal{P} \to \mathbb{R}^{N_\ell}$$
$$f \mapsto (f(a_i))_{i=1,N_\ell}^t$$

L'application ψ est linéaire de $\mathcal{P}$ dans $\mathbb{R}^{N_\ell}$, et, par hypothèse card $s\Sigma = \dim \mathcal{P}$. Donc ψ est une application linéaire continue de $\mathcal{P}$ dans $\mathbb{R}^{N_\ell}$, avec $\dim \mathcal{P} = \dim(\mathbb{R}^{N_\ell}) = N_\ell$. Si $(K, \Sigma, \mathcal{P})$ vérifie la condition (4.1) alors ψ est injective. En effet, si $\psi(f) = 0$, alors $f(a_i) = 0$, $\forall i = 1, \dots, N_\ell$ et donc par hypothèse, $f = 0$. Donc ψ est une application linéaire, ψ est injective de $\mathcal{P}$ dans $\mathbb{R}^{N_\ell}$ avec $\dim \mathcal{P} = N_\ell$. On en déduit que ψ est bijective. Donc toute fonction de $\mathcal{P}$ est entièrement déterminée par ses valeurs aux nœuds : $(K, \Sigma, \mathcal{P})$ est donc un élément fini de Lagrange.

On montre facilement que si la deuxième condition est vérifiée alors ψ est surjective. Donc ψ est bijective et $(K, \Sigma, \mathcal{P})$ est un élément fini de Lagrange. •

Proposition 4.7 — CS pour un élément fini de Lagrange. Soit $(\overline{K}, \overline{\Sigma}, \overline{\mathcal{P}})$ un élément fini de Lagrange, où $\overline{\Sigma}$ est l'ensemble des nœuds de $\overline{K}$ et $\overline{\mathcal{P}}$ un espace de fonctions de dimension finie et soit F une bijection de $\overline{K}$ dans K, où K est une maille d'un maillage éléments finis. On pose $\Sigma = F(\overline{\Sigma})$ et $\mathcal{P} = \{f : K \to \mathbb{R}; f \circ F \in \overline{\mathcal{P}}\}$ (voir figure 4.4). Alors le triplet $(K, \Sigma, \mathcal{P})$ est un élément fini de Lagrange.

Démonstration. On suppose que les hypothèses de la proposition sont réalisées et on veut montrer que l'ensemble Σ est $\mathcal{P}$-unisolvant. Soient $\Sigma = (a_1, \dots, a_{N_\ell})$ et $(\alpha_1, \dots, \alpha_{N_\ell}) \in \mathbb{R}^{N_\ell}$. On veut montrer qu'il existe une unique fonction $f \in \mathcal{P}$ telle que :

$$f(a_i) = \alpha_i \quad \forall i = 1, \dots, N_\ell$$

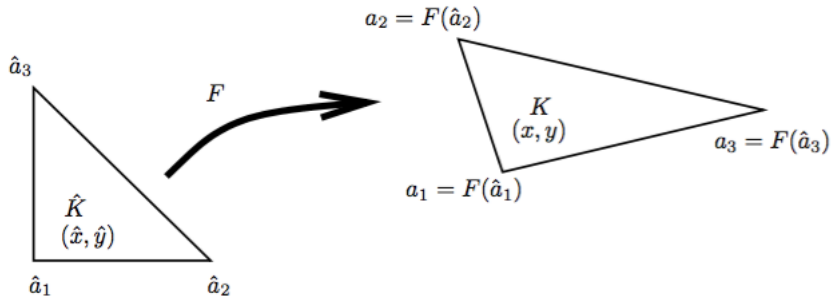


FIGURE 4.4 – Transformation F

Comme par hypothèse $(\overline{K}, \overline{\Sigma}, \overline{\mathcal{P}})$ est un élément fini de Lagrange, l'ensemble $\overline{\Sigma}$ est $\overline{\mathcal{P}}$)-unisolvant et il existe donc une unique fonction $\bar{f} \in \overline{\mathcal{P}}$ telle que :

$$\bar{f}(\bar{a}_i) = \alpha_i \quad \forall i = 1, \ldots, N_\ell$$

où $(\bar{a}_i)_{i=1,\ldots,N_\ell}$ désignent les nœuds de $\bar{K}$. Soit F la bijection de $\bar{K}$ sur K ; on pose $f = \bar{f} \circ F^{-1}$. Par hypothèse, $a_i = F(\bar{a}_i)$ et on a donc $f(a_i) = \bar{f} \circ F^{-1}(a_i) = \bar{f}(\bar{a}_i) = \alpha_i$. On a ainsi montré l'existence de f telle que $f(a_i) = \alpha_i$.

Montrons maintenant que f est unique. Supposons qu'il existe f et $g \in \mathcal{P}$ telles que :

$$f(a_i) = g(a_i) = \alpha_i \quad \forall i = 1, \ldots, N_\ell$$

Soit $h = f - g$ on a donc :

$$h(a_i) = 0 \quad \forall i = 1 \ldots N_\ell$$

On a donc $h \circ F(\bar{a}_i) = h(a_i) = 0$. Or $h \circ F \in \overline{\mathcal{P}}$ et comme l'ensemble $\bar{\Sigma}$ est $\overline{\mathcal{P}}$)-unisolvant, on en déduit que $h \circ F = 0$. Comme, pour tout $x \in K$, on a $h(x) = h \circ F \circ F^{-1}(x) = h \circ F(F^{-1}(x)) = 0$, on en conclut que $h = 0$. •

Définition 4.8 — Éléments affine-équivalents. Sous les hypothèses de la proposition 4.7, si la bijection F est affine, on dit que les éléments finis $(\hat{K}, \bar{\Sigma}, \overline{\mathcal{P}})$ et $(K, \Sigma, \mathcal{P})$ sont affine-équivalents.

Remarque 4.9 Soient $(\bar{K}, \bar{\Sigma}, \overline{\mathcal{P}})$ et $(K, \Sigma, \mathcal{P})$ deux éléments finis affine-équivalents. Si les fonctions de base locales de $(\bar{K}, \bar{\Sigma}, \overline{\mathcal{P}})$ (resp. de $(K, \Sigma, \mathcal{P})$) sont affines, alors celles de K (resp. $\bar{K}$) le sont aussi et on a :

$$\begin{cases} \bar{f}_i = f_i \circ F, \\ f_i = \bar{f}_i \circ F^{-1}, \end{cases} \qquad i = 1, \ldots, \operatorname{card} \Sigma.$$

La preuve de cette remarque fait l'objet de l'exercice 61.

Proposition 4.10 — Interpolation. Sous les hypothèses de la proposition 4.11, soient $\Pi_{\bar{K}}$ et Π_K les opérateurs d'interpolation respectifs sur $\bar{K}$ et K (voir définition 4.5). Soient $v \in C(K, \mathbb{R})$, alors on a $\Pi_K v \circ F = \Pi_{\bar{K}}(v \circ F)$.

Démonstration. Remarquons tout d'abord que $\Pi_K v \circ F$ et $\Pi_{\bar{K}}(v \circ F)$ sont toutes deux des fonctions définies de $\bar{K}$ à valeurs dans $\mathbb{R}$, voir figure 4.5. Remarquons ensuite que, par définition de l'interpolée, $\Pi_K v \in \mathcal{P}$. Comme $(\bar{K}, \bar{\Sigma}, \overline{\mathcal{P}})$ est l'élément de référence, on a donc $\Pi_K v \circ F \in \overline{\mathcal{P}}$. On a aussi, par définition de l'interpolée : $\Pi_{\bar{K}}(v \circ F) \in \overline{\mathcal{P}}$. On en déduit que $\Pi_K v \circ F$ et $\Pi_{\bar{K}}(v \circ F)$ sont toutes deux des fonctions de $\overline{\mathcal{P}}$. Comme l'ensemble $\bar{K}$ est $\bar{\Sigma}$)-unisolvant (car $(\bar{K}, \bar{\Sigma}, \overline{\mathcal{P}})$ est un élément fini de Lagrange), toute fonction de $\overline{\mathcal{P}}$ est uniquement déterminée par ses valeurs aux nœuds de $\bar{\Sigma}$. Pour montrer l'égalité de $\Pi_K v \circ F$ et $\Pi_{\bar{K}}(v \circ F)$, il suffit donc de montrer que :

$$\Pi_{\bar{K}}(v \circ F)(\bar{a}_i) = \Pi_K v \circ F(\bar{a}_i), \quad i = 1, \ldots, N_\ell, \tag{4.2}$$

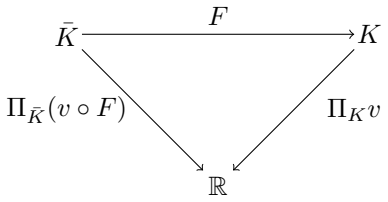


FIGURE 4.5 – Opérateurs d'interpolation $\Pi_{\bar{K}}$ et Π_K

où $N_\ell = \operatorname{card} \bar{\Sigma}$. Décomposons $\Pi_{\bar{K}}(v \circ F)$ sur les fonctions de base locales $(\bar{f}_j), j = 1, \ldots, N_\ell$; on obtient :

$$\Pi_{\bar{K}}(v \circ F)(\bar{a}_i) = \sum_{j=1}^{N_\ell} v \circ F(\bar{a}_j)\bar{f}_j(\bar{a}_i) = \sum_{j=1}^{N_\ell} v \circ F(\bar{a}_j)\delta_{ij} = v \circ F(\bar{a}_i) = v(a_i).$$

Mais on a aussi :

$$\Pi_K v \circ F(\bar{a}_i) = \Pi_K v(F(\bar{a}_i)) = \Pi_K v(a_i) = v(a_i)$$

d'où l'égalité. •

4.1.2 Construction de H_N et conformité

Nous allons considérer deux cas : le cas où l'espace H est l'espace H^1 tout entier et le cas où l'espace H est l'espace H_0^1

Cas des conditions de Dirichlet non homogènes

Plaçons-nous ici dans le cas où $H = H^1(\Omega)$, où $\Omega \subset \mathbb{R}^d$ est un ouvert borné polygonal (si $d = 2$, polyèdrique si $d = 3$). Soit $\mathcal{T}$ un maillage éléments finis, avec $\mathcal{T} = (K_\ell)_{\ell=1,\ldots,L}$, où les éléments finis K_ℓ sont fermés et tels que $\cup_{\ell=1}^{L} K_\ell = \bar{\Omega}$. Soit $\mathcal{S} = (S_i)_{i=1,\ldots,M}$ l'ensemble des nœuds du maillage éléments finis, avec $S_i \in \bar{\Omega}, \quad \forall i = 1, \ldots, M$. On cherche à construire une méthode d'éléments finis de Lagrange ; donc à chaque élément K_ℓ, $\ell = 1, \ldots, L$, est associé un ensemble de nœuds $\Sigma_\ell = \mathcal{S} \cap K_\ell$ et un espace $\mathcal{P}_\ell$ de polynômes. On veut que chaque triplet $(K_\ell, \Sigma_\ell, \mathcal{P}_\ell)$ soit un élément fini de Lagrange. On définit les fonctions de base globales $(\phi_i)_{i=1,\ldots,M}$ par :

$$\phi_i \,|_{K_\ell} \in \mathcal{P}_\ell \qquad \forall i = 1, \ldots, M; \qquad \forall \ell = 1; \ldots, L, \tag{4.3}$$

$$\phi_i(S_j) = \delta_{ij} \qquad \forall i = 1, \ldots, M, \qquad \forall j = 1, \ldots, M. \tag{4.4}$$

Chaque fonction ϕ_i est définie de manière unique, grâce au caractère unisolvant du couple $(\Sigma_\ell, \mathcal{P}_\ell)$, $\ell = 1, \ldots, M$. On pose $H_N = \operatorname{Vect}(\phi_1, \ldots, \phi_M)$; pour obtenir une méthode d'éléments finis conforme, il reste à s'assurer que $H_N \subset H^1$.

Une manière de construire l'espace H_N est de construire un maillage à partir d'un élément de référence, grâce à la proposition suivante, qui se déduit facilement de la proposition 4.7

Proposition 4.11 — Élément fini de référence. Soit $\mathcal{T}$ un maillage constitué d'éléments K. On appelle élément fini de référence un élément fini de Lagrange $(\bar{K}, \bar{\Sigma}, \overline{\mathcal{P}})$, où $\bar{\Sigma}$ est l'ensemble des nœuds de $\bar{K}$ et $\overline{\mathcal{P}}$ un espace de fonctions, de dimension finie, tel que, pour tout autre élément $K \in \mathcal{T}$, il existe une bijection $F : \bar{K} \to K$ telle que $\Sigma = F(\bar{\Sigma})$ et $\mathcal{P} = \{f : K \to \mathbb{R}; f \circ F \in \overline{\mathcal{P}}\}$ (voir figure 4.4). Le triplet $(K, \Sigma, \mathcal{P})$ est un élément fini de Lagrange.

Proposition 4.12 — Critère de conformité, cas H^1. Soit Ω un ouvert polygonal (ou polyèdrique) de $\mathbb{R}^d$, $d = 2$ ou 3. Soit $\mathcal{T} = (K_\ell)_{\ell=1,\ldots,L}$, un maillage éléments finis de $\Omega, \mathcal{S} = (S_i)_{i=1,\ldots,M}$ l'ensemble des nœuds de maillage (notons que les côtés de K_ℓ sont des arêtes en 2D et des faces en 3D). On se place sous les hypothèses de la proposition 4.11 ; soient $(\phi_i)_{i=1,\ldots,M}$ les fonctions de base globales, vérifiant (4.3) et (4.4) et on suppose de plus que les hypothèses suivantes sont vérifiées :

$$\text{Pour toute arête (ou face si d = 3) } \epsilon = K_{\ell_1} \cap K_{\ell_2}, \text{ on a } : \Sigma_{\ell_1} \cap \epsilon = \Sigma_{\ell_2} \cap \epsilon \text{ et } P_{\ell_1}\,|_\epsilon = P_{\ell_2}\big|_\epsilon, \tag{4.5}$$

où $\mathcal{P}_{\ell_1}|_\epsilon$ (resp. $\mathcal{P}_{\ell_2}|_\epsilon$) désigne l'ensemble des restrictions des fonctions de $\mathcal{P}_{\ell_1}$ (resp. $\mathcal{P}_{\ell_2}$) à ϵ),

$$\text{Si } \epsilon \text{ est un côté de } K_\ell, \text{ l'ensemble } \Sigma_\ell \cap \epsilon, \text{ est } \mathcal{P}_\ell|_\epsilon) - \text{unisolvant.} \tag{4.6}$$

alors $H_N \subset C(\bar{\Omega})$ et $H_N \subset H^1(\Omega)$. On a donc ainsi construit une méthode d'éléments finis conformes.

Démonstration. Pour montrer que $H_N \subset C(\bar{\Omega})$ et $H_N \subset H^1(\Omega)$, il suffit de montrer que pour chaque fonction de base globale ϕ_i, on a $\phi_i \in C(\bar{\Omega})$ et $\phi_i \in H^1(\Omega)$. Or par hypothèse, (4.3), chaque fonction ϕ_i est polynômiale par morceaux. De plus, grâce à l'hypothèse (4.5), on a raccord des polynômes sur les interfaces des éléments, ce qui assure la continuité de ϕ_i. Il reste à montrer que $\phi_i \in H^1(\Omega)$ pour tout $i = 1, \ldots, M$. Comme $\phi_i \in C(\bar{\Omega})$, il est évident que $\phi_i \in L^2(\Omega)$ (car Ω est un ouvert borné, donc $\phi_i \in L^\infty(\Omega) \subset L^2(\Omega)$.

Montrons maintenant que les dérivées faibles $D_j\phi_i, \quad j = 1, \ldots, d$, appartiennent à $L^2(\Omega)$. Par définition, la fonction ϕ_i admet une dérivée faible dans $L^2(\Omega)$ s'il existe une fonction $\psi_{i,j} \in L^2(\Omega)$ telle que :

$$\int_\Omega \phi_i(x)\partial_j\varphi(x)\,\mathrm{d}x = -\int_\Omega \psi_{ij}(x)\varphi(x)\,\mathrm{d}x, \tag{4.7}$$

pour toute fonction $\varphi \in C_c^1(\Omega)$ (on rappelle que $C_c^1(\Omega)$ désigne l'ensemble des fonctions de classe C^1 à support compact et que ∂_j désigne la dérivée classique par rapport à la j-ème variable). Or, comme $\bar{\Omega} = \bigcup_{\ell=1}^L K_\ell$, on a :

$$\int_\Omega \phi_i(x)D_j\varphi(x)\,\mathrm{d}x = \sum_{\ell=1}^L \int_{K_\ell} \phi_i(x)D_j\varphi(x)\,\mathrm{d}x. \tag{4.8}$$

Sur chaque élément K_ℓ, la fonction ϕ_i est polynômiale. On peut donc appliquer la formule de Green et on a :

$$\int_{K_\ell} \phi_i(x)\partial_j\varphi(x)\,\mathrm{d}x = \int_{\partial K_\ell} \phi_i(x)\varphi(x)n_j(x)d\gamma(x) - \int_{K_\ell} \partial_j\phi_i(x)\varphi(x)\,\mathrm{d}x,$$

où $n_j(x)$ est la j-ème composante du vecteur unitaire normal à ∂K_ℓ en x, extérieur à K_ℓ. Mais, si on note $\mathcal{E}_{int}$ l'ensemble des arêtes intérieures du maillage (i.e. celles qui ne sont pas sur le bord), on a :

$$\begin{aligned} X = \sum_{\ell=1}^L \int_{\partial K_\ell} \phi_i(x)\varphi(x)n_j(x)d\gamma(x) &= \int_{\partial\bar{\Omega}} \phi_i(x)\varphi(x)n_j(x)d\gamma(x) \\ &+ \sum_{\epsilon\in\mathcal{E}_{int}} \int \left[(\phi_i(x)\varphi(x)n_j(x))\,\big|_{K_{\ell_1}} + (\phi_i(x)\varphi(x)n_j(x))_{K_{\ell_2}}\right] d\gamma(x). \end{aligned} \tag{4.9}$$

où K_{ℓ_1} et K_{ℓ_2} désignent les deux éléments dont ϵ est l'interface.

Comme φ est à support compact,

$$\int_{\partial\Omega} \phi_i(x)\varphi(x)n_j(x)d\gamma(x) = 0$$

Comme ϕ_i et φ sont continues et comme $n_j(x)|_{K_{\ell_1}} = -n_j(x)|_{K_{\ell_2}}$ pour tout $x \in \epsilon$, on en déduit que $X = 0$. En reportant dans (4.8), on obtient donc que :

$$\int_\Omega \phi_i(x)\partial_j\varphi(x)\,\mathrm{d}x = -\sum_{\ell=1}^L \int_{K_\ell} \partial_j\phi_i(x)\varphi(x)\,\mathrm{d}x.$$

Soit $\psi_{i,j}$ la fonction de Ω dans $\mathbb{R}$ définie presque partout par

$$\psi_{ij}\,\Big|_{\overset{\circ}{K}_\ell} = -\partial_j\phi_i.$$

Comme $\partial_j\phi_i$ est une fonction polynômiale par morceaux, $\psi_{i,j} \in L^2(\Omega)$ qui vérifie (4.7), ce qui termine la démonstration. •

Cas des conditions de Dirichlet homogènes

Plaçons-nous mainteant dans le cas où $H = H_0^1(\Omega)$. On décompose alors l'ensemble $\mathcal{S}$ des nœuds du maillage $\mathcal{S} = \mathcal{S}_{int} \cup \mathcal{S}_{ext}$ où $\mathcal{S}_{int} = \{S_i, i = 1, \ldots, N\} \subset \Omega$ est l'ensemble des nœuds intérieurs à Ω et $\mathcal{S}_{ext} = \{S_i, i = N+1, \ldots, M\} \subset \partial\Omega$ est l'ensemble des nœuds de la frontière. Les fonctions de base globales sont alors les fonctions $\phi_i, i = 1, \ldots, N$ telles que

$$\phi_i\,|_{K_\ell} \in \mathcal{P}_\ell, \forall i = 1, \ldots, N, \quad \forall \ell = 1, \ldots, L \tag{4.10}$$

$$\phi_i(S_j) = \delta_{ij}, \forall j = 1, \ldots, N, \tag{4.11}$$

et on pose là encore $H_N = \mathrm{Vect}\{\phi_1, \ldots, \phi_N\}$. On a alors encore le résultat suivant :

Proposition 4.13 — Critère de conformité, cas H_0^1. Soit Ω un ouvert polygonal (ou polyèdrique) de $\mathbb{R}^d$, $d = 2$ ou 3. Soit $\mathcal{T} = (K_\ell)_{\ell=1,\dots,L}$ un maillage éléments finis de Ω, $\mathcal{S} = (S_i)_{i=1,\dots,M} = \mathcal{S}_{int} \cup \mathcal{S}_{ext}$ l'ensemble des nœuds du maillage. On se place sous les hypothèses de la proposition 4.7. On suppose que les fonctions de base globale $(\phi_i)_{i=1,\dots,M}$ vérifient (4.10) et (4.11) et que les conditions (4.5) et (4.6) sont vérifiées. Alors on a : $H_N \subset C(\bar{\Omega})$ et $H_N \subset H_0^1(\Omega)$

Démonstration. La preuve de cette proposition est laissée à titre d'exercice. •

Remarque 4.14 — Éléments finis conformes dans $H^2(\Omega)$. On a construit un espace d'approximation H_N inclus dans $C(\bar{\Omega})$. En général, on n'a pas $H_N \subset C^1(\bar{\Omega})$ et donc on n'a pas non plus $H_N \subset H^2(\Omega)$ (en dimension 1 d'espace, $H^2(\Omega) \subset C^1(\Omega)$). Même si on augmente le degré de l'espace des polynômes, on n'obtiendra pas l'inclusion $H_N \subset C^1(\bar{\Omega})$. Si on prend par exemple les polynômes de degré 2 sur les éléments, on n'a pas de condition pour assurer le raccord, des dérivées aux interfaces. Pour obtenir ce raccord, les éléments finis de Lagrange ne suffisent pas : il faut prendre des éléments de type Hermite, pour lesquels les degrés de liberté ne sont plus seulement les valeurs de la fonction aux nœuds, mais aussi les valeurs de ses dérivées aux nœuds. Les éléments finis de Hermite seront par exemple bien adaptés à l'approximation des problèmes elliptiques d'ordre 4, dont un exemple est l'équation :

$$\Delta^2 u = f \text{ dans } \Omega \tag{4.12}$$

où Ω est un ouvert borné de $\mathbb{R}^2$, $\Delta^2 u = \Delta(\Delta u)$, e avec des conditions aux limites adéquates, que nous ne détaillerons pas ici. On peut, en fonction de ces conditions aux limites, trouver un espace de Hilbert H et une formulation faible de (4.12), qui s'écrit :

$$\begin{cases} \displaystyle\int_\Omega \Delta u(x)\Delta\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x \\ u \in H, \quad \forall \varphi \in H. \end{cases}$$

Pour que cette formulation ait un sens, il faut que $\Delta u \in L^2(\Omega)$ et $\Delta\varphi \in L^2(\Omega)$ et donc que $H \subset H^2(\Omega)$. Pour construire une approximation par éléments finis conforme de ce problème, il faut donc choisir $H_N \subset H^2(\Omega)$ et le choix des éléments finis de Hermite semble donc indiqué.

4.2 Exemples

Pour chaque méthode d'élément fini de Lagrange, on définit :

1. un élément de référence $\bar{K}$
2. des fonctions de base locales sur $\bar{K}$
3. une bijection F_ℓ de K sur K_ℓ, pour $\ell = 1, \dots, L$, où L est le nombre déléments du maillage.

4.2.1 Élément fini de Lagrange P1 sur triangle ($d = 2$)

Le maillage du domaine est constitué de L triangles $(K_\ell)_{\ell=1,\dots,L}$ et les polynômes d'approximation sont de degré 1. paragraphÉlément fini de référence On choisit le triangle $\bar{K}$ de sommets $(0,0)$, $(1,0)$ et $(0,1)$ et $\overline{\mathcal{P}} = \{\psi : K \to \mathbb{R}(x,y) \mapsto ax + by + c, (a,b,c) \in \mathbb{R}^3\}$.

Proposition 4.15 — Unisolvance de l'élément fini P1. Soit $\bar{\Sigma} = (\bar{a}_i)_{i=1,2,3}$ avec $\bar{a}_1 = (0,0), \bar{a}_2 = (1,0)$ et $\bar{a}_3 = (0,1)$, et

$$\overline{\mathcal{P}} = \{\bar{\psi}; \bar{K} \to \mathbb{R}; (\bar{x},\bar{y}) \mapsto a + b\bar{x} + c\bar{y}, (a,b,c) \in \mathbb{R}^3\},$$

alors l'ensemble $\bar{\Sigma}$ est $\overline{\mathcal{P}}$)-unisolvant.

Démonstration. Soit $(\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3$ et $\psi \in \overline{\mathcal{P}}$. On suppose que $\psi(\bar{a}_i) = \alpha_i$, $i = 1, 2, 3$. La fonction ψ est de la forme $\psi(x,y) = a + bx + cy$ et on a donc :

$$\begin{cases} a = \alpha_1 \\ a + b = \alpha_2 \\ a + c = \alpha_3 \end{cases}$$

d'où $c = \alpha_1$, $b_1 = \alpha_2 - \alpha_1$ et $b_2 = \alpha_3 - \alpha_2$. La connaissance de ψ aux nœuds $(\bar{a}_i)_{i=1,2,3}$ détermine donc entièrement la fonction ψ. •

Fonctions de bases locales Les fonctions de base locales sur l'élément fini de référence $\bar{K}$ sont définies par $\bar{\phi}_i \in \overline{\mathcal{P}} \bar{\phi}_i(\bar{a}_j) = \delta_{ij}$, ce qui détermine les fonctions $\bar{\phi}_i$ de manière unique, comme on vient de le voir. On a donc

$$\begin{cases} \bar{\phi}_1(\bar{x}, \bar{y}) = 1 - \bar{x} - \bar{y} \\ \bar{\phi}_2(\bar{x}, \bar{y}) = \bar{x} \\ \bar{\phi}_3(\bar{x}, \bar{y}) = \bar{y} \end{cases}$$

Transformation F_ℓ On construit ici une bijection affine qui transforme $\bar{K}$ le triangle de référence en un autre triangle K du maillage. On cherche donc $\ell : \bar{K} \to K$, telle que

$$F_\ell(\bar{a}_i) = a_i \qquad i = 1, \ldots, 3$$

où $\Sigma = (a_i)_{i=1,2,3}$ est l'ensemble des sommets de K. Notons (x_i, y_i) les coordonnées de $a_i, i = 1, 2, 3$. Comme F_ℓ est une fonction affine de $\mathbb{R}^2$ dans $\mathbb{R}^2$, elle s'écrit sous la forme.

$$F_\ell(\bar{x}, \bar{y}) = (\beta_1 + \gamma_1 \bar{x} + \delta_1 \bar{y}, \beta_2 + \gamma_2 \bar{x} + \delta_2 \bar{y})$$

et on cherche $\beta_i, \gamma_i, \delta_i, \quad i = 1, 2$ tels que :

$$\begin{cases} F_\ell((0,0)) = (x_1, y_1) \\ F_\ell((1,0)) = (x_2, y_2) \\ F_\ell((0,1)) = (x_3, y_3). \end{cases}$$

Une résolution de système élémentaire amène alors à :

$$F_\ell(\bar{x}, \bar{y}) = \begin{pmatrix} x_1 + (x_2 - x_1)\bar{x} + (x_3 - x_1)\bar{y} \\ y_1 + (y_2 - y_1)\bar{x} + (y_3 - y_1)\bar{y} \end{pmatrix}$$

D'après la remarque 4.9, si on note $\bar{\phi}_k, k = 1, 2, 3$ les fonctions de base locales de l'élément de référence $(\bar{K}, \bar{\Sigma}, \overline{\mathcal{P}})$ et $\phi_k^{(\ell)}, k = 1, 2, 3$ les fonctions de base locales de l'élément $(K_\ell, \Sigma_\ell, \mathcal{P}_\ell)$, on a $\phi_k^{(\ell)} = \bar{\phi}_k \circ F_\ell^{-1}$

Si on note maintenant $(\phi_i)_{i=1,\ldots,N}$ les fonctions de base globales, on a :

$$\phi_i \Big|_{K_\ell} = \phi_k^{(\ell)},$$

où $i = \mathtt{ng}(\ell, k)$ est l'indice du k-ème nœud de l'élément ℓ dans la numérotation globale, et le tableau $\mathtt{ng}(L, N_\ell)$ ainsi introduit permet de relier la numérotation locale d'un nœud dans un élément à sa numérotation globale ; sa dimension est $L \times N_\ell$, où L est le nombre d'éléments du maillage et N_ℓ est le nombre de nœuds par élément. Notons que l'élément fini de Lagrange ainsi défini vérifie les critères de cohérence 4.5 et (4.6). Pour compléter la définition de l'espace d'approximation H_N, il ne reste qu'à déterminer les "nœuds liés", de la façon dont on a traité le cas de l'espace $H_0^1(\Omega)$.

Il faut également insister sur le fait que cet élément est très souvent utilisé, en raison de sa facilité d'implantation et de la structure creuse des systèmes linéaires qu'il génère. Il est particulièrement bien adapté lorsqu'on cherche des solutions dans l'espace $H^1(\Omega)$. Il se généralise facilement en trois dimensions d'espace, où on utilise alors des tétraèdres, avec toujours comme espace de polynôme l'espace des fonctions affines.

4.2.2 Élément fini triangulaire P2

Comme le titre du paragraphe l'indique, on considère un maillage triangulaire, et un espace de polynômes de degré 2 pour construire l'espace d'approximation.

Élément fini de référence On choisit comme élément fini de référence le triangle de sommets $(0,0)$, $(1,0)$ et $(0,1)$, voir Figure 4.6 et on prend pour Σ :

$$\bar{\Sigma} = \{(0,0), (1,0), (0,1), (1/2,1/2), (0,1/2), (1/2,0)\}$$

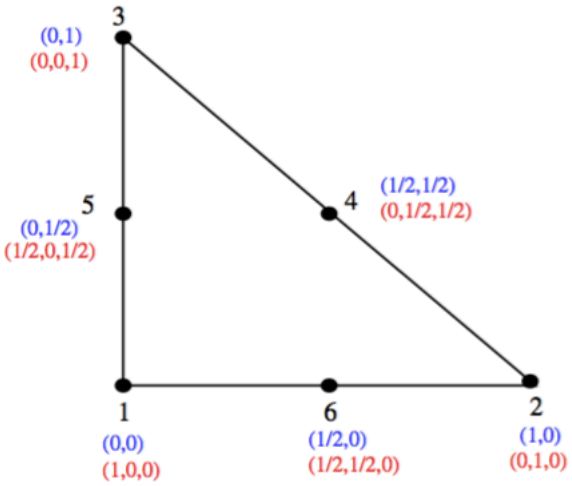


Figure 4.6 – Élément de référence pour les éléments finis P2 avec coordonnées cartésiennes (en bleu) et barycentriques (en rouge) des nœuds

Fonctions de base locales Les fonctions de base locales sont définies à partir des coordonnées barycentriques. On rappelle que les coordonnées barycentriques d'un point $\boldsymbol{x}$ du triangle K de sommets $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ et $\boldsymbol{a}_3$ sont les réels λ_1, λ_2, λ_3 tels que :

$$\boldsymbol{x} = \lambda_1 \boldsymbol{a}_1 + \lambda_2 \boldsymbol{a}_2 + \lambda_3 \boldsymbol{a}_3.$$

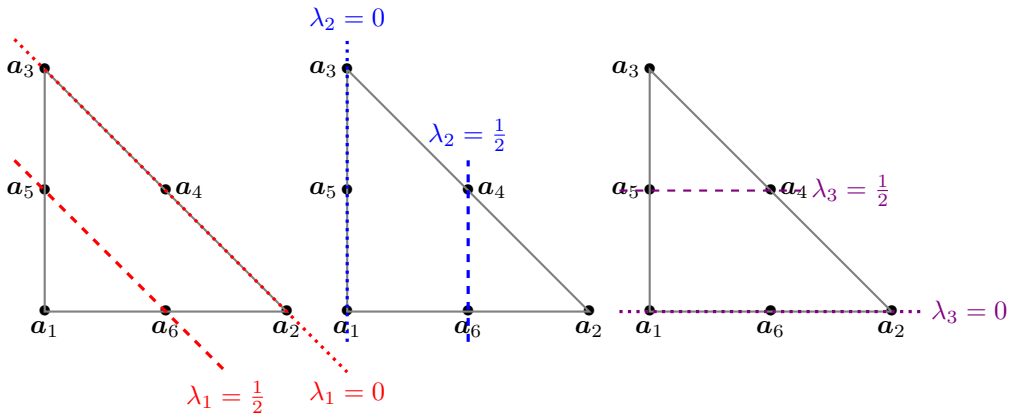


Figure 4.7 – Les droites $\lambda_i = \frac{1}{2}$ (en tirets) et $\lambda_i = 0$ (en pointillés).

Dans le cas du triangle de référence $\bar{K}$ de sommets (0,0), (1,0) et (0,1), les coordonnées barycentriques d'un point $\boldsymbol{x}$ de coordonnées cartésiennes x et y sont donc $\lambda_1 = 1 - x - y$, $\lambda_2 = x$, $\lambda_3 = y$. Par définition, on a $\sum_{i=1}^{3} \lambda_i = 1$ et $\lambda_i \geqslant 0$ (car le triangle K est l'enveloppe convexe de l'ensemble de ses sommets). On peut alors déterminer les fonctions de base en fonction des coordonnées barycentriques des six nœuds de $\bar{K}$ exprimés par leurs coordonnées barycentriques $\boldsymbol{a}_1 = (1,0,0)$, $\boldsymbol{a}_2 = (0,1,0)$, $\boldsymbol{a}_3 = (0,0,1)$, $\boldsymbol{a}_4 = (0, \frac{1}{2}, \frac{1}{2})$, $\boldsymbol{a}_5 = (\frac{1}{2}, 0, \frac{1}{2})$ et $\boldsymbol{a}_6 = (\frac{1}{2}, \frac{1}{2}, 0)$. Les fonctions de base ϕ_i, $i = 1, \ldots, 6$ appartiennent à l'espace des polynômes du second degré $\mathcal{P}_2$ et sont telles que

$$\phi_i(\boldsymbol{a}_j) = \delta_{ij}, \ \forall i, j = 1, \ldots, 6.$$

Commençons par ϕ_1 ; on veut que $\phi_1(\boldsymbol{a}_1) = 1$ et $\phi_i(\boldsymbol{a}_i) = 0$, $\forall i = 2, \ldots, 6$. La fonction ϕ_1 définie par $\phi_1(x, y) = 2\lambda_1(\lambda_1 - 1/2)$ convient (voir figure 4.7). Par symétrie, on définit $\phi_2(x, y) = 2\lambda_2(\lambda_2 - 1/2)$ et $\phi_3(x, y) = 2\lambda_3(\lambda_3 - 1/2)$. Les fonctions de base associées aux nœuds $\boldsymbol{a}_4, \boldsymbol{a}_5, \boldsymbol{a}_6$ sont alors $\phi_4(x, y) = 4\lambda_2\lambda_3$, $\phi_5(x, y) = 4\lambda_1\lambda_3$ et $\phi_6(x, y) = 4\lambda_1\lambda_2$. Il est facile de voir que ces fonctions forment une famille libre d'éléments de $\mathcal{P}_2$ et comme card $\bar{\Sigma} =$ card $\overline{\mathcal{P}}_2$, l'ensemble $\bar{\Sigma}$ est $\overline{\mathcal{P}}_2$ est bien unisolvant.

Transformation F_ℓ La bijection F_ℓ qui permet de passer de l'élément fini de référence $\bar{K}$ à l'élément K_ℓ a déjà été vue dans le cas de l'élément fini P1 c'est la fonction affine définie par :

$$F_\ell(x, y) = \begin{pmatrix} x_1 + (x_2 - x_1)x + (x_3 - x_1)y \\ y_1 + (y_2 - y_1)x + (y_3 - y_1)y \end{pmatrix}$$

où (x_i, y_i), $i = 1, 2, 3$ sont les coordonnées respectives des trois sommets du triangle K_ℓ. Comme cette transformation est affine, les coordonnées barycentriques restent inchangées par cette transformation.

On peut généraliser les éléments finis P1 et P2 aux éléments finis Pk sur triangles, pour $k \geqslant 1$. On prend toujours le même élément de référence, dont on divise chaque côté en k intervalles. Les extrémités de ces intervalles sont les nœuds du mailage. On a donc $3k$ nœuds, qu'on peut repérer par leurs coordonnées barycentriques, qui prennent les valeurs $0, \frac{1}{k}, \frac{2}{k}, \ldots, 1$. On peut montrer que si $u \in H^{k+1}$, alors

$$\|u_N - u\|_{H^1(\Omega)} \leqslant Ch^k \|u\|_{H^{k+1}(\Omega)} \tag{4.13}$$

4.2.3 Éléments finis sur quadrangles

Cas rectangulaire

On prend comme élément fini de référence le carré $\bar{K} = [-1\,;1] \times [-1\,;1]$ et comme nœuds les coins de ce carré $a_1 = (1, -1)$, $a_2 = (1, 1)$, $a_3 = (-1, 1)$ et $a_4 = (-1, -1)$. On prend comme espace de polynômes $\mathcal{P} = \{f : \bar{K} \to \mathbb{R}; f \in \mathcal{Q}_1\}$ où $\mathcal{Q}_1 = \{f : \mathbb{R}^2 \to \mathbb{R}; f(x, y) = a + bx + cy + dxy, (a, b, c, d) \in \mathbb{R}^4\}$. L'ensemble Σ est $\mathcal{P}$-unisolvant. Les fonctions de base locales sont les fonctions :

$$\phi_1(x, y) = -\frac{1}{4}(x + 1)(y - 1) \quad \phi_2(x, y) = \frac{1}{4}(x + 1)(y + 1)$$
$$\phi_3(x, y) = -\frac{1}{4}(x - 1)(y + 1) \quad \phi_4(x, y) = \frac{1}{4}(x - 1)(y - 1)$$

La transformation F_ℓ permet de passer de l'élément de référence carré $\bar{K}$ à un rectangle quelconque du maillage K_ℓ. Si on considère un rectangle K_ℓ parallèle aux axes, dont les nœuds sont notés (x_1, y_1), (x_2, y_1), (x_2, y_2), (x_1, y_2), les nœuds du rectangle K_ℓ, la bijection F_ℓ s'écrit :

$$F_\ell(x, y) = \frac{1}{2}\begin{pmatrix} (x_2 - x_1)x + x_2 + x_1 \\ (y_2 - y_1)y + y_2 + y_1 \end{pmatrix}$$

Considérons maintenant le cas d'un maillage quadrangulaire quelconque. Dans ce cas, on choisit toujours comme élément de référence le carré unité. La transformation F_ℓ qui transforme l'élément de référence en un quadrangle K_ℓ est toujours affine, mais par contre, les composantes de $F_\ell((x, y))$ dépendent maintenant de x et de y [exercice 60]. En conséquence, le fait que $f \in \mathcal{Q}_1$ n'entraîne plus que $f \circ F_\ell \in \mathcal{Q}_1$. Les fonctions de base seront donc des polynômes $\mathcal{Q}_1$ sur l'élément de référence $\bar{K}$, mais pas sur les éléments *courants* K_ℓ.

Éléments finis d'ordre supérieur Comme pour un maillage triangulaire, on peut choisir un espace de polynômes d'ordre supérieur, Q_k, pour les fonctions de base de l'élément de référence $\bar{K} = [-1\,;1] \times [-1\,;1]$. On choisit alors comme ensemble de nœuds : $\bar{\Sigma} = \bar{\Sigma}_k = \{(x, y) \in \bar{K}, (x, y) \in \{-1, -1 + \frac{1}{k}, -1 + \frac{1}{k}, \ldots, 1\}^2$. On peut montrer facilement que l'ensemble $\bar{\Sigma}_k$ est Q_k-unisolvant. Là encore, si la solution exacte de problème continu est suffisamment régulière, on peut démontrer la même estimation d'erreur (4.13) que dans le cas des triangles [4].

Exprimons par exemple l'espace des polynômes $\mathcal{Q}_2$. On a :

$$\mathcal{Q}_2 = \{f : \mathbb{R} \to \mathbb{R}; f(x) = a_1 + a_2 x + a_3 y + a_4 xy + a_5 x^2 + a_6 y^2 + a_7 xy^2 + a_8 x^2 y + a_9 x^2 y^2, a_i \in \mathbb{R}, i = 1, \ldots, 9\} \quad (4.14)$$

L'espace $\mathcal{Q}_2$ comporte donc neuf degrés de liberté. On a donc besoin de neuf nœuds dans $\bar{\Sigma}$ pour que l'ensemble $\bar{\Sigma}$ soit $\overline{\mathcal{Q}_2}$-unisolvant [exercice 66]. On peut alors utiliser comme nœuds sur le carré de référence $[-1\,;1] \times [-1\,;1]$:

$$\bar{\Sigma} = \{(-1,-1), (-1,0), (-1,-1), (0,-1), (0,0), (0,1), (1,-1), (1,0), (1,1)\}$$

En général, on préfère pourtant supprimer le nœud central (0,0) et choisir :

$$\Sigma^* = \bar{\Sigma} \setminus \{(0,0)\}$$

Il faut donc un degré de liberté en moins pour l'espace des polynômes. On définit alors :

$$\mathcal{Q}_2^* = \{f : \mathbb{R} \to \mathbb{R}; f(x) = a_1 + a_2 x + a_3 y + a_4 xy + a_5 x^2 + a_6 y^2 + a_7 xy^2 + a_8 x^2 y\} \text{ avec } a_i \in \mathbb{R},\ i = 1, \ldots, 8 \quad (4.15)$$

L'ensemble Σ^* est $\mathcal{Q}_2^*$-unisolvant [exercice 67] et on peut montrer que l'élément fini $Q2^*$ ainsi défini est aussi précis (et plus facile à mettre en œuvre que l'élément Q2).

4.3 Analyse d'erreur

4.3.1 Erreur d'interpolation et erreur de discrétisation

Soit $\Omega =]0\,;1[$, on considère un maillage classique, défini par les $N+2$ points $(x_i)_{i=0\ldots N+1}$, avec $x_0 = 0$ et $x_{N+1} = 1$ et on note $h_i = x_{i+1} - x_i$, $i = 0, \ldots, N+1$ et $h = \max\{|h_i|,\ i = 0, \ldots, N+1\}$. On va montrer que si $u \in H^2(]0\,;1[)$, alors on peut obtenir une majoration de l'erreur d'interpolation $\|u - u_I\|_{H^1}$, où u_I est l'interpolée de u, définie par

$$u_I = \sum_{i=1}^{N} u(x_i)\phi_i \quad (4.16)$$

où ϕ_i est la fonction de base associée au noeud x_i. On rappelle [exercice 37] que si $u \in H^1(]0\,;1[)$ alors u est continue. En particulier, on a donc $H^2(]0\,;1[) \subset C^1([0\,;1])$. Remarquons que ce résultat est lié à la dimension 1 [1]. On va démontrer le résultat suivant sur l'erreur d'interpolation.

Théorème 4.1 — Majoration de l'erreur d'interpolation, dimension 1. Soit $u \in H^2(]0\,;1[)$ et soit u_I son interpolée sur $H_N = \text{Vect}\{\phi_i,\ i = 1, \ldots, N\}$, où ϕ_i désigne la i-ème fonction de base élément fini P1 associée au nœud x_i d'un maillage éléments finis de $]0\,;1[$; soit u_I l'interpolée de u, définie par (4.16). Alors il existe $C_I \in \mathbb{R}$ ne dépendant que de u, tel que

$$\|u - u_I\|_{H^1} \leqslant C_I h.$$

Démonstration. On veut estimer

$$\|u - u_I\|_{H^1}^2 = |u - u_I|_0^2 + |u - u_I|_1^2$$

où $|v|_0 = \|v\|_{L^2}$ et $|v|_1 = \|Dv\|_{L^2}$. Calculons $|u - u_I|_1^2$:

$$|u - u_I|_1^2 = \int_0^1 |u' - u_I'|^2 \,\mathrm{d}x = \sum_{i=0}^{N} \int_{x_i}^{x_{i+1}} |u'(x) - u_I'(x)|^2 dx.$$

Or pour $x \in]x_i, x_{i+1}[$ il existe $\xi_i \in]x_i, x_{i+1}[$ tel que

$$u_I'(x) = \frac{u(x_{i+1}) - u(x_i)}{h_i} = u'(\xi_i),$$

On a donc :

$$\int_{x_i}^{x_{i+1}} |u'(x) - u_I'(x)|^2 \,\mathrm{d}x = \int_{x_i}^{x_{i+1}} |u'(x) - u'(\xi_i)|^2 \,\mathrm{d}x.$$

On en déduit que :

$$\int_{x_i}^{x_{i+1}} |u'(x) - u'_I(x)|^2 \,\mathrm{d}x = \int_{x_i}^{x_{i+1}} \left| \int_{\xi_i}^{x} u''(t)\,\mathrm{d}t \right|^2 \mathrm{d}x \leqslant \int_{x_i}^{x_{i+1}} \left[\int_{\xi_i}^{x} |u''(t)|^2 \,\mathrm{d}t \right] |x - \xi_i| \,\mathrm{d}x$$

par l'inégalité de Cauchy-Schwarz. Or $|x - \xi_i| \leqslant h_i$ et $\displaystyle\int_{\xi_i}^{x} |u''(t)|^2 \,\mathrm{d}t \leqslant \int_{x_i}^{x_{i+1}} |u''(t)|^2 \,\mathrm{d}t$, et donc

$$\int_{x_i}^{x_{i+1}} |u'(x) - u'_I(x)|^2 \,\mathrm{d}x \leqslant h_i^2 \int_{x_i}^{x_{i+1}} |u''(t)|^2 \,\mathrm{d}t.$$

En sommant ces inégalités pour $i = 0$ à N, on obtient alors

$$|u - u_I|_1^2 \leqslant h^2 \int_0^1 |u''(t)|^2 \,\mathrm{d}t. \tag{4.17}$$

Il reste maintenant à majorer $|u - u_I|_0^2 = \displaystyle\int_0^1 |u - u_I|^2 \,\mathrm{d}x$. Pour $x \in [x_i, x_{i+1}]$, on a

$$|u(x) - u_I(x)|^2 = \left[\int_{x_i}^{x} (u'(t) - u'_I(t)) \,\mathrm{d}t \right]^2.$$

Par l'inégalité de Cauchy-Schwarz, on a donc :

$$|u(x) - u_I(x)|^2 \leqslant |x - x_i| \int_{x_i}^{x} (u'(t) - u'_I(t))^2 \,\mathrm{d}t$$

et $|x - x_i| \leqslant h_i$ pour tout $x \in [x_i, x_{i+1}]$. Par des calculs similaires aux précédents, on obtient donc :

$$|u(x) - u_I(x)|^2 \leqslant \int_{x_i}^{x} h_i \left[\int_{x_i}^{x_{i+1}} |u''(t)|^2 \,\mathrm{d}t \right] \mathrm{d}x h_i \leqslant h_i^3 \int_{x_i}^{x_{i+1}} |u''(t)|^2 \,\mathrm{d}t.$$

En intégrant sur $[x_i, x_{i+1}]$, il vient :

$$\int_{x_i}^{x_{i+1}} |u(x) - u_I(x)|^2 \,\mathrm{d}x \leqslant h_i^4 \int_{x_i}^{x_{i+1}} (u''(t))^2 \,\mathrm{d}t,$$

et en sommant sur $i = 1, \ldots, N$:

$$\int_0^1 (u(x) - u_I(x))^2 \,\mathrm{d}x \leqslant h^4 \int_0^1 (u''(t))^2 \,\mathrm{d}t.$$

On a donc $|u - u_I|_0 \leqslant h^2 |u|_2$, où $|u|_2 = \|u''\|_{L^2}$, ce qui entraîne, avec (4.17) :

$$\|u - u_I\|_{H_1}^2 \leqslant h^4 |u|_2^2 + h^2 |u|_2^2 \leqslant (1 + h^2) h^2 |u|_2^2$$

On en déduit le résultat annoncé avec $C_I = \sqrt{2}|u|_2$. •

On en déduit, grâce au lemme de Céa (lemme 3.18) le résultat d'estimation d'erreur suivant :

Corollaire 4.16 — Estimation d'erreur des éléments P1 **en dimension 1.** Soit $\Omega =]0, 1[$; soit $f \in L^2(\Omega)$ et $u \in H_0^1(\Omega)$ l'unique solution du problème

$$\int_\Omega u'(x)\ v'(x)\,\mathrm{d}x = \int f(x) v(x)\ \,\mathrm{d}x,\ \forall v \in H_0^1(\Omega)$$

et $u_\mathcal{T}$, l'approximation éléments finis P1 obtenue sur un maillage admissible $\mathcal{T}$ de pas $h_\mathcal{T} = \max_{i=1,\ldots,N}\{|h_i|\}$. Alors il existe $C \in \mathbb{R}$ ne dépendant que de Ω et f tel que $\|u - u_\mathcal{T}\| < Ch$.

Démonstration. Par le théorème de régularité 3.4, on a $u \in H^2(\Omega)$ et donc on peut appliquer le théorème d'interpolation 4.1. On conclut par une inégalité triangulaire et le lemme de Céa. •

Ce résultat se généralise au cas de plusieurs dimensions d'espace [4] sous des conditions géométriques sur le maillage, nécessaires pour obtenir le résultat d'interpolation. Comme exemple, on donne ci-dessous sans démonstration un résultat d'interpolation pour un maillage triangulaire en deux dimensions d'espace, pour lequel intervient une condition d'angle.

Théorème 4.2 — Majoration de l'erreur d'interpolation, dimension 2, triangles. Soit Ω un ouvert polygonal convexe de $\mathbb{R}^2$. Soit $\mathcal{T}$ un maillage triangulaire de Ω. On note $\alpha_\mathcal{T}$ l'angle minimum de tous les triangles du maillage et on suppose que $\alpha_\mathcal{T} > 0$. Soit $u \in H^2(\Omega)$ et soit u_I

son interpolée sur l'espace vectoriel H_N engendré par les fonctions de base P1 associées aux nœuds du maillage $\mathcal{T}$. Il existe $C_I \in \mathbb{R}$ ne dépendant que de u, tel que

$$\|u - u_I\|_{H^1} \leqslant \frac{C_I}{\sin \alpha_{\mathcal{T}}} h.$$

On en déduit, toujours grâce au lemme de Céa, l'estimation d'erreur suivante.

Corollaire 4.17 — Estimation d'erreur des éléments P1 en dimension 2. Sous les hypothèses du théorème 4.17 ; soit $f \in L^2(\Omega)$ et $u \in H_0^1(\Omega)$ l'unique solution du problème

$$\int_\Omega \nabla u(x) \cdot \nabla v(x) \, \mathrm{d}x = \int f(x) v(x) \, \mathrm{d}x, \ \forall v \in H_0^1(\Omega)$$

et soit $u_{\mathcal{T}}$ l'approximation éléments finis P1 obtenue sur une famille de maillages triangulaires $\mathcal{T}$ de pas $h_{\mathcal{T}} = \max_{i=1,\dots,N}\{|h_i|\}$. On suppose qu'il existe $0 < \alpha < \pi$ tel que pour tout maillage $\mathcal{T}$ de la famille considérée, $\alpha_{\mathcal{T}} \geqslant \alpha$. Alors il existe $C \in \mathbb{R}$ ne dépendant que de Ω et f et α tel que $\|u - u_{\mathcal{T}}\| < Ch$.

Remarque 4.18 — Sur les techniques d'estimation d'erreur. Lorsqu'on a voulu montrer des estimations d'erreur pour la méthode des différences finies, on a utilisé le principe de positivité, la consistance et la stabilité en norme L^∞. En volumes finis et éléments finis, on n'utilise pas le principe de positivité. En volumes finis, la stabilité en norme L^2 est obtenue grâce à l'inégalité de Poincaré discrète et la consistance est en fait la consistance des flux. Notons qu'en volumes finis on se sert aussi de la conservativité des flux numériques pour la preuve de convergence. Enfin, en éléments finis, la stabilité est obtenue grâce à la coercivité de la forme bilinéaire et la consistance provient du contrôle de l'erreur d'interpolation.

Remarque 4.19 — Sur la positivité. Même si le principe de positivité n'est pas explicitement utilisé pour les preuves de convergence des éléments finis et volumes finis, il est toutefois intéressant de voir à quelles conditions ce principe est respecté, car il est parfois très important en pratique.

Reprenons d'abord le cas du schéma volumes finis sur un maillage $\mathcal{T}$ admissible pour la discrétisation de l'équation (3.1).

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

Rappelons que le schéma volumes finis s'écrit :

$$\sum_{K \in \mathcal{C}} \left(\sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_{int}} \tau_{K,L}(u_K - u_L) + \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_{ext}} \tau_{K,\sigma} u_K \right) = |K| f_K, \tag{4.18}$$

avec

$$\tau_{K,L} = \frac{|K|L|}{\mathrm{d}(x_K, x_L)} \text{ et } \tau_{K,\sigma} = \frac{|\sigma|}{\mathrm{d}(x_K, \partial\Omega)},$$

où $|K|$, (resp. $|\sigma|$) désigne la mesure de Lebesque en dimension d (resp. $d-1$) de K (resp. σ) et $\mathrm{d}(\,\cdot\,,\,\cdot\,)$ la distance euclidienne.

Notons que les coefficients $\tau_{K,L}$ et $\tau_{K,\sigma}$ sont positifs, grâce au fait que le maillage est admissible, et donc $\overrightarrow{x_K x_L} = d(x_K, x_L)\boldsymbol{n}_{KL}$, où $\overrightarrow{x_K x_L}$ désigne le vecteur d'extrémités x_K et x_L et $\boldsymbol{n}_{KL}$ la normale unitaire à $K|L$ sortante de K. En conséquence, le schéma (4.18) s'écrit comme une somme de termes d'échange entre les mailles K et L, avec des coefficients τ_{KL} positifs. C'est grâce à cette propriété que l'on montre facilement que le principe de positivité est vérifié, voir [6].

Considérons maintenant la méthode des éléments finis P1, pour la résolution du problème (3.1) sur maillage triangulaire. Si le maillage satisfait la condition faible de Delaunay, qui stipule que la somme de deux angles opposés à une même arête doit être inférieure à π, alors le principe du maximum est vérifiée (voir On sait [4]). Ce résultat peut se retrouver en écrivant le schéma éléments finis sous la forme d'un schéma volumes finis, voir [6].

4.3.2 Super convergence

On considère ici un ouvert Ω polygonal convexe de $\mathbb{R}^d$, $d \geqslant 1$ et on suppose que $f \in L^2(\Omega)$. On s'intéresse à l'approximation par éléments finis P1 de la solution $u \in H_0^1(\Omega)$ du problème (3.6). On suppose qu'on a démontré une estimation d'erreur ordre h en norme L^2 entre la solution exacte u et la solution approchée par éléments finis P1 comme par exemple sous les hypothèses des corollaires 4.16 ou 4.17. Si la solution u de (3.1) est dans H^2, il se produit un petit miracle car on peut montrer grâce à une technique astucieuse, dite méthode d'Aubin-Nitsche, que l'erreur de discrétisation en norme L^2 est en fait d'ordre 2.

Théorème 4.3 — Super convergence, méthode d'Aubin-Nitsche . Soit Ω un ouvert polygonal convexe de $\mathbb{R}^d$, $d \geqslant 1$; soit $f \in L^2(\Omega)$, et $u \in H_0^1(\Omega)$ solution de

$$\int_\Omega \nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x, \forall v \in H_0^1(\Omega). \tag{4.19}$$

On définit $u_\mathcal{T} \in H_{N_\mathcal{T}}$ la solution approchée obtenue par éléments finis P1, sur un maillage éléments finis $\mathcal{T}$ et soit $h_\mathcal{T} = \max_{K\in\mathcal{T}} \operatorname{diam} K$; l'espace H_N est donc l'espace vectoriel engendré par les fonctions de base associées aux nœuds du maillage ; on suppose qu'il existe $C > 0$ dépendant de f et éventuellement de la régularité du maillage tel que

$$\|u - u_\mathcal{T}\|_{H^1(\Omega)} \leqslant C\|f\|_{L^2(\Omega)} h_\mathcal{T}. \tag{4.20}$$

Alors il existe $C_s \in \mathbb{R}$ ne dépendant que de Ω et f tel que :

$$\|u - u_\mathcal{T}\|_{L^2(\Omega)} \leqslant C_s h_\mathcal{T}^2. \tag{4.21}$$

Démonstration. Soit maintenant $e_\mathcal{T} = u - u_\mathcal{T}$ et $w \in H_0^1(\Omega)$ vérifiant

$$\int_\Omega \nabla w(x) \cdot \nabla \psi(x)\,\mathrm{d}x = \int_\Omega e_\mathcal{T}(x)\psi(x)\,\mathrm{d}x, \forall \psi \in H_0^1(\Omega). \tag{4.22}$$

Notons que w est la solution faible du problème

$$\begin{cases} -\Delta w = e_\mathcal{T} \text{ dans } \Omega \\ w = 0 \text{ sur } \partial\Omega. \end{cases}$$

Comme $e_\mathcal{T} \in L^2(\Omega)$, par le théorème 3.4, il existe $C_r \in \mathbb{R}_+$ ne dépendant que Ω tel que

$$\|w\|_{H^2(\Omega)} \leqslant C_r \|e_\mathcal{T}\|_{L^2(\Omega)}.$$

Or, en choisissant $\psi = e_\mathcal{T}$ dans (4.22), on obtient

$$\|e_\mathcal{T}\|^2_{L^2(\Omega)} = \int_\Omega e_\mathcal{T}(x)e_\mathcal{T}(x)\,\mathrm{d}x = \int_\Omega \nabla w(x) \cdot \nabla e_\mathcal{T}(x)\,\mathrm{d}x. \tag{4.23}$$

Soit $w_\mathcal{T}$ la solution approchée par éléments finis P1 du problème (4.22), c.à.d solution de :

$$\begin{cases} w_\mathcal{T} \in H_N, \\ \displaystyle\int_\Omega \nabla w_\mathcal{T}(x) \cdot \nabla v(x)\,\mathrm{d}x = \int_\Omega e_\mathcal{T}(x)v(x)\,\mathrm{d}x, \forall v \in H_N. \end{cases} \tag{4.24}$$

Comme u est solution de (4.19) et $u_\mathcal{T}$ de

$$\int_\Omega \nabla u_\mathcal{T}(x) \cdot \nabla v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x, \forall v \in H_N \subset H_0^1(\Omega).$$

on a (voir remarque 3.19) que $u - u_\mathcal{T}$ vérifie :

$$\int_\Omega \nabla(u - u_\mathcal{T})(x) \cdot \nabla w_\mathcal{T}(x)\,\mathrm{d}x = 0;$$

on déduit de cette égalité et de (4.23) que :

$$\|e_\mathcal{T}\|^2_{L^2(\Omega} = \int_\Omega \nabla(w - w_\mathcal{T})(x) \cdot \nabla(u - u_\mathcal{T})(x)\,\mathrm{d}x \leqslant \|w - w_\mathcal{T}\|_{H^1(\Omega)} \|u - u_\mathcal{T}\|_{H^1(\Omega)}.$$

En utilisant l'hypothèse (4.20) pour $u - u_\mathcal{T}$ avec le second membre f et pour $w - w_\mathcal{T}$ avec le second membre $e_\mathcal{T}$, on a donc :

$$\|u - u_\mathcal{T}\|_{H^1(\Omega)} \leqslant C\|f\|_{L^2(\Omega)} h_\mathcal{T} \text{ et } \|w - w_\mathcal{T}\|_{H^1(\Omega)} \leqslant C\|e_\mathcal{T}\|_{L^2(\Omega)} h_\mathcal{T}.$$

On en déduit que :

$$\|e_\mathcal{T}\|_{L^2(\Omega)} \leqslant C^2\|f\|_{L^2(\Omega)} h_\mathcal{T}^2,$$

ce qui démontre le théorème. •

Ce théorème donne donc la super-convergence en 1D sous les hypothèses du théorème 4.1 et en 2D triangles sous les hypothèses du théorème 4.2.

Les estimations d'erreur obtenues au paragraphe précédent reposent sur la régularité H^2 de u. Que se passe-t-il si cette régularité n'est plus vérifiée ? Par exemple, si le domaine Ω possède un coin rentrant, on sait que dans ce cas, la solution u du problème (3.6) n'est plus dans $H^2(\Omega)$, mais dans un espace $H^{1+s}(\Omega)$, où s dépend de l'angle du coin rentrant. Considérons donc pour fixer les idées le problème

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \text{ où } \Omega \text{ est un ouvert} \\ u = 0 & \text{sur } \partial\Omega \text{ polygônal avec un coin rentrant.} \end{cases}$$

Pour approcher correctement la singularité, on peut raffiner le maillage dans le voisinage du coin. On peut également, lorsque cela est possible, modifier l'espace d'approximation pour tenir compte de la singularité. Dans le cas d'un polygône avec un coin rentrant par exemple, on sait trouver $\psi \in H^1_0(\Omega)$ (et $\psi \notin H^2(\Omega)$) telle que si u est solution de (3.6) avec $f \in L^2(\Omega)$, alors il existe un unique $\alpha \in \mathbb{R}$ tel que $u - \alpha\psi \in H^2(\Omega)$.

Examinons le cas d'une approximation par éléments finis de Lagrange. Dans le cas où u est régulière, l'espace d'approximation est

$$V_{\mathcal{T}} = \text{Vect}\{\phi_i, i = 1, N_{\mathcal{T}}\},$$

où $N_{\mathcal{T}}$ est le nombre de nœuds internes du maillage $\mathcal{T}$ de Ω considéré et $(\phi_i)_{i=1,N_{\mathcal{T}}}$ la famille des fonctions de forme associées aux nœuds.

Dans le cas d'une singularité portée par la fonction ψ introduite ci-dessus, on modifie l'espace V et on prend maintenant $V_{\mathcal{T}} = \text{Vect}\{\phi_i, i = 1, N_{\mathcal{T}}\} \oplus \mathbb{R}\psi$. Notons que $V_{\mathcal{T}} \subset H^1_0(\Omega)$, car $\psi \in H^1_0(\Omega)$. Reprenons maintenant l'estimation d'erreur. Grâce au lemme de Céa, on a toujours

$$\|u - u_{\mathcal{T}}\|_{H^1} \leqslant \frac{M}{\alpha}\|u - w\|_{H^1(\Omega)}, \forall w \in V_{\mathcal{T}}.$$

On a donc également :

$$\|u - u_{\mathcal{T}}\|_{H^1} \leqslant \frac{M}{\alpha}\|u - \alpha\psi - w\|_{H^1(\Omega)}, \forall w \in V_{\mathcal{T}}.$$

puisque $\alpha\psi + w \in V_{\mathcal{T}}$. Or, $u - \alpha\psi = \widetilde{u} \in H^2(\Omega)$. Donc $\|u - u_{\mathcal{T}}\|_{H^1} \leqslant \frac{M}{\alpha}\|\widetilde{u} - w\|_{H^1(\Omega)}, \forall w \in V_{\mathcal{T}}$. Grâce aux résultats d'interpolation qu'on a admis, si on note $\widetilde{u}_I$ l'interpolée de $\widetilde{u}$ dans $V_{\mathcal{T}}$, on a :

$$\|u - u_{\mathcal{T}}\|_{H^1(\Omega)} \leqslant \frac{M}{\alpha}\|\widetilde{u} - \widetilde{u}_I\|_{H^1(\Omega)} \leqslant \frac{M}{\alpha}C_2\|\widetilde{u}\|_{H^2(\Omega)}h.$$

On obtient donc encore une estimation d'erreur en h.

Examinons maintenant le système linéaire obtenu avec cette nouvelle approximation. On effectue un développement de Galerkin sur la base de $V_{\mathcal{T}}$. On pose

$$u_{\mathcal{T}} = \sum_{i=1,N_{\mathcal{T}}} u_i\phi_i + \gamma\psi.$$

Le problème discrétisé revient donc à chercher $(u_s)_{i=1,N_{\mathcal{T}}} \subset \mathbb{R}^N$ et $\gamma \in \mathbb{R}$ tel que :

$$\begin{cases} \displaystyle\sum_{j=1,N_{\mathcal{T}}} u_j \int_\Omega \nabla\phi_j(x) \cdot \nabla\phi_i(x)\,\mathrm{d}x + \gamma \int_\Omega \nabla\psi(x) \cdot \nabla\phi_i(x)\,\mathrm{d}x = \int_\Omega f(x)\phi_i(x)\,\mathrm{d}x \quad \forall i = 1, N_{\mathcal{T}} \\ \displaystyle\sum_{j=1,N_{\mathcal{T}}} u_j \int_\Omega \nabla\phi_i(x) \cdot \nabla\psi(x)\,\mathrm{d}x + \gamma \int_\Omega \nabla\psi(x) \cdot \nabla\psi(x)\,\mathrm{d}x = \int f(x)\psi(x)\,\mathrm{d}x. \end{cases}$$

On obtient donc un système linéaire de $N_{\mathcal{T}} + 1$ équations à $N_{\mathcal{T}} + 1$ inconnues.

4.4 Implémentation d'une méthode éléments finis

On construit ici le système linéaire pour un problème à conditions aux limites mixtes de manière à envisager plusieurs types de conditions aux limites.

4.4.1 Un problème avec conditions mixtes

Soit Ω un ouvert polygonal [1], on suppose que $\partial\Omega : \Gamma_0 \cup \Gamma_1$ avec $\text{mes}(\Gamma_0) \neq 0$. On va imposer des conditions de Dirichlet sur Γ_0 et des conditions de Fourier sur Γ_1 ; c'est ce qu'on appelle des conditions "mixtes". On se donne donc des fonctions $p : \Omega \to \mathbb{R}$, $g_0 : \Gamma_0 \to \mathbb{R}$ et $g_1 : \Gamma_1 \to \mathbb{R}$ et on cherche à approcher u solution de :

$$\begin{cases} -\operatorname{div}(p(x)\nabla u(x)) + q(x)u(x) = f(x) & x \in \Omega \\ u = g_0 & x \in \Gamma_0, \\ p(x)\nabla u(x) \cdot \boldsymbol{n}(x) + \sigma u(x) = g_1(x) & x \in \Gamma_1 \end{cases} \tag{4.25}$$

où $\mathbf{n}$ désigne le vecteur unitaire normal à $\partial\Omega$ extérieure à Ω. Pour assurer l'existence et unicité du problème (4.25) [exercice 46] on se place sous les hypothèses suivantes :

$$\begin{cases} p(x) \geqslant \alpha > 0 \quad \text{p.p.} \quad x \in \Omega \\ q \geqslant 0 \\ \sigma \geqslant 0 \\ \text{mes}(\Gamma_0) > 0 \end{cases} \tag{4.26}$$

Pour obtenir une formulation variationnelle, on introduit l'espace

$$H^1_{\Gamma_0,g_0} = \{u \in H^1(\Omega); u = g_0 \text{ sur } \Gamma_0\}$$

et l'espace vectoriel associé :

$$H = H^1_{\Gamma_0,0} = \{u \in H^1(\Omega); u = 0 \text{ sur } \Gamma_0\}$$

Notons que H est un espace de Hilbert. Par contre, attention, l'espace $H^1_{\Gamma_0,g_0}$ n'est pas un espace vectoriel. On va chercher u solution de (4.25) sous la forme $u = \widetilde{u} + u_0$, avec $u_0 \in H^1_{\Gamma_0,g_0}$ et $\widetilde{u} \in H^1_{\Gamma_0,0}$. Soit $v \in H$, on multiplie (4.25) par v et on intègre sur Ω. On obtient :

$$\int_\Omega -\operatorname{div}(p(x)\nabla u(x))v(x)\,\mathrm{d}x + \int_\Omega q(x)u(x)v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x, \quad \forall v \in H.$$

En appliquant la formule de Green, il vient alors :

$$\int_\Omega p(x)\nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x - \int_{\partial\Omega} p(x)\nabla u(x) \cdot \boldsymbol{n}v(x)d\gamma(x) + \int_\Omega q(x)u(x)v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x, \quad \forall v \in H.$$

Comme $v = 0$ sur Γ_0 on a :

$$\int_{\partial\Omega} p(x)\nabla u(x) \cdot \boldsymbol{n}v(x)d\gamma(x) = \int_{\Gamma_1} p\nabla u(x) \cdot \boldsymbol{n}v(x)d\gamma(x).$$

Mais sur Γ_1, la condition de Fourier s'écrit : $\nabla u \cdot \boldsymbol{n} = -\sigma u + g_1$, et on a donc

$$\int_\Omega p(x)\nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x + \int_{\Gamma_1} p(x)\sigma u(x)v(x)d\gamma(x) + \int_\Omega qu(x)v(x)\,\mathrm{d}x = \int_\Omega f(x)v(x)\,\mathrm{d}x + \int_{\Gamma_1} g_1(x)v(x)d\gamma(x). \tag{4.27}$$

On peut écrire cette égalité sous la forme :

$$a(u,v) = \widetilde{T}(v), \text{ avec } a(u,v) = a_\Omega(u,v) + a_{\Gamma_1}. \text{ où}$$

$$\begin{cases} a_\Omega(u,v) = \displaystyle\int_\Omega p(x)\nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x + \int_\Omega q(x)u(x)v(x)\,\mathrm{d}x \\ a_{\Gamma_1}(u,v) = \displaystyle\int_{\Gamma_1} p(x)\sigma(x)u(x)v(x)d\gamma(x) \end{cases}$$

1. Dans le cas où la frontière $\partial\Omega$ de Ω n'est pas polygonale mais courbe, il faut considérer des éléments finis dits "isoparamétriques" que nous verrons plus loin

et :

$$\widetilde{T}(v) = T_\Omega(v) + T_{\Gamma_1}(v) \text{ avec } T_\Omega(v) = \int_\Omega f(x)v(x)\,\mathrm{d}x \text{ et } T_{\Gamma_1} = \int_{\Gamma_1} g_1(x)v(x)d\gamma(x).$$

On en déduit une formulation faible associée à (4.25) :

$$\text{chercher } \widetilde{u} \in H \quad a(u_0 + \widetilde{u}, v) = \widetilde{T}(v), \quad \forall v \in H, \tag{4.28}$$

où $u_0 \in H^1(\Omega)$ est un relèvement de g_0, c'est-à-dire une fonction de $H^1(\Omega)$ telle que $u_0 = g_0$ sur Γ. Le problème (4.28) peut aussi s'écrire sous la forme :

$$\begin{cases} \widetilde{u} \in H, \\ a(\widetilde{u}, v) = T(v), \forall v \in H. \end{cases} \tag{4.29}$$

où $T(v) = \widetilde{T}(v) - a(u_0, v)$. Sous les hypothèses (4.26), on peut alors appliquer le théorème de Lax–Milgram 3.3 au problème (4.29) pour déduire l'existence et l'unicité de la solution de (4.28) ; notons que, comme la forme bilinéaire a est symétrique, ce problème admet aussi une formulation variationnelle :

$$J(u) = \min_{v \in H^1_{\Gamma_0, g_0}} J(v) \quad \text{avec} \quad J(v) = \frac{1}{2}a(v,v) + T(v), \quad \forall v \in H^1_{\Gamma_0, g_0}. \tag{4.30}$$

Dans ce cas, les méthodes de Ritz et Galerkin sont équivalentes. Remarquons que l'on peut choisir u_0 de manière abstraite, tant que u_0 vérifie $u_0 = g_0$ sur Γ_0 et $u_0 \in H^1$. Intéressons nous maintenant à la méthode d'approximation variationnelle. On approche l'espace H par $H_N = \mathrm{Vect}\{\phi_1, \ldots, \phi_N\}$ et on remplace (4.29) par :

$$\begin{cases} \widetilde{u}_N \in H_N \\ a(\widetilde{u}_N, \phi_i) = T(\phi_i) - a(u_0, \phi_i) \quad \forall i = 1, \ldots, N \end{cases} \tag{4.31}$$

On pose maintenant $\widetilde{u}_N = \sum_{j=1}^N \widetilde{u}_j \phi_j$. Le problème (4.31) est alors équivalent au système linéaire :

$$\mathcal{K}\widetilde{U} = \mathcal{G},$$

avec

$$\begin{cases} \mathcal{K}_{ij} = a(\phi_j, \phi_i) \quad i,j = 1, \ldots, N \\ \widetilde{U} = (\widetilde{u}_1, \ldots, \widetilde{u}_N) \\ \mathcal{G}_i = T(\phi_i) - a(u_0, \phi_i) \quad i = 1, \ldots, N \end{cases}$$

L'implantation numérique de la méthode d'approximation nécessite donc de :

1. construire $\mathcal{K}$ et $\mathcal{G}$,
2. résoudre $\mathcal{K}\widetilde{U} = \mathcal{G}$.

Commençons par la construction de l'espace H_N et des fonctions de base pour une discrétisation par éléments finis de Lagrange du problème (4.29).

4.4.2 Construction de l'espace H_N et des fonctions de base ϕ_i

On considère une discrétisation à l'aide déléments finis de Lagrange, qu'on note : $(K_\ell, \Sigma_\ell, \mathcal{P}_\ell) \qquad \ell = 1, \ldots, L$, où L est le nombre d'éléments. On note S_i, $i = 1, \ldots, M$, les nœuds du maillage et $\phi_1, \ldots, \phi_N$, les fonctions de base, avec $N \leqslant M$. On peut avoir deux types de nœuds :

— les nœuds libres : $S_i \notin \Gamma_0$. On a N nœuds libres

— les nœuds liés : $S_i \in \Gamma_0$. On a $M - N$ nœuds liés.

Notons qu'on a intérêt à mettre des nœuds à l'intersection de Γ_0 et Γ_1 (ce seront des nœuds liés). Grâce à ceci et à la cohérence globale et locale des éléments finis de Lagrange, on a $H_N \subset H$. On a donc bien des éléments finis conformes. Récapitulons alors les notations :

— M : nombre de nœuds total,

— N : nombre de nœuds libres,

— $M_0 = M - N$: nombre de nœuds liés,

— $J_0 = \{\text{indices des nœuds liés}\} \subset \{1, \ldots, M\}$. On a $\operatorname{card} J_0 = M_0$

— $J = \{$ indices des nœuds libres$\} = \{1 \ldots M\}$ J_0. On a $\operatorname{card} J = N$.

Pour la programmation des éléments finis, on a besoin de connaître, pour chaque nœud (local) de chaque élément, son numéro dans la numérotation globale. Pour cela on introduit comme dans la section 4.2.1 le tableau $\texttt{ng}(L, N_\ell)$, où L est le nombre d'éléments et N_ℓ est le nombre de nœuds par élément (on le suppose constant par souci de simplicité, N_ℓ peut en fait dépendre de L, le maillage pourrait par exemple être composé de triangles et de quadrangles). Pour tout $\ell \in \{1, \ldots, L\}$ et tout $r \in \{1, \ldots, N_\ell\}$, $\texttt{ng}(\ell, r)$ est alors le numéro global du r-ème nœud du ℓ-ème élément. On a également besoin de connaître les coordonnées de chaque nœud. On a donc deux tableaux x et y de dimension M, où $x(i), y(i)$ représentent les coordonnées du i-ème nœud. Notons que les tableaux $\texttt{ng}, x$ et y sont des données du mailleur (qui est un module externe par rapport au calcul éléments finis proprement dit). Pour les conditions aux limites, on se donne deux tableaux :

— `cf` : conditions de Fourier,

— `cd` : conditions de Dirichlet.

4.4.3 Construction de $\mathcal{K}$ et $\mathcal{G}$

On cherche à construire la matrice $\mathcal{K}$ d'ordre $(N \times N)$, définie par :

$$\mathcal{K}_{ij} = a(\phi_j, \phi_i) \qquad i, j \in J,$$

ainsi que le vecteur $\mathcal{G}$, défini par :

$$\mathcal{G}_i = T(\phi_i) - a(u_0, \phi_i) \qquad i \in J \quad \operatorname{card} J = N$$

La première question à résoudre est le choix de u_0. En effet, contrairement au cas unidimensionnel [exercice 41] il n'est pas toujours évident de trouver $u_0 \in H^1_{\Gamma_0, g_0}$. Pour se faciliter la tâche, on commet ce qu'on appelle un "crime variationnel", en remplaçant u_0 par $u_{0,N} = \sum_{j \in J_0}^{N} g_0(S_j)\phi_j$. Notons qu'on a pas forcément : $u_{0,N} \in H^1_{\Gamma_0, g_0}$; c'est en ce sens que l'on commet un "crime". Mais par contre, on a bien $u_{0,N}(S_j) = u_0(S_j)$ pour tout $j \in J_0$. On peut voir la fonction $u_{0,N}$ comme une approximation non conforme de $u_0 \in H^1_{\Gamma_0, g_0}$. On remplace donc $\mathcal{G}_i$ par : $\mathcal{G}_i = T(\phi_i) - \sum_{j \in J_0} g_0(S_j) a(\phi_j, \phi_i)$ Calculons maintenant $a(\phi_j, \phi_i)$ pour $j = 1, \ldots, M$, et $i = 1, \ldots, M$. Remarquons que l'on se sert pour l'implantation pratique de la méthode, des fonctions de forme associées aux nœuds liés même si dans l'écriture du problème discret théorique, on n'en avait pas besoin.

Calcul de $\mathcal{K}$ et $\mathcal{G}$

1. Calcul des contributions intérieures : on initialise les coefficients de la matrice $\mathcal{K}$ et les composantes par les contributions provenant de a_Ω et T_Ω.

$$\left.\begin{array}{l} \mathcal{K}_{ij} = a_\Omega(\phi_j, \phi_i) \\ \mathcal{G}_i = T_\Omega(\phi_i) \end{array}\right\} \begin{array}{l} i = 1, \ldots, N, \\ j = 1, \ldots, N. \end{array}$$

2. Calcul des termes de bord de Fourier. On ajoute maintenant à la matrice $\mathcal{K}$ les contributions de bord :

$$\mathcal{K}_{ij} \longleftarrow \mathcal{K}_{ij} + a_{\Gamma_i}(\phi_j, \phi_i) \quad i = 1, \ldots, N \quad j = 1, \ldots, N$$
$$\mathcal{G}_i \longleftarrow \mathcal{G}_i + T_{\Gamma_i}(\phi_i) \quad i = 1 \ldots M$$

3. Calcul des termes de bord de Dirichlet. On doit tenir compte du relèvement de la condition de bord :

$$\mathcal{G}_i \longleftarrow \mathcal{G}_i - \sum_{j \in J_0} g_0(S_j)\mathcal{K}_{ij} \quad \forall i \in J$$

Après cette affectation, les égalités suivantes sont vérifiées :

$$\mathcal{K}_{ij} = a(\phi_j, \phi_i) \qquad i, j \in J(\cup J_0)$$
$$\mathcal{G}_i = T(\phi_i) - a(u_{0,N}, \phi_i)$$

Il ne reste plus qu'à résoudre le système linéaire

$$\sum_{j\in J} \mathcal{K}_{ij}\alpha_j = \mathcal{G}_i \quad \forall i \in J \tag{4.32}$$

4. Prise en compte des nœuds liés. Pour des questions de structure de données, on inclut en général les nœuds liés dans la résolution du système et on résout donc le système linéaire d'ordre $M \geqslant N$ suivant :

$$\sum_{j=1,\dots,N} \widetilde{\mathcal{K}}_{ij}\alpha_j = \mathcal{G}_i \quad \forall i = 1,\dots,N \tag{4.33}$$

avec $\widetilde{\mathcal{K}}_{ij} = \mathcal{K}_{ij}$ pour $i,j \in J$, $\widetilde{\mathcal{K}}_{ij} = 0$ si $(i,j) \notin J^2$ et $i \neq j$ et $\widetilde{\mathcal{K}}_{ii} = 1$ si $i \notin J$. Ces deux systèmes sont équivalents, puisque les valeurs aux nœuds liées sont fixées.
Si on a numéroté les nœuds de manière à ce que tous les nœuds liés soient en fin de numérotation, c'est-à-dire si $J = \{1,\dots,N\}$ et $J_0 = \{N+1,\dots,M\}$, le système (4.33) est de la forme :

$$\left[\begin{array}{c|c} \mathcal{K} & 0 \\ \hline 0 & Id_M \end{array}\right], \quad U = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha u_N \\ -- \\ \alpha_{N+1} \\ \vdots \\ \alpha_M \end{pmatrix} \quad \text{et} \quad \mathcal{G} = \begin{pmatrix} \mathcal{G}_1 \\ \vdots \\ \mathcal{G}_N \\ -- \\ \mathcal{G}_{N+1} \\ \vdots \\ \mathcal{G}_M \end{pmatrix}$$

Dans le cas où la numérotation est quelconque, les nœuds liés ne sont pas forcément à la fin et pour obtenir le système linéaire d'ordre M (4.33) (donc incluant les inconnues α_i, $i \in J_0$, qui n'en sont pas vraiment) on peut adopter deux méthodes :

(a) Première méthode : on force les valeurs aux nœuds liés de la manière suivante :

$$\begin{aligned} &\mathcal{K}_{ii} \longleftarrow 1 \text{ pour tout } i \in J_0 \\ &\mathcal{K}_{ij} \longleftarrow 0 \text{ pour tout } i \in J_0,\ j \in \{1\dots M\},\ i \neq j \\ &\mathcal{G}_i \longleftarrow g_0(S_i) \text{ pour tout } i \in J_0 \end{aligned}$$

(b) Deuxième méthode : on force les valeurs aux nœuds liés de la manière suivante :

$$\begin{aligned} &\mathcal{K}_{ii} \longleftarrow 10^{20} \qquad \forall i \in J_0 \\ &\mathcal{G}_i \longleftarrow 10^{20} g_0(S_i) \qquad \forall i \in J_0 \end{aligned}$$

La deuxième méthode permet d'éviter l'affectation à 0 de coefficients extra-diagonaux de la matrice. Elle est donc un peu moins chère en temps de calcul.

Conclusion Après les calculs 1, 2, 3, 4, on a obtenu une matrice $\mathcal{K}$ d'ordre $M \times M$ et le vecteur $\mathcal{G}$ de $\mathbb{R}^M$. Soit $\alpha \in \mathbb{R}^M$ la solution du système $\mathcal{K}\alpha = \mathcal{G}$. Rappelons qu'on a alors :

$$u_N = \sum_{i=1}^{M} \alpha_i\phi_i = \sum_{i\in J} \alpha_i\phi_i + \sum_{i\in J_0} \alpha_i\phi_i = \widetilde{u}_N + u_0.$$

Remarque 4.20 — Numérotation des nœuds. Si on utilise une méthode itérative sans préconditionnement, la numérotation des nœuds n'est pas cruciale. Elle l'est par contre dans le cas d'une méthode directe et si on utilise une méthode itérative avec préconditionnement. Le choix de la numérotation s'effectue pour essayer de minimiser la largeur de bande.

4.4.4 Calcul de a_Ω et T_Ω et des matrices élémentaires

Détaillons maintenant le calcul des contributions intérieures, c'est-à-dire $a_\Omega(\phi_i,\phi_j)$ $i = 1,\dots M$, $j = 1,\dots,M$ et $T_\Omega(\phi_i)$ $i = 1,\dots,M$. Par définition,

$$a_\Omega(\phi_i,\phi_j) = \int_\Omega p(x)\nabla\phi_i(x)\cdot\nabla\phi_j(x)\,\mathrm{d}x + \int_\Omega q(x)\phi_i(x)\phi_j(x)\,\mathrm{d}x.$$

Décomposons Ω à l'aide du maillage éléments finis : $\Omega = \bigcup_{\ell=1}^{L} K_\ell$. En notant

$$\theta(\phi_i, \phi_j)(x) = p(x)\nabla\phi_i(x)\nabla\phi_j(x) + q(x)\phi_i(x)\phi_j(x),$$

on a : $a_\Omega(\phi_i, \phi_j) = \sum_{\ell=1}^{L} \int_{K_\ell} \theta(\phi_i, \phi_j)\,\mathrm{d}x$. Pour r et s numéros locaux de l'élément K_ℓ, on définit le coefficient (r, s) de la matrice élémentaire associée à K_ℓ par

$$k^\ell_{r,s} = \int_{K_\ell} \theta(\phi_s(x, y), \phi_r(x, y))\,\mathrm{d}x\,\mathrm{d}y. \tag{4.34}$$

On va calculer $k^\ell_{r,s}$ puis on calcule $a_\Omega(\phi_i, \phi_j)$, en effectuant un parcours sur les éléments, ce qui s'exprime par l'algorithme 1. On a ainsi construit complètement la matrice de rigidité $\mathcal{K}$. Il reste à savoir comment calculer le coefficient $k^\ell_{r,s}$ défini par (4.34). On commence par effectuer le calcul de l'intégrale sur l'élément de référence $\bar{K}$. On calcule ensuite la valeur du coefficient $k^\ell_{r,s}$ de la matrice élémentaire K^ℓ par des changements de variable à l'aide de la transformation F_ℓ (voir Figure 4.4).

initialisation : $\mathcal{K}_{ij} \longleftarrow 0$, $i = 1, \ldots M$, $j \leqslant i$;
Boucle sur les éléments ;
pour $\ell = 1$ *à* L **faire**
  pour $r = 1$ *à* N_ℓ **faire**
    $i = \mathtt{ng}(\ell, r)$ numéro global du nœud r de l'élément ℓ ;
    pour $s = 1$ *à* r **faire**
      calcul de $k^\ell_{r,s}$;
      $j = \mathtt{ng}(\ell, s)$;
      si $i \geqslant j$ **alors**
        $\mathcal{K}_{ij} \longleftarrow \mathcal{K}_{ij} + k^\ell_{r,s}$;
      sinon
        $\mathcal{K}_{ji} \longleftarrow \mathcal{K}_{ji} + k^\ell_{rs}$;
      fin
    fin
  fin
fin

Algorithme 1 : Assemblage de la matrice

Notons :

$$\begin{aligned} F_\ell(\bar{x}, \bar{y}) &= (x, y) \\ &= (a^\ell_0 + a^\ell_1\bar{x} + a^\ell_2\bar{y}, b^\ell_0 + b^\ell_1\bar{x} + b^\ell_2 y) \end{aligned} \tag{4.35}$$

Notons que les coefficients a^ℓ_i et b^ℓ_i sont déterminés à partir de la connaissance des coordonnées $(x(i), y(i))$ où $i = \mathtt{ng}(\ell, r)$. En effet, on peut déduire les coordonnées locales $x(r), y(r), r = 1, N_\ell$, des nœuds de l'élément ℓ, à partir des coordonnées globales des nœuds $(x(i), y(i))$ et du tableau $\mathtt{ng}(\ell, r) = i$.

Calculons maintenant le coefficient $k^\ell_{r,s}$ de la matrice élémentaire correspondant à l'élément courant K_ℓ, qui est défini par (4.34). Dans l'intégrale sur K_ℓ, on a $(x, y) = F_\ell(\bar{x}, \bar{y})$; donc par changement de variables, on a :

$$k^\ell_{r,s} = \int_{\bar{K}} \theta(\phi_s \circ F_\ell(\bar{x}, \bar{y}), \phi_r \circ F_\ell(\bar{x}, \bar{y}))\, Jac_{\bar{x},\bar{y}}(F_\ell)\, d\bar{x}d\bar{y}$$

ou $Jac_{\bar{x},\bar{y}}(F_\ell)$ désigne le Jacobien de F_ℓ en $(\bar{x}, \bar{y})$. Or, $\phi_s \circ F_\ell = \bar{\phi}_s$, et, puisque F_ℓ est définie par (4.35), on a :

$$Jac(F_\ell) = Det(DF_\ell) = \begin{vmatrix} a^\ell_1 & b^\ell_1 \\ a^\ell_2 & b^\ell_2 \end{vmatrix} = \left|a^\ell_1 b^\ell_2 - a^\ell_2 b^\ell_1\right|$$

donc $k^\ell_{r,s} = Jac(F_\ell)\bar{k}_{r,s}$, où

$$\bar{k}_{r,s} = \int_{\bar{K}} \theta(\bar{\phi}_s(\bar{x}, \bar{y}), \bar{\phi}_r(\bar{x}, \bar{y}))d\bar{x}d\bar{y}$$

Etudions maintenant ce qu'on obtient pour $\bar{k}_{r,s}$ dans le cas du problème modèle (4.25), on a :

$$\bar{k}_{r,s} = \int_{\bar{\ell}} \left[p(\bar{x},\bar{y}) \nabla\bar{\phi}_s(\bar{x},\bar{y}) \nabla\bar{\phi}_r(\bar{x},\bar{y}) + q(\bar{x},\bar{y}) \bar{\phi}_s(\bar{x},\bar{y}) \bar{\phi}_r(\bar{x},\bar{y}) \right] d\bar{x} d\bar{y}.$$

Les fonctions de base $\bar{\phi}_s$ et $\bar{\phi}_r$ sont connues ; on peut donc calculer $\bar{k}_{r,s}$ explicitement si p et q sont faciles à intégrer. Si les fonctions p et q ou les fonctions de base $\bar{\phi}$, sont plus compliquées, on calcule $\bar{k}_{r,s}$ en effectuant une intégration numérique. Rappelons que le principe d'une intégration numérique est d'approcher l'intégrale I d'une fonction continue donnée ψ,

$$I = \int_{\bar{\ell}} \psi(\bar{x},\bar{y}) d\bar{x} d\bar{y}, \text{ par } \widetilde{I} = \sum_{i=1}^{N_{\text{int}}} \omega_i(P_i) \psi(P_i),$$

où N_{int} est le nombre de points d'intégration, notés P_i, qu'on appelle souvent points d'intégration de Gauss, et les coefficients ω_i sont les poids associés. Notons que les points P_i et les poids ω_i sont indépendants de ψ. Prenons par exemple, dans le cas unidimensionnel, $\bar{K} = [0\,;1]$, $p_1 = 0$, $p_2 = 1$ et $\omega_1 = \omega_2 = \frac{1}{2}$. On approche alors

$$I = \int_0^1 \psi(x)\,\mathrm{d}x \text{ par } \widetilde{I} = \frac{1}{2}(\psi(0) + \psi(1)).$$

C'est la formule (bien connue) des trapèzes. Notons que dans le cadre d'une méthode d'éléments finis, il est nécessaire de s'assurer que la méthode d'intégration numérique choisie soit suffisamment précise pour que :

1. le système $\mathcal{K}\alpha = \mathcal{G}(N \times N)$ reste inversible,
2. l'ordre de convergence de la méthode reste le même.

Examinons maintenant des éléments en deux dimensions d'espace.

1. Élément fini P1 sur triangle. Prenons $N_{\text{int}} = 1$ (on a donc un seul point de Gauss), choisissons $P_1 = (1/3, 1/3)$, le centre de gravité du triangle $\bar{K}$ et $\omega_1 = 1$. On approche alors

   $$I = \int_{\bar{\ell}} \psi(\bar{x}) d\bar{x} \text{ par } \psi(P_1).$$

   On vérifiera que cette intégration numérique est exacte pour les polynômes d'ordre 1 (exercice 65).

2. Élément fini P2 sur triangles. On prend maintenant $N_{\text{int}} = 3$ et on choisit comme points de Gauss :

   $$P_1 = (1/2, 0),\ P_2 = (1/2, 1/2),\ P_3 = (0, 1/2),$$

   et les poids d'intégration $\omega_1 = \omega_2 = \omega_3 = 1/6$. On peut montrer que cette intégration numérique est exacte pour les polynômes d'ordre 2 [exercice 65].

Remarquons que, lors de l'intégration numérique du terme élémentaire

$$k^{\ell}_{r,s} = \int_{\bar{\ell}} \left[p(\bar{x},\bar{y})(F_\ell(\bar{x},\bar{y})) \nabla\bar{\phi}_r(\bar{x},\bar{y}) \cdot \nabla\bar{\phi}_s(\bar{x},\bar{y}) + q(\bar{x},\bar{y})(F_\ell(\bar{x},\bar{y})) \bar{\phi}_r(\bar{x},\bar{y}) \bar{\phi}_s \right] d\bar{x} d\bar{y},$$

on approche $k^{\ell}_{r,s}$ par

$$\bar{k}_{r,s} \simeq \sum_{i=1}^{N_{\text{int}}} \omega_i \left[p(F_\ell(P_i)) \nabla\bar{\phi}_r(P_i) \cdot \nabla\bar{\phi}_s(P_i) + q(F_\ell(P_i)) \bar{\phi}_r(P_i) \bar{\phi}_s(P_i) \right].$$

Les valeurs $\nabla\bar{\phi}_r(P_i)$, $\nabla\bar{\phi}_s(P_i)$, $\bar{\phi}_r(P_i)$ et $\bar{\phi}_s(P_i)$ sont calculées une fois pour toutes et dans la boucle sur ℓ, il ne reste donc plus qu'à évaluer les fonctions p et q aux points $F_\ell(P_i)$. Donnons maintenant un résumé de la mise en œuvre de la procédure d'intégration numérique (indépendante de ℓ). Les données de la procédure sont :

— les coefficients $\omega_i, i = 1, \ldots, N_{\text{int}}$,

— les coordonnées $(x_{\text{pg}}(i), y_{\text{pg}}(i))$, $i = 1, \ldots, N_{\text{int}}$ des points de Gauss,

— les valeurs de $\phi_r, \partial_x\phi_r$ et $\partial_y\phi_r$ aux points de Gauss, notées $\phi(r,i), \phi_x(r,i)$ et $\phi_y(r,i)$, pour $r = 1 \dots N_\ell$ et $i = 1, \dots, N_{\text{int}}$.

Pour ℓ donné, on cherche à calculer :

$$I = \int_{\bar{K}} p(F_\ell(\bar{x},\bar{y}))\frac{\partial\phi_r}{\partial\bar{x}}(\bar{x},\bar{y})\frac{\partial\phi_s}{\partial\bar{y}}(\bar{x},y)d\bar{x}d\bar{y} + \int_e q(F_\ell(\bar{x},\bar{y}))\phi_r(\bar{x},y)d\bar{x}d\bar{y}\phi_s(\bar{x},\bar{y}).$$

On propose l'algorithme 2. On procède de même pour le calcul du second membre :

$$T_\Omega(\phi_i) = \int_\Omega f(x,y)\phi_i'(x,y)\,\mathrm{d}x\,\mathrm{d}y = \sum_{\ell=1}^{L} g_\ell, \text{ où } g_\ell = \int_{K_\ell} f(x,y)\phi_i(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

On obtient l'algorithme 3. Il reste le calcul de g_ℓ^r qui se ramène au calcul de l'élément de référence

Initialisation $I \longleftarrow 0$;
pour $i = 1$ *à* N_{int} **faire**
  $p_i = p(F_\ell(P_i))$;
  $q_i = q(F_\ell(P_i))$;
  $I \longleftarrow I + \omega_i(p_i\phi_x(r,i)\phi_y(s,i) + q_i\phi(r,i)\phi(s,i))$
fin

Algorithme 2 : Assemblage de la matrice

Initialisation de $\mathcal{G}$ à 0 : $\mathcal{G}_i \leftarrow 0$, $i = 1$ à M ;
pour $\ell = 1$ *à* L **faire**
  pour $\ell = 1$ *à* L **faire**
    Calcul de $g_\ell^r = \int_{\ell_e} f(x,y)\phi_r(x,y)\,\mathrm{d}x\,\mathrm{d}y$;
    $i = \mathtt{ng}(\ell, r)$;
    $\mathcal{G}_i \longleftarrow \mathcal{G}_i + g_\ell^r$
  fin
fin

Algorithme 3 : Assemblage du second membre

par changement de variable. On a :

$$g_\ell^r = \int_{K_\ell} f(x,y)\phi_r(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\bar{K}} f\circ F_\ell(\bar{x},\bar{y})\bar{\phi}_r(\bar{x},\bar{y})\mathrm{Jac}_{\bar{x},\bar{y}}(F_\ell)d\bar{x}d\bar{y}.$$

L'intégration numérique est identique à celle effectuée pour $\bar{k}_{r,s}$.

4.4.5 Calcul de a_{Γ_1} et T_{Γ_1} — Contributions des arêtes de bord "Fourier"

Détaillons maintenant le calcul des contributions des bords où s'applique la condition de Fourier, c'est-à-dire $a_{\Gamma_1}(\phi_i,\phi_j)$ $i = 1,\dots M$, $j = 1,\dots,M$ et $T_{\Gamma_1}(\phi_i)$ $i = 1,\dots,M$. Par définition,

$$a_{\Gamma_1}(\phi_i,\phi_j) = \int_{\Gamma_1} p(x)\nabla\phi_i(x)\cdot\nabla\phi_j(x)\,\mathrm{d}x + \int_{\Gamma_1} q(x)\phi_i(x)\phi_j(x)\,\mathrm{d}x.$$

Notons que $a_{\Gamma_1}(\phi_i,\phi_j) = 0$ si ϕ_i et ϕ_j sont associées à des nœuds S_i, S_j de d'un élément sans arête commune avec les arêtes de la frontière. Soit $L1$ le nombre d'arêtes $\epsilon_k, k = 1,\dots,L1$ du maillage incluses dans Γ_1. Rappelons que les nœuds soumis aux conditions de Fourier sont repertoriés dans un tableau `cf`, de dimensions $(L1, 2)$, qui donne les informations suivantes :

1. $\mathtt{cf}(k,1)$ contient le numéro ℓ de l'élément K_ℓ auquel appartient l'arête ϵ_k.
2. $\mathtt{cf}(k,2)$ contient le premier numéro des nœuds de l'arête ϵ_k dans l'élément K_ℓ. On suppose que la numérotation des nœuds locaux a été effectuée de manière adroite, par exemple dans le sens trigonométrique. Dans ce cas, $\mathtt{cf}(k,2)$ détermine tous les nœuds de l'arête ϵ_k dans l'ordre, puisqu'on connait le nombre de nœuds par arête et le sens de numérotation des nœuds. Donnons des exemples pour trois cas différents, représentés sur la figure 4.8.
   (a) Dans le premier cas (à droite sur la figure), qui représente un élément fini P1, on a $\mathtt{cf}(k,2) = 3$ et le nœud suivant sur l'arête est 1.
   (b) Dans le second cas (au centre sur la figure), qui représente un élément fini P2, on a $\mathtt{cf}(k,2) = 3$ et les nœuds suivants sur l'arête sont 4 et 5.
   (c) Enfin dans l'élément P1 "de coin" représenté à gauche sur la figure, on a $\mathtt{cf}(k,1) = \ell$, $\mathtt{cf}(k',1) = \ell$, $\mathtt{cf}(k,2) = 1$, $\mathtt{cf}(k',2) = 2$.

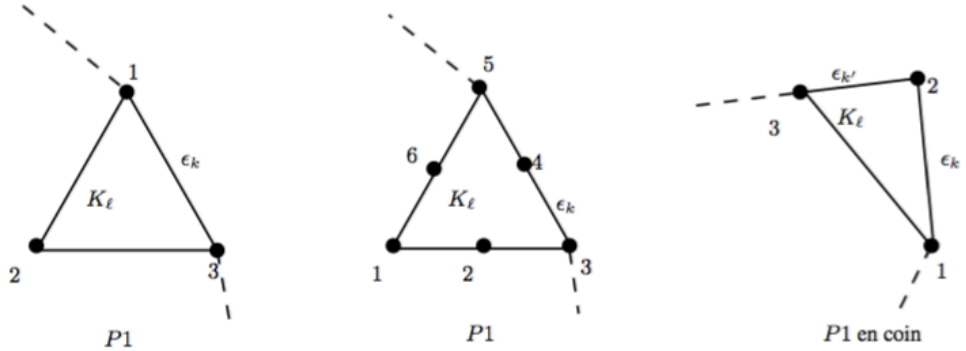


FIGURE 4.8 – Exemples de numérotation d'arête du bord

pour $k = 1 \dots L1$ **faire**
  $\ell = \texttt{cf}(k, 1)$;
  pour *chaque* $(r, s) \in S_k$ **faire**
    calcul de
    $I^\ell_{rs} = \int_{C_k} p(x)\sigma(x)\phi^\ell_r(x)\phi^\ell_s(x)\,\mathrm{d}x$;
    $i = \texttt{ng}(\ell, r)$;
    $j = \texttt{ng}(\ell, s)$;
    si $j \leqslant i$ **alors**
      $\mathcal{K}_{ij} \longleftarrow \mathcal{K}_{ij} + I^\ell_{rs}$
    sinon
      $\mathcal{K}_{ij} \longleftarrow \mathcal{K}_{ji} + I^\ell_{rs}$
    fin
  fin
fin

Algorithme 4 : Condition de Fourier

pour $i_0 = 1, \dots, M_0$ **faire**
  $j = \texttt{cd}(i_0)$;
  $a = g_0(S_j)$;
  $j = \texttt{ng}(\ell, s)$;
  si $i \leqslant j$ **alors**
    $\mathcal{G}_i \longleftarrow \mathcal{G}_i - a\mathcal{K}_{ij}$
  sinon
    $\mathcal{G}_i \longleftarrow \mathcal{G}_i - a\mathcal{K}_{ji}$
  fin
fin

Algorithme 5 : Condition de Fourier

Pour $k = 1, \dots, L1$, on note $\hat{S}_k$ l'ensemble des nœuds locaux de ϵ_k, donnés par $\texttt{cf}(k, 2)$ en appliquant la règle *ad hoc* (par exemple le sens trigonométrique). On peut alors définir :

$$S_k = \{(r, s) \in (\hat{S}_k)^2 / r < s\}$$

La prise en compte des conditions de Fourier est décrite dans l'algorithme 4.

Le calcul de I^ℓ_{rs} s'effectue sur l'élément de référence (avec éventuellement intégration numérique). De même, on a une procédure similaire pour le calcul de $T_{\Gamma_1} = \int_{\Gamma_1} p(x)g_1(x)v(x)d\gamma(x)$:

$$\mathcal{G}_i \longleftarrow \mathcal{G}_i + \int_{\Gamma_2} p(x)g_1(x)\phi_i(x)d\gamma(x)$$

4.4.6 Prise en compte des nœuds liés dans le second membre

Aprés les calculs précédents, on a maintenant dans $\mathcal{G}_i$:

$$\mathcal{G}_i = \int_\Omega f(x)\phi_i(x)\,\mathrm{d}x + \int_{\Gamma_1} p(x)g_1(x)\phi_i(x)d\gamma(x)$$

Il faut maintenant retirer du second membre, les combinaisons venant des nœuds liés :

$$\mathcal{G}_i \longleftarrow \mathcal{G}_i - \sum_{j \in J_0} g_0(S_j)a(\phi_j, \phi_i)$$

où J_0 est l'ensemble des indices des nœuds liés. On utilise pour cela le tableau `cd` qui donne les conditions, de Dirichlet, de dimension M_0 où $M_0 = card J_0$. Pour $i_0 = 1, \dots, M_0$, $\texttt{cd}(i_0) = j_0 \in J_0$ est le numéro du nœud lié dans la numérotation globale. La procédure est décrite dans l'algorithme 5.

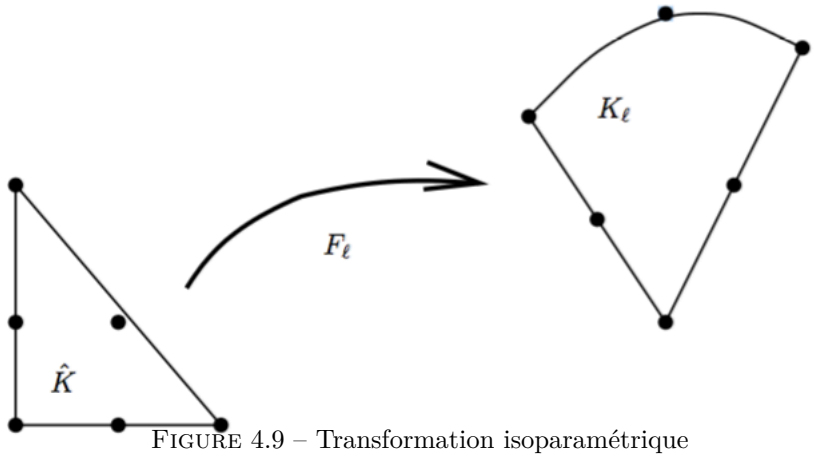


Figure 4.9 – Transformation isoparamétrique

4.4.7 Stockage de la matrice $\mathcal{K}$

Remarquons que la matrice $\mathcal{K}$ est creuse (et même très creuse), en effet $a(\phi_j, \phi_i) = 0$ dès que

$$\text{supp}(\phi_i) \cap \text{supp}(\phi_j) = \phi$$

Examinons une possibilité de stockage de la matrice $\mathcal{K}$. Soit NK le nombre d'éléments non nuls de la matrice $\mathcal{K}$ On peut stocker la matrice dans un seul tableau $KMAT$ en mettant bout à bout les coefficients non nuls de la première ligne, puis ceux de la deuxième ligne, etc... jusqu'à ceux de la dernieère ligne. Pour repérer les éléments de $\mathcal{K}$ dans le tableau $KMAT$, on a alors besoin de pointeurs. Le premier pointeur, nommé, IC est de dimension NK.

La valeur de $IC(k)$ est le numéro de la colonne de $K(k)$. On introduit alors le pointeur $IL(\ell), \ell = 1, \ldots, NL$, où NL est le nombre de lignes, où $IL(\ell)$ est l'indice dans $KMAT$ du début de la ℓ-ème ligne. L'identification entre $KMAT$ et $\mathcal{K}$ se fait alors par la procédure 6.

pour $k = 1 \ldots NK$ **faire**
  si $IL(m) \leqslant k < IL(m+1)$
    alors
      $KMAT(k) = \mathcal{K}_{m,IC(k)}$
  fin
fin

Algorithme 6 : Condition de Fourier

La matrice $\mathcal{K}$ est symétrique définie positive, on peut donc utiliser une méthode de type gradient conjugué préconditionné. Notons que la structure de la matrice dépend de la numérotation des nœuds. Il est donc important d'utiliser des algorithmes performants de maillage et de numérotation.

4.5 Éléments finis isoparamétriques

Dans le cas ou Ω est polygonal, si on utilise des éléments finis de type P2, les nœuds de la frontière sont effectivement sur la frontière même si on les calcule à partir de l'élément fini de référence. Par contre, si le bord est courbe, ce n'est plus vrai. L'utilisation d'éléments finis isoparamétriques va permettre de faire en sorte que tous les nœuds frontières soient effectivement sur le bord, comme sur la figure 4.9. Pour obtenir une transformation isoparamétrique, on définit

$$F_\ell : K \to K_\ell$$
$$(\bar{x}, \bar{y}) \mapsto (x, y)$$

à partir des fonctions de base de l'élément fini de référence :

$$x = \sum_{r=1}^{N_\ell} x_r \bar{\phi}_r(\bar{x}, \bar{y}), \; y = \sum_{r=1}^{N_\ell} y_r \bar{\phi}_r(\bar{x}, \bar{y}),$$

où N_ℓ est le nombre de nœuds de l'élément et (x_r, y_r) sont les coordonnées du r-ème nœud de K_ℓ. Remarquons que la transformation F_ℓ isoparamétrique P1 est identique à celle des éléments finis classiques. Par contre, la transformation isoparamétrique P2 n'est plus affine, alors qu'elle l'est en éléments finis classiques. Notons que les fonctions de base locales vérifient toujours

$$\phi_r^\ell \circ F_\ell = \phi_r, \forall \ell = 1, \ldots, L, \quad \forall r = 1, \ldots, N_\ell.$$

On peut alors se poser le problème de l'inversibilité de F_ℓ. On ne peut pas malheureusement démontrer que F_ℓ est inversible dans tous les cas, toutefois, cela s'avère être le cas dans la plupart des cas pratiques. L'intérêt de la transformation isoparamétrique est de pouvoir traiter les bords courbes, ainsi que les éléments finis Q1 sur quadrilatères. Notons que le calcul de ϕ_r^ℓ est toujours inutile, car on se ramène encore à l'élément de référence.

4.6 Exercices

4.6.1 Énoncés

corrigé p.169

Exercice 51 — Éléments finis P1 pour le problème de Dirichlet. Soit $f \in L^2(]0\,;1[)$. On s'intéresse au problème suivant :

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

dont on a étudié une formulation faible à l'exercice 39. Soient $N \in \mathbb{N}$, $h = 1/(N+1)$ et $x_i = ih$ pour $i = 0, \ldots, N+1$ et $K_i = [x_i, x_{i+1}]$ pour $i = 0, \ldots, N$. Soit $H_N = \{v \in C([0\,;1], \mathbb{R}) \text{ t.q. } v|_{K_i} \in \mathcal{P}_1, i = 0, \ldots, N \text{ et } v(0) = v(1) = 0\}$ où $\mathcal{P}_1$ désigne l'ensemble des polynômes de degré inférieur ou égal à 1.

1. Montrer que $H_N \subset H_0^1$.

2. Pour $i = 1, \ldots, N$, on pose :

$$\phi_i(x) = \begin{cases} 1 - \dfrac{|x - x_i|}{h} & \text{si } x \in K_i \cup K_{i-1} \\ 0 & \text{sinon} \end{cases}$$

Montrer que $\phi_i \in H_N$ pour tout $i = 1, \ldots, N$ et que H_N est engendré par la famille $\{\phi_1, \ldots, \phi_N\}$.

3. Donner le système linéaire obtenu en remplaçant H par H_N dans la formulation faible. Comparer avec le schéma obtenu par différences finies.

corrigé p.170

Exercice 52 — Conditions aux limites de Fourier et Neumann. Soit $f \in L^2(]0\,;1[)$. On s'intéresse au problème :

$$\begin{cases} -u''(x) + u(x) = f(x) & x \in]0\,;1[\\ u'(0) - u(0) = 0 \\ u'(1) = -1 \end{cases} \tag{4.36}$$

L'existence et l'unicité d'une solution faible ont été démontrées à l'exercice 42. On s'intéresse maintenant à la discrétisation de (4.36).

1. Écrire une discrétisation par différences finies pour un maillage non uniforme. Écrire le système linéaire obtenu.

2. Écrire une discrétisation de (4.36) par volumes finis pour un maillage non uniforme. écrire le système linéaire obtenu.

3. Écrire une discrétisation par éléments finis conformes de type Lagrange P1 de (4.36) pour un maillage non uniforme. Écrire le système linéaire obtenu.

Exercice 53 — Conditions aux limites de Fourier et Neumann, bis. Soit $f \in L^2(]0\,;1[)$. On s'intéresse au problème :

$$\begin{cases} -u''(x) - u'(x) + u(x) = f(x) \quad x \in]0\,;1[\\ u(0) + u'(0) = 0 \\ u(1) = 1 \end{cases}$$

La formulation faible de ce problème a été étudiée à l'exercice 43.

1. Écrire une discrétisation par éléments finis conformes de type Lagrange P1 pour un maillage uniforme. Écrire le système linéaire obtenu.

2. Écrire une discrétisation par volumes finis centrés pour un maillage uniforme. Écrire le système linéaire obtenu.

3. Écrire une discrétisation par différences finies centrés pour un maillage uniforme. Écrire le système linéaire obtenu.

4. Quel est l'ordre de convergence de chacune des méthodes étudiées aux questions précédentes ?

Exercice 54 — Éléments finis pour un problème périodique. Soit $\Omega =]0, 1[$. On cherche à déterminer $u \in H^2(\Omega)$ tel que

$$\begin{cases} -u'' + u = f \quad \text{p.p. dans } \Omega & \text{(4.37a)} \\ u(1) = u(0) & \text{(4.37b)} \\ u'(1) = u'(0) + p & \text{(4.37c)} \end{cases}$$

avec $f \in L^2(\Omega)$ et p un réel donné.

1. Montrer qu'il existe $C \in \mathbb{R}$ tel que pour tout $v \in H^1(\Omega)$, $\sup_{x\in[0\,;1]} |v(x)| \leqslant C\|v\|_{H^1(\Omega)}$. [On pourra commencer par montrer que $|v(x)| \leqslant |v(y)| + \int_0^1 |v'(t)|dt$ et intégrer cette inégalité entre 0 et 1 par rapport à y.]

2. Soit $V = \{v \in H^1(\Omega) : v(0) = v(1)\}$.

(a) Vérifier que V est un sous espace vectoriel de $H^1(\Omega)$.

(b) Donner une formulation faible associée au problème aux limites (4.37) de la forme $u \in V$ solution de

$$a(u, v) = T(v) \quad \forall v \in V \tag{4.38}$$

en explicitant la forme bilinéaire a et la forme linéaire T.

(c) Prouver l'existence et l'unicité de la solution de la formulation faible (4.38).

(d) Prouver l'équivalence de la formulation (4.37) avec la formulation faible (4.38), dans le cas d'une solution $u \in H^2(\Omega)$.

3. On cherche à résoudre numériquement ce problème par éléments finis P1. On se donne pour cela une discrétisation uniforme de l'intervalle $[0\,;1]$, de pas $h = 1/n$, avec $n \in \mathbb{N}^*$. Pour $i = 0, \ldots, n$ on pose $x_i = ih$. Soit $V_h = \{u \in V : u|_{[x_i, x_{i+1}]} \in \mathcal{P}_1, i = 0, \ldots, n-1\}$ où P1 désigne l'ensemble des polynômes de degré inférieur ou égal à 1. Soit ϕ la fonction définie de $\mathbb{R}$ dans $\mathbb{R}$ par :

$$\phi(x) = \begin{cases} 1 - x & \text{si } x \in [0\,;1] \\ 1 + x & \text{si } x \in [-1\,;0] \\ 0 & \text{sinon} \end{cases}$$

On définit les fonctions ϕ_i, $i = 1, \ldots, n$, de $[0\,;1]$ dans $\mathbb{R}$, par

$$\phi_i(x) = \phi\Big(\frac{x - x_i}{h}\Big) \quad \forall i = 1, \ldots, n-1 \quad \text{et} \quad \phi_n(x) = \max\Big(\phi\Big(\frac{x}{h}\Big), \phi\Big(\frac{x-1}{h}\Big)\Big)$$

(a) Représenter graphiquement les fonctions ϕ_i, $i = 1, \ldots, n$ et montrer que $(\phi_1, \ldots, \phi_n)$ forme une base de l'espace V_h.

(b) Vérifier que la méthode d'éléments finis ainsi choisie est une méthode conforme.

(c) Montrer que le problème faible approché qui consiste à trouver $u_h \in V_h$ solution de

$$a(u_h, v_h) = T(v_h) \quad \forall v_h \in V_h \tag{4.39}$$

est équivalent à déterminer $U_h \in \mathbb{R}^n$ tel que

$$A_h U_h = b_h$$

où A_h est une matrice $n \times n$ et $b_h \in \mathbb{R}^n$; donner l'expression de b_h et des coefficients de A_h en fonction de f et p.

4. Montrer qu'il existe $C > 0$ indépendant de u et u_h tel que

$$\|u - u_h\|_{H^1(\Omega)} \leqslant C \inf_{v_h \in V_h} \|u - v_h\|_{H^1(\Omega)}.$$

5. On définit l'opérateur d'interpolation $r_h : V \to V_h$ qui à tout $v \in V$ associe l'élément $r_h(v) \in V_h$ défini par $r_h(v)(x_i) = v(x_i)$ pour tout $i \in \{0, \dots, n-1\}$.

(a) Vérifier que r_h est défini de manière unique et montrer qu'il existe $\widetilde{C} \in \mathbb{R}_+$ ne dépendant ni de v ni de h tel que pour tout $v \in V \cap H^2(\Omega)$,

$$\|v - r_h(v)\|_{H^1(\Omega)} \leqslant \widetilde{C} h \|v\|_{H^2(\Omega)}.$$

(b) En déduire que u_h converge vers u lorsque h tend vers 0.

Exercice 55 — Éléments finis pour un problème avec conditions mixtes. Soit $f \in L^2(]0\,;1[$. On s'intéresse ici au problème

$$\begin{cases} -u''(x) + u(x) = f(x) \quad x \in=]0\,;1[\\ u(0) = 0 \\ u'(1) = 0 \end{cases} \tag{4.40}$$

Ce problème est un cas particulier du problème (3.54) étudié à l'exercice 46 en prenant $\Omega =]0\,;1[$, $p \equiv 1$, $q \equiv 1$, $\Gamma_0 = \{0\}$, $\Gamma_1 = \{1\}$, $g_0 \equiv 0$, $g_1 \equiv 0$ et $\sigma = 0$. On s'intéresse ici à la discrétisation du problème (4.40). Soient $N \in \mathbb{N}$, $h = 1/(N+1)$ et $x_i = ih$ pour $i = 0, \dots, N+1$ et $K_i = [x_i, x_{i+1}]$ pour $i = 0, \dots, N$. On cherche une solution approchée de (4.40), notée u_h, en utilisant les éléments finis $(K_i, \{x_i, x_{i+1}\}, \mathcal{P}_1)_{i=0}^N$.

1. Déterminer l'espace d'approximation V_h et montrer que les fonctions de base globales sont les fonctions Φ_i de $[0\,;1]$ dans $\mathbb{R}$ définies par $\Phi_i(x) = (1 - \frac{|x - x_i|}{h})^+$ pour $i = 1, \dots, N+1$.

2. Construire le système linéaire à résoudre et comparer avec les systèmes obtenus par différences finies et volumes finis.

3. A-t-on $u_h'(1) = 0$?

corrigé p.175

Exercice 56 — Éléments finis pour un problème de réaction-diffusion. Soit α un réel positif ou nul et f une fonction continue. On considère le problème suivant

$$\begin{cases} -u''(x) + \alpha u(x) = f(x) \quad x \in]0\,;1[\\ u'(0) = u(0) \\ u'(1) = 0 \end{cases} \tag{4.41}$$

où u'' désigne la dérivée seconde de u par rapport à x.

Dans toute la suite, on considère une subdivision uniforme de l'intervalle $[0\,;1]$: on note $h = 1/N$ où $N \geqslant 2$ est un entier fixé. On pose $x_i = ih$, pour $i = 0, \dots, N$.

Formulation variationnelle

1. Écrire une formulation variationnelle de (4.41).

2. On souhaite trouver $u \in H^1(]0\,;1[)$ tel que

$$\int_0^1 u'(x)v'(x)\,\mathrm{d}x + \alpha \int_0^1 u(x)v(x)\,\mathrm{d}x + u(0)v(0) = \int_0^1 f(x)v(x)\,\mathrm{d}x \qquad \forall v \in H^1(]0\,;1[) \tag{4.42}$$

(a) Déterminer le problème aux limites dont la formulation faible est (4.42).

(b) Montrer que si $\alpha > 0$, le problème (4.42) admet une unique solution.

Démontrer que ceci est encore vrai pour $\alpha = 0$ en appliquant l'inégalité de Poincaré à la fonction $v - v(0)$.

Discrétisation par éléments finis Soit V_h l'ensemble des fonctions continues sur $[0\,;1]$ dont la restriction à chaque intervalle $[x_i, x_{i+1}]$ est affine pour $0 \leqslant i \leqslant N-1$.

1. Quelle est la dimension de V_h ?

2. Donner la discrétisation éléments finis du problème (4.42).

3. Montrer que le problème discret ainsi obtenu admet une solution unique.

4. Écrire le problème discret sous la forme d'un système linéaire $AU = b$ en explicitant les dimensions des vecteurs et matrice et en donnant leur expression.

5. Donner une borne supérieure de l'erreur $\|u_h - u\|_{H^1(0,1)}$.

Exercice 57 — Un second membre irrégulier. Soit $\alpha \in C([0,1]) \cap C^1(]0,1[)$ une fonction telle que $\alpha(x) \geqslant \underline{\alpha} > 0$ pour tout $x \in]0,1[$. On considère le problème

$$(\alpha(x)u')'(x) = \frac{1}{x} \text{ sur }]0,1[, \tag{4.43}$$

$$u(0) = 0, u(1) = 0. \tag{4.44}$$

1. **Formulation faible** — On cherche une formulation faible du problème (4.43)-(4.44) dans $H_0^1(]0,1[)$.

   (a) Une première formulation faible — En utilisant la densité de $C_c^\infty(]0,1[)$ dans $H_0^1(]0,1[)$, montrer qu'une formulation faible possible est

   $$\begin{aligned} &u \in H_0^1(]0,1[), \\ &\int_0^1 \alpha(x)u'(x)v'(x)\,\mathrm{d}x = -\int_0^1 \ln x\ v'(x)\,\mathrm{d}x,\ \forall v \in H_0^1(]0,1[). \end{aligned} \tag{4.45}$$

   (b) Une deuxième formulation faible — On admet le résultat suivant (voir par exemple [1], exercice 6.15)

   Inégalité de Hardy Soit $p \in]1,\infty[$. On note L^p l'espace $L^p_{\mathbb{R}}(]0,\infty[)$, où $]0,\infty[$ est muni de la tribu des boréliens et de la mesure de Lebesgue.

   Soit $w \in L^p$. Pour $x \in]0,\infty[$, on pose $W(x) = \frac{1}{x}\int w\mathbb{1}_{]0,x[}\mathrm{d}\lambda$. Alors $W \in L^p$ et $\|W\|_{L^p} \leqslant \frac{p}{p-1}\|w\|_{L^p}$.

   i. Soit $\varphi \in H_0^1(]0,1[)$, montrer que la fonction $\Phi : x \mapsto \dfrac{\varphi(x)}{x}$ appartient à $L^2(]0,1[)$, et que

   $$\|\Phi\|_{L^2} \leqslant 2\|\varphi'\|_{L^2}.$$

   ii. En déduire qu'une formulation faible possible est

   $$\begin{aligned} &u \in H_0^1(]0,1[), \\ &\int_0^1 \alpha(x)u'(x)v'(x)\,\mathrm{d}x = \int_0^1 \frac{1}{x}v(x)\,\mathrm{d}x,\ \forall v \in H_0^1(]0,1[). \end{aligned} \tag{4.46}$$

   (c) Montrer que les formulations faibles (4.45) et (4.46) sont équivalentes et qu'elles admettent une solution unique $u \in H_0^1$.

2. **Eléments finis**

(a) Exemple sur un maillage grossier — On se donne un maillage uniforme de $]0,1[$ de pas $h = 1/3$. Donner le système linéaire obtenu par discrétisation par éléments finis de Lagrange P1 des formulations faibles (4.45) et (4.46).

(b) Analyse d'erreur — On se donne maintenant un maillage uniforme de $]0,1[$ de pas $h = \dfrac{1}{n+1}$. On note $x_i = ih$, $i = 1, \dots, n$ les noeuds du maillage et on pose $x_0 = 0$ et $x_{n+1} = 0$. On note V_h l'espace des fonctions continues, affines par mailles et nulles en 0 et en 1. Soit u_h la solution du problème approché par éléments finis P1 du problème (4.45). On note u_I l'interpolée de la solution exacte u dans V_h, c'est-à-dire la fonction affine par maille telle que $u_I(x_i) = u(x_i)$ pour $i = 0, \dots, n+1$.

i. Montrer qu'il existe $C \in \mathbb{R}_+$ ne dépendant que de α tel que

$$\|u - u_h\|_{H^1} \leqslant C\|u_I - u_h\|_{H^1}.$$

Dans les questions suivantes, on suppose que $\alpha(x) = 1$ pour tout $x \in [0,1]$.

ii. Calculer explicitement la solution u de (4.43)-(4.44).

iii. Montrer que

$$\int_h^1 |u'(x) - u_I'(x)|^2 \,\mathrm{d}x \leqslant h. \tag{4.47}$$

iv. Montrer qu'il existe $\beta > 0$ tel que

$$\int_0^h |u'(x) - u_I'(x)|^2 \,\mathrm{d}x = \beta h.$$

v. En déduire une estimation d'erreur entre solution exacte et approchée en norme $H^1(]0,1[)$.

vi. En s'inspirant de la technique d'Aubin-Nitsche vu en cours (théorème 4.3), obtenir une meilleure estimation de l'erreur en norme $L^2(]0,1[)$.

Exercice 58 — Sur la condensation de la masse (ou "mass lumping") en éléments finis. Soit $f \in L^2(]0,1[)$ (on rappelle que $L^2(]0,1[)$ désigne l'espace des (classes de) fonctions de carré intégrable sur $]0,1[$). On note $H^1(]0,1[) = \{v \in L^2(]0,1[) : v' \in L^2(]0,1[)\}$, où v' désigne la dérivée faible de v. On rappelle qu'une fonction $v \in L^2(]0,1[)$ admet une dérivée faible v' dans $L^2(]0,1[)$ s'il existe $w \in L^2(]0,1[)$ telle que

$$\int_0^1 v(x)\varphi'(x)\ \,\mathrm{d}x = -\int_0^1 w(x)\varphi(x)\ \,\mathrm{d}x. \tag{4.48}$$

pour toute fonction $\varphi \in C_c^1(]0,1[)$ où $C_c^1(]0,1[)$ désigne l'espace des fonctions de classe C^1 à support compact dans $]0,1[$. On note alors $v' = w$ la dérivée faible. On note $H^2(]0,1[) = \{v \in L^2(]0,1[), v' \in L^2(]0,1[) \text{ et } v'' \in L^2(]0,1[)\}$. On note $\|\cdot\|_{L^2}$ la norme $L^2(]0,1[)$ et $\|\cdot\|_{H^1}$ la norme H^1 définie par $\|v\|_{H^1}^2 = \|v\|_{L^2}^2 + \|v'\|_{L^2}^2$. Si $k \in \mathbb{N}$, on note $[\![1,k]\!]$ l'ensemble $\{1,2,\dots,k\}$.

On considère le problème suivant :

$$-u'' + u = f \quad \text{dans }]0,1[, \tag{4.49a}$$

$$u(0) = 0, \quad u(1) = 0. \tag{4.49b}$$

1. *Formulation faible.*

(a) Montrer que le problème (4.49) admet comme formulation faible :

$$\text{Trouver } u \in H_0^1(]0,1[) \ ; \ \forall v \in H_0^1(]0,1[),\ a(u,v) = \int_0^1 f(x)\ v(x)\ \,\mathrm{d}x. \tag{4.50}$$

avec $a(v,w) = \displaystyle\int_0^1 (v'(x)w'(x) + v(x)w(x))\,dx$ pour tout $(v,w) \in H_0^1(]0,1[)^2$ où $H_0^1(]0,1[) = \{v \in H^1(]0,1[);\ v(0) = v(1) = 0\}$.

(b) Montrer que le problème (4.50) admet une solution unique u. On admettra que $u \in H^2(]0,1[)$.

2. *Approximation par éléments finis* P1. On se donne un maillage uniforme constitué des points $x_i = ih$, $i \in [\![1, N]\!]$, avec $h = \dfrac{1}{N+1}$ et on considère l'approximation par éléments finis P1 du problème (4.50). On note $\phi_1, \ldots, \phi_N$ les fonctions de forme des éléments finis P1 associées aux noeuds $x_1, \ldots, x_N$.

(a) Rappeler l'expression des fonctions de forme $\phi_1, \ldots \phi_N$.

(b) Soient $v \in H^1_0(]0,1[) \cap H^2(]0,1[)$ et $v_I = \displaystyle\sum_{i=1}^{N} v(x_i)\phi_i$. Montrer qu'il existe $c_I \in \mathbb{R}_+$ ne dépendant que de v tel que

$$\|v - v_I\|_{H^1} \leqslant c_I h.$$

(c) Montrer que l'approximation par éléments finis P1 du problème (4.50) s'écrit

$$\text{Trouver } u_h \in V_h, \ ; \forall v \in V_h, \ a(u_h, v) = \int_0^1 f(x)\ v(x)\ \mathrm{d}x. \tag{4.51}$$

où V_h est un sous espace de $H^1_0(]0,1[)$ que l'on précisera.

(d) Montrer qu'il existe une unique solution au problème (4.51).

(e) Montrer que le problème (4.51) peut s'écrire comme un système linéaire $(A+M)U = B$, où A et M sont des matrices $N \times N$ correspondant respectivement à la discrétisation de $-u''$ et de u dans le problème (4.49), et $B \in \mathbb{R}^N$ (on explicitera les coefficients de ces matrices et de ce vecteur).

(f) Soit u la solution de (4.50) et u_h la solution de (4.51).

i. Montrer que $a(u-u_h, u-u_h) = a(u-u_h, u-v)$ pour tout $v \in V_h$ et en déduire qu'il existe $c_E \in \mathbb{R}_+$ ne dépendant que de u tel que $\|u-u_h\|_{H^1} \leqslant c_E h$.

ii. On note $e_h = u_h - u$. Soit $w \in H^1_0(]0,1[) \cap H^2(]0,1[)$ l'unique solution de

$$a(w,v) = \int_0^1 e_h(x)v(x)\ \mathrm{d}x,\ \forall v \in H^1_0(]0,1[),$$

et $w_h \in V_h$ l'unique solution de

$$a(w,v) = \int_0^1 e_h(x)v(x)\ \mathrm{d}x,\ \forall v \in V_h.$$

Montrer que

$$a(w_h - w, u_h - u) = \int_0^1 |e_h(x)|^2\ \mathrm{d}x.$$

iii. En déduire qu'il existe $c_S \in \mathbb{R}_+$ ne dépendant que de u tel que $\|u-u_h\|_{L^2} \leqslant c_S h^2$.

3. *Condensation de masse ou "mass lumping".* À partir des points x_i définis dans la question précédente, on définit pour $i \in [\![1, N]\!]$, les cellules $C_i =]x_i - \frac{h}{2}, x_i + \frac{h}{2}[$, et on note χ_i la fonction caractéristique associée à la cellule C_i, c.à.d. $\chi_i(x) = 1$ si $x \in C_i$ et $\chi_i(x) = 0$ sinon. Pour $v \in V_h$, on note $\Pi_h v$ la fonction constante par morceaux de $[0,1]$ dans $\mathbb{R}$, définie par $\Pi_h v = \sum_{i=1}^N v_i \chi_i$, où $v_i = v(x_i)$. On définit $a_h : V_h \times V_h \to \mathbb{R}$ par

$$a_h(v,w) = \int_0^1 v'(x)w'(x) + \Pi_h v(x)\Pi_h w(x)\ \mathrm{d}x,$$

et on considère maintenant le problème suivant.

$$\text{Trouver } \tilde{u}_h \in V_h\ ;\ \forall v \in V_h,\ a_h(\tilde{u}_h, v) = \int_0^1 f(x)\ v(x)\ \mathrm{d}x. \tag{4.52}$$

(a) Montrer que si $v \in V_h$, $\int_0^1 v(x)\ \mathrm{d}x = \int_0^1 \Pi_h v(x)\ dx$.

(b) Montrer que le problème (4.52) admet une solution unique $\tilde{u}_h \in V_h$ qui vérifie $\|\tilde{u}_h\|_{L^2} \leqslant c_M \|f\|_{L^2}$, où $c_M \in \mathbb{R}_+$ ne dépend pas de h.

(c) Montrer que le problème (4.52) peut s'écrire comme un système linéaire $(A + \tilde{M})\tilde{U} = b$, où A et $\tilde{M}$ sont des matrices $N \times N$ correspondant respectivement à la discrétisation effectuée dans (4.52) de $-u''$ et de u dans le problème (4.49), et $B \in \mathbb{R}^N$ (on explicitera les coefficients de ces matrices et de ce vecteur).

(d) On note $(m_{i,j})_{\substack{i\in[\![1,N]\!]\\ j\in[\![1,N]\!]}}$ et $(\tilde{m}_{i,j})_{\substack{i\in[\![1,N]\!]\\ j\in[\![1,N]\!]}}$ les coefficients des matrices M et $\tilde{M}$ pour tout $i \in [\![1, N]\!]$. Montrer que $\tilde{m}_{i,i} = \sum_{k=1}^{N} m_{i,k}$ pour tout $i \in [\![2, N-1]\!]$.

(e) Montrer que si $f \geqslant 0$, alors la solution $\tilde{U}$ du système linéaire $(A + \tilde{M})\tilde{U} = b$ vérifie $\tilde{U} \geqslant 0$, où l'inégalité s'entend composante par composante. Noter que cette propriété de positivité n'est pas vraie pour la méthode EF P1 sans mass lumping, voir à ce propos l'exercice 59.

(f) Pour $i \in [\![1, N]\!]$, on définit les "demi-cellules" C_i^+ et C_i^- par $C_i^+ =]x_i, x_i + \frac{h}{2}[$ et $C_i^- =]x_i - \frac{h}{2}, x_i[$. Soit $v \in V_h$; on pose $v_0 = v_{N+1} = 0$.

i. Montrer que

$$\Pi_h v(x) - v(x) = (v_i - v_{i+1})\phi_{i+1}(x),\ \forall x \in C_i^+, \forall i = 1, \dots, N,$$
$$\Pi_h v(x) - v(x) = (v_i - v_{i-1})\phi_{i-1}(x),\ \forall x \in C_i^-, \forall i = 1, \dots, N.$$

ii. En déduire qu'il existe $c_\Pi \in \mathbb{R}_+$ tel que $\|\Pi_h v - v\|_{L^2} \leqslant c_\Pi h \|v'\|_2$.

(g) Soit u la solution de (4.50), u_h la solution de (4.51) et $\tilde{u}_h$ la solution de (4.52).

i. Montrer qu'il existe $\tilde{c}_\Pi \in \mathbb{R}_+$ ne dépendant que de u tel que pour tout $v \in H^1$.

$$\left| \int_0^1 (\Pi_h \tilde{u}_h(x) \Pi_h v(x) - \tilde{u}_h(x) v(x)) \ \mathrm{d}x \right| \leqslant \tilde{c}_\Pi \|\tilde{u}_h\|_{H^1} \|v\|_{L^2}. \tag{4.53}$$

ii. En déduire qu'il existe $\tilde{c}_E$ ne dépendant que de u tel que $\|u - \tilde{u}_h\|_{L^2} \leqslant \tilde{c}_E h$.

(h) Montrer que si $f \geqslant 0$, alors la solution u de (4.50) est telle que $u \geqslant 0$.

4. *Volumes finis.* Proposer un maillage et un schéma volumes finis qui donne la même matrice $A + \tilde{M}$ que celle du système linéaire obtenu à partir de la formulation (4.52).

Exercice 59 — Pas de positivité sans mass lumping. Dans cet exercice, on propose un contre exemple pour montrer que les éléments finis P1 sans condensation de masse (ou mass lumping) ne respectent pas forcément le principe de positivité.

Soit $\varepsilon > 0$; on définit $a : H^1(]0,1[) \times H^1(]0,1[)$ par

$$a(u,v) = \int_0^1 u(x)v(x)\ \mathrm{d}x + \varepsilon \int_0^1 u'(x)v'(x)\ \mathrm{d}x,\ \forall (u,v) \in H^1(]0,1[)^2.$$

Soit $f \in C(]0,1[)$. On considère le problème

$$\text{Trouver } u \in H_0^1(]0,1[)\ ;\ \forall v \in H_0^1(]0,1[),\ a(u,v) = \int_0^1 f(x)\ v(x)\ \mathrm{d}x. \tag{4.54}$$

où $H_0^1(]0,1[) = \{v \in H^1(]0,1[);\ v(0) = v(1) = 0\}$.

1. Montrer que si $u \in C^2(]0,1[)$ vérifie (4.54), alors u est solution d'un problème aux limites qu'on explicitera.
2. Calculer la matrice A du système linéaire obtenu par l'approximation éléments finis P1 du problème (4.54) sur maillage uniforme.
3. Montrer que la matrice A est inversible ; on note $B = (b_{i,j}, i \in [\![1,N]\!], j \in [\![1,N]\!])$ son inverse. Montrer que $b_{i,i} > 0$.
4. On suppose dans cette question que le maillage est suffisamment grossier pour que la condition $\dfrac{h}{6} > \dfrac{\varepsilon}{h}$ soit vérifiée. En écrivant l'égalité $AB = \mathrm{Id}$, montrer que la matrice A n'est pas d'inverse à coefficients positifs.

5. En déduire que l'approximation par éléments finis P1 ne respecte pas le principe de positivité.
6. En quoi ce point peut-il être gênant ? Que peut-on faire pour obtenir la positivité ?

corrigé p.177

Exercice 60 — Éléments finis Q1. On considère le rectangle Ω de sommets $(-1,0)$, $(2,0)$, $(-1,1)$ et $(2,1)$. On s'intéresse à la discrétisation par éléments finis de l'espace fonctionnel $H^1(\Omega)$.

1. On choisit de découper Ω en deux éléments e_1 et e_2 définis par les quadrilatères de sommets respectifs $M_1(-1,1)$, $M_2(0,1)$, $M_5(1,0)$, $M_4(-1,0)$ et $M_2(0,1)$, $M_3(2,1)$, $M_6(2,0)$, $M_5(1,0)$. On prend comme nœuds les points $M_1,\dots,M_6$ et comme espace par élément l'ensemble des polynômes $\mathcal{Q}_1$. On note $\Sigma_1=\{M_4,M_5,M_2,M_1\}$ et $\Sigma_2=\{M_5,M_6,M_3,M_2\}$. On a construit la discrétisation $\{(e_1,\Sigma_1,\mathcal{Q}_1),(e_2,\Sigma_2,\mathcal{Q}_1)\}$.

(a) Montrer que les éléments $(e_1,\Sigma_1,\mathcal{Q}_1)$ et $(e_2,\Sigma_2,\mathcal{Q}_1)$ sont des éléments finis de Lagrange.

(b) Montrer que l'espace de dimension finie correspondant à cette discrétisation n'est pas inclus dans $H^1(\Omega)$ (construire une fonction de cet espace dont la dérivée distribution n'est pas dans L^2). Quelle est dans les hypothèses appelées en cours "cohérence globale" celle qui n'est pas vérifiée ?

2. On fait le même choix des éléments et des nœuds que dans la question 1. On introduit comme élément de référence e le carré de sommets $(\pm1,\pm1)$, Σ est l'ensemble des sommets de e et $\mathcal{P}=\mathcal{Q}_1$.

(a) Quelles sont les fonctions de base locales de (e,Σ,P). On note ces fonctions $\Phi_1,\dots,\Phi_4$.

(b) À partir des fonctions $\Phi_1,\dots,\Phi_4$, construire des bijections F_1 et F_2 de e dans e_1 et e_2. Les fonctions F_1 et F_2 sont elles affines ?

(c) On note $\mathcal{P}_{e_i}=\{f:e_i\to\mathbb{R}, f\circ F_i|_e\in\mathcal{Q}_1\}$ pour $i=1,2$ où les F_i sont définies à la question précédente. Montrer que les éléments $(e_1,\Sigma_1,\mathcal{P}_{e_1})$ et $(e_2,\Sigma_2,\mathcal{P}_{e_2})$ sont des éléments finis de Lagrange et que l'espace vectoriel construit avec la discrétisation $\{(e_1,\Sigma_1,\mathcal{P}_{e_1}),(e_2,\Sigma_2,\mathcal{P}_{e_2})\}$ est inclus dans $H^1(\Omega)$ (i.e. vérifier la "cohérence globale" définie en cours). On pourra pour cela montrer que si $S=e_1\cap e_2=\{(x,y);x+y=1\}$ alors $\{f|_S, f\in\mathcal{P}_{e_i}\}=\{f:S\to\mathbb{R}; f(x,y)=a+by, a,b\in\mathbb{R}\}$.

corrigé p.180

Exercice 61 — Éléments affine–équivalents. Soit Ω un ouvert polygonal de $\mathbb{R}^2$ et $\mathcal{T}$ un maillage de Ω. Soient $(\bar K,\bar\Sigma,\overline{\mathcal{P}})$ et $(K,\Sigma,\mathcal{P})$ deux éléments finis de Lagrange affine - équivalents. On suppose que les fonctions de base locales de $\bar K$ sont affines. Montrer que toute fonction de $\mathcal{P}$ est affine. En déduire que les fonctions de base locales de $(K,\Sigma,\mathcal{P})$ affines.

Exercice 62 — Assemblage de la matrice éléments finis. Soit $f\in L^2(\Omega)$, et soit Λ l'opérateur linéaire autoadjoint de matrice $(\lambda_{ij})_{i=1,2,j=1,2}$ symétrique définie positive dans la base canonique de $\mathbb{R}^2$, notée $(\vec v_1,\vec v_2)$. On considère le problème suivant, posé dans $\Omega=]0,1[\times]0,1[$:

Trouver $u\in H^1_0(\Omega)$ telle que

$$\forall v\in H^1_0(\Omega),\ \int_\Omega \Lambda\nabla u(x)\cdot\nabla v(x)\,\mathrm{d}x=\int_\Omega f(x)v(x)\,\mathrm{d}x,$$

où $f\in L^2(\Omega)$ est donnée, et Λ est l'opérateur linéaire autoadjoint de matrice $(\lambda_{ij})_{i=1,2,j=1,2}$ symétrique définie positive dans la base canonique de $\mathbb{R}^2$, notée $(\vec v_1,\vec v_2)$.

1. Le problème continu.
   a) Expliquer brièvement pourquoi ce problème est bien posé.
   b) Lorsqu'il possède une solution de classe C^2, de quel problème fort celle-ci est-elle également solution ?

On choisit, pour $n\in\mathbb{N}_*$, la discrétisation suivante : pour tout $i=1,2,...n$ et $j=1,2,...n$,

K_{ij1} est le triangle de sommets $\dfrac{i-1}{n}\vec v_1+\dfrac{j-1}{n}\vec v_2$, $\dfrac{i}{n}\vec v_1+\dfrac{j-1}{n}\vec v_2$, $\dfrac{i-1}{n}\vec v_1+\dfrac{j}{n}\vec v_2$

K_{ij2} est le triangle de sommets $\frac{i-1}{n}\vec{v}_1 + \frac{j}{n}\vec{v}_2$, $\frac{i}{n}\vec{v}_1 + \frac{j-1}{n}\vec{v}_2$, $\frac{i}{n}\vec{v}_1 + \frac{j}{n}\vec{v}_2$.

On recherche une solution approchée du problème par la méthode des éléments finis, constitués par :

— les $2n^2$ triangles K_{ij1} et K_{ij2},
— les degrés de liberté égaux aux valeurs aux sommets des triangles,
— l'espace vectoriel $\mathcal{P}^1(\mathbb{R}^2)$.

2. Donner l'expression des fonctions de forme locales sur le triangle de référence de sommets $(0,0),(h,0),(0,h)$, avec $h = \frac{1}{n}$.
3. Donner les matrices élémentaires d'assemblage des triangles K_{ij1} et K_{ij2}.
4. Donner l'expression de la matrice assemblée.

Exercice 63 — Éléments finis P2 en une dimension d'espace. On veut résoudre numériquement le problème aux limites suivant

$$\begin{cases} -u''(x) + u(x) = x^2 & 0 < x < 1 \\ u(0) = 0 \\ u'(1) = 1 \end{cases} \tag{4.55}$$

1. Donner une formulation faible du problème (4.55) ;
2. Démontrer que le problème (4.55) admet une unique solution ;
3. On partage l'intervalle $]0\,;1[$ en N intervalles égaux et on approche la solution par une méthode d'éléments finis de degré 2. écrire le système qu'il faut résoudre.

corrigé p.186

Exercice 64 — Éléments finis P1 sur maillage triangulaire. On veut résoudre numériquement le problème suivant :

$$\begin{cases} -\Delta u(x,y) = f(x,y) & (x,y) \in D = (0,a)\times(0,b) \\ u(x,y) = 0 & (x,y) \in \partial D \end{cases}$$

où f est une fonction donnée, appartenant à $L^2(D)$. Soient M, N deux entiers. On définit

$$\Delta x = \frac{a}{M+1} \qquad \Delta y = \frac{b}{N+1}$$

et on pose $x_k = k\Delta x$, $0 \leqslant k \leqslant M+1$, $y_\ell = \ell\Delta y$, $0 \leqslant \ell \leqslant N+1$. On note $T^0_{k+1/2,\ell+1/2}$ le triangle de sommets $(x_k, y_\ell), (x_{k+1}, y_\ell), (x_{k+1}, y_{\ell+1})$ et $T^1_{k+1/2,\ell+1/2}$, le triangle de sommets (x_k, y_ℓ), $(x_k, y_{\ell+1})$, (x_{k+1}, y_{v+1}). Écrire la matrice obtenue en discrétisant le problème avec les éléments finis triangulaires linéaires (utilisant le maillage précédent).

corrigé p.188

Exercice 65 — Intégration numérique.

1. Vérifier que l'intégration numérique à un point de Gauss, situé au centre de gravité du triangle, sur l'élément fini P1 sur triangle, est exacte pour les polynômes d'ordre 1.
2. Vérifier que l'intégration numérique à trois points de Gauss définis sur le triangle de référence par $p_1 = (1/2, 0)$, $p_2 = (1/2, 1/2)$, $p_3 = (0, 1/2)$ avec les poids d'intégration $\omega_1 = \omega_2 = \omega_3 = 1/6$, est exacte pour les polynômes d'ordre 2.

corrigé p.189

Exercice 66 — Éléments finis Q2. On note C le carré $[0\,;1]\times[0\,;1]$ de sommets $a_1 = (0,0)$, $a_2 = (1,0)$, $a_3 = (1,1)$, $a_4 = (0,1)$. On note $a_5 = (\frac{1}{2}, 0)$, $a_6 = (1, \frac{1}{2})$, $a_7 = (\frac{1}{2}, 1)$, $a_8 = (0, \frac{1}{2})$, $a_9 = (\frac{1}{2}, \frac{1}{2})$ et $\Sigma = \{a_i, 1 \leqslant i \leqslant 8\}$.

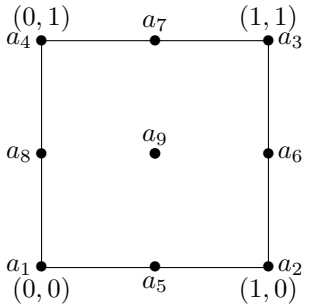


FIGURE 4.10 – Le carré unité et les sommets $a_1, \ldots, a_9$.

1. Montrer que pour tout $p \in \mathcal{P}_2$

$$\sum_{i=1}^{4} p(a_i) - 2\sum_{i=5}^{8} p(a_i) + 4) = 0.$$

2. En déduire une forme linéaire ϕ telle que si $p \in \mathcal{P} = \{p \in \mathcal{Q}_2, \phi(p) = 0\}$ et $p(a_i) = 0$ pour $i = 1, \ldots 8$, alors $p = 0$.

3. Calculer les fonctions de base de l'élément fini $(C, \Sigma, \mathcal{P})$, avec $\Sigma = \{a_1, \ldots, a_8\}$.

Exercice 67 — Éléments finis Q2*. Soit $C = [-1\,;1] \times [-1\,;1]$. On note $a_1, \ldots, a_8$ les nœuds de C définis par $a_1 = (-1, -1)$, $a_2 = (1, -1)$, $a_3 = (1, 1)$, $a_4 = (-1, 1)$, $a_5 = (0, -1)$, $a_6 = (1, 0)$, $a_7 = (0, 1)$ et $a_8 = (-1, 0)$. On rappelle que $\mathcal{Q}_2 = Vect\{1, x, y, xy, x^2, y^2, x^2y, xy^2, x^2y^2\}$ et que $\dim \mathcal{Q}_2 = 9$. On note $\mathcal{Q}_2^*$ l'espace de polynôme engendré par les fonctions $\{1, x, y, xy, x^2, y^2, x^2y, xy^2\}$.

1. Construire $(\varphi_i^*)_{i=1,\ldots,8} \subset Q_2^*$ tel que $\varphi_j^*(a_i) = \delta_{ij}$, $\forall i, j = 1, \ldots, 8$.

2. Montrer que l'ensemble $\Sigma = \{a_1, \ldots, a_8\}$ est $\mathcal{Q}_2^*$-unisolvant.

3. Soit $S = [-1\,;1] \times \{1\}$, $\Sigma_S = \Sigma \cap S$ et soit $\mathcal{P}$ l'ensemble des restrictions à S des fonctions de $\mathcal{Q}_2^*$, *i.e.* $\mathcal{P} = \{f|_S; f \in \mathcal{Q}_2^*\}$. Montrer que Σ_S est $\mathcal{P}$-unisolvant. La propriété est-elle vraie pour les autres arêtes de C ?

Exercice 68 — Discrétisation du bi-laplacien. La modélisation d'une poutre en charge encastrée à ses deux extrémités amène à s'intéresser au problème d'ordre 4 suivant (dit problème "biharmonique") :

$$u^{(4)}(x) = f(x),\ x \in]0\,;1[\tag{4.56a}$$

$$u(0) = 0, \quad u'(0) = 0, \qquad u(1) = 0, \quad u'(1) = 0. \tag{4.56b}$$

où $u^{(4)}$ désigne la dérivée quatrième de u par rapport à x, et f est une fonction continue.

Problème continu

1. On suppose (dans cette question seulement) que $f \equiv 1$. Calculer la solution exacte $\overline{u}$ de (4.56), et la représenter graphiquement (grossièrement).

2. Soit $H^2(]0\,;1[)$ l'ensemble des fonctions de carré intégrable dont les dérivées (faibles) première et seconde sont également de carré intégrable :

$$H^2(]0\,;1[) = \{u \in L^2(]0\,;1[), u' \in L^2(]0\,;1[) \text{ et } u'' \in L^2(]0\,;1[).\}$$

On rappelle également que les fonctions de $H^1(]0\,;1[)$ sont continues sur $[0\,;1]$.

(a) Montrer que $H^2(]0\,;1[) \subset C^1([0\,;1])$.

On définit alors :

$$H_0^2(]0\,;1[) = \{u \in H^2(]0\,;1[); u(0) = u(1) = 0,\ u'(0) = u'(1) = 0\}.$$

(b) Montrer que si $u \in C^4([0\,;1])$ est solution de (4.56), alors u est solution de :

$$\begin{aligned} &u \in H_0^2(]0\,;1[) \\ &\int_0^1 u''(x)v''(x)\ dx = \int_0^1 f(x)v(x)\ \mathrm{d}x,\ \forall v \in H_0^2(]0\,;1[, \end{aligned} \tag{4.57}$$

(c) Montrer que réciproquement, si u est solution de (4.57) et $u \in C^4([0\,;1])$, alors u est solution de (4.56).
On admettra pour la suite que le problème (4.57) admet une solution unique.

On cherche maintenant à trouver une solution approchée de la solution de (4.56) ou (4.57).

Discrétisation par différences finies

3. Soit $M > 2$ et $h = \frac{1}{M+1}$. On construit une subdivision de $[0\,;1]$, notée $(y_k)_{k=0,\ldots,M+1}$, définie par : $y_i = ih$ pour $i = 0, \ldots, M+1$. On note u_i l'inconnue discrète associée au point y_i, $i = 0, \ldots, M+1$.

(a) Soient $\varphi \in C^5(\mathbb{R})$, $x \in \mathbb{R}$ et $h > 0$, écrire les développements limités en x de $\varphi(x+2h)$, $\varphi(x+h)$,$\varphi(x-h)$, et $\varphi(x-2h)$ à l'ordre 5 en h.

(b) Par une combinaison linéaire adéquate, en déduire pour $i = 2, \ldots, M-1$, une approximation par différences finies de $u^{(4)}(y_i)$ en fonction de u_{i-2}, u_{i-1}, u_i, u_{i+1} et u_{i+2} qui soit consistante d'ordre 2.

(c) Ecrire un schéma de différences finies consistant pour la discrétisation de (4.56) sous la forme

$$\begin{aligned} &(\delta^{(4)}u)_i = f(y_i),\ i = 2, \ldots, M-1, \\ &u_0 = u_1 = 0, \\ &u_M = u_{M+1} = 0, \end{aligned}$$

où $(\delta^{(4)}u)_i$ est l'approximation consistante de $u^{(4)}(y_i)$ construite avec les inconnues discrètes u_{i-2}, u_{i-1}, u_i, u_{i+1} et u_{i+2} à la question 3.2.
Ecrire le schéma sous forme matricielle.

(d) Soit $\delta^{(4)}u \in \mathbb{R}^{M-2}$ dont les composantes sont les valeurs $(\delta^{(4)}u)_i$ pour $i = 2, \ldots, M-1$. Notons $(\delta^{(2)}u)_i = \frac{1}{h}(u_{i+1} + u_{i-1} - 2u_i)$ la discrétisation habituelle de $u''(y_i)$ par différences finies. A-t'on $(\delta^{(4)}u)_i = (\delta^{(2)}(\delta^{(2)}u))_i$ pour tout $i = 2, \ldots, M-1$?

Dans toute la suite, on considère le maillage suivant : pour $N > 2$ et $h = 1/N$, on définit les N mailles $(K_i)_{i=1,\ldots,N}$ par $K_i =]x_{i-1/2}, x_{i+1/2}[$, avec $x_{i+1/2} = ih$ pour $i = 0, \ldots, N$, et on note $x_i = (i-1/2)h$ pour $i = 1, \ldots, N$ les centres des mailles. On pose également $x_0 = 0$ et $x_{N+1} = 1$. On définit également des mailles "décalées" $K_{i+1/2} =]x_i, x_{i+1}[$, pour $i = 0, \ldots, N$.

Discrétisation par un schéma volumes finis

4. Soit $(u_i)_{i=1,N} \in \mathbb{R}^N$ et u la fonction de $]0\,;1[$ dans $\mathbb{R}$, constante par morceaux, définie par

$$u(x) = u_i \text{ pour tout } x \in K_i =]x_{i-1/2}, x_{i+1/2}[.$$

On définit $D_h u$ comme la fonction constante par morceaux sur les mailles décalées, définie par

$$D_h u(x) = \begin{cases} D_{i+1/2}u = \frac{1}{h}(u_{i+1} - u_i) \text{ pour tout } x \in K_{i+1/2} =]x_i, x_{i+1}[\text{ pour } i = 1, \ldots, N-1, \\ D_{1/2}u = \frac{2}{h}u_1 \text{ pour tout } x \in K_{1/2} =]x_0, x_1[, \\ D_{N+1/2}u = -\frac{2}{h}u_N \text{ pour tout } x \in K_{N+1/2} =]x_N, x_{N+1}[. \end{cases}$$

Enfin on définit $D_h^2 u$ comme la fonction constante par morceaux sur les mailles K_i, définie par :

$$D_h^2 u(x) = D_i^2 u = \frac{1}{h}(D_{i+1/2}u - D_{i-1/2}u) \text{ pour tout } x \in K_i =]x_{i-1/2}, x_{i+1/2}[\text{ pour } i = 1, \ldots, N.$$

(a) Calculer $D_i^2 u$ en fonction des valeurs u_j, $j = 1, \dots, N$.

(b) En déduire, pour $k = 1, \dots, N$, l'expression de la fonction $D_h^2 \chi_k$, où $\chi_k : \mathbb{R} \to \mathbb{R}$ est la fonction caractéristique de la maille K_k, c.à.d :

$$\chi_k(x) = \begin{cases} 1 & \text{si } x \in K_k \\ 0 & \text{sinon.} \end{cases}$$

On note H_h l'espace des fonctions constantes sur les mailles K_i, $i = 1, \dots, N$, et $H_{h,0}$ les fonctions de H_h nulles sur les mailles 1 et N. On considère le schéma numérique défini par la forme faible discrète suivante :

$$\text{Trouver } u \in H_{h,0} = \{u \in H_h; u_1 = u_N = 0\}, \tag{4.58}$$

$$\int_0^1 D_h^2 u(x)\ D_h^2 v(x)\, \mathrm{d}x = \int f(x)\ v(x)\, \mathrm{d}x, \forall v \in H_{h,0}. \tag{4.59}$$

(c) En prenant les fonctions caractéristiques des mailles K_i comme fonctions tests dans (4.59), montrer que le schéma (4.58)-(4.59) s'écrit aussi :

$$u \in H_{h,0} \tag{4.60a}$$

$$F_{i+1/2}(D_h^2 u) - F_{i-1/2}(D_h^2 u) = \int_{K_i} f(x)\, \mathrm{d}x, i = 1, \dots, N, \tag{4.60b}$$

$$F_{i+1/2}(D_h^2 u) = \frac{1}{h}(D_{i+1}^2 u - D_i^2 u), i = 1, \dots, N-1, \tag{4.60c}$$

$$F_{1/2}(D_h^2 u) = -\frac{2}{h} D_1^2 u \text{ et } F_{N+1/2}(D_h^2 u) = -\frac{2}{h} D_N^2 u. \tag{4.60d}$$

Expliquer pourquoi ce schéma peut prétendre à l'appellation "volumes finis".

Quelques propriétés du schéma volumes finis On se place ici sous les hypothèses et notations de la discrétisation par volumes finis.

5. Existence et unicité de la solution discrète.

(a) Montrer que

$$\forall u \in H_{h,0},\ -\int_0^1 u(x) D_h^2 u(x)\, \mathrm{d}x = \int_0^1 D_h u(x) D_h u(x)\, \mathrm{d}x.$$

(b) Soit $u \in H_{h,0}$; montrer que si $D_{i+\frac{1}{2}} u = 0$ pour tout $i = 1, \dots, N$ alors $u \equiv 0$.

(c) En déduire que si $f = 0$, et si u est solution de (4.58)-(4.59) alors $u = 0$.

(d) En déduire l'existence et l'unicité de u solution de (4.58)-(4.59).

6. Stabilité.

(a) *(Poincaré discret sur u).* Soit $u \in H_h$. Montrer que $\|u\|_{L^2(\Omega)} \leqslant \|D_h u\|_{L^2(\Omega)}$.

(b) *(Poincaré discret sur $D_h u$).* Soit $u \in H_{h,0}$. Montrer que $\|D_h u\|_{L^2(\Omega)} \leqslant \|D_h^2 u\|_{L^2(\Omega)}$.

(c) *(Estimation a priori sur la solution).* Soit $u \in H_{h,0}$ solution de (4.58)-(4.59). Montrer que $\|u\|_{L^2(\Omega)} \leqslant \|f\|_{L^2(\Omega)}$.

Discrétisation par un schéma éléments finis non conformes

7. On considère maintenant les fonctions de forme ϕ_k des éléments finis P1 associées aux noeuds x_k, $k = 1, \dots, N$.

(a) Donner l'expression des fonctions de forme ϕ_k pour $k = 1, \dots, N$.

Soit V_h l'espace engendré par les fonctions $\phi_1, \dots, \phi_N$ et $V_{h,0}$ l'espace engendré par les fonctions $\phi_2, \dots, \phi_{N-1}$.

Pour $\widetilde{v} \in V_{h,0}$, on définit $\widetilde{D}_h^2 \widetilde{v}$ comme la fonction de H_h définie par :

$$\widetilde{D}_h^2 \widetilde{v}(x) = -\frac{1}{h} \sum_{i=2}^{N-1} \widetilde{v}(x_i) \int_0^1 \phi_i'(s) \phi_k'(s)\ ds, \text{ pour tout } x \in K_k,\ k = 1, \dots, N.$$

On considère alors le schéma suivant pour l'approximation de (4.57).

$$\text{Trouver } \widetilde{u} \in V_{h,0}, \tag{4.61}$$

$$\int_0^1 \widetilde{D}_h^2\widetilde{u}(x)\widetilde{D}_h^2\widetilde{v}(x)\,\mathrm{d}x = \int f(x)\ \widetilde{v}(x)\,\mathrm{d}x, \forall \widetilde{v} \in V_{h,0}. \tag{4.62}$$

(b) Expliquer pourquoi le schéma (4.61)-(4.62) peut prétendre à l'appellation "éléments finis non conformes".

(c) Soit $(u_i)_{i=2,\dots,N-1} \in \mathbb{R}^{N-2}$ et soient $u = \sum_{k=2}^{N-1} u_k\chi_k \in H_{h,0}$ et $\widetilde{u} = \sum_{k=2}^{N-1} u_k\phi_k \in V_{h,0}$.

i. Montrer que $\widetilde{u}'$ est une fonction constante par morceaux sur les mailles décalées et comparer $\widetilde{u}'$ à $D_h u$.

ii. Calculer $\widetilde{D}_h^2\widetilde{u}$ et comparer $\widetilde{D}_h^2\widetilde{u}$ à $D_h^2 u$.

Estimation d'erreur

8. On note $\overline{u}$ la solution exacte de (4.56), et on suppose maintenant que $\overline{u} \in C^4([0\,;1])$. On note $u_h \in H_h$ la solution de (4.58)-(4.59), $\overline{u}_h$ la fonction de H_h définie par : $\overline{u}_i = \overline{u}(x_i)$, $i = 0,\dots,N$, et par e_h la fonction de H_h définie par : $e_i = \overline{u}_i - u_i$.

(a) *Equation sur l'erreur* – Montrer que e_h est solution de l'équation suivante :

$$e_h \in H_{h,0}, \tag{4.63}$$

$$\int_0^1 D_h^2 e_h(x)\ D_h^2 v(x)\,\mathrm{d}x = \int R_h(x)v(x)\,\mathrm{d}x, \forall v \in H_{h,0}, \tag{4.64}$$

où R_h est une fonction de H_h dont on donnera l'expression.

(b) *Consistance* – Montrer que $\|R_h\|_{L^2(\Omega)} \leqslant Ch^2$ où $C \geqslant 0$ ne dépend que de $\overline{u}$.

(c) *Estimation d'erreur* – Montrer que $\|e_h\|_{L^2(\Omega)} \leqslant \|R_h\|_{L^2(\Omega)}$ et en déduire que le schéma est convergent d'ordre 2.

(d) *Maillage non uniforme* – Expliquer pourquoi la technique d'estimation d'erreur proposée précédemment n'est possible que dans le cas d'un pas h uniforme. Réfléchir à la manière dont on pourrait procéder pour un pas non uniforme (on peut dans ce cas obtenir une estimation d'erreur d'ordre h).

4.6.2 Corrigés

Exercice 51– Éléments finis P1 pour le problème de Dirichlet.

1. Soit $v \in H_N$. Comme $H_N \subset C([0\,;1])$, on a $v \in L^2(]0\,;1[)$. D'autre part, comme $v|_{K_i} \in \mathcal{P}_1$, on a $v|_{K_i}(x) = \alpha_i x + \beta_i$ avec $\alpha_i, \beta_i \in \mathbb{R}$. Donc v admet une dérivée faible dans $L^2(]0\,;1[)$ et $Dv|_{K_i} = \alpha_i$ et on a donc :

$$\|Dv\|_{L^2} \leqslant \max_{i=1,\dots,N} |\alpha_i| < +\infty.$$

De plus $v(0) = v(1) = 0$ donc $v \in H_0^1(]0\,;1[)$. On en déduit que $H_N \subset H_0^1(]0\,;1[)$.

2. On a :

$$\phi_i(x) = \begin{cases} 1 - (x - x_i)/h & \text{si } x \in K_i \\ 1 + (x - x_i)/h & \text{si } x \in K_{i-1} \\ 0 & \text{si } x \in]0\,;1[\ K_i \cup K_{i-1} \end{cases}$$

On en déduit que $\phi_i|_{K_j} \subset P_1$ pour tout $j = 0,\dots,N$. De plus, les fonctions ϕ_i sont clairement continues. Pour montrer que $\phi_i \in H_N$, il reste à montrer que $\phi_i(0) = \phi_i(1) = 0$. Ceci est immédiat pour $i = 2,\dots,N-1$, car dans ce cas $\phi_i|_{K_0} = \phi_i|_{K_{N+1}} = 0$. On vérifie alors facilement que $\phi_1(0) = 1 - h/h = 0$ et $\phi_N(1) = 0$. Pour montrer que $H_N = \mathrm{Vect}\{\phi_1,\dots,\phi_N\}$, il suffit de montrer que $\{\phi_1,\dots,\phi_N\}$ est une famille libre de H_N. En effet, si $\sum_{i=1}^N a_i\phi_i = 0$ alors en particulier $\sum_{i=1}^N a_i\phi_i(x_k) = 0$ pour $k = 1,\dots,N$ et donc $a_k = 0$ pour $k = 1,\dots,N$.

3. Soit $u = \sum_{j=1}^{N} u_j \phi_j$ solution de

$$a(u, \phi_i) = T(\phi_i) \qquad \forall i = 1, \ldots, N$$

La famille $(u_j)_{j=1,\ldots,N}$ est donc solution du système linéaire

$$\sum_{j=1}^{N} \mathcal{K}_{i,j} u_j = \mathcal{G}_i \qquad i = 1, \ldots, N$$

où $\mathcal{K}_{i,j} = a(\phi_j, \phi_i)$ et $\mathcal{G}_i = T(\phi_i)$. Calculons $\mathcal{K}_{i,j}$ et $\mathcal{G}_i$; on a :

$$K_{i,j} = \int_0^1 \phi_j'(x)\phi_i'(x)dx \quad \text{avec} \quad \phi_i'(x) = \begin{cases} 1/h & \text{si } x \in]x_{i-1}x_i[\\ -1/h & \text{si } x \in]x_i, x_{i+1}[\\ 0 & \text{ailleurs} \end{cases}$$

On en déduit que :

$$\begin{cases} \mathcal{K}_{i,i} = \displaystyle\int_0^1 (\phi_i'(x))^2 \, \mathrm{d}x = 2h\frac{1}{h^2} = \frac{2}{h} & \text{pour} \quad i = 1, \ldots, N, \\ \mathcal{K}_{i,i+1} - \displaystyle\int_0^1 \phi_i'(x)\phi_{i+1}'(x)dx = -h \times \frac{1}{h^2} = -\frac{1}{h} & \text{pour} \quad i = 1, \ldots, N-1 \\ \mathcal{K}_{i,i-1} = \displaystyle\int_0^1 \phi_i'(x)\phi_{i-1}'(x)dx = -\frac{1}{h} & \text{pour} \quad i = 2, \ldots, N \\ \mathcal{K}_{i,j} = 0 & \text{pour} \quad |i-j| > 1 \end{cases}$$

Calculons maintenant $\mathcal{G}_i$:

$$\mathcal{G}_i = \int_{x_i - 1}^{x_{i+1}} f(x)\phi_i(x)dx$$

Si f est constante, on a alors

$$\mathcal{G}_i = f \int_{x_{i-1}}^{x_{i+1}} \phi_i(x) \, \mathrm{d}x = hf$$

Si f n'est pas constante, on procède à une intégration numérique. On peut, par exemple, utiliser la formule des trapèzes pour le calcul des intégrales :

$$\int_{x_{i-1}}^{x_i} f(x)\phi_i(x)dx \quad \text{et} \quad \int_{x_i}^{x_{i+1}} f(x)\phi_i(x)dx$$

On obtient alors $\mathcal{G}_i = hf(x_i)$. Le schéma est donc :

$$\begin{cases} 2u_i - u_{i-1} - u_{i+1} = h^2 f(x_i) \quad i = 1, \ldots, N \\ u_0 = 0 \\ u_{N+1} = 0 \end{cases}$$

C'est exactement le schéma différences finis avec un pas constant h.

Exercice 52.

1. Soit $(x_i)_{i=1,\ldots,N+1}$ une discrétisation de l'intervalle $[0\,;1]$ avec $0 = x_0 < x_1 < \cdots x_i < x_{i+1} < x_N < x_{N+1} = 1$. Pour $i = 1, \ldots, N$, on pose $h_{i+1/2} = x_{i+1} - x_i$. L'équation (4.36) au point x_i s'écrit $-u''(x_i) + u(x_i) = f(x)$. On écrit les développements de Taylor de $u(x_{i+1})$ et $u(x_{i-1})$: il existe $\zeta_i \in [x_i, x_{i+1}]$ et $\theta_i \in [x_{i-1}, x_i]$ tels que

$$u(x_{i+1}) = u(x_i) + h_{i+1/2}u'(x_i) + \frac{1}{2}h_{i+1/2}^2 u''(x_i) + \frac{1}{6}h_{i+1/2}^3 u'''(\zeta_i), \tag{4.65}$$

$$u(x_{i-1}) = u(x_i) - h_{i-1/2}u'(x_i) + \frac{1}{2}h_{i-1/2}^2 u''(x_i) - \frac{1}{6}h_{i-1/2}^3 u'''(\theta_i). \tag{4.66}$$

En multipliant la première égalité par $h_{i-1/2}$, la deuxième par $h_{i+1/2}$ et en additionnant :

$$u''(x_i) = \frac{2\big(h_{i-1/2}u(x_{i+1}) + h_{i+1/2}u(x_{i-1}) + (h_{i+1/2}+h_{i-1/2})u(x_i)\big)}{h_{i+1/2}h_{i-1/2}(h_{i+1/2}+h_{i-1/2})} - \frac{1}{6}\frac{h_{i+1/2}^2}{h_{i+1/2}+h_{i-1/2}}u'''(\zeta_i) + \frac{1}{6}\frac{h_{i-1/2}^2}{h_{i+1/2}+h_{i-1/2}}u'''(\theta_i). \quad (4.67)$$

En posant

$$\gamma_i = \frac{2}{h_{i+1/2}h_{i-1/2}(h_{i+1/2}+h_{i-1/2})}$$

on déduit donc l'approximation aux différences finies suivante pour tous les nœuds internes :

$$\gamma_i\big(h_{i-1/2}u_{i+1} + h_{i+1/2}u_{i-1} + (h_{i+1/2}+h_{i-1/2})u_i\big) + u_i = f(x_i) \qquad i = 1,\dots,N.$$

La condition de Fourier en 0 se discrétise par

$$\frac{u_1 - u_0}{h_{1/2}} - u_0 = 0$$

et la condition de Neumann en 1 par :

$$\frac{u_{N+1} - u_N}{h_{N+1/2}} = -1$$

On obtient ainsi un système linéaire carré d'ordre $N+1$.

2. On prend maintenant une discrétisation volumes finis non uniforme ; on se donne $N \in \mathbb{N}^\star$ et $h_1,\dots,h_N > 0$ tel que $\sum_{i=1}^N h_i = 1$. On pose $x_{1/2} = 0$, $x_{i+1/2} = x_{i-1/2} + h_i$ pour $i = 1,\dots,N$ (de sorte que $x_{N+1/2} = 1$),

$$h_{i+1/2} = \frac{h_{i+1}+h_i}{2} \text{ pour } i = 1,\dots,N-1 \quad \text{et} \quad f_i = \frac{1}{h_i}\int_{x_{i-1/2}}^{x_{i+1/2}} f(x)dx \text{ pour } i = 1,\dots,N$$

En intégrant la première équation de (4.36) et en approchant les flux $u'(x_{i+1/2})$ par le flux numérique $F_{i+1/2}$, on obtient le schéma suivant :

$$F_{i+1/2} - F_{i-1/2} + h_i u_i = h_i f_i \qquad i \in \{1,\dots,N\}, \quad (4.68)$$

où $(F_{i+1/2})_{i\in\{0,\dots,N\}}$ donné en fonction des inconnues discrètes $(u_1,\dots,u_N)$ par les expressions suivantes, tenant compte des conditions aux limites :

$$F_{i+1/2} = -\frac{u_{i+1}-u_i}{h_{i+1/2}} \qquad i \in \{1,\dots,N-1\} \quad (4.69)$$

$$F_{1/2} = -\frac{u_1 - u_0}{x_1/2} \quad (4.70)$$

$$F_{1/2} - u_0 = 0 \quad (4.71)$$

$$F_{N+1/2} = -1 \quad (4.72)$$

Notons que u_0 peut être éliminé des équations (4.70) et(4.71). On obtient ainsi un système linéaire de N équations à N inconnues :

$$-\frac{u_2-u_1}{h_{3/2}} + \frac{u_1}{1-\frac{x_1}{2}} + h_1u_1 = h_1f_1 \quad (4.73)$$

$$-\frac{u_{i+1}-u_i}{h_{i+1/2}} + \frac{u_i-u_{i-1}}{h_{i-1/2}} + h_iu_i = h_if_i \qquad i \in \{2,\dots,N-1\} \quad (4.74)$$

$$-1 + \frac{u_N-u_{N-1}}{h_{N-1/2}} + h_Nu_N = h_Nf_N \quad (4.75)$$

3. Comme pour les différences finies, on se donne $(x_i)_{i=1,\dots,N+1}$ une discrétisation de l'intervalle $[0\,;1]$, avec $0 = x_0 < x_1 < \cdots x_i < x_{i+1} < x_N < x_{N+1} = 1$. Pour $i = 1,\dots,N$, on pose $h_{i+1/2} = x_{i+1} - x_i$ et $K_{i+1/2} = [x_i, x_{i+1}]$ pour $i = 0,\dots,N$. On définit l'espace d'approximation

$H_N = \{v \in C([0\,;1],\mathbb{R})$ telle que $v|_{K_{i+1/2}} \in \mathcal{P}_1, i = 0,\dots,N\}$ où $\mathcal{P}_1$ désigne l'ensemble des polynômes de degré inférieur ou égal à 1. Remarquons que l'on a bien $H_N \subset H$.
Pour $i = 1,\dots,N$, on pose :

$$\begin{cases} \phi_i(x) = \dfrac{x - x_{i-1}}{h_{i-1/2}} & \text{si } x \in K_{i-1/2} \\ \phi_i(x) = \dfrac{x_{i+1} - x}{h_{i+1/2}} & \text{si } x \in K_{i+1/2} \\ \phi_i(x) = 0 & \text{sinon} \end{cases}$$

et on pose également :

$$\begin{cases} \phi_{N+1}(x) = \dfrac{x - x_N}{h_{N+1/2}} & \text{si } x \in K_{N+1/2} \\ \phi_{N+1}(x) = 0 & \text{sinon} \\ \phi_0(x) = \dfrac{x_1 - x}{h_{1/2}} & \text{si } x \in K_{1/2} \\ \phi_0(x) = 0 & \text{sinon} \end{cases}$$

On vérifie facilement que $\phi_i \in H_N$ pour tout $0 = 1,\dots,N+1$ et que $H_N = \text{Vect}\{\phi_0,\dots,\phi_{N+1}\}$. La formulation éléments finis consiste alors à trouver $u^{(N)} \in H_N$ tel que

$$a(u^{(N)}, v) = T(v) \qquad \forall v \in H_N \tag{4.76}$$

Pour construire le système linéaire à résoudre, on prend successivement $v = \phi_i$, $i = 0,\dots,N+1$ dans (4.76). Soit

$$u^{(N)} = \sum_{j=0}^{N+1} u_j \phi_j$$

solution de

$$a(u^{(N)}, \phi_i) = T(\phi_i) \qquad \forall i = 0,\dots,N+1.$$

La famille $(u_j)_{j=0,\dots,N+1}$ est donc solution du système linéaire

$$\sum_{j=0}^{N} \mathcal{K}_{i,j} u_j = \mathcal{G}_i \qquad i = 0,\dots,N+1,$$

où $\mathcal{K}_{i,j} = a(\phi_j, \phi_i)$ et $\mathcal{G}_i = T(\phi_i)$. Calculons $\mathcal{K}_{i,j}$ et $\mathcal{G}_i$; on a :

$$\mathcal{K}_{ij} = \int_0^1 \phi_j'(x)\phi_i'(x)\,\mathrm{d}x + \int_0^1 \phi_j(x)\phi_i(x)dx \quad \text{avec} \quad \phi_i'(x) = \begin{cases} 1/h_{i-1/2} & \text{si } x \in]x_{i-1}x_i[\\ -1/h_{i+1/2} & \text{si } x \in]x_i, x_{i+1}[\\ 0 & \text{ailleurs} \end{cases}$$

Donc :

— si $1 \leqslant i = j \leqslant N$, on a

$$\mathcal{K}_{i,i} = \int_0^1 (\phi_i'(x))^2\,\mathrm{d}x + \int_0^1 (\phi_i(x))^2\,\mathrm{d}x = \frac{1}{h_{i-1/2}} + \frac{1}{h_{i+1/2}} + \frac{h_{i-1/2}}{3} + \frac{h_{i+1/2}}{3}.$$

— si $i = j = N+1$, alors

$$\mathcal{K}_{N+1,N+1} = \int_0^1 (\phi_{N+1}'(x))^2\,\mathrm{d}x + \int_0^1 (\phi_{N+1}(x))^2\,\mathrm{d}x = \frac{1}{h_{N+1/2}} + \frac{h_{N+1/2}}{3}.$$

— si $i = j = 0$, alors

$$\mathcal{K}_{0,0} = \int_0^1 (\phi_0'(x))^2\,\mathrm{d}x + \int_0^1 (\phi_0(x))^2\,\mathrm{d}x + \phi_0^2 = \frac{1}{h_{1/2}} + \frac{h_{1/2}}{3} + 1.$$

— si $0 \leqslant i \leqslant N$ et $j = i+1$, on a :

$$\mathcal{K}_{i,i+1} = \int_0^1 \phi_i'(x)\phi_{i+1}'(x)\,\mathrm{d}x + \int_0^1 \phi_i(x)\phi_{i+1}(x)\,\mathrm{d}x = -\frac{1}{h_{i+1/2}} + \frac{h_{i+1/2}}{6}$$

La matrice étant symétrique, si $2 \leqslant i \leqslant N+1$ et $j = i-1$, on a :

$$\mathcal{K}_{i,i-1} = \mathcal{K}_{i-1,i} = -\frac{1}{h_{i-1/2}} + \frac{h_{i-1/2}}{6}.$$

Calculons maintenant $\mathcal{G}_i$.

$$\mathcal{G}_i = \int_{x_i-1}^{x_{i+1}} f(x)\phi_i(x)\,\mathrm{d}x + \phi_i(1).$$

— Si f est constante, on a alors

$$\mathcal{G}_i = f\int_{x_{i-1}}^{x_{i+1}} \phi_i(x)\,\mathrm{d}x + \phi_i(1) = 1/2(h_{i-1/2} + h_{i+1/2})f + \phi_i(1)$$

— Si f n'est pas constante, on procède à une intégration numérique. On peut, par exemple, utiliser la formule des trapèzes pour le calcul des intégrales

$$\int_{x_{i-1}}^{x_i} f(x)\phi_i(x)dx \quad \text{et} \quad \int_{x_i}^{x_{i+1}} f(x)\phi_i(x)dx$$

On obtient alors :

$$\mathcal{G}_i = 1/2(h_{i-1/2} + h_{i+1/2})f(x_i) + \phi_i(1).$$

Le schéma obtenu est donc :

$$\begin{aligned}
&\Big(\frac{1}{h_{i-1/2}} + \frac{1}{h_{i+1/2}} + \frac{h_{i-1/2}}{3} + \frac{h_{i+1/2}}{3}\Big)u_i + \Big(\frac{h_{i-1/2}}{6} - \frac{1}{h_{i-1/2}}\Big)u_{i-1} + \Big(\frac{h_{i+1/2}}{6} - \frac{1}{h_{i+1/2}}\Big)u_{i+1} \\
&\qquad = 1/2(h_{i-1/2} + h_{i+1/2})f(x_i) \qquad i = 1,\dots,N \\
&\Big(\frac{1}{h_{1/2}} + \frac{h_{1/2}}{3} + 1\Big)u_0 + \Big(\frac{1}{h_{i+1/2}} + \frac{h_{i+1/2}}{6}\Big)u_1 = 1/2h_{1/2}f(x_0) \\
&\Big(\frac{1}{h_{N+1/2}} + \frac{h_{N+1/2}}{3}\Big)u_{N+1}\Big(\frac{h_{N+1/2}}{6} - \frac{1}{h_{N+1/2}}\Big) = 1/2h_{N+1/2}f(x_{N+1}) + 1.
\end{aligned}$$

Exercice 55. 1. L'espace d'approximation V_h est l'ensemble des fonctions continues, affines sur chaque maille $K_i = [x_i, x_{i+1}]$ et nulles en 0. L'ensemble V_h est engendré par les fonctions de base éléments finis P1 aux nœuds $x_i, i = 1,\dots,N+1$. Ces fonctions s'écrivent

$$\phi_i(x) = \left(1 - \frac{|x - x_i|}{h}\right)^+ \qquad i = 1,\dots,N+1.$$

2. Le système linéaire à résoudre s'écrit :

$$\mathcal{K}u = \mathcal{G},$$

où $\mathcal{K}$ est une matrice d'ordre $N+1, \mathcal{G} \in \mathbb{R}^{N+1}$ et

$$\mathcal{K}_{i,j} = \int_0^1 [\phi_i'(x)\phi_j'(x) + \phi_i(x)\phi_j(x)]\,\mathrm{d}x,$$

$$\mathcal{G}_i = \int_0^1 f(x)\phi_i(x)\,\mathrm{d}x, \qquad i = 1,\dots,N+1.$$

Les intégrales $\int_0^1 \phi_i'(x)\phi_j'(x)dx$ et $\int_0^1 f(x)\phi_i(x)dx$ sont calculées par exemple à l'exercice 51, pour $i,j = 1,\dots,N$. Il reste à calculer $b_{i,j} = \int_0^1 \phi_i(x)\phi_j(x)dx$, pour $i,j = 1,\dots,N$ et les intégrales faisant intervenir le nœud $N+1$. En ce qui concerne le calcul de $b_{i,j}$ pour $i,j = 1,\dots,N$, quatre cas se présentent.

1. Si $j = i$, $b_{i,i} = \int_{x_{i-1}}^{x_i} \left(1 + \frac{x-x_i}{h}\right)^2 \mathrm{d}x + \int_{x_i}^{x_{i+1}} \left(1 - \frac{x-x_i}{h}\right)^2 \mathrm{d}x$, par changement de variable, $\xi = 1 + \frac{x-x_i}{h}$ dans la première intégrale et $\xi = 1 - \frac{x-x_i}{h}$ dans la seconde, on a donc :

$$b_{i,i} = 2h \int_0^1 \xi^2 d\xi = \frac{2h}{3}.$$

2. Si $j = i+1$, $b_{i,i+1} = \int_{x_i}^{x_{i+1}} \left(1 - \frac{x-x_i}{h}\right)\left(1 + \frac{x-x_{i+1}}{h}\right) dx$. Posons $\xi = \frac{x-x_i}{h}$, on a donc $\frac{x-x_{i+1}}{h} = \frac{x-x_i+x_i-x_{i+1}}{h} = \xi - 1$. Donc $b_{i,i+1} = \int_0^1 (1-\xi)\xi h d\xi = h\left[\frac{1}{2} - \frac{1}{3}\right] = \frac{h}{6}$.

3. De même, par symétrie, si $j = i-1$, $b_{i,i+1} = \frac{h}{6}$.

4. Dans tous les autres cas, $b_{i,j} = 0$ car le support de ϕ_i et ϕ_j sont disjoints.

On a donc finalement :

$$\mathcal{K}_{i,i} = \frac{2}{h} + \frac{2h}{3} \quad \text{pour } i = 1, \ldots, N$$

$$\mathcal{K}_{i,i+1} = -\frac{1}{h} + \frac{h}{6} \quad \text{pour } i = 1, \ldots, N$$

$$\mathcal{K}_{i-1,i} = -\frac{1}{h} + \frac{h}{6} \quad \text{pour } i = 2, \ldots, N+1$$

En ce qui concerne le nœud $N+1$, on a :

$$\int_0^1 \phi'_{N+1}(x)\phi'_{N+1}(x)\,\mathrm{d}x = h \text{ et } b_{N+1,N+1} = \int_{x_N}^{x_{N+1}} \left(1 + \frac{x - x_{N+1}}{h}\right)^2 \mathrm{d}x = h\int_0^1 \xi^2 d\xi = \frac{h}{3}$$

Donc $\mathcal{K}_{N+1,N+1} = \frac{1}{h} + \frac{h}{3}$.
D'autre part, avec une intégration numérique par la méthode des trapèzes pour le calcul de $\mathcal{G}_i$, on obtient

$$\mathcal{G}_i = hf(x_i) \qquad i = 1, \ldots, N$$
$$\mathcal{G}_{N+1} = \tfrac{h}{2} f(x_{N+1}).$$

Le schéma éléments finis s'écrit donc finalement :

$$\left(\frac{2}{h} + \frac{2h}{3}\right) u_i + \left(-\frac{1}{h} + \frac{h}{6}\right) u_{i-1} + \left(-\frac{1}{h} + \frac{h}{6}\right) u_{i+1} = hf(x_i), \quad i = 2, \ldots, N$$
$$\left(\frac{2}{h} + \frac{2h}{3}\right) u_1 + \left(-\frac{1}{h} + \frac{h}{6}\right) u_2 = hf(x_i)$$
$$\left(\frac{1}{h} + \frac{h}{3}\right) u_{N+1} + \left(-\frac{1}{h} + \frac{h}{6}\right) u_N = \frac{h}{2} f(x_{N+1}).$$

Le schéma différences finies pour le problème s'écrit :

$$\begin{cases} \dfrac{1}{h^2}[2u_i - u_{i-1} - u_{i+1}] + u_i = f(x_i) i = 1, \ldots, N \\ u_0 = 0 \\ u_{N+1} = u_N. \end{cases}$$

Ce qui s'écrit encore :

$$\begin{cases} \left(\dfrac{2}{h} + h\right) u_i - \dfrac{1}{h} u_{i-1} - \dfrac{1}{h} u_{i+1} = hf(x_i), \quad i = 2, \ldots, N \\ \left(\dfrac{2}{h} + h\right) u_1 - \dfrac{1}{h} u_2 = hf(x_1) \\ \left(\dfrac{1}{h} + h\right) u_N - \dfrac{1}{h} u_{N-1} = hf(x_N). \end{cases}$$

Donc le schéma EF P1 et le schéma différences finies ne sont pas équivalents. Pour discrétiser le problème par un schéma volumes finis, on commence par intégrer l'équation sur chaque maille

$K_i = [x_{i-\frac{1}{2}} x_{i+\frac{1}{2}}]$:

$$-u'(x_{i+\frac{1}{2}}) + u'(x_{i\frac{1}{2}}) + \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(x)\,\mathrm{d}x = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(x)\,\mathrm{d}x.$$

La discrétisation de cette équation donne alors, en fonction des inconnues discrètes $u_1, \ldots, u_N$ (en tenant compte des conditions aux limites).

$$\begin{cases} -\dfrac{u_{i+1}-u_i}{h} + \dfrac{u_i - u_{i-1}}{h} + hu_i = hf_i \qquad i = 2, \ldots, N-1 \\ -\dfrac{u_2-u_1}{h} + \dfrac{2u_1}{h} + hu_1 = hf_1 \\ \dfrac{u_N - u_{N-1}}{h} + hu_N = hf_N, \end{cases}$$

avec $f_i = \frac{1}{h}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(x)dx$. Ceci s'écrit encore :

$$\begin{cases} \left(\dfrac{2}{h} + h\right) u_i - \dfrac{1}{h}u_{i-1} - \dfrac{1}{h}u_{i+1} = hf(x_i), \quad i = 2, \ldots, N \\ \left(\dfrac{3}{h} + h\right) u_1 - \dfrac{1}{h}u_2 = hf(x_1) \\ \left(\dfrac{1}{h} + h\right) u_N - \dfrac{1}{h}u_{N-1} = hf(x_N). \end{cases}$$

Les schémas éléments finis, différences finies et volumes finis ne sont donc pas exactement les mêmes.

3. Dans le cas du schéma éléments finis, on a

$$u_h|_{[x_N, x_{N+1}]} = u_N\phi_{N-1} + u_{N+1}\phi_N.$$

Donc $u_h'(1) = \frac{1}{h}(u_{N+1} - u_N)$. Cette quantité n'est pas forcément égale à 0. (Faire le calcul par exemple dans le cas de deux mailles).

Exercice 56 — Éléments finis pour un problème de réaction-diffusion.

1. Soit v une fonction suffisamment régulière, on multiplie la première équation de (4.41) par v et on intègre sur $]0\,;1[$. En effectuant des intégrations par parties et en tenant compte des conditions aux limites, on obtient :

$$\int_0^1 u'(x)v'(x)\ \mathrm{d}x + \alpha\int_0^1 u(x)v(x)\ \mathrm{d}x + u(0)v(0) = \int_0^1 f(x)v(x)\ \mathrm{d}x.$$

Pour que les intégrales aient un sens, il suffit de prendre $u, v \in H^1(]0\,;1[)$, auquel cas les fonctions sont continues et donc les valeurs $u(0)$ et $v(0)$ ont aussi un sens. On en déduit qu'une formulation faible consiste à trouver $u \in H^1(]0\,;1[)$ tel que :

$$\int_0^1 u'(x)v'(x)\ \mathrm{d}x + \alpha\int_0^1 u(x)v(x)\ \mathrm{d}x + u(0)v(0) = \int_0^1 f(x)v(x)\ dx \qquad \forall v \in H^1(]0\,;1[)$$

Autrement dit, la formulation variationnelle consiste à trouver $u \in H^1(]0\,;1[)$ solution de :

$$J(u) = \min_{v\in H} J(v) \quad \text{avec} \quad J(v) = \frac{1}{2}a(v,v) - T(v)$$

où a est la forme bilinéaire définie par

$$a(u,v) = \int_0^1 u'(x)v'(x)\,\mathrm{d}x + \alpha\int_0^1 u(x)v(x)\,\mathrm{d}x + u(0)v(0)$$

et T la forme linéaire continue définie par

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

2. (a) Supposons u régulière et prenons d'abord $v \in C_c^1(]0\,;1[)$. On a alors

$$\int_0^1 u'(x)v'(x)\ \mathrm{d}x + \alpha \int_0^1 u(x)v(x)\ \mathrm{d}x + u(0)v(0) = \int_0^1 f(x)v(x)\ dx$$

et donc en intégrant par parties :

$$\int_0^1 (-u''(x) + \alpha u(x) - f(x))v(x)\ \mathrm{d}x = 0$$

Comme ceci est vrai pout toute fonction $v \in C_c(]0\,;1[)$, on en déduit que

$$-u''(x) + \alpha u(x) = f(x),\ x \in]0\,;1[$$

Prenons maintenant $v \in H^1(]0\,;1[)$, en intégrant par parties et tenant compte de ce qui précède, on obtient

$$(-u'(0) + u(0))v(0) + u'(1)v(1) = 0$$

Comme ceci est vrai pout toute fonction $v \in H^1(]0\,;1[)$, on en déduit que u vérifie (4.41).

(b) On peut appliquer le théorème de Lax Milgram ; en effet,

— la forme linéaire T est continue car $|T(v)| = |\int_0^1 f(x)v(x)|\ \mathrm{d}x \leqslant \|f\|_{L^2}\|v\|_{L^2}$ par l'inégalité de Cauchy–Schwarz et donc $|T(v)| \leqslant C\|v\|_{L^2}$ avec $C = \|f\|_{L^2}$.

— la forme bilinéaire a (qui est évidemment symétrique, ce qui n'est d'ailleurs pas nécessaire pour appliquer Lax-Milgram) est continue ; en effet :

$$|a(u,v)| \leqslant \|u'\|_{L^2}\|v'\|_{L^2} + \alpha\|u\|_{L^2}\|v\|_{L^2} + |u(0)||v(0)|;$$

or pour tout $x \in]0\,;1[$ $v(0) = v(x) + \int_0^x v'(t)dt$ et donc par inégalité triangulaire et par Cauchy–Schwarz, on obtient que $|v(0)| \leqslant |v(x)| + \|v'\|_{L^2}$. En intégrant cette inégalité entre 0 et 1, on obtient

$$|v(0)| \leqslant \|v\|_{L^1} + \|v'\|_{L^2} \leqslant \|v\|_{L^2} + \|v'\|_{L^2} \leqslant 2\|v\|_{H^1}.$$

La meme inégalité est évidemment vraie pour $u(0)$. On en déduit que :

$$\begin{aligned} a(u,v) &\leqslant \|u'\|_{L^2}\|v'\|_{L^2} + \alpha\|u\|_{L^2}\|v\|_{L^2} + 4\|u\|_{H^1}\|v\|_{H^1} \\ &\leqslant \|u\|_{H^1}\|v\|_{H^1} + \alpha\|u\|_{H^1}\|v\|_{H^1} + 4\|u\|_{H^1}\|v\|_{H^1} \\ &\leqslant (5+\alpha)\|u\|_{H^1}\|v\|_{H^1}, \end{aligned}$$

ce qui prouve que a est continue.

Montrons maintenant que a est coercive. Dans le cas où $\alpha > 0$, ceci est facile à vérifier, car on a

$$a(u,u) = \int_0^1 u'(x)^2\ \mathrm{d}x + \alpha \int_0^1 u(x)^2\ \mathrm{d}x + u(0)^2 \geqslant \min(\alpha, 1)\|u\|_{H^1}^2.$$

Par le théorème de Lax -Milgram, on peut donc conclure à l'existence et l'unicité de la solution de (4.42).

Dans le cas où $\alpha = 0$, on applique l'inégalité de Poincaré à la fonction $w = u - u(0)$, ce qui est licite car $w(0) = 0$; on a donc : $\|w\|_{L^2} \leqslant \|w'\|_{L^2}$ et donc $\|u'\|_{L^2} \geqslant \|u - u(0)\|_{L^2}$. En écrivant que $u = u - u(0) + u(0)$, en utilisant l'inégalité triangulaire qu'on élève au carré et le fait que $2ab \leqslant a^2 + b^2$, on en déduit que $a(u,u) \geqslant \|u - u(0)\|_{L^2}^2 + u(0)^2 \geqslant 1/2\|u\|_{L^2}^2$.

On écrit alors que

$$a(u,u) = \frac{1}{2}a(u,u) + 1/2a(u,u) \geqslant 1/2\|u'\|_{L^2}^2 + \frac{1}{4}\|u\|_{L^2}^2 \geqslant \frac{1}{4}\|u\|_{H^1}^2$$

ce qui montre que la forme bilinéaire a est encore coercive.

3. Une base de l'espace V_h est la famille des fonctions dites "chapeau", définies par

$$\varphi_i(x) = \min\Big(\frac{1}{h}(x - x_{i-1})^+, \frac{1}{h}(x_{i+1} - x)^+\Big) \qquad i = 1, \dots, N-1$$
$$\varphi_1(x) = \frac{1}{h}(x_1 - x)^+$$
$$\varphi_N(x) = \frac{1}{h}(x - x_{N+1})^+$$

On en déduit que l'espace V_h est de dimension $N + 1$.

4. Le problème discrétisé par éléments finis consiste à trouver $u_h \in V_h$ tel que :

$$\int_0^1 u_h'(x)v_h'(x)\ \mathrm{d}x + \alpha \int_0^1 u_h(x)v_h(x)\ \mathrm{d}x + u_h(0)v_h(0) = \int_0^1 f(x)v_h(x)\,dx \qquad \forall v_h \in V_h \quad (4.77)$$

5. Comme on a effectué une discrétisation par éléments finis conformes, le théorème de Lax Milgram s'applique à nouveau.

6. Commençons par le second membre $B = (b_i)_{0 \leqslant i \leqslant N}$ avec :

$$b_i = \int_0^1 f(x)\phi_i(x)dx$$

Calculons :

$$A_{i,j} = A_{j,i} = a(\phi_i, \phi_j)$$
$$= \int_0^1 \phi_i'(x)\phi_j'(x)\ \mathrm{d}x + \alpha \int_0^1 \phi_i(x)\phi_j(x)\ \mathrm{d}x + \phi_i(0)\phi_j(0) \text{ pour } i = 1, \dots, N$$

En raison de la forme des fonctions de base $(\phi_i)_{i=0,N}$, les seuls termes non nuls sont les termes $A_{i-1,i}$, $A_{i,i}$ et $A_{i,i+1}$. Après calculs, on obtient :

$$A = \begin{bmatrix} \frac{1}{h} + \frac{\alpha h}{3} + 1 & -\frac{1}{h} + \frac{\alpha h}{6} & 0 & 0 & \dots & 0 \\ -\frac{1}{h} + \frac{\alpha h}{6} & \frac{2}{h} + \frac{2\alpha h}{3} & -\frac{1}{h} + \frac{\alpha h}{6} & \ddots & & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & & & \\ & & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\frac{1}{h} + \frac{\alpha h}{6} & \frac{2}{h} + \frac{2\alpha h}{3} & -\frac{1}{h} + \frac{\alpha h}{6} \\ 0 & \dots & & 0 & -\frac{1}{h} + \frac{\alpha h}{6} & \frac{1}{h} + \frac{\alpha h}{3} \end{bmatrix}$$

7. Soit $u \in H^1(]0\,;1[)$ une solution de (4.42), alors $-u'' = f - \alpha u$ au sens des distibutions mais comme $f \in L^2(]0\,;1[)$ et $u \in L^2(]0\,;1[)$, on en déduit que $u \in H^2(]0\,;1[)$. On peut donc appliquer les résultats du cours. On a vu en cours que si u_I l'interpolée de u dans V_h, On a $\|u - u_h\|_{L^2(0,1)} \leqslant C\|u - u_I\|_{L^2(0,1)}$ où C est la racine carrée du rapport de la constante de continuité sur la constante de coercivité, c'est-à-dire :

$$C = \sqrt{\frac{5 + \alpha}{\min(\alpha, 1)}}$$

De plus, on a aussi vu que si $u \in H^2(]0\,;1[)$, l'erreur d'interpolation est d'ordre h ; plus précisément, on a :

$$\|u - u_I\|^2_{L^2(0,1)} \leqslant (1 + h^2)h^2\|u''\|^2_{L^2(0,1)}$$

On en déduit que :

$$\|u_h - u\|_{H^1(0,1)} \leqslant \sqrt{\frac{5 + \alpha}{\min(\alpha, 1)}}\sqrt{1 + h^2}\|u''\|_{L^2(0,1)}h$$

Exercice 60.

1. (a) On note x, y les deux variables de $\mathbb{R}^2$. L'espace $\mathcal{Q}_1$ est l'ensemble des polynômes de la forme $a + bx + cy + dxy$ avec a, b, c et $d \in \mathbb{R}$. On a donc $\dim \mathcal{Q}_1 = 4 = \operatorname{card} \Sigma_1 = \operatorname{card} \Sigma_2$. Pour montrer que $(e_1, \Sigma_1, \mathcal{Q}_1)$ est un élément fini de Lagrange, il suffit de montrer que $f \in \mathcal{Q}_1$ et $f|_{\Sigma_1} = 0$ implique $f = 0$. Soient donc a, b, c et $d \in \mathbb{R}$. On pose $f(x, y) = a + bx + cy + dxy$ pour $(x, y) \in e_1$ et on suppose que $f|_{\Sigma_1} = 0$, c'est-à-dire $f(-1, 1) = 0$, $f(0, 1) = 0$, $f(1, 0) = 0$ et $f(-1, 0) = 0$. On a donc :

$$\begin{cases} a - b + c - d = 0 \\ a + c = 0 \\ a + b = 0 \\ a - b = 0 \end{cases}$$

Les deux dernières équations entrainent $a = b = 0$, la troisième implique $c = 0$ et la première donne enfin que $d = 0$. On a donc montré que $f = 0$. On en déduit que $(e_1, \Sigma_1, \mathcal{Q}_1)$ est un élément fini de Lagrange. Pour montrer que $(e_2, \Sigma_2, \mathcal{Q}_1)$ est un élément fini de Lagrange, on procède de la même façon : soient a, b, c et $d \in \mathbb{R}$ et $f(x, y) = a + bx + cy + dxy$ pour $(x, y) \in e_2$. On suppose que $f|_{\Sigma_2} = 0$, c'est-à-dire $f(0, 1) = 0$, $f(2, 1) = 0$, $f(2, 0) = 0$ et $f(1, 0) = 0$. On a donc :

$$\begin{cases} a + c = 0 \\ a + 2b + c + 2d = 0 \\ a + 2b = 0 \\ a + b = 0 \end{cases}$$

Les deux dernières équations donnent $a = b = 0$, la première donne alors $c = 0$ et, finalement, la quatrième donne $d = 0$. On a donc montré que $f = 0$. On en déduit que $(e_2, \Sigma_2, \mathcal{Q}_1)$ est un élément fini de Lagrange.

(b) L'espace (de dimension finie) associé à cette discrétisation est engendré par les six fonctions de base globales. On va montrer que la fonction de base associée à M_1 (par exemple) n'est pas dans $H^1(\Omega)$. On note ϕ_1 cette fonction de base. On doit avoir $\phi_1|_{e_1} \in \mathcal{Q}_1$, $\phi_1|_{e_2} \in \mathcal{Q}_1$, $\phi_1(M_1) = 1$ ainsi que $\phi_1(M_i) = 0$ si $i \neq 1$. On en déduit que $\phi_1 = 0$ sur e_2 et $\phi_1(x, y) = -xy$ si $(x, y) \in e_1$. On a bien $\phi_1 \in L^2(\Omega)$ mais on va montrer maintenant que ϕ_1 n'a pas de dérivée faible dans $L^2(\Omega)$ (et donc que $\phi_1 \notin H^1(\Omega)$). On va s'intéresser à la dérivée faible par rapport à x (mais on pourrait faire un raisonnement similaire pour la dérivée faible par rapport à y). On suppose que ϕ_1 a une dérivée faible par rapport à x dans $L^2(\Omega)$ (et on va montrer que ceci mène à une contradiction). Supposons donc qu'il existe une fonction $\psi \in L^2(\Omega)$ telle que

$$I = \int_{-1}^{2} \int_{0}^{1} \phi_1(x, y) \frac{\partial \varphi}{\partial x}(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{-1}^{2} \int_{0}^{1} \psi(x, y) \varphi(x, y) \, \mathrm{d}x \, \mathrm{d}y \qquad \forall \varphi \in C_c^\infty(\Omega). \quad (4.78)$$

Soit $\varphi \in C_c^\infty(\Omega)$, comme ϕ_1 est nulle sur e_2, on a

$$I = \iint_{e_1} \phi_1(x, y) \frac{\partial \varphi}{\partial x}(x, y) \, \mathrm{d}x \, \mathrm{d}y$$

et donc :

$$I = \int_0^1 \left(\int_{-1}^{1-y} (-xy) \frac{\partial \varphi}{\partial x}(x, y) \, \mathrm{d}x \right) \mathrm{d}y.$$

Par intégration par parties, en tenant compte du fait que φ est à support compact sur Ω, on obtient :

$$\begin{aligned} I &= \int_0^1 \left[\int_{-1}^{1-y} y \, \varphi(x, y) \, \mathrm{d}x - (1 - y) y \varphi(1 - y, y) \right] \mathrm{d}y \\ &= \int_0^1 \int_{-1}^{2} y 1_{e_1}(x, y) \varphi(x, y) \, \mathrm{d}x - \int_0^1 (1 - y) y \varphi(1 - y, y) \, \mathrm{d}y. \end{aligned}$$

En posant $\widetilde{\psi}(x,y) = -\psi(x,y) + y1_{e_1}(x,y)$, on a $\widetilde{\psi} \in L^2(\Omega)$ et :

$$\int_0^1 (1-y)y\varphi(1-y,y)\,\mathrm{d}y = \int_{-1}^2 \int_0^1 \widetilde{\psi}(x,y)\varphi(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{4.79}$$

Pour aboutir à une contradiction, on va montrer que (4.79) est fausse pour certains $\varphi \in C_c^\infty(\Omega)$. On remarque tout d'abord qu'il existe $\varphi \in C_c^\infty(\Omega)$ tel que

$$\int_0^1 (1-y)(y)\varphi(1-y,y)\,\mathrm{d}y > 0.$$

(Il suffit de choisir $\varphi \in C_c^\infty(\Omega)$ tel que $\varphi \geqslant 0$ et $\varphi(1-y,y) > 0$ pour $y = 1/2$, par exemple.) On se donne maintenant une fonction $\rho \in C_c^\infty(\mathbb{R})$ tel que $\rho(0) = 1$ et $\rho = 0$ sur $[-1\,;1]^c$ et on écrit (4.79) avec φ_n au lieu de φ, où φ_n est définie par :

$$\varphi_n(x,y) = \varphi(x,y)\rho(n(x+y-1))$$

(noter que l'on a bien $\varphi_n \in C_c^\infty(\Omega)$ car $\rho \in C^\infty(\mathbb{R})$ et $\varphi \in C_c^\infty(\Omega)$) On a donc :

$$\int_0^1 (1-y)y\varphi_n(1-y,y)\,\mathrm{d}y = \int_{-1}^2 \int_0^1 \widetilde{\psi}(x,y)\varphi_n(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Le terme de gauche de cette égalité est indépendant de n et non nul car $\varphi_n(1-y,y) = \varphi(1-y,y)$ pour tout n et tout $y \in [0\,;1]$. Le terme de droite tend vers 0 quand $n \to \infty$ par convergence dominée car $\widetilde{\psi}\varphi_n \to 0$ p.p. et $|\widetilde{\psi}\varphi_n| \leqslant \|\rho\|_\infty |\widetilde{\psi}|\ |\varphi| \in L^1(\Omega)$. Ceci donne la contradiction désirée et donc que $\phi_1 \notin H^1(\Omega)$. L'hypothèse non vérifiée (pour avoir la cohérence globale) est l'hypothèse (4.6). En posant $S = \bar{e}_1 \cap \bar{e}_2$, on a $\Sigma_1 \cap S = \Sigma_2 \cap S = \{M_2, M_5\}$ et aussi, bien sûr, $\varphi_1|_S = \varphi_1|_S$ mais on remarque que l'ensemble ($\{M_2, M_5\}$ n'est pas $\mathcal{Q}_1|_S$-unisolvant car $\mathrm{card}(\{M_2, M_5\}) = 2$ et $\dim(\mathcal{Q}_1|_S) = 3$.

2. (a) Les quatre fonctions de base de $(e, \Sigma, \mathcal{P})$ sont :

$$\phi_1(x,y) = \frac{1}{4}(x+1)(y+1) \qquad \phi_2(x,y) = -\frac{1}{4}(x+1)(y-1)$$
$$\phi_3(x,y) = -\frac{1}{4}(x-1)(y+1) \quad \phi_4(x,y) = \frac{1}{4}(x-1)(y-1)$$

Construction de F_1 Pour $(x,y) \in e$, on pose

$$F_1(x,y) = M_1\phi_3(x,y) + M_2\phi_1(x,y) + M_5\phi_2(x,y) + M_4\phi_4(x,y)$$

ce qui donne

$$\begin{aligned} 4F_1(x,y) &= \begin{bmatrix} -1 \\ 1 \end{bmatrix}(1-x)(1+y) + \begin{bmatrix} 0 \\ 1 \end{bmatrix}(1+x)(1+y) + \begin{bmatrix} 1 \\ 0 \end{bmatrix}(1+x)(1-y) \\ &\qquad + \begin{bmatrix} -1 \\ 0 \end{bmatrix}(1-x)(1-y) \\ &= \begin{bmatrix} -1+3x-y-xy \\ 2(1+y) \end{bmatrix}. \end{aligned}$$

Pour $y \in [-1\,;1]$ fixé, la première composante de $F_1(x,y)$, qu'on note $F_{1,y}(x)$, est linéaire par rapport à x. Comme $F_{1,y}(-1) = -1$ et $F_{1,y}(-1) = \dfrac{1-y}{2}$, $F_{1,y}$ est une bijection de $[-1\,;1]$ dans $[-1\,;\frac{1-y}{2}]$. Donc pour $b \in [-1\,;1]$ donné, F_1 est une bijection de $[-1\,;1] \times \{b\}$ dans $[-1\,;(1-b)/2] \times \{\frac{1+b}{2}\}$. On en déduit que F_1 est une bijection de e dans e_1.

Construction de F_2 Pour $(x,y) \in e$, on pose

$$F_2(x,y) = M_2\phi_3(x,y) + M_3\phi_1(x,y) + M_6\phi_2(x,y) + M_5\phi_4(x,y)$$

ce qui donne

$$\begin{aligned}4F_2(x,y) &= \begin{bmatrix}0\\1\end{bmatrix}(1-x)(1+y) + \begin{bmatrix}2\\1\end{bmatrix}(1+x)(1+y) + \begin{bmatrix}2\\0\end{bmatrix}(1+x)(1-y)\\ &\qquad + \begin{bmatrix}1\\0\end{bmatrix}(1-x)(1-y)\\ &= \begin{bmatrix}5+3x-y+xy\\ 2+2y\end{bmatrix}\end{aligned}$$

Pour $y \in [-1\,;1]$ fixé, la première composante de $F_2(x,y)$, qu'on note $F_{2,y}(x)$ est linéaire par rapport à x et $F_{2,y}$ est une bijection de $[-1\,;1]$ dans $[\frac{1-y}{2}\,;2]$ et donc pour $b \in [-1\,;1]$ fixé, F_2 est une bijection de $[-1\,;1]\times\{b\}$ dans $[\frac{1-b}{2}\,;2]\times\left\{\frac{1+b}{2}\right\}$. On en déduit que F_2 est une bijection de e dans e_2.
Noter que les fonctions F_1 et F_2 ne sont pas affines.

(b) Les éléments $(e_1,\Sigma_1,\mathcal{P}_{e_1})$ et $(e_2,\Sigma_2,\mathcal{P}_{e_2})$ sont les éléments finis de Lagrange construits à partir de l'élément fini de Lagrange $(e,\Sigma,\mathcal{P})$ et des bijections F_1 et F_2 (de e dans e_1 et de e dans e_2), voir la proposition 4.11. Pour montrer que l'espace vectoriel construit avec $(e_1,\Sigma_1,\mathcal{P}_{e_1})$ et $(e_2,\Sigma_2,\mathcal{P}_{e_2})$ est inclus dans $H^1(\Omega)$, il suffit de vérifier la propriété de "cohérence globale" donnée dans la proposition 4.12. On pose

$$S = \bar{e}_1 \cap \bar{e}_2 = \{(x,y)\in\bar{\Omega}, x+y=1\} = \{(1-y,y), y\in[0\,;1]\}$$

On remarque tout d'abord que $\Sigma_1 \cap S = \Sigma_2 \cap S = \{M_2, M_5\}$. On détermine maintenant $\mathcal{P}_{e_1|S}$ et $\mathcal{P}_{e_2|S}$. Soient $f \in \mathcal{P}_{e_1}$ et $(x,y)\in S$ (c'est-à-dire $y\in[0\,;1]$ et $x+y=1$) : on a $f(x,y) = f\circ F_1(1,2y-1)$. (On a utilisé ici le fait que $F_1(\{1\}\times[-1\,;1]) = 5$). Par conséquent, $\mathcal{P}_{e_1|S}$ est l'ensemble des fonctions de S dans $\mathbb{R}$ de la forme $(x,y)\mapsto g(1,2y-1)$, où $g\in\mathcal{Q}_1$, c'est-à-dire l'ensemble des fonctions de S dans $\mathbb{R}$ de la forme :

$$(x,y)\mapsto \alpha+\beta+\gamma(2y-1)+\delta(2y-1)$$

avec α, β, γ, et $\delta\in\mathbb{R}$. On en déduit que $\mathcal{P}_{e_1}|_S$ est l'ensemble des fonctions de S dans $\mathbb{R}$ de la forme $(x,y)\mapsto a+by$ avec $a,b\in\mathbb{R}$. On a donc $\mathcal{P}_{e_1}|_S = \mathcal{P}_{e_2}|_S$. Ceci donne la condition (4.5). Enfin, la condition (4.6) est bien vérifiée ; en effet, l'ensemble Σ_1 est $\mathcal{P}_{e_1}|_S$)-unisolvant car un élément de $\mathcal{P}_{e_1}|_S$ est bien déterminé de manière unique par ses valeurs en $(0,1)$ et $(1,0)$.

Exercice 61 — éléments affine–équivalents. Si les fonctions de base de $(\bar{K},\bar{\Sigma},\overline{\mathcal{P}})$ sont affines, alors l'espace $\overline{\mathcal{P}}$ est constitué des fonctions affines, on peut donc écrire

$$\overline{\mathcal{P}} = \{\bar{f}:\bar{K}\to\mathbb{R}, \bar{x}=(\bar{x}_1,\bar{x}_2)\mapsto f(\bar{x}) = a_1\bar{x}_1 + a_2\bar{x}_2 + b\}.$$

Comme $(\bar{K},\bar{\Sigma},\overline{\mathcal{P}})$ et $((K,\Sigma,\mathcal{P})$ sont affines équivalents, on a par définition :

$$\mathcal{P} = \{f: K\to\mathbb{R}; f = \bar{f}\circ F^{-1}, \bar{f}\in\overline{\mathcal{P}}\},$$

où F est une fonction affine de $\bar{K}$ dans K la fonction F^{-1} est donc aussi affine et s'écrit donc sous la forme :

$$F^{-1}(x) = F^{-1}((x_1,x_2)) = (\alpha_1x_1+\alpha_1x_2++\gamma, \beta_1x_1+\beta_2x_2+\delta)$$

Donc si $f = \bar{f}\circ F^{-1}\in\mathcal{P}$, on a

$$f(x) = \bar{f}\circ F^{-1}((x_1,x_2)) = \bar{f}[(\alpha_1x_1+\alpha_2x_2+\gamma, \beta_1x_1+\beta_2x_2+\delta)] = A_1x_1+A_2x_2+B$$

où A_1, A_2 et $B\in\mathbb{R}^2$. On en déduit que f est bien affine. L'espace $\mathcal{P}$ est donc constitué de fonctions affines. Pour montrer que les fonctions de base locales sont affines, il suffit de montrer que l'espace $\mathcal{P}$ est constitué de toutes les fonctions affines. En effet, si f est affine, emphi.e. $f(x_1,x_2) = A_1x_1+A_2x_2+B$, avec $A_1,A_2,B\in\mathbb{R}^2$, on montre facilement que $\bar{f} : f\circ F\in\overline{\mathcal{P}}$, ce qui montre que $f\in\mathcal{P}$.

Exercice 62.

1. Le problème continu.

   a) On vérifie que la forme bilinéaire définie par $a(u,v) = \int_\Omega \Lambda\nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x$ est continue et coercive, que la forme linéaire définie par $T(v) = \int_\Omega f(x)v(x)\,\mathrm{d}x$ est continue et on applique le théorème de Lax Milgram.

   b)

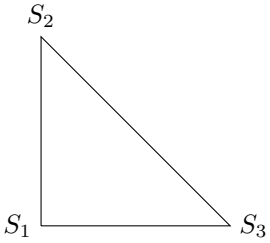
$$-\operatorname{div}\Lambda\nabla u = f \text{ dans } \Omega, u = 0 \text{ sur } \partial\Omega. \tag{4.80}$$

2.

$$\phi_1(x,y) = 1 - nx - ny,\ \phi_2(x,y) = nx,\ \phi_3(x,y) = ny.$$

3. Calculons les matrices élémentaires d'assemblage des triangles K_{ij1} et K_{ij2}.

   Triangle K_{ij1}
   La matrice élémentaire A sur le triangle K_{ij1} de sommets $S_1 = (0,0)$, $S_2 = (h,0)$ et $S_3 = (0,h)$ a pour coefficients $a_{i,j} = \int_{K_{ij1}} \Lambda\nabla\phi_i(x,y) \cdot \nabla\phi_j(x,y)\,\mathrm{d}x\,\mathrm{d}y$, où ϕ_1, ϕ_2, ϕ_3 sont les fonctions de forme du triangle K_{ij1} qu'on a calculées à la question 2. On a donc



$$\nabla\phi_1(x,y) = \begin{bmatrix}-n\\-n\end{bmatrix},\ \nabla\phi_2(x,y) = \begin{bmatrix}n\\0\end{bmatrix},\ \nabla\phi_3(x,y) = \begin{bmatrix}0\\n\end{bmatrix}.$$

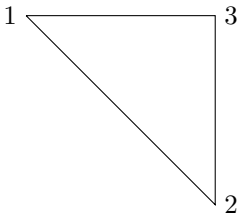
   Notons α, β, γ les coefficients de la matrice Λ : $\Lambda = \begin{bmatrix}\alpha & \beta\\ \beta & \gamma\end{bmatrix}$. En remarquant que l'aire du triangle K_{ij1} est $h^2/2 = 1/2n^2$, on a

$$\begin{aligned}
a_{1,1} &= \int_{K_{ij1}} \Lambda\nabla\phi_1(x,y)\cdot\nabla\phi_1(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{K_{ij1}} \begin{bmatrix}\alpha & \beta\\ \beta & \gamma\end{bmatrix}\begin{bmatrix}-n\\-n\end{bmatrix} \cdot \begin{bmatrix}-n\\-n\end{bmatrix} \frac{1}{2}(\alpha + 2\beta + \gamma),\\
a_{1,2} &= \int_{K_{ij1}} \Lambda\nabla\phi_1(x,y) \cdot \nabla\phi_2(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{K_{ij1}} \begin{bmatrix}\alpha & \beta\\ \beta & \gamma\end{bmatrix}\begin{bmatrix}-n\\-n\end{bmatrix} \cdot \begin{bmatrix}n\\0\end{bmatrix} = \frac{1}{2}(-\alpha - \beta),\\
a_{1,3} &= \int_{K_{ij1}} \Lambda\nabla\phi_1(x,y) \cdot \nabla\phi_2(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{K_{ij1}} \begin{bmatrix}\alpha & \beta\\ \beta & \gamma\end{bmatrix}\begin{bmatrix}-n\\-n\end{bmatrix} \cdot \begin{bmatrix}0\\n\end{bmatrix} = \frac{1}{2}(-\beta - \gamma),\\
a_{2,2} &= \int_{K_{ij1}} \Lambda\nabla\phi_2(x,y) \cdot \nabla\phi_2(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{K_{ij1}} \begin{bmatrix}\alpha & \beta\\ \beta & \gamma\end{bmatrix}\begin{bmatrix}n\\0\end{bmatrix} \cdot \begin{bmatrix}n\\0\end{bmatrix} = \frac{1}{2}\alpha,\\
a_{2,3} &= \int_{K_{ij1}} \Lambda\nabla\phi_2(x,y) \cdot \nabla\phi_3(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{K_{ij1}} \begin{bmatrix}\alpha & \beta\\ \beta & \gamma\end{bmatrix}\begin{bmatrix}n\\0\end{bmatrix} \cdot \begin{bmatrix}0\\n\end{bmatrix} = \frac{1}{2}\beta,\\
a_{3,3} &= \int_{K_{ij1}} \Lambda\nabla\phi_2(x,y) \cdot \nabla\phi_3(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{K_{ij1}} \begin{bmatrix}\alpha & \beta\\ \beta & \gamma\end{bmatrix}\begin{bmatrix}0\\n\end{bmatrix} \cdot \begin{bmatrix}0\\n\end{bmatrix} = \frac{1}{2}\gamma.
\end{aligned}$$

   Finalement,

$$A = \frac{1}{2}\begin{bmatrix}\alpha + 2\beta + \gamma & -\alpha - \beta & -\beta - \gamma\\ -\alpha - \beta & \alpha & \beta\\ -\beta - \gamma & \beta & \gamma\end{bmatrix}.$$

Triangle K_{ij2}



La matrice élémentaire B sur le triangle K_{ij2} de sommets $S_1 = (0, h)$, $S_2 = (h, h)$ et $S_3 = (h, 0)$ a pour coefficients

$$b_{i,j} = \int_{K_{ij2}} \Lambda \nabla \phi_i(x, y) \cdot \nabla \phi_j(x, y) \, \mathrm{d}x \, \mathrm{d}y,$$

où ϕ_1, ϕ_2, ϕ_3 sont les fonctions de forme du triangle K_{ij2} définies par

$$\phi_1(x, y) = 1 - nx, \ \phi_2(x, y) = 1 - ny, \ \phi_3(x, y) = nx + ny - 1.$$

On a donc

$$\nabla \phi_1(x, y) = \begin{bmatrix} -n \\ 0 \end{bmatrix}, \ \nabla \phi_2(x, y) = \begin{bmatrix} 0 \\ -n \end{bmatrix}, \ \nabla \phi_3(x, y) = \begin{bmatrix} n \\ n \end{bmatrix}.$$

Les coefficients de la matrice B sont donc :

$$b_{1,1} = \int_{K_{ij2}} \Lambda \nabla \phi_1(x, y) \cdot \nabla \phi_1(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{K_{ij2}} \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} -n \\ 0 \end{bmatrix} \cdot \begin{bmatrix} -n \\ 0 \end{bmatrix} = \frac{h^2}{2} n^2 \alpha = \frac{1}{2} \alpha,$$

$$b_{1,2} = \int_{K_{ij2}} \Lambda \nabla \phi_1(x, y) \cdot \nabla \phi_2(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{K_{ij2}} \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} -n \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -n \end{bmatrix} = \frac{1}{2} \beta,$$

$$b_{1,3} = \int_{K_{ij2}} \Lambda \nabla \phi_1(x, y) \cdot \nabla \phi_3(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{K_{ij2}} \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} -n \\ 0 \end{bmatrix} \cdot \begin{bmatrix} n \\ n \end{bmatrix} = -\frac{1}{2} (\alpha + \beta),$$

$$b_{2,2} = \int_{K_{ij2}} \Lambda \nabla \phi_2(x, y) \cdot \nabla \phi_2(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{K_{ij2}} \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} 0 \\ -n \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -n \end{bmatrix} = \frac{1}{2} \gamma,$$

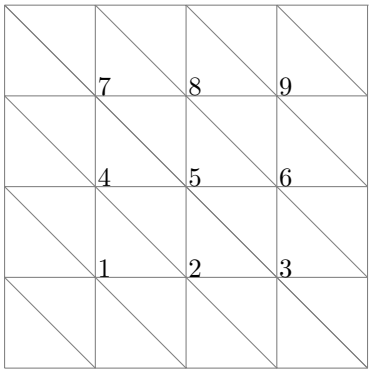
$$b_{2,3} = \int_{K_{ij2}} \Lambda \nabla \phi_2(x, y) \cdot \nabla \phi_3(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{K_{ij2}} \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} 0 \\ -n \end{bmatrix} \cdot \begin{bmatrix} n \\ n \end{bmatrix} = -\frac{1}{2} (\beta + \gamma),$$

$$b_{3,3} = \int_{K_{ij3}} \Lambda \nabla \phi_3(x, y) \cdot \nabla \phi_3(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{K_{ij2}} \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} n \\ n \end{bmatrix} \cdot \begin{bmatrix} n \\ n \end{bmatrix} = \frac{1}{2} (\alpha + 2\beta + \gamma).$$

Finalement, les matrices élémentaires des éléments K_{ij2} sont donc toutes identiques et égales à

$$B = \frac{1}{2} \begin{bmatrix} \alpha & -\alpha - \beta & -\beta - \gamma \\ \beta & \gamma & \beta + \gamma \\ -\alpha - \beta & \beta + \gamma & \alpha + 2\beta + \gamma \end{bmatrix}$$

4. Appelons $\mathcal{K}$ la matrice assemblée.



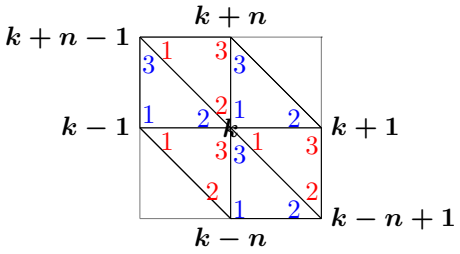
Comme on a des conditions aux limites de Dirichlet homogènes, on a N noeuds libres, avec $N = (n-1) \times (n-1)$, dont les coordonnées sont (ih, jh), $i, j = 1, \ldots, n-1$. Adoptons l'ordre lexicographique pour leur numérotation : le numéro du noeud (i, j) de la grille est donc $k(i, j) = (n-1)(j-1) + i$, voir l'exemple $n = 4$ sur la figure ci-contre. On initialise la matrice $\mathcal{K}_{k,k}$ à 0, et on écrit maintenant les coefficients non nuls en utilisant les matrices élémentaires. on a

Le support de la fonction de base du noeud k est constitué de 6 triangles dont les sommets sont les noeuds

- $k-1$, $k-n$, k, (triangle de type K_{ij2})
- $k-1$, k, $k+n-1$, (triangle de type K_{ij1})
- $k+n-1$, k, $k+n$, (triangle de type K_{ij2})
- k, $k+1$, $k+n$, (triangle de type K_{ij1})

— k, $k+1$, $k+n-1$, (triangle de type K_{ij2}).

Les numéros locaux correspondant sont donnés en bleu pour les noeuds des triangles de type K_{ij1} et en rouge pour ceux du type K_{ij2} sur la figure ci contre. En appliquant cette correspondance on obtient que, pour $i, j = 1, \ldots, n-1$, en notant $k = k(i,j)$,



$$\begin{aligned}\mathcal{K}_{k,k} &= a_{11} + b_{11} + a_{22} + b_{22} + a_{33} + b_{33} \\ &= 2(\alpha + \beta + \gamma)\end{aligned}$$

Puis pour $i = 2, \ldots, n-1$ et $j = 1, \ldots, n-1$,

$$\begin{aligned}\mathcal{K}_{k,k-1} &= a_{12} + b_{13} \\ &= -(\alpha + \beta) \ ;\end{aligned}$$

de même pour $i = 1, \ldots, n-2$ et $j = 1, \ldots, n-1$,

$$\begin{aligned}\mathcal{K}_{k,k+1} &= a_{12} + b_{13} \\ &= -(\alpha + \beta) \ ;\end{aligned}$$

pour $i = 1, \ldots, n-1$ et $j = 2, \ldots, n-1$,

$$\begin{aligned}\mathcal{K}_{k,k-n} &= a_{13} + b_{23} \\ &= -(\alpha + 2\beta + \gamma) \ ;\end{aligned}$$

de même pour $i = 1, \ldots, n-1$ et $j = 1, \ldots, n-2$,

$$\begin{aligned}\mathcal{K}_{k,k+n} &= a_{13} + b_{23} \\ &= -\frac{1}{2}(\alpha + 2\beta + \gamma) \ ;\end{aligned}$$

pour $i = 2 \ldots, n-1$ et $j = 1, \ldots, n-2$,

$$\begin{aligned}\mathcal{K}_{k,k+n-1} &= a_{23} + b_{12} \\ &= \beta \ ;\end{aligned}$$

pour $i = 1 \ldots, n-2$ et $j = 2, \ldots, n-1$,

$$\begin{aligned}\mathcal{K}_{k,k-n+1} &= a_{23} + b_{12} \\ &= \beta \ ;\end{aligned}$$

La matrice $\mathcal{K}$ est donc une matrice symétrique à structure bande, qui comporte 7 diagonales non nulles, et seulement 5 dans le cas où $\beta = 0$, c'est-à-dire lorsque la matrice Λ est diagonale. Noter que dans le cas où $\Lambda = \mathrm{Id}$, on retrouve la matrice différences finies pour le Laplacien (au $1/h^2$ près qui vient du second membre).

Exercice 63.

1. Pour obtenir une formulation faible, on considère une fonction test $\varphi \in C^2([0\,;1],\mathbb{R})$; on multiplie la première équation de (2.22) par φ et on intègre par partie, on obtient alors :

$$u'(1)\varphi(1) - u'(0)\varphi(0) + \int_0^1 (u'(x)\varphi'(x) + u(x)\varphi(x))dx = \int_0^1 x^2\varphi(x)\,\mathrm{d}x.$$

Comme $u'(1) = 0$, si on choisit $\varphi \in H = \{v \in H^1(]0\,;1[; v(0) = 0\}$, cette dernière égalité s'écrit :

$$\int_0^1 [u'(x)\varphi'(x) + u(x)\varphi(x)]\,\mathrm{d}x = \int_0^1 x^2\varphi(x)\;\mathrm{d}x - \varphi(1).$$

En posant

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

et

$$L(v) = \int_0^1 x^2 v(x)\,\mathrm{d}x + v(1),$$

on obtient la formulation variationnelle suivante :

$$\begin{cases} u \in H \\ a(u,v) = L(v) \quad \forall v \in H \end{cases} \tag{4.81}$$

On a donc montré que si u est solution de (3.46), alors u est solution de (4.81). Montrons maintenant la réciproque. Soit $u \in C^2([0\,;1])$ solution de (4.81) ; en intégrant par partie, il vient :

$$-u'(1)\varphi(1) + u'(0)\varphi(0) + \int_0^1 -u''(x)\varphi(x) + u(x)\varphi(x)\,\mathrm{d}x = \int_0^1 x^2(x)\varphi(x)\,\mathrm{d}x - \varphi(1).$$

Comme $\varphi(0) = 0$, on obtient donc :

$$\varphi(1)(-u'(1)+1) + \int_0^1 (-u''(x)\varphi(x) + u(x)\varphi(x))dx = \int_0^1 x^2\varphi(x)\,\mathrm{d}x \tag{4.82}$$

Si on choisit φ à support compact, on obtient :

$$\int_0^1 (-u''(x) + u(x) + x^2)\varphi(x)\,\mathrm{d}x = 0,$$

et comme cette égalité est vraie pour toute fonction φ à support compact sur $[0\,;1]$, on en déduit que

$$-u'' + u = x^2 \quad \text{p.p.}$$

En tenant compte de cette relation dans (4.82), on a donc $(-u'(1)+1)\varphi(1) = 0$ et comme ceci est vrai pour toute fonction $\varphi \in C^2([0\,;1],\mathbb{R})$, on en déduit que $u'(1) = 1$. Donc u est solution de (4.81).

2. Pour montrer que le problème (4.81) admet une unique solution, on va montrer que les hypothèses du théorème de Lax–Milgram sont satisfaites.

(a) Le sous espace H est fermé dans $H^1(]0\,;1[)$

(b) La forme bilinéaire a est en fait le produit scalaire sur H^1 et donc continue et coercive sur H^1.

(c) La forme linéaire L est continue sur $H^1(]0\,;1[)$: en effet, par l'inégalité de Cauchy-Schwarz,

$$\begin{aligned} |L(v)| = \Big|\int_0^1 x^2 v(x)\,\mathrm{d}x + v(1)\Big| &\leqslant \Big(\int_0^1 x^4 dx\Big)^{1/2} \|v(x)\|_{L^2(\Omega)} + |v(1)| \\ &\leqslant \frac{\sqrt{5}}{5}\|v(x)\|_{L^2(\Omega)} + |v(1)| \end{aligned} \tag{4.83}$$

Or

$$|v(1)| = \Big|\int_0^1 v'(s)ds\Big| \leqslant \Big(\int_0^1 |v'(s)|^2 ds\Big)^{1/2}$$

par l'inégalité de Cauchy-Schwarz et donc $|v(1)| \leqslant \|v\|_{H^1(\Omega)}$ (on pourrait aussi appliquer le théorème de trace 3.1 100). On en déduit que

$$|L(v)| \leqslant (1 + \frac{\sqrt{5}}{5})\|v\|_{H^1(\Omega)}$$

Donc le théorème de Lax–Milgram s'applique (d'ailleurs, le théorème de Riesz suffit).

3. On cherche des fonctions de base dans l'espace $\mathcal{P}_2 = \{P : x \mapsto ax^2 + bx + c,\ a, b, c \in \mathbb{R}\}$ des polynômes de degré 2 sur $\mathbb{R}$. Sur l'élément de référence $[0\,;1]$, on considère les nœuds $a_1 = 0$, $a_2 = 1/2$ et $a_3 = 1$ et les degrés de liberté des trois fonctions de base associées à ces nœuds sont les valeurs aux nœuds. Soient ϕ_1, ϕ_2 et ϕ_3 les fonctions de base locales associées aux nœuds a_1, a_2 et a_3, on a donc

$$\phi_1(x) = 2(x - 1/2)(x - 1),$$
$$\phi_2(x) = 4x(1 - x),$$
$$\phi_3(x) = 2x(x - 1/2).$$

Donnons nous maintenant un maillage de $[0\,;1]$, défini par

$$x_0 = 0, \qquad x_{i+1/2} = (i + \frac{1}{2})h, \quad i = 1, \ldots, N, \qquad x_i = ih, \quad i = 1, \ldots, N$$

On a donc un nœud lié ($x_0 = 0$) et $2N$ nœuds libres. On rappelle qu'une fonction de base globale est par définition une fonction continue dont la restriction à un élément est une fonction de base locale. Par "recollement" des fonctions de base locales, on obtient l'expression des fonctions de base globales : Pour $i = 1$ à N, on a :

$$\phi_i(x) = \begin{cases} \dfrac{2}{h^2}(x - x_{i-1/2})(x - x_{i-1}) & \text{si } x \in [x_{i-1}, x_i] \\ \dfrac{2}{h^2}(x - x_{i+1/2})(x - x_{i+1}) & \text{si } x \in [x_i, x_{i+1}] \\ 0 & \text{sinon .} \end{cases}$$

et

$$\phi_{i+1/2}(x) = \begin{cases} -\dfrac{4}{h^2}(x - x_i)(x - x_{i+1}) & \text{si } x \in [x_{i-1}, x_i] \\ 0 & \text{sinon .} \end{cases}$$

Notons que ces fonctions de forme ne sont pas de classe C^1. Remarquons que Supp $\phi_i = [x_{i-1}, x_{i+1}]$ et *Supp* $\phi_{i+1/2} = [x_i, x_{i+1}]$, pour$i = 1, \ldots, N - 1$. On en déduit que

$$a(\phi_i, \phi_{j+1/2}) = 0 \qquad \text{si } j \neq i \text{ ou } j \neq i - 1$$
$$a(\phi_i, \phi_j) = 0 \qquad \text{si } j > i + 1 \text{ ou } j < i - 1$$
$$a(\phi_{i+1/2}, \phi_{j+1/2}) = 0 \quad \text{dès que } j \neq i$$

Pour obtenir le système linéaire à résoudre, on approche H par

$$H_N = Vect\,\{\phi_i, \phi_{i+1/2},\ i = 1, \ldots, N,\}$$

dans la formulation faible (4.81) et on développe u sur la base des fonctions $(\phi_i, \quad \phi_{i+1/2})_{i=1,\ldots,N}$.

On obtient donc :

$$\begin{cases} u = \sum_{i=1}^{N} u_i \phi_i + \sum_{i=1}^{N} u_{i+1/2} \phi_{i+1/2} \in H_N \\ \sum_{j=1}^{N} u_j a(\phi_j, \phi_i) + \sum_{j=1}^{N} u_{j+1/2} a(\phi_{j+1/2}, \phi_i) = L(\phi_i) & i = 1, \dots, N \\ \sum_{j=1}^{N} u_j a(\phi_j, \phi_{i+1/2}) + \sum_{j=1}^{N} u_{j+1/2} a(\phi_{j+1/2}, \phi_{i+1/2}) = L(\phi_{i+1/2}) & i = 1, \dots, N \end{cases}$$

Si on "range" les inconnues discrètes dans l'ordre naturel $1/2, 1, \frac{3}{2} \dots i, i+1/2, i+1 \dots, N$, on obtient un système linéaire pentadiagonal, de la forme : $AU = b$ avec

$$U = (u_{1/2}, u_1, u_{\frac{3}{2}}, \dots, u_i, u_{i+1/2}, u_{i+1}, \dots, u_N) \tag{4.84}$$

$$b = (b_{1/2}, b_1, b_{\frac{3}{2}}, \dots, b_i, b_{i+1/2}, b_{i+1}, \dots, b_N), \tag{4.85}$$

où

$$b_\alpha = \int_0^1 x^2 \phi_\alpha(x) \, \mathrm{d}x \qquad \text{si } \alpha < N \tag{4.86}$$

$$b_N = \int_0^1 x^2 \phi_N(x) \, \mathrm{d}x - 1, \tag{4.87}$$

et A est une matrice pentadiagonale (5 diagonales non nulles) de coefficients

$$A_{\alpha\beta} = a(\phi_\alpha, \phi_\beta) = \int_0^1 (\phi_\alpha(x)\phi_\beta(x) + \phi'_\alpha(x)\phi'_\beta(x)) \, \mathrm{d}x, \text{ avec } \alpha, \beta = 1/2, 1, \dots, N - 1/2, N.$$

Exercice 64. La formulation faible du problème s'écrit :

$$\begin{cases} \int_D \nabla u(x) \nabla v(x) \, \mathrm{d}x = \int_D f(x) v(x) \, \mathrm{d}x, \forall v \in H_0^1(\Omega) \\ u \in H_0^1(\Omega) \end{cases}$$

On note $I = \{(k, \ell), 1 \leqslant k \leqslant M, 1 \leqslant \ell \leqslant N\}$ noter que Card $I = MN$). L'espace vectoriel de dimension finie dans lequel on cherche la solution approchée (en utilisant les éléments finis suggérés par l'énoncé) est donc $H = Vect\ \{\phi_i, i \in I\}$, où ϕ_i est la fonction de base globale associée au nœud i. Cette solution approchée s'écrit $u = \sum_{j \in I} u_j \phi_j$ où la famille $\{u_j, j \in I\}$ est solution du système linéaire :

$$\sum_{j \in I} a_{ij} u_j = b_i, \forall i \in I \tag{4.88}$$

avec $b_i = \int_D f(x, y) \phi_i(x, y) \, \mathrm{d}x \, \mathrm{d}y$, pour tout $i \in I$ et $a_{ij} = \int_D \nabla \phi_i(x, y) \nabla \phi_j(x, y) \, \mathrm{d}x \, \mathrm{d}y$, pour tous $i, j \in I$.

La matrice de ce système linéaire est donc donnée par le calcul de a_{ij} pour $i, j \in I$ et un ordre de numérotation des inconnues, plus précisément, soit $\varphi : I \to \{1, \dots, MN\}$ bijective. On note ψ la fonction réciproque de φ. Le système (4.88) peut alors s'écrire :

$$\sum_{n=1}^{MN} a_{i,\psi(n)} u_{\psi(n)} = b_i, \ \forall i \in I$$

ou encore :

$$\sum_{n=1}^{MN} a_{\psi(m),\psi(n)} u_{\psi(n)} = b_{\psi(m)}, \forall m \in \{1, \dots, MN\},$$

La famille $\{u_j, j \in I\}$ est donc solution de (4.88) si et seulement si $u_{\psi(n)} = \lambda_n$ pour tout $n \in \{1, \dots, MN\}$ où $\lambda = (\lambda_1, \dots, \lambda_{MN}) \in \mathbb{R}^{MN}$ est solution du système linéaire :

$$A\lambda = C$$

avec $C = (C_1, \ldots, C_{MN})$, $C_m = b_{\psi(m)}$ pour tout $m \in \{1, \ldots, MN\}$ et $A = (A_{m,n})_{m,n=1}^{MN} \in \mathbb{R}^{MN}$ avec $A_{m,n} = a_{\psi(m),\psi(n)}$ pour tout $m, n \in \{1, \ldots, MN\}$. Il reste donc à calculer a_{ij} pour $i, j \in I$. Un examen du support des fonctions ϕ_i et ϕ_j et le fait que le maillage soit à pas constant nous montrent que seuls quatre nombres différents peuvent apparaître dans la matrice :

1. si $i = j$, on pose $a_{ii} = \alpha$.

2. si $i = (k, \ell), j = (k \pm 1, \ell)$, on pose alors $a_{ij} = \beta$.

3. si $i = (k, \ell), j = (k, \ell \pm 1)$, on pose alors $a_{ij} = \gamma$.

4. si $i = (k, \ell), j = (k + 1, \ell + 1)$ ou $(k - 1, \ell - 1)$, on pose alors $a_{ij} = \delta$.

En dehors des quatre cas décrits ci-dessus, on a nécessairement $a_{ij} = 0$ (car les supports de ϕ_i et ϕ_j sont disjoints). Calculons maintenant α, β, γ et δ.

Calcul de β On prend ici $i = (k, \ell)$ et $j = (k+1, \ell)$ On calcule tout d'abord $\int_{T^0} \nabla\phi_i \cdot \nabla\phi_j \, \mathrm{d}x$ avec $T^0 = T^0_{k+1/2,j+1/2}$. Un argument d'invariance par translation permet de supposer que $x_k = y_\ell = 0$. On a alors

$$\phi_i(x, y) = \frac{\Delta x - x}{\Delta x} \text{ et } \phi_j(x, y) = \frac{x\ \Delta y - y\ \Delta x}{\Delta x\ \Delta y},$$

de sorte que

$$\nabla\phi_i(x, y) \cdot \nabla\phi_j(x, y) = -\left(\frac{1}{\Delta x}\right)^2.$$

On a donc

$$\int_{T^0} \nabla\phi_i \cdot \nabla\phi_j \, \mathrm{d}x = -\left(\frac{1}{\Delta x}\right)^2 \frac{\Delta x\ \Delta y}{2} = -\frac{\Delta y}{2\Delta x}.$$

Un calcul similaire donne l'intégrale de $\nabla\phi_i \cdot \nabla\phi_j$ sur le deuxième triangle commun aux supports de ϕ_i et ϕ_j. Sur ce deuxième triangle, formé par les points $(k, \ell), k + 1, \ell)$ et $(k, \ell - 1)$, noté T^2, on a

$$\phi_i(x, y) = 1 - \frac{x\ \Delta y - y\ \Delta x}{\Delta x\ \Delta y} \text{ et } \phi_j(x, y) = \frac{x}{\Delta x},$$

de sorte que

$$\nabla\phi_i(x, y) \cdot \nabla\phi_j(x, y) = -\left(\frac{1}{\Delta x}\right)^2 \text{ et}$$

$$\int_{T^2} \nabla\phi_i(x, y) \cdot \nabla\phi_j(x, y)\ dx\, \mathrm{d}y = -\left(\frac{1}{\Delta x}\right)^2 \frac{\Delta x\ \Delta y}{2} = -\frac{\Delta y}{2\Delta x}.$$

On a donc, finalement,

$$\beta = \int_D \nabla\phi_i(x, y) \cdot \nabla\phi_j(x, y) \, \mathrm{d}x \, \mathrm{d}y = -\frac{\Delta y}{\Delta x}.$$

Calcul de γ Le calcul de γ est le même que celui de β en changeant les rôles de Δx et Δy, on obtient donc

$$\gamma = -\frac{\Delta x}{\Delta y}$$

Calcul de δ On prend ici $i = (k, \ell)$ et $j = (k + 1, \ell + 1)$. On a donc, en notant $T^0 = T^0_{k+1/2,\ell+1/2}$ et $T^1 = T^1_{k+1/2,\ell+1/2}$,

$$\delta = \int_{T^0} \nabla\phi_i(x, y) \cdot \nabla\phi_j(x, y)\ \mathrm{d}x\, \mathrm{d}y + \int_{T^1} \nabla\phi_i(x, y) \cdot \nabla\phi_j(x, y) \, \mathrm{d}x \, \mathrm{d}y.$$

On peut supposer (par translation) que $x_k = 0 = y_\ell$. Sur T_1 , on a alors $\phi_i(x, y) = \frac{\Delta y - y}{\Delta y}$ et $\phi_j(x, y) = \frac{x}{\Delta x}$ de sorte que $\int_{T^1} \nabla\phi_i(x, y) \cdot \nabla\phi_j(x, y)\ dx\, \mathrm{d}y = 0$ (car $\nabla\phi_i \cdot \nabla\phi_j = 0$). En changeant les rôles de x et y, on a aussi $\int_{T^0} \nabla\phi_i(x, y) \cdot \nabla\phi_j(x, y)\ dx\, \mathrm{d}y = 0$. On a donc $\delta = 0$.

Calcul de α On prend ici $i = j = (k, \ell)$. On peut toujours supposer que $x_k = y_\ell = 0$. En reprenant les notations précédentes, on a, par raison de symétrie :

$$\alpha = \int_D \nabla\phi_i \cdot \nabla\phi_i \,\mathrm{d}x = 2\int_{T^0} |\nabla\phi_i|^2(x,y)\,\mathrm{d}x\,\mathrm{d}y + 2\int_{T^1} |\nabla\phi_i|^2(x,y)\,\mathrm{d}x\,\mathrm{d}y + 2\int_{T^2} |\nabla\phi_i|^2(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Sur T^0, on a $\phi_i(x,y) = \frac{\Delta x - x}{\Delta x}$ et donc

$$\int_{T^0} |\nabla\phi_i|^2(x,y)\,\mathrm{d}x\,\mathrm{d}y = \left(\frac{1}{\Delta x}\right)^2 \frac{\Delta x \Delta y}{2} = \frac{1}{2}\,\frac{\Delta y}{\Delta x}$$

Sur T_1, on a $\phi_1(x,y) = \frac{\Delta y - y}{\Delta y}$ et donc

$$\int_{T^2} |\nabla\phi_i|^2(x,y)\; dx\,\mathrm{d}y = \left[\left(\frac{1}{\Delta x}\right)^2 + \left(\frac{1}{\Delta y}\right)^2\right] \frac{\Delta x\;\Delta y}{2} = \frac{1}{2}\,\frac{\Delta y}{\Delta x} + \frac{1}{2}\,\frac{\Delta x}{\Delta y}$$

On en déduit

$$\alpha = 2\frac{\Delta x}{\Delta y} + 2\frac{\Delta y}{\Delta x}.$$

Exercice 65.

1. Soit K le triangle de référence, de sommets (0,0), (1,0) et (0,1). On veut montrer que si p est un polynôme de degré 1, alors

$$\iint_K p(x,y)\;\,\mathrm{d}x\,\mathrm{d}y = \iint_K \mathrm{d}x\,\mathrm{d}y p(x_G, y_G) \tag{4.89}$$

où (x_G, y_G) est le centre de gravité de K. Comme K est le triangle de sommets (0,0), (1,0) et (0,1), on a $x_G = y_G = 1/3$. Pour montrer (4.89), on va le montrer pour $p \equiv 1$, pour $p(x,y) = x$ et pour $p(x,y) = y$. On a

$$\iint_K \mathrm{d}x\,\mathrm{d}y = \int_0^1 \int_0^{1-x} \mathrm{d}y\,\mathrm{d}x = \frac{1}{2}.$$

On a donc bien (4.89) si $p \equiv 1$. Et

$$\iint_K x\;\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 x \int_0^{1-x} \mathrm{d}y\,\mathrm{d}x = \int_0^1 (x - x^2)\,\mathrm{d}x = \frac{1}{6}$$

Or si $p(x,y) = x$, on a $p(x_G, y_G) = \frac{1}{3}$ et donc on a encore bien (4.89). Le calcul de $\int\int_K y\,\mathrm{d}x\,\mathrm{d}y$ est identique ; on a donc bien montré que l'intégration numérique à un point de Gauss est exacte pour les polynômes d'ordre 1.

2. On veut montrer que pour tout polynôme p de degré 2, on a :

$$\iint_K p(x,y)\,\mathrm{d}x\,\mathrm{d}y = L(p), \text{ où on a posé } L(p) = \frac{1}{6}\left(p\left(1/2, 0\right) + p\left(1/2, 1/2\right) + p\left(0, 1/2\right)\right) \tag{4.90}$$

On va démontrer que (4.90) est vérifié pour tous les monômes de $\mathcal{P}_2$. Si $p \equiv 1$, on a $L(p) = 1/2$, et (4.90) est bien vérifiée. Si $p(x,y) = x$, on a $L(p) = 1/6$ et on a vu à la question 1 que

$$\iint_K x\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{6}.$$

On a donc bien (4.90). Par symétrie, si $p(x,y) = y$ vérifie aussi (4.90). Calculons maintenant

$$I = \iint_K xy\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \times \int_0^{1-x} y\,\mathrm{d}y\,\mathrm{d}x$$

On a donc

$$I = \int_0^1 \times \frac{(1-x)^2}{2}\,\mathrm{d}x = \frac{1}{2}\int_0^1 (x - 2x^2 + x^3)\,\mathrm{d}x = \frac{1}{24},$$

et si $p(x,y) = xy$, on a bien $L(p) = 1/6 \times 1/4$ et (4.90) est vérifiée. Il reste à vérifier que (4.90) est valide pour $p(x,y) = x^2$ (ou $p(x,y) = y^2$, par symétrie). Or,

$$J = \iint_K x^2 \,\mathrm{d}x\,\mathrm{d}y = \int_0^1 x^2 \int_0^{1-x} \mathrm{d}y\,\mathrm{d}x = \int_0^1 (x^2 - x^3)dx$$

donc $J = 1/3 - 1/4 = 1/12$. Et pour $p(x,y) = x^2$, on a bien :

$$L(p) = \frac{1}{6}\left(\frac{1}{4} + \frac{1}{4}\right) = \frac{1}{12}$$

Exercice 66.

1. Comme $p \in \mathcal{P}_2$, p est de la forme $p(x,y) = a + bx + cy + dxy + \alpha x^2 + \beta y^2$, on a par développement de Taylor (exact car $p''' = 0$) :

$$2p(a_9) - p(a_6) - p(a_8) = \frac{1}{4}p_{xx}(a_9) = \alpha,$$
$$2p(a_9) - p(a_5) - p(a_7) = \frac{1}{4}p_{yy}(a_9) = \beta,$$

d'où on déduit que

$$4p(a_9) - \sum_{i=5}^{8} p(a_i) = \alpha + \beta. \tag{4.91}$$

De même, on a :

$$2p(a_5) - p(a_1) - p(a_2) = \alpha$$
$$2p(a_7) - p(a_3) - p(a_4) = \alpha$$
$$2p(a_6) - p(a_2) - p(a_3) = \beta$$
$$2p(a_8) - p(a_1) - p(a_4) = \beta.$$

Ces quatre dernières égalités entraînent :

$$\sum_{i=5}^{8} p(a_i) - \sum_{i=1}^{4} p(a_i) = \alpha + \beta \tag{4.92}$$

De (4.91) et (4.92), on déduit que :

$$\sum_{i=1}^{4} p(a_i) - 2\sum_{i=5}^{8} p(a_i) + 4p(a_9) = 0.$$

2. La question précédente nous suggère de choisir $\phi : \mathcal{Q}_2 \to \mathbb{R}$ définie par

$$\phi(p) = \sum_{i=1}^{4} p(a_i) - 2\sum_{i=5}^{8} p(a_i) + 4p(a_9).$$

Soit $p \in \mathcal{P}$ tel que $p(a_i) = 0$, $i = 1, \ldots, 8$. Comme $p \in \mathcal{Q}_2$, p est une combinaison linéaire des fonctions de base $\varphi_1, \ldots, \varphi_9$, associées aux nœuds $a_1, \ldots, a_9$, et comme $p(a_i) = 0$, $i = 1, \ldots, 8$, on en déduit que $p = \alpha\varphi_9, \alpha \in \mathbb{R}$. On a donc $\phi(p) = \alpha\phi(\varphi_9) = 4\alpha = 0$, ce qui entraîne $\alpha = 0$. On a donc $p = 0$.

3. Calculons les fonctions de base $\varphi_1^*, \ldots, \varphi_8^*$ associées aux nœuds $a_1, \ldots, a_8$ qui définissent Σ. On veut que $\varphi_i^* \in \mathcal{P}$ et $\varphi_i^*(a_j) = \delta_{ij}$ pour $i, j = 1, \ldots, 8$. Or $\varphi_9(a_j) = 0$, $\forall i = 1, \ldots, 8$ et $\phi(\varphi_9) = 4$. Remarquons alors que pour $i = 1$ à 4 on a $p(\varphi_i) = 1$ et donc si $\varphi_i^* = \varphi_i - \varphi_9/4$ on a $p(\varphi_i^*) = 0$ et $\varphi_i^*(a_j) = \delta_{ij}$ pour $j = 1, \ldots, 8$. De même, pour $i = 5$ à 8, on a $p(\varphi_i) = -2$ et donc si $\varphi_i^* = \varphi_i + \varphi_9/2$, on a $p(\varphi_i^* = 0$ et $\varphi_i^*(a_j) = \delta_{ij}$ pour $j = 1, \ldots, 8$. On a ainsi trouvé les fonctions de base de l'élément fini $(C, \mathcal{P}, \Sigma)$. Notons que cet élément fini n'est autre que l'élément fini (C, Q_2^*, Σ) vu en cours (voir paragraphe 4.2.3) et que $\ker\phi = \mathcal{P} = Q_2^*$.

5

Problèmes hyperboliques

5.1 L'équation de transport

L'exemple le plus simple d'équation hyperbolique est une équation linéaire qu'on appelle équation de transport. Supposons par exemple, que l'on connaisse l'emplacement d'une nappe de pétrole due au dégazement intempestif d'un supertanker au large des côtes et que l'on cherche à prévoir son déplacement dans les heures à venir, par exemple pour la mise en œuvre efficace de barrages. On suppose connu $v : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}^2$, le champ des vecteurs vitesse des courants marins, qui dépend de la variable d'espace x et du temps t ; ce champ de vecteurs est donné par exemple par la table des marées (des exemples de telles cartes de courants sont données en Figure 5.1). À $t = 0$, on connaît

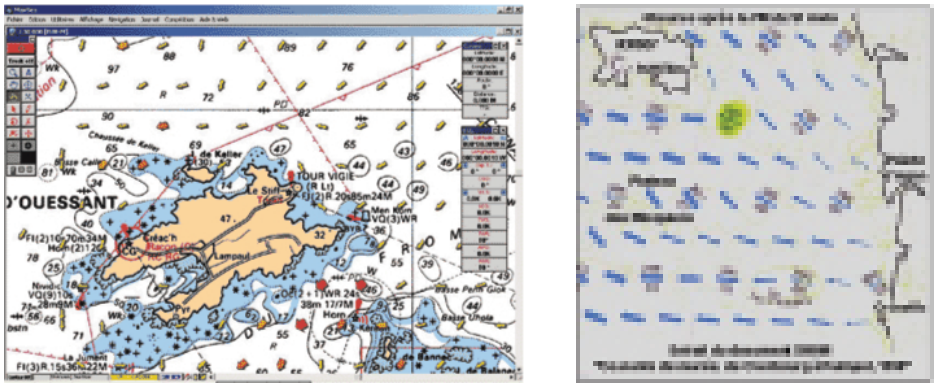


FIGURE 5.1 – Cartes de courants marins au large de côtes de Bretagne — source : SHOM

$\rho_0(x)$, la densité d'hydrocarbure initiale et on cherche à calculer la densité $\rho(x,t)$ d'hydrocarbure au point x et au temps t. On écrit alors l'équation de conservation de la masse :

$$\rho_t + \operatorname{div}(\rho v) = 0, \tag{5.1}$$

avec :

$$\rho_0(x) = \begin{cases} 1 & x \in A \\ 0 & x \in A^c \end{cases} \tag{5.2}$$

où A représente le lieu initial de la nappe de pétrole. Dans le cas d'un déplacement maritime, le vecteur $v : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}^2$ n'est évidemment pas constant (la marée n'est pas la même à Brest qu'à Saint Malo). De plus, le déplacement de la nappe dépend également du vent qui affecte donc le vecteur v. On supposera pourtant ici, pour simplifier l'exposé, que v est constant en espace et en temps. Alors le problème (5.1)-(5.2) admet comme solution :

$$\rho(x,t) = \rho_0(x - vt) \tag{5.3}$$

qui exprime le transport de la nappe à la distance vt du point de départ dans la direction de V, au temps t. En fait, il est clair que (5.3) n'est pas une solution *classique* de : comme ρ_0 n'est pas continue, la fonction ρ définie par (5.3) ne n'est pas non plus et ses dérivées partielles ne sont donc pas définies au sens classique.

Nous verrons par la suite comment on peut donner une formulation correcte des solutions de (5.1)-(5.2). Notons que les systèmes d'équations hyperboliques sont très importants en mécanique des

fluides ; les équations d'Euler, par exemple sont utilisées pour modéliser l'écoulement de l'air autour d'une aile d'avion.

Dans le cadre de ce cours, nous n'étudierons cependant que le cas des équations scalaires en une dimension d'espace, tout d'abord dans le cas relativement simple d'une équation linéaire (paragraphes 5.2 et 5.3), puis dans le cas nettement plus difficile d'une équation non linéaire (paragraphes 5.4 et 5.5).

5.2 Solutions classiques et solutions faibles, cas linéaire

Commençons par étudier le cas d'une équation hyperbolique linéaire :

$$\begin{cases} \partial_t u + c\partial_x u = 0 & x \in \mathbb{R} \quad t > 0 \\ u(x,0) = u_0(x) & x \in \mathbb{R}. \end{cases} \tag{5.4}$$

où la vitesse de transport $c \in \mathbb{R}$ et la condition initiale $u_0 : \mathbb{R} \to \mathbb{R}$ sont données. Le problème (5.4) s'appelle *problème de Cauchy*. On cherche $u : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$, solution de ce problème. Nous commençons par une étude succincte du problème continu, pour lequel on peut trouver une solution exacte explicite.

Définition 5.1 — Solution classique. On dit qu'une fonction $u : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ est solution classique du problème (5.4) si $u \in \mathcal{C}^1(\mathbb{R} \times \mathbb{R}_+, \mathbb{R})$ et u vérifie (5.4).

Une condition nécessaire pour avoir une solution classique est que $u_0 \in \mathcal{C}^1(\mathbb{R})$.

Théorème 5.1 Si $u_0 \in \mathcal{C}^1(\mathbb{R})$, alors il existe une unique solution classique du problème (5.4), qui s'écrit $u(x,t) = u_0(x - ct)$.

Démonstration. Pour montrer l'existence de la solution, il suffit de remarquer que u définie par (5.1) est de classe $\mathcal{C}^1$ et que $\partial_t u + c\partial_x u = 0$ en tout point. Pour montrer l'unicité de la solution, on va introduire la notion de caractéristique, qui est d'ailleurs aussi fort utile dans le cadre de la résolution numérique. Soit u solution classique de (5.4). On appelle droite caractéristique de (5.4) issue de x_0 la droite d'équation $x(t) = ct + x_0$, qui est illustrée sur la figure 5.2. Montrons que si u est solution de (5.4) alors u est constante sur la droite $\mathcal{D}_{x_0}$, pour tout $x_0 \in \mathbb{R}$. Soit $x_0 \in \mathbb{R}$ et φ_{x_0} la fonction de $\mathbb{R}_+$ dans $\mathbb{R}$ définie par $\varphi_{x_0}(t) = u(x_0 + ct, t)$. Dérivons φ_{x_0}

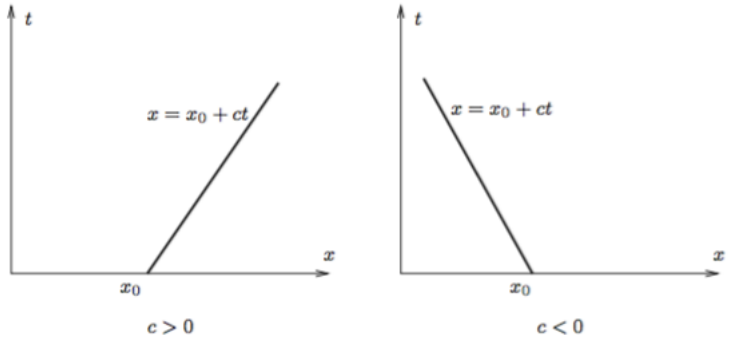


Figure 5.2 – Droites caractéristiques, cas linéaire

par rapport au temps $\varphi'_{x_0}(t) = c\partial_x u(x_0 + ct, t) + \partial_t u(x_0 + ct, t) = (\partial_t u + c\partial_x u)(x_0 + ct, t) = 0$ car u est solution de (5.4). On en déduit que $\varphi_{x_0}(t) = \varphi_{x_0}(0) = u_0(x_0) \quad \forall t \in \mathbb{R}_+$. On a donc $u(x_0 + ct, t) = u_0(x_0), \forall x_0 \in \mathbb{R}$, donc u est constante sur la droite caractéristique $\mathcal{D}_{x_0}$ et en posant $x = x_0 + ct$, $u(x,t) = u_0(x - ct)$ ce qui prouve l'existence et l'unicité de (5.4). •

Remarque 5.2 — Terme source. Le modèle physique peut amener à une équation avec terme source au second membre $f \in \mathcal{C}(\mathbb{R} \times \mathbb{R}_+, \mathbb{R})$:

$$\begin{cases} \partial_t u + c\partial_x u = f(x,t) \\ u(x,0) = u_0(x) \end{cases} \tag{5.5}$$

et $u_0 \in \mathcal{C}^1(\mathbb{R})$. Ceci peut modéliser un dégazage sur un temps plus long, comme dans le cas du Prestige sur les côtes de Galice en 2003 par exemple. Pour montrer l'unicité de la solution de (5.5), on suppose que u est solution classique et on pose $\varphi_{x_0}(t) = u(x_0 + ct, t)$. Par un calcul identique au précédent, on a $\varphi'_{x_0}(t) = f(x_0 + ct, t)$ donc :

$$\varphi_{x_0}(t) = \varphi_{x_0}(0) + \int_0^t f(x_0 + cs, s)\,\mathrm{d}s$$

On en déduit que :

$$u(x_0 + ct, t) = \varphi_{x_0}(0) + \int_0^t f(x_0 + cs, s)\,\mathrm{d}s$$

on pose alors $x = x_0 + ct$ et on obtient :

$$u(x,t) = u_0(x - ct) + \int_0^t f(x - c(t-s), s)\,\mathrm{d}s,$$

ce qui prouve l'unicité. On obtient alors l'existence en remarquant que la fonction $u(x,t)$ ainsi définie est effectivement solution de (5.5), car elle est de classe $\mathcal{C}^1$ et elle vérifie $\partial_t u + c\partial_x u = f$.

Dans ce qui précède, on a fortement utilisé le fait que u_0 est $\mathcal{C}^1$. Ce n'est largement pas toujours le cas dans la réalité. Que faire si, par exemple, $u_0 \in L^\infty(\mathbb{R})$?

Définition 5.3 — Solution faible. On dit que u est solution faible de (5.4) si $u \in L^\infty(\mathbb{R} \times \mathbb{R}_+, \mathbb{R})$ et u vérifie :

$$\int_{\mathbb{R}} \int_{\mathbb{R}_+} \big(u(x,t)\varphi_t(x,t) + cu(x,t)\varphi_x(x,t)\big)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}} u_0(x)\varphi(x,0)\,\mathrm{d}x = 0 \quad \forall \varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}_+, \mathbb{R}) \tag{5.6}$$

Notons que dans la définition ci-dessus, on note $\mathbb{R}_+ = [0\,;+\infty[$ et $\mathcal{C}^1_c(\mathbb{R} \times \mathbb{R}_+)$ l'ensemble des restrictions à $\mathbb{R} \times \mathbb{R}_+$ des fonctions $\mathcal{C}^1_c(\mathbb{R} \times \mathbb{R})$. On insiste sur le fait qu'on peut donc avoir $\varphi(x,0) \neq 0$. Voyons maintenant les liens entre solution classique et solution faible.

Proposition 5.4 Si u est solution classique de (5.4) alors u est solution faible. Réciproquement, si $u \in \mathcal{C}^1(\mathbb{R} \times]0\,;+\infty[) \cap \mathcal{C}(\mathbb{R} \times [0\,;+\infty[)$ est solution classique de (5.20) alors u est solution forte de (5.4).

La démonstration de cette proposition est effectuée dans le cadre plus général des équations hyperboliques non linéaires [Proposition 5.15]. Notons que si on prend $\varphi \in \mathcal{C}^1_c(\mathbb{R} \times]0\,;+\infty[, \mathbb{R})$ au lieu de $\varphi \in \mathcal{C}^1_c(\mathbb{R} \times [0\,;+\infty[, \mathbb{R})$ dans (5.6), on obtient $\partial_t u + c\partial_x u = 0$ mais on ne récupère pas la condition initiale. Il est donc essentiel de prendre des fonctions test dans $\mathbb{C}_c(\mathbb{R} \times [0\,;+\infty[, \mathbb{R})$.

Théorème 5.2 — Existence et unicité de la solution faible. Si $u_0 \in L^\infty_{loc}(\mathbb{R})$, il existe une unique fonction u solution faible de (5.4).

Démonstration. On va montrer que $u(x,t) = u_0(x - ct)$ est solution faible. Comme $u_0 \in L^\infty_{loc}(\mathbb{R})$, on a $u \in L^\infty(\mathbb{R} \times \mathbb{R}_+)$. Soit $\varphi \in \mathcal{C}^1_c(\mathbb{R} \times \mathbb{R}_+, \mathbb{R})$, on veut montrer que :

$$\iint_{\mathbb{R}\times\mathbb{R}_+} u(x,t)\varphi_t(x,t)\,\mathrm{d}x\,\mathrm{d}t + \iint_{\mathbb{R}\times\mathbb{R}_+} cu(x,t)\varphi_x(x,t)\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}} u_0(x)\varphi(x,0)\,\mathrm{d}x = 0.$$

Posons

$$A = \iint_{\mathbb{R}\times\mathbb{R}_+} u(x,t)\varphi_t(x,t)\,\mathrm{d}x\,\mathrm{d}t + \iint_{\mathbb{R}\times\mathbb{R}_+} cu(x,t)\varphi_x(x,t)\,\mathrm{d}x\,\mathrm{d}t$$

Si $u(x,t) = u_0(x-ct)$, on a donc :

$$A = \iint_{\mathbb{R}\times\mathbb{R}_+} \big(u_0(x-ct)\varphi_t(x,t) + cu_0(x-ct)\varphi_x(x,t)\big)\,\mathrm{d}x\,\mathrm{d}t$$

En appliquant le changement de variable $y = x - ct$ et en utilisant le théorème de Fubini, on obtient :

$$A = \int_{\mathbb{R}} u_0(y) \int_{\mathbb{R}_+} \big(\varphi_t(y+ct,t) + c\varphi_x(y+ct,t)\big)\,\mathrm{d}t\,\mathrm{d}y.$$

Posons alors $\psi_y(t) = \varphi(y+ct,t)$. On a donc :

$$A = \int_{\mathbb{R}} \left(u_0(y) \int_0^{+\infty} \psi_y'(t)\,\mathrm{d}t \right) \mathrm{d}y,$$

et comme ψ est à support compact sur $[0\,;+\infty[$,

$$A = -\int_{\mathbb{R}} u_0(y)\psi_y(0)\,\mathrm{d}y$$

donc finalement :

$$A = -\int_{\mathbb{R}} u_0(y)\varphi(y,0)\,\mathrm{d}y.$$

On a ainsi démontré que la fonction u définie par $u(x,t) = u_0(x-ct)$ est solution faible de l'équation (5.4). On a donc existence d'une solution faible. Montrons maintenant que celle-ci est unique. Soient u et v deux solutions faibles de (5.4). On pose $w = u - v$ et on va montrer que $w = 0$. Par définition, w satisfait :

$$\iint_{\mathbb{R}\times\mathbb{R}_+} w(x,t)(\varphi_t(x,t) + c\varphi_x(x,t))\,\mathrm{d}x\,\mathrm{d}t = 0, \quad \forall \varphi \in \mathcal{C}_c^1(\mathbb{R}\times\mathbb{R}^+,\mathbb{R}) \tag{5.7}$$

Par le lemme 5.5 donné ci-dessous, pour toute fonction $f \in C_c^\infty(\mathbb{R}\times\mathbb{R}_+^*)$ il existe $\varphi \in \mathcal{C}_c^1(\mathbb{R}\times\mathbb{R}_+,\mathbb{R})$, telle que $\varphi_t + c\varphi_x = f$ et on a donc par (5.7) :

$$\iint_{\mathbb{R}\times\mathbb{R}_+} w(x,t)f(x,t)\,\mathrm{d}x\,\mathrm{d}t = 0 \quad \forall f \in \mathcal{C}_c^1(\mathbb{R}\times\mathbb{R}_+^*,\mathbb{R})$$

Ceci entraîne que $w = 0$ p.p.. •

Lemme 5.5 — Résultat d'existence. Soit $f \in \mathcal{C}_c(\mathbb{R}\times\mathbb{R}_+^*,\mathbb{R})$, alors il existe $\varphi \in \mathcal{C}_c^1(\mathbb{R}\times\mathbb{R}_+,\mathbb{R})$ telle que $\varphi_t + c\varphi_x = f$.

Démonstration. Soit $f \in \mathcal{C}_c(\mathbb{R}\times\mathbb{R}_+^*,\mathbb{R})$ et $T > 0$ tel que $f(x,t) = 0$ si $t \geqslant T$. On considère le problème :

$$\begin{cases} \varphi_t + c\varphi_x = f \\ \varphi(x,T) = 0 \end{cases} \tag{5.8}$$

On vérifie facilement que le problème (5.8) admet une solution classique

$$\varphi(x,t) = -\int_t^T f(x-c(s-t),s)\,\mathrm{d}s$$

En effet, avec ce choix de φ, on a effectivement $\varphi \in \mathcal{C}_c^1(\mathbb{R}\times\mathbb{R}_+,\mathbb{R})$ et $\varphi_t + c\varphi_x = f$. De plus, comme f est à support compact, φ est à support compact. •

Remarque 5.6 — Sur les propriétés de la solution. Remarquons que la solution faible de (5.4) possède les propriétés suivantes :

1. Si $u_0 \geqslant 0$ p.p. alors $u \geqslant 0$ p.p. ;
2. $\|u(\,\cdot\,,t)\|_{L^p(\mathbb{R})} = \|u_0(x)\|_{L^p(\mathbb{R})}$, $\forall p \in [1\,;+\infty]$.

Lors de l'élaboration de schémas numériques pour la recherche d'une approximation, on s'attachera à vérifier que ces propriétés sont encore satisfaites par la solution approchée.

5.3 Schémas — cas linéaire

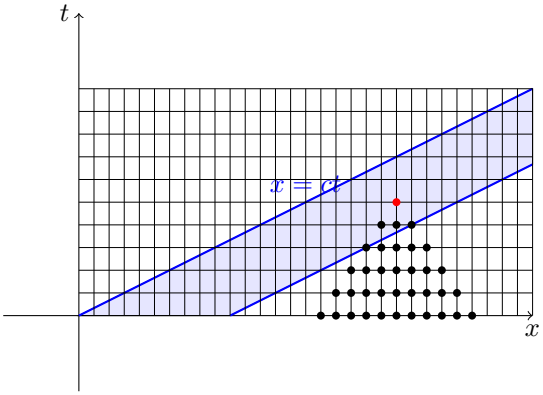
Soit $c \in \mathbb{R}_+^*$. On considère le problème de transport linéaire (5.4) avec $c = 1$:

$$\begin{cases} \partial_t u + c\partial_x u = 0 \\ u(x,0) = u_0(x) \quad u_0 \in L^\infty(\mathbb{R}). \end{cases} \tag{5.9}$$

On sait que la solution de ce problème s'écrit $u(x,t) = u_0(x-t)$. On rappelle que u est une solution classique si $u \in \mathcal{C}^1(\mathbb{R})$ et que u est une solution faible si $u_0 \in L^\infty(\mathbb{R})$. On va chercher à retrouver cette solution par une approximation numérique. Notons que dans le cas linéaire, l'utilisation d'un schéma numérique n'est évidemment pas utile puisque l'on connaît la solution exacte, mais nous commençons par ce cas par souci pédagogique.

5.3.1 La condition de Courant-Friedrichs-Lewy

Dans un article très célèbre publié en 1928, Richard Courant [1], Kurt Friedrichs [2] et Hans Lewy [3] démontrent l'existence de solutions à des équations aux dérivées partielles elliptiques et hyperboliques en passant à la limite sur ce qu'ils appellent à l'époque *équations aux différences*



5.3.2 Schéma explicite différences finies centrées

On choisit maintenant $c = 1$ dans (5.9), et on effectue une discrétisation espace temps en se donnant un pas de discrétisation en espace h et en posant : $x_i = ih$, $\forall i \in \mathbb{Z}$; de même on se donne un pas de discrétisation en temps k et on pose $t_n = nk$, $\forall n \in \mathbb{N}$. écrivons le schéma d'Euler explicite pour l'approximation de $\partial_t u$ et un schéma centré pour l'approximation de $\partial_x u$. On approche $\partial_t u(x_i, t_n)$ par $(u(x_i, t_{n+1}) - u(x_i, t_n))/k$ et $\partial_x u(x_i, t_n)$ par $(u(x_{i+1}, t_n) - u(x_{i-1}, t_n))/2h$. Le schéma centré s'écrit donc, en fonction des inconnues discrètes :

$$\begin{cases} \dfrac{u_i^{n+1} - u_i^n}{k} + \dfrac{u_{i+1}^n - u_{i-1}^n}{2h} = 0, \\ u_i^0 = u_0(x_i) \end{cases} \tag{5.10}$$

où on a supposé $u_0 \in \mathcal{C}^1(\mathbb{R})$. Ce schéma est *inconditionnellement instable* et il faut donc éviter de l'utiliser. Qu'entend-on par instable ? On peut montrer que :

1. Le schéma (5.10) ne respecte pas la positivité, car $u_0(x) \geqslant 0$, $\forall x$ n'entraine pas forcément $u_i^n \geqslant 0$. En effet si u_0 est telle que

$$\begin{cases} u_i^0 = 0 & \forall i \leqslant 0 \\ u_i^0 = 1 & \forall i > 0 \end{cases}$$

Alors :

$$u_i^{n+1} = u_i^n - \frac{k}{2h}(u_{i+1}^n - u_{i-1}^n)$$

1. mathématicien allemand naturalisé américain (1888–1972). Elève de D. Hilbert, il a émigré aux USA en 1933, fondateur du Courant Institute of Mathematics à NYU, il est célèbre pour ses travaux en mathématiques pour la physique.

2. mathématicien allemand naturalisé américain (1901–1982) spécialiste d'équations aux dérivées partielles et de méthodes numériques, cofondateur du Courant Institute de l'université de New York et détenteur de la National Medal of Science.

3. mathématicien allemand naturalisé américain (1904–1988) connu pour ses travaux sur les équations aux dérivées partielles. Il reçoit le prix Wolf de mathématiques en 19841.

donne, pour $n = 0$

$$u_0^1 = -\frac{k}{2h} < 0$$

2. Le schéma (5.10) n'est pas L^∞ stable puisque $\|u^n\|_\infty \leqslant C$ n'entraîne pas $\|u^{n+1}\|_\infty \leqslant C$;
3. Le schéma (5.10) n'est pas L^2 stable puisque $\|u^n\|_2 \leqslant C$ n'entraîne pas que $\|u^{n+1}\|_2 \leqslant C$;
4. Le schéma n'est pas stable au sens de Von Neumann. En effet, si $u_0(x) = \mathrm{e}^{ipx}$ où $i^2 = -1$ et $p \in \mathbb{Z}$, la solution exacte est $u(x,t) = \mathrm{e}^{ip(x-t)}$. Une discrétisation de u_0 s'écrit $u_j^0 = \mathrm{e}^{ipjh}$, $j \in \mathbb{Z}$. On a donc :

$$u_j^1 = u_j^0 - \frac{k}{2h}(u_{j+1}^0 - u_{j-1}^0) = \mathrm{e}^{ipjh} - \frac{k}{2h}(\mathrm{e}^{ip(j+1)h} - \mathrm{e}^{ip(j-1)h}) = \mathcal{J}_{k,h} u_j^0$$

avec $\mathcal{J}_{kh} = 1 - \frac{ik}{h}\sin ph$. On a donc $|\mathcal{J}_{kh}| > 1$ si $\sin ph \neq 0$, ce qui montre que le schéma n'est pas stable au sens de Von Neumann.

5. Le schéma (5.10) n'est pas convergent. En effet, on peut montrer qu'il existe u_0, k et h telle que la solution approchée $u_{h,k} : (u_i^n)_{i\in\mathbb{Z}}^{n\in\mathbb{N}}$ ne converge pas vers u lorsque h et k tendent vers 0.

5.3.3 Schéma différences finies décentré amont

On utilise toujours le schéma d'Euler explicite pour la discrétisation en temps, mais on approche maintenant $\partial_x u(x_i, t_n)$ par

$$\frac{u(x_i,t_n) - u(x_{i-1},t_n)}{h_{i-1/2}}$$

On considère de plus un pas de discrétisation variable, défini par $h_{i-1/2} = x_i - x_{i-1}$. Le schéma

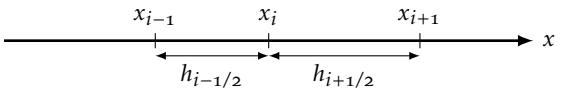


Figure 5.3 – Maillage volumes finis

par différences finies avec décentrement amont s'écrit :

$$\begin{cases} \dfrac{u_i^{n+1} - u_i^n}{k} + \dfrac{u_i^n - u_{i-1}^n}{h_{i-1/2}} = 0 \\ u(x,0) = u_0(x) \end{cases} \tag{5.11}$$

Proposition 5.7 — Stabilité du schéma décentré amont. Le schéma (5.11) est stable sous condition de Courant-Friedrichs-Levy (CFL)

$$k \leqslant \underline{h} = \inf_{i\in\mathbb{Z}} h_{i-1/2}, \tag{5.12}$$

c'est-à-dire que si $A \leqslant u_i^n \leqslant B$, alors $A \leqslant u_i^{n+1} \leqslant B$.

Démonstration. On a $u_i^{n+1} = u_i^n(1-\alpha_i) + \alpha_i u_{i-1}^n$ avec $\alpha_i = k/h_{i-1/2}$. Donc, si la condition (5.12) est vérifiée, u_i^{n+1} est une combinaison convexe de u_i^n et u_i^{n+1} et donc $u_i^{n+1} \in [u_{i-1}^n, u_i^n]$. •

Théorème 5.3 — Convergence du schéma décentré amont. On suppose que $u_0 \in \mathcal{C}^2(\mathbb{R},\mathbb{R})$ et que u_0, u_0' et u_0'' sont bornées. Soit $A = \inf_{x\in\mathbb{R}} u_0(x)$ et $B = \sup_{x\in\mathbb{R}} u_0(x)$. Soit $(u_i^n)_{i\in\mathbb{Z}\ nk\leqslant T}$ la solution du schéma (5.11) alors :

$$A \leqslant u_i^n \leqslant B \quad \forall i \in \mathbb{Z} \quad \forall n \in \mathbb{N}$$

Soit $\bar{u}_i^n = u(x_i, t_n)$, à u est la solution exacte de (5.9), alors :

$$\sup_{i\in\mathbb{Z},\, nk\leqslant T} |u_i^n - \bar{u}_i^n| \leqslant T C_{u_0}(k+h), \tag{5.13}$$

où $C_{u_0} \geqslant 0$ ne dépend que de u_0.

Démonstration. Le point 1 se démontre par récurrence sur n en utilisant le résultat de stabilité de la proposition 5.7. Le point 2 (estimation d'erreur) se démontre en introduisant l'erreur de consistance

$$R_i^n = \frac{\bar{u}_i^{n+1} - \bar{u}_i^n}{k} + \frac{\bar{u}_i^n - \bar{u}_{i-1}^n}{h_{i-1/2}}, \tag{5.14}$$

et en remarquant que celle-ci vérifie :

$$|R_i^n| \leqslant C_{u_0}(h+k), \tag{5.15}$$

où $h = \max_{i\in\mathbb{Z}} h_{i-1/2}$. On peut alors réécrire (5.14) de la manière suivante :

$$\bar{u}_i^{n+1} = \bar{u}_i^n\Big(1 - \frac{k}{h_{i-1/2}}\Big) + \frac{k}{h_{i-1/2}}\bar{u}_{i-1}^n + kR_i^n,$$

et en retranchant cette égalité au schéma numérique, qui s'écrit :

$$u_i^{n+1} = u_i^n\Big(1 - \frac{k}{h_{i-1/2}}\Big) + \frac{k}{h_{i-1/2}}u_{i-1}^n$$

on obtient :

$$u_i^{n+1} - \bar{u}_i^{n+1} = (u_i^n - \bar{u}_i^n)\Big(1 - \frac{k}{k_{i-1/2}}\Big) + (u_{i-1}^n - \bar{u}^n i - 1)\frac{k}{h_{i-1/2}} - kR_i^n$$

Grâce à la condition de CFL 5.12, les coefficients et sont positifs ou nuls. En utilisant la consistance du schéma (5.15), on obtient donc :

$$|u_i^{n+1} - \bar{u}_i^{n+1}| \leqslant |u_i^n - \bar{u}_i^n|\Big(1 - \frac{k}{h_{i-1/2}}\Big) + |u_{i-1}^n - \bar{u}_{i-1}^n|\frac{k}{h_{i-1/2}} + kC_{u_0}(\bar{h}+k) \tag{5.16}$$

On effectue alors l'hypothèse de récurrence :

$$\sup|u_i^p - \bar{u}_i^p| \leqslant pkC_{u_0}(k+h), \tag{5.17}$$

qui est vérifiée pour $k = 0$. Grâce à (5.16) et (5.17), on obtient :

$$|u_i^{p+1} - \bar{u}_i^{p+1}| \leqslant pkC_{u_0}(k+h) + k(C_{u_0}(k+h)$$

soit, finalement $|u_i^{p+1} - \bar{u}_i^{p+1}| \leqslant (p+1)kC_{u_0}(k+h)$. On a ainsi démontré l'estimation d'erreur (5.13). •

Remarque 5.8 — Décentrement. Pour une équation de transport telle que (5.9), le choix du décentrement est crucial. Ici, on a approché $\partial_x u(x_i)$ par $\frac{u_i - u_{i-1}}{h_{i-1/2}}$. Dans le cas où on étudie une équation de transport de type $\partial_t u + c\partial_x u = 0$ avec $c \in \mathbb{R}$, le choix décentré amont sera toujours :

$$\frac{u_i - u_{i-1}}{h} \quad \text{si } c > 0$$

par contre, si $c < 0$, le choix amont donnera

$$\frac{u_i - u_{i+1}}{h}$$

Regardons ce qui se passe si l'on effectue un *mauvais* décentrement. Considérons toujours l'équation $\partial_t u + \partial_x u = 0$. Effectuer le *mauvais décentrement* amène au schéma :

$$\frac{u_i^{n+1} - u_i^n}{k} + \frac{u_{i+1}^n - u_i^n}{h} = 0 \quad \Leftrightarrow \quad u_i^{n+1} = u_i^n\Big(1 + \frac{k}{h}\Big) - \frac{k}{h}u_{i+1}^n.$$

Examinons le comportement de la solution approchée donnée par le schéma si on prend une condition initiale u_0 telle que $u_0(x) = 0$, $\forall x \geqslant 0$. Dans ce cas, on sait que $u(x,t) \neq 0$ pour t assez grand, or après calculs on obtient :

$$u_{-1}^{n+1} = u_{-1}^n\Big(1 + \frac{k}{h}\Big) + 0 = u_{-1}^0\Big(1 + \frac{k}{h}\Big)^n$$

alors que $u_i^{n+1} = 0$, $\forall i \geqslant 0$. On en déduit que la solution approchée est très mauvaise.

Remarque 5.9 — Équation non linéaire avec donnée initiale.

1. Dans le cas non linéaire, la démonstration précédente de convergence ne s'adapte pas car les solutions ne sont pas régulières.
2. On a défini (5.11) pour $u_0 \in \mathcal{C}(\mathbb{R})$. Si $u_0 \notin \mathcal{C}(\mathbb{R})$, on peut prendre comme donnée initiale :

$$u_i^0 = \int_{x_{i-1/2}}^{x_{i+1/2}} u_0(x)\,\mathrm{d}x$$

5.3.4 Schéma volumes finis décentré amont

On considère toujours le problème (5.9), avec condition initiale $u_0 \in L^\infty(\mathbb{R})$. On se donne une discrétisation en espace, c'est-à-dire un ensemble de points $(x_{i+1/2})_{i\in\mathbb{Z}}$, tels que $x_{i+1/2} > x_{i-1/2}$ et on note $h_i = x_{i+1/2} - x_{i-1/2}$. On approche toujours la dérivée en temps par un schéma d'Euler explicite, on intègre (5.9) sur la maille $]x_{i-1/2}\,;x_{i+1/2}[$ et on obtient :

$$\int_{x_{i-1/2}}^{x_{i+1/2}} (\partial_t u + \partial_x u)\,\mathrm{d}x = 0.$$

En approchant $u(x_{i+1/2})$ (resp. $u(x_{i-1/2})$) par u_i^n (resp. u_{i-1}^n) et en approchant $\partial_t u$ par un schéma d'Euler explicite, on obtient :

$$\begin{cases} h_i \dfrac{u_i^{n+1} - u_i^n}{k} + u_i^n - u_{i-1}^n = 0, \\ u_i^0 = \dfrac{1}{h_i} \displaystyle\int_{x_{i-1/2}}^{x_{i+1/2}} u_0(x)\,\mathrm{d}x. \end{cases} \tag{5.18}$$

Proposition 5.10 Soit $(u_i^n)_{n\in\mathbb{N},\, i\in\mathbb{Z}}$ la solution de (5.18). Si $k \leqslant h = \min h_i$ et si $A \leqslant u_0(x) \leqslant B$, alors :

$$A \leqslant u_i^n \leqslant B \quad \forall i \in \mathbb{Z} \quad \forall n \in \mathbb{N}.$$

La démonstration est similaire à celle de la proposition (5.7).

Définition 5.11 — Solution approchée. Soit $\mathcal{T}$ un maillage volumes finis de $\mathbb{R}$ défini par $\mathcal{T} = (K_i)_{i\in\mathbb{Z}}$ avec $K_i =]x_{i-1/2}\,;x_{i+1/2}[$. On appelle solution approchée de (5.9) par le schéma (5.18) la fonction $u_{\mathcal{T},k} : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$, définie par

$$u_{\mathcal{T},k}(x,t) = u_i^n \text{ si } x \in K_i \text{ et } t \in [nk\,;nk+1[\tag{5.19}$$

On admettra le théorème de convergence suivant [exercice 5.11] :

Théorème 5.4 — Convergence du schéma 5.18. Soit $u_0 \in L^\infty(\mathbb{R})$, on suppose que $k \leqslant h = \inf(h_i)$, alors $u_{\mathcal{T},k}$ converge vers u dans $\mathcal{L}^1_{\mathrm{loc}}(\mathbb{R}\times\mathbb{R}_+)$ lorsque h (et k) tend vers 0, c'est-à-dire que l'on a :

$$\int_C |u_{\mathcal{T},k} - u|\,\mathrm{d}x\,\mathrm{d}t \to 0$$

pour tout compact C de $\mathbb{R}\times\mathbb{R}_+$ lorsque h (et k) tend vers 0.

5.4 Cas non linéaire

On se donne $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$ et $u_0 \in \mathcal{C}^1(\mathbb{R})$ et on considère maintenant l'équation hyperbolique non linéaire :

$$\begin{cases} \partial_t u + (f(u))_x = 0 \quad (x,t) \in \mathbb{R}\times\mathbb{R}_+, \\ u(x,0) = u_0(x). \end{cases} \tag{5.20}$$

Commençons par donner la définition de solution classique de ce problème même si, comme nous le verrons après, celle-ci n'a pas grand intérêt puisque le problème (5.20) n'a pas, en général de solution classique.

Définition 5.12 — Solution classique. On suppose que $u_0 \in \mathcal{C}^1(\mathbb{R})$ et $f \in \mathcal{C}^2(\mathbb{R},\mathbb{R})$. Alors u est solution classique de (5.20) si $u \in \mathcal{C}^1(\mathbb{R}\times\mathbb{R}_+,\mathbb{R})$ et u vérifie

$$\begin{cases} (\partial_t u + (f(u))_x)(x,t) = 0 & \forall (x,t) \in \mathbb{R}\times\mathbb{R}_+, \\ u(x,0) = u_0(x) & \forall x \in \mathbb{R}. \end{cases}$$

Avant d'énoncer le théorème de non existence, rappelons que dans le cas d'une équation différentielle du type non linéaire,

$$\begin{cases} x'(t) = f(x(t)) \quad t \in \mathbb{R}_+, \\ x(0) = x_0, \end{cases}$$

si on note $T_{\max}$ le temps d'existence de la solution et si $T_{\max} < +\infty$ alors $\|x(t)\| \to +\infty$ lorsque $t \to T_{\max}$. Donnons maintenant la définition des courbes caractéristiques de l'équation (5.20), qui permet le lien entre les équations hyperboliques non linéaires et les équations différentielles ordinaires.

Définition 5.13 — Courbe caractéristique. On appelle courbe caractéristique du problème (5.20) issue de $x_0 \in \mathbb{R}$, la courbe définie par le problème de Cauchy suivant :

$$\begin{cases} x'(t) = f'(u(x(t),t)) \\ x(0) = x_0 \end{cases} \tag{5.21}$$

Théorème 5.5 — Non existence. Soit $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$, on suppose que f' n'est pas constante, alors il existe $u_0 \in C_c^\infty(\mathbb{R})$ telle que (5.20) n'admette pas de solution classique.

Démonstration. Comme $f \in \mathcal{C}^2(\mathbb{R},\mathbb{R})$, on a $f' \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$ et donc le théorème de Cauchy-Lipschitz s'applique. Il existe donc une solution maximale $x(t)$ définie sur $[0\,;T_{\max}[$ et $x(t)$ tend vers l'infini lorsque t tend vers $T_{\max}$ si $T_{\max} < +\infty$. Les quatre étapes de la démonstration sont les suivantes :

1. $u(x(t),t) = u_0(x_0)$, $\forall t \in [0\,;T_{\max}[$ et donc que toute solution de (5.20) est constante sur les caractéristiques.
2. Les courbes caractéristiques sont des droites.
3. $T_{\max} = +\infty$ et donc $u(x,t) = u_0(x_0)$, $\forall t \in [0\,;+\infty[$.
4. On en déduit alors qu'on n'a pas de solution classique de (5.20).

Détaillons maintenant ces étapes.

1. Soit $\varphi(t) = u(x(t),t)$; par dérivation, on obtient $\varphi'(t) = \partial_t u(x(t),t) + \partial_x u(x(t),t)x'(t)$. Comme x vérifie (5.21), ceci entraîne $\varphi'(t) = \partial_t u(x(t),t) + f'(u(x(t),t))\partial_x u(x(t),t)$ et donc $\varphi'(t) = (\partial_t u + (f(u))_x)(x(t),t) = 0$. La fonction φ est donc constante et on a $u(x(t),t) = \varphi(t) = \varphi(0) = u(x(0),0) = u(x_0,0) = u_0(x_0)$, $\forall t \in [0\,;T_{\max}[$.
2. Comme $u(x(t),t) = u_0(x_0)$, $\forall t \in [0\,;T_{\max}[$, on a donc $x'(t) = f'(u_0(x_0))$. Donc en intégrant, on obtient que le système (5.21) décrit la droite d'équation :
$$x(t) = f'(u_0(x_0))t + x_0. \tag{5.22}$$
3. Puisque x vérifie (5.22), on a donc
$$\lim_{t \to T_{\max}} |x(t)| < +\infty$$
On en déduit que $T = T_{\max}$.
4. Comme f' est non constante, il existe v_0, v_1 tel que $f'(v_0) > f'(v_1)$, et on peut construire $u_0 \in \mathcal{C}_c^\infty(\mathbb{R},\mathbb{R})$ telle que $u_0(x_0) = v_0$ et $u_0(x_1) = v_1$, où x_0 et x_1 sont donnés et $x_0 < x_1$, voir figure 5.4. Supposons

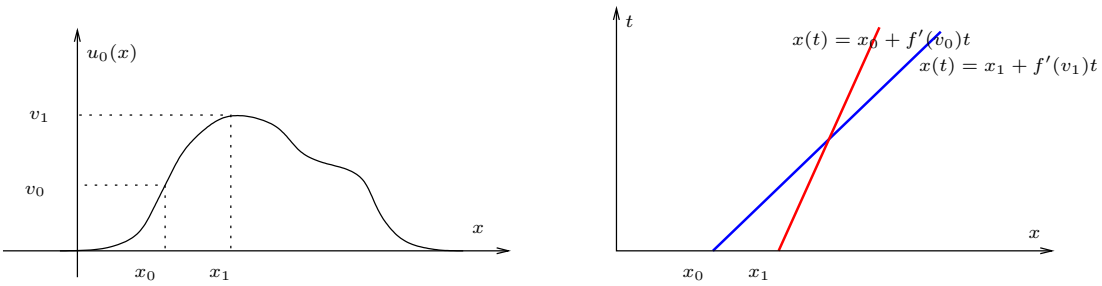


Figure 5.4 – Droites caractéristiques — cas non linéaire

que u soit solution classique avec cette donnée initiale alors $u(x_0 + f'(u_0(x_0))t, t) = u_0(x_0) = v_0$ et $u(x_1 + f'(u_0(x_1))t, t) = u_0(x_1) = v_1$. Soit T tel que $x_0 + f'(v_0)T = x_1 + f'(v_1)T = \bar{x}$, c'est-à-dire
$$T = \frac{x_1 - x_0}{f'(v_0) - f'(v_1)}.$$
On a alors $u(\bar{x}, T) = u_0(x_0) = v_0 = u_0(x_1) = v_1$, ce qui est impossible.

On en conclut que (5.20) n'admet pas de solution classique pour cette donnée initiale. •

Définition 5.14 — Solution faible. Soit $u_0 \in L^\infty(\mathbb{R})$ et $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$, On appelle solution faible de (5.20) une fonction $u \in L^\infty(\mathbb{R}\times\mathbb{R}_+)$ telle que

$$\iint_{\mathbb{R}\times\mathbb{R}_+} \big(u(x,t)\varphi_t(x,t) + f(u(x,t))\varphi_x(x,t)\big)\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}} u_0(x)\varphi(x,0)\,\mathrm{d}x = 0 \quad \forall \varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}_+,\mathbb{R}). \tag{5.23}$$

Donnons maintenant les liens entre solution classique et solution faible.

Proposition 5.15 Soient $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$ et $u_0 \in \mathcal{C}(\mathbb{R},\mathbb{R})$ des fonctions données.

1. Si u est solution classique de (5.20) alors u est solution faible de (5.20).
2. Si $u \in \mathcal{C}^1(\mathbb{R}\times\,]0\,;+\infty[) \cap \mathcal{C}^1(\mathbb{R}\times[0\,;+\infty[)$ est solution faible de (5.20) alors u est solution classique de (5.20).
3. Soit $\sigma \in \mathbb{R}$, $D_1 = \{(x,t) \in \mathbb{R}\times\mathbb{R}_+ ; x < \sigma t\}$ et $D_2 = \{(x,t) \in \mathbb{R}\times\mathbb{R}_+ ; x > \sigma t\}$. Alors si $u \in C(\mathbb{R}\times\mathbb{R}_+)$ est telle que $u_{|D_i} \in \mathcal{C}^1(D_i,\mathbb{R})$, $i = 1,2$ et que (5.20) est vérifié pour tout $(x,t) \in D_i$, $i = 1,2$, alors u est solution faible de (5.20).

Démonstration.

1. Supposons que u est solution classique de (5.20), c'est-à-dire de :
$$\begin{cases} \partial_t u + (f(u))_x = 0 \quad (x,t) \in \mathbb{R}\times\mathbb{R}_+ \\ u(x,0) = u_0(x) \end{cases}$$
Soit $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}_+,\mathbb{R})$. Multiplions (5.20) par φ et intégrons sur $\mathbb{R}\times\mathbb{R}_+$. On obtient :
$$\int_{\mathbb{R}}\int_{\mathbb{R}_+} \partial_t u(x,t)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}}\int_{\mathbb{R}_+} (f(u))_x(x,t)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x = 0.$$
L'application du théorème de Fubini et une intégration par parties donnent alors :
$$\int_{\mathbb{R}} u(x,0)\varphi(x,0)\,\mathrm{d}x - \int_{\mathbb{R}}\int_{\mathbb{R}_+} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x - \int_{\mathbb{R}_+}\int_{\mathbb{R}} f(u)(x,t)\varphi_x(x,t)\,\mathrm{d}x\,\mathrm{d}t = 0,$$
(car supp(φ) est compact). Et on obtient donc bien la relation (5.23), grâce à la condition initiale $u(x,0) = u_0(x)$.
2. Soit donc u une solution faible de (5.20), qui vérifie de plus $u \in \mathcal{C}^1(\mathbb{R}\times\,]0\,;+\infty[) \cap \mathcal{C}(\mathbb{R}\times[0\,;+\infty[)$. On a donc suffisamment de régularité pour intégrer par parties dans (5.23).
Commençons par prendre φ à support compact dans $\mathbb{R}\times\,]0\,;+\infty[$. On a donc $\varphi(x,0) = 0$ et une intégration par parties dans (5.23) donne :
$$-\int_{\mathbb{R}}\int_{\mathbb{R}_+} \partial_t u(x,t)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x - \int_{\mathbb{R}_+}\int_{\mathbb{R}} (f(u))_x(x,t)\varphi(x,t)\,\mathrm{d}x\,\mathrm{d}t = 0.$$
On a donc :
$$\int_{\mathbb{R}}\int_{\mathbb{R}_+} \big(\partial_t u(x,t) + (f(u))_x(x,t)\big)\,\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x = 0 \quad \forall \varphi \in \mathcal{C}^1_c(\mathbb{R}\times\,]0\,;+\infty[).$$
Comme $\partial_t u + (f(u))_x$ est continue, on en déduit que $\partial_t u + (f(u))_x = 0$. En effet, on on rappelle que si $\int_R f(x)\varphi(x)\,\mathrm{d}x = 0$ pour toute fonction φ continue de $\mathbb{R}$ dans $\mathbb{R}$, alors $f = 0$ p.p. ; si de plus f est continue, alors $f = 0$ partout.
On prend alors $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}_+)$. Dans ce cas, une intégration par parties dans (5.23) donne
$$\int_{\mathbb{R}} u(x,0)\varphi(x,0)\,\mathrm{d}x - \int_{\mathbb{R}}\int_{\mathbb{R}_+} \big(\partial_t u(x,t) + (f(u))_x(x,t)\big)\,\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x - \int_{\mathbb{R}} u_0(x)\varphi(x,0)\,\mathrm{d}x = 0.$$
Mais on vient de montrer que $\partial_t u + (f(u))_x = 0$. On en déduit que
$$\int_{\mathbb{R}} (u_0(x) - u(x,0))\varphi(x,0)\,\mathrm{d}x = 0 \quad \forall \varphi \in \mathcal{C}^1_c(\mathbb{R}).$$
Comme u est continue, ceci entraîne $u(x,0) = u_0(x)$. Donc u est solution classique de (5.20).
3. Soit $u \in \mathcal{C}(\mathbb{R}\times\mathbb{R}_+)$ telle que $u|_{D_i}$ vérifie (5.20), pour tout $(x,t) \in D_i$. Montrons que u est solution faible. Pour cela, calculons :
$$X = \int_{\mathbb{R}}\int_{\mathbb{R}_+} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}_+}\int_{\mathbb{R}} f(u)(x,t)\varphi_x(x,t)\,\mathrm{d}x\,\mathrm{d}t. \tag{5.24}$$

On a donc $X = X_1 + X_2$, avec

$$X_1 = \int_{\mathbb{R}} \int_{\mathbb{R}_+} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x \text{ et } X_2 = \int_{\mathbb{R}_+} \int_{\mathbb{R}} (f(u))(x,t)\varphi_x(x,t)\,\mathrm{d}x\,\mathrm{d}t.$$

Calculons X_1 ; comme u n'est de classe $\mathcal{C}^1$ que sur chacun des domaines D_i, on n'a pas le droit d'intégrer par parties sur $\mathbb{R} \times \mathbb{R}_+$ entier. On va donc décomposer l'intégrale sur D_1 et D_2 ; supposons par exemple $\sigma < 0$, voir figure 5.5 (le cas $\sigma > 0$ se traite de façon similaire). On a alors $D_2 = \{(x,t); x \in \mathbb{R}_- \text{ et } 0 < t < \frac{x}{\sigma}\}$ et $D_1 = \mathbb{R}_+ \times \mathbb{R}_+ \cup \{(x,t); x \in \mathbb{R}_- \text{ et } \frac{x}{\sigma} < t < +\infty\}$. On a donc :

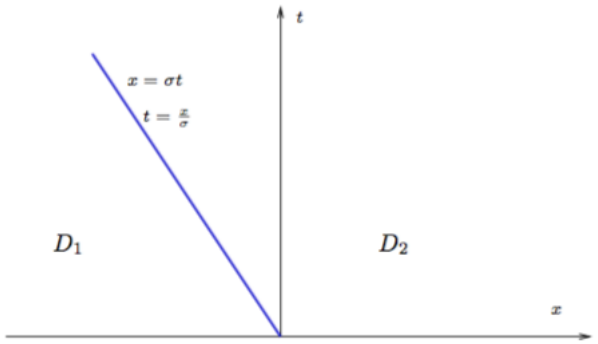


Figure 5.5 – Domaines D_1 et D_2

$$X_1 = \int_{\mathbb{R}_-} \int_0^{x/\sigma} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}_-} \int_{\frac{x}{\sigma}}^{+\infty} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x.$$

Comme u est de classe $\mathcal{C}^1$ sur chacun des domaines, on peut intégrer par parties, ce qui donne :

$$\begin{aligned} X_1 = &\int_{\mathbb{R}_-} u\Big(x, \frac{x}{\sigma}\Big)\varphi\Big(x, \frac{x}{\sigma}\Big)\,\mathrm{d}x - \int_{\mathbb{R}_-} u(x,0)\varphi(x,0)\,\mathrm{d}x - \int_{\mathbb{R}_-} \int_0^{\frac{x}{\sigma}} \partial_t u(x,t)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x \\ &+ \int_{\mathbb{R}_-} u\Big(x, \frac{x}{\sigma}\Big)\varphi\Big(x, \frac{x}{\sigma}\Big)\,\mathrm{d}x - \int_{\mathbb{R}_-} \int_{\frac{x}{\sigma}}^{+\infty} \partial_t u(x,t)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x \\ &+ \int_{\mathbb{R}_+} -u(x,0)\varphi(x,0)\,\mathrm{d}x - \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \partial_t u(x,t)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x. \end{aligned} \quad (5.25)$$

En simplifiant, il vient :

$$X_1 = -\int_{\mathbb{R}} u(x,0)\varphi(x,0)\,\mathrm{d}x - \iint_{D_1} \partial_t u(x,t)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x - \iint_{D_2} \partial_t u(x,t)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x.$$

On décompose de même X_2 sur $D_1 \cup D_2$, en remarquant maintenant que $D_1 = \{(x,t) \in \mathbb{R} \times \mathbb{R}_+; x < \sigma t\}$

et $D_2 = \{(x,t) \in \mathbb{R} \times \mathbb{R}_+ ; x > \sigma t\}$:

$$X_2 = \int_{\mathbb{R}_+} \int_{-\infty}^{\sigma t} f(u)(x,t)\partial_x \varphi(x,t)\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}_+} \int_{\sigma t}^{+\infty} f(u)(x,t)\partial_x \varphi(x,t)\,\mathrm{d}x\,\mathrm{d}t.$$

La fonction u est de classe $\mathcal{C}^1$ sur chacun des domaines, on peut là encore intégrer par parties. Comme φ est à support compact sur $\mathbb{R} \times \mathbb{R}_+$, on obtient après simplification :

$$X_2 = -\iint_{D_1} (f(u))_x(x,t)\varphi(x,t)\,\mathrm{d}x\,\mathrm{d}t - \iint_{D_2} (f(u))_x(x,t)\varphi(x,t)\,\mathrm{d}x\,\mathrm{d}t.$$

Comme $\partial_t u + (f(u))_x = 0$ sur D_1 et D_2, on a donc :

$$X = X_1 + X_2 = -\int_{\mathbb{R}} u(x,0)\varphi(x,0)\,\mathrm{d}x,$$

ce qui prouve que u est solution faible de (5.20).

Remarquons que le calcul de X défini en (5.24) peut se faire de manière beaucoup plus rapide en écrivant X sous la forme :

$$X = \int_{\mathbb{R}\times\mathbb{R}_+} \mathbf{V} \cdot \nabla_{x,t}\varphi\,\mathrm{d}x\,\mathrm{d}t, \text{ où } \mathbf{V} = \begin{pmatrix} f(u) \\ u \end{pmatrix} \text{ et} \nabla_{x,t}\varphi = \begin{pmatrix} \partial_x \varphi \\ \partial_t \varphi \end{pmatrix},$$

En effectuant une intégration par parties (multidimensionnelle) et en remarquant que $\mathrm{div}(\mathbf{V}) = \partial_x(f(u)) + \partial_t u = 0$ sur D_1 et sur D_2, on obtient que

$$X = \int_{\partial D_1} \mathbf{V} \cdot \mathbf{n}_1\ \mathrm{d}\gamma + \int_{\partial D_2} \mathbf{V} \cdot \mathbf{n}_2\ \mathrm{d}\gamma,$$

où $\mathbf{n}_i$ désigne le vecteur normal unitaire à ∂D_i, extérieur à D_i, et $\mathrm{d}\gamma$ désigne le sympole d'intégration sur la frontière. Mais comme $\mathbf{V}$ est continue, ceci se simplifie en

$$X = -\int_{\mathbb{R}} u_0(x)\varphi(x,0)\,\mathrm{d}x.$$

•

Notons qu'il existe souvent plusieurs solutions faibles. On a donc besoin d'une notion supplémentaire pour les distinguer. C'est la notion de solution entropique, qui nous permettra d'obtenir l'unicité. Donnons tout d'abord un exemple de non-unicité de la solution faible. Pour cela on va considérer une équation modèle, appelée *équation de Burgers*, qui s'écrit

$$\partial_t u + (u^2)_x = 0 \tag{5.26}$$

Pour calculer les solutions du problème de Cauchy associé à cette équation de manière analytique, on considère une donnée initiale particulière, qui sécrit

$$u_0(x) = \begin{cases} u_g & \text{si } x < 0, \\ u_d & \text{si } x > 0, \end{cases}$$

Ces données initiales définissent un problème de Cauchy particulier, qu'on appelle problème de Riemann, que nous étudierons plus en détails par la suite.

Considérons alors le problème suivant dit *problème de Riemann* [définition 5.22] pour l'équation de Burgers :

$$\begin{cases} \partial_t u + (u^2)_x = 0 \\ u_0(x) = \begin{cases} u_g = -1 & \text{si } x < 0 \\ u_d = 1 & \text{si } x > 0 \end{cases} \end{cases} \tag{5.27}$$

On cherche une solution faible de la forme :

$$u(x,t) = \begin{cases} u_g & \text{si } x < \sigma t \\ u_d & \text{si } x > \sigma t \end{cases} \tag{5.28}$$

Notons que cette éventuelle solution est discontinue au travers de la droite d'équation $x = \sigma t$ dans le plan (x,t). On remplace $u(x,t)$ par ces valeurs dans (5.23). Après calculs [exercice 77 et proposition 5.23] on s'aperçoit que u est solution faible si la condition suivante, dite *condition de Rankine et Hugoniot*, est vérifiée :

$$\sigma(u_d - u_g) = (f(u_d) - f(u_g)) \tag{5.29}$$

ce qui avec la condition initiale particulière choisie ici donne $2\sigma = 1^2 - (-1)^2 = 0$. On peut trouver d'autres solutions faibles : en effet, on sait que sur les caractéristiques, qui ont pour équation $x = x_0 + f'(u_0(x_0))t$, la fonction u est constante. Comme $f'(u) = 2u$, les caractéristiques sont donc des droites de pente -2 si $x_0 < 0$ et de pente 2 si $x_0 > 0$. Construisons ces caractéristiques sur la figure 5.6 : Dans la zone du milieu, où l'on a représenté un point d'interrogation, on cherche u sous

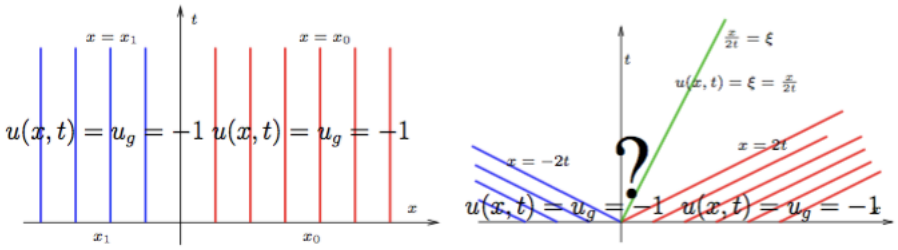


Figure 5.6 – Problème de Riemann pour l'équation de Burgers

la forme $u(x,t) = \varphi(x/t)$ et telle que u soit continue sur $\mathbb{R} \times \mathbb{R}_+$. La fonction u suivante convient :

$$u(x,t) = \begin{cases} -1 & \text{si } x < -2t \\ \dfrac{x}{2t} & \text{si } -2t < x < 2t \\ 1 & \text{si } x > 2t \end{cases} \tag{5.30}$$

Comment choisir la *bonne* solution faible, entre (5.28) et(5.30) ? Comme les problèmes hyperboliques sont souvent obtenus en négligeant les termes de diffusion dans des équations paraboliques, une technique pour choisir la solution est de chercher la limite du problème de diffusion associé qui s'écrit :

$$\partial_t u + (f(u))_x - \varepsilon \partial^2_{xx} u = 0, \tag{5.31}$$

lorsque le terme de diffusion devient négligeable, c'est-à-dire lorsque ε tend vers 0. Soit u_ε la solution de (5.31) (on admettra l'existence et l'unicité de u_ε). On peut montrer que u_ε tend vers u lorsque ε tend vers 0, où u est la *solution faible entropique* de (5.31), définie comme suit.

Définition 5.16 — Solution entropique. Soit $u_0 \in L^\infty(\mathbb{R})$ et $f \in \mathcal{C}^1(\mathbb{R})$, on dit que $u \in L^\infty(\mathbb{R} \times \mathbb{R}_+)$ est solution entropique de (5.31) si pour toute fonction $\eta \in \mathcal{C}^1(\mathbb{R})$ convexe, appelée *entropie* et pour toute fonction $\phi \in \mathcal{C}^1(\mathbb{R})$ telle que $\phi' = f'\eta'$, appelé *flux d'entropie*, on a :

$$\int_{\mathbb{R}} \int_{\mathbb{R}_+} (\eta(u)\varphi_t + \phi(u)\varphi_x)\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}} \eta(u_0(x))\varphi(x,0)\,\mathrm{d}x \geqslant 0 \quad \forall \varphi \in \mathcal{C}^1_c(\mathbb{R} \times \mathbb{R}_+, \mathbb{R}_+). \tag{5.32}$$

Remarque 5.17 — Condition initiale. Noter que dans la définition 5.16, on prend une fois de plus $\varphi \in \mathcal{C}^1_c(\mathbb{R} \times \mathbb{R}_+, \mathbb{R}_+)$ de manière à bien prendre en compte la condition initiale ; ceci n'est pas toujours fait de cette manière dans les travaux plus anciens sur le sujet, mais entraîne des difficultés lorqu'on s'intéresse à la convergence des schémas numériques.

On admettra le théorème suivant :

Théorème 5.6 — Kruskov. Soient $u_0 \in L^\infty(\mathbb{R})$ et $f \in \mathcal{C}^1(\mathbb{R})$ alors il existe une unique solution entropique de (5.20) au sens de la définition 5.16.

Proposition 5.18 Si u est solution classique de (5.20), alors u est solution entropique.

Démonstration. Soit $u \in \mathcal{C}^1(\mathbb{R}\times\mathbb{R}_+,\mathbb{R})$, Soit $\eta \in \mathcal{C}^1(\mathbb{R})$, convexe, une entropie et ϕ tel que $\phi' = f'\eta'$, le flux associé. Multiplions (5.20) par $\eta'(u)$:

$$\eta'(u)\partial_t u + f'(u)\partial_x u\eta'(u) = 0$$

Soit encore, puisque $\phi' = f'\eta'$,

$$(\eta(u))_t + \phi'(u)\partial_x u - 0$$

On a donc finalement :

$$(\eta(u))_t + (\phi(u))_x - 0 \tag{5.33}$$

De plus, comme $u(x,0) = u_0(x)$, on a aussi : $\eta(u(x,0)) = \eta(u_0(x))$. Soit $\varphi \in \mathcal{C}^1_c(\mathbb{R}\times\mathbb{R}_+,\mathbb{R}_+)$, on multiplie (5.33) par φ, on intègre sur $\mathbb{R}\times\mathbb{R}_+$ et on obtient (5.32) (avec égalité) en intégrant par parties. Dans le cas d'une solution classique, l'inégalité d'entropie est une égalité. •

On a de même le résultat suivant :

Proposition 5.19 Si u est solution faible entropique de (5.20), alors u est solution faible.

Démonstration. Il suffit de prendre $\eta(u) = u$ et $\eta(u) = -u$ dans (5.32) pour se convaincre du résultat. •

On déduit de la proposition 5.18 et du théorème 5.6 de Kruskov, que si on a plusieurs solutions faibles au problème 5.20 et que l'une d'entre elles est régulière, alors cette dernière est forcément la solution entropique. Enfin, la caractérisation suivante, que l'on admettra, est souvent utilisée en pratique :

Proposition 5.20 — Entropies de Kruskov. Soit $u_0 \in L^\infty(\mathbb{R})$ et $f \in \mathcal{C}^1(\mathbb{R})$, alors $u \in L^\infty(\mathbb{R}\times\mathbb{R}_+)$ est solution entropique de (5.31) au sens de la définition 5.16 si et seulement si pour tout $\kappa \in \mathbb{R}$, alors (5.32) est vérifiée avec η_κ définie par $\eta_\kappa(s) = |s-\kappa|$ et ϕ_κ, flux d'entropie associée, défini par :

$$\phi_\kappa(u) = \max(f(u),\kappa) - \min(f(u),\kappa).$$

Notons que $\eta\kappa$ n'est pas de classe $\mathcal{C}^1$.

Notons que les solutions d'une équation hyperbolique non linéaire respectent les bornes de la solution initiale. Plus précisément, on a le résultat suivant, qu'on admettra :

Proposition 5.21 — Estimation L^∞. Si $u_0 \in L^\infty(\mathbb{R})$ et soit A et $B \in \mathbb{R}$ tels que $A \leqslant u_0 \leqslant B$ p.p.. Soit $f \in \mathcal{C}^1(\mathbb{R})$, alors la solution entropique $u \in L^\infty(\mathbb{R}\times\mathbb{R}_+)$ de (5.20) vérifie $A \leqslant u(x) \leqslant B$ p.p. dans $\mathbb{R}\times\mathbb{R}_+$.

Cette propriété est essentielle dans les phénomènes de transport et il est donc important qu'elle soit préservée pour la solution approchée donnée par un schéma numérique.

Avant d'aborder l'étude des schémas numériques pour les équations hyperboliques, nous terminons par un résultat sur les solutions du problème de Riemann, dont nous nous sommes d'ailleurs servis pour montrer la non unicité des solutions faibles de (5.27).

Définition 5.22 — Problème de Riemann. Soient $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$, on appelle problème de Riemann avec données u_g, $u_d \in \mathbb{R}$, le problème suivant :

$$\begin{cases} \partial_t u + (f(u))_x = 0 \quad x \in \mathbb{R} \quad t > 0 \\ u(0,x) = \begin{cases} u_g & \text{si } x < 0 \\ u_d & \text{si } x > 0 \end{cases} \end{cases} \tag{5.34}$$

Lorsque la fonction f est convexe ou concave, les solutions du problème de Riemann se calculent facilement ; en effet, on peut montrer le résultat suivant [exercice 78] :

Proposition 5.23 Soit $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$ strictement convexe et soient u_g et $u_d \in \mathbb{R}$.

1. Si $u_g > u_d$, on pose
$$\sigma = \frac{[f(u)]}{[u]} \tag{5.35}$$
avec $[f(u)] = f(u_d) - f(u_g)$ et $[u] = u_d - u_g$, alors la fonction u définie par
$$\begin{cases} u(x,t) = u_g & \text{si } x < \sigma t \\ u(x,t) = u_d & \text{si } x > \sigma t \end{cases} \tag{5.36}$$
est l'unique solution entropique de (5.34). Une solution de la forme (5.36) est appellée une *onde de choc*.
2. Si $u_g < u_d$, alors la fonction u définie par
$$\begin{cases} u(x,t) = u_g & \text{si } x < f'(u_g)t \\ u(x,t) = u_d & \text{si } x > f'(u_d)t \\ u(x,t) = \xi & \text{si } x = f'(\xi)t \ \text{ avec } \ u_g < \xi < u_d \end{cases} \tag{5.37}$$
est l'unique solution entropique de (5.34). Notons que dans ce cas, la solution entropique est continue. Une solution de la forme (5.37) est appelée une *onde de détente*.

Démonstration. Cherchons u sous la forme (5.36). Commençons par déterminer σ pour que u soit solution faible. On suppose, pour fixer les idées, que $\sigma > 0$ (mais le même raisonnement marche pour $\sigma < 0$). Soit $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}\times\mathbb{R}_+,\mathbb{R})$. On veut montrer que
$$X = X_1 + X_2 = -\int_{\mathbb{R}} u(x,0)\varphi(x,0)\,\mathrm{d}x,$$
où
$$X_1 = \int_{\mathbb{R}}\int_{\mathbb{R}_+} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x \quad \text{et} \quad X_2 = \int_{\mathbb{R}}\int_{\mathbb{R}_+} f((u(x,t))\varphi_x(x,t)\,\mathrm{d}t\,\mathrm{d}x$$
Calculons donc X_1 et X_2 :
$$\begin{aligned} X_1 &= \int_{-\infty}^0\int_0^{+\infty} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_0^{+\infty}\int_0^{\frac{x}{\sigma}} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_0^{+\infty}\int_{\frac{x}{\sigma}}^{+\infty} u(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x \\ &= -\int_{\infty}^0 u_g\varphi(x,0)\,\mathrm{d}x + \int_0^{+\infty} u_d\Big(\varphi\Big(x,\frac{x}{\sigma}\Big) - \varphi(x,0)\Big)\,\mathrm{d}x + \int_0^{+\infty} -u_g\varphi\Big(x,\frac{x}{\sigma}\Big)\,\mathrm{d}x \\ &= -\int_{\mathbb{R}} u(x,0)\varphi(x,0)\,\mathrm{d}x + \int_0^{+\infty}(u_d-u_g)\varphi\Big(x,\frac{x}{\sigma}\Big)\,\mathrm{d}x \end{aligned}$$
De même
$$\begin{aligned} X_2 &= \int_0^{+\infty}\int_{-\infty}^{\sigma t} f(u)\varphi_x(x,t)\,\mathrm{d}x\,\mathrm{d}t + \int_0^{+\infty}\Big(\int_{\sigma t}^{+\infty} f(u)(x,t)\varphi_x(x,t)\Big)\,\mathrm{d}x\,\mathrm{d}t \\ &= \int_0^{+\infty} f(u_g)\varphi(\sigma t,t)\,\mathrm{d}t - \int_0^{+\infty} f(u_d)\varphi(\sigma t,t)\,\mathrm{d}t. \end{aligned}$$
En posant $[u] = u_d - u_g$ et $[f(u)] = f(u_d) - f(u_g)$, on obtient :
$$\begin{aligned} X + \int_{\mathbb{R}} u(x,0)\varphi(x,0)\,\mathrm{d}x &= \int_0^{+\infty}[u]\varphi\Big(x,\frac{x}{\sigma}\Big)\,\mathrm{d}x - \int_0^{+\infty}[f(u)]\varphi(\sigma t,t)\,\mathrm{d}t \\ &= \int_0^{+\infty}[u]\varphi(\sigma t,t)\sigma\,\mathrm{d}t - \int_0^{+\infty}[f(u)]\varphi(\sigma t,t)\,\mathrm{d}t. \end{aligned}$$
On en déduit que
$$X + \int_{\mathbb{R}} u(x,0)\varphi(x,0)\,\mathrm{d}x = 0 \text{ si } \sigma[u] - [f(u)] = 0,$$
ce qui est vrai si la condition suivante, dite de *Rankine et Hugoniot* :
$$\sigma[u] = [f(u)] \tag{5.38}$$
est vérifiée.

Voyons maintenant si u est bien solution entropique. Pour cela, on considère $\eta \in \mathcal{C}^1$ une *entropie* et $\phi \in \mathbb{C}$ le flux d'entropie associé tels que $\phi' = \eta' f'$. Le même calcul que le précédent, en remplaçant u par $\eta(u)$ et $f(u)$ par $\phi(u)$ donne que :

$$\int_{\mathbb{R}}\int_{\mathbb{R}_+} \eta(u)(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}_+}\int_{\mathbb{R}} \phi(u)(x,t)\varphi_x(x,t)\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}} \eta(u_0(x))\,\mathrm{d}x = \int_0^{+\infty} (\sigma[\eta(u)] - [\phi(u)])\varphi(\sigma t,t)\,\mathrm{d}t. \quad (5.39)$$

Pour que u soit solution entropique, il faut (et il suffit) donc que

$$\sigma[\eta(u)] \geqslant [\phi(u)] \quad (5.40)$$

Il reste à vérifier que cette inégalité est vérifiée pour σ donné par (5.38), c'est-à-dire

$$\frac{f(u_d)-f(u_g)}{u_d-u_g}(\eta(u_d)-\eta(u_g)) \geqslant \phi(u_d)-\phi(u_g)$$

Ceci s'écrit encore :

$$(f(u_d)-f(u_g))(\eta(u_d)-\eta(u_g)) \leqslant (\phi(u_d)-\phi(u_g))(u_d-u_g).$$

Cette inégalité est vérifiée en appliquant le lemme suivant avec $b = u_g > u_d = a$. •

Lemme 5.24 Soient $a, b \in \mathbb{R}$ tels que $a < b$, soient f et $\eta \in \mathcal{C}^1(\mathbb{R})$ des fonctions convexes et $\phi \in \mathcal{C}^1(\mathbb{R})$ telle que $\phi' = \eta' f'$, alors :

$$\int_a^b \phi'(s)(b-a)\,\mathrm{d}s \geqslant \int_a^b f'(s)\,\mathrm{d}s \int_a^b \eta'(s)\,\mathrm{d}s$$

Démonstration. On a

$$\int_a^b \phi'(x)\,\mathrm{d}x = \int_a^b f'(x)\eta'(x)\,\mathrm{d}x = \int_a^b f'(x)(\eta'(x)-\eta'(y))\,\mathrm{d}x + \int_a^b f'(x)\eta'(y)\,\mathrm{d}x, \quad \forall y \in \mathbb{R}$$

On a donc, en intégrant par rapport à y entre a et b :

$$(b-a)\int_a^b \phi'(x)\,\mathrm{d}x = \int_a^b\int_a^b f'(x)(\eta'(x)-\eta'(y))\,\mathrm{d}x\,\mathrm{d}y + \int_a^b f'(x)\,\mathrm{d}x \int_a^b \eta'(y)\,\mathrm{d}y$$

Or

$$\int_a^b\int_a^b f'(x)[\eta'(x)-\eta'(y)]\,\mathrm{d}x\,\mathrm{d}y = \int_a^b\int_a^b f'(y)(\eta'(y)-\eta'(x))\,\mathrm{d}x\,\mathrm{d}y$$

et donc

$$(b-a)\int_a^b\int_a^b \phi'(x)\,\mathrm{d}x = \int_a^b\int_a^b (f'(x)-f'(y)(\eta'(x)-\eta'(y)\,\mathrm{d}x\,\mathrm{d}y + \Big(\int_a^b f'(x)\,\mathrm{d}x\Big)\Big(\int_a^b \eta'(y(y)\,\mathrm{d}y\Big).$$

Comme f' et η' sont croissantes, la première intégrale du second membre est nulle et on a donc bien le résultat annoncé.

On vérifie facilement que la fonction u définie par (5.37) est continue sur $\mathbb{R}\times\mathbb{R}_+^*$ et qu'elle vérifie $\partial_t u + (f(u))_x = 0$ dans chacun des domaines D_1, D_2 et D_3 définis par

$$D_1 = \{t>0, x < f'(u_g)t\}, D_2 = \{t>0, f'(u_g)t < x < f'(u_d)t\} \text{ et } D_3 = \{t>0, x > f'(u_d)t\}.$$

Donc par le point 3 de la proposition 5.15, on sait que u est solution faible (mais attention, ce n'est pas une solution classique car u n'est pas forcément $\mathcal{C}^1$ sur $\mathbb{R}\times\mathbb{R}_+$ tout entier).

Soit $\eta \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$ une entropie (convexe) et ϕ le flux d'entropie associé, comme $\partial_t u + (f(u))_x = 0$ dans D_i pour $i = 1$ à 3, en multipliant par $\eta'(u)$, on a également que $(\eta(u))_t + (\phi(u))_x = 0$ dans D_i pour $i = 1$ à 3. Soit maintenant $\varphi \in \mathcal{C}_c^1(\mathbb{R}\times\mathbb{R}_+,\mathbb{R}_+)$, on va montrer que

$$\int_{\mathbb{R}}\int_{\mathbb{R}_+} (\eta(u))(x,t)\varphi_t(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}}\int_{\mathbb{R}} (\phi(u))(x,t)\varphi_x(x,t)\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}} \eta(u_0(x))\varphi(x,0)\,\mathrm{d}x = 0$$

(dans le cas d'une solution continue, l'inégalité d'entropie est une égalité). En effet, en intégrant par parties les trois termes précédents sur D_1, D_2, D_3, comme on l'a fait dans les questions 1 et 2, comme la fonction u est continue, les traces des fonctions sur le bord des domaines s'annulent deux à deux et il ne reste donc que la condition initiale. On montre ainsi (faire le calcul pour s'en convaincre...) que

$$\int_{\mathbb{R}}\int_{\mathbb{R}_+} (\eta(u))(x,t)\varphi_x(x,t)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}}\int_{\mathbb{R}_+} \phi(u)(x,t)(\varphi_x(x,t))\,\mathrm{d}x\,\mathrm{d}t = -\int_{\mathbb{R}} \eta(u_0(x))\varphi(x,0),$$

ce qui prouve que u est la solution entropique. •

5.5 Schémas, cas non linéaire

On se donne $u_0 \in L^\infty(\mathbb{R}$ et $f \in \mathcal{C}^1(\mathbb{R})$ et on cherche à trouver une approximation de la solution entropique du problème (5.20). On utilise les mêmes notations que pour le schéma (5.18). En intégrant l'équation $\partial_t u + (f(u))_x = 0$ sur une maille K_i, on obtient, au temps $t = t_n$:

$$\int_{K_i} \partial_t u(x, t_n)\,\mathrm{d}x\,\mathrm{d}t + f(u(x_{i+1/2}, t)) - f(u(x_{i-1/2}, t_n)) = 0.$$

En utilisant le schéma d'Euler explicite pour la discrétisation de la dérivée temporelle et en notant $f^n_{i+1/2}$ le flux numérique, c'est-à-dire l'approximation de $f(u(x_{i+1/2}, t_n))$ on obtient le schéma numérique suivant :

$$\begin{cases} h_i \dfrac{u_i^{n+1} - u_i^n}{k} + f^n_{i+1/2} - f^n_{i-1/2} = 0 \\ u_i^0 = \dfrac{1}{h_i} \displaystyle\int_{K_i} u_0(x)\,\mathrm{d}x. \end{cases} \tag{5.41}$$

Pour que ce schéma soit complètement défini, il reste à préciser $f^n_{i+1/2}$ en fonction des inconnues discrètes u^n_i. Un premier choix possible est le schéma centré,

$$f^n_{i+1/2} = \frac{f(u^n_{i+1}) + f(u^n_i)}{2}$$

dont on a vu qu'il est à proscrire, puisque, dans le cas linéaire, il est instable. Rappelons que dans le cas linéaire, le choix décentré amont donne $f^n_{i+1/2} = f(u^n_i)$ si $f(u) = u$ et $f^n_{i+1/2} = f(u^n_{i+1})$ si $f(u) = -u$. On va s'intéresser aux schémas les plus simples à trois points, c'est-à-dire que l'équation associée à l'inconnue u^n_i fait intervenir les trois inconnues discrètes u^n_i, u^n_{i-1} et u^n_{i+1}. Le flux numérique g s'écrit sous la forme $f^n_{i+1/2} = g(u^n_i, u^n_{i+1})$. Pour obtenir un bon schéma, on va choisir un flux monotone, au sens suivant :

Définition 5.25 On dit que qu'une fonction g définie de $\mathbb{R}^2$ dans $\mathbb{R}$ est un flux monotone pour la discrétisation de (5.20), si

1. g est consistante par rapport à f, c'est-à-dire $g(u, u) = f(u)$,
2. g est croissante par rapport à la première variable et décroissante par rapport à la deuxième variable,
3. g est lipschitzienne sur $[A\,;B]$, où $A = \inf_{\mathbb{R}} u_0$ et $B = \sup_{\mathbb{R}} u_0$.

Remarque 5.26 — Flux monotones et schémas monotones. Si le schéma 5.18 est à flux monotone et s'il vérifie la condition de CFL, on peut alors montrer que le schéma est monotone, c'est-à-dire qu'il s'écrit sous la forme $u^{n+1}_i = H(u^n_{i-1}, u^n_i, u^n_{i+1})$ où H est une fonction croissante de ses trois arguments.

Cas où f est monotone Pour illustrer le choix de g, supposons par exemple que f soit croissante. Un choix très simple consiste alors à prendre $g(u^n_i, u^n_{i+1}) = f(u^n_i)$. On vérifie que dans ce cas, les trois conditions ci-dessus sont vérifiées, ce schéma est dit décentré amont. On vérifiera qu'on retrouve le schéma décentré amont exposé dans le cas linéaire. De même si f est décroissante on peut facilement vérifier que le choix $g(u^n_i, u^n_{i+1}) = f(u^n_{i+1})$ convient.

Schéma à décomposition de flux Le schéma à décomposition de flux, appelé aussi *flux splitting* en anglais, consiste comme le nom l'indique à décomposer $f = f_1 + f_2$, où f_1 est croissante et f_2 décroissante et à prendre $g(u^n_i, u^n_{i+1}) = f_1(u^n_i) + f_2(u^n_{i+1})$.

Schéma de Lax Friedrich Le schéma de Lax Friedrich consiste à modifier le schéma centré de manière à le rendre stable. On écrit donc :

$$g(u^n_i, u^n_{i+1}) = \frac{1}{2}(f(u^n_i) + f(u^n_{i+1})) + D(u^n_i - u^n_{i+1})$$

où $D \geqslant 0$ est il faut avoir D suffisamment grand pour que g soit croissante par rapport à la première variable et décroissante par rapport à la seconde variable.

Schéma de Godunov Le schéma de Godunov [4] est un des schémas les plus connus pour les équations hyperboliques non linéaires. De nombreux schémas pour les systèmes ont été inspirés par ce schéma. Le flux numérique du schéma de Godunov s'écrit :

$$g(u_i^n, u_{i+1}^n) = f(w_R(u_i^n, u_{i+1}^n)) \tag{5.42}$$

où $w_R(u_i^n, u_{i+1}^n)$ est la solution en 0 du problème de Riemann avec conditions u_i^n, u_{i+1}^n, qui s'écrit

$$\begin{cases} \partial_t u + (f(u))_x = 0 \\ u_0(x) = \begin{cases} u_g = u_i^n & w < 0 \\ u_d = u_{i+1}^n & 0 \end{cases} \end{cases}$$

On peut montrer que le flux de Godunov (5.42) vérifie les conditions de la définition 5.25. Une manière de le faire est de montrer que le flux de Godunov s'écrit sous la forme équivalente suivante (voir exercice 81).

$$g(u_i^n, u_{i+1}^n) = \begin{array}{l} \min\{f(\xi), \xi \in [u_i^n, u_{i+1}^n]\} \text{ if } u_i^n \leqslant u_{i+1}^n, \\ \max\{f(\xi), \xi \in [u_{i+1}^n, u_i^n]\} \text{ if } u_{i+1}^n \leqslant u_i^n. \end{array} \tag{5.43}$$

Schéma de Murman Une manière de simplifier le schéma de Godunov est de remplacer la résolution du problème de Riemann linéaire. On prend alors $g(u_i^n, u_{i+1}^n) = f(\widetilde{w}_R(u_i^n, u_{i+1}^n)$ où $\widetilde{w}_R(u_i^n, u_{i+1}^n)$ est solution de

$$\begin{cases} \partial_t u + \alpha \partial_x u = 0 \\ u_0(x) = \begin{cases} u_i^n & x < 0 \\ u_{i+1}^n & x > 0 \end{cases} \end{cases}$$

Comme le problème est linéaire, la solution de ce problème est connue : $u(x,t) = u_0(x - \alpha t)$. Le schéma est donc très simple, malheureusement, le schéma de Murman n'est pas un schéma monotone [exercice (80)] car le flux n'est pas monotone par rapport aux deux variables. De fait on peut montrer que les solutions approchées peuvent converger vers des solutions non entropiques. On peut alors envisager une procédure de *correction d'entropie...*

On admettra les résultats de convergence et de stabilité des schémas à flux monotone résumés dans le théorème suivant, voir [6, Chapter 5] pour les différentes démonstrations possibles.

Théorème 5.7 — Stabilité et convergence. Soit $(u_i^n)_{i\in\mathbb{Z}, n\in\mathbb{N}}$ donnée par le schéma

$$\begin{cases} h_i \dfrac{u_i^{n+1} - u_i^n}{k} + g(u_i^n, u_{i+1}^n) - g(u_{i-1}^n, u_i^n) = 0 \\ u_i^0 = \dfrac{1}{h_i} \displaystyle\int_{K_i} u_0(x)\,\mathrm{d}x \end{cases}$$

On suppose que g est un flux monotone au sens de la définition 5.25. On suppose de plus que $k \leqslant \alpha h/(2M)$ et $\alpha h \leqslant h_i \leqslant h$, $\forall i$ où M est la constante de Lipschitz de g sur $[A\,;B]$ et A et B sont tels que $A \leqslant u_0(x) \leqslant B$ p.p. On a alors $A \leqslant u_i^n \leqslant B$ p.p. et $\|u_{\tau,k}\| \leqslant \|u_0\|_\infty$. Sous les mêmes hypothèses, si on note $u_{\tau,k}$ la solution approchée définie par (5.19), alors $u_{\tau,k}$ tend vers u, solution entropique de (5.20) dans $\mathcal{L}^1_{\text{loc}}(\mathbb{R}\times\mathbb{R}_+)$ lorsque h (et k) tend vers 0.

5.6 Exercices

5.6.1 Énoncés

Exercice 69 — Problème linéaire en dimension 1. Calculer la solution faible du problème $\partial_t u - 2\partial_x u = 0$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$ avec :

$$u(x,0) = \begin{cases} 0 & \text{si } x < 0, \\ 1 & \text{sinon.} \end{cases} \tag{5.44}$$

4. Sergei K. Godunov est un mathématicien russe né en 1929, membre de l'Académie des Sciences russe, en poste au Sobolev Institute of Mathematics, Novosibirsk, Sibérie

1. Tracer sur un graphique la solution a $t = 0$ et à $t = 1$, en fonction de x. Cette solution faible est-elle solution classique de (5.44) ?

2. Même question en remplaçant la condition initiale par $u(x, 0) = \sin x$.

Exercice 70 — Problème linéaire en dimension 2. *Corrigé en page 217*
Soit $\mathbf{v} \in \mathbb{R}^2$ et soit $u_0 \in C^1(\mathbb{R}^2, \mathbb{R})$. On considère le problème de Cauchy suivant :

$$\begin{cases} \partial_t u + \operatorname{div}(\mathbf{v}u) = 0, \\ u(x, 0) = u_0(x), \end{cases} \tag{5.45}$$

Calculer la solution du problème (5.45) (en fonction de u_0) en tout point $(x, t) \in \mathbb{R}^2 \times \mathbb{R}$

Exercice 71 — Schéma de Lax–Wendroff. Soit $u_0 \in C_c^\infty(\mathbb{R}, \mathbb{R})$ (ensemble des fonctions continues à support compact) et $T > 0$, et $a > 0$. On considère le problème suivant :

$$\partial_t u(x,t) + a\partial_x u(x,t) = 0,\ x \in \mathbb{R},\ t \in [0,T], \tag{5.46a}$$

$$u(x,0) = u_0(x). \tag{5.46b}$$

1 Rappeler l'expression de la solution unique du problème (5.46) en fonction de u_0.

On se donne un pas de temps constant k, avec $k = \frac{T}{N+1}$ ($N \in \mathbb{N}$). On se donne également un pas d'espace constant h, et des points de discrétisation en espace, $(x_i)_{i\in\mathbb{Z}}$ tels que $x_{i+1} - x_i = h$ pour tout i. On pose $t_n = nk$, pour $n \in \{0, \dots, N+1\}$. On cherche une approximation de $u(x_j, t_n)$ pour $n \in \{0, \dots, N+1\}$ et $i \in \mathbb{Z}$. On pose $\lambda = \frac{ak}{h}$.

2 Rappeler l'expression de la solution unique du problème (5.46) en fonction de u_0. Soit u la solution de (5.46). Montrer que pour tout $j \in \mathbb{Z}$ et pour tout $n \in \mathbb{N}$,

$$u(x_j, t_{n+1}) = u(x_j, t_n) - ak\partial_x u(x_j, t_n) + \frac{1}{2}a^2k^2\partial_{xx}^2 u(x_j, t_n + k^3 R_{j,n},\ \text{avec } |R_{j,n}| \leqslant C_{u_0}, \tag{5.47}$$

où $C_{u_0} \in \mathbb{R}$ ne dépend que de u_0, et $u(x_j, t_{n+1}) = u(x_{j-1}, t_n)$ si $\lambda = 1$.

3 Montrer pourquoi l'égalité (5.47) suggère le schéma suivant, dit de Lax-Wendroff :

$$\begin{cases} u_j^{(n+1)} = u_j^{(n)} - \frac{1}{2}\lambda(u_{j+1}^{(n)} - u_{j-1}^{(n)}) + \frac{1}{2}\lambda^2(u_{j+1}^{(n)} - 2u_j^{(n)} + u_{j-1}^{(n)}),\ j \in \mathbb{Z},\ n > 0, \\ u_j^{(0)} = u_0(x_j);\ j \in \mathbb{Z}. \end{cases} \tag{5.48}$$

avec $u_j^{(0)} = u_0(x_j)$ pour tout $j \in \mathbb{Z}$. Donner l'ordre de consistance du schéma (distinguer les cas $\lambda \neq 1$ et $\lambda = 1$).

4 On prend comme condition initiale $u^0(x) = e^{ipx}$, pour $p \in \mathbb{Z}$ fixé (avec $i^2 = -1$). Pour $j \in \mathbb{Z}$, calculer la valeur $u_j^{(1)}$ donnée par le schéma (5.48) en fonction de $u_j^{(0)}$ et en déduire le facteur d'amplification ξ_p, tel que $u_j^{(1)} = \xi_p u_j^{(0)}$. Montrer que le schéma est stable au sens de Von Neumann sous une condition de CFL qu'on explicitera.

5 Montrer par un contre exemple que si $\lambda \neq 1$, la norme L^∞ de la solution approchée n'est pas décroissante. [On pourra par exemple prendre dans le case $\lambda < 1$, $u_j^{(0)} = 1$ pour $j \leqslant 0$, $u_j^{(0)} = 0$ pour $j \geqslant 0$, et calculer $u_0^{(1)}$ et chercher ensuite un contre-exemple pour le cas $\lambda > 1$.]

Exercice 72 — Stabilité du schéma amont dans le cas linéaire. *Corrigé en page 218*
On considère le problème hyperbolique linéaire (5.9), avec $u_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, dont on calcule une solution approchée par le schéma volumes finis amont (5.18). Montrer que ce schéma est stable pour les normes L^1, L^2 et L^∞, c.à.d. que la solution approchée satisfait les propriétés suivantes :

1. $\|u_{\mathcal{T},k}(.,n)\|_{L^1(\mathbb{R})} \leqslant \|u_0\|_{L^1(\mathbb{R})}, \forall n \in \mathbb{N}$,

2. $\|u_{\mathcal{T},k}(.,n)\|_{L^2(\mathbb{R})} \leqslant \|u_0\|_{L^2(\mathbb{R})}, \forall n \in \mathbb{N}$,

3. $\|u_{\mathcal{T},k}(.,n)\|_{L^\infty(\mathbb{R})} \leqslant \|u_0\|_{L^\infty(\mathbb{R})}, \forall n \in \mathbb{N}$,

où $u_{\mathcal{T},k}$ désigne la solution approchée calculée par le schéma (voir (5.19)).

Exercice 73 — Convergence des schémas DFDA et VFDA dans le cas linéaire. *Corrigé en page 219*
Soit $u_0 \in C^2(\mathbb{R},\mathbb{R})$ et $T \in \mathbb{R}_+^\star$. On suppose que u_0, u_0' et u_0'' sont bornées sur $\mathbb{R}$. On considère le problème suivant :

$$\partial_t u(x,t) + \partial_x u(x,t) = 0 \quad x \in \mathbb{R} \quad t \in [0,T] \tag{5.49}$$
$$u(x,0) = u_0(x) \tag{5.50}$$

Ce problème admet une et une seule solution classique, notée u. On se donne un pas de temps, k, avec $k = \frac{T}{N+1}$ ($N \in \mathbb{N}$), et des points de discrétisation en espace, $(x_i)_{i\in\mathbb{Z}}$. On pose $t_n = nk$, pour $n \in \{0,\dots,N+1\}$ et $h_{i+\frac12} = x_{i+1} - x_i$, pour $i \in \mathbb{Z}$. On note $\overline{u}_i^n = u(t_n, x_i)$ (pour $n \in \{0,\dots,N+1\}$ et $i \in \mathbb{Z}$) et on cherche une approximation de $\overline{u}_i^n$.

1. Soient $\alpha, \beta \in \mathbb{R}$. On suppose que, pour un certain $h \in \mathbb{R}$, $\alpha h \leqslant h_{i+1/2} \leqslant \beta h$, pour tout $i \in \mathbb{Z}$. On considère, dans cette question le schéma suivant, appelé DFDA (pour Différences Finies Décentré Amont) :

$$\frac{u_i^{n+1} - u_i^n}{k} + \frac{1}{h_{i-1/2}}(u_i^n - u_{i-1}^n) = 0 \quad n \in \{0,\dots,N\} \quad i \in \mathbb{Z} \tag{5.51}$$
$$u_i^0 = u_0(x_i) \quad i \in \mathbb{Z} \tag{5.52}$$

(a) (Stabilité) Montrer que $k \leqslant \alpha h \Rightarrow \inf(u_0) \leqslant u_i^n \leqslant \sup(u_0)$, $\forall n \in \{0,\dots,N+1\}$, $\forall i \in \mathbb{Z}$.

(b) (Convergence) Montrer que, si $k \leqslant \alpha h$, on a :

$$\sup_{i\in\mathbb{Z}} |u_i^n - \overline{u}_i^n| \leqslant CT(k+h) \quad \forall n \in \{0,\dots,N+1\}$$

où C ne dépend que de u_0 et β.

2. On suppose maintenant que x_i est le centre de la maille $M_i = [x_{i-1/2}, x_{i+1/2}]$, pour $i \in \mathbb{Z}$. On pose $h_i = x_{i+1/2} - x_{i-1/2}$. Soient $\alpha, \beta \in \mathbb{R}$. On suppose que, pour un certain $h \in \mathbb{R}$, $\alpha h \leqslant h_i \leqslant \beta h$, pour tout $i \in \mathbb{Z}$. On considère, dans cette question le schéma suivant, appelé VFDA (pour Volumes Finis Décentré Amont) :

$$h_i \frac{u_i^{n+1} - u_i^n}{k} + (u_i^n - u_{i-1}^n) = 0 \quad n \in \{0,\dots,N\} \quad i \in \mathbb{Z} \tag{5.53}$$
$$u_i^0 = \frac{1}{h_i}\int_{M_i} u_0(x)dx \quad i \in \mathbb{Z} \tag{5.54}$$

(a) (Stabilité) Montrer que $k \leqslant \alpha h \Rightarrow \inf(u_0) \leqslant u_i^n \leqslant \sup(u_0)$, $\forall n \in \{0,\dots,N+1\}$, $\forall i \in \mathbb{Z}$.

(b) Etudier la consistance du schéma au sens DF.

(c) (Convergence) On pose $\overline{\overline{u}}_i^n = u(t_n, x_{i+\frac12})$. Montrer que si $k \leqslant \alpha h$, on a :

$$\sup_{i\in\mathbb{Z}} |u_i^n - \overline{\overline{u}}_i^n| \leqslant C_1(k+h) \quad \forall n \in \{0,\dots,N+1\}$$

où C_1 ne dépend que de u_0, β et T. En déduire que :

$$\sup_{i\in\mathbb{Z}} |u_i^n - \overline{u}_i^n| \leqslant C_2(k+h) \quad \forall n \in \{0,\dots,N+1\}$$

où C_2 ne dépend que de u_0, β et T.

Exercice 74 — Équation linénaire et solution faible. *Corrigé en page 220*
Soit $u_0 \in L^\infty(\mathbb{R}) \cap L^2(\mathbb{R})$ et $T \in \mathbb{R}_+^\star$. On considère le problème suivant :

$$\partial_t u(x,t) + \partial_x u(x,t) = 0 \quad x \in \mathbb{R} \quad t \in [0,T] \tag{5.55}$$
$$u(x,0) = u_0(x) \tag{5.56}$$

Ce problème admet une et une seule solution faible, notée u. On se donne un pas de temps, k, avec $k = \frac{T}{N+1}$ ($N \in \mathbb{N}$) et on pose $t_n = nk$, pour $n \in \{0,\dots,N+1\}$; On se donne des points de discrétisation en espace, $(x_i)_{i\in\mathbb{Z}}$ et on suppose que x_i est le centre de la maille $M_i = [x_{i-1/2}, x_{i+1/2}]$,

pour $i \in \mathbb{Z}$. On pose $h_i = x_{i+1/2} - x_{i-1/2}$ et $h_{i+1/2} = x_{i+1} - x_i$. On suppose qu'il existe $\alpha, \beta \in \mathbb{R}$ et $h \in \mathbb{R}$ tels que, $\alpha h \leqslant h_i \leqslant \beta h$, pour tout $i \in \mathbb{Z}$. On considère le schéma (5.53) et (5.54).

1. (Stabilité L^∞) Montrer que $k \leqslant \alpha h \Rightarrow |u_i^n| \leqslant \|u_0\|_\infty$, $\forall n \in \{0, \dots, N+1\}$, $\forall i \in \mathbb{Z}$.

2. Montrer que, pour tout $n = 0, \dots, N$, on a $u_i^n \to 0$ lorsque $i \to +\infty$ ou $i \to -\infty$.

3. (Estimation "BV faible") Soient $\zeta > 0$. Montrer que :

$$k \leqslant (1-\zeta)\alpha h \Rightarrow \sum_{n=0,\dots,N} \sum_{i \in \mathbb{Z}} k(u_i^n - u_{i-1}^n)^2 \leqslant C(\zeta, u_0),$$

où $C(\zeta, u_0)$ ne dépend que de ζ et u_0 (multiplier (5.53) par ku_i^n et sommer sur i et n.)

4. (convergence) On pose $\mathcal{T} = (M_i)_{i \in \mathbb{Z}}$ et on définit la solution approchée sur $[0,T] \times \mathbb{R}$, notée $u_{\mathcal{T},k}$, donnée par (5.53),(5.54), par $u_{\mathcal{T},k}(t,x) = u_i^n$, si $x \in \mathcal{M}_i$ et $t \in [t_n, t_{n+1}[$.
On admet que $u_{\mathcal{T},k} \to u$, pour la topologie faible-$\star$ de $L^\infty(]0,T[\times\mathbb{R})$, quand $h \to 0$, avec $k \leqslant (1-\zeta)\alpha h$ (ζ fixé). Montrer que u est la solution faible de (5.55)-(5.56).

Remarque 5.27 — VF, DF et convergence forte. On peut montrer le même résultat avec (5.51) au lieu de (5.53). On peut aussi montrer (cf. la suite du cours...) que la convergence est forte dans $L^p_{loc}(]0,T[\times\mathbb{R})$, pour tout $p < \infty$.

Exercice 75 — Construction d'une solution faible. *Corrigé en page 223*

1. Construire une solution faible du problème $\partial_t u + (u^2)_x = 0$ avec

$$u(x,0) = u_0(x) = \begin{cases} 1 & x < 0 \\ 1-x & x \in [0\,;1] \\ 0 & x > 1 \end{cases}$$

2. Même question pour le problème $\partial_t u + (u^2)_x = 0$ avec

$$u(x,0) = u_0(x) = \begin{cases} 0 & x < 0 \\ 1-x & x \in [0\,;1] \\ 1 & x > 1 \end{cases}$$

Exercice 76 — Problème de Riemann. Soit f la fonction de $\mathbb{R}$ dans $\mathbb{R}$ définie par $f(s) = s^4$. Soit u_d et u_g des réels. Calculer la solution entropique du problème de Riemann (5.34) avec données u_d et u_g en fonction de u_d et u_g.

Exercice 77 — Non unicité des solutions faibles. *Corrigé en page 223*
On considère l'équation

$$\begin{cases} \partial_t u + (u^2)_x = 0 \\ u(0,x) = \begin{cases} u_g & si\ x < 0 \\ u_d & si\ x > 0 \end{cases} \end{cases} \tag{5.57}$$

avec $u_g < u_d$.

1. Montrer qu'il existe $\sigma \in \mathbb{R}$ tel que si

$$u(t,x) = \begin{cases} u_g & x < \sigma t \\ u_d & x > \sigma t \end{cases}$$

alors u est solution faible de (5.57). Vérifier que u n'est pas solution entropique de (5.57).

2. Montrer que u définie par :

$$\begin{cases} u(t,x) = u_g & x < 2u_g t \\ u(t,x) = \dfrac{x}{2t} & 2u_g t \leqslant x \leqslant 2u_d t \\ u(t,x) = u_d & x > 2u_d t \end{cases}$$

alors u est solution faible entropique de (5.57).

Exercice 78 — Problème de Riemann.

1. Déterminer la solution entropique de (5.34) dans le cas où f est strictement concave.

2. On se place dans le cas où f est convexe puis concave : plus précisément, on considère $f \in C^2(\mathbb{R},\mathbb{R})$ avec

(a) $f(0)=0,\ f'(0)=f'(1)=0$

(b) $\exists a \in]0,1[$, tel que f est strictement convexe sur$]0,a[$, f est strictement concave sur $]a,1[$.

On supposera de plus $u_g=1,\ u_d=0$.

(a) Soit b l'unique élément $b \in]a,1[$ tel que $\frac{f(b)}{b}=f'(b)$; montrer que u définie par :

$$\begin{cases} u(t,x)=1 & x \leqslant 0 \\ u(t,x)=\xi & x=f'(\xi)t,\ b<\xi<1 \\ u(t,x)=0 & x>f'(b)t \end{cases}$$

est la solution faible entropique de (5.34) (sous les hypothèses précédentes).

(b) Construire la solution entropique du problème de Riemann dans le cas $f(u)=\frac{u^2}{u^2+\frac{(1-u)^2}{4}}$ et $u_g,\ u_d \in [0\,;1]$. [Compliqué. On distinguera plusieurs cas.)

Exercice 79 — Stabilité de schémas numériques. *Corrigé en page 224*
Soient $f \in C^1(\mathbb{R},\mathbb{R})$ et $u_0 \in L^\infty(\mathbb{R})$. On considère le problème suivant :

$$\partial_t u(x,t)+(f(u))_x(x,t)=0,\ x\in\mathbb{R},\ t\in[0,T], \tag{5.58}$$
$$u(x,0)=u_0(x). \tag{5.59}$$

On utilise ci dessous les notations du cours. On discrétise le problème (5.58),(5.59) par l'un des schémas vu en cours ("Flux-splitting", "Godunov", "Lax-Friedrichs modifié" et "Murman"). Montrer qu'il existe M (dépendant de la fonction "flux numérique" et de u_0) tel que $k \leqslant Mh_i$, pour tout $i\in\mathbb{Z}$, implique :

1. $\|u^{n+1}\|_\infty \leqslant \|u^n\|_\infty$ pour tout $n\in\mathbb{N}$.

2. (Plus difficile) $\sum_{i\in\mathbb{Z}}|u_{i+1}^{n+1}-u_i^{n+1}| \leqslant \sum_{i\in\mathbb{Z}}|u_{i+1}^n-u_i^n|$ pour tout $n\in\mathbb{N}$. (Cette estimation n'est intéressante que si $\sum_{i\in\mathbb{Z}}|u_{i+1}^0-u_i^0|<\infty$, ce qui n'est pas toujours vrai pour $u_0\in L^\infty(\mathbb{R})$. Cela est vrai si u_0 est une fonction à "variation bornée".)

Exercice 80 — Schéma de Murman. *Corrigé en page 225*
Soient $f\in C^1(\mathbb{R},\mathbb{R})$ et $u_0\in L^\infty(\mathbb{R})$. On suppose que $A\leqslant u_0\leqslant B$, p.p. sur $\mathbb{R}$. On s'intéresse au problème suivant :

$$\partial_t u(x,t)+(f(u))_x(x,t)=0,\ x\in\mathbb{R},\ t\in\mathbb{R}_+, \tag{5.60}$$
$$u(x,0)=u_0(x),\ x\in\mathbb{R}. \tag{5.61}$$

Pour discrétiser le problème (5.60)-(5.61), on se donne un pas d'espace $h>0$ et un pas de temps $k>0$. On pose $M_i=]ih,ih+h[$ et on note u_i^n l'approximation recherchée de la solution exacte dans la maille M_i à l'instant nk. On considère le schéma de Murmann :

$$h\frac{u_i^{n+1}-u_i^n}{k}+(f_{i+\frac12}^n-f_{i-1/2}^n)=0,\ n\in\mathbb{N},\ i\in\mathbb{Z}, \tag{5.62}$$
$$u_i^0=\frac{1}{h}\int_{M_i}u_0(x)\,\mathrm{d}x,\ i\in\mathbb{Z}, \tag{5.63}$$

avec $f_{i+1/2}^n=g(u_i^n,u_{i+1}^n)$ et $g\in C(\mathbb{R}\times\mathbb{R},\mathbb{R})$ définie par $g(a,a)=f(a)$ et, pour $a\neq b$,

$$g(a,b)=\begin{cases} f(a) & \text{si } \dfrac{f(b)-f(a)}{b-a}\geqslant 0, \\ f(b) & \text{si } \dfrac{f(b)-f(a)}{b-a}<0. \end{cases}$$

1. (Stabilité) Montrer qu'il existe M, ne dépendant que de f, A et B (on donnera la valeur de M en fonction de f, A et B) t.q. pour $k\leqslant Mh$ on ait :

(a) (Stabilité L^∞) $A \leqslant u_i^n \leqslant B$, pour tout $n \in \mathbb{N}$ et tout $i \in \mathbb{Z}$,

(b) (Stabilité BV) $\sum_{i\in\mathbb{Z}} |u_{i+1}^{n+1} - u_i^{n+1}| \leqslant \sum_{i\in\mathbb{Z}} |u_{i+1}^n - u_i^n|$ pour tout $n \in \mathbb{N}$. (Cette estimation n'est intéressante que si $\sum_{i\in\mathbb{Z}} |u_{i+1}^0 - u_i^0| < \infty$, ce qui n'est pas toujours vrai pour $u_0 \in L^\infty(\mathbb{R})$. Cela est vrai si u_0 est une fonction à "variation bornée".)

2. On prend, dans cette question, $f(s) = s^2$.

(a) (Non monotonie) Montrer que si $A < 0$ et $B > 0$, la fonction g n'est pas "croissante par rapport à son premier argument et décroissante par rapport à son deuxième argument" sur $[A, B]^2$.

(b) (Exemple de non convergence) Donner un exemple de non convergence du schéma. Plus précisément, donner u_0 t.q., pour tout $h > 0$ et tout $k > 0$, on ait $u_i^n = u_i^0$ pour tout $i \in \mathbb{Z}$ et pour tout $n \in \mathbb{N}$ (la solution discrète est donc "stationnaire") et pourtant $u(., T)$ (u est la solution exacte de (5.60)-(5.61)) est différent de u_0 pour tout $T > 0$ (la solution exacte n'est donc pas stationnaire).

3. (Schéma "ordre 2", question plus difficile) Pour avoir un schéma "plus précis", on pose maintenant $p_i^n = \text{minmod}(\frac{u_{i+1}^n - u_{i-1}^n}{2h}, 2\frac{u_{i+1}^n - u_i^n}{h}, 2\frac{u_i^n - u_{i-1}^n}{h})$ et on remplaçe, dans le schéma précédent, $f_{i+1/2}^n = g(u_i^n, u_{i+1}^n)$ par $f_{i+1/2}^n = g(u_i^n + (h/2)p_i^n, u_{i+1}^n - (h/2)p_{i+1}^n)$. Reprendre les 2 questions précédentes (c'est-à-dire : "Stabilité L^∞", "Stabilité BV", "non monotonie" et "Exemple de non convergence").

Exercice 81 — Flux monotones et schéma de Godunov. Soient $f \in C^1(\mathbb{R}, \mathbb{R})$ et $u_0 \in L^\infty(\mathbb{R})$; on considère l'équation hyperbolique non linéaire :

$$\begin{cases} \partial_t u + (f(u))_x = 0, \quad (x,t) \in \mathbb{R} \times \mathbb{R}_+, \\ u(x,0) = u_0(x). \end{cases} \tag{5.64}$$

On se donne un maillage $(K_i =]x_{i-1/2}, x_{i+1/2}[)_{i\in\mathbb{Z}}$ de $\mathbb{R}$ et $k > 0$ et, pour $i \in \mathbb{Z}$, on définit une condition initiale approchée :

$$u_i^0 = \frac{1}{h_i} \int u_0(x)dx \quad \text{avec} \quad h_i = x_{i+1/2} - x_{i-1/2}$$

Pour calculer la solution entropique de l'équation (5.64), on considère un schéma de type volumes finis explicite à trois points, défini par un flux numérique g, fonction de deux variables.

1. Écrire le schéma numérique (i.e. donner l'expression de u_i^{n+1} en fonction des $(u_i^n)_{i\in\mathbb{Z}}$).

2. On suppose dans cette question que le flux g est monotone et lipschitzien en ses deux variables, c.à.d. qu'il existe $M \geqslant 0$ tel que pour tout $(x, y, z) \in \mathbb{R}^3$, $|g(x, z) - g(y, z)| \leqslant M|x - y|$ et $|g(x, y) - g(x, z)| \leqslant M|y - z|$. Montrer que le schéma numérique de la question précédente peut s'écrire sous la forme

$$u_i^{n+1} = H(u_{i-1}^n, u_i^n, u_{i+1}^n),$$

où H est une fonction croissante de ses trois arguments si k satisfait une condition de type $k \leqslant Ch_i$ pout tout i, où C est une constante à déterminer.

3. Montrer que si la fonction g est croissante par rapport à son premier argument et décroissante par rapport au second et si $a, b \in \mathbb{R}$ sont tels que $a \leqslant b$, alors $g(a, b) \leqslant g(\xi, \xi)$ pour tout $\xi \in [a, b]$.

4. En déduire que si le flux g est monotone, alors il vérifie la propriété suivante :

$$\forall (a,b) \in \mathbb{R}^2, \begin{cases} g(a,b) \leqslant \min\limits_{s\in[a,b]} f(s) \text{ si } a \leqslant b \\ g(a,b) \geqslant \max\limits_{s\in[a,b]} f(s) \text{ si } a \geqslant b. \end{cases}$$

5. Soit g un flux monotone qui est tel que pour tous $a, b \in \mathbb{R}$, il existe $u_{a,b}$ dans l'intervalle d'extrémités a et b, tel que $g(a, b) = f(u_{a,b})$. Montrer que

$$\forall (a,b) \in \mathbb{R}^2, \begin{cases} g(a,b) = \min\limits_{s\in[a,b]} f(s) \text{ si } a \leqslant b \\ g(a,b) = \max\limits_{s\in[b,a]} f(s) \text{ si } a \geqslant b. \end{cases}$$

Dans la question suivante on étudie le schéma de Godounov ; on admet pour cela un lemme de comparaison, qui est une conséquence de l'unicité de la solution entropique.

Lemme 5.28 — Principe de comparaison. Soit u solution entropique de (5.64), et v solution entropique de la même équation, où l'on a remplacé la donnée initiale u_0 par une autre donnée initiale $v_0 \in L^\infty(\mathbb{R})$. Si $u_0 \geqslant v_0$ p.p. alors $u \geqslant v$ p.p..

6. Connaissant une solution approchée u^n au temps t_n, la solution approchée au temps t_{n+1} donnée par le schéma de Godunov "historique" est définie par la moyenne par maille de la solution exacte du problème (5.64) avec donnée initiale u^n :

$$u_i^{n+1} = \frac{1}{h_i}\int_{K_i} u_G^n(x)\,\mathrm{d}x \tag{5.65}$$

où u_G^n est la solution entropique du problème (5.64) avec donnée initiale u^n, i.e. $u_0(x) = u_i^n$ pour $x \in K_i =]x_{i-1/2}, x_{i+1/2}[$.

(a) Montrer que sous une condition de type CFL qu'on précisera, on a

$$u_i^{n+1} = u_i^n - \frac{k}{h}(g_G(u_i^n, u_{i+1,n}) - g_G(u_{i-1}^n, u_{i,n})) \tag{5.66}$$

où $g_G(u_i^n, u_{i+1,n}) = f(w_R(0, u_i^n, u_i^{+1})$, et $w_R(0, u_i^n, u_i^{+1})$ est la solution exacte en $x = 0$ du problème de Riemann :

$$\begin{cases} \partial_t u + (f(u))_x = 0, \quad (x,t) \in \mathbb{R}\times\mathbb{R}_+, \\ u(x,0) = \begin{cases} u_i^n \text{ si } x < 0 \\ u_{i+1}^n \text{ si } x > 0 \end{cases} \end{cases} \tag{5.67}$$

(b) Montrer que le schéma s'écrit sous la forme

$$u_i^{n+1} = H_G(u_{i-1}^n, u_i^n, u_{i+1}^n)$$

où H_G est une fonction croissante de ses trois arguments.

(c) En déduire que g_G est un flux monotone.

(d) Montrer que le flux de Godunov ainsi défini est aussi donné par l'expression

$$g_G(a,b) = \begin{cases} \min_{s\in[a,b]} f(s) \text{ si } a \leqslant b, \\ \max_{s\in[b,a]} f(s) \text{ si } b \leqslant a. \end{cases}$$

Exercice 82 — Schémas pour les problèmes hyperboliques. Soient $f \in C^2(\mathbb{R},\mathbb{R})$, $T > 0$ et $u_0 \in L^\infty(\mathbb{R}) \cap BV(\mathbb{R})$; on cherche une approximation de la solution de l'equation hyperbolique avec condition initiale :

$$\partial_t u(x,t) + (f(u))_x(x,t) = 0,\ x \in \mathbb{R},\ t \in [0,T], \tag{5.68}$$

$$u(x,0) = u_0(x). \tag{5.69}$$

On note h (resp. $k = \frac{1}{N+1}$) le pas (constant, pour simplifier) de la discrétisation en espace (resp. en temps) et u_i^n la valeur approchée recherchée de u au temps nk dans la maille $M_i = [(i-1/2)h, (i+1/2)h]$, pour $n \in \{0,\dots,N+1\}$ et $i \in \mathbb{Z}$. On considère le schéma obtenu par une discrétisation par volumes finis explicite à trois points :

$$\frac{u_i^{n+1} - u_i^n}{k} + \frac{1}{h}(f_{i+\frac{1}{2}}^n - f_{i-1/2}^n) = 0,\ n \in \{0,\dots,N+1\},\ i \in \mathbb{Z}, \tag{5.70}$$

$$u_i^0 = \frac{1}{h}\int_{M_i} u_0(x)\,\mathrm{d}x, \tag{5.71}$$

avec $f_{i+\frac{1}{2}}^n = g(u_i^n, u_{i+1}^n)$, où $g \in C^1(\mathbb{R},\mathbb{R})$.

1. Montrer que le schéma (5.70),(5.71) possède la propriété de "consistance des flux" ssi g est telle que :

$$g(s,s) = f(s),\ \forall s \in \mathbb{R}. \tag{5.72}$$

2. Montrer que le schéma, vu comme un schéma de différences finies, est, avec la condition (5.72), d'ordre 1 (c.à.d. que l'erreur de consistance est majorée par $C(h+k)$, où C ne dépend que de f et de la solution exacte, que l'on suppose régulière). Montrer que si le pas d'espace est non constant, la condition (5.72) est (en général) insuffisante pour assurer que le schéma (5.70),(5.71) (convenablement modifié) est consistant au sens des différences finies et que le schéma est alors d'ordre 0.

3. On étudie, dans cette question, le schéma de Godunov, c.à.d. qu'on prend :

$$g(u_g, u_d) = f\big(u_{u_g,u_d}(0,t)\big),$$

où u_{u_g,u_d} est la solution du problème de Riemann :

$$\begin{aligned} \partial_t u(x,t) + (f(u))_x(x,t) &= 0, \quad (x,t) \in \mathbb{R}\times\mathbb{R}_+ && (5.73)\\ u(x,0) &= u_g \text{ si } x<0, && (5.74)\\ u(x,0) &= u_d \text{ si } x>0. && (5.75) \end{aligned}$$

(a) Montrer que le schéma (5.70), (5.71) peut s'écrire :

$$u_i^{n+1} = u_i^n + C_i(u_{i+1}^n - u_i^n) + D_i(u_{i-1}^n - u_i^n),$$

avec : $C_i = \frac{k}{h}\frac{f(u_i^n)-g(u_i^n,u_{i+1}^n)}{u_{i+1}^n-u_i^n} \geqslant 0$ et $D_i = \frac{k}{h}\frac{g(u_{i-1}^n,u_i^n)-f(u_i^n)}{u_{i-1}^n-u_i^n} \geqslant 0$.

(b) On pose $A = \|u_0\|_\infty$, $M = \sup_{s\in[-A,A]=} |f'(s)|$ et h le pas (constant) d'espace. On suppose que k et h vérifient la condition de CFL :

$$k \leqslant \frac{h}{2M}.$$

On note u^n la fonction définie par : $u^n(x) = (u_i^n)$ si $x \in M_i$; montrer que :

— Stabilité L^∞ : $\|u^{n+1}\|_\infty \leqslant \|u^n\|_\infty (\leqslant \ldots \leqslant \|u^0\|_\infty), \forall\, n \in \{0,\ldots,N+1\}$. $(E1)$

— Stabilité BV : $\|u^{n+1}\|_{BV} \leqslant \|u^n\|_{BV} (\leqslant \ldots \leqslant \|u^0\|_{BV}), \forall\, n \in \{0,\ldots,N+1\}$. $(E2)$

On rappelle que, comme u_n est une fonction constante par morceaux, on a :

$$\|u^n\|_{BV} = \sum_{i\in\mathbb{Z}} |u_{i+1}^n - u_i^n|.$$

(c) Remarque : on peut montrer (ce n'est pas facile) que si on a la condition "CFL" le schéma de Godunov converge.

4. On suppose maintenant $f(u) = au$, $a \in \mathbb{R}$ et on prend $g(\lambda,\mu) = \frac{\lambda+\mu}{2}$ (schéma centré). Montrer que pour tous $k, h > 0$, les conditions $(E1)$ et $(E2)$ sont fausses, c.à.d. qu'il existe $u_0 (\in L^\infty \cap BV)$ t.q. $\|u^1\|_\infty \not\leqslant \|u_0\|_\infty$ et $\|u_1\|_{BV} \not\leqslant \|u_0\|_{BV}$.

5. On étudie maintenant un schéma de type "MUSCL", i.e. On prend dans le schéma (5.70) $f_{i+1/2}^n = f(u_i^n + \frac{h}{2}p_i^n)$, où :

$$p_i^n = \begin{cases} \dfrac{\varepsilon_i^n}{2h}\min\big(|u_{i+1}^n - u_{i-1}^n|, 4|u_{i+1}^n - u_i^n|, 4|u_i^n - u_{i-1}^n|\big), \text{ où } \varepsilon_i^n = \operatorname{sign}(u_{i+1}^n - u_{i-1}^n) \\ \qquad\qquad \text{si } \operatorname{sign}(u_{i+1}^n - u_{i-1}^n) = \operatorname{sign}(u_{i+1}^n - u_i^n) = \operatorname{sign}(u_i^n - u_{i-1}^n) \\ 0 \text{ sinon.} \end{cases}$$

(a) Montrer que $\frac{1}{h}(f_{i+1/2}^n - f_{i-1/2}^n)$ est une approximation d'ordre 2 de $\big(f(u)\big)_x(x_i,t_n)$ aux points où $u \in C^2$ et $\partial_x u \neq 0$.

(b) Montrer que sous une condition de type $k \leqslant Ch$, où C ne dépend que de u_0 et f, les conditions de stabilité $(E1)$ et $(E2)$ sont vérifiées.

Exercice 83 — Éléments finis pour une équation hyperbolique. Soit $f \in C^1(\mathbb{R},\mathbb{R})$, $u_0 \in C(\mathbb{R})$ t.q. u_0 bornée ; on considère la loi de conservation scalaire suivante :

$$\frac{\partial u}{\partial t}(x,t) + \frac{\partial}{\partial x}(f(u))(x,t) = 0 \quad x \in \mathbb{R} \quad t \in \mathbb{R}_+, \tag{5.76}$$

avec la condition initiale :

$$u(x,0) = u_0(x).$$

On se donne un pas de discrétisation en temps constant k, on note $t_n = nk$ pour $n \in \mathbb{N}$, et on cherche à approcher $u(.,t_n)$. On note $u^{(n)}$ la solution approchée recherchée.

1. Montrer qu'une discrétisation par le schéma d'Euler explicite en temps amène au schéma en temps suivant :

$$\frac{1}{k}(u^{(n+1)} - u^{(n)}) + \frac{\partial}{\partial x}\Big(f(u^{(n)})\Big)(x) = 0,\ x \in \mathbb{R},\ n \in \mathbb{N}^*, \tag{5.77}$$

$$u^0(x) = u_0(x). \tag{5.78}$$

On cherche à discrétiser (5.77) par une méthode d'éléments finis. On se donne pour cela une famille de points $(x_i)_{i\in\mathbb{Z}} \subset \mathbb{R}$, avec $x_i < x_{i+1}$.

2. On introduit les fonctions de forme P1, notées $\Phi_i, i \in \mathbb{Z}$, des éléments finis associés au maillage donné par la famille de points $(x_i)_{i\in\mathbb{Z}}$; on effectue un développement de Galerkin de $u^{(n)}$ sur ces fonctions de forme dans (5.77) et (5.78) ; on multiplie l'équation ainsi obtenue par chaque fonction de forme et on approche le terme $f(\sum_{j\in\mathbb{Z}} u_j^{(n)}\Phi_j)$ par $\sum_{j\in\mathbb{Z}} f(u_j^{(n)})\Phi_j$ et on intègre sur $\mathbb{R}$. Montrer qu'on obtient ainsi un système d'équations de la forme :

$$\sum_{j\in\mathbb{Z}} a_{i,j}\frac{u_j^{(n+1)} - u_j^{(n)}}{k} + \sum_{j\in\mathbb{Z}} b_{i,j} f(u_j^{(n)}) = 0 \quad i \in \mathbb{Z} \quad n \in \mathbb{N}^* \tag{5.79}$$

$$u_i^0 = u_0(x_i) \quad i \in \mathbb{Z} \tag{5.80}$$

(les $a_{i,j}$ et $b_{i,j}$ sont à déterminer).

3. On effectue une *condensation de la matrice de masse* (ou mass lumping en anglais) , c'est-à-dire qu'on remplace les $a_{i,j}$ dans (5.79) par $\widetilde{a}_{i,j}$ avec $\widetilde{a}_{i,j} = 0$ si $i \neq j$ et $\widetilde{a}_{i,i} = \sum_{j\in\mathbb{Z}} a_{i,j}$. Montrer que le schéma ainsi obtenu est identique à un schéma volumes finis sur le maillage $(K_i)_{i\in\mathbb{Z}}$ où $K_i =]x_{i-1/2}, x_{i+1/2}[$, $x_{i+1/2} = (x_i + x_{i+1})/2$, avec approximation centrée du flux.

4. Montrer que ce schéma est instable, dans un (ou plusieurs) sens à préciser.

5. On remplace le flux numérique centré $F_{i+1/2}$ du schéma volumes finis obtenu à la question 3 par $G_{i+1/2} = F_{i+1/2} + D_{i+1/2}(u_i^{(n)} - u_{i+1}^{(n)})$. Montrer que l'approximation du flux reste consistante et que si les $D_{i+1/2}$ sont bien choisis, le nouveau schéma est stable sous une condition de CFL à préciser. On considère maintenant la même équation de conservation, mais sur $\mathbb{R}^2$ (avec $f \in C^1(\mathbb{R},\mathbb{R}^2)$, $u_0 \in C(\mathbb{R}^2)$, bornée :

$$\partial_t u(x,t) + \mathrm{div}(f(u))(x,t) = 0 \quad x \in \mathbb{R}^2 \quad t \in \mathbb{R}_+ \tag{5.81}$$

$$u(x,0) = u_0(x) \tag{5.82}$$

Soit $\mathcal{T}$ un maillage en triangles de $\mathbb{R}^2$, admissible pour une discrétisation par éléments finis P1. Soit $\mathcal{S}$ l'ensemble des nœuds de ce maillage et $(\Phi_j)_{j\in\mathcal{S}}$ la famille des fonctions de forme éléments finis bilinéaires P1. En conservant la même discrétisation en temps, on cherche une approximation de $u(.,t_n)$ dans l'espace engendré par les fonctions Φ_j.

6. Montrer qu'en suivant la même démarche qu'aux questions 2 et 3, on aboutit au schéma :

$$\frac{u_i^{(n+1)} - u_i^{(n)}}{k}\int_{\mathbb{R}^2}\Phi_i(x)\,\mathrm{d}x - \sum_{j\in\mathcal{S}} f(u_j^{(n)}) \cdot \int_{\mathbb{R}^2}\Phi_j(x)\nabla\Phi_i(x)\,\mathrm{d}x = 0, n \in \mathbb{N}^*$$

7. Montrer que ce schéma peut encore s'écrire :

$$\frac{u_i^{(n+1)} - u_i^{(n)}}{k}\int_{\mathbb{R}^2}\Phi_i(x)\,\mathrm{d}x + \sum_{j\in\mathcal{S}} E_{i,j} = 0, \tag{5.83}$$

avec

$$E_{i,j} = 1/2(f(u_i^{(n)}) + f(u_j^{(n)})) \cdot \int_{\mathbb{R}^2}(\Phi_i(x)\nabla\Phi_j(x) - \Phi_j(x)\nabla\Phi_i(x))\,\mathrm{d}x.$$

Montrer que ce schéma est instable.

8. Dans le schéma (5.83), on remplace $E_{i,j}$ par

$$\widetilde{E}^n_{i,j} = E^n_{i,j} + D_{i,j}(u^n_i - u^n_j),$$

où $D_{i,j} = D_{j,i}$ (pour que le schéma reste conservatif). Montrer que pour un choix judicieux de $D_{i,j}$, le schéma ainsi obtenu est à flux monotone et stable sous condition de CFL.

5.6.2 Corrigés

Exercice 69.

1. En appliquant les résultats de la section 5.2, la solution faible du problème s'écrit $u(x,t) = u_0(x+2t)$ pour $x \in \mathbb{R}$ et $t \in \mathbb{R}_+$, c'est-à-dire

$$u(x,t) = \begin{cases} 0 & \text{si } x < -2t \\ 1 & \text{si } x > -2t \end{cases} \tag{5.84}$$

La représentation graphique de la solution a $t = 0$ et à $t = 1$, en fonction de x est donnée en Figure 5.7. Cette solution faible n'est pas solution classique de (5.84) car elle n'est pas continue,

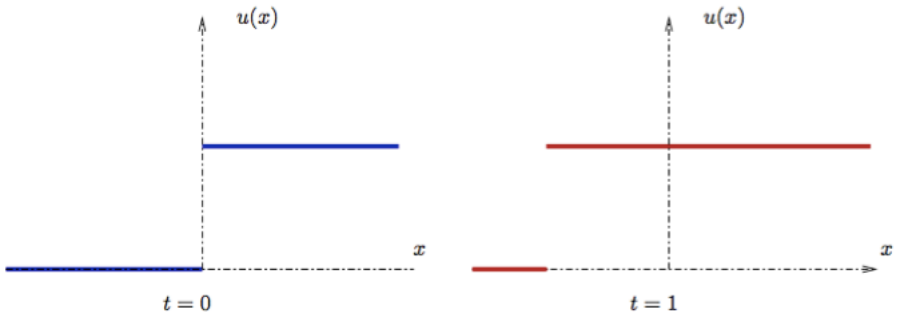


Figure 5.7 – Représentation graphique de la solution

donc ses dérivées en temps et espace ne sont pas définies partout.

2. Dans le cas où $u_0(x) = \sin x$, la solution faible du problème s'écrit $u(x,t) = \sin(x+2t)$, pour $x \in \mathbb{R}$ et $t \in \mathbb{R}_+$, et cette solution est régulière, donc solution classique.

Exercice 70. Pour $(x,t) \in \mathbb{R}^2 \times \mathbb{R}$, on pose $u(x,t) = u_0(x - \mathbf{v}t)$. Comme $u_0 \in C^1(\mathbb{R},\mathbb{R})$, on a $u \in C^1(\mathbb{R}^2 \times \mathbb{R}_+, \mathbb{R})$; on peut donc calculer les dérivées partielles de u par rapport à au temps t, qu'on notera $\partial_t u$ et par rapport aux deux variables d'espace x_1 et x_2, qu'on notera $\partial_1 u$ et $\partial_2 u$. On a $\partial_t u(x,t) = \nabla u_0(x - \mathbf{v}t) \cdot \mathbf{v}$. Or $\mathrm{div}(\mathbf{v}u) = \mathbf{v} \cdot \nabla u$ car $\mathbf{v}$ est constant et $\nabla u = \nabla u_0$. On en déduit que $\partial_t u(x,t) + \mathrm{div}(\mathbf{v}u)(x,t) = 0$ et donc u est solution (classique) de (5.45).

Exercice 71. 1. On a vu en cours qu'il y a une solution unique au problème 5.46, qui s'écrit $u(x,t) = u_0(x - at)$, et qui est donc de classe C^∞. Comme u est de classe C^∞, on peut effectuer un développement de Taylor à l'ordre 3 en temps :

$$u(x_j, t_{n+1}) = u(x_j,t_n) + k\partial_t u(x_j,t_n) + \frac{1}{2}k^2\partial^2_{tt}u(x_j,t_n) + c_u k^3 \tag{5.85}$$

où c_u.ne dépend que de u. Mais u est solution de l'équation de transport (5.46a), donc :

$$\partial_t u(x_j,t_n) + a\partial_x u(x_j,t_n) = 0 \tag{5.86}$$

Par dérivation de (5.46a) par rapport à t, on obtient : $\partial^2_{tt}u(x,t) + a\partial_{tx}u(x,t) = 0$, puis, par rapport à x : $\partial_{xt}u(x,t) + a\partial^2_{xx}u(x,t) = 0$. On en déduit que :

$$\partial^2_{tt}u(x_j,t_n) - a^2\partial^2_{xx}u(x_j,t_n) = 0. \tag{5.87}$$

On déduit le résultat souhaité de (5.85), (5.86) et (5.87), , avec $C_u = ac_u$.

Supposons maintenant que $\lambda = 1$. Soit $j \in \mathbb{Z}$ et $n \in \mathbb{N}$. Pour $s \in \mathbb{R}$, posons $\psi(s) = u(x_j+as, t_n+s)$. On a donc $\psi(0) = u(x_j,t_n)$ et $\psi(k) = u(x_j + ak, t_n + k) = u(x_{j+1}, t_{n+1})$ car $h = ak$. Comme la fonction u est régulière, la fonction ψ l'est aussi et on a : $\psi'(s) = \partial_t u(x_j + as, t_n + s) + a\partial_x u(x_j + as, t_n + s) = 0$ puisque u est solution de (5.46a). On en déduit que ψ est constante, et donc en particulier que $\psi(0) = \psi(k)$. On a donc bien $u(x_j,t_n) = u(x_{j+1}, t_{n+1})$.

2. L'égalité (5.47) suggère de rechercher une solution approchée de (5.46a) par une approximation numérique de l'équation (approchée) :

$$u(x_j,t_{n+1}) = u(x_j,t_n) - ak\partial_x u(x_j,t_n) + \frac{1}{2}a^2k^2\partial^2_{xx}u(x_j,t_n). \tag{5.88}$$

Soit u^n_j l'inconnue discrète censée approchée $u(x_j,t_n)$. Il est naturel d'approcher la dérivée partielle $\partial_x u(x_j,t_n)$ par le quotient différentiel $\frac{1}{2h}(u^{(n)}_{j+1} - u^{(n)}_{j-1})$; notons que c'est une approximation centrée, donc d'ordre 2 en espace. De même, il est naturel d'approcher la dérivée partielle $\partial^2_{xx}u(x_j,t_n)$ par le quotient différentiel $\frac{1}{h^2}(u^{(n)}_{j+1} - 2u^{(n)}_j + u^{(n)}_{j-1})$, donnant lui aussi une approximation d'ordre 2. En remplaçant chaque terme de (5.88) par son approximation, on obtient le schéma (5.48).

Lorsque $\lambda \neq 1$, on déduit immédiatement de (5.85) que le schéma est d'ordre 2. Lorsque $\lambda = 1$, le schéma s'écrit $u^{(n+1)}_j = u^{(n)}_{j-1}$. Or on a vu à la question 1 que la solution exacte vérifie $u(x_j,t_{n+1}) = u(x_{j-1},t_n)$. On en déduit que le schéma est d'ordre infini.

3. Par définition, $u^{(0)}_j = e^{ipjh}$. En appliquant le schéma (5.48) à $n = 0$, on a donc :

$$\begin{aligned} u^{(1)}_j &= e^{ipjh} - \frac{1}{2}\lambda(e^{ip(j+1)h} - e^{ip(j-1)h}) + \frac{1}{2}\lambda^2(e^{ip(j+1)h} - 2e^{ipjh} + e^{ip(j-1)h}) \\ &= u^{(0)}_j(1 - \frac{1}{2}\lambda e^{iph} + \frac{1}{2}\lambda e^{-iph} + \frac{1}{2}\lambda^2 e^{iph} - \lambda^2 + \frac{1}{2}\lambda^2 e^{-iph}) \\ &= u^{(0)}_j(1 - \lambda^2 - i\lambda\sin(ph) + \lambda^2\cos(ph)) \end{aligned}$$

On a donc : $\xi_p = 1 - \lambda^2 + \lambda^2\cos(ph) - i\lambda\sin(ph)$. Calculons le module de ξ_p :

$$\begin{aligned} |\xi_p|^2 &= (1 - \lambda^2 + \lambda^2\cos(ph))^2 + \lambda^2\sin^2(ph) \\ &= (1 - 2\lambda^2(1 - \cos ph) + \lambda^2(1 - \cos^2 ph) + \lambda^4(1 - \cos ph)^2 \\ &= (1 + \lambda^2(1 - \cos ph)(-2 + 1 - \cos ph) + \lambda^4(1 - \cos ph)^2 \\ &= 1 + \lambda^2(\lambda^2 - 1)(1 - \cos ph)^2 \end{aligned}$$

Si $\lambda \leqslant 1$, $\lambda^2(\lambda^2 - 1)(1 - \cos ph)^2 \leqslant 0$ et donc $\sup_p |\xi_p| \leqslant 1$ donc le schéma est stable. Si $\lambda > 1$, $\lambda^2(\lambda^2 - 1) > 0$ et donc $\sup_p |\xi_p| > 1$ donc le schéma est instable.

4. Si on prend $u^{(0)}_j = 1$ pour $j \geqslant 0$, $u^{(0)}_j = 0$ pour $j < 0$, le schéma de Lax-Wendroff donne : $u^{(1)}_0 = 1 + \lambda(1 - \lambda) > 1$ si $\lambda < 1$, ce qui montre que $\max_j |u^1_j| > \max_j |u^0_j|$.

Si on prend $u^{(0)}_j = 1$ pour $j \neq 0$, $u^{(0)}_0 = 0$, le schéma de Lax-Wendroff donne : $u^{(1)}_0 = \lambda^2 > 1$ si $\lambda > 1$, ce qui montre que $\max_j |u^1_j| > \max_j |u^0_j|$.

Notons que ces deux contrexemples montrent que la norme infinie de la solution approchée n'est pas décroissante, mais ceci ne démontre pas qu'elle n'est pas bornée. Ceci est une autre paire de manches...

Exercice 72. On considère le problème hyperbolique linéaire (5.9), avec $u_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, dont on calcule une solution approchée par le schéma volumes finis amont (5.18). Montrer que ce schéma est stable pour les normes L^2 et L^∞, c'est-à-dire que la solution approchée satisfait les propriétés suivantes :

1. $\|u_{\mathcal{T},k}(.,n)\|_{L^2(\mathbb{R})} \leqslant \|u_0\|_{L^2(\mathbb{R})}, \forall n \in \mathbb{N}$,

2. $\|u_{\mathcal{T},k}(.,n)\|_{L^\infty(\mathbb{R})} \leqslant \|u_0\|_{L^\infty(\mathbb{R})}, \forall n \in \mathbb{N}$,
où $u_{\mathcal{T},k}$ désigne la solution approchée calculée par le schéma (voir (5.19)). Le schéma (5.18) s'écrit encore :

$$h_i(u_i^{n+1} - u_i^n) = k(u_i^n - u_{i-1}^n).$$

Multiplions par u_i^{n+1}. On obtient :

$$\frac{h_i(u_i^{n+1} - u_i^n)^2}{2} + \frac{h_i(u_i^{n+1})^2}{2} - \frac{h_i(u_i^n)^2}{2} + k(u_i^{n+1} - u_i^n)(u_i^n - u_{i-1}^n) + ku_i^n(u_i^n - u_{i-1}^n) = 0.$$

Exercice 73.

1. (a) Le schéma numérique s'écrit :

$$u_i^{n+1} = (1 - \frac{k}{h_{i-\frac{1}{2}}})u_i^n + \frac{k}{h_{i-\frac{1}{2}}}u_{i-1}^n \tag{5.89}$$

Comme $k \leqslant \alpha h \leqslant h_{i-\frac{1}{2}}$ on a $\frac{k}{h_{i-\frac{1}{2}}} \in [0\,;1]$ On a donc $\min(u_i^n, u_{i-1}^n) \leqslant u_i^{n+1} \leqslant \max(u_i^n, u_{i-1}^n)$ d'où on déduit que $\min_j(u_j^n) \leqslant u_i^{n+1} \leqslant \max_j(u_j^n)$, $\forall i \in \mathbb{Z}$, puis, par récurrence sur n, que $\inf u_0 \leqslant u_i^n \leqslant \sup u_0$, $\forall i \in \mathbb{Z}$, $\forall n \in \mathbb{N}$.

(b) Par définition de l'erreur de consistance, on a :

$$\frac{\bar{u}_i^{n+1} - \bar{u}_i^n}{k} = \partial_t u(x_i, tn) + R_i^n \quad \text{où} \quad |R_i^n| \leqslant \|u_{tt}\|_\infty k$$
$$\frac{\bar{u}_i^n - \bar{u}_{i-1}^n}{h_{i-\frac{1}{2}}} = \partial_x u(x_i, tn) + S_i^n, |S_i^n| \leqslant \|\partial_{xx}^2 u\|_\infty h\beta$$

En posant $e_i^n = \bar{u}_i^n - u_i^n$, on a donc

$$\frac{e_i^{n+1} - e_i^n}{k} + \frac{1}{h_{i-\frac{1}{2}}}(e_i^n - e_{i-1}^n) = R_i^n + S_i^n \leqslant C(u_0, \beta)(h + k),$$

avec $C(u_0\beta) = \|u_0''\|_\infty \max(\beta, 1)$, car $(u(t,x) = u_0(x,t))$ et donc $\|u_{tt}\|_\infty = \|\partial_{xx}^2 u\|_\infty = \|u_0''\|_\infty$. On pose $C(u_0, \beta) = C$, on obtient alors

$$e_i^{n+1} = \Big(1 - \frac{k}{h_{i-\frac{1}{2}}}\Big)e_i^n + \frac{k}{h_{i-\frac{1}{2}}}e_{i-1}^n + Ck(h + k)$$

donc $\sup_i |e_i^{n+1}| \leqslant \sup_j |e_j^n| + Ck(h + k)$. Par récurrence sur n, on en déduit

$$\sup_i |e_i^n| \leqslant Ckn(h+k) \text{ et donc } \sup_i |e_i^n| \leqslant CT(h+k) \text{ si } 0 \leqslant n \leqslant N+1, \text{ où } (N+1)k = T.$$

2. (a) On a $\inf u_0 \leqslant u_i^0 \leqslant \sup u_0$ puis, pa récurrence :

$$u_i^{n+1} = \Big(1 - \frac{k}{h_i}\Big)u_i^n + \frac{k}{h_i}u_{i-1}^n.$$

Comme $k \leqslant \alpha h \leqslant h_i$ on en déduit comme en 1)a) que $\inf(u_0) \leqslant u_i^h \leqslant \sup(u_0)$.

(b) Consistance

$$\frac{\bar{u}_i^{n+1} - \bar{u}_i^n}{k} = \partial_t u(x_i, t_n) + R_i^n, |R_i^n| \leqslant \|u_{tt}\|_\infty k$$

mais

$$\frac{\bar{u}_i^n - \bar{u}_{i-1}^n}{h_i} = \frac{\bar{u}_i^n - \bar{u}_{i-1}^n}{h_{i-\frac{1}{2}}} \frac{h_{i-\frac{1}{2}}}{h_i} = [\partial_x u(x_i, t_n) + S_i^n]\frac{h_{i-\frac{1}{2}}}{h}, \text{ avec } |S_i^n| \leqslant \|\partial_{xx}^2 u\|_\infty \beta h,$$

donc

$$\frac{\bar{u}_i^{n+1} - \bar{u}_i^n}{k} + \frac{\bar{u}_i^n - \bar{u}_{i-1}^n}{h_i} = (\partial_t u + \partial_x u)(x_i, t_n) + R_i^n + S_i^n \frac{h_{i-\frac{1}{2}}}{h_i} + T_i^n = R_i^n + S_i^n \frac{h_{i-\frac{1}{2}}}{h_i} + T_i^n$$

avec :

$$|R_i^n| \leqslant \|u_{tt}\|_\infty k$$
$$\left|\frac{h_{i-\frac{1}{2}}}{h_i}\right| |S_i^n| \leqslant \|\partial_{xx}^2 u\|_\infty \frac{\beta h}{\alpha h}\beta h = \frac{\beta^2}{\alpha}\|\partial_{xx}^2 u\|_\infty h$$
$$T_i^n = \partial_x u(x_i,t_n)\frac{h_{i-\frac{1}{2}}-h_i}{h_i} = \partial_x u(x_i,t_n)\frac{h_{i-1}-h_i}{2h_i}$$

En prenant par exemple un pas tel que $h_i = h$ si i est pair et $h_i = h/2$ si i est impair, on voit T_i^n ne tend pas vers 0 lorsque h tend vers 0 ; le schéma apparait donc comme non consistant au sens des différences finies.

(c) Convergence. On a

$$\frac{\bar{\bar{u}}_i^{n+1} - \bar{\bar{u}}_i^n}{k} = \partial_t u(x_{i+\frac{1}{2}},t_n) + R_i^n, |R_i^n| \leqslant \|u_{tt}\|_\infty k \tag{5.90}$$
$$\frac{\bar{\bar{u}}_i^n - \bar{\bar{u}}_{i-1}^n}{h_i} = \partial_x u(x_{i+\frac{1}{2}},t_n) + S_i^n, |S_i^n| \leqslant \|\partial_{xx}^2 u\|_\infty \beta h \tag{5.91}$$

donc, avec $f_i^n = \bar{\bar{u}}_i^n - u_i^n$

$$f_i^{n+1} = (1-\frac{k}{h_i})f_i^n + f_{i-1}^n(\frac{k}{h_i}) + k(S_i^n + R_i^n)$$

On a donc :

$$\sup_i |f_i^{n+1}| \leqslant \sup_i |f_i^n| + k\|u_0''\|_\infty(k+\beta h) \leqslant \sup_i |f_i^n| + kC_1(k+h)$$

avec $C_1 = \|u_0''\|_\infty \max(\beta,1)$ et par récurrence sur n

$$\sup_i |f_i^n| \leqslant C_1 nk(k+h) + \|u_0'\|_\infty h\beta$$

car $\sup_i |f_i^0| \leqslant \|u_0'\|_\infty h\beta$. D'où on déduit que

$$\sup_i |f_i^n| \leqslant C_1 T(k+h) + \|u_0'\|_\infty \beta h \leqslant C_2(k+h) \quad 0 \leqslant n \leqslant N+1$$

avec $C_2 = C_1 T + \beta\|u_0'\|_\infty$ Il reste à remarquer que $|\bar{u}_i^n - \bar{\bar{u}}_i^n| \leqslant \|u_0'\|_\infty \beta h$ pour avoir

$$\sup_i |\bar{u}_i^n - u_i^n| \leqslant C_3(h+k) \quad \text{avec} \quad C_3 = C_2 + \beta\|u_0'\|_\infty = \|u_0''\|_\infty \max(\beta,1)T + 2\beta\|u_0'\|_\infty$$

Exercice 74.

1. On remarque d'abord que $|u_i^0| \in [-\|u_0\|_\infty, \|u_0\|_\infty]$. On a vu dans l'exercice 73 que $u_i^{n+1} \in [u_i^n, u_{i-1}^n[$ ou $[u_{i-1}^n, u_i^n]$. On en déduit par une récurrence sur n que $u_i^n \in [-\|u_0\|_\infty, \|u_0\|_\infty]$, $\forall i$, $\forall n \geqslant 0$.

2. On va utiliser le fait que $u_0 \in L^2$ et montrer la propriété par récurrence sur n. Pour $n = 0$, on a :

$$|u_i^0|^2 \leqslant \int_{x_{i-\frac{1}{2}}}^{x_{i-\frac{1}{2}}} (u_0(x))^2 dx \frac{1}{h_i} \to 0 \text{ lorsque } i \to \pm\infty \tag{5.92}$$

En effet, comme $u_0 1_{[x,x+\eta[} \to 0$ p.p. $u_0 1_{[x,x+\eta[} \leqslant u_0 \in L^2$ donc $\int_x^{x+\eta} |u_0|^2 \,\mathrm{d}x \to 0$ lorsque $x \to +\infty$ par convergence dominée, pour tout $\eta > 0$. De plus, $h \geqslant \alpha h$ d'où on déduit que (5.92) est vérifiée. On conclut ensuite par une récurrence immédiate sur n, que :

$$|u_i^{n+1}| \leqslant \max(|u_i^n||u_{i-1}^n|) \to 0 \text{ quand } i \to \pm\infty. \tag{5.93}$$

3. On veut montrer que $\sum_{n=0}^N \sum_{i\in\mathbb{Z}} k(u_i^n - u_{i-1}^n)^2 \leqslant C(\zeta, u_0)$. On multiplie le schéma par ku_i^n, on obtient :

$$h_i(u_i^{n+1} - u_i^n)u_i^n + (u_i^n - u_{i-1}^n)ku_i^n = 0,$$

ce qu'on peut réécrire :

$$h_i\Big(-\frac{(u_i^{n+1}-u_i^n)^2}{2}+\frac{(u_i^{n+1})^2}{2}-\frac{(u_i^n)^2}{2}\Big)+k\Big(\frac{(u_i^n-u_{i-1}^n)^2}{2}+\frac{(u_i^n)^2}{2}-\frac{(u_{i-1}^n)^2}{2}\Big)=0$$

Comme $|u_i^{n+1}-u_i^n|=k|u_i^n-u_{i-1}^n|/h_i$, ceci s'écrit aussi :

$$k\Big(1-\frac{k}{h_i}\Big)(u_i^n-u_{i-1}^n)^2+h_i(u_i^{n+1})^2-h_i(u_i^n)^2+k(u_i^n)^2-k(u_{i-1}^n)^2=0$$

et comme $k/h\leqslant 1-\zeta$, on a donc $1-k/h_i\geqslant\zeta$ et

$$\zeta(u_i^n-u_{i-1}^n)^2+h_i(u_i^{n+1})^2-h_i(u_i^n)^2+(u_i^n)^2-(u_{i-1}^n)^2\leqslant 0$$

en sommant pour $i\in\{-M,\dots,M\}$ et $h\in\{0,\dots,N\}$, on obtient alors :

$$\zeta\sum_{i=-M}^{M}\sum_{n=0}^{M}(u_i^n-u_{i-1}^n)^2+\alpha h\sum_{n=0}^{N}(u_M^n)^2-\beta h\sum_{n=0}^{N}(u_{-M-1}^n)^2\leqslant\sum_{i=-M}^{M}(u_i^0)^2.$$

En remarquant que

$$k\sum_{i=-M}^{M}(u_i^0)^2\leqslant\sum_{i=-M}^{M}h_i(u_i^0)^2\leqslant\|u_0\|_2^2$$

(voir (5.92)) et que $u_{-M}^n\to 0$ qd $M\to\infty$ (voir (5.93)), on en déduit

$$\zeta k\sum_{i=-\infty}^{\infty}\sum_{n=0}^{N}(u_i^n-u_{i-1}^n)^2\leqslant\|u_0\|_2^2,$$

donc $C=\frac{\|u_0\|_2^2}{\zeta}$ convient.

4. (Convergence) Pour montrer la convergence, on va passer à la limite sur le schéma numérique. On aura pour cela besoin du lemme suivant :

Lemme 5.29 Soit $(u_n)_{n\in\mathbb{N}}$ une suite bornée dans $L^\infty(\mathbb{R})$. Si $u_n\to u$ dans $L^\infty(\mathbb{R})$ pour la topologie faible $*$ lorsque $n\to+\infty$, (c.à.d

$$\int_{\mathbb{R}}u_n(x)\varphi(x)dx\underset{n\to+\infty}{\to}\int_{\mathbb{R}}u(x)\varphi(x)dx\qquad\forall\varphi\in L^1(\mathbb{R})$$

et $v_n\to v$ dans L^1 lorsque $n\to+\infty$, alors

$$\int u_n(x)v_n(x)dx\underset{n\to+\infty}{\to}\int u(x)v(x)\,\mathrm{d}x.$$

Démonstration.

$$\Big|\int u_n(x)v_n(x)dx-\int u(x)v(x)dx\Big|\leqslant\|u_n\|_\infty\|v_n-v\|_1+\Big|\int u_n(x)v(x)dx-\int u(x)v(x)dx\Big|$$
$$\leqslant C\|u_n-v\|_1+\Big|\int u_n(x)v_n(x)dx-\int u(x)v(x)dx\Big|\underset{n\to+\infty}{\to}0\quad(5.94)$$

car $(u_n)_n$ est bornée dans L^∞. •

On multiplie le schéma numérique par $k\varphi_i^n$, $\varphi\in C_c^\infty(\mathbb{R}\times[0,T[)$ et $\varphi_i^n=\varphi(x,t_n)$ et en somme sur i et n (toutes les sommes sont finies car φ est à support compact) ; on obtient :

$$\sum_{i\in\mathbb{Z}}\sum_{n\in\mathbb{N}}\frac{u_i^{n+1}-u_i^n}{k}kh_i\varphi_i^n+\sum_{i\in\mathbb{Z}}\sum_{n=1}^{N}(u_i^n-u_{i-1}^n)k\varphi_i^n=0.$$

Comme $\varphi_i^n=0$ si $n\geqslant N+1$, on a :

$$\sum_{i\in\mathbb{Z}}\sum_{n=1}^{N}h_iu_i^n(\varphi_i^{n-1}-\varphi_i^n)-\sum_i u_i^0\varphi_i^0h_i+\sum_{i\in\mathbb{Z}}\sum_n(\varphi_i^n-\varphi_{i+1}^n)u_i^nk=0.$$

Or :

$$T_1 = \sum_{i\in\mathbb{Z}} u_i^0 \varphi_i^0 h_i = \sum_i \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u_0(x)\varphi_0(x_i)\,\mathrm{d}x \xrightarrow[h\to 0]{} \int u_0\varphi\,\mathrm{d}x.(\text{ avec } \varphi_0 = \varphi(.,0))$$

car $\sum_i \varphi_0(x_i) 1_{]x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[} \to \varphi(.,0)$ dans L^1 quand $h \to 0$. et

$$T_2 = \sum_{i\in\mathbb{Z}}\sum_{n=1} h_i u_i^n \frac{\varphi_i^{n-1} - \varphi_i^n}{k} k = -\int_{\mathbb{R}_+}\int_{\mathbb{R}} u_{\mathcal{T},k}\psi_{\mathcal{T},k}\,\mathrm{d}x\mathrm{d}t$$

Soit

$$\psi_{\mathcal{T},k}(x,t) = \sum_{i\in\mathbb{Z}}\sum_{n=1}^{N} \frac{\varphi_i^{n-1} - \varphi_i^n}{k} 1_{]x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[} 1_{]nk,(n+1)k[}.$$

En effet, pour $x \in \mathbb{R}$ et $t > 0$,

$$\left|\frac{\varphi_i^{n-1} - \varphi_i^n}{k} - \varphi_t(x,t)\right| \leqslant k\|\varphi_{tt}\|_\infty \quad \text{si} \quad (x,t) \in]x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[\times[nk(n+1)k[, \text{ pour } n \geqslant 1$$

On a donc donc $\psi_{\mathcal{T},k} \to \varphi_t$ p.p. sur $\mathbb{R}\times]0,T[$, De plus, et $|\psi_{\mathcal{T},k}| \leqslant \|\varphi_t\|_\infty 1_K$ si $\beta h \leqslant 1$, où $K = [-a-1, a+1]\times[0,T]$, et a est tel que $\varphi = 0$ sur $([-a,a]\times[0,T])^c$ Donc, par convergence dominée, $\psi_{\mathcal{T},k} \to -\varphi_t$ dans $L^1(\mathbb{R}\times]0,T[)$ lorsque $k \to 0$. Comme $u_{\mathcal{T},k}$ converge vers u dans L^infty faible $*$, on en déduit par le lemme 5.29 que :

$$T_2 == -\int_{\mathbb{R}_+}\int_{\mathbb{R}} u_{\mathcal{T},k}(x,t)\psi_{\mathcal{T},k}(x,t)\,\mathrm{d}x\mathrm{d}t \xrightarrow[h,k\to 0]{} -\int_{\mathbb{R}_+}\int_{\mathbb{R}} u(x,t)\varphi_t(x,t)\,\mathrm{d}x\mathrm{d}t.$$

Aussi :

$$\begin{aligned} T_3 &= \sum_{i\in\mathbb{Z}}\sum_{n\in\mathbb{N}} \frac{\varphi_i^n - \varphi_{i+1}^n}{h_i} u_i^n k h_i = \sum_{i\in\mathbb{Z}}\sum_{n\in\mathbb{N}} k\varphi_i^n(u_i^n - u_{i-1}^n) \\ &= \sum_{i\in\mathbb{Z}}\sum_{n\in\mathbb{N}} k\varphi_{i-1}^n(u_i^n - u_{i-1}^n) + \sum_{i\in\mathbb{Z}}\sum_{n\in\mathbb{N}} k(\varphi_i^n - \varphi_{i-\frac{1}{2}}^n)(u_i^n - u_{i-1}^n) = T_4 + T_5 \end{aligned}$$

avec :

$$T_4 = \sum_i\sum_n k h_i \frac{\varphi_{i-\frac{1}{2}}^n - \varphi_{i+\frac{1}{2}}^n}{h} u_i^n = \iint u_{\mathcal{T}k}(x)\chi_{\mathcal{T}k}(x)dx$$

où

$$\chi_{\mathcal{T}k} = \frac{\varphi_{i-\frac{1}{2}}^n - \varphi_{i+\frac{1}{2}}^n}{h_i} \text{ sur }]x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[\times]t_n, t_{n+1}[;$$

et donc $\chi_{\mathcal{T}k} \to -\varphi_x$ dans $L^1(\mathbb{R}\times]0,1[)$ et

$$T_4 \to -\iint u(x)\varphi_x(x)\,\mathrm{d}x\mathrm{d}t \text{ lorsque } h \to 0$$

Aussi :

$$T_5 \leqslant \sum_{i=M}^{M_2}\sum_{n=0}^{N} k\beta h\|\varphi_x\|_\infty(u_i - u_{i-1}^h) \leqslant \beta k h\|\varphi_x\|_\infty \sum_{n=0}^{N}\sum_{i=M}^{M_2}(u_i^n - u_{i-1}^n)$$

si $\beta h \leqslant 1$, où M_1 et M_2 sont tels que $i \notin \{M_1,\dots,M_2\} \Rightarrow]x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[\subset [-a,a]^c$ et $\varphi = 0$ sur $([-a,a]\times[0,T])^c$. On a donc :

$$\begin{aligned} T_5 &\leqslant \beta k h\|\varphi_x\|\left(\sum_{i=M_1}^{M_2}\sum_{n=0}^{N}(u_i^n - u_{i-1}^n)^2\right)^{1/2}\left(\sum_{n=0}^{N}\sum_{i=M_1}^{M_2} 1\right)^{1/2} \leqslant \beta k h\|\varphi_x\|_\infty \frac{\sqrt{c}}{\sqrt{k}}\left(\sum_{n=0}^{N}\sum_{i=M_1}^{M_2} 1\right)^{1/2} \\ &\leqslant \beta\sqrt{k}h\|\varphi_x\|_\infty\sqrt{c}\sqrt{N+1}\sqrt{M_2 - M_1} \quad (M_2 - M_1)\alpha h \leqslant 2a \\ &\leqslant \beta\sqrt{k}h\|\varphi_x\|_\infty\sqrt{c}\frac{\sqrt{T}}{\sqrt{k}} \quad \frac{\sqrt{2}a}{\sqrt{\alpha}h} = \beta\|\psi_x\|_\infty\sqrt{c}\frac{\sqrt{T}}{\sqrt{\alpha}}\sqrt{h} \to 0 \text{ quand } h \to 0 \end{aligned}$$

On en déduit que $T_3 \to -\iint u(x)\varphi_x(x)dx$ quand $h \to 0$. Comme $T_1 + T_2 + T_3 = 0$, on a donc

$$\int_{\mathbb{R}}\int_{\mathbb{R}_+} u(x,t)\varphi_t(x;t)\,\mathrm{d}xdt + \int_{\mathbb{R}}\int_{\mathbb{R}_+} u(x,t)\varphi_x(x,t)\,\mathrm{d}xdt + \int_{\mathbb{R}} u_0(x)\varphi(x,.)\,\mathrm{d}x = 0$$

et donc u est solution faible de (5.55)-(5.56).

Exercice 75.

1. Dans le premier cas, la solution est facile à construire par la méthode des caractéristiques, pour tout $t < 1/2$. En effet, les droites caractéristiques sont d'équation : $x(t) = 2u_0(x_0)t + x_0$, c'est-à-dire

$$x(t) = \begin{cases} 2t + x_0 & \text{si } x_0 < 0 \\ 2(1-x_0)t + x_0 & \text{si } x_0 \in]0,1[\\ 0 & \text{si } x_0 > 1 \end{cases}$$

Les droites caractéristiques se rencontrent à partir de $t = 1/2$, il y alors apparition d'un choc, dont la vitesse est donnée par la relation de Rankine–Hugoniot :

$$\sigma(u_g - u_d) = (u_g^2 - u_d^2), \text{ et donc } \sigma = u_g + u_d = 1.$$

La solution entropique est donc :

$$u(x,t) = \begin{cases} 1 & \text{si } t < 1/2 \text{ et } x < 2t \quad \text{ou si } t > 1/2 \text{ et } x < t + 1/2, \\ \dfrac{x-1}{2t-1} & \\ 0 & \text{si } t < 1/2 \text{ et } x > 1 \quad \text{ou si } t > 1/2 \text{ et } x > t + 1/2. \end{cases}$$

2. On pourra montrer que la fonction définie par les formules suivantes est la solution pour $t < 1/2$ (c'est-à-dire avant que les droites caractéristiques ne se rencontrent, la solution contient deux zones de détentes) :

$$\begin{cases} u(x,t) = 0 & \text{si } x < 0 \\ u(x,t) = \dfrac{x}{2t} & \text{si } 0 < x < 2t \\ u(x,t) = \dfrac{1-x}{1-2t} & \text{si } 2t < x < 1 \\ u(x,t) = \dfrac{x-1}{2t} & \text{si } 1 < x < 1 + 2t \\ u(x,t) = 1 & \text{si } 1 + 2t < x \end{cases}$$

En $t = 1/2$, on pourra vérifier qu'un choc apparaît en $x = 1$ et se propage à la vitesse 1. On obtient alors pour $t > 1/2$ la solution suivante :

$$\begin{cases} u(x,t) = 0 & \text{si } x < 0 \\ u(x,t) = \dfrac{x}{2t} & \text{si } 0 < x < 1/2 + t \\ u(x,t) = \dfrac{x-1}{2t} & \text{si } 1/2 + t < x < 1 + 2t \\ u(x,t) = 1 & \text{si } 1 + 2t < x \end{cases}$$

Remarquons que, bien que la solution initiale soit discontinue, la solution entropique est continue pour $t \in]0, 1/2[$.

Exercice 77.

1. La question 1 découle du point 1 de la proposition 5.23 (il faut que σ satisfasse la condition de Rankine-Hugoniot.

2. La question 2 découle du point 2 de la proposition 5.23.

Exercice 79. Les quatre schémas s'écrivent sous la forme :

$$u_i^{n+1} = u_i^n - \frac{k}{h_i}(g(u_i^n, u_{i+1}^n) - g(u_i^n, u_i^n)) + \frac{k}{h_i}(g(u_{i-1}^n, u_i^n) - g(u_i^n, u_i^n))$$

soit encore

$$u_i^{n+1} = u_i^n + C_i^n(u_{i+1}^n - u_i^n) + D_i^n(u_{i-1}^n - u_i^n),$$

avec

$$C_i^n = \frac{k}{h_i}\frac{g(u_i^n,u_i^n)-g(u_i^n,u_{i+1}^n)}{u_{i+1}^n-u_i^n} \quad \text{si } u_i^n \neq u_{i+1}^n (0 \text{ sinon })$$

$$D_i^n = \frac{k}{h_i}\frac{g(u_{i-1}^n,u_i^n)-g(u_i^n,u_i^n)}{u_{i-1}^n-u_i^n} \quad \text{si } u_i^n \neq u_{i+1}^n (0 \text{ sinon })$$

On suppose que $A \leqslant u_0 \leqslant B$ p.p. et on remarque qu'il existe $L \in \mathbb{R}_+$ tel que :

$$\left.\begin{array}{l} |g(a,b) - g(a,c)| \leqslant L|b-c| \\ |g(b,a) - g(c,a)| \leqslant L|b-c| \end{array}\right\} \qquad \forall a,b,c \in [A,B]$$

(On laisse le lecteur vérifier qu'un tel L existe pour les 4 schémas considérés).

1. Dans le cas des 3 premiers schémas (FS, Godunov et LFM), la fonction g est croissante par rapport au 1er argument et décroissante par rapport au 2ème argument. Donc si $u_i^n \in [A,B], \forall i$ (pour n fixé), on a $C_i^n \geqslant 0 \quad D_i^n \geqslant 0$. En prenant $2k \leqslant Lh_i \quad \forall i$ on a aussi : $C_i^n, D_i^n \leqslant \frac{1}{2}$ et donc u_i^{n+1} est une combinaison convexe de $u_{i-1}^n, u_i^n, u_{i+1}^n$ donc $u_i^{n+1} \in [A,B] \quad \forall i$ (et aussi $\|u^{n+1}\|_\infty \leqslant \|u^n\|_\infty$). Par récurrence sur n on en déduit :

$$u_i^n \in [A,B] \quad \forall i, \forall n \text{ si } k \leqslant uh_i \forall i \text{ avec } M = \frac{L}{2}$$

Dans le dernier cas (Murman), on a
$g(a,b) = f(a)$ si $\frac{f(b)-f(a)}{b-a} \geqslant 0 \quad (a \neq b)$, $g(a,b) = f(b)$ si $\frac{f(b)-f(a)}{b-a} < 0 \quad (a \neq b)$ et $g(a,a) = f(a)$.
Si $\frac{f(u_{i+1}^n)-f(u_i^n)}{u_{i+1}^n-u_i^n} \geqslant 0$, on a : $g(u_i^n, u_{i+1}^n) = f(u^n)$, donc $C_i^n = 0$.
Si $\frac{f(u_{i+1}^n)-f(u_i^n)}{u_{i+1}^n-u_i^n} < 0$, on a : $g(u_i^n, u_{i+1}^n) = f(u_{i+1}^n), C_i^n = \frac{-f(u_{i+1}^n)+f(u^n)}{u_{i+1}^n-u_i^n} > 0$, et $C_i^n \leqslant \frac{1}{2}$ si $k \leqslant Mh_i$ avec $M = \frac{L}{2}$ (L est ici la constante de Lipschitz de f).
Le même calcul vaut pour D_i^n et on conclut comme précédemment car

$$u^{n+1} = (1 - C_i^n - D_i^n)u^n + C_i^n u_{i+1}^n + D_i^n u_{i-1}^n$$

2. On reprend la formule de 1) et la même limitation sur k (pour les 4 schémas). On a :

$$u_{i+1}^{n+1} = u_{i+1}^n + C_{i+1}^n(u_{i+2}^n - u_{i+1}^n) + D_{i+1}^n(u_i^n - u_{i+1}^n)$$
$$u_i^{n+1} = u_i^n + C_i^n(u_{i+1}^n - u_i^n) + D_i^n(u_{i-1}^n - u_i^n)$$

et donc, en soustrayant membre à membre :

$$u_{i+1}^{n+1} - u_i^{n+1} = (u_{i+1}^n - u_i^n)\underbrace{(1 - C_i^n - D_{i+1}^n)}_{\geqslant 0} + \underbrace{C_{i+1}^n}_{\geqslant 0}(u_{i+2}^n - u_{i+1}^n) + \underbrace{D_i^n}_{\geqslant 0}(u_i^n - u_{i-1}^n)$$

Par inégalité triangulaire, on a donc :

$$|u_{i+1}^{n+1} - u_i^{n+1}| \leqslant |u_{i+1}^n - u_i^n|(1 - C_i^n - D_{i+1}^n) + C_{i+1}^n|u_{i+2}^n - u_{i+1}^n| + D_i^n|u_i^n - u_{i-1}^n|$$

Sommons alors entre $i = -P$ à P :

$$\sum_{i=-P}^{P} \left|u_{i+1}^{n+1} - u_i^{n+1}\right| \leqslant \sum_{i=-P}^{P} \left|u_{i+1}^n - u_i^n\right| - \sum_{i=-P}^{P} C_i^n \left|u_{i+1}^n - u_i^n\right| + \sum_{i=-P}^{P} C_{i+1}^n \left|u_{i+2}^n - u_{i+1}^n\right|$$
$$- \sum_{i=-P}^{P} D_{i+1}^n \left|u_{i+1}^n - u_i^n\right| + \sum_{i=-P}^{P} D_i^n \left|u_i^n - u_{i-1}^n\right|.$$

En regroupant :

$$\sum_{i=-P}^{P} \left|u_{i+1}^{n+1} - u_i^{n+1}\right| \leqslant \sum_{i=-P}^{P} \left|u_{i+1}^n - u_i^n\right| + C_{P+1}^n \left|u_{P+2}^n - u_{P+1}^n\right| + D_{-P}^n \left|u_{-P}^n - u_{-P-1}^n\right|.$$

Or $C_{P+1}^n \in [0\,;1]$ et $D_{-P}^n \in [0\,;1]$ donc

$$\sum_{i=-P}^{P} \left|u_{i+1}^{n+1} - u_i^{n+1}\right| \leqslant \sum_{i=-P-1}^{P+1} \left|u_{i+1}^n - u_i^n\right| \leqslant \sum_{i=-\infty}^{+\infty} \left|u_{i+1}^n - u_i^n\right|.$$

Il ne reste plus qu'a faire tendre P vers $+\infty$ pour obtenir le résultat.

Exercice 80.

1. Cette question a été complètement traitée dans l'exercice 79.
Les estimations sont vérifiées avec $M = L/2$ où L est la constante de Lipschitz de f sur $[A, B]$.

2. Remarquons que si $f(s) = s^2$ alors $\frac{f(b)-f(a)}{b-a} = b + a$.

(a) Soit $\bar{b} \in]0, B[, \bar{a} \in]A, 0[$ tel que $\bar{b} + \bar{a} > 0$, (par exemple $\bar{a} = -\epsilon/2, \bar{b} = \epsilon$ avec $0 < \epsilon < \min(-A, B)$). Soit $\alpha \in]0, \bar{a} + \bar{b}[$. Pour $a \in [\bar{a} - \alpha, \bar{a} + \alpha]$, on a $\bar{b} + a > 0$ et donc $g(a, \bar{b}) = f(a) = a^2$, ce qui prouve que sur l'ensemble $[\bar{a} - \alpha, \bar{a} + \alpha] \times \{\bar{b}\}$, la fonction g est décroissante par rapport à a.

(b) Soit n u_0 définie par :

$$u_0 = \begin{cases} -1 \text{ sur } \mathbb{R}_- \\ +1 \text{ sur } \mathbb{R}_+ \end{cases}$$

de sorte que

$$u_i^0 = \begin{cases} +1 & \text{si } i \geqslant 0 \\ -1 & \text{si } i < 0 \end{cases}$$

Comme $f(u_i^0) = +1$, $\forall i$ on a $u_i^1 = u_i^0$, $\forall i$ et donc $u_i^n = u_i^0$ pour tout i et pour tout n. Par une récurrence facile, la solution approchée est donc stationnaire. La solution exacte n'est pas stationnaire (voir proposition 5.23, cas où f est strictement convexe et $u_g < u_d$).

Exercice 81.

1. Le schéma s'écrit :

$$u_i^{n+1} = u_i^n - \frac{h_i}{k}(g(u_i^n, u_{i+1}^n) - g(u_i^n, u_{i-1}^n))$$

2. On peut écrire le schéma sous la forme

$$u_{i+1}^n = u_i^n + C_i^n(u_{i+1}^n - u_i^n) + D_i^n(u_{i-1}^n - u_i^n) = (1 - C_i^n - D_i^n)u_i^n + C_i^n u_{i+1}^n + D_i^n u_{i-1}^n$$

avec :

$$\begin{aligned} C_i^n &= \frac{h_i}{k}\frac{g(u_i^n, u_{i+1}^n) - g(u_i^n, u_i^n)}{u_{i+1}^n - u_i^n} \quad \text{si } u_{i+1}^n \neq u_i^n \text{ (et 0 sinon)} \\ D_i^n &= \frac{h_i}{k}\frac{g(u_i^n, u_{i+1}^n) - g(u_i^n, u_i^n)}{u_{i+1}^n - u_i^n} \quad \text{si } u_{i-1}^n \neq u_i^n \text{ (et 0 sinon)} \end{aligned} \tag{5.95}$$

Remarquons que $C_i^n \geqslant 0$ et $D_i^n \geqslant 0$ car g est monotone. On en déduit que H définie par

$$H(u_{i-1}^n, u_i^n, u_{i+1}^n) = (1 - C_i^n - D_i^n)u_i^n + C_i^n u_{i+1}^n + D_i^n u_{i-1}^n,$$

où les coefficients C_i^n et D_i^n sont définis par (5.95), est une fonction croissante de ses arguments si $1 - C_i^n - D_i^n \geqslant 0$, ce qui est vérifié si $k \leqslant \frac{h_i}{2M}$ pour tout $i \in \mathbb{Z}$.

3. Comme $a \leqslant \xi$, on a $g(a, b) \leqslant g(\xi, b)$; de même, comme $\xi \leqslant b$, on a $g(\xi, b) \leqslant g(\xi, \xi)$.

4. D'après la question précédente, si $a \leqslant b$, on a bien $g(a,b) \leqslant \min\{g(\xi,\xi), \xi \in [a,b]\}$ et comme $g(\xi,\xi) = f(\xi)$, on a le résultat souhaité. Si $a \geqslant b$, alors on vérifie facilement que $g(a,b) \geqslant g(\xi,\xi)$ pour tout $\xi \in [b,a]$, ce qui prouve le résultat.

5. Comme $g(a,b) = f(u_{a,b})$, on a $\min_{s\in[a,b]} f(s) \leqslant g(a,b)$ si $a \leqslant b$ et $g(a,b) \leqslant \max_{s\in[b,a]} f(s)$ si $a \geqslant b$. On a donc égalité dans les inégalités de la question 3.

6. **(a)** Le résultat s'obtient en intégrant sur le pavé $K_i \times [t_n t_{n+1}[\ldots$

(b) D'après la question précédente, le schéma est effectivement un schéma à trois points et qui peut donc s'écrire sous la forme

$$u_i^{n+1} = H_G(u_{i-1}^n, u_i^n, u_{i+1}^n).$$

Le fait que H_G est une fonction croissante de ses trois arguments se déduit de la vision "historique" de Godunov et du lemme de comparaison : en effet si l'on prend par exemple $\widetilde{u}^n = (\widetilde{u}_i^n)_{i\in\mathbb{Z}}$ avec $\widetilde{u}_j^n = u_j^n$ pour tout $j \neq i$ et $\widetilde{u}_i^n \geqslant u_i^n$, alors par principe de comparaison $\widetilde{u}_G^n \geqslant u_G^n$ et donc $\widetilde{u}_i^{n+1} \geqslant u_i^{n+1}$. On a ainsi montré que H_G est croissante par rapport à son deuxième argument. On montre de manière similaire que H_G est croissante par rapport à son premier et troisième argument.

(c) Comme H_G est une fonction croissante de ses trois arguments, en fixant u_i^n et u_{i+1}^n , on déduit immédiatement de l'expression (5.66) que g_G est une fonction croissante du premier argument ; de même, en fixant u_i^n et u_{i-1}^n, on déduit que g_G est une fonction décroissante du deuxième argument. Le flux g_G est donc monotone.

(d) Comme $g_G(a,b) = f(w_R(0,a,b))$, il existe c dans l'intervalle d'extrémités a et b tel que $g_G(a,b) = f(c)$. Le résultat de la questions 5 donne alors que

$$g_G(a,b) = \begin{cases} \min_{s\in[a,b]} f(s) \text{ si } a \leqslant b, \\ \max_{s\in[b,a]} f(s) \text{ si } b \leqslant a. \end{cases}$$

Références

[1] Haïm Brezis. *Analyse Fonctionnelle : Théorie et Applications*. Paris : Masson, 1983.

[2] Philippe Ciarlet. « Basic error estimates for elliptic problems ». *Finite Element Methods*. Tome 2. Handbook of Numerical Analysis. Elsevier, 1991, pages 17-351.
doi : 10.1016/S1570-8659(05)80039-0.

[3] Philippe Ciarlet. *Introduction à l'analyse numérique et à l'optimisation*. Paris : Masson, 1982.

[4] Philippe Ciarlet. *The Finite Element Method for Elliptic Problems*. Paris : North-Holland, Amsterdam, 1978.

[5] Philippe Ciarlet, Bernadette Miara et Jean-Marie Thomas. *Exercices d'analyse numérique matricielle et d'optimisation avec solution*. Paris : Dunod, 1982. isbn : 2100055879.

[6] Robert Eymard, Thierry Gallouët et Raphaèle Herbin. *Finite Volume Methods. Handbook of Numerical Analysis*. Amsterdam : North-Holland, 1978, pages 713-1020.

[7] Thierry Gallouët et Raphaèle Herbin. *Équations aux dérivées partielles*. 2021.
eprint : hal.archives-ouvertes.fr/cel-01196782v4.

[8] Thierry Gallouët et Raphaèle Herbin. *Mesure, Intégration, Probabilités*. Ellipses, 2014.
oai : cel.archives-ouvertes.fr:cel-00637007.

[9] Edwige Godlewski et Pierre-Arnaud Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Tome 118. Applied Mathematical Sciences. New York : Springer, 1996.

[10] Sergeï Godounov et Valerii Platonov. *Résolution numérique des problèmes multidimensionnels de la dynamique des gaz*. Moscou : Mir, 1976.

[11] Karl Gustafson et Takehisa Abe. « The third boundary condition—was it robin's ? » *The Mathematical Intelligencer* 20 (1 1998), pages 63-71. issn : 0343-6993.
doi : 10.1007/BF03024402.

[12] Raphaèle Herbin. *Analyse numérique*. 2011.
oai : cel.archives-ouvertes.fr:cel-00092967.

[13] Dietmar Kröner. *Numerical schemes for conservation laws in two dimensions*. Advances in Numerical Mathematics. Stuttgart : John Wiley et Sons, 1997.

[14] Randall LeVeque. *Numerical methods for conservation laws*. Stuttgart : Birkhauser Verlag, 1990.

[15] Alfio Quarteroni, Riccardo Sacco et Fausto Saleri. *Numerical mathematics*. Stuttgart : Springer, 2000.

[16] Jacques Rappaz et Marco Picasso. *Introduction à l'analyse numérique*. Lausanne : Presses Polytechniques et Universitaires Romandes, 1998.

[17] Pierre-Arnaud Raviart et Jean-Marie Thomas. *Introduction à l'analyse numérique des équations aux dérivées partielles*. Paris : Dunod, 2004. isbn : 978-2100486458.