Formation en mathématiques commune à Cachan et P7

Rapport de stage de licence : Etude de la convergence de schémas aux volumes finis

P., J.-C. BERTRAND et J. GUERAND

pierre.bertrand@ens-cachan.fr, jessica.guerand@ens-cachan.fr

1er juillet 2012

sous la responsabilité de
Daniel Bouche et Frédéric P. Pascal

Introduction

De nombreux schémas numériques existent pour approcher les solutions des équations aux dérivées partielles. Ces schémas sont obtenus à partir de la méthode des différences finies, des éléments finis ou de celle des volumes finis et permettent d'obtenir des résultats de convergence satisfaisants. Pendant notre stage, nous nous sommes surtout intéressés à la méthode des volumes finis, qui respecte les lois de conservation de la physique. Nous avons étudié cette méthode dans le cadre de problèmes de convection. Bien que des résultats récents aient fait avancer la recherche en proposant des schémas convergents pour des problèmes simples tels que la convection linéaire en dimension 2, de nombreuses questions restent encore ouvertes surtout pour les problèmes en dimension 3, et les ordres de convergence.
Au début du stage, par la lecture des livres [1] et [2], nous nous sommes familiarisés avec la théorie des EDP et la méthode des différences finies qui présente une bonne introduction à celle des volumes finis et dont l'analyse mathématique est bien décrite. Nous avons au travers de ces livres étudié les notions de consistance, stabilité et convergence. Par ailleurs, après avoir repéré quelques erreurs dans le livre [1] nous les avons corrigées. Ayant notamment observé une ambiguïté dans la démonstration d'un lemme utile pour le théorème de Lax, nous proposons un éclaircissement de cette preuve en annexe. Vous pouvez également trouver en annexe notre preuve de la stabilité du schéma de l'équation de Poisson laissée en exercice dans le chapitre 10 du livre [2].
Nous avons ensuite étudié la méthode des volumes finis. Elle consiste à discrétiser l'intégration de l'équation considérée sur chaque volume d'un maillage. Nous l'avons tout d'abord appliquée à l'équation de convection linéaire en nous aidant des articles [3] et [4]. Nous avons ainsi appris de nouvelles notions liées aux volumes finis et avec le phénomène de supra-convergence. Ce dernier se manifeste par un ordre de convergence de l'erreur globale supérieur à celui de l'erreur de troncature. Pour montrer qu'un schéma est supra-convergent, on peut utiliser un correcteur du défaut de consistance et évaluer une erreur de troncature corrigée définie dans la partie 1. Nous avons appliqué cette méthode du correcteur pour le problème de convection linéaire en dimension 2 et 3 et obtenu des résultats en dimension 2 et des contre-exemples de certaines conjectures en dimension 3. Tout ceci est détaillé dans la partie 1. Ensuite, nous avons étudié l'équation de transport avec terme source en dimension 1 et montré la convergence de schémas adaptés à cette équation, notamment pour celui présenté dans l'article [5]. D'autre part, il a été remarqué numériquement que le schéma de cet article ne conservait pas la solution stationnaire. Enfin, nous avons étudié un schéma de type volumes finis adapté à l'équation des ondes en dimension 1 et 2 sous la forme d'un système d'équations de transport. Nous avons démontré la convergence de ce schéma en dimension 1 et nous avons vu les limites du correcteur en tentant de l'appliquer à la dimension 2.

Table des matières

Chapitre 1

Convergence d'un schéma aux volumes finis pour l'équation d'advection

1.1 Introduction

Dans un premier temps, nous présentons le théorème de Lax, et la méthode du correcteur, utiles pour prouver la convergence de schéma. Puis nous appliquons cette méthode au cas d'un maillage en dimension 2 pour démontrer la convergence d'un schéma discrétisant l'équation d'advection, introduit en début de partie. Nous simplifions par ailleurs l'expression du correcteur trouvée dans l'article [4] dans le cas de la dimension 2. Nous montrons également que cette simplification n'est pas valable en dimension 3. Enfin, nous calculons d'une manière originale le correcteur dans le cas d'un triangle quelconque raffiné uniformément.

1.2 Théorème de Lax, correcteur et obtention d'un schéma aux volumes finis

1.2.1 Théorème de Lax et supra-convergence

Notons L l'opérateur de différentiation associé à l'équation différentielle que l'on considère, h le pas maximal de discrétisation et L_h l'opérateur discret. En supposant u solution de l'équation différentielle $L(u) = F$, on définit l'erreur de consistance par

$$\epsilon_h = L_h(u) - F.$$

En supposant u_h solution de l'équation discrétisée $L_h(u_h) = F$, on définit l'erreur globale par

$$e_h = u_h - u.$$

Comme L_h est linéaire, on observe que

$$||L_h(e_h)|| = -\epsilon_h.$$

Un schéma est dit convergent si l'erreur globale e_h tend vers 0 quand h tend vers 0. La stabilité d'un schéma est la condition $||L_h||^{-1} \leq C$ pour une constante C indépendante de h, pour une certaine norme. Pour un schéma stable, on obtient donc

$$||e_h|| \leq c||\epsilon_h||.$$

Rappelons alors le théorème de Lax-Richtmyer qui donne une condition suffisante de convergence.

Théorème 1 (Théorème de Lax-Richtmyer) *Le schéma L_h est convergent si il est stable et consistant.*

Malheureusement, l'hypothèse de consistance n'est pas toujours vérifiée malgré la convergence du schéma. Le schéma est alors dit supra-convergent. Rappelons la définition générale de la supra-convergence.

Définition 1 *Un schéma est dit supra-convergent lorsque l'erreur de convergence se comporte mieux que l'erreur de consistance.*

Il existe une méthode pour montrer qu'un schéma est supra-convergent, le principe du correcteur.

1.2.2 Principe du correcteur

Grâce au théorème de Lax-Richtmyer, on sait que l'erreur globale a un ordre de convergence supérieur à celui de l'erreur de consistance lorsque le schéma est stable. Une condition suffisante de supra convergence est l'existence d'un correcteur γ vérifiant les propriétés suivantes :
i) $\gamma = \mathcal{O}(h^p)$
i) $\underline{\epsilon}_h = \epsilon_h - L_h(\gamma)$ est l'erreur de consistance corrigée
iii) $L_h(\gamma) = \mathcal{O}(h^{p-1})$
iv) $\underline{\epsilon}_h = \mathcal{O}(h^p)$
Grâce au théorème de Lax-Richtmyer, si le schéma est stable alors $||e_h + \gamma|| = \mathcal{O}(h^p)$ donc $||e_h|| = \mathcal{O}(h^p)$. En particulier, lorsque p est nul, l'erreur de consistance ϵ_h ne tend pas vers 0 et pourtant le schéma est convergent d'ordre 1. Cette méthode permet ainsi parfois de montrer la convergence d'un schéma sans disposer de la consistance.

1.2.3 Equation, notations et obtention du schéma

On considère l'équation d'advection scalaire

$$\begin{cases} \frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{a}u) = 0, & (x,t) \in \Omega \times]0, +\infty[\\ u(x,0) = \phi(x), & x \in \Omega \\ u(x,t) = \psi(x,t), & (x,t) \in \partial\Omega^- \times [0, +\infty[\end{cases}$$

où $\mathbf{a}$ est un vecteur non-nul, Ω (voir figure 1.1) est un domaine polygonal borné de $\mathbb{R}^d$, $\partial\Omega^-$ est $\{x \in \partial\Omega, \mathbf{a} \cdot \mathbf{n}(x) < 0\}$ avec $\mathbf{n}$ la normale unitaire extérieure à Ω, ϕ et ψ devant vérifier la condition de compatibilité

$$\psi(x,0) = \phi(x), \qquad x \in \partial\Omega^-.$$

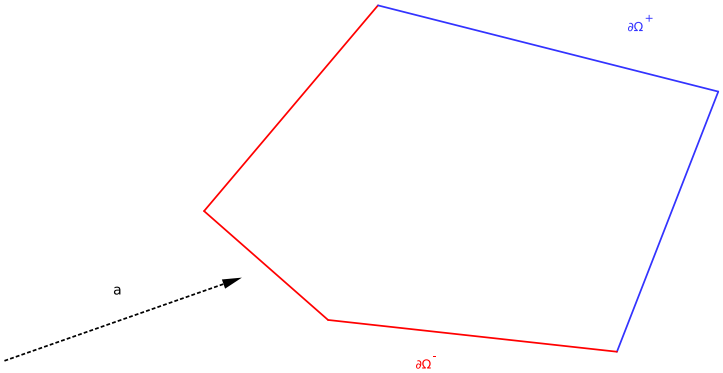

FIGURE 1.1 – Domaine Ω

On pose $\mathcal{T} = \{K_j : j = 1, .., N\}$ une partition de Ω en polyèdres K_j telle que l'hyperface entre deux volumes adjacents soit incluse dans un hyperplan. Pour $j \in [1, N]$, on définit les ensembles
– $\mathcal{N}_0(j) = \{k : |K_k \cap K_j| \neq 0\}$ des indices des volumes voisins de K_j
– $\mathcal{N}_b(j) = \{l : F_l$ hyperface de K_j et $F_l \subset \partial\Omega\}$ des indices des hyperfaces de bord de K_j
– $\mathcal{N}(j) = \mathcal{N}_0(j) \cup \mathcal{N}_b(j)$ des indices des volumes voisins et des hyperfaces de bord
Pour k $\in \mathcal{N}_0(j)$ on désigne par $\mathbf{n_{j,k}}$ la normale unitaire sur $K_j \cap K_k$ pointant de K_j vers K_k et on pose $\mathbf{N_{j,k}} = |K_j \cap K_k|\mathbf{n_{j,k}}$. Si $l \in \mathcal{N}_b(j)$, on désigne par K_l le symétrique de K_j par rapport à F_l et on garde les mêmes notations. On distingue $\mathcal{N}_j^+$ et $\mathcal{N}_j^-$ selon le signe de $\mathbf{a} \cdot \mathbf{n}_{j,k}$. On résume les notations en figure 1.2.

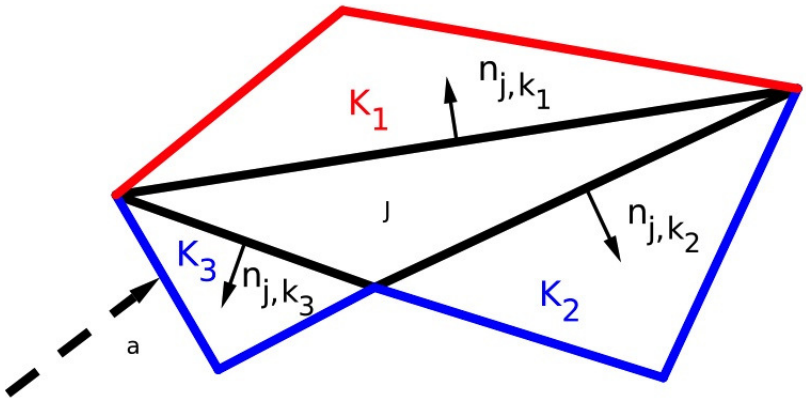

FIGURE 1.2 – Notations

On utilise alors le schéma suivant de type volume fini explicite en temps, décentré amont où u_j^n est une approximation de $\frac{1}{K_j}\int_{K_j} u(x_j,t^n)dx$,

$$\frac{u_j^{n+1}-u_j^n}{\Delta t^n}+\frac{1}{|K_j|}\left(\sum_{k\in\mathcal{N}^+(j)}\mathbf{a}.\mathbf{N}_{j,k}u_j^n+\sum_{k\in\mathcal{N}^-(j)}\mathbf{a}.\mathbf{N}_{j,k}u_k^n\right)=0,$$

où $\Delta t^n = t^n - t^{n-1}$. Les auteurs des articles [3] et [4] montrent la convergence de ce schéma en utilisant la méthode du correcteur dans le cas $p=1$. Nous allons maintenant détailler les points que nous avons simplifiés ou les nouveaux résultats démontrés. On définit la norme p par

$$||\xi||_p=\left(\sum_{j=1}^N |K_j||\xi_j|^p\right)^{1/p},$$

et la norme infinie par

$$||\xi||_\infty=\max_{1\le j\le N}|\xi_j|.$$

On introduit l'opérateur $\mathcal{L}^n$ définit de $\mathbb{R}^N$ dans $\mathbb{R}^N$,

$$\mathcal{L}_j^n((\xi_k)_{k=1}^N)=\xi_j-\frac{\Delta t_n}{|K_j|}\left(\sum_{k\in\mathcal{N}^+(j)}\mathbf{a}.\mathbf{N}_{j,k}\xi_j+\sum_{k\in\mathcal{N}_0^-(j)}\mathbf{a}.\mathbf{N}_{j,k}\xi_k\right).$$

On montre grâce à la formule de la divergence l'égalité suivante qui sera utilisée par la suite

$$\sum_{k\in\mathcal{N}_{j,k}}\mathbf{a}\cdot\mathbf{N}_{j,k}=0. \tag{1.1}$$

Pour montrer la convergence, il suffisait, d'après l'article [3] de montrer l'inégalité

$$||e^n||_p\le||e^0||_p+\sum_{i=0}^{n-1}\Delta t^i||\epsilon^i||_p,$$

satisfaisant la CFL

$$\Delta t^n\le\frac{\min_j|K_j|}{\sum\limits_{k\in\mathcal{N}^+(j)}\mathbf{a}.\mathbf{N}_{j,k}}. \tag{1.2}$$

Cette inégalité découle aisément de l'inégalité de stabilité montrée dans la proposition 2 de la partie suivante.

1.3 Démonstration de l'inégalité de stabilité

Nous avons montré l'inégalité de stabilité énoncée dans les articles [3] et [4].

Proposition 2 *Inégalité de stabilité*

$$||\mathcal{L}^n(\xi)||_p \leq ||\xi||_p, \quad \forall \xi \in \mathbb{R}^N$$

Preuve. On se ramène au cas où ξ n'a que des composantes positives car $||\mathcal{L}^n(\xi)||_p \leq ||\mathcal{L}^n(|\xi|)||_p$ par inégalité triangulaire. Dans le cas où $p = 1$, on a

$$\begin{aligned} |K_j||\mathcal{L}_j^n(\xi)| &= |K_j| \left| (1 - \tfrac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k})\xi_j - \tfrac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_k \right|, \\ &= |Kj|\xi_j - \Delta t^n \sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_j - \Delta t^n \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_k. \end{aligned}$$

En sommant sur j, on obtient

$$\begin{aligned} \sum_{j=1}^N |K_j||\mathcal{L}_j^n(\xi)| &= \sum_{j=1}^N |K_j|\xi_j - \Delta t^n \left(\sum_{j=1}^N \sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_j + \sum_{j=1}^N \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_k \right), \\ &\leq \textstyle\sum_{j=1}^N |K_j|\xi_j - \Delta t^n \left(\sum_{j=1}^N \sum_{k \in \mathcal{N}_0^+(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_j + \sum_{j=1}^N \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_k \right), \end{aligned}$$

puis en utilisant $\mathbf{a}.\mathbf{N}_{j,k}\xi_j = -\mathbf{a}.\mathbf{N}_{k,j}\xi_j$,

$$-\sum_{j=1}^N \sum_{k \in \mathcal{N}_0^+(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_j = \sum_{j=1}^N \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k}\xi_k,$$

d'où

$$\sum_{j=1}^N |K_j||\mathcal{L}_j^n(\xi)| \leq \sum_{j=1}^N |K_j|\xi_j.$$

Dans le cas où $p = \infty$, on définit l'indice j_0 tel que $|\mathcal{L}_{j_0}^n(\xi)| = \max_{1 \leq j \leq N} |\mathcal{L}_j^n(\xi)|$.
Comme $1 - \frac{\Delta t^n}{|K_{j_0}|} \sum_{k \in \mathcal{N}^+(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k}$ est positif d'après la CFL (1.2) et comme $-\sum_{k \in \mathcal{N}_0^-(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k}\xi_k$ est aussi positif, par définition, on a

$$\begin{aligned} |\mathcal{L}_{j_0}^n(\xi)| &= \xi_{j_0} \left(1 - \tfrac{\Delta t^n}{|K_{j_0}|} \sum_{k \in \mathcal{N}^+(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k} \right) - \tfrac{\Delta t^n}{|K_{j_0}|} \sum_{k \in \mathcal{N}_0^-(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k}\xi_k, \\ &\leq \left(1 - \tfrac{\Delta t^n}{|K_{j_0}|} \sum_{k \in \mathcal{N}^+(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k} - \tfrac{\Delta t^n}{|K_{j_0}|} \sum_{k \in \mathcal{N}_0^-(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k} \right) ||\xi||_\infty, \\ &\leq ||\xi||_\infty, \end{aligned}$$

car $1 - \frac{\Delta t^n}{|K_{j_0}|} \sum_{k \in \mathcal{N}^+(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k} - \frac{\Delta t^n}{|K_{j_0}|} \sum_{k \in \mathcal{N}_0^-(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k} = (1 + \frac{\Delta t^n}{|K_{j_0}|} \sum_{k \in \mathcal{N}_b^-(j_0)} \mathbf{a}.\mathbf{N}_{j_0,k})$ est dans $[0, 1]$, d'après la CFL (1.2). On en déduit alors que

$$||\mathcal{L}^n(\xi)||_\infty \leq ||\xi||_\infty.$$

Dans le cas où $p > 1$, on va montrer que

$$||\mathcal{L}^n(\xi)||_p^p \leq ||\mathcal{L}^n(\xi^p)||_1 \leq ||\xi^p||_1 = ||\xi||_p^p.$$

L'égalité est triviale et la deuxième inégalité résulte du cas $p = 1$. Il nous reste à montrer la première inégalité. On sait que

$$||\mathcal{L}^n(\xi)||_p^p = \sum_{j=1}^{n} |K_j| \left(\left(1 - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k} \right) \xi_j - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k} \xi_k \right)^p.$$

Or par convexité de la fonction $x \mapsto x^p$ pour $x \geq 0$ et $p > 1$, en prenant $y \geq 0$,

$$\left(\left(1 - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k} \right) \xi_j - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k} \xi_k - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}_b^-(j)} \mathbf{a}.\mathbf{N}_{j,k} y \right)^p$$

$$\leq \left(1 - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k} \right) \xi_j^p - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k} \xi_k^p - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}_b^-(j)} \mathbf{a}.\mathbf{N}_{j,k} y^p,$$

car $1 - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k} - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k} - \frac{\Delta t^n}{|K_j|} \sum_{k \in \mathcal{N}_b^-(j)} \mathbf{a}.\mathbf{N}_{j,k} = 1$, d'après (1.1) avec des termes positifs donc en prenant $y = 0$, en sommant sur j et en multipliant par $|K_j|$, on obtient le résultat recherché $||\mathcal{L}^n(\xi)||_p^p \leq ||\mathcal{L}^n(\xi^p)||_1$. ■

1.4 Existence de γ

Dans l'article [3], le correcteur γ est défini à partir d'un correcteur géométrique $\mathbf{\Gamma}$ par $\gamma_j^n = \mathbf{\Gamma}_j \cdot \nabla u(g_j, t^n)$ où g_j est le centre de gravité du volume K_j et où $g_{j,k}$ désigne quant à lui le centre de gravité de la face entre K_k et K_j. L'existence d'un vecteur $\mathbf{\Gamma}$ vérifiant le d système linéaire

$$\sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k}(\mathbf{\Gamma}_j - g_{j,k} + g_j) + \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k}(\mathbf{\Gamma}_k - g_{j,k} + g_k) = 0 \tag{1.3}$$

assure celle d'un correcteur γ associé au schéma que l'on considère.
On peut réécrire l'équation (1.3)

$$\sum_{k \in \mathcal{N}(j)} \mathbf{a}^+(\mathbf{N}_{j,k})(\mathbf{\Gamma}_j - g_{j,k} + g_j) + \sum_{k \in \mathcal{N}_0(j)} \mathbf{a}^-(\mathbf{N}_{j,k})(\mathbf{\Gamma}_k - g_{j,k} + g_k) = 0.$$

On a utilisé les notations

$$\mathbf{a}^+(N_{j,k}) = \frac{\mathbf{a} \cdot \mathbf{N}_{j,k} + sign(\mathbf{a} \cdot \mathbf{N}_{j,k}) \mathbf{a} \cdot \mathbf{N}_{j,k}}{2} \quad \text{et} \quad \mathbf{a}^-(N_{j,k}) = \frac{\mathbf{a} \cdot \mathbf{N}_{j,k} - sign(\mathbf{a} \cdot \mathbf{N}_{j,k}) \mathbf{a} \cdot \mathbf{N}_{j,k}}{2}$$

Les auteurs de l'article [4] prouvent l'existence et l'unicité de $\mathbf{\Gamma}$.
On définit alors

$$(BX)_j = \frac{\sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k} X_k}{\sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k}}$$

ainsi que

$$\Delta_j = \frac{\sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k}(g_{j,k} - g_j) + \sum_{k \in \mathcal{N}_0^-(j)} \mathbf{a}.\mathbf{N}_{j,k}(g_{j,k} - g_k)}{\sum_{k \in \mathcal{N}^+(j)} \mathbf{a}.\mathbf{N}_{j,k}}.$$

Le correcteur Γ vérifie alors

$$(I - B)\mathbf{\Gamma} = \Delta.$$

L'article [4] montrait que $|Spec(B)| < 1$, on avait ainsi l'existence de γ sous la forme d'une éventuelle somme infinie. Cependant, nous avons montré que la matrice B était nilpotente en dimension 2 et ne l'était pas en dimension 3. On définit pour cela la notion de cycle dans le cas où les volumes sont des triangles et l'on prouve une série de lemmes.

Théorème 3 *La matrice B est nilpotente pour un maillage en dimension 2.*

Pour démontrer ce théorème, prouvons d'abord trois lemmes utiles.

Définition 2 *Un cycle de taille r est un ensemble de triangles (ou tétraèdres)* $K_0, K_1, ..., K_{r-1}$ *tel qu'il existe une permutation* σ *telle que* $a.N_{\sigma(i[r])\sigma(i+1[r])} < 0$ *pour tout entier i dans* $[0, r-1]$.

On voit sur la figure 1.3, qu'il semble difficile d'obtenir un cycle en dimension 2. On peut voir sur la figure 1.5,

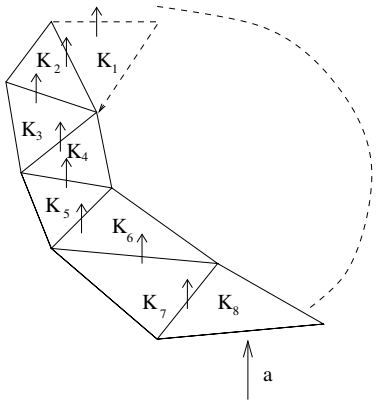

FIGURE 1.3 – Cycle impossible en 2D

qu'on obtient un $\mathcal{N}_0^-$ vide au moins pour ce maillage. Généralisons ce résultat avec le lemme suivant.

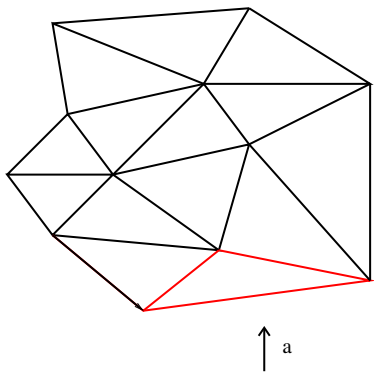

FIGURE 1.4 – Existence d'un $\mathcal{N}_0^-$ vide

Lemme 4 *Pour un maillage donné en dimension 2, il existe un triangle du bord dont* $\mathcal{N}_0^-$ *est vide.*

Preuve. On procède par récurrence sur le nombre N de triangles du maillage. Pour $N = 1$, c'est évident. Pour $N \geq 1$, on suppose le résultat vrai pour tout $k \leq N$. On considère un maillage M de $N+1$ triangles. On utilise le lemme 4.5 (valable également pour les maillages avec trous car il suffit de mailler les trous) de l'article [4] qui nous permet, s'il existe au moins un côté intérieur au maillage non parallèle à $\mathbf{a}$, de séparer le maillage par une ligne brisée en deux maillages non vide A et B comme on peut le voir sur la figure 1.5, sinon $\mathcal{N}_0^-$ est vide pour tout triangle. Le vecteur $\mathbf{a}$ est dirigé de A vers B au niveau de la ligne brisée. On considère le maillage A, il possède au moins un triangle T avec un $\mathcal{N}_0^-$ vide par hypothèse de récurrence . Ce triangle T est sur le bord entrant de A et donc de M. ■
On peut alors montrer le lemme suivant.

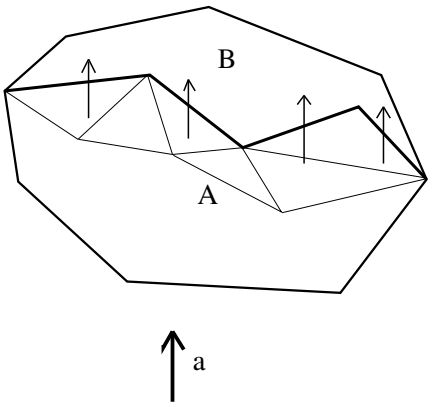

FIGURE 1.5 – Ligne brisée séparant le maillage en A et B

Lemme 5 *Pour tout $r \in \mathbb{N}^\star$, il n'existe pas de cycle de longueur r.*

Preuve. Par l'absurde on suppose qu'il existe un cycle C de longueur r. On considère le maillage constitué par le cycle. D'après le lemme 4, un triangle du cycle possède un $\mathcal{N}_0^-$ vide, ce qui contredit la définition du cycle. ■

Lemme 6 *Soit M un maillage,il existe un cycle si et seulement si B est non nilpotente.*

Preuve. On suppose qu'il existe un cycle.
On note $B = (B_{ij})_{0 \leq i,j \leq N}$ avec $a_{ij} \geq 0$ et $B^N = \sum_{k_1,\ldots,k_{N+1}} B_{k_1 k_2} ... B_{k_N k_{N+1}} E_{k_1 k_{N+1}}$.
Comme il existe un cycle, il existe $r_1, r_2, ..., r_p$ tel que $B_{r_1 r_2} B_{r_2 r_3} ... B_{r_{p-1} r_p} B_{r_p r_1} \neq 0$, donc B^N, qui est une somme de terme positif, possède un coefficient strictement positif (celui où on met le cycle bout à bout). On déduit alors que B^N est non nulle, donc B est non nilpotente.
Pour la réciproque, on montre la contraposée. On suppose qu'il n'y a aucun cycle. Donc pour tous $k_1, k_2, ..., k_r$ entiers compris entre 1 et N, $\prod_{i=1}^{r} B_{k_i\, k_{i+1}} = 0$,où $B_{k_r\, k_{r+1}} = B_{k_r\, k_1}$. Or comme $B = \sum_{i,j} a_{ij} E_{ij}$, on déduit que $B^N = \sum_{k_1,\ldots,k_{N+1}} B_{k_1 k_2} ... B_{k_N k_{N+1}} E_{k_1 k_{N+1}}$. D'après le lemme des tiroirs, il existe donc $k_i = k_j$ avec $i < j$ pour tous les termes donc tous les termes de la somme sont nuls. Par suite, on obtient $B^N = 0$. ■

Démonstration. Montrons le théorème dans le cas de la dimension 2. On note $B = (B_{ij})_{i,j}$ avec $B_{ij} = \begin{cases} \varepsilon_{i,j} \neq 0 \, si \, \mathbf{a}.\mathbf{N}_{ij} < 0 \\ 0 \, sinon \end{cases}$. D'après les lemmes 5 et 6, B est nilpotente. ■
On aurait aimé que B soit aussi nilpotente comme le montre nos recherches sur les ombrelles et les parapluies (annexe). En réalité, B n'est pas nilpotente. Pour montrer que B n'est pas nécessairement nilpotente, il suffit d'exhiber un cycle en dimension 3 pour un certain vecteur $\mathbf{a}$, d'après le lemme 6. Donnons un contre-exemple au

théorème dans le cas de la dimension 3. En prenant le vecteur $\mathbf{a} = (0, 0, 1)$

On donne le tableau des points du maillage

points	1	2	3	4	5	6	7	8	9
x	0	$\frac{1}{2}$	$-\frac{1}{2}$	0	1	-1	1	- 1	0
y	0	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$	$\sqrt{3}$	0	0	$\sqrt{3} - \frac{1}{\sqrt{3}}$	$\sqrt{3} - \frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$
z	0	0	0	0	0	0	0	0	0

points	10	11	12	13	14	15	16	17	18
x	$\frac{5}{8}$	$\frac{1}{2}$	$-\frac{1}{8}$	$-\frac{1}{2}$	$-\frac{7}{8}$	$-\frac{3}{4}$	$-\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{4}$
y	$\frac{\sqrt{3}}{4}$	$-\frac{1}{8}$	$\frac{1}{2\sqrt{3}}$	$\frac{\sqrt{3}}{16}$	$\frac{\sqrt{3}}{4} + \frac{1}{8}$	$\frac{3\sqrt{3}}{4} - \frac{1}{3\sqrt{3}}$	$\frac{3\sqrt{3}}{4}$	$\frac{3\sqrt{3}}{4}$	$\frac{3\sqrt{3}}{4} - \frac{2}{3\sqrt{3}}$
z	1	1	1	1	1	1	1	1	1

La liste des tétraèdres est la suivante : (2 5 7 10) ; (2 1 5 11) ; (9 1 5 12) ; (1 6 9 13) ; (1 3 6 14) ; (3 6 8 15) ; (3 4 8 16) ; (2 3 4 17) ; (2 4 7 18) ; (2 5 10 11) ; (5 1 11 12) ; (1 9 12 13) ; (1 6 13 14) ; (3 6 14 15) ; (3 8 15 16) ; (3 4 16 17) ; (2 4 17 18) ; (2 7 18 10). Voici enfin un aperçu du cycle en dimension 3, en figure 1.6.

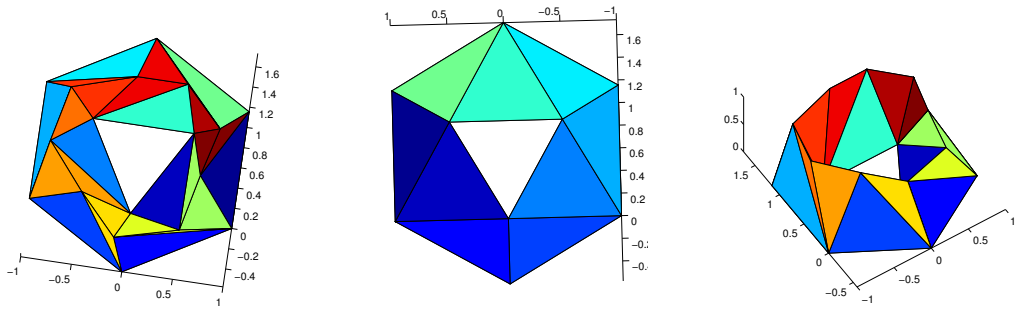

FIGURE 1.6 – Trois vues du Cycle 3D

1.5 Estimation de Γ dans le cas d'un triangle

On considère un triangle quelconque de longueur de plus grand coté l que l'on raffine uniformément r fois (la longueur du coté des triangles du maillage est $\frac{l}{2^r}$. Chaque petit triangle j possède un correcteur $\mathbf{\Gamma}_j$ et le but est de montrer que $\mathbf{\Gamma}_j = \mathcal{O}(\frac{1}{2^r})$. Les auteurs de l'article [4] le démontrent avec de lourds calculs dus au fait que le triangle est quelconque. Nous procédons d'une autre manière. On commence par montrer le théorème suivant

Théorème 7 *Si Ω' est l'image par f de Ω avec f affine alors $\mathbf{\Gamma}'$ = f($\mathbf{\Gamma}$).*

Ainsi, il suffit de faire les calculs sur un triangle équilatéral puis de passer d'un triangle équilatéral à un triangle quelconque par f linéaire. Si le correcteur d'un triangle équilatéral est un $\mathcal{O}(\frac{1}{2^r})$ alors celui d'un triangle quelconque l'est aussi.

1.5.1 Passage d'un triangle équilatéral à un triangle quelconque (démonstration du théorème)

On montre que si Ω' est l'image par f de Ω avec f affine alors $\mathbf{\Gamma}'$ = f($\mathbf{\Gamma}$). Il s'agit de montrer que si $\mathbf{\Gamma}$ vérifie (1.3) alors f($\mathbf{\Gamma}$) vérifie (1.3) où $\mathbf{a}$ est devenu f($\mathbf{a}$), les $g_{i,j}$ sont devenus f($g_{i,j}$) et $\mathbf{N}_{i,j}$ est la normale entre les triangles de l'image du pavage. La seule difficulté réside dans $\mathbf{a} \cdot \mathbf{N}_{i,j}$, le reste étant linéaire. On montre donc le

Lemme 8 *Pour tout couple de points (M_1, M_2) de $\mathbb{R}^2$, pour toute fonction affine f de $\mathbb{R}^2$ dans $\mathbb{R}^2$ on a la relation*

$$det(J_f)a \cdot M_1M_2^{\perp} = f(a) \cdot [f(M_1)f(M_2)]^{\perp}.$$

Preuve. On pose

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Et l'on définit A matrice de taille (2,1) ainsi que F matrice de taille (2,2) représentant respectivement $\vec{a}$ et la partie linéaire de f.

On a $J\vec{u} = \vec{u}^{\perp}$
Ainsi

$$\begin{aligned} & f(a) \cdot [f(M_1)f(M_2)]^{\perp} = f(a) \cdot [f(M_1) - f(M_2)]^{\perp} \\ & = f(a) \cdot [Jf(M_1) - Jf(M_2)] \\ & = A^T F^T JF(M_2 - M_1) \\ & = A^T F^T JFJ^{-1}(M_2 - M_1)^{\perp} \end{aligned}$$

On calcule donc JFJ^{-1}

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -c & -d \\ a & b \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$
$$= Com(F),$$

s'en suit $F^t JFJ^{-1} = det(f)I_2$ et le lemme est ainsi établi. ■

Démonstration du théorème 7. On montre que si $\mathbf{\Gamma}$ vérifie (1.3) alors $\mathbf{\Gamma}'$ vérifie (1.3) pour $f(\Omega)$. Réécrivons (1.3)

$$\sum_{k\in\mathcal{N}^+(j)} \mathbf{a}\cdot\mathbf{N}_{j,k}(\mathbf{\Gamma}_j - g_{j,k} + g_j) + \sum_{k\in\mathcal{N}_0^-(j)} \mathbf{a}\cdot\mathbf{N}_{j,k}(\mathbf{\Gamma}_k - g_{j,k} + g_k) = 0.$$

On obtient alors par linéarité de f

$$\sum_{k\in\mathcal{N}^+(j)} \mathbf{a}\cdot\mathbf{N}_{j,k} f(\mathbf{\Gamma}_j - g_{j,k} + g_j) + \sum_{k\in\mathcal{N}_0^-(j)} \mathbf{a}\cdot\mathbf{N}_{j,k} f(\mathbf{\Gamma}_k - g_{j,k} + g_k) = 0.$$

Soit

$$\sum_{k\in\mathcal{N}^+(j)} \mathbf{a}\cdot\mathbf{N}_{j,k}(\mathbf{\Gamma}'_j - g'_{j,k} + g'_j) + \sum_{k\in\mathcal{N}_0^-(j)} \mathbf{a}\cdot\mathbf{N}_{j,k}(\mathbf{\Gamma}'_k - g'_{j,k} + g'_k) = 0.$$

On multiplie alors par $det(J_f)$ pour avoir

$$\sum_{k\in\mathcal{N}^+(j)} det(J_f)\mathbf{a}\cdot\mathbf{N}_{j,k}(\mathbf{\Gamma}'_j - g'_{j,k} + g'_j) + \sum_{k\in\mathcal{N}_0^-(j)} det(J_f)\mathbf{a}\cdot\mathbf{N}_{j,k}(\mathbf{\Gamma}'_k - g'_{j,k} + g'_k) = 0.$$

En utilisant le lemme (8), l'égalité se réécrit

$$\sum_{k\in\mathcal{N}^+(j)} f(a)\cdot N'_{j,k}(\mathbf{\Gamma}'_j - g'_{j,k} + g'_j) + \sum_{k\in\mathcal{N}_0^-(j)} f(a)\cdot N'_{j,k}(\mathbf{\Gamma}'_k - g'_{j,k} + g'_k) = 0.$$

L'expression précédente signifie exactement que $f(\mathbf{\Gamma})$ est solution de (1.3) pour $f(\Omega)$. On a donc montré par unicité de $\mathbf{\Gamma}'$, $\mathbf{\Gamma}' = f(\mathbf{\Gamma})$. ■
Le but de la partie suivante est de démontrer que le correcteur $\mathbf{\Gamma}$ est un $\mathcal{O}(\frac{1}{2^r})$ dans le cas d'un triangle équilatéral.

1.5.2 Cas d'un triangle équilatéral

On se donne un triangle équilatéral divisé r fois uniformément. On considère un petit triangle du maillage de coté $\frac{l}{2^r}$ que l'on appelle triangle j. On suppose ce triangle suffisamment loin du bord pour avoir autant de voisins que dans la figure 1.7. On veut exprimer le correcteur de ce triangle en fonction de ceux de ses voisins. Pour le bord, on connait le correcteur qui vaut $\epsilon = \mathcal{O}(\frac{1}{2^r})$. On s'occupe ici seulement du cas où $\vec{a}$ rentre par deux cotés. Le deuxième cas (lorsque $\vec{a}$) rentre par un coté est disponible en annexe.

Intérieur. L'angle θ est ici entre 0 et $\frac{\pi}{3}$. On s'occupe d'abord du cas où le triangle n'est pas sur le bord, on isole une telle structure dans le maillage (figure 1.7) : on calcule $\mathbf{\Gamma}_j^+$ en fonction de ceux qui l'entourent. Le but est d'obtenir une formule de récurrence entre des $\mathbf{\Gamma}^+$ uniquement (il s'agit donc d'éliminer les $\mathbf{\Gamma}^-$). On utilise la formule (1.3)

$$\sum_{k\in\mathcal{N}^+(j)} \mathbf{a}\cdot\mathbf{N}_{j,k}\,(\mathbf{\Gamma}_j - g_{j,k} + g_j) + \sum_{k\in\mathcal{N}_0^-(j)} \mathbf{a}\cdot\mathbf{N}_{j,k}(\mathbf{\Gamma}_k - g_{j,k} + g_k) = 0$$

On indice $\mathbf{\Gamma}$ par un - si le triangle est vers le bas et par un + sinon, (1.3) se traduit de la façon suivante

(1) pour $\mathbf{\Gamma}_j^+$:

$$\sin(\frac{\pi}{3} + \theta)\mathbf{\Gamma}_j^+ - \sin(\theta)\mathbf{\Gamma}_1^- - \sin(\theta')\mathbf{\Gamma}_2^- = \vec{X}$$

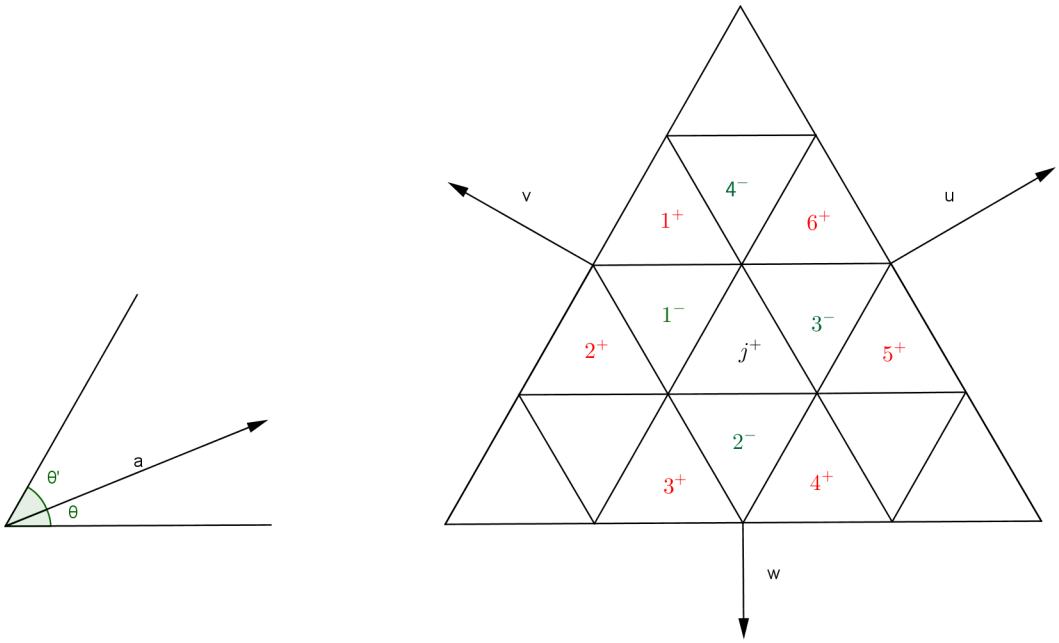

FIGURE 1.7 – Voisins du petit triangle j

(2) pour $\mathbf{\Gamma}_1^-$:

$$(\sin(\theta)+\sin(\theta'))\,\mathbf{\Gamma}_1^- = \sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_2^+ - \vec{X}$$

(3) pour $\mathbf{\Gamma}_2^-$:

$$(\sin(\theta)+\sin(\theta'))\mathbf{\Gamma}_2^- = \sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_3^+ - \vec{X}$$

où l'on a défini $\vec{X}$ ainsi

$$\vec{X} = \frac{\sqrt{3}}{6}\frac{l}{2^r}\left(\sin(\frac{\pi}{3}+\theta)\vec{u}+\sin(\theta')\vec{v}+\sin(\theta)\vec{w}\right).$$

On pose $S=\sin(\theta)+\sin(\theta')$, $p=\frac{\sin(\theta)}{S}$ et l'on injecte les valeurs de $\mathbf{\Gamma}_1^-$ et $\mathbf{\Gamma}_2^-$ obtenues grâce à (1) et (2) dans l'expression de $\mathbf{\Gamma}_j^+$. On remarque que les $\vec{X}$ se simplifient de sorte qu'il ne reste finalement que

$$\sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_j^+ = \frac{\sin(\theta)\sin(\frac{\pi}{3}+\theta)}{S}\mathbf{\Gamma}_2^+ + \frac{\sin(\theta')\sin(\frac{\pi}{3}+\theta)}{S}\mathbf{\Gamma}_3^+.$$

En simplifiant par $\sin(\theta+\frac{\pi}{3})$ on obtient

$$\mathbf{\Gamma}_j^+ = \frac{\sin(\theta)}{S}\mathbf{\Gamma}_2^+ + \frac{\sin(\theta')}{S}\mathbf{\Gamma}_3^+.$$

il vient enfin

$$\mathbf{\Gamma}_j^+ = p\mathbf{\Gamma}_2^+ + (1-p)\mathbf{\Gamma}_3^+. \tag{1.4}$$

Bord gauche. Pour le bord, on utilise là encore (1.3). On considère par exemple que 1^+ est un petit triangle sur le bord gauche du maillage :

(1) pour $\mathbf{\Gamma}_1^+$:

$$\sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_1^+ = \sin(\theta)\mathbf{\Gamma}_1^- + \frac{\frac{l}{2^r}\sqrt{3}}{6}(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta)\vec{w}) + \epsilon$$

(2) pour $\mathbf{\Gamma}_1^-$:

$$(\sin(\theta)+\sin(\theta'))\mathbf{\Gamma}_1^- = \sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_2^+ - \frac{\frac{l}{2^r}\sqrt{3}}{6}\left(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v} + \sin(\theta)\vec{w}\right)$$

On remplace $\mathbf{\Gamma}_1^-$ pour obtenir

$$\sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_1^+ = \frac{\sin(\frac{\pi}{3}+\theta)\sin(\theta)}{S}\mathbf{\Gamma}_2^+ + \frac{\frac{l}{2^r}\sqrt{3}}{6}\left(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta)\vec{w} - \frac{\sin(\theta)}{S}\left(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v} + \sin(\theta)\vec{w}\right)\right) + \epsilon.$$

Soit encore

$$\mathbf{\Gamma}_1^+ = \frac{\sin(\theta)}{S}\mathbf{\Gamma}_2^+ + \frac{\frac{l}{2^r}\sqrt{3}\sin(\theta')}{6S\sin(\frac{\pi}{3}+\theta)}(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta)\vec{w}) - \frac{\frac{l}{2^r}\sqrt{3}}{6S\sin(\frac{\pi}{3}+\theta)}\sin(\theta')\sin(\theta)\vec{v} + \frac{\epsilon}{\sin(\frac{\pi}{3}+\theta)}.$$

Et enfin

$$\mathbf{\Gamma}_1^+ = p\mathbf{\Gamma}_2^+ + (1-p)\frac{\frac{l}{2^r}\sqrt{3}}{6\sin(\frac{\pi}{3}+\theta)}\left(\sin(\frac{\pi}{3}+\theta)\vec{u} - \sin(\theta)\vec{v} + \sin(\theta)\vec{w}\right) + \frac{\epsilon}{\sin(\frac{\pi}{3}+\theta)}. \tag{1.5}$$

Bord bas. On s'occupe par exemple de 4^+ supposé être un petit triangle sur le bord bas du maillage :

(1) pour $\mathbf{\Gamma}_4^+$:

$$\sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_4^+ = \sin(\theta')\mathbf{\Gamma}_2^- + \frac{\frac{l}{2^r}\sqrt{3}}{6}(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v}) + \epsilon$$

(2) pour $\mathbf{\Gamma}_2^-$:

$$(\sin(\theta)+\sin(\theta'))\mathbf{\Gamma}_2^- = \sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_3^+ - \frac{\frac{l}{2^r}\sqrt{3}}{6}(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v} + \sin(\theta)\vec{w})$$

On remplace $\mathbf{\Gamma}_2^-$ pour obtenir

$$\sin(\frac{\pi}{3}+\theta)\mathbf{\Gamma}_4^+ = \frac{\sin(\frac{\pi}{3}+\theta)\sin(\theta')}{S}\mathbf{\Gamma}_3^+ + \frac{\frac{l}{2^r}\sqrt{3}}{6}\left(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v} - \frac{\sin(\theta')}{S}\left(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v} + \sin(\theta)\vec{w}\right)\right) + \epsilon.$$

Soit encore

$$\mathbf{\Gamma}_4^+ = \frac{\sin(\theta')}{S}\mathbf{\Gamma}_3^+ + \frac{\frac{l}{2^r}\sqrt{3}\sin(\theta)}{6S\sin(\frac{\pi}{3}+\theta)}(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v}) - \frac{\frac{l}{2^r}\sqrt{3}}{6S\sin(\frac{\pi}{3}+\theta)}\sin(\theta)\sin(\theta')\vec{w} + \frac{\epsilon}{\sin(\frac{\pi}{3}+\theta)}.$$

Et enfin

$$\mathbf{\Gamma}_4^+ = (1-p)\mathbf{\Gamma}_3^+ + p\frac{\frac{l}{2^r}\sqrt{3}}{6\sin(\frac{\pi}{3}+\theta)}\left(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v} - \sin(\theta')\vec{w}\right) + \frac{\epsilon}{\sin(\frac{\pi}{3}+\theta)}. \tag{1.6}$$

1.5.3 Résumé des calculs

On numérote les triangles par des couples (m, n) où m désigne l'abscisse du triangle dans le quadrillage et n son ordonnée comme le montre la figure 1.8. Les formules obtenues précédemment et en annexe se traduisent alors de la façon suivante.

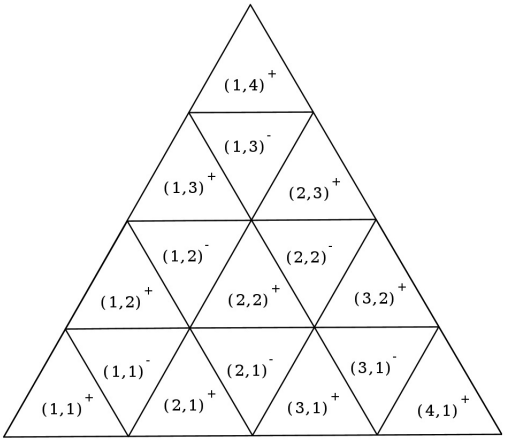

FIGURE 1.8 – Numérotation des petits triangles

$\vec{a}$ rentre par deux cotés

pour $m > 1$ et $n > 1$:

$$\mathbf{\Gamma}^+_{m,n} = p\mathbf{\Gamma}^+_{m-1,n} + (1-p)\mathbf{\Gamma}^+_{m,n-1} \qquad (cf(1.4))$$

pour $m = 1$ et $n > 1$:

$$\mathbf{\Gamma}^+_{1,n} = p\mathbf{\Gamma}^+_{1,n-1} + (1-p)\frac{\frac{l}{2^r}\sqrt{3}}{6\sin(\frac{\pi}{3}+\theta)}(\sin(\frac{\pi}{3}+\theta)\vec{u} - \sin(\theta)\vec{v} + \sin(\theta)\vec{w}) + \frac{\epsilon}{\sin(\frac{\pi}{3}+\theta)} \qquad (cf(1.5))$$

pour $m > 1$ et $n = 1$:

$$\mathbf{\Gamma}^+_{m,1} = (1-p)\mathbf{\Gamma}^+_{m-1,1} + p\frac{\frac{l}{2^r}\sqrt{3}}{6\sin(\frac{\pi}{3}+\theta)}(\sin(\frac{\pi}{3}+\theta)\vec{u} + \sin(\theta')\vec{v} - \sin(\theta')\vec{w}) + \frac{\epsilon}{\sin(\frac{\pi}{3}+\theta)} \qquad (cf(1.6))$$

pour $m = 1$ et $n = 1$:

$$\mathbf{\Gamma}^+_{1,1} = \frac{\frac{l}{2^r}\sqrt{3}}{6}\vec{u} + \epsilon$$

où p est celui défini précédemment à savoir

$$p = \frac{\sin(\theta)}{\sin(\theta)+\sin(\theta')}.$$

$\vec{a}$ rentre par un coté

pour $m > 1$ et $n > 1$:

$$\mathbf{\Gamma}^+_{m,n} = p\mathbf{\Gamma}^+_{m,n-1} + (1-p)\mathbf{\Gamma}^+_{m+1,n-1} \qquad (cf(3.5))$$

pour $m \geq 1$ et $n = 1$:

$$\mathbf{\Gamma}^+_{m,1} = \frac{\frac{l}{2^r}\sqrt{3}}{6}(p\vec{u} + (1-p)\vec{v}) + \frac{\epsilon}{\sin(\frac{\pi}{3}+\theta)+\sin(\theta')} \qquad (cf(3.6))$$

où p est celui défini en annexe à savoir

$$p = \frac{\sin(\frac{\pi}{3}+\theta)}{\sin(\theta')+\sin(\frac{\pi}{3}+\theta)}.$$

1.5.4 Résultat principal

Nous avons montré que $\mathbf{\Gamma}$ est un $O(\frac{l}{2^r})$ au bord et qu'à l'intérieur, il peut s'exprimer comme un barycentre de $\mathbf{\Gamma}$ qui sont sur les bords.
Il s'ensuit que $\mathbf{\Gamma}$ est un $O(\frac{l}{2^r})$ sur tout le domaine triangulaire. La démonstration présentée nécessite des calculs bien plus simples que celle présentée dans l'article [4]. Passer par un triangle équilatéral et utiliser une application linéaire simplifie bien les calculs.

1.6 Conclusion

Cette partie nous a permis de mieux appréhender les caractéristiques du correcteur géométrique que l'on utilisera dans les parties suivantes. Etablir la nilpotence de B (qui est d'ordre au plus le nombre de triangle) en dimension 2 permet d'avoir une expression exacte et finie du correcteur $\mathbf{\Gamma}$. Le cas de la dimension 3 reste compliqué et n'est pas analogue au précédent puisque le résultat de nilpotence n'est plus vrai. De nombreuses propriétés ne sont plus valables. Par ailleurs on a pu appliquer la formule du correcteur au cas d'un maillage triangulaire et simplifier la démonstration antérieure du papier [4] par une nouvelle approche (triangle équilatéral puis tranformation linéaire).

Chapitre 2

Etude de l'équation de transport avec terme source

2.1 Introduction

Dans une seconde partie de notre stage, nous nous sommes intéressés à une nouvelle équation, l'équation de transport en dimension en dimension 1 avec terme source. Cette équation présentée dans l'article [5] est la suivante

$$\begin{cases} \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} + b(u)z'(x) = 0 \\ u(0,x) = u_0(x) \end{cases} \tag{2.1}$$

où b et z sont supposées suffisamment régulières (au moins $\mathcal{C}^1$).
Il s'agit de l'équation de Saint-Venant pour un fond non constant, c'est la dérivée de la hauteur z' qui intervient. On sait discrétiser la partie sans le terme source de l'équation de manière à obtenir un schéma adéquat et ce quelque soit le signe de a que l'on choisira ici positif. Lorsque a est positif, on décentre en amont et en aval quand a est négatif. L'auteur de cet article applique la même discrétisation au terme z' qu'elle décentre en amont et présente ce schéma comme une discrétisation valable. Cependant, on ne voyait pas pourquoi il fallait discrétiser z' ainsi. On a alors montré que deux autres schémas ainsi que celui-ci étaient convergents. Ensuite lors de test numériques avec Matlab, on a vu que pour le même exemple que celui de l'article, aucun des trois schémas ne conservait la solution stationnaire. On a enfin trouvé un nouveau schéma convergent qui conserve au moins numériquement la solution stationnaire. On utilise pour la suite un maillage non uniforme, comme on peut le voir sur la figure 2.1.

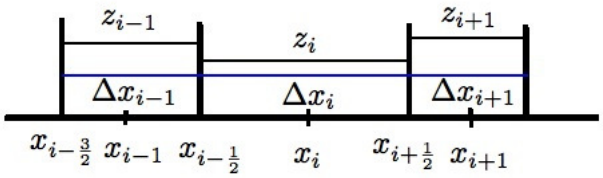

FIGURE 2.1 – Maillage non uniforme

2.2 Obtention de la consistance par la méthode d'un correcteur

On présente successivement les trois schémas et on montre leur consistance en exhibant un correcteur qui satisfait les conditions précédemment exposées. On en présente un dans cette partie et on laisse en annexe les deux autres.

2.2.1 Schéma avec z décentré à gauche

On étudie le schéma suivant proposé dans l'article [5]

$$S(v) = \frac{v_i^{n+1} - v_i^n}{\Delta t} + a\frac{v_i^n - v_{i-1}^n}{\Delta x_i} + b(v_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i}\,. \tag{2.2}$$

On suppose disposer de u solution de l'équation et l'on applique S au vecteur de composantes $u_i^n = u(x_i, n\cdot\Delta t)$. On définit

$$\epsilon_i^n = \frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{u_i^n - u_{i-1}^n}{\Delta x_i} + b(u_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i}.$$

On applique le schéma S à un vecteur w avec $w_i^n = u_i^n + \gamma_i^n$, où γ_i^n sera défini ultérieurement et est cherché sous la forme d'un $\mathcal{O}(\Delta x_i)$ (preuve par annalyse-synthèse), on suppose qu'un tel γ_i^n existe, on pose alors l'erreur de consistance corrigée

$$\delta_i^n = \frac{w_i^{n+1} - w_i^n}{\Delta t} + a\frac{w_i^n - w_{i-1}^n}{\Delta x_i} + b(w_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i}.$$

On utilise une formule de Taylor à l'ordre 1 sur b supposée C^1 et l'on obtient

$$\delta_i^n = \frac{w_i^{n+1} - w_i^n}{\Delta t} + a\frac{w_i^n - w_{i-1}^n}{\Delta x_i} + b(u_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i} + \gamma_i^n b'(u_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i} + \mathcal{O}(\Delta x_i),$$

qui se réécrit en utilisant une formule de Taylor sur z

$$\delta_i^n = \frac{w_i^{n+1} - w_i^n}{\Delta t} + a\frac{w_i^n - w_{i-1}^n}{\Delta x_i} + b(u_i^n)z'(x_i) + \gamma_i^n b'(u_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i} - b(u_i^n)\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}z'(x_i) + \mathcal{O}(\Delta x_i).$$

En développant comme précédemment, on obtient

$$\delta_i^n = \frac{\gamma_i^{n+1} - \gamma_i^n}{\Delta t} + a\frac{\gamma_i^n - \gamma_{i-1}^n}{\Delta x_i} + \gamma_i^n b'(u_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i} - a\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}\frac{\partial u}{\partial x}(x_i, t_n) - b(u_i^n)\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}z'(x_i)$$
$$+\mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i).$$

On définit γ_i^n par $\gamma_i^n = \frac{\Delta x_i}{2}\frac{\partial u}{\partial x}(x_i, t_n) + \frac{\Delta x_i}{2a}b(u_i^n)z'(x_i)$.
Une formule de Taylor donne alors
i) $\frac{\gamma_i^{n+1} - \gamma_i^n}{\Delta t} = \mathcal{O}(\Delta x_i)$
ii) $a\frac{\gamma_i^n - \gamma_{i-1}^n}{\Delta x_i} = a\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}\frac{\partial u}{\partial x}(x_i, t_n) + b(u_i^n)\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}z'(x_i) + \mathcal{O}(\Delta x_i)$
iii) $\gamma_i^n b'(u_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i} = \mathcal{O}(\Delta x_i)$.
Avec un tel vecteur γ, on a finalement

$$S(u + \gamma) = \mathcal{O}(\Delta x) + \mathcal{O}(\Delta t).$$

2.2.2 Conclusion

On dispose à ce stade de la consistance des trois schémas et l'on a remarqué qu'aucun des schémas ne semblait préférable à un autre. On veut désormais montrer la convergence des schémas. Nous a d'abord montré la convergence dans le cas où b' est borné, puis nous avons trouvé une démonstration du cas général.

2.3 Démonstration de la convergence pour le schéma de l'article dans le cas où b' est borné.

On montre ici que le schéma de l'article est convergent dans le cas où b' est borné, il est facile de voir que l'on peut le montrer pour les trois schémas précédemment proposés de la même manière.

Théorème 9 *Le schéma de l'article [5] avec z' décentré à gauche est convergent lorsque b' est borné.*

Preuve. On pose δ_i^n l'erreur de consistance corrigée par $\gamma_i^n = \frac{\Delta x_i}{2a}\frac{\partial u}{\partial t}(x_i, t^n)$ et définie par $\delta_i^n = S(u_i^n + \gamma_i^n)$ où $u_i^n = u(x_i, t^n)$ avec u solution exacte de l'équation (2.1). On a montré que $\delta_i^n = \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i)$. Soit (v_i^n) la solution discrète vérifiant $S(v_i^n) = 0$ et $v_i^0 = u(x_i, 0)$ alors

$$S(u_i^n + \gamma_i^n) - S(v_i^n) = \delta_i^n. \tag{2.3}$$

On note $e_i^n = u_i^n + \gamma_i^n - v_i^n$. D'après la formule de Taylor-Lagrange, b étant de classe $\mathcal{C}^1$ il existe $\theta_i^n \in]u_i^n + \gamma_i^n, v_i^n[$ tel que $b(u_i^n + \gamma_i^n) - b(v_i^n) = (u_i^n + \gamma_i^n - v_i^n)b'(\theta_i^n)$. Comme b' est borné on sait que $b'(\theta_i^n)$ est borné pour tout (i, n). La formule (2.3) s'écrit alors

$$\frac{e_i^{n+1} - e_i^n}{\Delta t} + a\frac{e_i^n - e_{i-1}^n}{\Delta x_i} + e_i^n b'(\theta_i^n)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i} = \delta_i^n.$$

En posant e^n le vecteur des e_i^n et δ^n le vecteur des δ_i^n, on obtient

$$e^{n+1} = (A + B_n)e^n + \Delta t\delta^n, \tag{2.4}$$

avec A une matrice triangulaire inférieure d'éléments non nuls $a_{ii} = 1 - a\frac{\Delta t}{\Delta x_i}$ et $a_{ii-1} = a\frac{\Delta t}{\Delta x_i}$ et B_n la matrice diagonale telle que

$$b_{ii} = -b'(\theta_i^n)\frac{\Delta t}{\Delta x_i}(z(x_i) - z(x_{i-1})) = \mathcal{O}(\Delta t).$$

En admettant la CFL $0 < a\frac{\Delta t}{\Delta x_i} \leq 1$, on montre la convergence dans les normes suivantes.
En norme infinie, on montre que

$$|||A + B_n||| \leq |||A||| + |||B_n||| \leq 1 + M\Delta t.$$

où $M = \sup_{(i,n)} |b'(\theta_i^n)|(\sup_{x\in[\alpha,\beta]} |z'(x)|\frac{\alpha+1}{2} + J)$ et J est le nombre de mailles et $\Delta x_{i+1} \leq \alpha\Delta x_i$ par quasi-uniformité du maillage.
D'après (2.4), on a

$$||e^{n+1}|| \leq (1 + M\Delta t)||e^n|| + \Delta t||\delta^n||,$$

en utilisant le lemme de Gronwall discret,

$$||e^n|| \leq ||e^0||e^{MT} + \Delta t e^{MT}\sum_{k=0}^{n-1} ||\delta^k||.$$

On en déduit que

$$||e^n||_\infty = \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i).$$

En norme $p \geq 1$, on a que $|e_i^n| \leq ||e^n||_\infty$,

d'où

$$(\sum_{i=1}^{J} \Delta x_i |e_i^n|^p)^{\frac{1}{p}} \leq (\sum_{i=1}^{J} \Delta x_i ||e^n||_\infty^p)^{\frac{1}{p}},$$

$$||e^n||_p \leq ||e^n||_\infty (\sum_{i=1}^{J} \Delta x_i)^{\frac{1}{p}}.$$

Or $(\sum_{i=1}^{J} \Delta x_i)^{\frac{1}{p}}$ est une constante et $||e^n||_\infty = \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i)$, donc finalement

$$||e^n||_p = \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i).$$

■

2.4 Démonstration de la convergence pour le schéma de l'article dans le cas général

Montrons maintenant le cas général.

Théorème 10 *L'erreur de convergence corrigée e_i^n est un $\mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i)$.*

Preuve. On montre par récurrence sur n que l'on a

$$||e^n||_\infty \leq (1 + \frac{A}{N})^n \beta_0 = (1+\alpha)^n \beta_0$$

avec $A = \frac{1+p}{2}||z'||_\infty ||b'||_\infty T$. Où l'on a utilisé les notations suivantes :
* $\beta_0 = ||e^0||_\infty$
* T le temps maximal séparé en N intervalles de longueur Δt
* $\frac{\Delta x_{i-1}}{\Delta x_i} \leq p$ (quasi-uniformité du maillage)
* $||u||_\infty = \sup\{|u(x,t)|, x \in [a;b], t \in [0;T]\}$
* $||b'||_\infty = \sup\{|b'(x)|, x \in [-||u||_\infty - 1 - \exp(A)\beta_0; ||u||_\infty + 1 + \exp(A)\beta_0]\}$
* $||z'||_\infty = \sup\{|z'(x)|, x \in [a;b]\}$

Pour $n = 0$, ça marche par définition de β_0.
Pour l'hérédité, on a la formule suivante due au schéma utilisé (2.2)

$$e_i^{n+1} = e_i^n(1 - \frac{a\Delta t}{\Delta x_i}) + a e_{i-1}^n \frac{\Delta t}{\Delta x_i} + [b(u_i^n + \gamma_i^n) - b(v_i^n)] \frac{z(x_i) - z(x_{i-1})}{\Delta x_i} \Delta t.$$

On passe alors à la norme infinie et l'on obtient avec Taylor la formule suivante

$$||e^{n+1}||_\infty \leq ||e^n||_\infty (1 + \frac{1+p}{2}||b'(\theta^n)||_\infty ||z'||_\infty \frac{T}{N})$$

Il faut alors remarquer que $\theta_i^n \in [u_i^n + \gamma_i^n; v_i^n]$ ou $[v_i^n; u_i^n + \gamma_i^n]$.
Soit avec l'hypothèse de récurrence $\theta_i^n \in [u_i^n + \gamma_i^n; u_i^n + \gamma_i^n + (1+\alpha)^n\beta_0]$ ou $[u_i^n + \gamma_i^n - (1+\alpha)^n\beta_0; u_i^n + \gamma_i^n]$.
Comme $\gamma_i^n = \mathcal{O}(\Delta x_i)$ on peut supposer $||\gamma_i^n|| \leq 1$.
On a alors $\theta_i^n \in [-||u||_\infty - 1 - \exp(A)\beta_0; ||u||_\infty + 1 + \exp(A)\beta_0]$.
On obtient alors le résultat escompté à savoir

$$||e^{n+1}||_\infty \leq (1+\alpha)||e^n||_\infty = (1+\alpha)^{n+1}\beta_0.$$

On peut donc affirmer que l'on a $||e^n||_\infty \leq \exp(A)\beta_0$ pour tout n. L'erreur de convergence corrigée (avec le correcteur) est donc bien un $\mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i)$ en norme infinie. En norme $p \geq 1$, on montre comme pour le cas où b' était borné que

$$||e^n||_p = \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i).$$

■

Remarque : En remplaçant la discrétisation du terme source $b(v_i^n)\frac{z(x_i)-z(x_{i-1})}{\Delta x_i}$ par $b(v_i^n)\frac{z(x_{i+1})-z(x_i)}{\Delta x_i}$ ou par $b(v_i^n)z'(x_i)$, la démonstration reste la même.

2.5 Exemples numériques

On a vu dans la section précédente que les trois schémas étaient convergents. On s'en convainc maintenant avec des exemples numériques.

2.5.1 Convergence

On prend $b(x) = x^2$, $z(x) = x$, $a = 1$ sur l'intervalle $[0,1]$. La condition initiale est $u(x,0) = \frac{1}{x+1}$. On divise $[0,1]$ de façon non uniforme. La solution exacte est $\frac{2}{x+t+1}$ pour une condition au bord $u(0,t) = \frac{2}{a+t}$ et l'on voit numériquement que lorsque le pas diminue, la solution discrète obtenue par le schéma de l'article [5] se rapproche de la solution exacte comme on peut le voir sur les figures 2.2. La solution exacte est en bleu à $T = 1.5$ et la

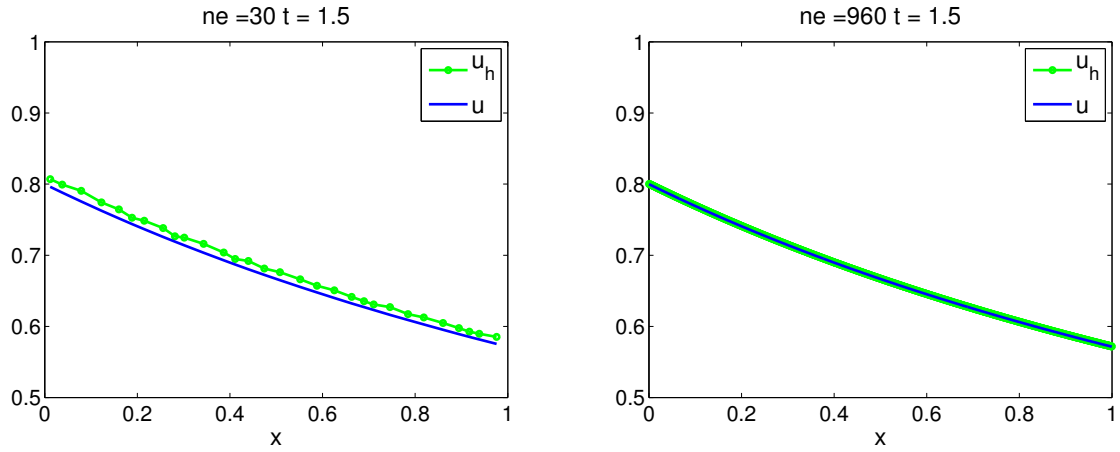

FIGURE 2.2 – 30 pas d'espace (gauche) et 960 pas d'espace (droite)

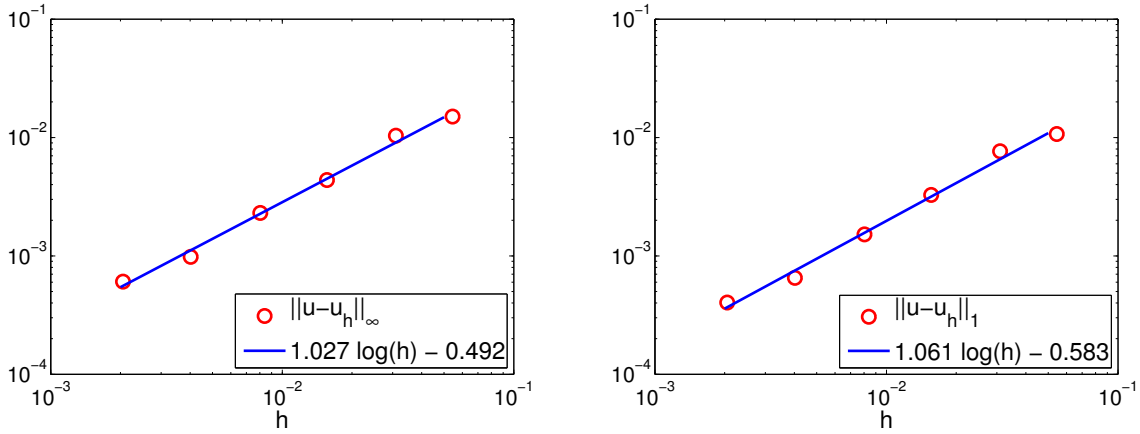

FIGURE 2.3 – Norme ∞ (gauche) et Norme 1 (droite)

solution approchée est en vert.

On voit que les normes infini et 1 de l'erreur se comportent comme prévu en $\mathcal{O}(h)$ sur les figures 2.3. La solution exacte est $\frac{2}{x+t+1}$ pour une condition au bord : $u(0,t) = \frac{2}{a+t}$ et l'on voit numériquement que lorsque le pas diminue, la solution discrète obtenue par le schéma de l'article [5] se rapproche de la solution exacte.

2.5.2 Non conservation de la solution stationnaire

L'exemple de l'article [5] est $z(x) = sin(\pi x)$, $b(x) = x$, $a = 1$ sur l'intervalle $[0, 1]$. Pour la condition initiale est $u(x, 0) = u_0(x)e^{-\frac{1}{a}sin(\pi x)}$ on sait calculer la solution stationnaire qui est

$$u_0(x - at)e^{\frac{1}{a}(-sin(\pi x)+sin(\pi(x-at)))} = e^{-\frac{1}{a}sin(\pi x)} = u_0(x).$$

Comme la solution est périodique sur $[0, 1]$, les conditions au bord sont périodiques et imposées. On applique alors le schéma de l'article [5] à la solution stationnaire exacte et l'on observe que lorsque T augmente, la solution s'éloigne, sur les figures 2.4. Le schéma ne conserve donc pas la solution stationnaire.

2.5.3 Obtention d'un schéma conservant la solution stationnaire

On s'est demandé comment trouver un schéma comblant le défaut observé sur les trois précédents. On a eu une démarche descendante, sachant ce que le schéma devait vérifier, on en a trouvé un. La partie numérique semble

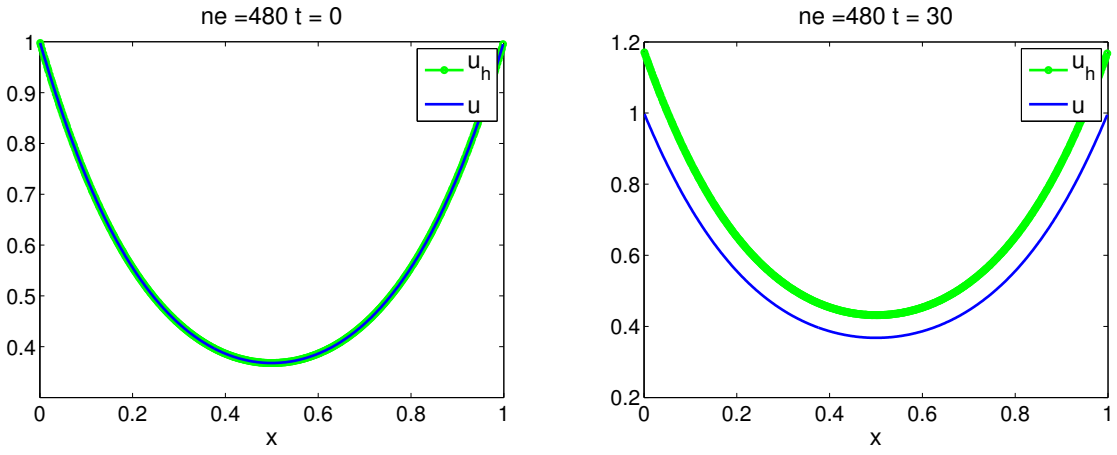

FIGURE 2.4 – T = 0 (gauche) et T = 30 (droite)

confirmer sa validité mais, après avoir montré sa convergence, on n'a pas su montrer qu'il conservait la solution stationnaire. On veut que le schéma discrétise l'équation considérée lorsque $\frac{\partial u}{\partial t} = 0$. On l'écrit sous la forme

$$\frac{v_i^{n+1} - v_i^n}{\Delta t} + a\frac{v_i^n - v_{i-1}^n}{\Delta x_i} + \Sigma_i^n = 0.$$

On a en intégrant lorsque $\frac{\partial u}{\partial t} = 0$,

$$a\int_{x_{i-1}}^{x_i} \frac{\partial u^\infty(x,t)}{\partial x}dx + \int_{x_{i-1}}^{x_i} b(u^\infty)z'(x)dx = 0$$

$$a\int_{x_{i-1}}^{x_i} \frac{du^\infty(x,t)}{dx}dx + \int_{x_{i-1}}^{x_i} b(u^\infty)z'(x)dx = 0$$

$$a\frac{u_i^\infty - u_{i-1}^\infty}{\Delta x_i} + \frac{1}{\Delta x_i}\int_{x_{i-1}}^{x_i} b(u^\infty)z'(x)dx = 0$$

$$a\frac{u_i^\infty - u_{i-1}^\infty}{\Delta x_i} + \frac{1}{\Delta x_i}\left(\int_{x_{i-1}}^{x_{i-\frac{1}{2}}} b(u^\infty)z'(x)dx + \int_{x_{i-\frac{1}{2}}}^{x_i} b(u^\infty)z'(x)dx\right) = 0$$

On ne connait z' qu'en $x_{i-\frac{1}{2}}$ et il vaut $2\frac{z(x_i)-z(x_{i-1})}{\Delta x_i + \Delta x_{i-1}}$ et u^∞ est connu sur $[x_{i-1}, x_{i-\frac{1}{2}}]$ comme valant u_{i-1}^∞ et sur $[x_{i-\frac{1}{2}}, x_i]$ comme valant u_i^∞. On ne peut donc approcher la formule précédente que par

$$a\frac{u_i^\infty - u_{i-1}^\infty}{\Delta x_i} + \frac{1}{\Delta x_i}\left(\Delta x_{i-1}b(u_{i-1}^\infty) + \Delta x_i b(u_i^\infty)\right)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i + \Delta x_{i-1}}.$$

Le schéma est donc le suivant

$$S(v) = \frac{v_i^{n+1} - v_i^n}{\Delta t} + a\frac{v_i - v_{i-1}}{\Delta x_i} + \frac{1}{\Delta x_i}\left(\Delta x_{i-1}b(v_{i-1}) + \Delta x_i b(v_i)\right)\frac{z(x_i) - z(x_{i-1})}{\Delta x_i + \Delta x_{i-1}}. \tag{2.5}$$

2.5.4 Vérifications numériques

On vérifie la convergence avec les mêmes conditions que pour les trois schémas précédents, sur les figures 2.5.

On voit numériquement qu'il conserve la solution stationnaire de l'article [5] contrairement au schéma de ce même article, sur les figures 2.6.

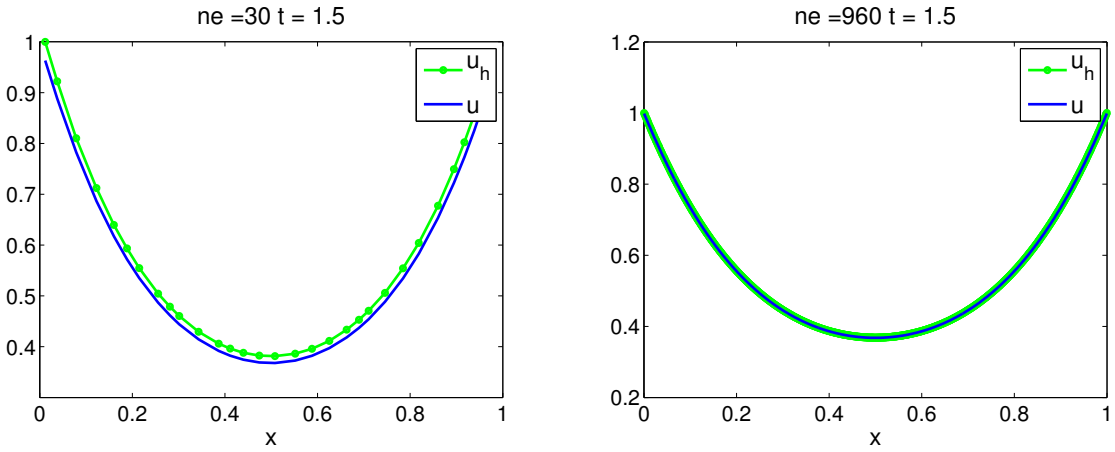

FIGURE 2.5 – 30 pas d'espace (gauche) et 960 pas d'espace (droite)

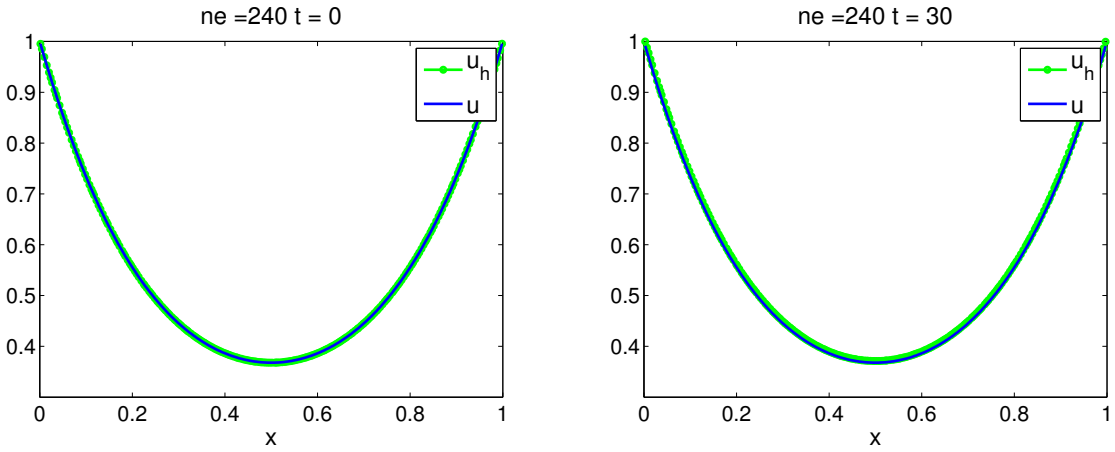

FIGURE 2.6 – T = 0 (gauche) et T = 30 (droite)

2.5.5 Eléments de démonstration de la convergence du nouveau schéma

On suit la même méthode que pour les autres schémas.
– Le nouveau correcteur est $\gamma_i^n = \frac{\Delta x_{i-1}}{2} \frac{\partial u}{\partial x})_i^n + \frac{\Delta x_{i-1}}{2a} b(u_i^n) z'(x_i)$.
– On montre que l'erreur e_i^n est bornée.
– On montre que l'erreur se comporte comme un $\mathcal{O}(h)$.

2.6 Conclusion

Dans cette partie, on a vu à l'oeuvre numériquement différents schémas et on les a comparés. On a ainsi pu vérifier la convergence des schémas remarquer leurs propriétés respectives. Notamment, nous avons vu que le schéma de l'article [5] ne conservait pas la solution stationnaire pour des conditions aux bords périodiques malgré la convergence du schéma. C'est ainsi que l'absence de conservation de la solution stationnaire des schémas initiaux nous a amenés à en trouver un nouveau. On s'est ainsi familiarisé avec l'analyse numérique sur Matlab et on en a constaté l'utilité. On a pu se poser des questions sur un article de recherche [5], en voir les limites et se demander comment aller au delà des résultats qui y sont présentés. Ainsi, on a démontré la convergence du schéma de l'article, on a vu qu'il ne conservait pas la solution stationnaire et l'on en a construit un qui la conservait tout en restant convergent.

Chapitre 3

Equation des ondes en dimension 2

3.1 Introduction

Dans cette partie, on a appliqué un schéma de type volume fini pour obtenir une discrétisation convergente de l'équation des ondes en dimension 2. On commence par discrétiser l'équation des ondes en dimension 1 et l'on démontre la convergence du schéma obtenu. Pour la dimension 2, on procède par analogie avec l'article [3] pour obtenir l'équation du correcteur $\mathbf{\Gamma}$. Cette équation est bien plus compliquée qu'en dimension 1 donc on l'a implémentée en Matlab pour pouvoir l'étudier plus facilement.

3.2 Ondes en dimension 1

On s'intéresse au système suivant

$$\begin{cases} \frac{\partial h}{\partial t} + \frac{\partial u}{\partial x} &= 0 \\ \frac{\partial u}{\partial t} + c^2 \frac{\partial h}{\partial x} &= 0 \end{cases}.$$

On peut montrer facilement que h est la fonction qui vérifie l'équation des ondes. On pose

$$A = \begin{pmatrix} 0 & 1 \\ c^2 & 0 \end{pmatrix}, \quad U = \begin{pmatrix} h \\ u \end{pmatrix}, \quad P = \begin{pmatrix} 1 & -1 \\ c & c \end{pmatrix} \quad et \quad \Delta = \begin{pmatrix} c & 0 \\ 0 & -c \end{pmatrix}.$$

On peut écrire le système sous forme matricielle comme

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} = 0.$$

La matrice A est diagonalisable et l'on a ainsi $A = P\Delta P^{-1}$. En posant alors $\tilde{U} = P^{-1}U$ il vient succesivement

$$\frac{\partial U}{\partial t} + P\Delta P^{-1}\frac{\partial U}{\partial x} = 0,$$

puis

$$P^{-1}\frac{\partial U}{\partial t} + \Delta P^{-1}\frac{\partial U}{\partial x} = 0,$$

et enfin

$$\frac{\partial \tilde{U}}{\partial t} + \Delta\frac{\partial \tilde{U}}{\partial x} = 0.$$

Si l'on pose $\tilde{U} = \begin{pmatrix} \tilde{h} \\ \tilde{u} \end{pmatrix}$, on trouve le système suivant

$$\begin{cases} \frac{\partial \tilde{h}}{\partial t} + c\frac{\partial \tilde{h}}{\partial x} &= 0 \\ \frac{\partial \tilde{u}}{\partial t} - c\frac{\partial \tilde{u}}{\partial x} &= 0 \end{cases}.$$

Les équations étant indépendantes, une discrétisation décentrée à gauche pour $\tilde{h}$ et à droite pour $\tilde{u}$ donne le système convergent

$$\left\{\begin{array}{rcl} \frac{\tilde{h}_i^{n+1}-\tilde{h}_i^n}{\Delta t}+c\frac{\tilde{h}_i^n-\tilde{h}_{i-1}^n}{\Delta x_i} & = & 0 \\ \\ \frac{\tilde{u}_i^{n+1}-\tilde{u}_i^n}{\Delta t}-c\frac{\tilde{u}_{i+1}^n-\tilde{u}_i^n}{\Delta x_i} & = & 0 \end{array}\right. .$$

On sait que $U_i^n = P\tilde{U}_i^n$ et cela permet d'obtenir

$$\begin{array}{rcl} \begin{pmatrix} \tilde{h}_i^n \\ \tilde{u}_i^n \end{pmatrix} & = & \frac{1}{2c}\begin{pmatrix} c & 1 \\ -c & 1 \end{pmatrix}\begin{pmatrix} h_i^n \\ u_i^n \end{pmatrix} \\ \\ & = & \frac{1}{2c}\begin{pmatrix} ch_i^n + u_i^n \\ -ch_i^n + u_i^n \end{pmatrix} \end{array}$$

En sommant et en retranchant on obtient les deux équations suivantes

$$\left\{\begin{array}{rcl} 2\frac{h_i^{n+1}-h_i^n}{\Delta t}+\frac{2ch_i^n-ch_{i-1}^n-ch_{i+1}^n+u_{i+1}^n-u_{i-1}^n}{\Delta x_i} & = & 0 \\ \\ 2\frac{u_i^{n+1}-u_i^n}{\Delta t}+\frac{2u_i^n-u_{i-1}^n-u_{i+1}^n+ch_{i+1}^n-ch_{i-1}^n}{\Delta x_i} & = & 0 \end{array}\right.$$

On a montré dans la partie 2, théorème (9), (cas où le terme source est nul) que l'on a une telle erreur de convergence

$$||\tilde{\mathcal{U}}^n - \tilde{\mathcal{V}}^n|| = \mathcal{O}(h),$$

avec

$$\tilde{\mathcal{U}}^n = \begin{pmatrix} \tilde{h}(x_1,t^n) \\ \tilde{u}(x_1,t^n) \\ \tilde{h}(x_2,t^n) \\ \tilde{u}(x_2,t^n) \\ \vdots \\ \tilde{h}(x_d,t^n) \\ \tilde{u}(x_d,t^n) \end{pmatrix} \quad et \quad \tilde{\mathcal{V}}^n = \begin{pmatrix} \tilde{h}_1^n \\ \tilde{u}_1^n \\ \tilde{h}_2^n \\ \tilde{u}_2^n \\ \vdots \\ \tilde{h}_d^n \\ \tilde{u}_d^n \end{pmatrix}.$$

On définit de la même façon $\mathcal{U}^n$ et $\mathcal{V}^n$. En posant $Q = diag(P)$, on vérifie facilement que $\mathcal{U}^n = Q\,\tilde{\mathcal{U}}^n$ et $\mathcal{V}^n = Q\,\tilde{\mathcal{V}}^n$. D'où l'on déduit

$$||\mathcal{U}^n - \mathcal{V}^n|| \leq |||Q||| \; ||\tilde{\mathcal{U}}^n - \tilde{\mathcal{V}}^n||.$$

Soit enfin

$$||\mathcal{U}^n - \mathcal{V}^n|| = \mathcal{O}(h).$$

Le correcteur trouvé dans la partie précédente pour la discrétisation du système associé à $\tilde{U}$ est

$$\tilde{\mathbf{\Gamma}} = \begin{pmatrix} \mathbf{\Gamma}_1 \\ \\ \mathbf{\Gamma}_2 \end{pmatrix} \quad = \quad \begin{pmatrix} \frac{\Delta x_i}{2}\frac{\partial \tilde{h}}{\partial x}(x_i,t_n) \\ \\ -\frac{\Delta x_i}{2}\frac{\partial \tilde{u}}{\partial x}(x_i,t_n) \end{pmatrix}.$$

En notant $\mathcal{L}$ l'opérateur linéaire discrétisé associé au schéma de $\tilde{U}$ et $\mathcal{S}$ celui du schéma de U, on remarque que l'on a

$$\mathcal{S}(\xi) = \mathcal{L}(P^{-1}(\xi)).$$

On sait que $\mathcal{L}(\tilde{U} + \tilde{\mathbf{\Gamma}}) = \mathcal{O}(\Delta x)$ et on cherche un $\mathbf{\Gamma} = \mathcal{O}(\Delta x)$ tel que $\mathcal{S}(U + \mathbf{\Gamma}) = \mathcal{O}(\Delta x)$ soit

$$\mathcal{S}(U + \mathbf{\Gamma}) = \mathcal{L}(P^{-1}(U + \Gamma)) = \mathcal{L}(\tilde{U} + P^{-1}\Gamma) = \mathcal{O}(\Delta x).$$

On constate que $\Gamma = P\tilde{\Gamma}$ convient.

$$\Gamma = \frac{\Delta x_i}{2}\begin{pmatrix} \frac{\partial \tilde{h}}{\partial x}(x_i,t_n) + \frac{\partial \tilde{u}}{\partial x}(x_i,t_n) \\ \\ c\frac{\partial \tilde{h}}{\partial x}(x_i,t_n) - c\frac{\partial \tilde{u}}{\partial x}(x_i,t_n) \end{pmatrix} \quad = \quad \frac{\Delta x_i}{2}\begin{pmatrix} \frac{\frac{\partial u}{\partial x}(x_i,t_n)}{c} \\ \\ c\frac{\partial h}{\partial x}(x_i,t_n) \end{pmatrix}$$

$$= \begin{pmatrix} 0 & \frac{1}{c} \\ c & 0 \end{pmatrix} \quad \frac{\Delta x_i}{2} \frac{\partial \tilde{U}}{\partial x}(x_i, t_n) \quad = \quad \frac{A}{c} \frac{\Delta x_i}{2} \frac{\partial \tilde{U}}{\partial x}(x_i, t_n)$$

Finalement, vue la diagonalisation de A, on remarque que le correcteur peut s'écrire de la façon suivante $\Gamma = sign(A)\frac{\Delta x_i}{2}\frac{\partial \tilde{U}}{\partial x}(x_i, t_n)$.
En dimension 1, on se ramène à un système découplé et l'on sait exprimer le correcteur de chaque partie puis le correcteur global. En dimension 2, on va voir que l'on ne sait pas exprimer le correcteur et que l'on sait seulement écrire l'équation qu'il doit vérifier.

3.3 Obtention d'un schéma discrétisant l'équation des ondes en dimension 2

On s'intéresse à l'équation des ondes en dimension 2

$$\frac{\partial^2 H}{\partial t^2} - c^2 \left(\frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial y^2} \right) = 0.$$

On est alors amené à résoudre le système suivant

$$\begin{cases} \frac{\partial h}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0 \\ \frac{\partial u}{\partial t} + c^2 \frac{\partial h}{\partial x} &= 0 \\ \frac{\partial v}{\partial t} + c^2 \frac{\partial h}{\partial y} &= 0 \end{cases}$$

On pose $U = (h, u, v)^T$ et on a alors

$$\frac{\partial U}{\partial t} \quad + \quad \begin{pmatrix} 0 & 1 & 0 \\ c^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{\partial U}{\partial x} \quad + \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ c^2 & 0 & 0 \end{pmatrix} \frac{\partial U}{\partial y} = 0.$$

On définit $F(U) = (F_1(U), F_2(U)) = (A_1 U, A_2 U)$ pour avoir l'expression suivante du flux

$$\Phi(U_K^n, U_L^n, n_{K,L}) = |K \cap L|(\frac{F(U_K^n) + F(U_L^n)}{2} \cdot n_{K,L} - sign(F)\frac{F(U_L^n) - F(U_K^n)}{2} \cdot n_{K,L}).$$

On remarque que l'on a $F(U) \cdot n_{K,L} = A_{K,L} U$ où $A_{K,L} = \begin{pmatrix} 0 & n_1 & n_2 \\ c^2 n_1 & 0 & 0 \\ c^2 n_2 & 0 & 0 \end{pmatrix}$, où n_1 et n_2 sont les composantes de $n_{K,L}$ et dépendent donc de K et L.
On diagonalise alors $A_{K,L}$ par $A_{K,L} = P_{K,L} \Delta P_{K,L}^{-1}$ avec

$$P_{K,L} = \begin{pmatrix} -\frac{1}{c} & 0 & \frac{1}{c} \\ n_1 & n_2 & n_1 \\ n_2 & -n_1 & n_2 \end{pmatrix}, \quad P_{K,L}^{-1} = \begin{pmatrix} -\frac{c}{2} & \frac{n_1}{2} & \frac{n_2}{2} \\ 0 & n_2 & -n_1 \\ \frac{c}{2} & \frac{n_1}{2} & \frac{n_2}{2} \end{pmatrix} \quad et \quad \Delta_{K,L} = \Delta = \begin{pmatrix} -c & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c \end{pmatrix}$$

Remarque. Ici, contrairement à la dimension 1, les bases de diagonalisation dépendent du vecteur n.

On peut donc maintenant écrire $\Phi_{K,L} = \frac{|K \cap L|}{2}(A_{K,L}(U_K^n + U_L^n) - |A_{K,L}|(U_L^n - U_K^n))$ et l'on définit

$$\Phi_K = \sum_{L \in N(K)} \Phi_{K,L}.$$

En appliquant un schéma de type volumes finis, on obtient alors

$$\frac{U_K^{n+1} - U_K^n}{\Delta t} + \Phi_K = 0.$$

On remarque que $sign(A_{K,L}) = \frac{A_{K,L}}{c}$ et on peut alors simplifier le schéma

$$\frac{U_K^{n+1} - U_K^n}{\Delta t} + \sum_{L \in N(K)} \frac{|K \cap L|}{2}(A_{K,L}(U_K^n + U_L^n) - \frac{A_{K,L}^2}{c}(U_L^n - U_K^n)) = 0.$$

$$\frac{U_K^{n+1} - U_K^n}{\Delta t} + \sum_{L \in N(K)} |K \cap L|(\frac{cA_{K,L} + A_{K,L}^2}{2c} U_K^n + \frac{cA_{K,L} - A_{K,L}^2}{2c} U_L^n) = 0.$$

On calcule

$$A_{K,L}^2 = \begin{pmatrix} c^2 & 0 & 0 \\ 0 & c^2 n_1^2 & c^2 n_1 n_2 \\ 0 & c^2 n_1 n_2 & c^2 n_2^2 \end{pmatrix}.$$

Le schéma est donc le suivant

$$\frac{U_K^{n+1} - U_K^n}{\Delta t} + \sum_{L \in N(K)} |K \cap L|(A_{K,L}^+ U_K^n + A_{K,L}^- U_L^n) = 0. \tag{3.1}$$

Avec

$$A_{K,L}^+ = \frac{1}{2}\begin{pmatrix} c & n_1 & n_2 \\ c^2 n_1 & c n_1^2 & c n_1 n_2 \\ c^2 n_2 & c n_1 n_2 & c n_2^2 \end{pmatrix} \quad et \quad A_{K,L}^- = \frac{1}{2}\begin{pmatrix} -c & n_1 & n_2 \\ c^2 n_1 & -c n_1^2 & -c n_1 n_2 \\ c^2 n_2 & -c n_1 n_2 & -c n_2^2 \end{pmatrix}.$$

Tous les calculs faits jusqu'à présent supposaient que le volume K ne touchait pas le bord. On calcule maintenant le flux $\Phi_{K,L}$ lorsque K touche $\partial\Omega$, on note alors L son voisin fictif miroir selon la face qui touche $\partial\Omega$. Il vient successivement les égalités suivantes

$$\begin{aligned}
\Phi_{K,L} &= A_{K,L}\frac{U_K + U_L}{2} - |A_{K,L}|\frac{U_L - U_K}{2} \\
&= P_{K,L}\Delta P_{K,L}^{-1}\frac{U_K + U_L}{2} - P_{K,L}|\Delta|P_{K,L}^{-1}\frac{U_L - U_K}{2} \\
&= P\left[\Delta\frac{\tilde{U}_K + \tilde{U}_L}{2} - |\Delta|\frac{\tilde{U}_L - \tilde{U}_K}{2}\right] \\
&= P\left[\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c \end{pmatrix}\tilde{U}_K \quad - \quad \begin{pmatrix} -c & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\tilde{U}_L\right] \\
&= P\left[c\tilde{U}_K \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad - \quad c\tilde{U}_L \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right] \\
&= c\tilde{U}_K \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} r_3 \quad - c\tilde{U}_L \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} r_1 \\
&= A_{K,L}\left(\tilde{U}_K \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} r_3 \quad + \quad \tilde{U}_L \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} r_1\right)
\end{aligned}$$

En prenant $\tilde{U}_L = \begin{pmatrix} \tilde{h} \\ \tilde{u}_K \\ \tilde{v}_K \end{pmatrix}$ et en supposant $\tilde{h}_L$ donné, on peut réécrire ce qui précède

$$\Phi_{K,L} = A_{K,L}\left[\tilde{U}_L \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} r_1 \quad + \quad \tilde{U}_L \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} r_3\right].$$

Ainsi, la formule qui définit le schéma (3.1) reste correcte en prenant, lorsque L est fictif, la valeur du flux $\Phi_{K,L}$ écrite ci-dessus. Cela suppose de disposer de $\tilde{h}$ sur le bord du domaine.
Le but est maintenant de démontrer la convergence du schéma obtenu et donc, pour commencer, la consistance. On suit le même raisonnement que dans [3], en séparant l'erreur de consistance en une partie consistante et une à corriger.

3.4 Obtention de l'équation du correcteur

On suppose désormais que U est la solution exacte de l'équation initiale. On définit l'erreur de consistance

$$E_K^n = \frac{U_K^{n+1} - U_K^n}{\Delta t} + \sum_{L \in N(K)} |K \cap L|(\frac{cA_{K,L} + A_{K,L}^2}{2c} U_K^n + \frac{cA_{K,L} - A_{K,L}^2}{2c} U_L^n).$$

On pose $A_{K,L}^+ = \frac{cA_{K,L}+A_{K,L}^2}{2c}$ et $A_{K,L}^- = \frac{cA_{K,L}-A_{K,L}^2}{2c}$. On écrit $E_K^n = G_K^n + I_K^n$ avec
* $G_K^n = \frac{U_K^{n+1}-U_K^n}{\Delta t} + \frac{1}{|K|}\sum_{L\in N(K)} |K \cap L| A_{K,L} U_{K,L}^n$
* $I_K^n = \frac{1}{|K|}\sum_{L\in N(K)} |K \cap L|(A_{K,L}^+(U_K^n - U_{K,L}^n) + A_{K,L}^-(U_L^n - U_{K,L}^n))$.

3.4.1 G_K^n est consistant

On montre que G_K^n est consistant. On sait que

$$\sum_{L \in N(K)} |K \cap L| A_{K,L} = 0.$$

Ainsi on a successivement

$$\begin{aligned}
\sum_{L\in N(K)} |K \cap L| A_{K,L} U_{K,L}^n &= \sum_{L\in N(K)} |K \cap L| A_{K,L}(U_{K,L}^n - U_K^n), \\
&= \sum_{L\in N(K)} A_{K,L}((G_{K,L} - G_K) \odot \nabla U_K^n) + \mathcal{O}(h^3), \\
&= \sum_{L\in N(K)} |K \cap L| A_{K,L}(G_{K,L} \odot \nabla U_K^n) + \mathcal{O}(h^3). \qquad (3.2)
\end{aligned}$$

Où l'on a posé
* $G_{K,L} = \begin{pmatrix} g_{K,L}^x g_{K,L}^y \\ g_{K,L}^x g_{K,L}^y \\ g_{K,L}^x g_{K,L}^y \end{pmatrix}$
* $g_{K,L}$ le centre de gravité de $K \cap L$
* $G_K = \begin{pmatrix} g_K^x g_K^y \\ g_K^x g_K^y \\ g_K^x g_K^y \end{pmatrix}$
* g_K le centre de gravité de K
*

$$\begin{aligned}
G_{K,L} \odot \nabla U_K^n &= \begin{pmatrix} g_{K,L} \cdot \nabla h_K^n \\ g_{K,L} \cdot \nabla u_K^n \\ g_{K,L} \cdot \nabla v_K^n \end{pmatrix} \\
&= \begin{pmatrix} g_{K,L}|_x \frac{\partial h}{\partial x})_K^n \\ g_{K,L}|_x \frac{\partial u}{\partial x})_K^n \\ g_{K,L}|_x \frac{\partial v}{\partial x})_K^n \end{pmatrix} + \begin{pmatrix} g_{K,L}|_y \frac{\partial h}{\partial y})_K^n \\ g_{K,L}|_y \frac{\partial u}{\partial y})_K^n \\ g_{K,L}|_y \frac{\partial v}{\partial y})_K^n \end{pmatrix} \\
&= g_{K,L}|_x \frac{\partial U}{\partial x})_K^n + g_{K,L}|_y \frac{\partial U}{\partial y})_K^n.
\end{aligned}$$

On montre que l'on a les deux formules suivantes à partir de la formule de la divergence

$$\sum_{L\in N(K)} |K \cap L| A_{K,L} g_{K,L}|_x = A_1|K| \quad et \quad \sum_{L\in N(K)} |K \cap L| A_{K,L} g_{K,L}|_y = A_2|K|.$$

Dès lors, (3.2) se réécrit $|K \cap L| \left(A_1 \frac{\partial U}{\partial x})_K^n + A_2 \frac{\partial U}{\partial y})_K^n + \mathcal{O}(h^3)\right)$ et G_K^n est alors égal à

$$\frac{\partial U}{\partial t})_K^n + A_1 \frac{\partial U}{\partial x})_K^n + A_2 \frac{\partial U}{\partial y})_K^n + \mathcal{O}(h) + \mathcal{O}(\Delta t).$$

Comme U est solution de l'équation, on a bien montré que G_K^n est consistant.

3.4.2 Correction de I_K^n

Pour rendre le terme I_K^n consistant, on introduit un correcteur γ_K^n que l'on cherche à déterminer. On définit alors

$$\tilde{I}_K^n = \frac{1}{|K|} \sum_{L \in N(K)} |K \cap L|(A_{K,L}^+(U_K^n + \gamma_K^n - U_{K,L}^n) + A_{K,L}^-(U_L^n + \gamma_L^n - U_{K,L}^n)).$$

Comme $U_L^n = U_{K,L}^n$,

$$\tilde{I}_K^n = \frac{1}{|K|} \sum_{L \in N(K)} |K \cap L| A_{K,L}^+(U_K^n + \gamma_K^n - U_{K,L}^n) + \sum_{L \in N_0(K)} |K \cap L| A_{K,L}^-(U_L^n + \gamma_L^n - U_{K,L}^n).$$

$$\tilde{I}_K^n = \frac{1}{|K|} \sum_{L \in N(K)} |K \cap L| A_{K,L}^+(\gamma_K^n + (G_K - G_{K,L}) \odot \nabla U_K^n) + \sum_{L \in N_0(K)} |K \cap L| A_{K,L}^-(\gamma_L^n + (G_L - G_{K,L}) \odot \nabla U_L^n).$$

On pose $\gamma_K^n = \Gamma_k^n \odot \nabla U_K^n$ et il vient alors

$$\tilde{I}_K^n = \frac{1}{|K|} \sum_{L \in N(K)} |K \cap L| A_{K,L}^+((\Gamma_K^n + G_K - G_{K,L}) \odot \nabla U_K^n) + \sum_{L \in N_0(K)} |K \cap L| A_{K,L}^-((\Gamma_L^n + G_L - G_{K,L}) \odot \nabla U_L^n).$$

On cherche alors Γ solution de l'équation suivante, car $\nabla U_K^n - \nabla U_L^n = \mathcal{O}(h)$,

$$\sum_{L \in N(K)} |K \cap L| A_{K,L}^+(\Gamma_K^n + G_K - G_{K,L}) + \sum_{L \in N_0(K)} |K \cap L| A_{K,L}^-(\Gamma_L^n + G_L - G_{K,L}) = 0. \tag{3.3}$$

On remarque qu'il y a une très grande ressemblance entre cette formule et l'équation du correcteur (1.3) de la partie 1 grâce aux notations choisies. Pour autant, elle est tout de même bien plus compliquée et les vecteurs sont devenus des matrices. Ces différences sont décrites ci-après.

3.5 Etude de Γ

On remarque tout d'abord que Γ_K est maintenant une matrice $(3, 2)$. On a implémenté en Matlab un programme (voir annexe) qui étant donné un maillage calcule les matrices A et B telles que :

$$A\Gamma = B.$$

A est une matrice carrée de taille $3J$ où J est le nombre de volumes, B est une matrice $(3J, 2)$. On a pu voir que la matrice A est en général non inversible mais que Γ existe. Pour autant, Γ n'est pas toujours un $\mathcal{O}(h)$ et ainsi, la méthode du correcteur semble ne pas fonctionner ici. On exploite les propriétés de la matrice A en annexe.

3.6 Conclusion

Dans cette partie, on a pu appliquer les méthodes vues précédemment pour obtenir un schéma convergent en dimension 1. Pour la dimension 2, on a pu définir un problème et le résoudre numériquement après avoir introduit l'équation d'un correcteur géométrique. Les résultats numériques se sont révélés décevants (A non inversible). Grâce aux résultats numériques, on a émis de nombreuses conjectures qui permettent de mieux cerner les phénomènes à l'oeuvre, de voir que le correcteur géométrique obtenue numériquement n'était pas satisfaisant car sa norme ne semble pas tendre vers 0 lorsqu'on raffine le maillage, en concluant que les problèmes observés devaient être dûs aux conditions aux bords difficiles à définir.

Conclusion

Ce stage nous a permis de rentrer dans l'univers de la recherche, et de nous initier au métier de chercheur. On a passé beaucoup de temps à chercher des démonstrations et on s'est aperçu que ce n'était pas toujours évident. Notamment lorsqu'on a une conjecture, on ne sait pas si on doit privilégier la recherche d'une preuve ou d'un contre-exemple. Pour autant, chacune des deux directions aide à mieux cerner le problème. De plus, les rendez-vous de stage avec M. Pascal et M. Bouche, à raison de deux par semaine, nous encourageaient à obtenir rapidement de nouveaux résultats ou à défaut de nouvelles questions. Lors de ces rendez-vous, durant lesquels on nous indiquait de nouvelles directions de recherche, nous avons appris à chercher en équipe dans une ambiance conviviale. Lorsque nous obtenions de nouveaux résultats, il nous fallait les rédiger en Latex puis les envoyer à M. Pascal qui nous les corrigeait avec une exigence proche de celle requise pour toute publication. Nous avons ainsi appris les bases de la rédaction d'articles scientifiques. Nous avons ainsi pu découvrir toutes les facettes du métier de chercheur. Lorsque le stage était à plein temps, nous avons pu assister à des groupes de travail et à chaque fois présenter une partie de nos travaux et les questions soulevées par ces derniers. Nous tenons maintenant à remercier nos deux encadrants de stage, M. Pascal et M. Bouche, pour le temps et la confiance qu'ils nous ont accordés ainsi que les conseils qu'ils nous ont prodigués. Nous passions en effet plusieurs heures par semaine dans le bureau de M. Pascal. Les thèmes des présentations abordés lors des groupes de travail, étaient plutôt libres, ce qui témoigne de la confiance que l'on nous accordait. Enfin, nous n'aurions pas obtenu tous ces résultats sans la précieuse aide de nos deux encadrants.

Annexe

Corrections apportées aux livres [1] et [2]

Démonstration de la proposition 4.4.2 du livre [1]

On rapppelle le problème de Cauchy d'ordre 1 en temps noté (4.20)

$$\left\{ \begin{array}{rcl} \frac{\partial u}{\partial t} & = & Au, \quad t > 0 \\ u(0) & = & u_0 \end{array} \right.$$

où A est un opérateur différentiel linéaire selon x. On rappelle que la tranformée de Fourier s'écrit $\hat{u}(t,\xi) = e^{tA(i\xi)}\hat{u}_0(\xi)$ et que $g(h\xi) = 1 - c\frac{k}{h}(e^{ih\xi} - 1)$. Le livre [1] donnait la proposition suivante.

Proposition 11 *Si le schéma aux différences finies est consistant avec le problème (4.20) alors on a*

$$\lim_{(k,h)\to(0,0)} \frac{1}{k}(e^{kA(i\xi)} - g(h\xi)) = 0.$$

Preuve. On pose $f(t,x) = e^{st}e^{ix\xi}$ et on définit $P_{k,h}(e^{st_n}e^{ix_j\xi}) = p_{k,h}(s,\xi)e^{st_n}e^{ix_j\xi}$. Pour toute fonction ϕ régulière, on a

$$(P\phi)_j^n = \frac{\partial \phi}{\partial t}(t_n, x_{_j}) - A\phi(t_n, x_j),$$

d'où $(Pf)_j^n = [s - A(i\xi)]e^{st_n}e^{ix_n\xi}$. Par hypothèse de consistance, on obtient $(P_{k,h}f - Pf)_j^n = \varepsilon(k,h)$, d'où

$$[p_{k,h}(s,\xi) - (s - A(i\xi))]e^{st_n}e^{ix_n\xi} = \varepsilon(k,h),$$

donc

$$p_{k,h}(s,\xi) - (s - A(i\xi)) = \varepsilon(k,h), \forall s, \forall \xi. \tag{3.4}$$

Or par définition de $g(h\xi)$, $P_{k,h}(g(h\xi)^n e^{ix_j\xi}) = 0$ et comme $g(h\xi)^n = e^{nk\frac{ln(g(h\xi))}{k}}$, on en déduit que $p_{k,h}(\frac{ln(g(h\xi))}{k}, \xi) = 0$. Ainsi, en posant $s = \frac{ln(g(h\xi))}{k}$ dans (1), on obtient

$$\frac{ln(g(h\xi))}{k} - A(i\xi) = \varepsilon(k,h),$$

d'où enfin

$$g(h\xi) = e^{k(A(i\xi)+\varepsilon(k,h))} = e^{kA(i\xi)}(1 + k\varepsilon(k,h) + \mathcal{O}(k^2)).$$

■

Démontration de la stabilité du schéma centré pour l'équation de Poisson en dimension 2

Dans le livre [2], il y avait la démonstration de la stabilité du schéma centré pour l'équation de Poisson en dimension 1. Celle pour la dimension 2 était laissé en exercice. On rappelle le problème de Poisson en dimension 2,

$$\left\{ \begin{array}{rcll} -\Delta u(x) & = & f(x), & x \in \Omega =]0,1[\times]0,1[, \\ u(x) & = & 0 & x \in \Gamma \end{array} \right.$$

On rappelle le schéma, appelé aussi schéma à cinq points du laplacien

$$-\frac{u_{i+1,j}-2u_{i,j}+u_{i-1,j}}{h_1^2}-\frac{u_{i,j+1}-2u_{i,j}+u_{i,j-1}}{h_2^2}=f(ih_1,jh_2), i\in\{1,...,N_1\}, j\in\{1,...,N_2\},$$

$$u_{i,j}=0, i\in\{0,N_1+1\}, j\in\{0,N_2+1\}.$$

On peut écrire ce système sous forme matricielle (voir [2])

$$Au=b,$$

où u est la solution discrète du problème. Pour obtenir la stabilité, il suffit que $||A^{-1}||_\infty$ soit borné. On montre facilement, comme fait dans l'article [2] pour la dimension 1, que A est définie positive et monotone en dimension 2 (son inverse est positif).

Proposition 12 *$||A^{-1}||_\infty$ est borné.*

Preuve. On considère le problème suivant

$$\begin{cases} -\Delta v(x)=1, & x\in]0,1[, y\in]0,1[, \\ v(x,0)=sin(\pi x) & x\in]0,1[\\ v(x,1)=e^\pi sin(\pi x) & x\in]0,1[\\ v(0,y)=-\frac{y(y-1)}{2} & y\in]0,1[\\ v(1,y)=-\frac{y(y-1)}{2} & y\in]0,1[\end{cases}$$

avec des conditions aux bords positives. La solution de ce problème est $v(x,y)=e^{\pi y}sin(\pi x)-\frac{y(y-1)}{2}$. On note ϵ l'erreur de consistance ; on a alors par définition $Av=b+\epsilon$, où b est un vecteur avec que des 1. Comme $||\epsilon||_\infty\leq\frac{h^2}{6}||\frac{\partial^4 u}{\partial x^4}||_\infty$, pour h assez petit, on sait que l'on a $b+\epsilon\geq\frac{1}{2}b$.
Or A est monotone donc $||A^{-1}||_\infty=||A^{-1}b||_\infty$. De plus, $v=A^{-1}(b+\epsilon)$ d'où $\frac{1}{2}||A^{-1}||_\infty=\frac{1}{2}||A^{-1}b||_\infty\leq||A^{-1}(b+\epsilon)||_\infty=||v||_\infty$. On en déduit enfin

$$||A^{-1}||_\infty\leq 2||v||_\infty\leq\frac{9}{8}.$$

■

Annexe du chapitre 1

Tentative pour montrer que B est nilpotente en dimension 3

On énonce le lemme 4.5 de l'article [4] en dimension 3

Lemme 13 *On considère un maillage de tétraèdre sans trou, topologiquement parlant. On suppose qu'il existe au moins une face interne d'un tétraèdre qui ne contient pas le vecteur* $\mathbf{a}$*. Alors il existe un plan brisé composé de faces du maillage séparant le maillage en deux sous-maillages non vide A et B, tel que pour chaque face du plan brisé, la normale à cette face orientée de A vers B ait un produit scalaire avec* $\mathbf{a}$ *positif. De manière équivalente, le plan brisé vu du côté de A est éclairé si* $\mathbf{a}$ *indique la direction de la lumière.*

On ne considère dans la suite que des triangles et tétraèdes qui vérifient les hypothèses de ce lemme. On introduit quelques définitions.

Définition 3 *On dit que r triangles K_1,K_2,...,K_r sont consécutifs si $\forall i\in[1,r-1]$ K_i et K_{i+1} ont une arête en commun et si pour tout triplet de trois triangles, aucune arête n'est commune aux trois triangles.*

Définition 4 *On appelle éventail de taille r, un ensemble de r triangles K_1,K_2,...,K_r consécutifs partageant un même sommet s. On dit que l'éventail est bien étendu si lorsqu'on considère l'ensemble des tétraèdres possèdant au moins l'une des faces K_1,K_2,...,K_r, obtenant ainsi un polyèdre T séparé par l'union des K_1,K_2,...,K_r en deux polyèdres dont l'un noté A est non vide, on a $\forall i\in[1,r]$ $a.N_{A,K_i}<0$. Ou de manière équivalente, l'éventail vu du côté de A est éclairé si* $\mathbf{a}$ *indique la direction de la lumière.*

Définition 5 *On appelle parapluie issu du sommet intérieur s, un éventail bien étendu de taille r qui boucle, i.e. telle que K_1 et K_r aient une arête en commun.*

Définition 6 *On appelle ombrelle de s, l'ensemble des tétraèdres ayant s pour sommet.*

On montre un premier lemme.

Lemme 14 *Soit M un maillage, T un tédraèdre, K une face interne et A une arête interne de K non colinéaire à* $\mathbf{a}$*. Soit* $\mathbf{N}$ *la normale sortante de K. On suppose que* $\mathbf{a}.\mathbf{N} < 0$*. On considère le plan* $P = (\mathbf{a}, A)$*. Il existe un tétraèdre T' dans le demi-espace limité par P ne contenant pas T, tel que T' possède une face K' possédant l'arête A et en désignant* $\mathbf{N}'$ *la normale sortante de K',* $\mathbf{a}.\mathbf{N}' < 0$*.*

Preuve. Même argument que le lemme 4.5 de l'article [4] en 2D. ■

Lemme 15 *(faux) Tout éventail bien étendu issu d'un sommet interne s peut se compléter en parapluie.*

Preuve. On considère l'ombrelle de s. Soit K_1,K_2,...,K_r , un éventail bien étendu de taille r, et T_1,T_2,...,T_r des tédraèdres possédant respectivement les faces K_1,K_2,...,K_r . On note $K_i = s_i s s_{i+1}$ et P_i le plan (a, ss_i). On utilise le lemme 13 pour prolonger l'éventail de manière bien étendue. On montre par l'absurde que l'on peut prolonger l'éventail avec des triangles dont au moins un sommet est dans le demi-espace limité par P_1 ne contenant pas T_1. En effet, si tel n'était pas le cas, on n'aurait aucun sommet de l'ombrelle dans ce demi-espace, donc pas de tétraèdre et s ne serait pas un sommet interne. On considère le nouvel éventail K_1,K_2,...,K_p bien étendu dont s_{p+1} est dans le demi-espace limité par P_1 ne contenant pas T_1. Puis on prolonge récursivement l'éventail à l'aide du lemme 1 en considérant l'intersection du demi-espace limité par P_1 ne contenant pas T_1 et du demi-espace limité par P_{p+1} ne contenant pas T_p. Tant qu'il y a un sommet de l'ombrelle dans cette intersection on prolonge l'éventail. Dès qu'il n'y a plus de sommet, après construction des triangles K_{p+1},K_{p+2},...,K_n, alors $s_1 s s_{n+1}$ est un triangle du maillage et on obtient un parapluie. Cette dernière affirmation est en fait fausse et le lemme 15 aussi. Il y a des contre-exemples. ■
La démonstration suivante ne peut donc aboutir comme on l'aurait souhaité.
Démonstration. Si le maillage initial n'a aucun sommet interne, alors tous les tétraèdres sont au bord et on utilise le lemme 13 de manière récurrente, en partant d'un triangle qui a au moins deux faces au bord, et en appliquant le lemme 13 à chaque nouveau triangle construit à partir d'une arête interne. Si le maillage initial a au moins un sommet interne, on construit un parapluie autour de ce sommet, puis on considère chaque sommet interne du parapluie et on complète l'éventail (formé par les triangles du plan brisé déjà construit) issu chacun de ces sommets à l'aide du lemme 14. On s'arrête dès qu'il n'y a plus de sommet interne issu de parapluie, non complété. On obtient alors un plan brisé séparant le maillage en deux maillages non vide. ■
Le lemme 4.5 en 3D (lemme 13) est alors faux, on en donne un contre-exemple dans la partie 1. En effet, on donne un exemple de cycle en 3D alors que si ce lemme était vrai, cela impliquerait l'existence d'un tétraèdre $\mathcal{N}_0^-$ vide, qui impliquerait qu'il n'existe pas de cycle en 3D.

Calcul de Γ pour un triangle lorsque a rentre par un seul coté

La figure ne change pas et l'on considère les mêmes triangles, simplement, le vecteur $\mathbf{a}$ n'étant pas dans la même direction (figure 3.1), on s'attend à ce que $\mathbf{\Gamma}_j^+$ ne dépende pas des mêmes correcteurs.

Intérieur. Nous conservons le même schéma pour plus de simplicité, l'angle θ est désormais supérieur à $\frac{\pi}{3}$ et toujours inférieur à $\frac{\pi}{2}$. On traduit toujours (1.3) :

(1) pour $\mathbf{\Gamma}_j^+$:

$$(\sin(\theta') + \sin(\frac{\pi}{3} + \theta))\mathbf{\Gamma}_j^+ - \sin(\theta)\mathbf{\Gamma}_2^- = \vec{X}$$

(2) pour $\mathbf{\Gamma}_2^-$:

$$\sin(\theta)\mathbf{\Gamma}_2^- = \sin(\theta')\mathbf{\Gamma}_4^+ + \sin(\frac{\pi}{3} + \theta)\mathbf{\Gamma}_3^+ - \vec{X}$$

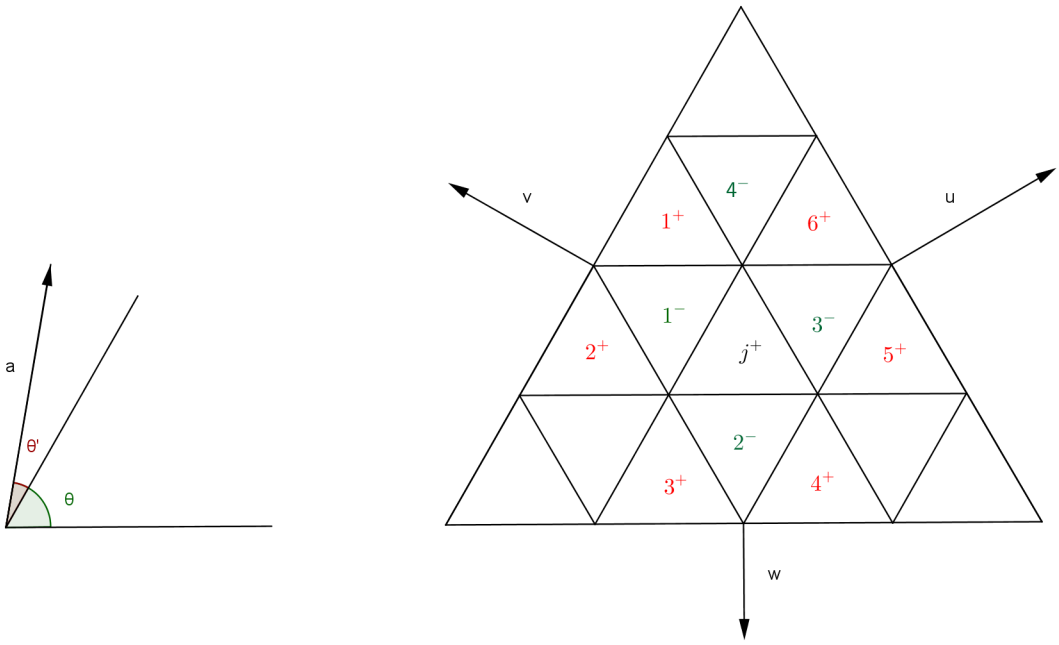

FIGURE 3.1 – Voisins du petit triangle j considéré.

où l'on a posé

$$\vec{X} = \frac{\sqrt{3}}{6}\frac{l}{2^r}\left(\sin(\frac{\pi}{3}+\theta)\vec{u} - \sin(\theta')\vec{v} + \sin(\theta)\vec{w}\right).$$

En remplaçant alors $\mathbf{\Gamma}_2^-$ par son expression dans (2) dans celle de $\mathbf{\Gamma}_j^+$ dans (1) il vient

$$\mathbf{\Gamma}_j^+ = \frac{\sin(\theta')}{S}\mathbf{\Gamma}_4^+ + \frac{\sin(\frac{\pi}{3}+\theta)}{S}\mathbf{\Gamma}_3^+,$$

avec $S = \sin(\theta') + \sin(\frac{\pi}{3}+\theta)$. On pose alors p = $\frac{\sin(\frac{\pi}{3}+\theta)}{S}$ pour obtenir

$$\mathbf{\Gamma}_j^+ = p\mathbf{\Gamma}_3^+ + (1-p)\mathbf{\Gamma}_4^+. \tag{3.5}$$

Bord bas. On considère par exemple 4^+ sur le bord bas du maillage :

$$(\sin(\frac{\pi}{3}+\theta) + \sin(\theta'))\mathbf{\Gamma}_4^+ = \frac{\frac{l}{2^r}\sqrt{3}}{6}(\sin(\frac{\pi}{3})\vec{u} + \sin(\theta')\vec{v}) + \epsilon$$

Soit

$$\mathbf{\Gamma}_4^+ = \frac{\frac{l}{2^r}\sqrt{3}}{6}(p\vec{u} + (1-p)\vec{v}) + \frac{\epsilon}{\sin(\frac{\pi}{3}+\theta) + \sin(\theta')}. \tag{3.6}$$

Annexe du chapitre 2

Convergence de deux autres schémas pour l'équation de transport avec terme source

Schéma avec z'

On s'intéresse à l'équation différentielle suivante

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} + b(u)z'(x) = 0.$$

Le schéma étudié est le suivant

$$S(v) = \frac{v_i^{n+1} - v_i^n}{\Delta t} + a\frac{v_i^n - v_{i-1}^n}{\Delta x_i} + b(v_i^n)z'(x_i).$$

On suppose disposer de u solution de l'équation et l'on applique S au vecteur de composantes $u_i^n = u(x_i, n\cdot\Delta t)$. On définit

$$\epsilon_i^n = \frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{u_i^n - u_{i-1}^n}{\Delta x_i} + b(u_i^n)z'(x_i).$$

On applique le schéma S à un vecteur w avec $w_i^n = u_i^n + \gamma_i^n$, où γ_i^n sera défini ultérieurement et est cherché sous la forme d'un $\mathcal{O}(\Delta x_i)$, on pose alors l'erreur de consistance corrigée

$$\delta_i^n = \frac{w_i^{n+1} - w_i^n}{\Delta t} + a\frac{w_i^n - w_{i-1}^n}{\Delta x_i} + b(w_i^n)z'(x_i).$$

On utilise une formule de Taylor à l'ordre 1 sur b supposée C^1 et l'on obtient comme $\gamma_i^n = \mathcal{O}(\Delta x_i)$

$$\delta_i^n = \frac{w_i^{n+1} - w_i^n}{\Delta t} + a\frac{w_i^n - w_{i-1}^n}{\Delta x_i} + b(u_i^n)z'(x_i) + \gamma_i^n b'(u_i^n)z'(x_i) + \mathcal{O}(\Delta x_i).$$

On utilise une formule de Taylor sur u et l'on obtient les deux formules suivantes :
i) $\frac{w_i^{n+1}-w_i^n}{\Delta t} = \frac{\gamma_i^{n+1}-\gamma_i^n}{\Delta t} + \frac{\partial u}{\partial t}(x_i, t_n) + \mathcal{O}(\Delta t)$
ii) $a\frac{w_i^n-w_{i-1}^n}{\Delta x_i} = a\frac{\gamma_i^n-\gamma_{i-1}^n}{\Delta x_i} + a\frac{\partial u}{\partial x}(x_i, t_n) - a\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}\frac{\partial u}{\partial x}(x_i, t_n) + \mathcal{O}(\Delta x_i)$
Comme u est solution exacte de l'équation, on a

$$\delta_i^n = \frac{\gamma_i^{n+1} - \gamma_i^n}{\Delta t} + a\frac{\gamma_i^n - \gamma_{i-1}^n}{\Delta x_i} + \gamma_i^n b'(u_i^n)z'(x_i) - a\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}\frac{\partial u}{\partial x}(x_i, t_n) + \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i).$$

On choisit alors $\gamma_i^n = \frac{\Delta x_i}{2}\frac{\partial u}{\partial x}(x_i, t_n)$ et l'on obtient avec une formule de Taylor :
– $\frac{\gamma_i^{n+1}-\gamma_i^n}{\Delta t} = \mathcal{O}(\Delta x_i)$
– $a\frac{\gamma_i^n-\gamma_{i-1}^n}{\Delta x_i} = a\frac{\Delta x_i-\Delta x_{i-1}}{2\Delta x_i}\frac{\partial u}{\partial x}(x_i, t_n) + \mathcal{O}(\Delta x_i)$
– $\gamma_i^n b'(u_i^n)z'(x_i) = \mathcal{O}(\Delta x_i)$
Avec un tel vecteur γ, on a donc

$$S(u + \gamma) = \mathcal{O}(\Delta x) + \mathcal{O}(\Delta t).$$

Schéma avec z décentré à droite

On étudie un troisième schéma

$$S(v) = \frac{v_i^{n+1} - v_i^n}{\Delta t} + a\frac{v_i^n - v_{i-1}^n}{\Delta x_i} + b(v_i^n)\frac{z(x_{i+1}) - z(x_i)}{\Delta x_i}.$$

On suppose disposer de u solution de l'équation et l'on applique S au vecteur composé de $u_i^n = u(x_i, n\cdot\Delta t)$. On définit l'erreur de consistance

$$\epsilon_i^n = \frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{u_i^n - u_{i-1}^n}{\Delta x_i} + b(u_i^n)\frac{z(x_{i+1}) - z(x_i)}{\Delta x_i}.$$

On applique le schéma S à un vecteur w avec $w_i^n = u_i^n + \gamma_i^n$, où γ_i^n sera défini ultérieurement et est cherché sous la forme d'un $\mathcal{O}(\Delta x_i)$ on pose alors l'erreur de consistance corrigée

$$\delta_i^n = \frac{w_i^{n+1} - w_i^n}{\Delta t} + a\frac{w_i^n - w_{i-1}^n}{\Delta x_i} + b(w_i^n)\frac{z(x_{i+1}) - z(x_i)}{\Delta x_i}.$$

On utilise une formule de Taylor à l'ordre 1 sur b supposée C^1 pour obtenir

$$\epsilon_i^n = \frac{w_i^{n+1} - w_i^n}{\Delta t} + a\frac{w_i^n - w_{i-1}^n}{\Delta x_i} + b(u_i^n)\frac{z(x_{i+1}) - z(x_i)}{\Delta x_i} + \gamma_i^n b'(u_i^n)\frac{z(x_{i+1}) - z(x_i)}{\Delta x_i} + \mathcal{O}(\Delta x_i),$$

qui se réécrit en utilisant une formule de Taylor sur z

$$\epsilon_i^n = \frac{w_i^{n+1} - w_i^n}{\Delta t} + a\frac{w_i^n - w_{i-1}^n}{\Delta x_i} + b(u_i^n)z'(x_i) + \gamma_i^n b'(u_i^n)\frac{z(x_{i+1}) - z(x_i)}{\Delta x_i} - b(u_i^n)\frac{\Delta x_i - \Delta x_{i+1}}{2\Delta x_i}z'(x_i) + \mathcal{O}(\Delta x_i).$$

En développant comme précédemment, on obtient

$$\epsilon_i^n = \frac{\gamma_i^{n+1} - \gamma_i^n}{\Delta t} + a\frac{\gamma_i^n - \gamma_{i-1}^n}{\Delta x_i} + \gamma_i^n b'(u_i^n)\frac{z(x_{i+1}) - z(x_i)}{\Delta x_i} - a\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}\frac{\partial u}{\partial x}(x_i, t_n) - b(u_i^n)\frac{\Delta x_i - \Delta x_{i+1}}{2\Delta x_i}z'(x_i)$$
$$+\mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i).$$

On définit γ_i^n par $\gamma_i^n = \frac{\Delta x_i}{2}\frac{\partial u}{\partial x}(x_i, t_n) - \frac{\Delta x_{i+1}}{2a}b(u_i^n)z'(x_i)$. Une formule de Taylor donne alors :
i) $\frac{\gamma_i^{n+1} - \gamma_i^n}{\Delta t} = \mathcal{O}(\Delta t)$
ii) $a\frac{\gamma_i^n - \gamma_{i-1}^n}{\Delta x_i} = a\frac{\Delta x_i - \Delta x_{i-1}}{2\Delta x_i}\frac{\partial u}{\partial x}(x_i, t_n) + b(u_i^n)\frac{\Delta x_i - \Delta x_{i+1}}{2\Delta x_i}z'(x_i) + \mathcal{O}(\Delta x_i)$
iii) $\gamma_i^n b'(u_i^n)\frac{z(x_{i+1}) - z(x_i)}{\Delta x_i} = \mathcal{O}(\Delta x_i)$
Avec un tel vecteur γ, on a finalement

$$S(u + \gamma) = \mathcal{O}(\Delta x) + \mathcal{O}(\Delta t).$$

Annexe du chapitre 3

Propriétés de A

On a d'abord montrer que les blocs $(3, 3)$ diagonaux étaient inversibles.

Lemme 16 *Les blocs diagonaux* $(3, 3)$ *de* A *sont inversibles.*

Preuve. On raisonne par l'absurde. Un tel bloc s'écrit

$$\begin{pmatrix} c^2 \sum_{L\in\mathcal{N}(K)} |K \cap L| & 0 & 0 \\ 0 & c^2 \sum_{L\in\mathcal{N}(K)} |K \cap L| n_1^2 & c^2 \sum_{L\in\mathcal{N}(K)} |K \cap L| n_1 n_2 \\ 0 & c^2 \sum_{L\in\mathcal{N}(K)} |K \cap L| n_1 n_2 & c^2 \sum_{L\in\mathcal{N}(K)} |K \cap L| n_2^2 \end{pmatrix}.$$

Or on montre que $\left(\sum_{L\in\mathcal{N}(K)} |K \cap L| n_1^2\right)\left(\sum_{L\in\mathcal{N}(K)} |K \cap L| n_2^2\right) - \left(\sum_{L\in\mathcal{N}(K)} |K \cap L| n_1 n_2\right)^2 > 0$, en utilisant le fait que le cas d'égalité de l'inégalité de Cauchy-Schwartz n'est pas possible ici. Sinon le vecteur des n_1 serait colinéaire au vecteur des n_2. D'où pour K fixé, il existerait λ tel que pour tout L, $n_1 = \lambda n_2$ et comme $n_1^2 + n_2^2 = 1$, $n_2^2 = \frac{1}{1+\lambda^2}$, il n'y a donc que deux vecteurs n possibles (car λ a un signe constant) pour le volume K alors qu'il y a au moins trois arêtes. Contradiction ■

La matrice A est alors inversible pour les maillages ne comportant qu'un seul volume. On a ensuite calculé le nombre de valeurs propres nulles de A et on a essayé de le relier au maillage considéré. Des résultats numériques nous ont amenés à conjecturer le résultat suivant
Conjecture. Le nombre de valeurs propres nulles de A est $\frac{T - A_b}{2} + 1$. On utilise les notations suivantes

* T est le nombre de triangle
* A_r est le nombre d'arêtes
* A_b est le nombre d'arêtes de bord
* A' est le nombre d'arêtes intérieures
* S est le nombre de sommets.

Pour que la formule ait un sens il faut que $T - A_b$ soit pair. C'est ce que l'on montre grâce à la formule d'Euler.

Théorème 17 *$T - A_b$ est pair.*

Preuve. On sait $T - A_r + S = 1$ et l'on a de plus $2A' + A_b = 3T$, ainsi

$$2T - 2A_r + 2S = 2,$$

$$2A_r - A_b = 3T.$$

En sommant, on obtient

$$2T - A_b + 2S = 3T + 2,$$

soit enfin

$$T + A_b = 2S - 2.$$

Comme $T + A_b$ est pair, $T - A_b$ aussi. ■

Remarque. On a $N = \frac{T-A_b}{2} + 1 = \frac{T-3T+2A'}{2} + 1 = A' - T + 1$. T est le nombre de blocs $(3,3)$ diagonaux qui sont tous non nuls ($\sum_{L \in N(K)} |K \cap L| A^+_{K,L} \neq 0$) de la matrice A comme le montre l'équation vérifiée par Γ. A' est quant à lui le nombre de blocs $(3,3)$ non nuls au dessus de la diagonale ($A^-_{K,L} \neq 0$). *Remarque 2.* On montre même que $N = \frac{T-A_b}{2} + 1 = S_{int}$ pour les maillages sans trou où S_{int} est le nombre de sommets intérieurs, en utilisant le fait que $A_b = S_b$ où S_b est le nombre de sommets au bord. On illustre cette propriété à l'aide d'un maillage implémenté sur Matlab sur la figure 3.2.

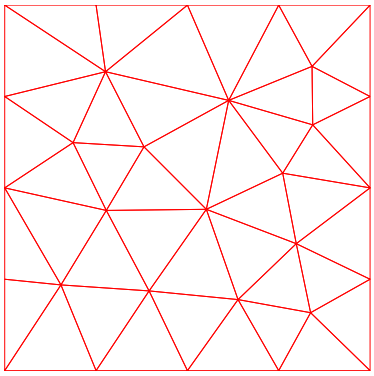

FIGURE 3.2 – $T = 42$, $A_b = 16$, $S_{int} = 14$

Une expression intéressante du nombre de valeurs propres nulles

Rappelons que tout ceci n'est, pour le moment, qu'une conjecture. On note $T_{i,j}$ un trajet minimal en nombre de volumes parcourus entre les volumes i et j. Puis, $N_{i,j}$ le nombre de triangles non parcourus par $T_{i,j}$. On a

$$N = \min_{i,j}(N_{i,j}).$$

Programme calculant la matrice A

On a adapté un programme rédigé par M. PASCAL au cas de la dimension 2. En effet, les similitudes remarquées précédemment entre l'équation du correcteur en dimension 1 et 2 permettent d'avoir le même type de programme en changeant les dimensions des matrices. Voici la partie du programme que l'on a modifiée, c'est celle qui calcule la matrice A étant données de nombreuses caractéristiques du maillage considéré (diponibles dans mesh).

Code Matlab

```
%
function [A,M,F] = assemb_AF(beta,mesh)
% assemb_A(beta,mesh) assemble la matrice de connectivite de la methode de VF
% pour le pb de convection
%    du/dt   + div (beta u) = 0
% sur le maillage mesh (structure contenant les champs
% (nbs,nbt,elm_som,som_coo,som_zon))

% Copyright (c) 2005 by Frederic Pascal, ENS de Cachan

% Initialisation
A = sparse(mesh.nbt,mesh.nbt);
M = sparse(mesh.nbt,mesh.nbt);
F = zeros(mesh.nbt,2);

% beta ligne -> beta colonne
s=size(beta);
if s(2)==2
    beta=beta';
end

% beta -> beta par arete
if prod(s)==2
    beta = beta*ones(1,mesh.nba);
end

% Normales mesurees
fac_nor_mes= mesh.fac_nor(:,:).*(mesh.fac_mes(:)*ones(1,2));

% On parcourt les aretes/faces
for nf = 1:mesh.nba

    % elements adjacents
    nfe = mesh.fac_elm(nf,:);
    % N=normale*mesure
    nfn= fac_nor_mes(nf,:);
    % centre de gravite
    nfg = mesh.fac_gra(nf,:);

    % beta*N
    bn = nfn*beta(:,nf);

    if (nfe(2) ~= 0)
        % arete/face interieure

        ie1 = nfe(1);
        ie2 = nfe(2);
```

```
        if (bn > 0)
            g = mesh.elm_gra(ie1,:);

            M(ie1,ie1) = M(ie1,ie1) + bn;

            A(ie1,ie1) = A(ie1,ie1) + bn;
            A(ie2,ie1) = A(ie2,ie1) - bn;

            F(ie1,:) = F(ie1,:) + bn * (nfg - g);
            F(ie2,:) = F(ie2,:) - bn * (nfg - g);

        elseif (bn < 0)
            g = mesh.elm_gra(ie2,:);

            M(ie2,ie2) = M(ie2,ie2) - bn;

            A(ie2,ie2) = A(ie2,ie2) - bn;
            A(ie1,ie2) = A(ie1,ie2) + bn;

            F(ie2,:) = F(ie2,:) - bn * (nfg - g);
            F(ie1,:) = F(ie1,:) + bn * (nfg - g);

        end

    else
        % arete/face de bord

        ie1 = nfe(1);

        if (bn > 0)
            g = mesh.elm_gra(ie1,:);

            M(ie1,ie1) = M(ie1,ie1) + bn;

            A(ie1,ie1) = A(ie1,ie1) + bn;

            F(ie1,:) = F(ie1,:) + bn * (nfg - g);
        end

    end
end

disp('Fin assemblage Matrices')
```

Bibliographie

[1] Laurent Di Menza. *Analyse numérique des équations aux dérivés partielles.* Enseignement des Mathématiques (Cassini) 24. Paris : Cassini. xii, 221 p., 2009.

[2] Brigitte Lucquin. *Equations aux dérivées partielles et leurs approximations. Niveau M1.* Paris : Ellipses. v, 227 p., 2004.

[3] Frédéric Pascal. On supra-convergence of the finite volume method for the linear advection problem. In *Paris-Sud Working Group on Modelling and Scientific Computing 2006–2007*, volume 18 of *ESAIM Proc.*, pages 38–47. EDP Sci., Les Ulis, 2007.

[4] Daniel Bouche, Jean-Michel Ghidaglia, and Frédéric Pascal. Error estimate and the geometric corrector for the upwind finite volume method applied to the linear advection equation. *SIAM J. Numer. Anal.*, 43(2) :578–603 (electronic), 2005.

[5] Chiara Simeoni. Remarks on the consistency of upwind source at interface schemes on nonuniform grids. *J. Sci. Comput.*, 48(1-3) :333–338, 2011.