



**HAL**  
open science

# Schémas boîte hermitiens - Algorithmes rapides pour la discrétisation des équations aux dérivés partielles

Ali Abbas

► **To cite this version:**

Ali Abbas. Schémas boîte hermitiens - Algorithmes rapides pour la discrétisation des équations aux dérivés partielles. Mathématiques générales [math.GM]. Université Paul Verlaine - Metz, 2011. Français. NNT : 2011METZ024S . tel-01749061v1

**HAL Id: tel-01749061**

**<https://hal.univ-lorraine.fr/tel-01749061v1>**

Submitted on 29 Mar 2018 (v1), last revised 16 Mar 2016 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

**Université Paul Verlaine-Metz**  
**Ecole doctorale IAEM Lorraine**

THÈSE en vue de l'obtention du  
DOCTORAT DE MATHÉMATIQUES

Ali ABBAS

sous la direction de

**Prof. Jean-Pierre CROISILLE**

**Schémas boîte hermitiens - Algorithmes rapides  
pour la discrétisation des équations aux dérivées  
partielles**

**Soutenue le 9 novembre 2011 devant le jury composé de :**

M. Olivier BOTELLA,	Maître de Conférences à l'Université de Nancy, Examineur
Mme Paola CINNELLA,	Professeur à l'ENSAM, Examinatrice
M. Jean-Pierre CROISILLE,	Professeur à l'Université de Metz, Directeur de thèse
M. Bruno DESPRES,	Professeur à l'Université de Paris 6, Rapporteur
M. Vincent GIOVANGIGLI,	Directeur de recherches à l'Ecole Polytechnique, Rapporteur
M. Alain LERAT,	Professeur à l'ENSAM, Examineur
M. Dong YE,	Professeur à l'Université de Metz. Examineur



# Table des matières

<b>1</b>	<b>Algèbre des schémas aux différences sur <math>\Omega = (a, b)</math></b>	<b>5</b>
1.1	Fonctions de grille . . . . .	5
1.2	Opérateurs aux différences . . . . .	7
1.3	Notations vectorielles et matricielles . . . . .	10
<b>2</b>	<b>Schémas boîte hermitiens en dimension 1</b>	<b>17</b>
2.1	Le schéma boîte de Keller . . . . .	17
2.2	Schéma boîte hermitien HB1 . . . . .	19
2.2.1	Design du schéma HB1 . . . . .	19
2.2.2	Expression matricielle du schéma HB1 . . . . .	23
2.2.3	Stabilité . . . . .	32
2.2.4	Principe du maximum . . . . .	34
2.2.5	Résultats numériques . . . . .	36
2.3	Schéma boîte hermitien HB2 . . . . .	39
2.3.1	Design du schéma HB2 . . . . .	39
2.3.2	Expression matricielle du schéma HB2 . . . . .	40
2.3.3	Principe du maximum pour le schéma HB2 . . . . .	45
2.3.4	Relation entre les schémas HB1 et HB2 . . . . .	48
2.3.5	Résultats numériques . . . . .	52
<b>3</b>	<b>Schémas boîte hermitiens sur géométrie cartésienne</b>	<b>57</b>
3.1	Introduction . . . . .	57
3.2	Opérateurs aux différences en deux dimensions . . . . .	58
3.3	Notations algébriques . . . . .	59
3.3.1	Notations vectorielles . . . . .	63
3.4	Fonctions de grille et algèbre de Kronecker . . . . .	64
3.5	Schéma boîte hermitien en dimension 2 . . . . .	66
3.5.1	Introduction . . . . .	66
3.5.2	Définition du schéma pour le problème de Poisson . . . . .	67
3.5.3	Exemples d'approximations de $\frac{\partial u}{\partial n}$ sur $\partial\Omega$ . . . . .	70
3.6	Forme matricielle du schéma boîte hermitien en dimension 2 . . . . .	73
3.6.1	Forme matricielle générale des dérivées au bord . . . . .	74
3.6.2	Forme matricielle générale du schéma boîte hermitien . . . . .	78
3.7	Le schéma boîte hermitien HB1 en dimension 2 . . . . .	84

3.7.1	Forme matricielle du schéma HB1	84
3.7.2	Algorithme rapide pour le schéma HB1	86
3.8	Mise en oeuvre du schéma HB1	87
3.8.1	Introduction	87
3.8.2	Détails de l'implémentation	87
3.8.3	Résultats numériques pour le schéma HB1	91
3.9	Le schéma boîte hermitien HB2 en dimension 2	95
3.9.1	Forme matricielle du schéma HB2	95
3.9.2	Algorithme rapide pour le schéma HB2	97
3.10	Mise en oeuvre du schéma HB2	100
3.10.1	Introduction	100
3.10.2	Détail de l'implémentation de l'étape 3 (résolution du système auxiliaire)	100
3.10.3	Résultats numériques pour le schéma HB2	104
<b>4</b>	<b>Algorithmes de type boîte hermitien</b>	<b>111</b>
4.1	Introduction	111
4.2	Définition du schéma (HB) pour certains problèmes elliptiques non séparables	112
4.3	Forme matricielle du schéma boîte hermitien	115
4.4	Algorithme de Concus et Golub pour le schéma boîte hermitien	119
4.4.1	Introduction	119
4.4.2	Schéma itératif utilisant le schéma boîte hermitien	120
4.4.3	Résultats numériques	124
4.5	Algorithmes de type "Local Defect Correction" pour les schémas boîte hermitiens	127
4.5.1	Introduction	127
4.5.2	Schéma boîte hermitien et algorithme LDC en dimension 1	132
4.5.3	Résultats numériques	134
4.5.4	Schéma boîte hermitien avec LDC en dimension 2	137
4.5.5	Interpolation de bord par spline cubique	139
4.5.6	Mise en oeuvre de l'algorithme LDC	143
4.5.7	Résultats numériques	143
<b>5</b>	<b>Schémas boîte hermitiens en dimension 3</b>	<b>147</b>
5.1	Introduction	147
5.2	Opérateurs aux différences en trois dimensions	147
5.3	Notations algébriques	150
5.4	Fonctions de grille et algèbre de Kronecker	154
5.5	Schéma boîte hermitien en dimension 3	161
5.5.1	Introduction	161
5.5.2	Définition du schéma hermitien (HB) pour le problème de Poisson dans $\Omega = (a, b)^3$	161
5.5.3	Exemples d'approximations de $\frac{\partial u}{\partial n}$ sur $\partial\Omega$	165
5.5.3.1	Schéma HB1	165

---

5.5.3.2	Schéma HB2 . . . . .	167
5.6	Forme matricielle du schéma boîte hermitien en dimension trois : cas général . . . . .	169
5.6.1	Introduction . . . . .	169
5.6.2	Forme matricielle générale des dérivées au bord . . . . .	169
5.7	Dérivées normales . . . . .	170
5.8	Dérivées tangentielles . . . . .	173
5.9	Expression du gradient $(U_x, U_y, U_z)$ en fonction de $U$ . . . . .	177
5.10	Forme matricielle générale du schéma boîte hermitien . . . . .	178
5.11	Solveurs rapides . . . . .	185
5.11.1	Solveur rapide pour le schéma HB1 . . . . .	185
5.11.2	Résultats numériques pour le schéma HB1 . . . . .	186
5.11.3	Solveur rapide pour le schéma HB2 . . . . .	187
5.11.4	Résultats numériques pour le schéma HB2 . . . . .	192





# Introduction générale

La conception de schémas aux différences compacts d'ordre élevé pour l'équation de Laplace dans un carré ou dans un cube est un thème classique en Mathématiques Appliquées et calcul scientifique. Les livres [10], [15], [20], [21], [24], [29], [33] traitent ce problème comme une question éducative de base sur les méthodes numériques. Ces dernières années, on a noté un regain d'intérêt sur cette question. On renvoie à [11] pour une technique d'utilisation d'un solveur rapide dans la résolution de quelques problèmes elliptiques non séparables. Dans [38] un schéma différence finie compact d'ordre quatre est utilisé avec un algorithme multigrille pour la résolution de l'équation de Poisson en trois dimensions. L'algorithme dans [8] est basé sur un solveur utilisant la décomposition de la solution en une partie analytique et autre spectrale, cet algorithme est optimal dans des cas particuliers. Dans [36], un schéma compact utilisant la procédure multigrille V-cycle et l'extrapolation de Richardson donne une précision d'ordre six pour la solution sur la grille grossière. On renvoie à [30] pour quelques utilisations de solveur de Poisson rapide dans l'étude des Streamers, quelques limitations de schéma différence finie standard sont mentionnées à la fin de cet article.

Au delà du design de schémas numériques particuliers, qui traitent de la précision et de la stabilité, la nécessité d'un solveur rapide efficace est un point crucial pour les calculs pratiques. Sur cette question on renvoie à [5]. L'utilisation de tels solveurs dans des géométries canoniques reste essentiel dans les codes de calcul en physique. Des exemples sont la mécanique des fluides (Equations de Navier-Stokes compressibles et incompressibles), [4], [16], [18], l'équation de Helmholtz [7], [9], les calculs en astrophysique, [28], en géophysique [34].

Dans cette thèse, on introduit un schéma compact d'ordre quatre sur une grille cartésienne pour le problème de Poisson dans un carré et dans un cube, dont le design est basé sur le travail préliminaire de Croisille [12]. On appelle le schéma "schéma boîte hermitien" dans la suite car il combine une approche volumes finies (boîte) avec un calcul hermitien de la dérivée comme dans les méthodes mixtes. Il fournit par construction une

approximation de  $u$  et  $\nabla u$ . De plus seulement les moyennes du terme source  $f(x, y)$  sont utilisées au lieu des valeurs ponctuels  $f(x_i, y_j)$ . C'est une propriété intéressante dans l'approximation de problèmes en électromagnétisme.

La partie essentielle du mémoire a consisté en la mise au point d'une procédure de résolution rapide utilisant la décomposition de la matrice du schéma en une somme d'une matrice diagonale (dans une base spectrale) et d'une matrice auxiliaire de rang petit. La résolution est effectuée par un algorithme de résolution directe utilisant la formule de Sherman-Morrison-Woodbury et la FFT (voir [6] ou [3] pour un contexte différent). De plus on a poussé l'implémentation jusqu'à la mise en oeuvre en 3 dimensions dans un cube, afin de prouver l'intérêt de la méthode proposée pour les applications.

## Plan et présentation des résultats

### Chapitre 1 : Algèbre des schémas aux différences sur $\Omega = (a, b)$

Dans ce chapitre nous introduisons les notations utilisées dans la thèse. On introduit la notion de fonction de grille qu'on utilise comme modèle de fonction discrète aux noeuds du maillage. On rappelle aussi quelques opérateurs aux différences usuels. Pour approcher les moyennes sur les cellules de maille on introduit l'opérateur de Simpson. Nous rappelons ainsi quelques notations vectorielles et matricielles nécessaires pour la construction des schémas boîte. On présente les propriétés importantes des valeurs et vecteurs propres de la matrice de Laplace  $T$  (opérateur à trois points), qui joue un rôle fondamental dans les algorithmes de résolution rapide des schémas.

### Chapitre 2 : Schémas boîte hermitiens en dimension 1

Dans ce chapitre, nous rappelons la notion de schéma boîte de Keller pour l'équation de Poisson en dimension 1. On rappelle aussi le schéma boîte en dimension 1 introduit par J.P. Croisille [12] qui par construction utilise des approximations d'ordre 1 des dérivées au bord. Ce schéma (appelé HB1) possède une matrice  $\mathcal{H}_1$  diagonalisable dans la base de vecteurs propres de la matrice du Laplacien  $T$ . En pratique, on observe un ordre 2 de convergence pour la solution et sa dérivée. On propose ensuite sur le même principe un schéma boîte d'ordre plus élevé qui utilise des approximations d'ordre 3 des dérivées au bord. On donne une forme matricielle générale du schéma boîte pour le problème de Poisson en dimension 1. On en déduit la matrice du nouveau schéma boîte obtenu (appelé HB2). La matrice  $\mathcal{H}_2$  associée au schéma HB2 s'écrit comme somme d'une matrice diagonale et d'une matrice de petit rang. Des résultats numériques montrent que ce schéma est d'ordre 4 pour la solution et sa dérivée. Les courbes de convergence

données à la fin de ce chapitre illustrent l'intérêt du nouveau schéma HB2 par rapport au schéma HB1 et au schéma différence finie classique à trois points (appelé DF2).

### Chapitre 3 : Schémas boîte hermitiens sur géométrie cartésienne

On présente une généralisation en deux dimensions de la construction du schéma boîte 1D pour le problème de Poisson sur un maillage carré. Après avoir rappelé les opérateurs aux différences et les notations algébriques et vectorielles nécessaires, on introduit un nouveau schéma de type boîte hermitien sur une grille cartésienne pour le problème de Poisson sur un carré. On observe un ordre de convergence de 4 pour la solution et son gradient. L'effort principal a porté sur la conception d'une procédure de résolution rapide basée sur la décomposition de la matrice associée comme une somme d'une matrice diagonale (dans une base convenable) et une matrice de petit rang. On présente ainsi une procédure de résolution rapide basée sur la formule de Sherman-Morrison-Woodbury. Des résultats numériques montrent la convergence d'ordre 4 de ce schéma en normes  $L^2$  et  $L^\infty$  discrètes.

### Chapitre 4 : Algorithmes de type boîte hermitien

On introduit deux types d'algorithmes utilisant les schémas boîte hermitiens en dimension 2 sur un maillage cartésien dans un carré.

- La première application concerne certains problèmes elliptiques non séparables. La méthode itérative proposée utilise les méthodes de "shift" et de "scaling" de Concus [11]. On présente trois algorithmes itératifs qui utilisent comme solveur de base les schémas pour le problème de Poisson précédemment définis. Le premier algorithme utilise le schéma HB1, le deuxième utilise le schéma HB2 et le troisième algorithme utilise le schéma HB1 dans l'initialisation et le schéma HB2 pour les itérations. Ce dernier algorithme est très intéressant lorsque la solution à calculer est oscillante. On présente plusieurs résultats numériques prouvant l'efficacité des algorithmes proposés.
- La deuxième application concerne la méthode LDC introduite par W. Hackbusch [22]. L'algorithme proposé dans cette partie utilise seulement deux niveaux de grille et le domaine de raffinement a été choisi a priori. La grille grossière  $\Omega^H$  qui est une discrétisation de tout le domaine de calcul  $\Omega$  et une grille fine  $\Omega_l^h$  qui est une discrétisation d'un sous-domaine  $\Omega_l \subset \Omega$  où la solution est supposée posséder de forts gradients. Le travail est inspiré de celui d'Anthonissen [2] qui utilise cette technique avec le schéma de Laplace cinq points. Ici, on utilise cette méthode avec le schéma HB2. Une interpolation spline  $P^{h,H}$  est utilisée en dimension 2 pour obtenir des valeurs artificielles sur l'interface  $\Gamma$  entre la grille grossière et la grille fine. Des résultats

numériques montrent l'intérêt de la méthode LDC combinée avec le schéma HB2.

### **Chapitre 5 : Schémas boîte hermitiens en dimension 3**

Le but de ce chapitre est d'obtenir un nouveau schéma boîte pour le problème de Poisson en dimension 3. La difficulté principale concerne l'obtention de la forme matricielle des vecteurs correspondant aux conditions de bord. Après avoir obtenu la forme matricielle générale du schéma boîte en 3D pour le problème de Poisson, on donne un algorithme de résolution des schémas boîte HB1 et HB2. Le code de calcul est effectué en Matlab. Comme en dimension 2, les schémas HB1 et HB2 sont respectivement d'ordre 2 et 4 pour la solution et son gradient.

# Chapitre 1

## Algèbre des schémas aux différences sur $\Omega = (a, b)$

### 1.1 Fonctions de grille

On considère un intervalle  $\Omega = (a, b)$ ,  $a < b$ , de longueur  $L = b - a$ . Les lettres latines sont utilisées pour des fonctions continues, comme  $u(x), v(x)$ ,  $x \in \Omega$ . En général, nos fonctions seront à valeurs réelles, de sorte que le produit scalaire dans  $L^2(\Omega)$  est donné par

$$(u, v) = \int_a^b u(x)v(x)dx. \quad (1.1)$$

La norme sur  $L^2(\Omega)$  est donnée par

$$\|u\|_{L^2(\Omega)} = (u, u)^{1/2} = \left( \int_a^b u^2(x)dx \right)^{1/2}. \quad (1.2)$$

La norme sur  $L^\infty(\Omega)$  est

$$\|u\|_{L^\infty(\Omega)} = \sup_{a \leq x \leq b} |u(x)|. \quad (1.3)$$

On considère une grille différences finies régulière,

$$a = x_0 < x_1 < \dots < x_{N-1} < x_N = b, \quad (1.4)$$

où

$$x_i = a + ih, \quad i = 0, 1, \dots, N, \quad h = \frac{b-a}{N}. \quad (1.5)$$

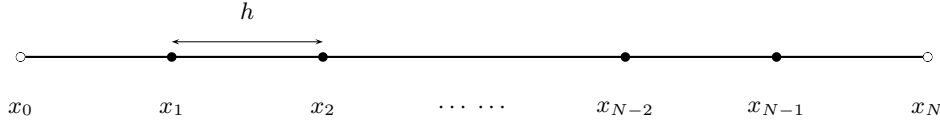


FIG. 1.1: Grille en dimension 1. Les deux points de bord sont  $x_0, x_N$  (notés par “o”). Les points  $x_i, i = 1, \dots, N - 1$  sont les points intérieurs (notés par “•”).

Les points intérieurs sont les points  $x_1, x_2, \dots, x_{N-1}$ . Les deux points de bord sont  $x_0$  et  $x_N$  (figure 1.1).

On discrétise une fonction donnée  $u(x)$  par des valeurs appropriées définies aux points de grille. Nous distinguons deux niveaux de représentation des fonctions aux points de grille.

(i) Si  $u(x)$  est une fonction donnée sur  $\Omega = (a, b)$ , on note  $u^*$  la fonction de grille définie par  $u_i^* = u(x_i)$ , soit

$$u^* = (u_0, u_1, \dots, u_N). \quad (1.6)$$

Du manière générale, une fonction de grille est une fonction définie sur la grille discrète  $\{x_i\}_{i=0}^N$ . Lorsqu’une fonction de grille est considérée avec  $h > 0$ , fixé ( $h = \frac{L}{N}$ ), elle s’identifie avec les valeurs qu’elle prend aux points  $x_0, x_1, \dots, x_N$ ,

$$u = (u_0, u_1, \dots, u_N). \quad (1.7)$$

On note par  $l_h^2$  l’espace des fonctions de grille, avec  $h > 0$  fixé. Cet espace est muni d’un produit scalaire et d’une norme associée

$$(u, v)_h = h \sum_{j=0}^N u_j v_j, \quad |u|_h^2 = h \sum_{j=0}^N u_j^2. \quad (1.8)$$

Le sous-espace des fonctions de grille ayant des conditions au bord nulles en  $x_0 = a$  et  $x_N = b$ , est notée par  $l_{h,0}^2$ . Pour des fonctions de grille  $u, v \in l_{h,0}^2$  on a

$$(u, v)_h = h \sum_{j=1}^{N-1} u_j v_j. \quad (1.9)$$

On définit aussi la norme max pour une fonction de grille  $u$

$$\|u\|_\infty = \max_{0 \leq i \leq N} |u_i|. \quad (1.10)$$

(ii) On note par des lettres capitales les vecteurs et matrices dans les espaces vectoriels réels.  $U \in \mathbb{R}^{N+1}$  est

$$U = [u_0, u_1, u_2, \dots, u_N]^T \in \mathbb{R}^{N+1}. \quad (1.11)$$

La norme euclidienne dans  $\mathbb{R}^{N+1}$  est notée par  $|U|$ . Pour  $A \in \mathbb{M}_{N+1}(\mathbb{R})$ , matrice réelle d'ordre  $N + 1$ , on note

$$|A|_2 = \sup_{U \neq 0} \frac{|AU|}{|U|}. \quad (1.12)$$

Si  $A$  est symétrique, alors  $|A|_2 = \rho(A)$ , où  $\rho(A)$  est le rayon spectral de  $A$ ,

$$\rho(A) = \max_i |\lambda_i|, \quad (1.13)$$

où les  $\lambda_i$  sont les valeurs propres de  $A$ . La norme sup de  $U \in \mathbb{R}^{N+1}$  est

$$\|U\|_\infty = \max_{0 \leq i \leq N} |U_i|. \quad (1.14)$$

La norme induite sur  $\mathbb{M}_{N+1}(\mathbb{R})$  est

$$\|A\|_\infty = \sup_{U \neq 0} \frac{|AU|_\infty}{|U|_\infty} = \max_{1 \leq i \leq N+1} \sum_{j=1}^{N+1} |A_{i,j}|. \quad (1.15)$$

La distinction entre fonctions de grille et vecteurs a plusieurs avantages. D'abord, elle permet de faire une distinction claire entre les deux classes d'opérateurs :

- *Opérateurs aux différences finies*, qui agissent naturellement sur des fonctions de grille.
- *Matrices* qui agissent naturellement sur des vecteurs.

Les fonctions de grille contiennent toutes les échelles nécessaires dans le contexte physique, alors que les vecteurs sont sans dimension. De plus, le raisonnement au niveau discret est plus naturel avec les fonctions de grilles. Il s'effectue d'une façon abstraite à l'aide d'opérateurs aux différences. Par contre le codage effectif est effectué dans le cadre de l'algèbre linéaire.

## 1.2 Opérateurs aux différences

On introduit quelques opérateurs aux différences usuels qui agissent sur des fonctions de grille.

- *Dérivée centrée*

L'opérateur  $\delta_x$  est défini par

$$\delta_x u_j = \frac{u_{j+1} - u_{j-1}}{2h}, \quad 1 \leq j \leq N-1. \quad (1.16)$$

On le prolonge en  $j = 0, j = N$  par  $\delta_x u_0 = \delta_x u_N = 0$ .

• *Dérivées décentrées*

Les opérateurs  $\delta_x^+$  et  $\delta_x^-$  sont définis par

$$\begin{cases} \delta_x^+ u_j = \frac{u_{j+1} - u_j}{h}, & 0 \leq j \leq N-1, \\ \delta_x^- u_j = \frac{u_j - u_{j-1}}{h}, & 1 \leq j \leq N. \end{cases} \quad (1.17)$$

On les prolonge en  $j = 0, j = N$  par  $\delta_x^+ u_0 = 0, \delta_x^- u_N = 0$ .

• *Laplacien à trois points*

Le laplacien à trois points est

$$\delta_x^2 u_j = \frac{u_{j+1} + u_{j-1} - 2u_j}{h^2}, \quad 1 \leq j \leq N-1. \quad (1.18)$$

On le prolonge en  $j = 0, j = N$  par  $\delta_x^2 u_0 = \delta_x^2 u_N = 0$ .

**Proposition 1.1.** Soit  $u(x) \in C^4(\Omega)$  et  $u^*$  sa fonction de grille correspondante. On a

$$-\delta_x^2 u_i^* + u''(x_i) = -\frac{h^2}{4!} \left( u^{(4)}(\eta_i) + u^{(4)}(\varsigma_i) \right), \quad \eta_i, \varsigma_i \in (x_{i-1}, x_{i+1}). \quad (1.19)$$

**Preuve.** On note  $t \in l_{h,0}^2$  la fonction de grille “troncature”,

$$t(x_i) = -\delta_x^2 u_i^* + u''(x_i), \quad i = 1, \dots, N. \quad (1.20)$$

On estime  $t(x_i)$  à l'aide de la formule de Taylor-Lagrange de la façon suivante. En développant  $u(x_i + h)$  et  $u(x_i - h)$ ,  $1 \leq i \leq N-1$ , on obtient

$$u(x_i + h) = u(x_i) + hu'(x_i) + \frac{h^2}{2!}u''(x_i) + \frac{h^3}{3!}u^{(3)}(x_i) + \frac{h^4}{4!}u^{(4)}(\eta_i), \quad (1.21)$$

et

$$u(x_i - h) = u(x_i) - hu'(x_i) + \frac{h^2}{2!}u''(x_i) - \frac{h^3}{3!}u^{(3)}(x_i) + \frac{h^4}{4!}u^{(4)}(\varsigma_i), \quad (1.22)$$



avec  $\eta_i \in (x_i, x_i + h)$  et  $\varsigma_i \in (x_i - h, x_i)$ . Donc,

$$\begin{aligned} t(x_i) &= -\frac{u(x+h) + u(x-h) - 2u(x)}{h^2} + u''(x_i) \\ &= -\frac{h^2}{4!} \left( u^{(4)}(\eta_i) + u^{(4)}(\varsigma_i) \right), \quad 1 \leq i \leq N-1. \end{aligned}$$

■

• *Opérateur de Simpson  $\sigma_x$*

L'opérateur  $\sigma_x$  est défini par

$$\sigma_x u_j = \frac{1}{6}u_{j-1} + \frac{2}{3}u_j + \frac{1}{6}u_{j+1}, \quad 1 \leq j \leq N-1. \quad (1.23)$$

On le prolonge au bord par  $\sigma_x u_0 = \sigma_x u_N = 0$ .

**Proposition 1.2.** *On suppose ici que  $u(x)$  est au moins de classe  $C^4$ . Sur un intervalle  $[x_i - h, x_i + h]$ , l'erreur  $e_i(h)$  commise par l'approximation de Simpson définie par*

$$e_i(h) = \frac{1}{6}u(x_i - h) + \frac{2}{3}u(x_i) + \frac{1}{6}u(x_i + h) - \frac{1}{2h} \int_{x_i-h}^{x_i+h} u(x) dx, \quad (1.24)$$

est tel que

$$e_i(h) = \mathcal{O}(h^4). \quad (1.25)$$

**Preuve.** En utilisant la formule de Taylor-Lagrange on obtient

$$\begin{cases} u(x_i + h) = u(x_i) + hu'(x_i) + \frac{h^2}{2}u''(x_i) + \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\varsigma_i), & \varsigma_i \in (x_i, x_i + h), \\ u(x_i - h) = u(x_i) - hu'(x_i) + \frac{h^2}{2}u''(x_i) - \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\eta_i), & \eta_i \in (x_i - h, x_i), \end{cases} \quad (1.26)$$

donc

$$\frac{1}{6}u(x_{i-1}) + \frac{2}{3}u(x_i) + \frac{1}{6}u(x_{i+1}) = u(x_i) + \frac{h^2}{6}u^{(2)}(x_i) + \frac{h^4}{144}[u^{(4)}(\eta_i) + u^{(4)}(\varsigma_i)]. \quad (1.27)$$

Pour  $x \in [x_{i-1}, x_{i+1}]$  en utilisant à nouveau la formule de Taylor-Lagrange on obtient

$$u(x) = u(x_i) + (x - x_i)u'(x_i) + \frac{(x - x_i)^2}{2}u^{(2)}(x_i) + \frac{(x - x_i)^3}{6}u^{(3)}(x_i) + \dots \quad (1.28)$$

En intégrant (1.28) sur  $[x_{i-1}, x_{i+1}]$  on obtient

$$\begin{aligned} \int_{x_{i-1}}^{x_{i+1}} u(x) dx &= \int_{x_{i-1}}^{x_{i+1}} \left( u(x_i) + (x - x_i)u'(x_i) + \frac{(x - x_i)^2}{2}u^{(2)}(x_i) + \frac{(x - x_i)^3}{6}u^{(3)}(x_i) + \dots \right) \\ &= 2hu(x_i) + \frac{h^3}{3}u^{(2)}(x_i) + \mathcal{O}(h^5), \end{aligned}$$

donc

$$\frac{1}{2h} \int_{x_{i-1}}^{x_{i+1}} u(x) dx = u(x_i) + \frac{h^2}{6}u^{(2)}(x_i) + \mathcal{O}(h^4). \quad (1.29)$$

Il résulte de (1.27) et (1.29) que

$$\frac{1}{6}u(x_{i-1}) + \frac{2}{3}u(x_i) + \frac{1}{6}u(x_{i+1}) - \frac{1}{2h} \int_{x_{i-1}}^{x_{i+1}} u(x) dx = \mathcal{O}(h^4), \quad (1.30)$$

d'où le résultat annoncé. ■

### 1.3 Notations vectorielles et matricielles

Les notations vectorielles et matricielles utilisées sont les suivantes :

- *Notations vectorielles*

On considère les fonctions de grille  $u = [u_0, u_1, \dots, u_N]^T$ . On note par  $U \in \mathbb{R}^{(N-1)}$  le vecteur correspondant aux points intérieurs

$$U = [u_1, u_2, \dots, u_{N-1}]^T, \quad (1.31)$$

et par  $\tilde{U} \in \mathbb{R}^{(N+1)}$  le vecteur correspondant aux points intérieurs et aux points de bord

$$\tilde{U} = [u_0, u_1, u_2, \dots, u_N]^T. \quad (1.32)$$

On note  $(e_j)_{1 \leq j \leq N-1}$  la base canonique de  $\mathbb{R}^{N-1}$ ,

$$e_j = [\delta_{1j}, \delta_{2j}, \dots, \delta_{Nj}]^T, \quad (1.33)$$

où  $\delta_{ij}$  désigne le symbole de Kronecker :

$$\delta_{ij} = \begin{cases} 1, & \text{si } i = j, \\ 0, & \text{si } i \neq j. \end{cases} \quad (1.34)$$

Le produit scalaire canonique de  $\mathbb{R}^{N-1}$  est noté

$$\langle x, y \rangle = \sum_{i=1}^{N-1} x_i y_i, \quad x, y \in \mathbb{R}^{N-1}. \quad (1.35)$$

• Opérateur  $\text{vec}_1$

**Définition 1.** Soient  $v = (v_i)_{0 \leq i \leq N} \in l_h^2$  une fonction de grille et  $(e_i)_{1 \leq i \leq N-1}$  la base canonique de  $\mathbb{R}^{(N-1)}$  définie en (1.33). On définit l'opérateur  $\text{vec}_1$  par

$$\begin{aligned} \text{vec}_1 : L_h^2 &\longrightarrow \mathbb{R}^{(N-1)} \\ v &\longrightarrow V = \text{vec}_1(v) \end{aligned}$$

avec

$$\text{vec}_1(v) = \sum_{i=1}^{N-1} e_i v_i. \quad (1.36)$$

Matriciellement, l'opérateur  $\text{vec}_1$  crée à partir d'une fonction de grille  $v = (v_i)_{0 \leq i \leq N}$ , un vecteur de taille  $(N-1)$  formé des points intérieurs de  $v$  (c'est-à-dire les points  $(v_i)_{1 \leq i \leq N-1}$ ). On peut écrire

$$V = \text{vec}_1(v) = \text{vec}_1[v_i]_{1 \leq i \leq N-1} = [v_1, v_2, \dots, v_{N-1}]^T \in \mathbb{R}^{(N-1)}. \quad (1.37)$$

• Matrice de l'opérateur aux différences centrées

Pour  $u \in l_{h,0}^2$  on peut aisément voir que

$$\delta_x u_j = \frac{1}{2h} (Ku)_j, \quad (1.38)$$

où  $K = (K_{i,m})_{1 \leq i,m \leq N-1}$  est la matrice antisymétrique donnée par

$$K_{i,m} = \begin{cases} \text{sign}(m-i), & \text{si } |m-i| = 1, \\ 0, & \text{si } |m-i| \neq 1, \end{cases} \quad (1.39)$$

ou encore

$$K = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & -1 & 0 & 1 \\ 0 & \dots & 0 & -1 & 0 \end{bmatrix} \quad (1.40)$$

- *Matrice du Laplacien centré*

En agissant sur les fonctions de grille  $u \in l_{h,0}^2$ , la matrice  $T$  correspondant à l'opérateur de Laplace à trois points,  $-h^2\delta_x^2$ , est la matrice symétrique définie positive définie par

$$T_{i,m} = \begin{cases} 2, & \text{si } |m - i| = 0, \\ -1, & \text{si } |m - i| = 1, \\ 0, & \text{si } |m - i| \geq 2, \end{cases} \quad (1.41)$$

c'est-à-dire

$$T = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix} \in \mathbb{M}_{N-1}(\mathbb{R}). \quad (1.42)$$

Les vecteurs propres et les valeurs propres de la matrice  $T$  jouent un rôle important dans la suite.

**Proposition 1.3.** *Les vecteurs propres  $(Z^k)_{1 \leq k \leq N-1}$  et les valeurs propres  $(\lambda_k)_{1 \leq k \leq N-1}$  de la matrice  $T$  sont*

$$\begin{cases} Z_j^k = \left(\frac{2}{N}\right)^{1/2} \sin\left(\frac{kj\pi}{N}\right), & 1 \leq j \leq N-1, & (a) \\ \lambda_k = 4 \sin^2\left(\frac{k\pi}{2N}\right). & & (b) \end{cases} \quad (1.43)$$

**Preuve.** On considère le problème

$$\begin{cases} -\delta_x^2 z_j^k - \lambda_k z_j^k = 0, & 1 \leq j \leq N-1, & (a) \\ z_0^k = z_N^k = 0, & & (b) \end{cases} \quad (1.44)$$

où  $z_j^k$  représente la  $j$ -ème composante du vecteur propre  $z^k$ ,  $k = 1, 2, \dots, N-1$ , et  $\delta_x^2$  est donné en (1.18). L'équation (1.44) s'écrit

$$z_{j-1}^k - 2 \cos \theta_k z_j^k + z_{j+1}^k = 0, \quad (1.45)$$

où  $\theta_k$  est défini en fonction de  $\lambda_k$  par

$$\lambda_k = -2 + 2 \cos \theta_k. \quad (1.46)$$

La solution générale de (1.45) est (récurrence à 3 points)

$$z_j^k = c_1 \cos(j\theta_k) + c_2 \sin(j\theta_k), \quad (1.47)$$

où  $c_1$  et  $c_2$  sont des constantes arbitraires. La condition (1.44)<sub>b</sub> entraîne  $c_1 = 0$  et  $\theta_k = \pi \frac{k}{N}$ , de sorte que

$$\begin{cases} z^k = \left[ \sin \frac{\pi k}{N}, \sin \frac{2\pi k}{N}, \dots, \sin \frac{(N-1)\pi k}{N} \right]^T, \\ \lambda_k = 4 \sin^2 \left( \frac{k\pi}{2N} \right). \end{cases} \quad (1.48)$$

On en déduit (1.43)<sub>b</sub>. On vérifie que

$$\langle z^k, z^l \rangle = \frac{N}{2} \delta_{kl}, \quad (1.49)$$

$\delta_{ki}$  est le symbole de kronecker (1.34). En notant  $Z^k \in \mathbb{R}^{N-1}$  le vecteur de composantes

$$Z_j^k = \left( \frac{2}{N} \right)^{1/2} z_j^k, \quad 1 \leq k \leq N-1, \quad (1.50)$$

et on a

$$\langle Z^k, Z^l \rangle = \delta_{kl}, \quad (1.51)$$

on obtient (1.43)<sub>a</sub>. ■

D'après (1.51), les vecteurs  $(Z^k)_{1 \leq k \leq N-1}$  forment une base orthonormale de  $\mathbb{R}^{N-1}$ . Tout vecteur  $V \in \mathbb{R}^{N-1}$  se décompose donc sur  $(Z^k)_{1 \leq k \leq N-1}$  sous la forme

$$V = \sum_{k=1}^{N-1} v_k Z^k, \quad (1.52)$$

avec

$$v_k = \langle v, Z^k \rangle, \quad k = 1, 2, \dots, N-1. \quad (1.53)$$

On introduit la matrice de passage de la base  $(e_k)_{1 \leq k \leq N-1}$  à la base des vecteurs propres  $(Z^k)_{1 \leq k \leq N-1}$ , notée  $Z$ . Cette matrice est symétrique orthonormale car toutes ses colonnes sont orthonormales d'après (1.49) donc on a

$$Z^{-1} = Z^T = Z. \quad (1.54)$$

On a

$$T = Z \Lambda Z, \quad (1.55)$$

où  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{N-1})$  est la matrice diagonale formée des valeurs propres de  $T$  données en (1.43)<sub>b</sub>.

**Corollaire 1.3.1.** *La matrice  $T$  vérifie*

$$\begin{cases} \|T\|_2 = 4 + \mathcal{O}\left(\frac{1}{N^2}\right), & (a) \\ \|T^{-1}\|_2 = \frac{\pi^2}{N^2} + \mathcal{O}\left(\frac{1}{N^4}\right). & (b) \end{cases} \quad (1.56)$$

**Preuve.** La matrice  $T$  est symétrique donc

$$\|T\|_2 = \max_{1 \leq k \leq N-1} \lambda_k = \sin^2\left(\frac{(N-1)\pi}{2N}\right). \quad (1.57)$$

On a

$$\sin^2\left(\frac{(N-1)\pi}{2N}\right) = 1 - \cos^2\left(\frac{(N-1)\pi}{2N}\right). \quad (1.58)$$

Or

$$\cos\left(\frac{(N-1)\pi}{2N}\right) = \cos\left(\frac{\pi}{2} - \frac{\pi}{2N}\right) = \sin\left(\frac{\pi}{2N}\right), \quad (1.59)$$

on a  $\sin(x) = x - \frac{x^3}{6} + \dots$ , ce qui donne

$$x - \frac{x^3}{6} \leq \sin x \leq x. \quad (1.60)$$

Donc,

$$1 - \left(\frac{\pi}{2N}\right)^2 \leq \sin^2\left(\frac{(N-1)\pi}{2N}\right) \leq 1 - \left(\frac{\pi}{2N} - \frac{1}{6}\left(\frac{\pi}{2N}\right)^3\right)^2. \quad (1.61)$$

On déduit de (1.61)

$$4 \sin^2\left(\frac{(N-1)\pi}{2N}\right) = 4 - \frac{\pi^2}{N^2} + \mathcal{O}\left(\frac{1}{N^4}\right), \quad (1.62)$$

alors

$$\|T\|_2 = 4 + \mathcal{O}\left(\frac{1}{N^2}\right). \quad (1.63)$$

d'où (1.56)<sub>a</sub>. En plus,

$$\|T^{-1}\|_2 = \min_{1 \leq k \leq N-1} \lambda_k = 4 \sin^2\left(\frac{\pi}{2N}\right) = \frac{\pi^2}{N^2} + \mathcal{O}\left(\frac{1}{N^4}\right), \quad (1.64)$$

d'où (1.56)<sub>b</sub>. ■

• La matrice symétrique définie positive  $P_s$  correspond à l'opérateur  $\sigma_x$  défini en (1.23)

est déduite de  $T$  par

$$P_s = I - \frac{1}{6}T. \quad (1.65)$$

Donc

$$(P_s)_{i,j} = \begin{cases} \frac{2}{3}, & i = j, \\ \frac{1}{6}, & |i - j| = 1, \\ 0, & |i - j| \geq 2, \end{cases} \quad (1.66)$$

soit

$$P_s = \begin{bmatrix} \frac{2}{3} & \frac{1}{6} & 0 & \dots & 0 \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ 0 & \dots & 0 & \frac{1}{6} & \frac{2}{3} \end{bmatrix} \in \mathbb{M}_{N-1}(\mathbb{R}). \quad (1.67)$$

**Lemme 1.3.2.** La matrice  $P_s$  vérifie les propriétés suivantes

$$\begin{cases} \|P_s\|_2 = 1 - \frac{\pi^2}{6N^2} + \mathcal{O}\left(\frac{1}{N^4}\right), & (a) \\ \|P_s^{-1}\|_2 = 3 + \mathcal{O}\left(\frac{1}{N^4}\right). & (b) \end{cases} \quad (1.68)$$

**Preuve.** La matrice  $P_s$  est symétrique alors

$$\begin{aligned} \|P_s\|_2 &= \max_{1 \leq k \leq N-1} \lambda_k(P_s) \\ &= 1 - \frac{1}{6} \lambda_{\min}(T) \\ &= 1 - \frac{\pi^2}{6N^2} + \mathcal{O}\left(\frac{1}{N^4}\right), \end{aligned}$$

d'où (1.68)<sub>a</sub>. De la même manière on démontre (1.68)<sub>b</sub>. ■

• Les deux matrices  $F_1$  et  $F_2 \in \mathbb{M}_{N-1}(\mathbb{R})$  suivantes sont utilisées dans la suite

$$F_1 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix}, \quad F_2 = \begin{bmatrix} -1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix}. \quad (1.69)$$

On a

$$\begin{cases} F_1 = e_1 e_1^T + e_{N-1} e_{N-1}^T = [e_1, e_{N-1}] \begin{bmatrix} e_1^T \\ e_{N-1}^T \end{bmatrix}, & (a) \\ F_2 = -e_1 e_1^T + e_{N-1} e_{N-1}^T = [-e_1, e_{N-1}] \begin{bmatrix} e_1^T \\ e_{N-1}^T \end{bmatrix}. & (b) \end{cases} \quad (1.70)$$

**Lemme 1.3.3.** *Les matrices  $F_1$  et  $F_2$  vérifient les relations*

$$\begin{cases} F_2 \cdot F_1 = F_2, & (a) \\ F_2 \cdot F_2 = F_1. & (b) \end{cases} \quad (1.71)$$

**Preuve.**

$$\begin{aligned} F_2 \cdot F_1 &= (-e_1 e_1^T + e_{N-1} e_{N-1}^T)(e_1 e_1^T + e_{N-1} e_{N-1}^T) \\ &= -e_1 e_1^T e_1 e_1^T - e_1 e_1^T e_{N-1} e_{N-1}^T + e_{N-1} e_{N-1}^T e_1 e_1^T + e_{N-1} e_{N-1}^T e_{N-1} e_{N-1}^T \\ &= -e_1 e_1^T + 0 + 0 + e_{N-1} e_{N-1}^T \\ &= F_2, \end{aligned}$$

d'où (1.71)<sub>a</sub>.

$$\begin{aligned} F_2 \cdot F_2 &= (-e_1 e_1^T + e_{N-1} e_{N-1}^T)(-e_1 e_1^T + e_{N-1} e_{N-1}^T) \\ &= e_1 e_1^T e_1 e_1^T - e_1 e_1^T e_{N-1} e_{N-1}^T - e_{N-1} e_{N-1}^T e_1 e_1^T + e_{N-1} e_{N-1}^T e_{N-1} e_{N-1}^T \\ &= e_1 e_1^T + 0 + 0 + e_{N-1} e_{N-1}^T \\ &= F_1, \end{aligned}$$

d'où (1.71)<sub>b</sub>. ■



## Chapitre 2

# Schémas boîte hermitiens en dimension 1

### 2.1 Le schéma boîte de Keller

La notion de schéma boîte a été introduite par H.B. Keller dans [25]. Il s'agit à l'origine d'une méthode numérique pour l'approximation de problèmes monodimensionnels de diffusion. Le schéma de Keller s'apparente d'une part à la méthode des différences finies car les inconnues discrètes sont situées aux noeuds du maillage. Elle s'apparente d'autre part à la méthode des volumes finis car la discrétisation s'opère en moyennant les équations sur les cellules du maillage.

Dans le cas des schémas boîte, le maillage est utilisé à la fois pour localiser les inconnues au niveau des noeuds du maillage et pour localiser les volumes de contrôle sur lesquels on effectue les moyennes des équations. L'inconvénient est que le nombre de noeuds peut être différent du nombre de volumes en général : a priori, on n'a pas le bon décompte entre nombre d'inconnues et nombre d'équations. Ceci n'est le cas ni pour la méthode de différences finies, où la discrétisation se fait au niveau des noeuds, ni pour la méthode de volumes finies, où les inconnues sont situées au niveau des volumes de contrôle. Cependant pour des opérateurs elliptiques, H.B. Keller [25] a remarqué qu'en considérant une *formulation mixte*, on obtient un décompte correct entre nombre d'équations et nombre d'inconnues. L'illustration la plus simple est le problème elliptique :

$$\begin{cases} -u'' = f, & x \in (0, 1), \\ u(0) = u(1) = 0. \end{cases} \quad (2.1)$$

La formulation mixte de (2.1) s'obtient en introduisant  $v(x) = u'(x)$

$$\begin{cases} -v' = f, & x \in (0, 1), & (a) \\ v = u', & x \in [0, 1], & (b) \\ u(0) = u(1) = 0. & & (c) \end{cases} \quad (2.2)$$

En intégrant les équations (2.2)<sub>a,b</sub> sur des boîtes  $\mathcal{K}_j = [x_{j-1}, x_j]$  de longueur  $h$  (figure 2.1) on obtient

$$\begin{cases} -\frac{v_j - v_{j-1}}{h} = \bar{\Pi} f_{j-1/2}, & 1 \leq j \leq N, \\ \bar{\Pi} v_{j-1/2} = \frac{u_j - u_{j-1}}{h}, & 1 \leq j \leq N, \\ u_0 = u_N = 0, \end{cases} \quad (2.3)$$

où  $\bar{\Pi} v_{j-1/2} = \frac{1}{h} \int_{x_{j-1}}^{x_j} v(x) dx$ . Dans (2.3) on a  $2(N+1)$  équations et également  $2(N+1)$  inconnues. Donc

$$\text{nombre d'équations} = \text{nombre d'inconnues}. \quad (2.4)$$

C'est-à-dire  $2 \times (N)$  boîtes + 2 C.L. =  $2N$  inconnues + 2 inconnues.

Ici, nous introduisons un schéma du type précédent, mais qui utilise des “boîtes recouvrantes”  $K_j = [x_{j-1}, x_{j+1}]$  de longueur  $2h$  (figure 2.2) au lieu des boîtes de longueur  $h$ . Ceci permet d'obtenir une précision locale d'ordre 4. Le prix à payer est la nécessité de définir une approximation de  $u'(0)$  et  $u'(1)$ , qui ne peut être déduite simplement de (2.3).

Dans la section 2.2, nous donnons une première version du schéma qui correspond à

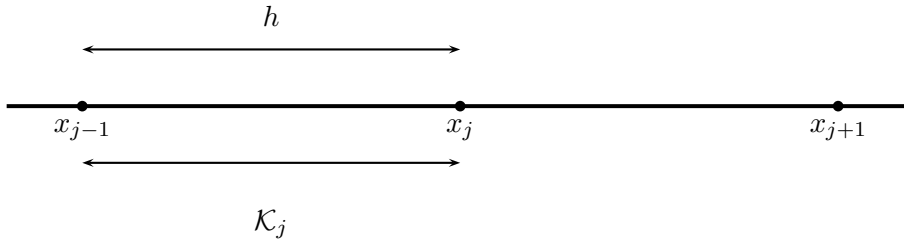


FIG. 2.1: Le stencil du schéma boîte de Keller :  $\mathcal{K}_j = [x_{j-1}, x_j]$

un choix simple de l'approximation de  $u'(0)$  et  $u'(1)$ . On montre que le schéma obtenu de cette approximation possède une matrice diagonale dans la base spectrale de  $-\delta_x^2$  donnée en (1.43)<sub>a</sub>. De plus ce schéma vérifie un principe du maximum et du minimum discret. On termine cette section en donnant quelques tests numériques prouvant l'ordre 2 de convergence du schéma pour les inconnues  $u_j \simeq u(x_j)$  et  $u_{x,j} \simeq u'(x_j)$ . Dans la section 2.3, nous introduisons une deuxième version qui correspond à un choix plus

précis de  $u'(0)$  et  $u'(1)$ . La matrice de ce schéma est une perturbation de rang petit de la matrice du premier schéma. On donne une forme matricielle générale du schéma boîte hermitien en dimension 1. Après avoir étudié quelques propriétés de ce schéma, nous donnons quelques résultats numériques prouvant l'intérêt du schéma. Des courbes de convergence comparent plusieurs schémas numériques entre eux.

## 2.2 Schéma boîte hermitien HB1

### 2.2.1 Design du schéma HB1

On considère le problème elliptique

$$\begin{cases} -u''(x) = f(x), & x \in (a, b), \\ u(a) = g_a, u(b) = g_b. \end{cases} \quad (2.5)$$

Ce problème est bien posé en  $H_0^1(a, b)$  avec donnée  $f \in L^2(a, b)$ , ceci implique que  $u(x)$  est continue et que  $u'(x) \in L^2(a, b)$ . En utilisant la variable auxiliaire  $v = u'(x)$ , (2.5) peut être réécrite sous forme mixte :

$$\begin{cases} v'(x) + f(x) = 0, & x \in (a, b), & (a) \\ v(x) - u'(x) = 0, & x \in [a, b], & (b) \\ u(a) = g_a, u(b) = g_b. & & (c) \end{cases} \quad (2.6)$$

On appelle “schéma boîte hermitien” (HB en abrégé), le schéma obtenu de la façon suivante. On intègre les équations (2.6)<sub>a,b</sub> sur des “boîtes” constitués des cellules  $K_j = [x_{j-1}, x_{j+1}]$ , de longueur  $2h$  (figure 2.2). A la différence du schéma de Keller, les “boîtes”

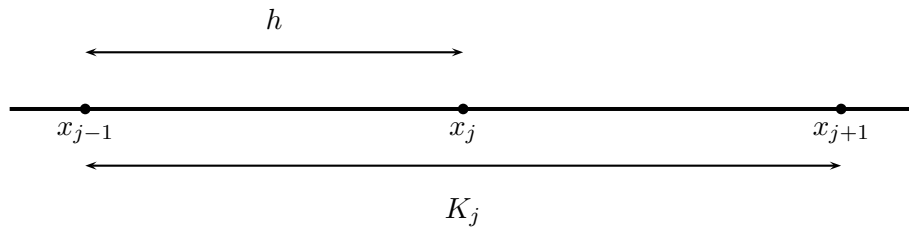


FIG. 2.2: Le stencil du schéma boîte hermitien :  $K_j = [x_{j-1}, x_{j+1}]$

$K_j, K_{j+1}$  se recouvrent,  $K_j \cap K_{j+1} = [x_j, x_{j+1}]$ . On suppose que  $u(x)$  est une solution exacte  $C^1[a, b]$  de (2.5) et  $v(x) := u'(x) \in C^0[a, b]$  est sa dérivée. En intégrant l'équation

conservative (2.6)<sub>a</sub> sur les boîtes  $K_j$ ,  $1 \leq j \leq N - 1$ , on obtient les identités

$$-\frac{v(x_{j+1}) - v(x_{j-1}))}{2h} = \Pi^0 f_j, \quad 1 \leq j \leq N - 1, \quad (2.7)$$

où  $\Pi^0$  est l'opérateur de moyenne sur  $K_j = [x_{j-1}, x_{j+1}]$ , défini par

$$\Pi^0 v_j = \frac{1}{2h} \int_{x_{j-1}}^{x_{j+1}} v(x) dx. \quad (2.8)$$

De même, l'identité (2.6)<sub>b</sub> est intégrée sur les boîtes  $K_j$ ,

$$\frac{1}{2h} \int_{x_{j-1}}^{x_{j+1}} u'(x) dx = \frac{u(x_{j+1}) - u(x_{j-1}))}{2h}, \quad 1 \leq j \leq N - 1. \quad (2.9)$$

Donc la solution exacte  $u(x)$  et sa dérivée  $v(x) = u'(x)$  vérifient les  $2N$  identités,

$$\begin{cases} v(x_{j+1}) - v(x_{j-1}) = -2h(\Pi^0 f)_j, & 1 \leq j \leq N - 1, & (a) \\ 2h(\Pi^0 v)_j - [u(x_{j+1}) - u(x_{j-1})] = 0, & 1 \leq j \leq N - 1, & (b) \\ u(x_0) = g_a, u(x_N) = g_b. & & (c) \end{cases} \quad (2.10)$$

Pour avoir le bon décompte entre le nombre d'inconnues et le nombre d'équations du système (2.10), on ajoute les deux équations  $v_0 = u'(a), v_N = u'(b)$ . La solution exacte  $u(x)$  du problème continu (2.5) et sa dérivée  $v(x)$  satisfont le système

$$\begin{cases} v(x_{j+1}) - v(x_{j-1}) = -2h\Pi^0 f_j, & 1 \leq j \leq N - 1, & (a) \\ 2h(\Pi^0 v)_j - [u(x_{j+1}) - u(x_{j-1})] = 0, & 1 \leq j \leq N - 1, & (b) \\ v(x_0) = u'(x_0), v(x_N) = u'(x_N), & & (c) \\ u(x_0) = g_a, u(x_N) = g_b. & & (d) \end{cases} \quad (2.11)$$

Le système (2.11) constitue un ensemble de  $2(N + 1)$  identités vérifiées par la solution exacte de (2.6). On considère les inconnues  $(u_j, u_{x,j})$ ,  $0 \leq j \leq N$ , où  $u_j \simeq u(x_j)$ ,  $u_{x,j} \simeq v(x_j)(= u'(x_j))$ . Le système (2.11) consiste en :

- (2.11)<sub>a</sub> donne les  $(N - 1)$  relations

$$u_{x,j+1} - u_{x,j-1} = -2h\Pi^0 f_j, \quad 1 \leq j \leq N - 1. \quad (2.12)$$

Aucune approximation n'a encore été effectuée à ce stade. Noter que  $\Pi^0 f_j$  doit être approché par une formule de quadrature convenable.

• Dans (2.11)<sub>b</sub>, pour  $1 \leq j \leq N - 1$ , la moyenne de  $v(x)$  sur  $K_j = [x_{j-1}, x_{j+1}]$  est

$$\Pi^0 v_j = \frac{1}{2h} \int_{x_{j-1}}^{x_{j+1}} v(t) dt. \quad (2.13)$$

On effectue l'approximation de  $\Pi^0 v_j$  par la formule de Simpson (1.23),

$$\Pi^0 v_j \simeq \sigma_x v_j = \frac{1}{6} v_{j-1} + \frac{2}{3} v_j + \frac{1}{6} v_{j+1}, \quad 1 \leq j \leq N - 1, \quad (2.14)$$

alors l'équation (2.11)<sub>b</sub>, au niveau discret, devient

$$\frac{1}{6} u_{x,j-1} + \frac{2}{3} u_{x,j} + \frac{1}{6} u_{x,j+1} = \frac{u_{j+1} - u_{j-1}}{2h}, \quad 1 \leq j \leq N - 1. \quad (2.15)$$

Les équations (2.12) et (2.15) s'expriment sous la forme

$$\begin{cases} -\frac{u_{x,j+1} - u_{x,j-1}}{2h} = \Pi^0 f_j, & 1 \leq j \leq N - 1, \quad (a) \\ \frac{1}{6} u_{x,j-1} + \frac{2}{3} u_{x,j} + \frac{1}{6} u_{x,j+1} = \frac{u_{j+1} - u_{j-1}}{2h}, & 1 \leq j \leq N - 1. \quad (b) \end{cases} \quad (2.16)$$

Utilisant les notations différences finies, (2.16) se réécrit

$$\begin{cases} -\delta_x u_{x,j} = \Pi^0 f_j, & 1 \leq j \leq N - 1, \quad (a) \\ \sigma_x u_{x,j} = \delta_x u_j, & 1 \leq j \leq N - 1. \quad (b) \end{cases} \quad (2.17)$$

Le système (2.17) constitue les deux identités de base du schéma boîte hermitien "HB".

On considère à présent la question de l'élimination de  $u_x$  dans (2.17). En pratique on approchera  $\Pi^0 f_j$  par la formule du Simpson, c'est-à-dire,

$$\Pi^0 f_j \simeq \sigma_x f_j = \frac{1}{6} f_{j-1} + \frac{2}{3} f_j + \frac{1}{6} f_{j+1}, \quad 1 \leq j \leq N - 1. \quad (2.18)$$

**Proposition 2.1.** *Supposons que  $(u_j)_{j \in \mathbb{Z}}$ ,  $(u_{x,j})_{j \in \mathbb{Z}}$  soient deux suites satisfaisant*

$$\begin{cases} -\delta_x u_{x,j} = \Pi^0 f_j, & j \in \mathbb{Z}, \quad (a) \\ \sigma_x u_{x,j} = \delta_x u_j, & j \in \mathbb{Z}. \quad (b) \end{cases} \quad (2.19)$$

Alors  $(u_j)_{j \in \mathbb{Z}}$  vérifie

$$-\frac{u_{j+2} + u_{j-2} - 2u_j}{4h^2} = \frac{1}{36} f_{j-2} + \frac{2}{9} f_{j-1} + \frac{1}{2} f_j + \frac{2}{9} f_{j+1} + \frac{1}{36} f_{j+2}. \quad (2.20)$$

**Preuve.** Notons par  $E_j$ ,  $j \in \mathbb{Z}$ , la  $j$ -ème équation du (2.19)<sub>a</sub>. En faisant la somme

$$\frac{1}{6}E_{j-1} + \frac{2}{3}E_j + \frac{1}{6}E_{j+1} = \frac{1}{6}\sigma_x f_{j-1} + \frac{2}{3}\sigma_x f_j + \frac{1}{6}\sigma_x f_{j+1}, \quad (2.21)$$

et en utilisant (2.19)<sub>b</sub>, on trouve que le terme de gauche du (2.21) vérifie

$$\begin{cases} \frac{1}{6}E_{j-1} + \frac{2}{3}E_j + \frac{1}{6}E_{j+1} &= -\frac{1}{6}\delta_x u_{x,j-1} - \frac{2}{3}\delta_x u_{x,j} - \frac{1}{6}\delta_x u_{x,j+1} \\ &= -\frac{u_{j+2} + u_{j-2} - 2u_j}{4h^2}. \end{cases} \quad (2.22)$$

Le terme de droite de (2.21) vérifie

$$\begin{cases} \frac{1}{6}\sigma_x f_{j-1} + \frac{2}{3}\sigma_x f_j + \frac{1}{6}\sigma_x f_{j+1} &= \frac{1}{6}\left(\frac{1}{6}f_{j-2} + \frac{2}{3}f_{j-1} + \frac{1}{6}f_j\right) \\ &+ \frac{2}{3}\left(\frac{1}{6}f_{j-1} + \frac{2}{3}f_j + \frac{1}{6}f_{j+1}\right) \\ &+ \frac{1}{6}\left(\frac{1}{6}f_j + \frac{2}{3}f_{j+1} + \frac{1}{6}f_{j+2}\right) \\ &= \frac{1}{36}f_{j-2} + \frac{2}{9}f_{j-1} + \frac{1}{2}f_j + \frac{2}{9}f_{j+1} + \frac{1}{36}f_{j+2}, \end{cases}$$

d'où l'égalité (2.20) pour  $j \in \mathbb{Z}$ . ■

En revanche, pour  $0 \leq j \leq N$ , le système (2.17) vérifie la relation (2.20) seulement pour  $2 \leq j \leq N-2$ . Il consiste en  $2(N-1)$  équations pour  $2N$  inconnues. Les identités (2.17)<sub>a,b</sub> nécessitent donc une approximation de la dérivée  $v(x)$  aux points de bord  $x_0$  et  $x_N$ . Evidemment, on ne connaît pas les dérivées au bord et donc les équations (2.11)<sub>c</sub> seront remplacées par des approximations des dérivées au bord que l'on décrit à présent.

• *Première approximation des dérivées au bord  $u_{x,0}, u_{x,N}$*

Supposons que le prolongement par imparité aux points du bord de la solution exacte  $u(x)$  soit une fonction régulière. Alors la dérivée  $v(x) = u'(x)$  est paire. Ceci donne que les points  $v_{x,-1}, v_{x,N+1}$  qui sont à l'extérieur du domaine  $[a, b] = [x_0, x_N]$  vérifient

$$\begin{cases} v_{-1} = v_1, \\ v_{N+1} = v_{N-1}. \end{cases} \quad (2.23)$$

On peut prolonger  $\sigma_x$  (1.23) au bord par

$$\begin{cases} \sigma_x v_0 = \frac{1}{6}v_{-1} + \frac{2}{3}v_0 + \frac{1}{6}v_1 = \frac{2}{3}v_0 + \frac{1}{3}v_1, & (a) \\ \sigma_x v_N = \frac{1}{6}v_{N-1} + \frac{2}{3}v_N + \frac{1}{6}v_{N+1} = \frac{2}{3}v_N + \frac{1}{3}v_{N-1}. & (b) \end{cases} \quad (2.24)$$

Or

$$\left\{ \begin{array}{l} \sigma_x v_0 \simeq \frac{1}{2h} \int_{x_{-1}}^{x_1} v(x) dx = \frac{1}{h} \int_{x_0}^{x_1} u'(x) dx = \frac{u_1 - u_0}{h}, \quad (a) \\ \sigma_x v_N \simeq \frac{1}{2h} \int_{x_{N-1}}^{x_{N+1}} v(x) dx = \frac{1}{h} \int_{x_{N-1}}^{x_N} u'(x) dx = \frac{u_N - u_{N-1}}{h}, \quad (b) \end{array} \right. \quad (2.25)$$

on déduit de (2.24) et (2.25) les approximations de  $u_{x,0}$  et  $u_{x,N}$  suivantes :

$$\left\{ \begin{array}{l} u_{x,0} = \frac{3}{2} \left( \frac{u_1 - u_0}{h} - \frac{1}{3} u_{x,1} \right), \\ u_{x,N} = \frac{3}{2} \left( \frac{u_N - u_{N-1}}{h} - \frac{1}{3} u_{x,N-1} \right). \end{array} \right. \quad (2.26)$$

Le schéma HB (2.11) s'exprime donc en terme des fonctions de grille  $(u, u_x) \in l_h^2 \times l_h^2$  par : trouver  $(u, u_x) \in l_h^2 \times l_h^2$  tel que

$$\left\{ \begin{array}{l} -\delta_x u_{x,j} = \Pi^0 f_j, \quad 1 \leq j \leq N-1, \quad (a) \\ \frac{2}{3} u_{x,N} + \frac{1}{3} u_{x,N-1} = \frac{u_N - u_{N-1}}{h}, \quad (b) \\ \frac{1}{6} u_{x,j-1} + \frac{2}{3} u_{x,j} + \frac{1}{6} u_{x,j+1} = \delta_x u_j, \quad 1 \leq j \leq N-1, \quad (c) \\ \frac{2}{3} u_{x,0} + \frac{1}{3} u_{x,1} = \frac{u_1 - u_0}{h}, \quad (d) \\ u_0 = g_a, \quad u_N = g_b. \quad (e) \end{array} \right. \quad (2.27)$$

Le système (2.27) s'appelle *schéma HB1*.

## 2.2.2 Expression matricielle du schéma HB1

Commençons par déterminer la forme matricielle du système (2.17). Ensuite, on en déduira la forme matricielle du schéma HB1 (2.27). Dans ce but, on dénote les valeurs de bord par (voir figure 2.3)

$$\left\{ \begin{array}{l} U_L = u_L = u_0, \\ U_R = u_R = u_N, \\ U_{x,L} = u_{x,L} = u_{x,0}, \\ U_{x,R} = u_{x,R} = u_{x,N}. \end{array} \right. \quad (2.28)$$

On rappelle que les vecteurs  $U, U_x \in \mathbb{R}^{N-1}$  sont donnés en (1.31).

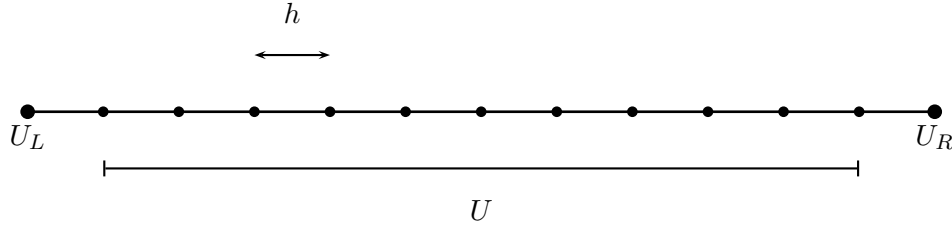


FIG. 2.3: Les notations en dimension 1 sur  $\Omega = (a, b)$ . La grande boule à gauche désigne le noeud  $x_0 = a$ , celle à droite désigne le noeud  $x_N = b$ . Les petites boules désignent les noeuds  $x_i$ ,  $1 \leq i \leq 11$

**Lemme 2.2.1.** Les équations (2.17)<sub>a</sub> s'écrivent matriciellement sous la forme

$$-\frac{1}{2h}KU_x - \frac{1}{2h}F_2(e_1U_{x,L} + e_{N-1}U_{x,R}) = F, \quad (2.29)$$

où  $F \in \mathbb{R}^{N-1}$ ,

$$F = [\Pi^0 f_1, \Pi^0 f_2, \dots, \Pi^0 f_{N-1}]^T, \quad (2.30)$$

et  $K$  est la matrice donnée en (1.40).

**Preuve.** (2.17)<sub>a</sub> donne

$$\left\{ \begin{array}{l} -\frac{1}{2h}u_{x,N} + \frac{1}{2h}u_{x,N-2} = \Pi^0 f_{N-1}, \\ -\frac{u_{x,N-1} - u_{x,N-3}}{2h} = \Pi^0 f_{N-2}, \\ \dots \\ -\frac{u_{x,3} - u_{x,1}}{2h} = \Pi^0 f_2, \\ -\frac{1}{2h}u_{x,2} + \frac{1}{2h}u_{x,0} = \Pi^0 f_1. \end{array} \right. \quad (2.31)$$

Ce qui est équivalent à

$$\left\{ \begin{array}{l} -\frac{1}{2h}(KU_x)_{N-1} - \frac{1}{2h}U_{x,N} = \Pi^0 f_{N-1}, \\ -\frac{1}{2h}(KU_x)_{N-2} = \Pi^0 f_{N-2}, \\ \dots \\ -\frac{1}{2h}(KU_x)_2 = \Pi^0 f_2, \\ -\frac{1}{2h}(KU_x)_1 + \frac{1}{2h}U_{x,0} = \Pi^0 f_1. \end{array} \right. \quad (2.32)$$



Matriciellement (2.32) s'écrit

$$-\frac{1}{2h}KU_x + \frac{1}{2h}e_1U_{x,L} - \frac{1}{2h}e_{N-1}U_{x,R} = [\Pi^0 f_1, \Pi^0 f_2, \dots, \Pi^0 f_{N-2}, \Pi^0 f_{N-1}]^T. \quad (2.33)$$

En plus on a,

$$\begin{aligned} \frac{1}{2h}e_1U_{x,L} - \frac{1}{2h}e_{N-1}U_{x,R} &= -\frac{1}{2h}(-e_1e_1^T + e_{N-1}e_{N-1}^T)(e_1U_{x,L} + e_{N-1}U_{x,R}) \\ &= -\frac{1}{2h}F_2(e_1U_{x,L} + e_{N-1}U_{x,R}). \end{aligned}$$

Enfin, (2.33) devient

$$-\frac{1}{2h}KU_x - \frac{1}{2h}F_2(e_1U_{x,L} + e_{N-1}U_{x,R}) = [\Pi^0 f_1, \Pi^0 f_2, \dots, \Pi^0 f_{N-2}, \Pi^0 f_{N-1}]^T,$$

d'où (2.29). ■

**Lemme 2.2.2.** *Les équations (2.17)<sub>b</sub> s'écrivent matriciellement sous la forme*

$$P_s U_x + \frac{1}{6}(e_1 U_{x,L} + e_{N-1} U_{x,R}) = \frac{1}{2h} K U + \frac{1}{2h} F_2(e_1 U_L + e_{N-1} U_R), \quad (2.34)$$

avec  $P_s$ ,  $K$ ,  $F_2$  les matrices données en (1.67), (1.40), (1.70) respectivement,  $U_L$ ,  $U_R$  sont les conditions de Dirichlet au bord et  $U_{x,L}$ ,  $U_{x,R}$  sont les dérivées au bord.

**Preuve.** (2.17)<sub>b</sub> s'écrit

$$\left\{ \begin{array}{l} \frac{1}{6}u_{x,N} + \frac{2}{3}u_{x,N-1} + \frac{1}{6}u_{x,N-2} = \frac{u_N - u_{N-2}}{2h}, \\ \frac{1}{6}u_{x,N-1} + \frac{2}{3}u_{x,N-2} + \frac{1}{6}u_{x,N-3} = \frac{u_{N-1} - u_{N-3}}{2h}, \\ \dots\dots\dots \\ \frac{1}{6}u_{x,3} + \frac{2}{3}u_{x,2} + \frac{1}{6}u_{x,1} = \frac{u_3 - u_1}{2h}, \\ \frac{1}{6}u_{x,2} + \frac{2}{3}u_{x,1} + \frac{1}{6}u_{x,0} = \frac{u_2 - u_0}{2h}. \end{array} \right. \quad (2.35)$$

Ceci est équivalent à

$$\left\{ \begin{array}{l} (P_s U_x)_{N-1} + \frac{1}{6} U_{x,N} = \frac{1}{2h} U_N - \frac{1}{2h} (KU)_{N-1}, \\ (P_s U_x)_{N-2} = \frac{1}{2h} (KU)_{N-2}, \\ \dots\dots \\ (P_s U_x)_2 = \frac{1}{2h} (KU)_2, \\ (P_s U_x)_1 + \frac{1}{6} U_{x,0} = \frac{1}{2h} (KU)_1 - \frac{1}{2h} U_0. \end{array} \right. \quad (2.36)$$

En utilisant les notations (2.28), on obtient

$$P_s U_x + \frac{1}{6} e_1 U_{x,L} + \frac{1}{6} e_{N-1} U_{x,R} = \frac{1}{2h} KU - \frac{1}{2h} e_1 U_L + \frac{1}{2h} e_{N-1} U_R. \quad (2.37)$$

On peut aisément vérifier que

$$\begin{aligned} -\frac{1}{2h} e_1 U_L + \frac{1}{2h} e_{N-1} U_R &= \frac{1}{2h} (-e_1 e_1^T + e_{N-1} e_{N-1}^T) (e_1 U_L + e_{N-1} U_R) \\ &= \frac{1}{2h} F_2 (e_1 U_L + e_{N-1} U_R). \end{aligned}$$

Donc (2.37) se réécrit

$$P_s U_x + \frac{1}{6} (e_1 U_{x,L} + e_{N-1} U_{x,R}) = \frac{1}{2h} KU + \frac{1}{2h} F_2 (e_1 U_L + e_{N-1} U_R). \quad (2.38)$$

■

L'identité (2.34) permet de calculer le vecteur  $U_x$  en fonction des inconnues  $U$  et des données au bord  $U_L$  et  $U_R$ ,

$$U_x = P_s^{-1} \left( \frac{1}{2h} KU + \frac{1}{2h} F_2 (e_1 U_L + e_{N-1} U_R) - \frac{1}{6} (e_1 U_{x,L} + e_{N-1} U_{x,R}) \right). \quad (2.39)$$

En remplaçant  $U_x$  dans (2.29) on obtient

$$-\frac{1}{4h^2} (K P_s^{-1} K) U - \frac{1}{2h} (F_2 - \frac{1}{6} K P_s^{-1}) (e_1 U_{x,L} + e_{N-1} U_{x,R}) - \frac{1}{2h} K P_s^{-1} F_2 (e_1 U_L + e_{N-1} U_R) = F. \quad (2.40)$$

On peut écrire

$$\left\{ \begin{array}{l} e_1 U_L + e_{N-1} U_R = [U_L, 0, \dots, 0, U_R]^T, \quad (a) \\ e_1 U_{x,L} + e_{N-1} U_{x,R} = [U_{x,L}, 0, \dots, 0, U_{x,R}]^T. \quad (b) \end{array} \right. \quad (2.41)$$

Noter que dans (2.40) la matrice  $KP_s^{-1}K$  n'est pas inversible et ceci nous empêche d'utiliser une méthode de résolution par le complément de Schur par exemple. Pour cela, en considérant le cas du problème (2.5) et les approximations (2.26) des dérivées au bord, on exprime dans le lemme 2.2.3 suivant le vecteur  $e_1U_{x,L} + e_{N-1}U_{x,R} \in \mathbb{R}^{(N-1)}$  dans (2.29) et (2.39) en fonction des vecteurs  $U, U_x \in \mathbb{R}^{(N-1)}$  et des données au bord  $U_L$  et  $U_R \in \mathbb{R}$ .

**Lemme 2.2.3.** *En utilisant les approximations (2.26) des dérivées  $U_{x,L}$  et  $U_{x,R}$  au bord, le vecteur  $e_1U_{x,L} + e_{N-1}U_{x,R}$  s'écrit matriciellement*

$$e_1U_{x,L} + e_{N-1}U_{x,R} = \frac{1}{h}\mathcal{A}_1U - \mathcal{B}_1U_x + \frac{1}{h}\mathcal{C}_1(e_1U_L + e_{N-1}U_R), \quad (2.42)$$

où les matrices  $\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1$  sont

$$\begin{cases} \mathcal{A}_1 = -\frac{3}{2}F_2, & (a) \\ \mathcal{B}_1 = \frac{1}{2}F_1, & (b) \\ \mathcal{C}_1 = -\frac{3}{2}F_1, & (c) \end{cases} \quad (2.43)$$

avec les matrices  $F_1, F_2$  sont données en (1.70).

**Preuve.** D'après (2.26), on peut écrire les dérivées au bord,  $u_{x,0}$  et  $u_{x,N}$  sous la forme,

$$\begin{cases} u_{x,0} = \frac{1}{h}\left(\frac{3}{2}e_1\right)^T U - \left(\frac{1}{2}e_1\right)^T U_x - \frac{1}{h}\left(\frac{3}{2}e_1\right)^T [U_L, 0, \dots, 0, U_R]^T, \\ u_{x,N} = \frac{1}{h}\left(-\frac{3}{2}e_{N-1}\right)^T U - \left(\frac{1}{2}e_{N-1}\right)^T U_x + \frac{1}{h}\left(\frac{3}{2}e_{N-1}\right)^T [U_L, 0, \dots, 0, U_R]^T, \end{cases} \quad (2.44)$$

où  $e_1, e_{N-1}$  sont respectivement le premier et le dernier élément de la base canonique de  $\mathbb{R}^{N-1}$ . Les vecteurs  $U, U_x \in \mathbb{R}^{N-1}$  sont donnés en (1.31), et  $U_L = g_a, U_R = g_b$  sont les conditions de Dirichlet au bord. On a donc

$$\begin{aligned} e_1U_{x,L} + e_{N-1}U_{x,R} &= \frac{3}{2h}(e_1e_1^T - e_{N-1}e_{N-1}^T)U - \frac{1}{2}(e_1e_1^T + e_{N-1}e_{N-1}^T)U_x \\ &+ \frac{3}{2h}(-e_1e_1^T + e_{N-1}e_{N-1}^T)[U_L, 0, \dots, 0, U_R]^T \\ &= \frac{1}{h}\mathcal{A}_1U - \frac{1}{2}\mathcal{B}_1U_x + \frac{1}{h}\mathcal{C}_1(e_1U_L + e_{N-1}U_R), \end{aligned}$$

où  $\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1 \in \mathbb{M}_{N-1}(\mathbb{R})$  sont telles que

$$\begin{cases} \mathcal{A}_1 = \frac{3}{2}(e_1 e_1^T - e_{N-1} e_{N-1}^T) = -\frac{3}{2}F_2, & (a) \\ \mathcal{B}_1 = \frac{1}{2}(e_1 e_1^T + e_{N-1} e_{N-1}^T) = \frac{1}{2}F_1, & (b) \\ \mathcal{C}_1 = \frac{3}{2}(e_1 e_1^T + e_{N-1} e_{N-1}^T) = \frac{3}{2}F_1, & (c) \end{cases} \quad (2.45)$$

d'où le résultat (2.42). ■

En utilisant le lemme 2.2.3, on peut établir une nouvelle forme matricielle du vecteur  $U_x$ . En effet, on remplace dans (2.34), le vecteur  $e_1 U_{x,L} + e_{N-1} U_{x,R} \in \mathbb{R}^{(N-1)}$  par son expression (2.42). On obtient après quelques calculs

$$\begin{cases} U_x = \frac{1}{2h}(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(K - \frac{1}{3}\mathcal{A}_1)U \\ + \frac{1}{2h}(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(F_2 - \frac{1}{3}\mathcal{C}_1)(e_1 U_L + e_{N-1} U_R). \end{cases} \quad (2.46)$$

**Proposition 2.2.** *Le schéma HB1 (2.27) s'écrit matriciellement : trouver  $U \in \mathbb{R}^{N-1}$  tel que*

$$\frac{1}{h^2}\mathcal{H}_1 U = F - \frac{1}{h^2}\mathcal{G}_1(e_1 U_L + e_{N-1} U_R), \quad (2.47)$$

avec  $F \in \mathbb{R}^{(N-1)}$  donné en (2.30) et

$$\begin{cases} \mathcal{H}_1 = -\frac{1}{4}(K - F_2\mathcal{B}_1)(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(K - \frac{1}{3}\mathcal{A}_1) - \frac{1}{2}F_2\mathcal{A}_1, & (a) \\ \mathcal{G}_1 = -\frac{1}{4}(K - F_2\mathcal{B}_1)(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(F_2 - \frac{1}{3}\mathcal{C}_1) - \frac{1}{2}F_2\mathcal{C}_1, & (b) \end{cases} \quad (2.48)$$

où les matrices  $\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1$  sont données au lemme 2.2.3.

**Preuve.** En remplaçant le vecteur  $e_1 U_{x,L} + e_{N-1} U_{x,R} \in \mathbb{R}^{(N-1)}$  dans (2.29) par sa valeur en (2.42) on obtient

$$\begin{cases} -\frac{1}{2h}K U_x - \frac{1}{2h}F_2\left(\frac{1}{h}\mathcal{A}_1 U - \mathcal{B}_1 U_x + \frac{1}{h}\mathcal{C}_1(e_1 U_L + e_{N-1} U_R)\right) & = -\frac{1}{2h}(K - F_2\mathcal{B}_1)U_x - \frac{1}{2h^2}F_2\mathcal{A}_1 U \\ & - \frac{1}{2h^2}F_2\mathcal{C}_1(e_1 U_L + e_{N-1} U_R) \\ & = F. \end{cases} \quad (2.49)$$

En remplaçant  $U_x \in \mathbb{R}^{(N-1)}$  dans (2.49) par sa valeur en (2.46), on obtient

$$\begin{cases} -\frac{1}{4h^2}(K - F_2\mathcal{B}_1)(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(K - \frac{1}{3}\mathcal{A}_1)U \\ -\frac{1}{4h^2}(K - F_2\mathcal{B}_1)(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(F_2 - \frac{1}{3}\mathcal{C}_1)(e_1U_L + e_{N-1}U_R) \\ -\frac{1}{2h^2}F_2\mathcal{A}_1U - \frac{1}{2h^2}F_2\mathcal{C}_1(e_1U_L + e_{N-1}U_R) = F. \end{cases} \quad (2.50)$$

En utilisant les matrices  $\mathcal{H}_1$  et  $\mathcal{G}_1$  en (2.48) on déduit (2.47). ■

**Corollaire 2.2.4.** *La matrice  $\mathcal{H}_1$  (2.48)<sub>a</sub> est symétrique.*

**Preuve.** En utilisant le lemme 1.3.3,

$$F_2\mathcal{B}_1 = F_2(\frac{1}{2}F_1) = -\frac{1}{3}\mathcal{A}_1.$$

La matrice  $\mathcal{H}_1$  en (2.48)<sub>a</sub> devient

$$\mathcal{H}_1 = -\frac{1}{4}(K + \frac{1}{3}\mathcal{A}_1)(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(K - \frac{1}{3}\mathcal{A}_1) - \frac{1}{2}F_2\mathcal{A}_1. \quad (2.51)$$

Les matrices  $(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}$ ,  $F_2\mathcal{A}_1$  sont symétriques car les matrices  $P_s$ ,  $\mathcal{A}_1$ ,  $\mathcal{B}_1$ ,  $F_2$  le sont. La matrice  $(K - \frac{1}{3}\mathcal{A}_1)$  vérifie

$$(K - \frac{1}{3}\mathcal{A}_1)^T = -K - \frac{1}{3}\mathcal{A}_1, \quad (2.52)$$

car la matrice  $K$  est antisymétrique et la matrice  $\mathcal{A}_1$  est symétrique. Enfin,

$$\begin{aligned} (\mathcal{H}_1)^T &= \left( -\frac{1}{4}(K + \frac{1}{3}\mathcal{A}_1)(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(K - \frac{1}{3}\mathcal{A}_1) - \frac{1}{2}F_2\mathcal{A}_1 \right)^T \\ &= -\frac{1}{4} \left( (K + \frac{1}{3}\mathcal{A}_1)(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(K - \frac{1}{3}\mathcal{A}_1) \right)^T - \frac{1}{2}(F_2\mathcal{A}_1)^T \\ &= -\frac{1}{4}(K - \frac{1}{3}\mathcal{A}_1)^T((P_s - \frac{1}{6}\mathcal{B}_1)^{-1})^T(K + \frac{1}{3}\mathcal{A}_1)^T - \frac{1}{2}(F_2\mathcal{A}_1)^T \\ &= -\frac{1}{4}(-K - \frac{1}{3}\mathcal{A}_1)(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}(-K + \frac{1}{3}\mathcal{A}_1) - \frac{1}{2}F_2\mathcal{A}_1 \\ &= \mathcal{H}_1. \end{aligned}$$

■

Le résultat principal de cette section est le suivant.

**Proposition 2.3** (Forme diagonale de  $\mathcal{H}_1$ ). *La matrice  $\mathcal{H}_1$  d'ordre  $(N-1)$  en (2.47) s'écrit*

$$\mathcal{H}_1 = P_s^{-1} \left( T - \frac{1}{4} T^2 \right), \quad (2.53)$$

*elle est donc symétrique définie positive diagonalisable dans la base  $(Z^k)_{1 \leq k \leq N-1}$  de vecteurs propres (1.43) de la matrice  $T$  (1.42).*

**Preuve.** Soient  $T$  la matrice de Laplace,  $u \in l_{h,0}^2$  la fonction de grille solution du schéma HB1 (2.27). On note  $U = \text{vec}_1(u) \in \mathbb{R}^{(N-1)}$  le vecteur solution du schéma HB1. On peut facilement démontrer que le vecteur  $(T^2 - 4T)U \in \mathbb{R}^{(N-1)}$  vérifie

$$\begin{cases} \left( (T^2 - 4T)U \right)_{N-1} = 3u_{N-1} - u_{N-3}, & (a) \\ \left( (T^2 - 4T)U \right)_j = u_{j+2} + u_{j-2} - 2u_j, \quad 2 \leq j \leq N-2, & (b) \\ \left( (T^2 - 4T)U \right)_1 = 3u_1 - u_3. & (c) \end{cases} \quad (2.54)$$

Notons par  $E_j$ ,  $1 \leq j \leq N-1$ , la  $j$ -ème équation du (2.27)<sub>a</sub>. On a

$$\begin{cases} \frac{2}{3}E_{N-1} + \frac{1}{6}E_{N-2} = \frac{2}{3}\Pi^0 f_{N-1} + \frac{1}{6}\Pi^0 f_{N-2}, \\ \frac{1}{6}E_{j-1} + \frac{2}{3}E_j + \frac{1}{6}E_{j+1} = \frac{1}{6}\Pi^0 f_{j-1} + \frac{2}{3}\Pi^0 f_j + \frac{1}{6}\Pi^0 f_{j+1}, \\ \frac{2}{3}E_1 + \frac{1}{6}E_2 = \frac{2}{3}\Pi^0 f_1 + \frac{1}{6}\Pi^0 f_2, \end{cases} \quad (2.55)$$

D'autre part, on a

$$\begin{cases} \frac{2}{3}E_{N-1} + \frac{1}{6}E_{N-2} = -\frac{2}{3} \frac{u_{x,N} - u_{x,N-2}}{2h} - \frac{1}{6} \frac{u_{x,N-1} - u_{x,N-3}}{2h} \\ = \frac{1}{2h} \left( -\frac{2}{3}u_{x,N} + \frac{2}{3}u_{x,N-2} - \frac{1}{6}u_{x,N-1} + \frac{1}{6}u_{x,N-3} \right) \\ = \frac{1}{2h} \left( -\frac{2}{3}u_{x,N} - \frac{1}{3}u_{x,N-1} + \frac{1}{6}u_{x,N-1} + \frac{2}{3}u_{x,N-2} + \frac{1}{6}u_{x,N-3} \right). \end{cases} \quad (2.56)$$

En utilisant les relations suivantes déduites de (2.27)<sub>b,c</sub>

$$\begin{cases} \frac{2}{3}u_{x,N} + \frac{1}{3}u_{x,N-1} = \frac{u_N - u_{N-1}}{h}, \\ \frac{1}{6}u_{x,N-1} + \frac{2}{3}u_{x,N-2} + \frac{1}{6}u_{x,N-3} = \frac{u_{N-1} - u_{N-3}}{2h}, \end{cases}$$

on obtient de (2.56)

$$\begin{cases} \frac{2}{3}E_N + \frac{1}{3}E_{N-1} &= \frac{1}{2h} \left( -\frac{u_N - u_{N-1}}{h} + \frac{u_{N-1} - u_{N-3}}{2h} \right) \\ &= \frac{1}{2h} \left( \frac{3u_{N-1} - u_{N-3}}{2h} \right) - \frac{u_N}{2h^2}. \end{cases} \quad (2.57)$$

En utilisant (2.54)<sub>a</sub>, (2.57) donne

$$\frac{2}{3}E_N + \frac{1}{3}E_{N-1} = \frac{1}{4h^2} \left( (T^2 - 4T)U \right)_{N-1} - \frac{u_N}{2h^2}. \quad (2.58)$$

De la même manière, on démontre

$$\frac{2}{3}E_1 + \frac{1}{3}E_2 = \frac{1}{4h^2} \left( (T^2 - 4T)U \right)_1 + \frac{u_0}{2h^2}. \quad (2.59)$$

D'après (2.22) on a

$$\begin{aligned} \frac{1}{6}E_{j-1} + \frac{2}{3}E_j + \frac{1}{6}E_{j+1} &= -\frac{u_{j+2} + u_{j-2} - 2u_j}{4h^2} \\ &= \left( (T^2 - 4T)U \right)_j, \quad 2 \leq j \leq N-2. \end{aligned}$$

On déduit de (2.56) que le vecteur solution  $U$  du schéma HB1 (2.27) vérifie

$$\begin{cases} \frac{1}{4h^2} \left( (T^2 - 4T)U \right)_{N-1} = \frac{2}{3}\Pi^0 f_{N-1} + \frac{1}{6}\Pi^0 f_{N-2} - \frac{u_N}{2h^2}, \\ \frac{1}{4h^2} \left( (T^2 - 4T)U \right)_j = \frac{1}{6}\Pi^0 f_{j-1} + \frac{2}{3}\Pi^0 f_j + \frac{1}{6}\Pi^0 f_{j+1}, \quad 2 \leq j \leq N-2, \\ \frac{1}{4h^2} \left( (T^2 - 4T)U \right)_1 = \frac{2}{3}\Pi^0 f_1 + \frac{1}{6}\Pi^0 f_2 + \frac{u_0}{2h^2}, \end{cases} \quad (2.60)$$

Matriciellement, comme  $u_0 = u_N = 0$ , (2.60) s'écrit

$$-\frac{(T^2 - 4T)}{4h^2}U = P_s F, \quad (2.61)$$

avec  $F = \text{vec}_1(\Pi^0 f)$ . Ce qui est équivalent à

$$\frac{-P_s^{-1}(T^2 - 4T)}{4h^2}U = F. \quad (2.62)$$

En comparant avec (2.47), on trouve

$$\mathcal{H}_1 = \frac{-P_s^{-1}(T^2 - 4T)}{4}, \quad (2.63)$$

d'où (2.53). ■

**Corollaire 2.2.5.** *Les valeurs propres  $(\mu_k)_{1 \leq k \leq N-1}$  de la matrice  $\mathcal{H}_1$  sont telles que*

$$\mu_k = \frac{\sin^2\left(\frac{k\pi}{N}\right)}{\frac{2}{3} + \frac{1}{3} \cos\left(\frac{k\pi}{N}\right)}. \quad (2.64)$$

**Preuve.** Il résulte de la proposition 2.3 que les valeurs propres de la matrice  $\mathcal{H}_1$  vérifient

$$\mu_k = \frac{\lambda_k - \frac{1}{4}\lambda_k^2}{1 - \frac{1}{6}\lambda_k}, \quad (2.65)$$

avec  $\lambda_k$ ,  $1 \leq k \leq N-1$  les valeurs propres de la matrice  $T$  données en (1.43)<sub>b</sub>. Alors

$$\begin{aligned} \mu_k &= \frac{4 \sin^2\left(\frac{k\pi}{2N}\right) - \frac{16}{4} \sin^4\left(\frac{k\pi}{2N}\right)}{1 - \frac{2}{3} \sin^2\left(\frac{k\pi}{2N}\right)} \\ &= \frac{\frac{4}{4} \sin^2\left(\frac{2k\pi}{2N}\right)}{1 - \frac{1}{3} (2 - 2 \cos^2\left(\frac{k\pi}{2N}\right))} \\ &= \frac{\sin^2\left(\frac{k\pi}{N}\right)}{1 - \frac{1}{3} (1 - \cos\left(\frac{k\pi}{N}\right))} \\ &= \frac{\sin^2\left(\frac{k\pi}{N}\right)}{\frac{2}{3} + \frac{1}{3} \cos\left(\frac{k\pi}{N}\right)}. \end{aligned}$$

■

### 2.2.3 Stabilité

On déduit de la proposition 2.3 les corollaires suivants :

**Corollaire 2.2.6** (Existence et unicité de la solution du schéma HB1). *Le schéma HB1 (2.27) possède une solution unique  $(U, U_x) \in \mathbb{R}^{N-1} \times \mathbb{R}^{N-1}$ .*



**Preuve.** D'après (2.65), toutes les valeurs propres  $(\mu_k)_{1 \leq k \leq N-1}$  sont strictement positives donc la matrice  $\mathcal{H}_1$  est inversible et d'après (2.53) son inverse est donné par :

$$\mathcal{H}_1^{-1} = (T - \frac{1}{4}T^2)^{-1}P_s. \quad (2.66)$$

Donc le schéma HB1 (2.27) admet une unique solution  $U \in \mathbb{R}^{N-1}$ . De plus le vecteur gradient  $U_x \in \mathbb{R}^{N-1}$  peut être calculé par (2.46).

**Corollaire 2.2.7** (Stabilité  $L^2$  du schéma HB1). *Le schéma HB1 est stable au sens suivant*

$$\|U\|_2 \leq \frac{1}{4}\|F\|_2, \quad (2.67)$$

avec  $F \in \mathbb{R}^{(N-1)}$  donné en (2.30).

**Preuve.** D'après la proposition 2.3, les vecteurs propres  $(Z^k)_{1 \leq k \leq N-1}$  de la matrice  $\mathcal{H}_1$  forment une base orthonormale de  $\mathbb{R}^{N-1}$ . Donc pour tout  $U \in \mathbb{R}^{N-1}$ ,

$$U = \sum_{i=1}^{N-1} \alpha_i Z^i, \quad (2.68)$$

avec

$$\alpha_i = \langle U, Z^i \rangle, \quad 1 \leq i \leq N-1. \quad (2.69)$$

On en déduit

$$\frac{1}{h^2}(\mathcal{H}_1 U, U) = \frac{1}{h^2}(\mathcal{H}_1 \sum_{i=1}^{N-1} \alpha_i Z^i, \sum_{j=1}^{N-1} \alpha_j Z^j) \quad (2.70)$$

$$= \frac{1}{h^2} \sum_{i=1}^{N-1} \mu_i \alpha_i^2 \quad (2.71)$$

$$\geq \frac{1}{h^2} \min_{1 \leq i \leq N-1} (\mu_i) \sum_{i=1}^{N-1} \alpha_i^2 \quad (2.72)$$

$$\geq \frac{1}{h^2} \min_{1 \leq i \leq N-1} (\mu_i) \|U\|_2^2. \quad (2.73)$$

On a

$$\begin{aligned} \frac{1}{h^2} \min_{1 \leq i \leq N-1} (\mu_i) &= \frac{1}{h^2} \frac{\sin^2(\frac{\pi}{N})}{2 \frac{1}{3} + \frac{1}{3} \cos(\frac{\pi}{N})} \\ &\geq \frac{1}{h^2} \frac{4}{N^2} = 4(\text{indépendant de } h), \end{aligned}$$

car  $\frac{2}{\pi}x \leq \sin(x)$ . Donc

$$\frac{1}{h^2}(\mathcal{H}_1 U, U) \geq 4\|U\|_2^2. \quad (2.74)$$

En utilisant l'inégalité de Cauchy-Schwarz on obtient

$$\|U\|_2 \leq \frac{1}{4}\|F\|_2. \quad (2.75)$$

■

## 2.2.4 Principe du maximum

Rappelons d'abord le principe du maximum et du minimum pour l'opérateur  $\delta_x^2$ .

**Lemme 2.2.8.** Soit  $u = (u_i)_{0 \leq i \leq N} \in l_h^2$ . Supposons que  $-\delta_x^2 u_i \leq 0$ ,  $1 \leq i \leq N-1$ , alors

$$\max_{1 \leq i \leq N-1} u_i \leq \max(u_0, u_N). \quad (\text{Principe du maximum}). \quad (2.76)$$

De même, si  $-\delta_x^2 u_i \geq 0$ ,  $1 \leq i \leq N-1$ ,  $u$  vérifie

$$\min_{1 \leq i \leq N-1} u_i \geq \min(u_0, u_N). \quad (\text{Principe du minimum}). \quad (2.77)$$

**Preuve.** Voir le livre [4].

Le schéma HB1 vérifie le principe du maximum et du minimum discret suivant.

**Proposition 2.4** (Principe du maximum et du minimum). Soit  $u = (u_i)_{0 \leq i \leq N} \in l_h^2$  la solution du schéma HB1 (2.27). Supposons que  $f_i \leq 0$ ,  $\forall 1 \leq i \leq N-1$ , alors

$$\max_{1 \leq i \leq N-1} u_i \leq \max\{u_0, u_N\}. \quad (\text{Principe du maximum}) \quad (2.78)$$

De même, si  $f_i \geq 0$ ,  $1 \leq i \leq N-1$ , alors

$$\min_{1 \leq i \leq N-1} u_i \geq \min\{u_0, u_N\}. \quad (\text{Principe du minimum}) \quad (2.79)$$

Cela signifie que  $u$  atteint ses valeurs maxima et minima aux points de bord  $x_0$  et  $x_N$  quand  $f = 0$ .

**Preuve.** Supposons que  $f_i \leq 0$ ,  $\forall 1 \leq i \leq N-1$  alors

$$\Pi^0 f_i \leq 0, \quad 1 \leq i \leq N-1, \quad (2.80)$$

d'après (2.20)

$$-\frac{u_{i+2} + u_{i-2} - 2u_i}{4h^2} \leq 0, \quad \forall 2 \leq i \leq N-2, \quad (2.81)$$

Si  $N$  est pair, d'après le principe du maximum du lemme 2.2.8 on obtient

$$\begin{cases} \max_{1 \leq k \leq \frac{N-2}{2}} u_{2k} \leq \max\{u_0, u_N\}, \\ \max_{1 \leq k \leq \frac{N-2}{2}} u_{2k+1} \leq \max\{u_1, u_{N-1}\}. \end{cases} \quad (2.82)$$

De même, si  $N$  est impair, d'après le principe du maximum du lemme 2.2.8,

$$\begin{cases} \max_{1 \leq k \leq \frac{N-1}{2}} u_{2k} \leq \max\{u_0, u_{N-1}\}, \\ \max_{1 \leq k \leq \frac{N-1}{2}} u_{2k+1} \leq \max\{u_1, u_N\}. \end{cases} \quad (2.83)$$

Il en résulte pour tout  $N$ ,

$$\max_{2 \leq i \leq N-2} u_i \leq \max\{u_0, u_1, u_{N-1}, u_N\}. \quad (2.84)$$

Donc pour démontrer (2.131) il suffit de démontrer que

$$\max\{u_1, u_{N-1}\} \leq \max\{u_0, u_N\}. \quad (2.85)$$

Supposons

$$\max\{u_1, u_{N-1}\} = u_1. \quad (2.86)$$

D'après (2.27)<sub>a</sub>,

$$-\delta_x u_{x,j} = \Pi^0 f_j \leq 0, \quad \forall 1 \leq j \leq N-1, \quad (2.87)$$

cela donne  $u_{x,0} \leq u_{x,2}$  et  $u_{x,1} \leq u_{x,3}$ . D'après (2.27)<sub>c,d</sub> on a

$$\begin{cases} \frac{2}{3}u_{x,0} + \frac{1}{3}u_{x,1} = \frac{u_1 - u_0}{h}, \\ \frac{1}{6}u_{x,1} + \frac{2}{3}u_{x,2} + \frac{1}{3}u_{x,3} = \frac{u_3 - u_1}{2h} \leq 0, \text{ car } u_3 \leq u_1. \end{cases} \quad (2.88)$$

Ensuite, on a

$$\frac{2}{3}u_{x,0} + \frac{1}{3}u_{x,1} = \frac{u_1 - u_0}{h} \leq \frac{1}{6}u_{x,1} + \frac{2}{3}u_{x,2} + \frac{1}{6}u_{x,3} = \frac{u_3 - u_1}{2h} \leq 0, \quad (2.89)$$

car  $u_3 \leq u_1$ , donc  $u_1 \leq u_0$ . De plus

$$\max\{u_1, u_{N-1}\} \leq \max\{u_0, u_N\}. \quad (2.90)$$

Supposons

$$\max\{u_1, u_{N-1}\} = u_{N-1}. \quad (2.91)$$

(2.27)<sub>b,c</sub> donnent

$$\begin{cases} \frac{2}{3}u_{x,N} + \frac{1}{3}u_{x,N-1} = \frac{u_N - u_{N-1}}{h}, \\ \frac{1}{6}u_{x,N-3} + \frac{2}{3}u_{x,N-2} + \frac{1}{3}u_{x,N-1} = \frac{u_{N-1} - u_{N-2}}{2h} \geq 0, \text{ car } u_{N-1} \geq u_{N-2}. \end{cases} \quad (2.92)$$

En plus, d'après (2.87), on a  $u_{x,N-2} \leq u_{x,N}$  et  $u_{x,N-3} \leq u_{x,N-1}$  alors

$$\frac{2}{3}u_{x,N} + \frac{1}{3}u_{x,N-1} = \frac{u_N - u_{N-1}}{2h} \geq \frac{1}{6}u_{x,N-3} + \frac{2}{3}u_{x,N-2} + \frac{1}{6}u_{x,N-1} = \frac{u_{N-1} - u_{N-3}}{2h} \geq 0, \quad (2.93)$$

car  $u_{N-1} \geq u_{N-3}$ , alors  $u_N \geq u_{N-1}$  ensuite

$$\max\{u_1, u_{N-1}\} \leq \max\{u_0, u_N\}. \quad (2.94)$$

Il résulte de (2.84) et (2.85) que

$$\max_{1 \leq i \leq N-1} u_i \leq \max\{u_0, u_N\}, \quad (2.95)$$

d'où (2.131). De la même manière on démontre (2.79). ■

## 2.2.5 Résultats numériques

Dans cette section on présente quelques résultats numériques prouvant l'intérêt du schéma HB1. On résoud le problème de Poisson (2.5) sur  $\Omega = (0, 1)$ . La solution exacte donne les conditions de Dirichlet sur la frontière. Le système linéaire résolu est (2.47). On observe numériquement que le schéma HB1 est d'ordre 2 pour  $u$  et  $u_x$ . Le calcul est effectué en Matlab en utilisant la méthode directe pour inverser la matrice  $\mathcal{H}_1$ . Les erreurs entre la solution exacte  $u_{ex}$  et la solution calculée  $u$ , ainsi que, l'erreur entre le gradient exacte  $u_{x,ex}$  et le gradient calculé  $u_x$ , sont estimés par les normes  $L^2$  et  $L^\infty$

$N$	$\ u - u_{ex}\ _h$	$\ u - u_{ex}\ _\infty$	$\ u_x - u_{x,ex}\ _h$	$\ u_x - u_{x,ex}\ _\infty$
128	1.639(-4)	3.9853(-4)	6.39(-4)	8.036(-4)
Taux de conv.	2	1.99	2	2
256	4.098(-5)	1(-4)	1.592(-4)	2.008(-4)
Taux de conv.	2	1.99	2	2
512	1.024(-5)	2.505(-5)	3.9748(-5)	3.0201(-5)
Taux de conv.	2	1.99	2	2
1024	2.561(-6)	6.268(-6)	9.929(-6)	1.255(-5)
Taux de conv.	2	1.99	2	2
2048	6.404(-7)	1.568(-6)	9.4814(-6)	3.1375(-6)

TAB. 2.1: Erreur et taux de convergence de Schéma HB1 avec  $u_{ex}(x) = \cos(\pi x^2)$  et  $u_{x,ex}(x) = -2\pi x \sin(\pi x^2)$

$N$	$\ u - u_{ex}\ _h$	$\ u - u_{ex}\ _\infty$	$\ u_x - u_{x,ex}\ _h$	$\ u_x - u_{x,ex}\ _\infty$
128	7.836(-5)	1.907(-4)	3.058(-4)	3.855(-4)
Taux de conv.	2	1.99	2	2
256	1.957(-5)	4.778(-5)	7.607(-5)	9.56(-5)
Taux de conv.	2	1.99	2	2
512	4.8929(-6)	1.1963(-5)	1.898(-5)	2.3976(-5)
Taux de conv.	2	1.99	2	2
1024	1.2223(-6)	2.993(-6)	4.7411(-6)	5.9926(-6)
Taux de conv.	2	1.99	2	2
2048	3.0578(-7)	7.4866(-7)	1.1848(-6)	1.4981(-6)

TAB. 2.2: Erreur et taux de convergence de Schéma HB1 avec  $u_{ex}(x) = \sin(\pi x^3)$  et  $u_{x,ex}(x) = 3\pi x^2 \cos(\pi x^3)$

discrètes définies par

$$\left\{ \begin{array}{l} \|u - u_{ex}\|_h = \left[ h \sum_{i=1}^{N-1} (u_i - u_{ex}(x_i))^2 \right]^{\frac{1}{2}}, \quad (a) \\ \|u - u_{ex}\|_\infty = \max_{i=1, \dots, N-1} |u_i - u_{ex}(x_i)|. \quad (b) \end{array} \right. \quad (2.96)$$

**Test 1 :** On considère la fonction  $u_{ex}(x) = \cos(\pi x^2)$ . On a  $f(x) = 4\pi^2 x^2 \cos(\pi x^2)$ . Les résultats numériques sont donnés dans la table 2.1.

**Test 2 :** On considère la solution exacte  $u_{ex}(x) = \sin(\pi x^3)$ . Les résultats numériques sont donnés dans la table 2.2. Le schéma HB1 est d'ordre 2 pour les deux inconnues avec  $u_{x,ex}(x) = 3\pi x^2 \cos(\pi x^3)$ .

$N$	$\ u - u_{ex}\ _h$	$\ u - u_{ex}\ _\infty$	$\ u_x - u_{x,ex}\ _h$	$\ u_x - u_{x,ex}\ _\infty$
128	5.602(-6)	7.922(-6)	5.131(-10)	6.771(-10)
Taux de conv.	2.00	2.00	4.00	4.00
256	1.4005(-5)	1.980(-6)	3.196(-11)	4.231(-11)
Taux de conv.	2.00	2.00	4.00	3.99
512	3.501(-7)	4.951(-7)	1.995(-12)	2.647(-12)
Taux de conv.	2.00	2.00	4.00	4.00
1024	8.753(-8)	1.237(-7)	1.242(-13)	1.648(-13)
Taux de conv.	2.00	2.00	4.26	4.06
2048	2.188(-8)	3.094(-8)	6.476(-15)	9.881(-15)

TAB. 2.3: Erreur et taux de convergence de Schéma HB1 avec  $u_{ex}(x) = \exp(-(x - \frac{1}{2})^2)$ 

$N$	$\ u - u_{ex}\ _h$	$\ u - u_{ex}\ _\infty$	$\ u_x - u_{x,ex}\ _h$	$\ u_x - u_{x,ex}\ _\infty$
128	9.054(-5)	5.374(-4)	4.4(-3)	2.29(-2)
Taux de conv.	4.07	4.08	4.10	4.14
256	5.365(-6)	3.162(-5)	2.554(-5)	1.3(-3)
Taux de conv.	4.01	4.02	4.02	3.95
512	3.312(-7)	1.949(-6)	1.572(-5)	8.408(-5)
Taux de conv.	4.00	4.00	4.00	4.00
1024	2.063(-8)	1.214(-7)	9.79(-7)	5.23(-8)
taux de conv.	4.00	4.00	4.00	4.00
2048	1.288(-9)	7.581(-9)	6.113(-8)	3.267(-7)

TAB. 2.4: Erreur et taux de convergence de Schéma HB1 avec  $u_{ex}(x) = \exp(-1000(x - \frac{1}{2})^2)$ 

**Test 3 :** La solution exacte est  $u_{ex}(x) = \exp(-(x - \frac{1}{2})^2)$ . Le table 2.3 montre que le schéma HB1 est d'ordre 2 pour les deux inconnues.

**Test 4 :** La table 2.4 montre qu'on a un taux de convergence de 4 dans certains cas, tel que  $u_{ex}(x) = \exp(-1000(x - \frac{1}{2})^2)$ . Cette fonction est nulle au voisinage des points de bord.

## 2.3 Schéma boîte hermitien HB2

### 2.3.1 Design du schéma HB2

Dans la section 2.2, on a vu que les relations (2.17) sont d'ordre 4. La perte de précision qui donne un schéma numérique d'ordre 2 pour  $u$  et  $u_x$  provient des conditions limites d'ordre 1 (2.26).

Dans cette section, nous cherchons à construire des schémas boîte d'ordre plus élevé. Le principe est de remplacer la formule d'approximation d'ordre 1 des dérivées aux bords par la formule d'ordre 3 décrite dans la proposition 2.5 suivante.

**Proposition 2.5.** *Soit  $u(x)$  une fonction régulière sur  $[a, b]$ . Les dérivées  $u_{x,0}$  et  $u_{x,N}$  aux points  $x_0$  et  $x_N$  respectivement, vérifient les relations d'ordre 3 suivantes :*

$$\begin{cases} \frac{1}{3}u_{x,0} + \frac{2}{3}u_{x,1} = \frac{1}{h} \left( \frac{1}{6}u_2 + \frac{2}{3}u_1 - \frac{5}{6}u_0 \right) + \mathcal{O}(h^3), & (a) \\ \frac{1}{3}u_{x,N} + \frac{2}{3}u_{x,N-1} = \frac{1}{h} \left( \frac{5}{6}u_N - \frac{1}{6}u_{N-2} - \frac{2}{3}u_{N-1} \right) + \mathcal{O}(h^3). & (b) \end{cases} \quad (2.97)$$

**Preuve.** Montrons (2.97)<sub>a</sub>. Par la méthode des coefficients indéterminés, on considère a priori au voisinage de  $x_0 = a$  une relation de la forme

$$\alpha u_{x,0} + (1 - \alpha)u_{x,1} = \beta \frac{u_1 - u_0}{h} + \gamma \frac{u_2 - u_0}{2h} + \mathcal{O}(h^n). \quad (2.98)$$

On cherche les 3 paramètres  $\alpha$ ,  $\beta$ ,  $\gamma$  tel que  $n$  soit maximum dans (2.98). On a

$$\begin{cases} \delta_x^+ u_0 = \frac{u_1 - u_0}{h}, \\ \delta_x u_1 = \frac{u_2 - u_0}{2h}. \end{cases} \quad (2.99)$$

Les formules de Taylor-Lagrange donnent les identités formelles

$$\begin{cases} u_1 = u_0 + hu_{x,0} + \frac{h^2}{2}u_0^{(2)} + \frac{h^3}{6}u_0^{(3)} + \mathcal{O}(h^4), \\ u_2 = u_0 + 2hu_{x,0} + \frac{(2h)^2}{2}u_0^{(2)} + \frac{(2h)^3}{6}u_0^{(3)} + \mathcal{O}(h^4), \\ u_{x,1} = u_{x,0} + hu_0^{(2)} + \frac{h^2}{2}u_0^{(3)} + \mathcal{O}(h^3). \end{cases} \quad (2.100)$$

On obtient le système suivant :

$$\begin{cases} \beta + \gamma = 1, \\ \alpha + \frac{\beta}{2} + \gamma = 1, \\ \frac{\alpha}{2} + \frac{\beta}{6} + \frac{2}{3}\gamma = \frac{1}{2}. \end{cases} \quad (2.101)$$

Ce qui donne la solution unique

$$\alpha = \frac{1}{3}, \beta = \frac{2}{3}, \gamma = \frac{1}{3}, \quad (2.102)$$

d'où (2.97)<sub>a</sub>. De la même manière au voisinage du point  $x_N = b$ , on obtient (2.97)<sub>b</sub>. ■

La proposition 2.5 représente la deuxième approximation des dérivées aux points du bord. Le schéma boîte hermitien considéré ici pour le problème de Poisson (2.5) est donc formé des équations (2.17) et de deux équations (2.97). Le schéma s'écrit : trouver  $(u, v) \in l_h^2 \times l_h^2$  tel que

$$\begin{cases} -\delta_x u_{x,j} = \Pi^0 f_j, & 1 \leq j \leq N-1, \quad (a) \\ \sigma_x u_{x,j} = \delta_x u_j, & 1 \leq j \leq N-1, \quad (b) \\ \frac{1}{3}u_{x,0} + \frac{2}{3}u_{x,1} = \frac{1}{h} \left( \frac{1}{6}u_2 + \frac{2}{3}u_1 - \frac{5}{6}u_0 \right), & (c) \\ \frac{1}{3}u_{x,N} + \frac{2}{3}u_{x,N-1} = \frac{1}{h} \left( \frac{5}{6}u_N - \frac{1}{6}u_{N-2} - \frac{2}{3}u_{N-1} \right), & (d) \\ u_0 = g_a, \quad u_N = g_b. & (e) \end{cases} \quad (2.103)$$

Ce schéma est appelé *Schéma HB2* dans la suite.

### 2.3.2 Expression matricielle du schéma HB2

Dans cette section, on établit la forme matricielle générale du schéma boîte hermitien pour le problème (2.5), puis on en déduit celle du schéma HB2 (2.103). Ensuite, on étudie quelques propriétés du système linéaire correspondant.

Afin d'obtenir une forme matricielle générale du schéma boîte hermitien on utilise la proposition suivante.

**Proposition 2.6.** *Pour toute approximation linéaire des dérivées au bord  $U_{x,L}$  et  $U_{x,R}$  en fonction de  $U$ ,  $U_x$  et des données au bord  $U_L$ ,  $U_R$ , il existe des matrices  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C} \in$*



$\mathbb{M}_{N-1}(\mathbb{R})$  telles que

$$e_1 U_{x,L} + e_{N-1} U_{x,R} = \frac{1}{h} \mathcal{A}U - \mathcal{B}U_x + \frac{1}{h} \mathcal{C}(e_1 U_L + e_{N-1} U_R). \quad (2.104)$$

**Preuve.** Toute approximation de  $u_{x,0}$  en fonction de  $u = (u_i)_{0 \leq i \leq N}$ ,  $u_x = (u_{x,i})_{0 \leq i \leq N}$  et  $u_0$ ,  $u_N$  possède la forme générale

$$u_{x,0} = \sum_{i=1}^{i=N-1} \alpha_i^0 u_i + \sum_{j=1}^{N-1} \beta_j^0 u_{x,j} + \gamma_0^0 u_0 + \gamma_N^0 u_N, \quad (2.105)$$

avec  $\alpha_i^0, \beta_j^0, \gamma_0^0, \gamma_N^0 \in \mathbb{R}$ .

De même, toute approximation de  $u_{x,N}$  en fonction de  $u = (u_i)_{0 \leq i \leq N}$ ,  $u_x = (u_{x,i})_{0 \leq i \leq N}$  et  $u_0$ ,  $u_N$ , possède la forme générale

$$u_{x,N} = \sum_{i=1}^{N-1} \alpha_i^N u_i + \sum_{j=1}^{N-1} \beta_j^N u_{x,j} + \gamma_0^N u_0 + \gamma_N^N u_N, \quad (2.106)$$

avec  $\alpha_i^N, \beta_j^N, \gamma_0^N, \gamma_N^N \in \mathbb{R}$ .

Multiplions (2.105) par  $e_1$  de deux côtés, on obtient

$$e_1 u_{x,0} = \sum_{i=1}^{N-1} \alpha_i^0 e_1 u_i + \sum_{j=1}^{N-1} \beta_j^0 e_1 u_{x,j} + \gamma_0^0 e_1 u_0 + \gamma_N^0 e_1 u_N. \quad (2.107)$$

Multiplions (2.106) par  $e_{N-1}$  de deux côtés, on obtient

$$e_{N-1} u_{x,N} = \sum_{i=1}^{N-1} \alpha_i^N e_{N-1} u_i + \sum_{j=1}^{N-1} \beta_j^N e_{N-1} u_{x,j} + \gamma_0^N e_{N-1} u_0 + \gamma_N^N e_{N-1} u_N. \quad (2.108)$$

En utilisant les notations vectorielles (2.28) et en sommant (2.107) et (2.108), on obtient

$$\left\{ \begin{aligned} e_1 U_{x,L} + e_{N-1} U_{x,R} &= \sum_{i=1}^{N-1} (\alpha_i^0 e_1 + \alpha_i^N e_{N-1}) U_i + \sum_{j=1}^{N-1} (\beta_j^0 e_1 + \beta_j^N e_{N-1}) U_{x,j} \\ &+ (\gamma_0^0 e_1 + \gamma_0^N e_{N-1}) U_L + (\gamma_N^0 e_1 + \gamma_N^N e_{N-1}) U_R. \end{aligned} \right. \quad (2.109)$$

Or on a  $U_i = e_i^T U$ ,  $U_{x,j} = e_j^T U_x$ ,  $U_L = e_1^T e_1 U_L$ ,  $U_R = e_{N-1}^T e_{N-1} U_R$ . Donc (2.109) donne

$$\begin{cases} e_1 U_{x,L} + e_{N-1} U_{x,R} &= \sum_{i=1}^{N-1} (\alpha_i^0 e_1 e_i^T + \alpha_i^N e_{N-1} e_i^T) U + \sum_{j=1}^{N-1} (\beta_j^0 e_1 e_j^T + \beta_j^N e_{N-1} e_j^T) U_x \\ &+ (\gamma_0^0 e_1 e_1^T + \gamma_0^N e_{N-1} e_1^T) e_1 U_L + (\gamma_N^0 e_1 e_{N-1}^T + \gamma_N^N e_{N-1} e_{N-1}^T) e_{N-1} U_R. \end{cases} \quad (2.110)$$

En notant  $\mathcal{A}, \mathcal{B}, \mathcal{C} \in M_{N-1}(\mathbb{R})$  les matrices

$$\begin{cases} \mathcal{A} = h \sum_{i=1}^{N-1} (\alpha_i^0 e_1 e_i^T + \alpha_i^N e_{N-1} e_i^T), \\ \mathcal{B} = \sum_{j=1}^{N-1} (\beta_j^0 e_1 e_j^T + \beta_j^N e_{N-1} e_j^T), \\ \mathcal{C} = h(\gamma_0^0 e_1 e_1^T + \gamma_0^N e_{N-1} e_1^T + \gamma_N^0 e_1 e_{N-1}^T + \gamma_N^N e_{N-1} e_{N-1}^T), \end{cases} \quad (2.111)$$

nous déduisons le résultat annoncé. ■

**Lemme 2.3.1.** *Le vecteur  $U_x \in \mathbb{R}^{(N-1)}$  possède la forme matricielle générale*

$$U_x = \frac{1}{h} \mathcal{D} U + \frac{1}{h} \mathcal{E} (e_1 U_L + e_{N-1} U_R), \quad (2.112)$$

où les matrices  $\mathcal{D} = \mathcal{D}(\mathcal{A}, \mathcal{B})$  et  $\mathcal{E} = \mathcal{E}(\mathcal{B}, \mathcal{C}) \in \mathbb{M}_{N-1}(\mathbb{R})$  sont définies par

$$\begin{cases} \mathcal{D} = \frac{1}{2} (P_s - \frac{1}{6} \mathcal{B})^{-1} (K - \frac{1}{3} \mathcal{A}), \\ \mathcal{E} = \frac{1}{2} (P_s - \frac{1}{6} \mathcal{B})^{-1} (F_2 - \frac{1}{3} \mathcal{C}). \end{cases} \quad (2.113)$$

**Preuve.** En remplaçant dans (2.34) le vecteur  $e_1 U_{x,L} + e_{N-1} U_{x,R} \in \mathbb{R}^{(N-1)}$  par sa forme matricielle générale en (2.104) on obtient

$$P_s U_x + \frac{1}{6h} \mathcal{A} U - \frac{1}{6} \mathcal{B} U_x + \frac{1}{6h} (e_1 U_L + e_{N-1} U_R) = \frac{1}{2h} K U + \frac{1}{2h} F_2 (e_1 U_L + e_{N-1} U_R). \quad (2.114)$$

Donc

$$\begin{aligned} U_x &= \frac{1}{2h} (P_s - \frac{1}{6} \mathcal{B})^{-1} (K - \frac{1}{3} \mathcal{A}) U \\ &+ \frac{1}{2h} (P_s - \frac{1}{6} \mathcal{B})^{-1} (F_2 - \frac{1}{3} \mathcal{C}) (e_1 U_L + e_{N-1} U_R). \end{aligned}$$

Ce qu'il fallait démontrer. ■

**Proposition 2.7.** *En utilisant l'approximation (2.104), la forme matricielle générale du schéma HB associée au système (2.11) est : Trouver  $U \in \mathbb{R}^{(N-1)}$  tel que*

$$\frac{1}{h^2}\mathcal{H}U = F - \frac{1}{h^2}\mathcal{G}(e_1U_L + e_{N-1}U_R), \quad (2.115)$$

où les matrices  $\mathcal{H} = \mathcal{H}(\mathcal{A}, \mathcal{B})$  et  $\mathcal{G} = \mathcal{G}(\mathcal{B}, \mathcal{C}) \in \mathbb{M}_{N-1}(\mathbb{R})$  sont

$$\begin{cases} \mathcal{H} = -\frac{1}{2}\left(K\mathcal{D} + F_2(\mathcal{A} - \mathcal{B}\mathcal{D})\right), \\ \mathcal{G} = -\frac{1}{2}\left(K\mathcal{E} + F_2(\mathcal{C} - \mathcal{B}\mathcal{E})\right), \end{cases} \quad (2.116)$$

ou de façon équivalente en utilisant (2.113),

$$\begin{cases} \mathcal{H} = -\frac{1}{4}(K - F_2\mathcal{B})(P_s - \frac{1}{6}\mathcal{B})^{-1}(K - \frac{1}{3}\mathcal{A}) - \frac{1}{2}F_2\mathcal{A}, & (a) \\ \mathcal{G} = -\frac{1}{4}(K - F_2\mathcal{B})(P_s - \frac{1}{6}\mathcal{B})^{-1}(F_2 - \frac{1}{3}\mathcal{C}) - \frac{1}{2}F_2\mathcal{C}. & (b) \end{cases} \quad (2.117)$$

**Preuve.** Après avoir remplacé le vecteur  $e_1U_{x,L} + e_{N-1}U_{x,R} \in \mathbb{R}^{(N-1)}$  dans (2.29) par sa forme matricielle générale en (2.104), on obtient

$$-\frac{1}{2h}KU_x - \frac{1}{2h^2}F_2\mathcal{A}U + \frac{1}{2h}F_2\mathcal{B}U_x - \frac{1}{2h^2}F_2(e_1U_L + e_{N-1}U_R) = F. \quad (2.118)$$

En regroupant les termes, on obtient

$$-\frac{1}{2h}(K - F_2\mathcal{B})U_x - \frac{2}{h^2}F_2\mathcal{A}U - \frac{1}{2h^2}F_2\mathcal{C}(e_1U_L + e_{N-1}U_R). \quad (2.119)$$

En substituant  $U_x$  par sa valeur en (2.112), ceci donne

$$-\frac{1}{h^2}(K\mathcal{D} + F_2\mathcal{A} - F_2\mathcal{B}\mathcal{D})U + \frac{1}{h^2}(K\mathcal{E} - F_2\mathcal{B}\mathcal{E} + F_2\mathcal{C})(e_1U_L + e_{N-1}U_R) = F, \quad (2.120)$$

d'où (2.115). ■

La proposition 2.7 donne la forme matricielle du schéma HB associée à (2.11) une fois que les matrices  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$ , sont connues. Ces dernières correspondent aux approximations des dérivées au bord. Dans la première approximation (2.26), la forme matricielle du schéma HB1 (2.47) est obtenue en utilisant les matrices  $\mathcal{A}_1$ ,  $\mathcal{B}_1$ ,  $\mathcal{C}_1$  données en (2.43). Pour la deuxième approximation (2.97) on commence par le lemme suivant.

**Lemme 2.3.2.** *Les conditions aux limites (2.97) s'écrivent matriciellement sous la forme*

$$e_1 U_{x,L} + e_{N-1} U_{x,R} = \frac{1}{h} \mathcal{A}_2 U - \mathcal{B}_2 U_x + \frac{1}{h} \mathcal{C}_2 (e_1 U_L + e_{N-1} U_R), \quad (2.121)$$

où les matrices  $\mathcal{A}_2$ ,  $\mathcal{B}_2$ ,  $\mathcal{C}_2 \in \mathbb{M}_{N-1}(\mathbb{R})$  sont définies par

$$\begin{cases} \mathcal{A}_2 = -2F_2 + \frac{1}{2}(e_1 e_2^T - e_{N-1} e_{N-2}^T), & (a) \\ \mathcal{B}_2 = 2F_1, & (b) \\ \mathcal{C}_2 = \frac{5}{2}F_2, & (c) \end{cases} \quad (2.122)$$

et les matrices  $F_1$ ,  $F_2 \in \mathbb{M}_{N-1}(\mathbb{R})$  sont données en (1.70) et  $g_a$ ,  $g_b$  sont les conditions de Dirichlet au bord.

**Preuve.** D'après (2.97) les dérivées  $u_{x,0}$  et  $u_{x,N}$  s'écrivent

$$\begin{cases} u_{x,0} = \frac{1}{h}(2e_1 + \frac{1}{2}e_2)^T U - 2e_1^T U_x - \frac{5}{2}e_1^T (e_1 g_a + e_{N-1} g_b), \\ u_{x,N} = \frac{1}{h}(-2e_{N-1} - \frac{1}{2}e_{N-2})^T U - 2e_{N-1}^T U_x + \frac{5}{2}e_{N-1}^T (e_1 g_a + e_{N-1} g_b). \end{cases} \quad (2.123)$$

On a  $g_a = U_L$ ,  $g_b = U_R$ , (2.123) se réécrit

$$\begin{aligned} [u_{x,0}, 0, \dots, 0, u_{x,N}]^T &= \frac{1}{h}(2e_1 e_1^T + \frac{1}{2}e_1 e_2^T - 2e_{N-1} e_{N-1}^T - \frac{1}{2}e_{N-1} e_{N-2}^T)U \\ &\quad - 2(e_1 e_1^T + e_{N-1} e_{N-1}^T)U_x \\ &\quad + \frac{5}{2}(-e_1 e_1^T + e_{N-1} e_{N-1}^T)[g_a, 0, \dots, 0, g_b]^T \\ &= \frac{1}{h} \mathcal{A}_2 U - \frac{1}{2} \mathcal{B}_2 U_x + \frac{1}{h} \mathcal{C}_2 (e_1 U_L + e_{N-1} U_R), \end{aligned}$$

où les matrices  $\mathcal{A}_2$ ,  $\mathcal{B}_2$ ,  $\mathcal{C}_2 \in \mathbb{M}_{N-1}(\mathbb{R})$  sont définies par

$$\begin{cases} \mathcal{A}_2 = (2e_1 e_1^T + \frac{1}{2}e_1 e_2^T - 2e_{N-1} e_{N-1}^T - \frac{1}{2}e_{N-1} e_{N-2}^T), & (a) \\ \mathcal{B}_2 = 2(e_1 e_1^T + e_{N-1} e_{N-1}^T) = 2F_1, & (b) \\ \mathcal{C}_2 = \frac{5}{2}(-e_1 e_1^T + e_{N-1} e_{N-1}^T) = \frac{5}{2}F_2. & (c) \end{cases} \quad (2.124)$$

On en déduit (2.121). ■

Noter que les deux matrices  $\mathcal{A}_2$ ,  $\mathcal{B}_2$  vérifient les propriétés suivantes :

$$\begin{cases} \mathcal{A}_2 = \mathcal{A}_1 + \delta \mathcal{A}_1, & (a) \\ \mathcal{B}_2 = \mathcal{B}_1 + \delta \mathcal{B}_1, & (b) \end{cases} \quad (2.125)$$

avec

$$\begin{cases} \delta\mathcal{A}_1 = \frac{1}{2}(e_1e_1^T - e_{N-1}e_{N-1}^T + e_1e_2^T - e_{N-1}e_{N-2}^T) = -\frac{1}{2}F_2 + \frac{1}{2}(e_1e_2^T - e_{N-1}e_{N-2}^T), & (a) \\ \delta\mathcal{B}_1 = \frac{3}{2}(e_1e_1^T + e_{N-1}e_{N-1}^T) = \frac{3}{2}F_1. & (b) \end{cases} \quad (2.126)$$

Une fois que l'on dispose des matrices  $\mathcal{A}_2$ ,  $\mathcal{B}_2$  et  $\mathcal{C}_2$ , en utilisant (2.117), les matrices  $\mathcal{H}_2 = \mathcal{H}_2(\mathcal{A}_2, \mathcal{B}_2)$ ,  $\mathcal{G}_2 = \mathcal{G}_2(\mathcal{B}_2, \mathcal{C}_2)$  associées au schéma HB2 sont déduites de (2.117)

$$\begin{cases} \mathcal{H}_2 = -\frac{1}{4}(K - F_2\mathcal{B}_2)(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}(K - \frac{1}{3}\mathcal{A}_2) - \frac{1}{2}F_2\mathcal{A}_2, & (a) \\ \mathcal{G}_2 = -\frac{1}{4}(K - F_2\mathcal{B}_2)(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}(F_2 - \frac{1}{3}\mathcal{C}_2) - \frac{1}{2}F_2\mathcal{C}_2. & (b) \end{cases} \quad (2.127)$$

(2.113) donne

$$\begin{cases} \mathcal{D}_2 = \frac{1}{2}(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}(K - \frac{1}{3}\mathcal{A}_2), \\ \mathcal{E}_2 = \frac{1}{2}(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}(F_2 - \frac{1}{3}\mathcal{C}_2). \end{cases} \quad (2.128)$$

Enfin, d'après la proposition 2.7, le schéma HB2 (2.103) s'écrit en fonction de  $U \in \mathbb{R}^{(N-1)}$  sous la forme du système linéaire suivant :

$$\frac{1}{h^2}\mathcal{H}_2U = F - \frac{1}{h^2}\mathcal{G}_2(e_1U_L + e_{N-1}U_R). \quad (2.129)$$

D'après le lemme 2.3.1 le vecteur  $U_x \in \mathbb{R}^{(N-1)}$  vérifie

$$U_x = \frac{1}{h}\mathcal{D}_2U + \frac{1}{h}\mathcal{E}_2(e_1U_L + e_{N-1}U_R). \quad (2.130)$$

### 2.3.3 Principe du maximum pour le schéma HB2

Le schéma HB2 (2.103) vérifie le principe du maximum et du minimum discret suivant.

**Proposition 2.8** (Principe du maximum et du minimum). *Soit  $u = (u_i)_{0 \leq i \leq N} \in l_h^2$  la solution du schéma HB2 (2.103). Supposons que  $f_i \leq 0$ ,  $\forall 1 \leq i \leq N-1$ , alors*

$$\max_{1 \leq i \leq N-1} u_i \leq \max\{u_0, u_N\}. \quad (\text{Principe du maximum}) \quad (2.131)$$

De même, si  $f_i \geq 0$ ,  $\forall 1 \leq i \leq N-1$ ,  $u$  vérifie

$$\min_{1 \leq i \leq N-1} u_i \geq \min\{u_0, u_N\}. \quad (\text{Principe du minimum}) \quad (2.132)$$

Cela signifie que  $u$  atteint ses valeurs maxima et minima aux points de bord  $x_0$  et  $x_N$ .

**Preuve.** Supposons que  $f_i \leq 0, \forall 1 \leq i \leq N - 1$  alors

$$\Pi^0 f_i \leq 0, \quad 1 \leq i \leq N - 1. \quad (2.133)$$

D'après (2.20)

$$-\frac{u_{i+2} + u_{i-2} - 2u_i}{4h^2} \leq 0, \quad \forall 2 \leq i \leq N - 2. \quad (2.134)$$

Si  $N$  est pair, d'après le principe du maximum du lemme 2.2.8 on obtient

$$\begin{cases} \max_{1 \leq k \leq \frac{N-2}{2}} u_{2k} \leq \max\{u_0, u_N\}, \\ \max_{1 \leq k \leq \frac{N-2}{2}} u_{2k+1} \leq \max\{u_1, u_{N-1}\}. \end{cases} \quad (2.135)$$

De la même façon si  $N$  est impair, d'après le lemme 2.2.8 on a

$$\begin{cases} \max_{1 \leq k \leq \frac{N-1}{2}} u_{2k} \leq \max\{u_0, u_{N-1}\}, \\ \max_{1 \leq k \leq \frac{N-1}{2}} u_{2k+1} \leq \max\{u_1, u_N\}. \end{cases} \quad (2.136)$$

Il en résulte que pour tout  $N$ ,

$$\max_{2 \leq i \leq N-2} u_i \leq \max\{u_0, u_1, u_{N-1}, u_N\}. \quad (2.137)$$

Donc pour démontrer (2.131) il suffit de démontrer que

$$\max\{u_1, u_{N-1}\} \leq \max\{u_0, u_N\}. \quad (2.138)$$

Supposons par exemple que

$$\max\{u_1, u_{N-1}\} = u_1. \quad (2.139)$$

On a

$$-\delta_x u_{x,j} = \Pi^0 f_j \leq 0, \quad \forall 1 \leq j \leq N - 1. \quad (2.140)$$

Cela donne  $u_{x,0} \leq u_{x,2}$  et  $u_{x,1} \leq u_{x,3}$ . Le schéma HB2 (2.103) vérifie

$$\begin{cases} \frac{1}{3}u_{x,0} + \frac{2}{3}u_{x,1} = \frac{2}{3}\frac{u_1 - u_0}{h} + \frac{1}{3}\frac{u_2 - u_0}{2h}, \\ \frac{1}{6}u_{x,0} + \frac{2}{3}u_{x,1} + \frac{1}{6}u_{x,2} = \frac{u_2 - u_0}{2h}. \end{cases} \quad (2.141)$$

De plus

$$\frac{1}{3}u_{x,0} + \frac{2}{3}u_{x,1} \leq \frac{1}{6}u_{x,0} + \frac{2}{3}u_{x,1} + \frac{1}{6}u_{x,2} \leq \frac{u_2 - u_0}{2h}.$$

Donc

$$\frac{2}{3} \frac{u_1 - u_0}{h} + \frac{1}{3} \frac{u_2 - u_0}{2h} \leq \frac{u_2 - u_0}{2h}.$$

En utilisant  $u_2 \leq u_0$ , ceci donne

$$\frac{2}{3} \frac{u_1 - u_0}{h} \leq \frac{2}{3} \frac{u_2 - u_0}{2h} \leq 0, \quad (2.142)$$

donc

$$u_1 \leq u_0, \quad (2.143)$$

d'où (2.138). Supposons à présent que

$$\max\{u_1, u_{N-1}\} = u_{N-1}. \quad (2.144)$$

Les équations (2.103)<sub>b,d</sub> donnent

$$\begin{cases} \frac{1}{3}u_{x,N} + \frac{2}{3}u_{x,N-1} = \frac{2}{3} \frac{u_N - u_{N-1}}{h} + \frac{1}{3} \frac{u_N - u_{N-2}}{2h}, \\ \frac{1}{6}u_{x,N-2} + \frac{2}{3}u_{x,N-1} + \frac{1}{6}u_{x,N} = \frac{u_N - u_{N-2}}{2h}. \end{cases} \quad (2.145)$$

De plus d'après (2.140),  $u_{x,N-2} \leq u_{x,N}$ , donc

$$\frac{1}{3}u_{x,N} + \frac{2}{3}u_{x,N-1} \geq \frac{1}{6}u_{x,N-2} + \frac{2}{3}u_{x,N-1} + \frac{1}{6}u_{x,N} = \frac{u_N - u_{N-2}}{2h}, \quad (2.146)$$

car  $u_{N-1} \geq u_{N-3}$ , donc

$$\frac{2}{3} \frac{u_N - u_{N-1}}{h} + \frac{1}{3} \frac{u_N - u_{N-2}}{2h} \geq \frac{u_N - u_{N-2}}{2h}, \quad (2.147)$$

ce qui donne

$$\frac{2}{3} \frac{u_N - u_{N-1}}{h} \geq \frac{2}{3} \frac{u_N - u_{N-2}}{2h} \geq 0, \quad (2.148)$$

car  $u_N \geq u_{N-2}$ . On en déduit finalement

$$u_{N-1} \leq u_N, \quad (2.149)$$

d'où

$$\max\{u_1, u_{N-1}\} \leq \max\{u_0, u_N\}. \quad (2.150)$$

Enfin

$$\max_{1 \leq i \leq N-1} u_i \leq \max\{u_0, u_N\}, \quad (2.151)$$

d'où (2.131). De la même manière on démontre (2.132). ■

**Corollaire 2.3.3** (existence et unicité). *Le schéma HB2 (2.103) admet une solution unique  $(u, v) \in l_{h,0}^2 \times l_h^2$ .*

**Preuve.** (2.103) est un système linéaire de  $2N$  équations à  $2N$  inconnues. Pour démontrer l'existence et l'unicité de la solution, il suffit de démontrer  $\Pi^0 f_i = 0$ ,  $1 \leq i \leq N-1$ , implique  $u = v = 0$ . Soit  $(u, v) \in l_{h,0}^2 \times l_h^2$  tel que

$$\left\{ \begin{array}{ll} -\delta_x v_j = 0, & 1 \leq j \leq N-1, \quad (a) \\ \frac{1}{3}v_N + \frac{1}{3}v_{N-1} = \frac{2}{3}\frac{u_N - u_{N-1}}{h} + \frac{1}{3}\frac{u_N - u_{N-2}}{2h}, & j = N, \quad (b) \\ \frac{1}{6}v_{j-1} + \frac{2}{3}v_j + \frac{1}{6}v_{j+1} = \frac{u_{j+1} - u_{j-1}}{2h}, & 1 \leq j \leq N-1, \quad (c) \\ \frac{1}{3}v_0 + \frac{2}{3}v_1 = \frac{2}{3}\frac{u_1 - u_0}{h} + \frac{1}{3}\frac{u_2 - u_0}{2h}, & j = 0, \quad (d) \\ u_0 = 0, u_N = 0. & (e) \end{array} \right. \quad (2.152)$$

En utilisant le principe du maximum discret de la proposition 2.8, on obtient

$$u_i = 0, \quad \forall 1 \leq i \leq N-1, \quad (2.153)$$

donc  $u = 0$ , alors  $U = \text{vec}_1(u) = 0 \in \mathbb{R}^{(N-1)}$ , ce qui prouve que la matrice  $\mathcal{H}_2$  associée au schéma HB2 est de noyau nul ( $\text{Ker}(\mathcal{H}_2) = \{0\}$ ). Puisque c'est une matrice carrée, elle est inversible, d'où l'existence et l'unicité de la solution du schéma HB2. ■

### 2.3.4 Relation entre les schémas HB1 et HB2

Commençons par rappeler le résultat d'algèbre linéaire suivant.

**Formule 1** (Formule de Sherman-Morrison-Woodbury). *Soit  $N \geq 1$ . Supposons que  $A, B \in \mathbb{M}_N(\mathbb{R})$  soient deux matrices inversibles telles que*

$$A = B + RS^T, \quad (2.154)$$



avec  $R$ , et  $S \in \mathbb{M}_{N,n}(\mathbb{R})$ ,  $n \leq N$ . Alors l'inverse de la matrice  $A$  peut être écrit comme

$$A^{-1} = B^{-1} - B^{-1}R(I + S^T B^{-1}R)^{-1}S^T B^{-1}. \quad (2.155)$$

Dans le cas où  $n \ll N$ , la matrice  $A$  est une perturbation de la matrice  $B$  de la forme  $A = B + \delta B$  avec  $\text{rg}(\delta B)$  "petit". Dans le cas où on peut facilement calculer l'inverse de la matrice  $B$ , la formule (2.155) donne une méthode efficace pour inverser  $A$ . On utilisera cette formule pour inverser les matrices associées au schéma HB2 en deux et trois dimensions dans la suite. La formule (2.154) suggère l'algorithme de résolution de  $AU = b$  suivant qui utilise (2.155).

**Algorithme 1. - Etape 1 :** Calcul de  $V_1 = B^{-1}b$ . Stocker  $V_1$  dans  $V$ .

- **Etape 2 :** Calcul de  $V_2 = S^T V_1$ .

- **Etape 3 :** Calcul de  $V_3 = (I + S^T B^{-1}R)^{-1}V_2$ .

- **Etape 4 :** Calcul de  $V_4 = RV_3$ .

- **Etape 5 :** Calcul de  $V_5 = B^{-1}V_4$ .

- **Etape 6 :** Calcul de  $U = V - V_5$ .

**Proposition 2.9 (Lien entre les schémas HB1 et HB2).** La matrice  $\mathcal{H}_2$  s'écrit comme une perturbation de la matrice  $\mathcal{H}_1$  (2.53) sous la forme

$$\mathcal{H}_2 = \mathcal{H}_1 + \delta\mathcal{H}, \quad (2.156)$$

où  $\delta\mathcal{H}$  est la matrice de rang 6,

$$\delta\mathcal{H} = RS^T, \quad (2.157)$$

avec  $R, S \in M_{N-1,6}(\mathbb{R})$  sont donnés en (2.174) et (2.172).

Pour démontrer la proposition 2.9, on se sert du lemme suivant.

**Lemme 2.3.4.** L'inverse de la matrice  $(P_s - \frac{1}{6}\mathcal{B}_2)$  est donné par

$$(P_s - \frac{1}{6}\mathcal{B}_2)^{-1} = (P_s - \frac{1}{6}\mathcal{B}_1)^{-1} + \begin{bmatrix} v_1, v_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}, \quad (2.158)$$

où les deux vecteurs  $v_1, v_2 \in \mathbb{R}^{N-1}$  sont

$$\begin{cases} v_1 = \frac{1}{2(\alpha + \beta)^{\frac{1}{2}}} (P_s - \frac{1}{6}\mathcal{B}_1)^{-1} \left( \frac{\sqrt{2}}{2}e_1 - \frac{\sqrt{2}}{2}e_{N-1} \right), \\ v_2 = \frac{1}{2(\alpha - \beta)^{\frac{1}{2}}} (P_s - \frac{1}{6}\mathcal{B}_1)^{-1} \left( \frac{\sqrt{2}}{2}e_1 + \frac{\sqrt{2}}{2}e_{N-1} \right), \end{cases} \quad (2.159)$$

et les scalaires  $\alpha, \beta$  sont

$$\begin{cases} \alpha = 1 - \frac{1}{4}e_1^T (P_s - \frac{1}{6}\mathcal{B}_1)^{-1} e_1, \\ \beta = \frac{1}{4}e_{N-1}^T (P_s - \frac{1}{6}\mathcal{B}_1)^{-1} e_1. \end{cases} \quad (2.160)$$

#### Preuve du lemme 2.3.4.

La preuve est basée sur la formule de Sherman-Morrison-Woodbury (2.155). En utilisant l'expression de la matrice  $\mathcal{B}_2$  en (2.125)<sub>b</sub>, on obtient :

$$(P_s - \frac{1}{6}\mathcal{B}_2)^{-1} = (P_s - \frac{1}{6}\mathcal{B}_1 - \frac{1}{4}[e_1, e_{N-1}][e_1, e_{N-1}]^T)^{-1}. \quad (2.161)$$

Puisque  $[e_1, e_{N-1}]$  est une matrice de rang 2, la formule de Sherman-Morrison-Woodbury permet de calculer l'inverse (2.161) par

$$\begin{aligned} (P_s - \frac{1}{6}\mathcal{B}_1 - \frac{1}{4}[e_1, e_{N-1}][e_1, e_{N-1}]^T)^{-1} &= (P_s - \frac{1}{6}\mathcal{B}_1)^{-1} \\ &\quad - \frac{1}{4}(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}[e_1, e_{N-1}](A^{-1})[e_1, e_{N-1}]^T(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}, \end{aligned}$$

où la matrice  $A \in \mathbb{M}_2(\mathbb{R})$  est

$$A = (I_2 - \frac{1}{4}[e_1, e_{N-1}]^T(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}[e_1, e_{N-1}])^{-1}, \quad (2.162)$$

avec  $I_2$  la matrice identité d'ordre 2.

On écrit

$$A = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix}. \quad (2.163)$$

Les scalaires  $\alpha, \beta$  sont

$$\begin{cases} \alpha = 1 - \frac{1}{4}e_1^T (P_s - \frac{1}{6}\mathcal{B}_1)^{-1} e_1, \\ \beta = \frac{1}{4}e_{N-1}^T (P_s - \frac{1}{6}\mathcal{B}_1)^{-1} e_1. \end{cases} \quad (2.164)$$

L'inverse de la matrice d'ordre  $A$  est donné par

$$A^{-1} = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix}^{-1} = \frac{1}{\alpha^2 - \beta^2} \begin{bmatrix} \alpha & -\beta \\ -\beta & \alpha \end{bmatrix},$$

$A^{-1}$  se met sous forme diagonale par

$$\frac{1}{\alpha^2 - \beta^2} \begin{bmatrix} \alpha & -\beta \\ -\beta & \alpha \end{bmatrix} = \frac{1}{\alpha^2 - \beta^2} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \alpha + \beta & 0 \\ 0 & \alpha - \beta \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

On obtient finalement

$$-\frac{1}{4}(P_s - \frac{1}{6}\mathcal{B}_1)^{-1}[e_1, e_{N-1}](A^{-1})[e_1, e_{N-1}]^T(P_s - \frac{1}{6}\mathcal{B}_1)^{-1} = \begin{bmatrix} v_1, v_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix},$$

où les vecteurs  $v_1, v_2$  sont donnés en (2.159). On en déduit (2.161). ■

**Preuve de la proposition 2.9.** En utilisant (1.71) et (2.125) on obtient

$$\begin{cases} F_2\mathcal{A}_2 = F_2\mathcal{A}_1 - \frac{1}{2}F_1 - \frac{1}{2}(e_1e_2^T + e_{N-1}e_{N-2}^T), \\ F_2\mathcal{B}_2 = F_2\mathcal{B}_1 + \frac{3}{2}F_2. \end{cases} \quad (2.165)$$

Donc la matrice  $\mathcal{H}_2$  en (2.127<sub>a</sub>) peut s'écrire sous la forme

$$\mathcal{H}_2 = \mathcal{H}_2^{(a)} + \mathcal{H}_2^{(b)}, \quad (2.166)$$

avec

$$\begin{cases} \mathcal{H}_2^{(a)} = -\frac{1}{4}(K - F_2\mathcal{B}_1)(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}(K - \frac{1}{3}\mathcal{A}_1) + \frac{3}{8}F_2(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}(K - \frac{1}{3}\mathcal{A}_1) - \frac{1}{2}F_2\mathcal{A}_1, \\ \mathcal{H}_2^{(b)} = \frac{1}{4}F_1 + \frac{1}{4}(e_1e_2^T + e_{N-1}e_{N-2}^T) - \frac{1}{24}(K - F_2\mathcal{B}_2)(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}(F_2 - e_1e_2^T + e_{N-1}e_{N-2}^T). \end{cases} \quad (2.167)$$

En remplaçant  $(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}$  par sa valeur donnée en (2.161), on obtient

$$\mathcal{H}_2^{(a)} = \mathcal{H}_1 + w_1w_1^T + w_2w_2^T + e_1s_1^T + e_{N-1}s_2^T, \quad (2.168)$$

et

$$\mathcal{H}_2^{(b)} = r_1(e_1 + e_2)^T + r_2(e_{N-1} + e_{N-2})^T. \quad (2.169)$$

Les six vecteurs  $w_1, w_2, r_1, r_2, s_1, s_2 \in \mathbb{R}^{(N-1)}$  sont

$$\begin{cases} w_1 = -\frac{1}{2}(K + \frac{1}{3}\mathcal{A}_1)v_1, \\ w_2 = -\frac{1}{2}(K + \frac{1}{3}\mathcal{A}_1)v_2, \end{cases} \quad (2.170)$$

$$\begin{cases} r_1 = -\frac{1}{24}(-K + \frac{1}{4}F_2\mathcal{B}_2)(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}e_1 + \frac{1}{4}e_1, \\ r_2 = \frac{1}{24}(-K + \frac{1}{4}F_2\mathcal{B}_2)(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}e_{N-1} + \frac{1}{4}e_{N-1}, \end{cases} \quad (2.171)$$

$$\begin{cases} s_1 = -\frac{3}{8}(-K - \frac{1}{3}\mathcal{A}_2^T)(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}e_1, \\ s_2 = \frac{3}{8}(-K - \frac{1}{3}\mathcal{A}_2^T)(P_s - \frac{1}{6}\mathcal{B}_2)^{-1}e_{N-1}. \end{cases} \quad (2.172)$$

Les scalaires  $\alpha, \beta$  et les vecteurs  $v_1, v_2 \in \mathbb{R}^{N-1}$  sont donnés en (2.159) et (2.160). En sommant (2.168) et (2.169) on obtient

$$\mathcal{H}_2 = \mathcal{H}_1 + RS^T, \quad (2.173)$$

où  $RS^T$  est la perturbation définie par

$$\begin{cases} R = [w_1, w_2, e_1, e_{N-1}, r_1, r_2] \in \mathbb{M}_{N-1,6}(\mathbb{R}), \\ S = [w_1, w_2, s_1, s_2, e_1 + e_2, e_{N-1} + e_{N-2}] \in \mathbb{M}_{N-1,6}(\mathbb{R}). \end{cases} \quad (2.174)$$

■

Noter que la non-symétrie de la matrice  $\mathcal{H}_2$  est uniquement due à la matrice de perturbation  $\delta\mathcal{H} = RS^T$ , qui reflète les conditions de bord. De plus, la forme (2.173) de la matrice  $\mathcal{H}_2$  joue un rôle très important en deux et trois dimensions dans les chapitres 3 et 5 pour obtenir un solveur très rapide .

### 2.3.5 Résultats numériques

Afin de tester le schéma HB2 en dimension 1, on donne d'abord quelques résultats numériques pour les exemples déjà traités dans la section 2.2.5. La solution exacte  $u_{ex}(x)$  fournit les conditions de Dirichlet au bord. On résoud le système linéaire (2.129). Le calcul est effectué en Matlab en utilisant la méthode directe pour inverser la matrice  $\mathcal{H}_2$  (2.48). Les erreurs entre la solution exacte  $u_{ex}$  et la solution calculée  $u$  ainsi que l'erreur entre le gradient exacte  $u_{x,ex}$  et le gradient calculé  $u_x$ , sont calculés en utilisant

$N$	$\ u - u_{ex}\ _h$	$\ u - u_{ex}\ _\infty$	$\ u_x - u_{x,ex}\ _h$	$\ u_x - u_{x,ex}\ _\infty$
128	1.221(-7)	2.451(-7)	3.269(-7)	4.915(-7)
Taux de conv.	3.95	3.95	3.95	3.95
256	7.854(-9)	1.578(-8)	2.114(-8)	3.171(-8)
Taux de conv.	3.98	3.98	3.97	3.97
512	4.978(-10)	1(-9)	1.343(-9)	2.031(-9)
Taux de conv.	3.99	3.99	3.98	3.98
1024	3.134(-11)	6.301(-11)	8.466(-11)	1.268(-10)
Taux de conv.	3.99	3.99	3.99	3.99
2048	1.970(-12)	3.961(-12)	5.314(-12)	7.961(-12)

TAB. 2.5: Erreur et taux de convergence de Schéma HB2 avec  $u_{ex}(x) = \cos(\pi x^2)$ 

$N$	$\ u - u_{ex}\ _h$	$\ u - u_{ex}\ _\infty$	$\ u_x - u_{x,ex}\ _h$	$\ u_x - u_{x,ex}\ _\infty$
128	7.838(-7)	1.590(-6)	2.637(-6)	3.733(-6)
Taux de conv.	4.00	4.00	4.01	4.00
256	4.895(-8)	9.922(-8)	1.631(-7)	2.322(-7)
Taux de conv.	4.00	4.00	4.00	4.00
512	3.056(-9)	6.191(-9)	1.013(-8)	1.446(-8)
Taux de conv.	4.00	4.00	4.00	4.00
1024	1.908(-10)	3.865(-10)	6.315(-10)	9.021(-10)
Taux de conv.	3.99	3.99	4.00	4.00
2048	1.193(-11)	2.416(-11)	3.941(-11)	5.634(-11)

TAB. 2.6: Erreur et taux de convergence de Schéma HB2 avec  $u_{ex}(x) = \sin(\pi x^3)$ 

les normes  $L^2$  et  $L^\infty$  discrètes définies par

$$\begin{cases} \|u - u_{ex}\|_h = \left[ h \sum_{i=1}^{N-1} (u_i - u_{ex}(x_i))^2 \right]^{\frac{1}{2}}, & (a) \\ \|u - u_{ex}\|_\infty = \max_{i=1, \dots, N-1} |(u_i - u_{ex}(x_i))|. & (b) \end{cases} \quad (2.175)$$

**Test 1 :** On considère la solution exacte régulière  $u_{ex}(x) = \cos(\pi x^2)$  (conditions de Dirichlet non homogènes). La table 2.5 montre que le schéma HB2 est d'ordre 4 pour les deux inconnues,  $u$  et  $u_x$ . Le schéma HB1 est d'ordre 2 dans ce cas (voir figure 2.4).

**Test 2 :** On considère la fonction exacte  $u_{ex}(x) = \sin(\pi x^3)$ . Les résultats sont indiqués dans la table 2.6. On observe la convergence d'ordre 4 du schéma HB2 pour les deux inconnues  $u$  et  $u_x$ . Le schéma HB1 est aussi d'ordre 4 pour ce cas (voir figure 2.5).

**Test 3 :** Le schéma HB2 est d'ordre 4 pour les deux inconnues  $u$  et  $u_x$ , avec la solution

$N$	$\ u - u_{ex}\ _h$	$\ u - u_{ex}\ _\infty$	$\ u_x - u_{x,ex}\ _h$	$\ u_x - u_{x,ex}\ _\infty$
128	2.674(-10)	4.645(-10)	5.131(-10)	6.771(-10)
Taux de conv.	3.97	4.00	4.00	4.00
256	1.702(-11)	2.903(-11)	3.196(-11)	4.231(-11)
Taux de conv.	3.98	4.00	4.00	3.99
512	1.074(-12)	1.813(-12)	1.994(-12)	2.645(-12)
Taux de conv.	3.93	3.95	3.99	3.95
1024	7.034(-14)	1.170(-13)	1.247(-13)	1.705(-13)
Taux de conv.	3.68	3.66	4.01	3.89
2048	5.471(-15)	9.214(-15)	7.726(-15)	1.143(-14)

TAB. 2.7: Erreur et taux de convergence de Schéma HB2 avec  $u_{ex}(x) = \exp(-(x - \frac{1}{2})^2)$ 

$N$	$\ u - u_{ex}\ _h$	$\ u - u_{ex}\ _\infty$	$\ u_x - u_{x,ex}\ _h$	$\ u_x - u_{x,ex}\ _\infty$
128	9.054(-5)	5.374(-4)	4.400(-3)	2.290(-2)
Taux de conv.	4.07	4.08	4.10	4.13
256	5.365(-6)	3.162(-5)	2.554(-4)	1.300(-3)
Taux de conv.	4.01	4.02	4.02	3.95
512	3.312(-7)	1.949(-6)	1.572(-5)	8.408(-5)
Taux de conv.	4.00	4.00	4.00	4.00
1024	2.063(-8)	1.214(-7)	9.790(-7)	5.223(-5)
Taux de conv.	4.00	4.00	4.00	4.00
2048	1.288(-9)	7.581(-9)	6.113(-8)	3.267(-7)

TAB. 2.8: Erreur et taux de convergence de Schéma HB2 avec  $u_{ex}(x) = \exp(-1000(x - \frac{1}{2})^2)$ 

exacte  $u_{ex}(x) = \exp(-(x - \frac{1}{2})^2)$ . Les résultats numériques sont donnés dans la table 2.7. Le schéma HB1 est d'ordre 2 dans ce cas (figure 2.6).

**Test 4 :** On considère la solution exacte  $u_{ex}(x) = \exp(-1000(x - \frac{1}{2})^2)$ . Les résultats numériques sont donnés dans la table 2.8. Pour ce cas, le schéma HB1 donne les mêmes résultats numériques que le schéma HB2 (voir figure 2.7).

Afin d'illustrer graphiquement la vitesse de convergence de chacun de ces schémas, on présente les courbes de convergence sur la même figure pour chacun des quatre cas testés. Dans les tests 1, 2, et 3, le schéma HB1 et le schéma différence finie standard (DF2) possèdent presque les mêmes courbes de convergence. Cependant, le schéma HB1 calcule la dérivée ce qui n'est pas le cas pour le schéma classique. Dans le test 4,

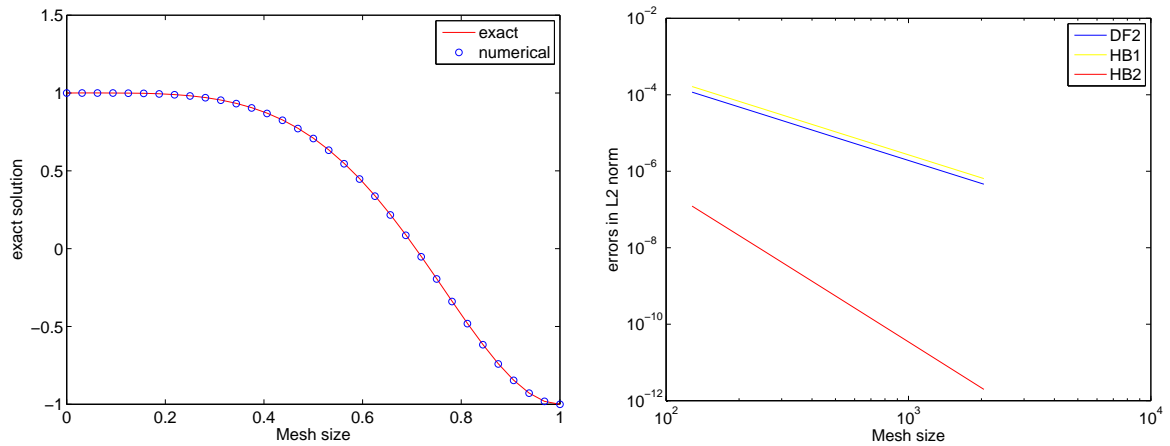


FIG. 2.4: A gauche, les solutions exacte et numérique avec  $u_{ex}(x) = \cos(\pi x^2)$ . A droite, les courbes de convergence de 3 schémas.

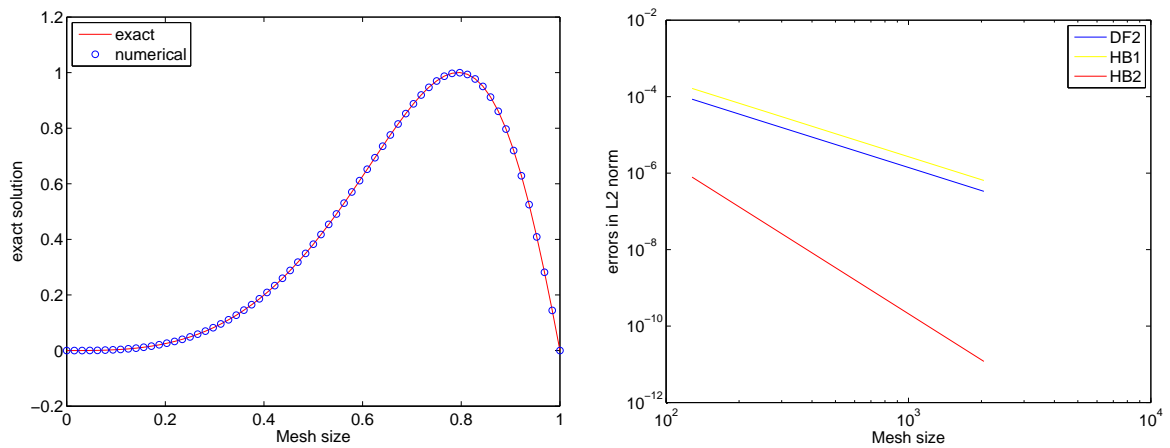


FIG. 2.5: A gauche, les solutions exacte et numérique avec  $u_{ex}(x) = \sin(\pi x^3)$ . A droite, les courbes de convergence de 3 schémas.

le schéma HB1 et le schéma HB2 possèdent les mêmes courbes de convergence. Ce cas est particulier parce que la fonction exacte est nulle au voisinage de bord. Dans tous les tests proposés, le schéma HB2 est meilleur que les schémas HB1 et DF2.

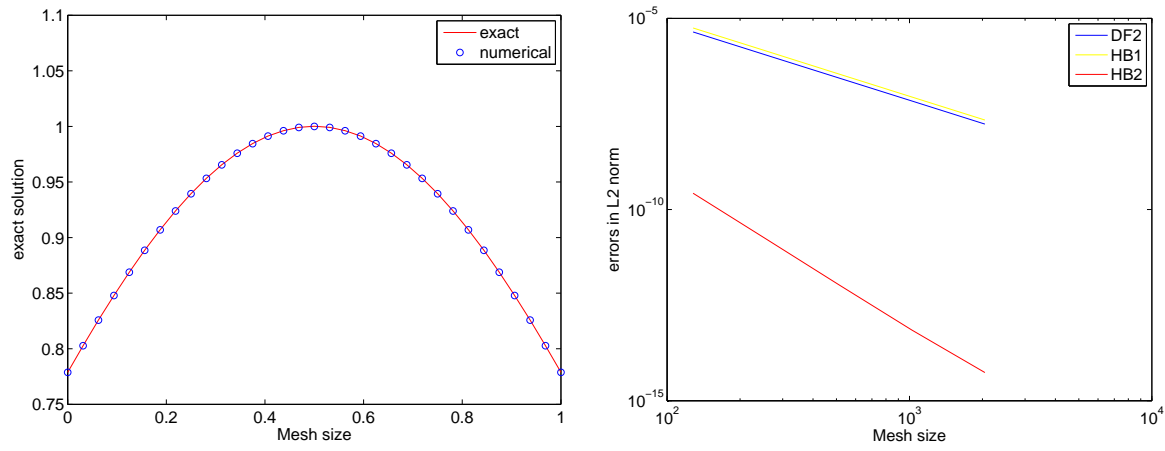


FIG. 2.6: A gauche, les solutions exacte et numérique avec  $u_{ex}(x) = \exp(-(x - \frac{1}{2})^2)$ .  
A droite, les courbes de convergence de 3 schémas.

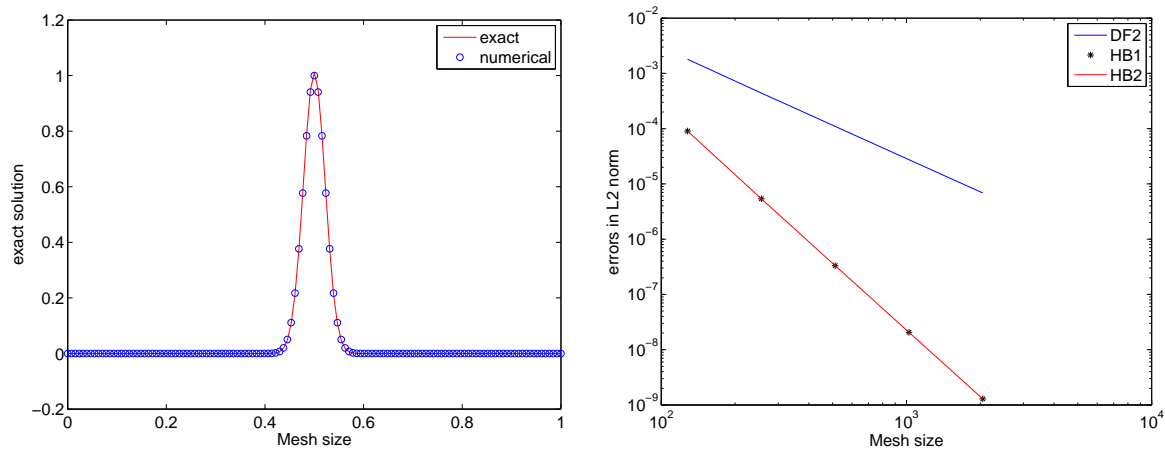


FIG. 2.7: A gauche, les solutions exacte et numérique avec  $u_{ex}(x) = \exp(-1000(x - \frac{1}{2})^2)$ . A droite, les courbes de convergence de 3 schémas.



## Chapitre 3

# Schémas boîte hermitiens sur géométrie cartésienne

### 3.1 Introduction

Dans ce chapitre, on introduit un nouveau schéma aux différences de type *boîte hermitien* sur une grille cartésienne. On se concentre sur le problème du Poisson dans un carré. L'ordre de précision obtenu en pratique est 4 pour les deux inconnues  $u$  et  $\nabla u$ . Ceci est particulièrement important pour les problèmes physiques dans lesquels on souhaite évaluer un champ (électrique, magnétique, vitesse, ...).

Une fois que le schéma est mis au point, l'effort principal porte sur la conception d'une procédure de résolution rapide. Celle-ci est basée sur la décomposition de la matrice du schéma comme somme d'une matrice diagonale  $H_1$  dans la base spectrale  $(Z^k)_{1 \leq k \leq N-1}$  (1.43) et d'une matrice auxiliaire  $\delta H_1$  de petit rang. On utilise ensuite la formule de Sherman-Morrison-Woodbury (2.155) pour définir un algorithme rapide sur le modèle de l'algorithme 1. De nombreux exemples numériques mettent en évidence l'intérêt du schéma ainsi défini.

### 3.2 Opérateurs aux différences en deux dimensions

On considère sur le carré  $\Omega = (a, b)^2$  la grille de différences finies régulière

$$\begin{cases} x_i = a + ih, & 0 \leq i \leq N, \\ y_j = a + jh, & 0 \leq j \leq N. \end{cases} \quad (3.1)$$

Le pas du maillage est  $h = \frac{b-a}{N}$ . En chaque point  $(x_i, y_j)_{0 \leq i, j \leq N}$  sont localisées les inconnues  $u_{i,j}$ .

- *Fonctions de grille*

On note  $L_h^2$  l'espace des fonctions de grille  $(u_{i,j})_{0 \leq i, j \leq N}$  et par  $L_{h,0}^2$  le sous-espace de  $L_h^2$  telles que  $u_{i,j} = 0$  pour  $i \in \{0, N\}$  ou  $j \in \{0, N\}$ .

- Le produit scalaire et la norme sur  $L_h^2$  sont définis par

$$(u, v)_h = h^2 \sum_{i,j=0}^N u_{i,j} v_{i,j}, \quad |u|_h^2 = h^2 \sum_{i,j=0}^N u_{i,j}^2. \quad (3.2)$$

- *Dérivées centrées*

Les opérateurs  $\delta_x$  et  $\delta_y$  sont définis par

$$\begin{cases} \delta_x u_{i,j} = \frac{u_{i+1,j} - u_{i-1,j}}{2h}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, \\ \delta_y u_{i,j} = \frac{u_{i,j+1} - u_{i,j-1}}{2h}, & 0 \leq i \leq N, \quad 1 \leq j \leq N-1. \end{cases} \quad (3.3)$$

On les prolonge aux points de bord  $i \in \{0, N\}$ ,  $j \in \{0, N\}$  par

$$\begin{cases} \delta_x u_{0,j} = \delta_x u_{N,j} = 0, & 0 \leq j \leq N, \\ \delta_y u_{i,0} = \delta_y u_{i,N} = 0, & 0 \leq i \leq N. \end{cases} \quad (3.4)$$

- *Dérivées décentrées*

Les opérateurs  $\delta_x^+$ ,  $\delta_x^-$ ,  $\delta_y^+$ ,  $\delta_y^-$  sont définis par

$$\begin{cases} \delta_x^+ u_{i,j} = \frac{u_{i+1,j} - u_{i,j}}{h}, & 0 \leq i \leq N-1, \quad 0 \leq j \leq N, & (a) \\ \delta_x^- u_{i,j} = \frac{u_{i,j} - u_{i-1,j}}{h}, & 1 \leq i \leq N, \quad 0 \leq j \leq N, & (b) \\ \delta_y^+ u_{i,j} = \frac{u_{i,j+1} - u_{i,j}}{h}, & 0 \leq i \leq N, \quad 0 \leq j \leq N-1, & (c) \\ \delta_y^- u_{i,j} = \frac{u_{i,j} - u_{i,j-1}}{h}, & 0 \leq i \leq N, \quad 1 \leq j \leq N. & (d) \end{cases} \quad (3.5)$$

On les prolonge aux points de bord  $i \in \{0, N\}$ ,  $j \in \{0, N\}$  par

$$\begin{cases} \delta_x^+ u_{N,j} = 0, & 0 \leq j \leq N, \\ \delta_x^+ u_{0,j} = 0, & 0 \leq j \leq N, \\ \delta_y^+ u_{i,N} = 0, & 0 \leq i \leq N, \\ \delta_y^- u_{i,0} = 0, & 0 \leq i \leq N. \end{cases} \quad (3.6)$$

- *Laplacien à trois points*

Les opérateurs de Laplace  $\delta_x^2$  et  $\delta_y^2$  sont définis par

$$\begin{cases} \delta_x^2 u_{i,j} = \frac{u_{i+1,j} + u_{i-1,j} - 2u_{i,j}}{h^2}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, \\ \delta_y^2 u_{i,j} = \frac{u_{i,j+1} + u_{i,j-1} - 2u_{i,j}}{h^2}, & 1 \leq j \leq N-1, \quad 0 \leq i \leq N. \end{cases} \quad (3.7)$$

On les prolonge aux points de bord  $i \in \{0, N\}$ ,  $j \in \{0, N\}$  par

$$\begin{cases} \delta_x^2 u_{0,j} = \delta_x^2 u_{N,j} = 0, & 0 \leq j \leq N, \\ \delta_y^2 u_{i,0} = \delta_y^2 u_{i,N} = 0, & 0 \leq i \leq N. \end{cases} \quad (3.8)$$

- *Opérateurs de Simpson*

Les opérateurs de Simpson  $\sigma_x$  et  $\sigma_y$  sont définis par

$$\begin{cases} \sigma_x u_{i,j} = \frac{1}{6}u_{i-1,j} + \frac{2}{3}u_{i,j} + \frac{1}{6}u_{i+1,j}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, \\ \sigma_y u_{i,j} = \frac{1}{6}u_{i,j-1} + \frac{2}{3}u_{i,j} + \frac{1}{6}u_{i,j+1}, & 0 \leq i \leq N, \quad 1 \leq j \leq N-1. \end{cases} \quad (3.9)$$

On les prolonge aux points de bord  $i \in \{0, N\}$ ,  $j \in \{0, N\}$ , par

$$\begin{cases} \sigma_x u_{0,j} = \sigma_x u_{N,j} = 0, & 0 \leq j \leq N, \\ \sigma_y u_{i,0} = \sigma_y u_{i,N} = 0, & 0 \leq i \leq N. \end{cases} \quad (3.10)$$

### 3.3 Notations algébriques

Dans cette section, on précise les notations vectorielles et matricielles utiles dans la suite.

- Opérateur  $\text{vec}_2$

**Définition 2.** Soient  $v = (v_{i,j})_{0 \leq i,j \leq N} \in L_h^2$  une fonction de grille et  $(e_i)_{1 \leq i \leq N-1}$  la base canonique de  $\mathbb{R}^{(N-1)}$  définie par

$$(e_i)_j = \begin{cases} 1, & \text{si } j = i, \\ 0, & \text{sinon.} \end{cases} \quad (3.11)$$

On définit l'opérateur  $\text{vec}_2$  par

$$\begin{aligned} \text{vec}_2 : L_h^2 &\longrightarrow \mathbb{R}^{(N-1)^2} \\ v &\longrightarrow V = \text{vec}_2(v) \end{aligned}$$

avec

$$\text{vec}_2(v) = \text{vec}_2[v_{i,j}]_{1 \leq i,j \leq N-1} = \sum_{i,j=1}^{N-1} (e_i \otimes e_j) v_{i,j}. \quad (3.12)$$

Matriciellement, l'opérateur  $\text{vec}_2$  crée à partir d'une matrice  $(v_{i,j})_{0 \leq i,j \leq N}$  de taille  $N+1$  un vecteur de taille  $(N-1)^2$  formé des points intérieurs de cette matrice (c'est-à-dire les points  $(v_{i,j})_{1 \leq i,j \leq N-1}$ ) dans l'ordre lexicographique sur  $(i, j)$ . On peut écrire

$$V = [v_{1,1}, v_{1,2}, \dots, v_{1,N-1}, v_{2,1}, v_{2,2}, \dots, v_{2,N-1}, \dots, v_{N-1,1}, v_{N-1,2}, \dots, v_{N-1,N-1}]^T. \quad (3.13)$$

**Exemple :** Soient  $\Omega = (0, 1)^2$ ,  $x_i = ih$ ,  $y_j = jh$ ,  $0 \leq i, j \leq 3$ , et  $h = 1/N = 1/3$ . et  $f(x, y) = \ln(x + 2y + 1)$ , on définit la fonction de grille  $f^* = f_{i,j}^* = \ln(x_i + 2y_j + 1)$ , avec  $i, j = 0, 1, 2, 3$ . On suppose  $u = f^*$ , alors

$$U = \text{vec}_2(u) = [\ln(x_1 + 2y_1 + 1), \ln(x_1 + 2y_2 + 1), \ln(x_2 + 2y_1 + 1), \ln(x_2 + 2y_2 + 1)]^T. \quad (3.14)$$

**Définition 3.** Soient  $v = (v_{i,j})_{0 \leq i,j \leq N}$  et  $(e_i)_{1 \leq i \leq N}$  la base canonique de  $\mathbb{R}^{(N-1)}$ . On définit les vecteurs  $V_{i,:}$  et  $V_{:,j} \in \mathbb{R}^{(N-1)}$  par

$$V_{i,:} = \text{vec}_1[v_{i,j}]_{1 \leq j \leq N-1} = \sum_{j=1}^{N-1} v_{i,j} e_j, \quad (3.15)$$

et

$$V_{:,j} = \text{vec}_1[v_{i,j}]_{1 \leq i \leq N-1} = \sum_{i=1}^{N-1} v_{i,j} e_i. \quad (3.16)$$

On peut écrire

$$\begin{cases} V_{i,:} = [v_{i,1}, v_{i,2}, \dots, v_{i,N-1}]^T, & (a) \\ V_{:,j} = [v_{1,j}, v_{2,j}, \dots, v_{N-1,j}]^T. & (b) \end{cases} \quad (3.17)$$

**Lemme 3.3.1** (Relations entre les opérateurs  $\text{vec}_1$  et  $\text{vec}_2$ ). Soient  $v = (v_{i,j})_{0 \leq i,j \leq N} \in L_h^2$  et  $V = \text{vec}_2(v) \in \mathbb{R}^{(N-1)^2}$  alors

$$\begin{cases} V = \sum_{j=1}^{N-1} V_{:,j} \otimes e_j, & (a) \\ V = \sum_{i=1}^{N-1} e_i \otimes V_{i,:}. & (b) \end{cases} \quad (3.18)$$

**Preuve.** On a par définition

$$V = \sum_{i,j=1}^{N-1} (e_i \otimes e_j) v_{i,j},$$

alors

$$\begin{aligned} V &= \sum_{i,j=1}^{N-1} (e_i v_{i,j} \otimes e_j) \\ &= \sum_{j=1}^{N-1} \left( \sum_{i=1}^{N-1} e_i v_{i,j} \otimes e_j \right) \\ &= \sum_{j=1}^{N-1} (V_{:,j} \otimes e_j), \end{aligned}$$

d'où (3.18)<sub>a</sub>. De la même manière on démontre (3.18)<sub>b</sub>. ■

• Algèbre de Kronecker

**Définition 4.** Soient  $B$  une matrice de taille  $m \times n$  et  $C$  une matrice de taille  $p \times q$ ,  $m, n, p, q \in \mathbb{N}$ . Leur produit de Kronecker est la matrice  $B \otimes C$  de taille  $mp \times nq$  définie par

$$B \otimes C = \begin{bmatrix} b_{11}C & \dots & b_{1n}C \\ \vdots & \vdots & \vdots \\ b_{m1}C & \dots & b_{mn}C \end{bmatrix}. \quad (3.19)$$

Si  $A$  est une matrice de taille  $l \times k$ .  $A \otimes B \otimes C$  est une matrice de taille  $lmp \times knq$ , définie par

$$A \otimes B \otimes C = \begin{bmatrix} a_{11}(B \otimes C) & \dots & a_{1k}(B \otimes C) \\ \vdots & \vdots & \vdots \\ a_{l1}(B \otimes C) & \dots & a_{lk}(B \otimes C) \end{bmatrix}. \quad (3.20)$$

Le produit de Kronecker intervient dans de nombreux contextes d'algèbre linéaire. On renvoie aux livres [17, 23, 27] pour une présentation générale. On se limite à rappeler les trois propriétés suivantes.

**Proposition 3.1.** Soient  $A, B, C, D$ , quatre matrices quelconques et  $\alpha$  un scalaire, alors

$$(i) \alpha(A \otimes B) = \alpha A \otimes B = A \otimes \alpha B. \quad (3.21)$$

Si les produits  $AB$  et  $CD$  sont bien définis, on a

$$(ii) (A \otimes B)(C \otimes D) = AC \otimes BD. \quad (3.22)$$

Si de plus les matrices  $A$  et  $B$  sont inversibles alors le produit kroneckerien  $A \otimes B$  est inversible et vérifie

$$(iii) (A \otimes B)^{-1} = A^{-1} \otimes B^{-1}. \quad (3.23)$$

On donne dans la proposition suivante des propriétés élémentaires du produit kroneckerien dont on aura besoin pour l'écriture matricielle de notre schéma boîte hermitien.

**Proposition 3.2.** Soient  $I_N$  la matrice identité d'ordre  $N$ ,  $v$  un vecteur dans  $\mathbb{R}^N$  et  $A$  une matrice de  $N$  lignes alors on a les deux égalités,

$$(i) (I \otimes v)A = (A \otimes v).$$

$$(ii) (v \otimes I)A = (v \otimes A).$$

**Preuve.** On montre (i). Pour toute matrice  $A \in \mathbb{M}_N(\mathbb{R})$  on a

$$A = A \otimes 1. \quad (3.24)$$

En utilisant (3.22) on obtient

$$(I \otimes v)A = (I \otimes v)(A \otimes 1) = (A \otimes v), \quad (3.25)$$

d'où (i). De la même manière on démontre (ii). ■

### 3.3.1 Notations vectorielles

Soit  $u(x, y)$  une fonction régulière sur  $\Omega = (a, b)^2$ . En chaque point  $(x_i, y_j)$ ,  $0 \leq i, j \leq N$ , de la grille de différences finies précisée au début de la section 3.2, trois inconnues  $u_{i,j} \simeq u(x_i, y_j)$ ,  $u_{x,i,j} \simeq \frac{\partial u}{\partial x}(x_i, y_j)$ ,  $u_{y,i,j} \simeq \frac{\partial u}{\partial y}(x_i, y_j)$  sont localisées. Les vecteurs  $U$ ,  $U_x$ ,  $U_y \in \mathbb{R}^{(N-1)^2}$  sont respectivement

$$\begin{cases} U = \text{vec}_2[u_{i,j}]_{1 \leq i, j \leq N-1} = [u_{1,1}, \dots, u_{1,N-1}, u_{2,1}, \dots, u_{2,N-1}, \dots, u_{N-1,1}, \dots, u_{N-1,N-1}]^T, \\ U_x = \text{vec}_2[u_{x,i,j}]_{1 \leq i, j \leq N-1} = [u_{x,1,1}, \dots, u_{x,1,N-1}, u_{x,2,1}, \dots, u_{x,2,N-1}, \dots, u_{x,N-1,1}, \dots, u_{x,N-1,N-1}]^T, \\ U_y = \text{vec}_2[u_{y,i,j}]_{1 \leq i, j \leq N-1} = [u_{y,1,1}, \dots, u_{y,1,N-1}, u_{y,2,1}, \dots, u_{y,2,N-1}, \dots, u_{y,N-1,1}, \dots, u_{y,N-1,N-1}]^T. \end{cases} \quad (3.26)$$

Soient  $U_B$  et  $U_T \in \mathbb{R}^{(N-1)}$  associés aux valeurs de l'inconnue  $u$  avec ordonnés  $y = a$  et  $y = b$  respectivement. De même,  $U_L$  et  $U_R \in \mathbb{R}^{(N-1)}$  sont les vecteurs associés aux valeurs de l'inconnue  $u$  avec abscisses  $x = a$  et  $x = b$  respectivement (figure 3.1),

$$\begin{cases} U_L = \text{vec}_1[u_{0,j}]_{1 \leq j \leq N-1} = [u_{0,1}, \dots, u_{0,N-1}]^T, \\ U_R = \text{vec}_1[u_{N,j}]_{1 \leq j \leq N-1} = [u_{N,1}, \dots, u_{N,N-1}]^T, \\ U_B = \text{vec}_1[u_{i,0}]_{1 \leq i \leq N-1} = [u_{1,0}, \dots, u_{N-1,0}]^T, \\ U_T = \text{vec}_1[u_{i,N}]_{1 \leq i \leq N-1} = [u_{1,N}, \dots, u_{N-1,N}]^T. \end{cases} \quad (3.27)$$

Les vecteurs de dérivées au bord par rapport à  $x$  sont notés par

$$\begin{cases} U_{x,L} = \text{vec}_1[u_{x,0,j}]_{1 \leq j \leq N-1} = [u_{x,0,1}, u_{x,0,2}, \dots, u_{x,0,N-2}, u_{x,0,N-1}]^T, \\ U_{x,R} = \text{vec}_1[u_{x,N,j}]_{1 \leq j \leq N-1} = [u_{x,N,1}, u_{x,N,2}, \dots, u_{x,N,N-2}, u_{x,N,N-1}]^T, \\ U_{x,B} = \text{vec}_1[u_{x,i,0}]_{1 \leq i \leq N-1} = [u_{x,1,0}, u_{x,2,0}, \dots, u_{x,N-2,0}, u_{x,N-1,0}]^T, \\ U_{x,T} = \text{vec}_1[u_{x,i,N}]_{1 \leq i \leq N-1} = [u_{x,1,N}, u_{x,2,N}, \dots, u_{x,N-2,N}, u_{x,N-1,N}]^T. \end{cases} \quad (3.28)$$

De même, les vecteurs de dérivées au bord par rapport à  $y$  sont notés par

$$\begin{cases} U_{y,L} = \text{vec}_1[u_{y,0,j}]_{1 \leq j \leq N-1} = [u_{y,0,1}, u_{y,0,2}, \dots, u_{y,0,N-2}, u_{y,0,N-1}]^T, \\ U_{y,R} = \text{vec}_1[u_{y,N,j}]_{1 \leq j \leq N-1} = [u_{y,N,1}, u_{y,N,2}, \dots, u_{y,N,N-2}, u_{y,N,N-1}]^T, \\ U_{y,B} = \text{vec}_1[u_{y,i,0}]_{1 \leq i \leq N-1} = [u_{y,1,0}, u_{y,2,0}, \dots, u_{y,N-2,0}, u_{y,N-1,0}]^T, \\ U_{y,T} = \text{vec}_1[u_{y,i,N}]_{1 \leq i \leq N-1} = [u_{y,1,N}, u_{y,2,N}, \dots, u_{y,N-2,N}, u_{y,N-1,N}]^T. \end{cases} \quad (3.29)$$

On note les quatres valeurs de Dirichlet aux sommets par

$$U_{LB} = u_{0,0}, \quad U_{LT} = u_{0,N}, \quad U_{RB} = u_{N,0}, \quad U_{RT} = u_{N,N}. \quad (3.30)$$

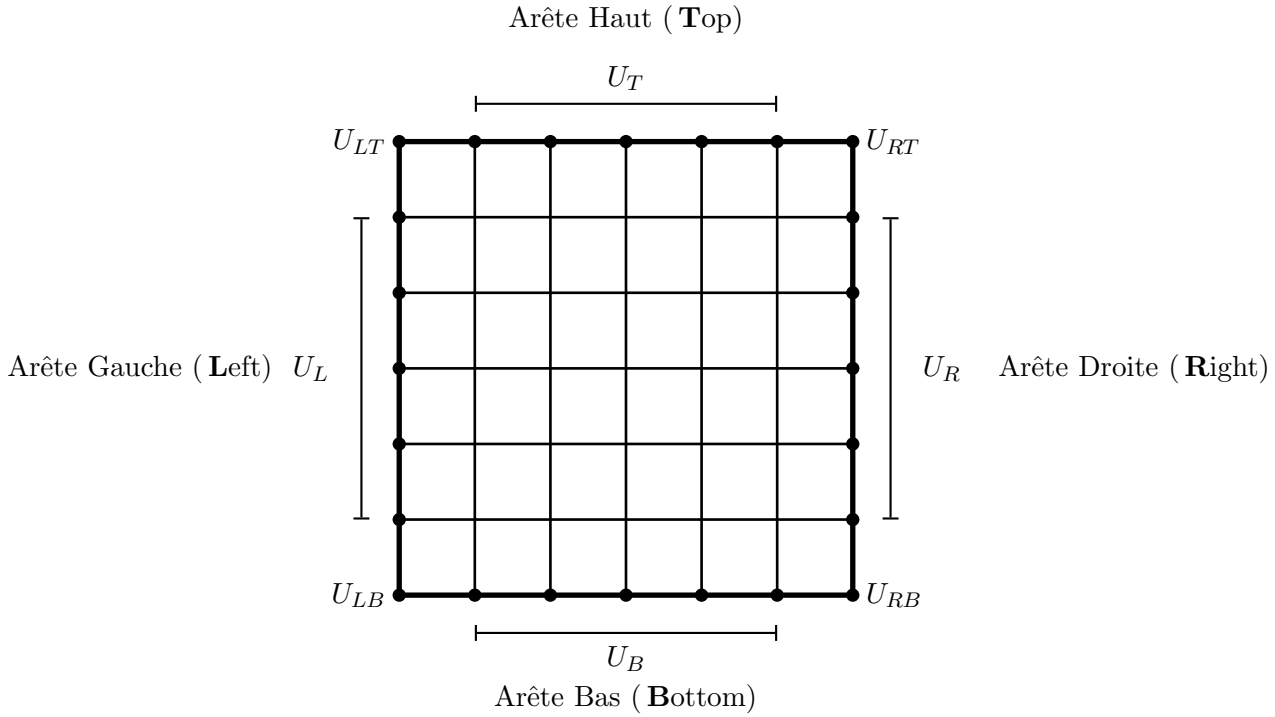


FIG. 3.1: Notations de la grille en dimension 2. Les lettres  $L$  et  $R$  correspondent à “Left” et “Right”. Les lettres  $B$  et  $T$  correspondent à “Bottom” et “Top”.

Les dérivées sur les quatres sommets sont notées par

$$\begin{cases} (U_{x,LB}, U_{y,LB}) = (u_{x,0,0}, u_{y,0,0}), \\ (U_{x,RB}, U_{y,RB}) = (u_{x,N,0}, u_{y,N,0}), \\ (U_{x,LT}, U_{y,LT}) = (u_{x,0,N}, u_{y,0,N}), \\ (U_{x,RT}, U_{y,RT}) = (u_{x,N,N}, u_{y,N,N}). \end{cases} \quad (3.31)$$

### 3.4 Fonctions de grille et algèbre de Kronecker

Les deux lemmes suivants jouent un rôle fondamental dans la dérivation du schéma boîte hermitien en dimension 2. Dans la suite,  $e_0$  et  $e_N$  sont deux vecteurs nuls de  $\mathbb{R}^{N-1}$ .



**Lemme 3.4.1.** Soit  $u = (u_{i,j})_{0 \leq i,j \leq N} \in L_h^2$ . Les deux opérateurs des dérivées centrées (3.3) s'écrivent matriciellement sous la forme

$$\begin{cases} \text{vec}_2(\delta_x u) = \text{vec}_2[\delta_x u_{i,j}]_{1 \leq i,j \leq N-1} = \frac{1}{2h}(K \otimes I)U - \frac{1}{2h}(e_1 \otimes I)U_L + \frac{1}{2h}(e_{N-1} \otimes I)U_R, & (a) \\ \text{vec}_2(\delta_y u) = \text{vec}_2[\delta_y u_{i,j}]_{1 \leq i,j \leq N-1} = \frac{1}{2h}(I \otimes K)U - \frac{1}{2h}(I \otimes e_1)U_B + \frac{1}{2h}(I \otimes e_{N-1})U_T. & (b) \end{cases} \quad (3.32)$$

**Preuve.** On a

$$\delta_x u_{i,j} = \frac{u_{i+1,j} - u_{i-1,j}}{2h}, \quad 1 \leq i, j \leq N-1, \quad (3.33)$$

donc

$$\text{vec}_2(\delta_x u) = [\delta_x u_{1,1}, \delta_x u_{1,2}, \dots, \delta_x u_{1,N-1}, \delta_x u_{2,1}, \dots, \delta_x u_{2,N-1}, \dots, \delta_x u_{N-1,1}, \dots, \delta_x u_{N-1,N-1}]^T, \quad (3.34)$$

ce qui donne

$$\begin{aligned} \text{vec}_2(\delta_x u) &= \frac{1}{2h}[u_{2,1} - u_{0,1}, \dots, u_{2,N-1} - u_{0,N-1}, u_{3,1} - u_{1,1}, \dots, u_{N,1} - u_{N-2,1}, \dots, u_{N,N-1} - u_{N-2,N-1}]^T \\ &= \frac{1}{2h}[u_{2,1}, \dots, u_{2,N-1}, u_{3,1} - u_{1,1}, \dots, u_{3,N-1} - u_{1,N-1}, \dots, -u_{N-2,1}, \dots, -u_{N-2,N-1}]^T \\ &\quad - \frac{1}{2h}[u_{0,1}, u_{0,2}, \dots, u_{0,N-1}, 0, \dots, 0, \dots, 0, \dots, 0]^T \\ &\quad + \frac{1}{2h}[0, 0, \dots, 0, 0, \dots, 0, \dots, u_{N,1}, u_{N,2}, \dots, u_{N,N-1}]^T \\ &= \frac{1}{2h}(K \otimes I)\text{vec}_2(u) - \frac{1}{2h}(e_1 \otimes I)U_L + \frac{1}{2h}(e_{N-1} \otimes I)U_R, \end{aligned}$$

d'où (3.32)<sub>a</sub>. De la même manière on démontre (3.32)<sub>b</sub>. ■

**Lemme 3.4.2.** Soit  $u = (u_{i,j})_{0 \leq i,j \leq N} \in L_h^2$ . Les deux opérateurs de Simpson (3.9) s'écrivent matriciellement sous la forme

$$\begin{cases} \text{vec}_2(\sigma_x u) = \text{vec}_2[\sigma_x u_{i,j}]_{1 \leq i,j \leq N-1} = (P_s \otimes I)U + \frac{1}{6}(e_1 \otimes I)U_L + \frac{1}{6}(e_{N-1} \otimes I)U_R, & (a) \\ \text{vec}_2(\sigma_y u) = \text{vec}_2[\sigma_y u_{i,j}]_{1 \leq i,j \leq N-1} = (I \otimes P_s)U + \frac{1}{6}(I \otimes e_1)U_B + \frac{1}{6}(I \otimes e_{N-1})U_T. & (b) \end{cases} \quad (3.35)$$

**Preuve.** On a par définition

$$\text{vec}_2(\sigma_x u) = \sum_{i,j=1}^{N-1} (e_i \otimes e_j)(\sigma_x u)_{i,j},$$

alors

$$\begin{aligned}
\text{vec}_2(\sigma_x u) &= \sum_{i,j=1}^{N-1} (e_i \otimes e_j) \left( \frac{1}{6} u_{i-1,j} + \frac{2}{3} u_{i,j} + \frac{1}{6} u_{i+1,j} \right) \\
&= \sum_{i,j=1}^{N-1} \left( \frac{1}{6} e_i u_{i-1,j} \otimes e_j \right) + \sum_{i,j=1}^{N-1} \left( \frac{2}{3} e_i u_{i,j} \otimes e_j \right) + \sum_{i,j=1}^{N-1} \left( \frac{1}{6} e_i u_{i+1,j} \otimes e_j \right) \\
&= \sum_{j=1}^{N-1} \frac{1}{6} e_1 u_{0,j} \otimes e_j + \sum_{j=1}^{N-1} \sum_{i=2}^{N-1} \frac{1}{6} e_i u_{i-1,j} \otimes e_j + \sum_{i,j=1}^{N-1} \frac{2}{3} e_i u_{i,j} \otimes e_j + \sum_{j=1}^{N-1} \sum_{i=1}^{N-2} \frac{1}{6} e_i u_{i+1,j} \otimes e_j \\
&\quad + \sum_{j=1}^{N-1} \left( \frac{1}{6} e_{N-1} u_{N,j} \otimes e_j \right) \\
&= \frac{1}{6} (e_1 \otimes I) \sum_{j=1}^{N-1} u_{0,j} e_j + \sum_{j=1}^{N-1} \left( \sum_{i=1}^{N-1} \left( \frac{1}{6} e_{i+1} u_{i,j} + \frac{2}{3} e_i u_{i,j} + \frac{1}{6} e_{i-1} u_{i,j} \right) \otimes e_j \right) \\
&\quad + \frac{1}{6} (e_{N-1} \otimes I) \sum_{j=1}^{N-1} u_{N,j} e_j \\
&= \frac{1}{6} (e_1 \otimes I) U_L + \sum_{j=1}^{N-1} \left( \sum_{i=1}^{N-1} P_s e_i \otimes e_j \right) u_{i,j} + \frac{1}{6} (e_{N-1} \otimes I) U_R \\
&= \frac{1}{6} (e_1 \otimes I) U_L + (P_s \otimes I) \sum_{i,j=1}^{N-1} (e_i \otimes e_j) u_{i,j} + \frac{1}{6} (e_{N-1} \otimes I) U_R \\
&= \frac{1}{6} (e_1 \otimes I) U_L + (P_s \otimes I) U + \frac{1}{6} (e_{N-1} \otimes I) U_R,
\end{aligned}$$

d'où (3.35)<sub>a</sub>. De la même manière on démontre (3.35)<sub>b</sub>. ■

## 3.5 Schéma boîte hermitien en dimension 2

### 3.5.1 Introduction

On a vu dans la section 2.2 que le principe du schéma boîte hermitien (HB) repose sur :

- La formulation mixte (2.6).
- L'évaluation des moyennes sur  $[x_{j-1}, x_{j+1}]$  de (2.7) et (2.9).

Ici, on généralise ce point de vue pour obtenir un schéma boîte hermitien en dimension 2. On se limite au problème de Poisson dans un carré mais le principe du schéma est généralisable à d'autres types de problèmes. Dans la section 3.5.2 on établit un système d'équations qui sera *la base du schéma HB en dimension 2*. Pour compléter ce

système d'équations, on donne dans la section 3.5.3 deux exemples d'approximations des dérivées normales  $\frac{\partial u}{\partial n}$  sur  $\partial\Omega$ , déduites des approximations de dérivées au bord utilisées au chapitre 2 pour construire les deux schémas HB1 et HB2.

Commençons par rappeler le théorème de la divergence dont on a besoin dans la suite.

Si  $\Omega$  est un ouvert polygonal et  $\vec{u}(x, y)$  un champ de vecteurs régulier, alors

$$\int_{\Omega} \operatorname{div} \vec{u} \, d\Omega = \int_{\partial\Omega} \vec{u} \cdot \vec{n} \, d\sigma, \quad (3.36)$$

avec  $\vec{u}$  est le vecteur normal unitaire à  $\partial\Omega$  dirigé vers l'extérieur.

### 3.5.2 Définition du schéma pour le problème de Poisson

On considère le problème de Poisson dans  $\Omega = (a, b)^2$ ,

$$\begin{cases} -\partial_x^2 u - \partial_y^2 u = f, & \text{dans } \Omega, & (a) \\ u = g, & \text{sur } \partial\Omega. & (b) \end{cases} \quad (3.37)$$

La fonction  $f(x, y)$  est le terme source et la fonction  $g(x, y)$  correspond aux données de Dirichlet. En posant  $v^1 = \partial_x u$  et  $v^2 = \partial_y u$ , (3.37) se réécrit sous la forme mixte

$$\begin{cases} -(\partial_x v^1 + \partial_y v^2) = f, & \text{dans } \Omega, & (a) \\ v^1 - \partial_x u = 0, & \text{dans } \bar{\Omega}, & (b) \\ v^2 - \partial_y u = 0, & \text{dans } \bar{\Omega}, & (c) \\ u = g, & \text{sur } \partial\Omega. & (d) \end{cases} \quad (3.38)$$

Le domaine  $\Omega$  est discrétisé par la grille (3.1). Pour obtenir la forme discrète de (3.38), on utilise les boîtes  $K_{i,j}$  centrées aux points  $(x_i, y_j)$  de longueur  $2h$  (figure 3.2),

$$K_{i,j} = [x_i - h, x_i + h] \times [y_j - h, y_j + h], \quad \forall 1 \leq i, j \leq N - 1. \quad (3.39)$$

- Les moyennes de  $f$  sur chaque boîte  $K_{i,j}$  sont données par

$$\Pi^0 f_{i,j} = \frac{1}{4h^2} \int_{K_{i,j}} f(x, y) \, dx \, dy, \quad 1 \leq i, j \leq N - 1. \quad (3.40)$$

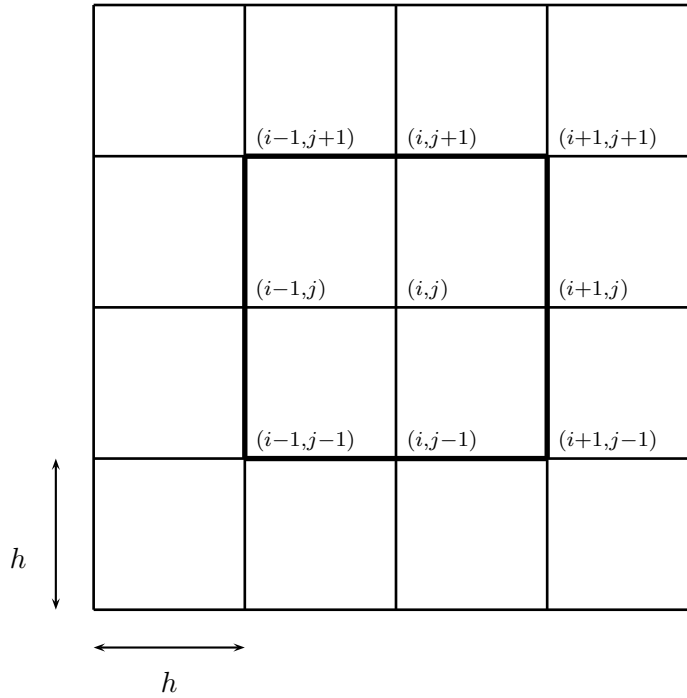


FIG. 3.2: Schéma HB en dimension deux. Le flux est intégré sur le contour noir de la boîte  $K_{i,j} = [x_i - h, x_i + h] \times [y_j - h, y_j + h]$

En utilisant le théorème de la divergence (3.36), (3.38)<sub>a</sub> devient (le vecteur normal dirigé vers l'extérieur de  $\partial K_{i,j}$  est  $\vec{n}(n_1, n_2)$  )

$$-\frac{1}{4h^2} \int_{\partial K_{i,j}} (v^1 n_1 + v^2 n_2) d\sigma = \Pi^0 f_{i,j}, \quad 1 \leq i, j \leq N-1. \quad (3.41)$$

Le terme de gauche de (3.41) consiste en quatre intégrales monodimensionnelles correspondant aux quatre faces de  $\partial K_{i,j}$ . Chaque intégrale est approchée par la formule de Simpson. Ceci donne l'approximation suivante de (3.41),

$$\left\{ \begin{array}{l} -\frac{1}{(2h)^2} \left\{ \begin{array}{l} \left[ \frac{1}{6} u_{x,i+1,j-1} + \frac{2}{3} u_{x,i+1,j} + \frac{1}{6} u_{x,i+1,j+1} \right] \\ - \left[ \frac{1}{6} u_{x,i-1,j-1} + \frac{2}{3} u_{x,i-1,j} + \frac{1}{6} u_{x,i-1,j+1} \right] \\ + \left[ \frac{1}{6} u_{y,i-1,j+1} + \frac{2}{3} u_{y,i,j+1} + \frac{1}{6} u_{y,i+1,j+1} \right] \\ - \left[ \frac{1}{6} u_{y,i-1,j-1} + \frac{2}{3} u_{y,i,j-1} + \frac{1}{6} u_{y,i+1,j-1} \right] \end{array} \right\} = \Pi^0 f_{i,j}, \quad 1 \leq i, j \leq N-1. \end{array} \right. \quad (3.42)$$

En utilisant les opérateurs aux différences  $\delta_x$ ,  $\delta_y$ ,  $\sigma_x$ ,  $\sigma_y$ , (3.42) s'écrit

$$-(\sigma_y \circ \delta_x)u_{x,i,j} - (\sigma_x \circ \delta_y)u_{y,i,j} = \Pi^0 f_{i,j}, \quad 1 \leq i, j \leq N-1, \quad (3.43)$$

On utilise la notation  $\circ$  pour la composition des opérateurs aux différences. (3.43) donne  $(N-1)^2$  équations.

- La relation entre  $u$  et  $v^1$  est déduite de (3.38)<sub>b</sub>,

$$\frac{1}{2h} \int_{x_{i-1}}^{x_{i+1}} v^1(x, y_j) dx = \frac{u(x_{i+1}, y_j) - u(x_{i-1}, y_j)}{2h}, \quad 1 \leq i \leq N-1, \quad 0 \leq j \leq N. \quad (3.44)$$

La relation entre  $u$  et  $v^2$  est déduite de (3.38)<sub>c</sub>

$$\frac{1}{2h} \int_{y_{j-1}}^{y_{j+1}} v^2(x_i, y) dy = \frac{u(x_i, y_{j+1}) - u(x_i, y_{j-1})}{2h}, \quad 0 \leq i \leq N, \quad 1 \leq j \leq N-1. \quad (3.45)$$

(3.44) et (3.45) donnent  $2(N-1)(N+1)$  équations. Les termes de gauche de (3.44) et (3.45) sont approchés par la formule de Simpson, ce qui donne

$$\begin{cases} \frac{1}{6}u_{x,i-1,j} + \frac{2}{3}u_{x,i,j} + \frac{1}{6}u_{x,i+1,j} = \frac{u_{i+1,j} - u_{i-1,j}}{2h}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, \\ \frac{1}{6}u_{y,i,j-1} + \frac{2}{3}u_{y,i,j} + \frac{1}{6}u_{y,i,j+1} = \frac{u_{i,j+1} - u_{i,j-1}}{2h}, & 1 \leq j \leq N-1, \quad 0 \leq i \leq N. \end{cases} \quad (3.46)$$

En utilisant les opérateurs  $\delta_x$ ,  $\delta_y$ ,  $\sigma_x$ ,  $\sigma_y$ , (3.46) s'écrit

$$\begin{cases} \sigma_x u_{x,i,j} = \delta_x u_{i,j}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, \\ \sigma_y u_{y,i,j} = \delta_y u_{i,j}, & 1 \leq j \leq N-1, \quad 0 \leq i \leq N. \end{cases} \quad (3.47)$$

(3.47) donne  $2(N-1)(N+1)$  équations.

- Les conditions limites naturelles sont

$$\begin{cases} u_{i,0} = g(x_i, y_0), & 0 \leq i \leq N, & \text{Arête Bas "B"} \\ u_{i,N} = g(x_i, y_N), & 0 \leq i \leq N, & \text{Arête Haut "T"} \\ u_{0,j} = g(x_0, y_j), & 1 \leq j \leq N-1, & \text{Arête Gauche "L"} \\ u_{N,j} = g(x_N, y_j), & 1 \leq j \leq N-1, & \text{Arête Droite "R"} \end{cases} \quad (3.48)$$

Ce qui donne  $4N$  équations.

Finalement, en regroupant les systèmes (3.43), (3.47), (3.48), on obtient

$$\left\{ \begin{array}{ll} -(\sigma_y \circ \delta_x)u_{x,i,j} - (\sigma_x \circ \delta_y)u_{y,i,j} = \Pi^0 f_{i,j}, & 1 \leq i, j \leq N-1, \quad (a) \\ \sigma_x u_{x,i,j} = \delta_x u_{i,j}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, \quad (b) \\ \sigma_y u_{y,i,j} = \delta_y u_{i,j}, & 1 \leq j \leq N-1, \quad 0 \leq i \leq N, \quad (c) \\ u_{i,0} = g(x_i, y_0), & 0 \leq i \leq N, \quad (d) \\ u_{i,N} = g(x_i, y_N), & 0 \leq i \leq N, \quad (e) \\ u_{0,j} = g(x_0, y_j), & 1 \leq j \leq N-1, \quad (f) \\ u_{N,j} = g(x_N, y_j), & 1 \leq j \leq N-1. \quad (g) \end{array} \right. \quad (3.49)$$

Le système (3.49) est la base du schéma HB en dimension 2. Il constitue de  $(N-1)^2 + 2(N-1)(N+1) + 4N$  équations. Par ailleurs, on a  $3(N+1)^2$  inconnues correspondant aux  $u = (u_{i,j})_{0 \leq i,j \leq N}$ ,  $u_x = (u_{x,i,j})_{0 \leq i,j \leq N}$ ,  $u_y = (u_{y,i,j})_{0 \leq i,j \leq N} \in L_h^2$ . Pour obtenir le bon décompte entre le nombre d'équations et le nombre d'inconnues on doit effectuer des approximations de dérivées normales  $u_{x,0,j}$ ,  $u_{x,N,j}$ ,  $u_{y,i,0}$ ,  $u_{y,i,N}$ ,  $0 \leq i, j \leq N$ . C'est ce que l'on fait dans la section suivante.

### 3.5.3 Exemples d'approximations de $\frac{\partial u}{\partial n}$ sur $\partial\Omega$

Dans les sections 2.2.1 et 2.3.1, on a vu que le schéma HB en dimension 1 pour le problème de Poisson nécessite la donnée d'une approximation de  $\frac{\partial u}{\partial n}$  sur le bord du domaine de calcul. Ceci peut sembler non naturel, puisque le problème de Poisson ne nécessite normalement que la donnée de  $u$  sur  $\partial\Omega$ . Ici, l'approximation est nécessitée par l'approximation de  $\nabla u$  à l'intérieur du domaine. C'est le même problème pour une approximation de type spline, où l'on doit également choisir une certaine approximation au bord de  $u'(x)$  pour définir entièrement  $u'(x)$  à l'intérieur du domaine.

On a donné au chapitre 2 deux exemples d'approximation de la dérivée au bord pour un problème de dimension 1 (2.1). Dans le premier exemple (2.26),  $u_{x,0}$  et  $u_{x,N}$  sont évaluées par les relations

$$\left\{ \begin{array}{l} \frac{2}{3}u_{x,0} + \frac{1}{3}u_{x,1} = \frac{u_1 - u_0}{h}, \\ \frac{2}{3}u_{x,N} + \frac{1}{3}u_{x,N-1} = \frac{u_N - u_{N-1}}{h}. \end{array} \right. \quad (3.50)$$

En utilisant (3.50), on a défini le schéma boîte hermitien appelé HB1 (2.27). Les résultats numériques donnés au section 2.2.5 montrent que ce schéma est d'ordre 2

pour  $u$  et  $u_x$ . Pour améliorer l'ordre de précision on a introduit une seconde approximation (2.97),

$$\begin{cases} \frac{1}{3}u_{x,0} + \frac{2}{3}u_{x,1} = \frac{1}{h} \left( \frac{1}{6}u_2 + \frac{2}{3}u_1 - \frac{5}{6}u_0 \right), \\ \frac{1}{3}u_{x,N} + \frac{2}{3}u_{x,N-1} = \frac{1}{h} \left( \frac{5}{6}u_N - \frac{1}{6}u_{N-2} - \frac{2}{3}u_{N-1} \right). \end{cases} \quad (3.51)$$

(3.51) est d'ordre 3 en  $x_0 = a$  et  $x_N = b$ . On a obtenu le schéma HB2 (2.103) qui est d'ordre 4 pour  $u$  et  $u_x$ . Chacune de ces deux approximations (3.50) et (3.51) peut être considérée comme une version discrète de l'opérateur appelé "Dirichlet-to-Neumann" :

$$\left( u|_{(a,b)}, u_x|_{(a,b)}, u(a), u(b) \right) \longrightarrow \left( u'(a), u'(b) \right). \quad (3.52)$$

Dans cette partie, nous généralisons (3.50) et (3.51) au cas du problème de Poisson dans un carré. Les dérivées normales par rapport à  $x$ ,  $u_{x,0,j}$ ,  $u_{x,N,j}$ ,  $0 \leq j \leq N$  et les dérivées normales par rapport à  $y$ ,  $u_{y,i,0}$ ,  $u_{y,i,N}$ ,  $0 \leq i \leq N$ , sont écrites en fonction des inconnues aux points intérieurs et des données de bord.

• *Exemple 1* : On généralise (3.50) de la façon suivante. Le long des quatres arêtes du carré  $\Omega = (a, b)^2$  on considère les identités

$$\begin{cases} \frac{2}{3}u_{x,0,j} + \frac{1}{3}u_{x,1,j} = \frac{u_{1,j} - u_{0,j}}{h}, & 0 \leq j \leq N, \quad \text{Arête Gauche "L"} \\ \frac{2}{3}u_{x,N,j} + \frac{1}{3}u_{x,N-1,j} = \frac{u_{N,j} - u_{N-1,j}}{h}, & 0 \leq j \leq N, \quad \text{Arête Droite "R"} \\ \frac{2}{3}u_{y,i,0} + \frac{1}{3}u_{y,i,1} = \frac{u_{i,1} - u_{i,0}}{h}, & 0 \leq i \leq N, \quad \text{Arête Bas "B"} \\ \frac{2}{3}u_{y,i,N} + \frac{1}{3}u_{y,i,N-1} = \frac{u_{i,N} - u_{i,N-1}}{h}, & 0 \leq i \leq N. \quad \text{Arête Haut "T"} \end{cases} \quad (3.53)$$

En complétant le système (3.49) par les équations (3.53) on obtient le schéma HB1 en dimension 2. Trouver  $u, u_x, u_y \in L_h^2$  solution du système

$$\left\{ \begin{array}{ll}
 -(\sigma_y \circ \delta_x)u_{x,i,j} - (\sigma_x \circ \delta_y)u_{y,i,j} = \Pi^0 f_{i,j}, & 1 \leq i, j \leq N-1, \quad (a) \\
 \sigma_x u_{x,i,j} = \delta_x u_{i,j}, & 1 \leq i, j \leq N-1, \quad 0 \leq j \leq N, \quad (b) \\
 \sigma_y u_{y,i,j} = \delta_y u_{i,j}, & 1 \leq i, j \leq N-1, \quad 0 \leq i \leq N, \quad (c) \\
 \frac{2}{3}u_{x,0,j} + \frac{1}{3}u_{x,1,j} = \frac{u_{1,j} - u_{0,j}}{h}, & 0 \leq j \leq N, \quad (d) \\
 \frac{2}{3}u_{x,N,j} + \frac{1}{3}u_{x,N-1,j} = \frac{u_{N,j} - u_{N-1,j}}{h}, & 0 \leq j \leq N, \quad (e) \\
 \frac{2}{3}u_{y,i,0} + \frac{1}{3}u_{y,i,1} = \frac{u_{i,1} - u_{i,0}}{h}, & 0 \leq i \leq N, \quad (f) \\
 \frac{2}{3}u_{y,i,N} + \frac{1}{3}u_{y,i,N-1} = \frac{u_{i,N} - u_{i,N-1}}{h}, & 0 \leq i \leq N, \quad (g) \\
 u_{i,0} = g(x_i, y_0), & 0 \leq i \leq N, \quad (h) \\
 u_{i,N} = g(x_i, y_N), & 0 \leq i \leq N, \quad (l) \\
 u_{0,j} = g(x_0, y_j), & 1 \leq j \leq N-1, \quad (m) \\
 u_{N,j} = g(x_N, y_j), & 1 \leq j \leq N-1. \quad (n)
 \end{array} \right. \quad (3.54)$$

• *Exemple 2* : De la même façon, (3.51) est généralisée sur les quatre arêtes du  $\Omega = (a, b)^2$  par les quatres séries des relations

$$\left\{ \begin{array}{ll}
 \frac{1}{3}u_{x,0,j} + \frac{2}{3}u_{x,1,j} = \frac{1}{h} \left( \frac{1}{6}u_{2,j} + \frac{2}{3}u_{1,j} - \frac{5}{6}u_{0,j} \right), & 0 \leq j \leq N, \quad \text{Arête Gauche "L"} \\
 \frac{1}{3}u_{x,N,j} + \frac{2}{3}u_{x,N-1,j} = \frac{1}{h} \left( \frac{5}{6}u_{N,j} - \frac{1}{6}u_{N-2,j} - \frac{2}{3}u_{N-1,j} \right), & 0 \leq j \leq N, \quad \text{Arête Droite "R"} \\
 \frac{1}{3}u_{y,i,0} + \frac{2}{3}u_{y,i,1} = \frac{1}{h} \left( \frac{1}{6}u_{i,2} + \frac{2}{3}u_{i,1} - \frac{5}{6}u_{i,0} \right), & 0 \leq i \leq N, \quad \text{Arête Bas "B"} \\
 \frac{1}{3}u_{y,i,N} + \frac{2}{3}u_{y,i,N-1} = \frac{1}{h} \left( \frac{5}{6}u_{i,N} - \frac{1}{6}u_{i,N-2} - \frac{2}{3}u_{i,N-1} \right), & 0 \leq i \leq N. \quad \text{Arête Haut "T"}
 \end{array} \right. \quad (3.55)$$



En complétant le système (3.49) par (3.55) on obtient le schéma HB2 en dimension 2. Trouver  $u, u_x, u_y \in L_h^2$  solution du système

$$\left\{ \begin{array}{ll}
 -(\sigma_y \circ \delta_x)u_{x,i,j} - (\sigma_x \circ \delta_y)u_{y,i,j} = \Pi^0 f_{i,j}, & 1 \leq i, j \leq N-1, \quad (a) \\
 \sigma_x u_{x,i,j} = \delta_x u_{i,j}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, \quad (b) \\
 \sigma_y u_{y,i,j} = \delta_y u_{i,j}, & 1 \leq j \leq N-1, \quad 0 \leq i \leq N, \quad (c) \\
 \frac{1}{3}u_{x,0,j} + \frac{2}{3}u_{x,1,j} = \frac{1}{h} \left( \frac{1}{6}u_{2,j} + \frac{2}{3}u_{1,j} - \frac{5}{6}u_{0,j} \right), & 0 \leq j \leq N, \quad (d) \\
 \frac{1}{3}u_{x,N,j} + \frac{2}{3}u_{x,N-1,j} = \frac{1}{h} \left( \frac{5}{6}u_{N,j} - \frac{1}{6}u_{N-2,j} - \frac{2}{3}u_{N-1,j} \right), & 0 \leq j \leq N, \quad (e) \\
 \frac{1}{3}u_{y,i,0} + \frac{2}{3}u_{y,i,1} = \frac{1}{h} \left( \frac{1}{6}u_{i,2} + \frac{2}{3}u_{i,1} - \frac{5}{6}u_{i,0} \right), & 0 \leq i \leq N, \quad (f) \\
 \frac{1}{3}u_{y,i,N} + \frac{2}{3}u_{y,i,N-1} = \frac{1}{h} \left( \frac{5}{6}u_{i,N} - \frac{1}{6}u_{i,N-2} - \frac{2}{3}u_{i,N-1} \right), & 0 \leq i \leq N, \quad (g) \\
 u_{i,0} = g(x_i, y_0), & 0 \leq i \leq N, \quad (h) \\
 u_{i,N} = g(x_i, y_N), & 0 \leq i \leq N, \quad (l) \\
 u_{0,j} = g(x_0, y_j), & 1 \leq j \leq N-1, \quad (m) \\
 u_{N,j} = g(x_N, y_j), & 1 \leq j \leq N-1. \quad (n)
 \end{array} \right. \quad (3.56)$$

Le système (3.56) contient  $3(N+1)^2$  inconnues et  $(N-1)^2 + 2(N-1)(N+1) + 6(N+1) + 2(N-1) = 3(N+1)^2$  équations. Ceci donne un décompte correct entre les équations et les inconnues.

## 3.6 Forme matricielle du schéma boîte hermitien en dimension 2

Dans cette section, on montre que la structure algébrique des schémas HB1 et HB2 en (3.54) et (3.56) respectivement, peut être exprimée de manière naturelle en utilisant le formalisme du produit de Kronecker.

On introduit à présent toutes les identités dont on a besoin pour obtenir la forme matricielle du schéma boîte hermitien en dimension 2. Dans la suite de cette section, on considère dans  $L_h^2$ , les trois fonctions de grille  $u = (u_{i,j})$ ,  $u_x = (u_{x,i,j})$ ,  $u_y = (u_{y,i,j})$ ,  $0 \leq i, j \leq N$ . Pour les vecteurs on utilise les notations vectorielles données dans la partie 3.3.1 (voir figure 3.1).

### 3.6.1 Forme matricielle générale des dérivées au bord

On a vu dans la section 2.3.2 la relation suivante des dérivées au bord

$$e_1 U_{x,L} + e_{N-1} U_{x,R} = \frac{1}{h} \mathcal{A}U - \mathcal{B}U_x + \frac{1}{h} \mathcal{C}(e_1 U_L + e_{N-1} U_R). \quad (3.57)$$

Dans la relation (3.57),  $U_{x,L}$  et  $U_{x,R}$  sont les dérivées par rapport à  $x$  aux points  $x = a$  et  $x = b$  respectivement. Le vecteur  $U \in \mathbb{R}^{(N-1)}$  est associé à la fonction de grille  $u = (u_i)_{0 \leq i \leq N}$ , c'est-à-dire

$$U = \text{vec}_1(u) = \text{vec}_1[u_i]_{1 \leq i \leq N-1} = [u_1, u_2, \dots, u_{N-1}]^T. \quad (3.58)$$

Les vecteurs  $U_L$  et  $U_R$  sont les conditions de Dirichlet aux points  $x = a$  et  $x = b$  respectivement.  $\mathcal{A}$ ,  $\mathcal{B}$  et  $\mathcal{C}$  sont des matrices d'ordre  $(N-1)$ .

En dimension 2, on va déduire de (3.57), une écriture matricielle générale pour toutes les dérivées de bord. En appliquant (3.57) ligne par ligne et colonne par colonne on obtient

$$\begin{cases} e_1 U_{x,L,j} + e_{N-1} U_{x,R,j} = \frac{1}{h} \mathcal{A}U_{:,j} - \mathcal{B}U_{x,:,j} + \frac{1}{h} \mathcal{C}(e_1 U_{L,j} + e_{N-1} U_{R,j}), & 0 \leq j \leq N, \quad (a) \\ e_1 U_{y,i,B} + e_{N-1} U_{y,i,T} = \frac{1}{h} \mathcal{A}U_{i,:} - \mathcal{B}U_{y,i,:} + \frac{1}{h} \mathcal{C}(e_1 U_{i,B} + e_{N-1} U_{i,T}), & 0 \leq i \leq N. \quad (b) \end{cases} \quad (3.59)$$

Ces relations généralisent de façon naturelle (3.57) pour  $i$  ou  $j$  fixé.

**Lemme 3.6.1.** • Les dérivées normales par rapport à  $x$ ,  $U_{x,LB}$ ,  $U_{x,RB}$ ,  $U_{x,LT}$ ,  $U_{x,RT} \in \mathbb{R}$  sur les sommets du  $\Omega = (a, b)^2$  vérifient

$$(i) \begin{cases} e_1 U_{x,LB} + e_{N-1} U_{x,RB} = \frac{1}{h} \mathcal{A}U_B - \mathcal{B}U_{x,B} + \frac{1}{h} \mathcal{C}(e_1 U_{LB} + e_{N-1} U_{RB}), & (a) \\ e_1 U_{x,LT} + e_{N-1} U_{x,RT} = \frac{1}{h} \mathcal{A}U_T - \mathcal{B}U_{x,T} + \frac{1}{h} \mathcal{C}(e_1 U_{LT} + e_{N-1} U_{RT}). & (b) \end{cases} \quad (3.60)$$

• Les dérivées normales par rapport à  $y$ ,  $U_{y,LB}$ ,  $U_{y,LT}$ ,  $U_{y,RB}$ ,  $U_{y,RT} \in \mathbb{R}$  sur les sommets du  $\Omega = (a, b)^2$  vérifient

$$(ii) \begin{cases} e_1 U_{y,LB} + e_{N-1} U_{y,LT} = \frac{1}{h} \mathcal{A}U_L - \mathcal{B}U_{y,L} + \frac{1}{h} \mathcal{C}(e_1 U_{LB} + e_{N-1} U_{LT}), & (a) \\ e_1 U_{y,RB} + e_{N-1} U_{y,RT} = \frac{1}{h} \mathcal{A}U_R - \mathcal{B}U_{y,R} + \frac{1}{h} \mathcal{C}(e_1 U_{RB} + e_{N-1} U_{RT}). & (b) \end{cases} \quad (3.61)$$

**Preuve.** Pour  $j = 0$ , (3.59)<sub>a</sub> donne

$$e_1 U_{x,L,0} + e_{N-1} U_{x,R,0} = \frac{1}{h} \mathcal{A} U_{:,0} - \mathcal{B} U_{x,;,0} + \frac{1}{h} \mathcal{C} (e_1 U_{L,0} + e_{N-1} U_{R,0}). \quad (3.62)$$

D'après la définition 3 on a

$$\begin{cases} U_{:,0} = [u_{1,0}, u_{2,0}, \dots, u_{N-1,0}]^T = U_B, \\ U_{x,;,0} = [u_{x,1,0}, u_{x,2,0}, \dots, u_{x,N-1,0}]^T = U_{x,B}. \end{cases} \quad (3.63)$$

En utilisant les notations de la partie 3.3.1 on obtient (3.60)<sub>a</sub>. De la même manière on démontre (3.60)<sub>b</sub> et (3.61)<sub>a,b</sub>. ■

**Lemme 3.6.2.** • *Les dérivées tangentielles par rapport à  $x$  sur les arêtes “B” et “T” du  $\Omega = (a, b)^2$  vérifient*

$$\begin{cases} U_{x,B} = \frac{1}{h} \mathcal{D} U_B + \frac{1}{h} \mathcal{E} (e_1 U_{LB} + e_{N-1} U_{RB}), & (a) \\ U_{x,T} = \frac{1}{h} \mathcal{D} U_T + \frac{1}{h} \mathcal{E} (e_1 U_{LT} + e_{N-1} U_{RT}). & (b) \end{cases} \quad (3.64)$$

• *De même, les dérivées tangentielles par rapport à  $y$  sur les arêtes “L” et “R” du  $\Omega = (a, b)^2$  vérifient*

$$\begin{cases} U_{y,L} = \frac{1}{h} \mathcal{D} U_L + \frac{1}{h} \mathcal{E} (e_1 U_{LB} + e_{N-1} U_{LT}), & (a) \\ U_{y,R} = \frac{1}{h} \mathcal{D} U_R + \frac{1}{h} \mathcal{E} (e_1 U_{RB} + e_{N-1} U_{RT}). & (b) \end{cases} \quad (3.65)$$

Les matrices  $\mathcal{D} = \mathcal{D}(\mathcal{A}, \mathcal{B})$  et  $\mathcal{E} = \mathcal{E}(\mathcal{B}, \mathcal{C})$  sont données en (2.113).

**Preuve.** Pour  $j = 0$ , (3.49)<sub>b</sub> donne

$$\sigma_x u_{x,i,0} = \delta_x u_{i,0}, \quad 1 \leq i \leq N-1, \quad (3.66)$$

soit

$$\frac{1}{6} u_{x,i-1,0} + \frac{2}{3} u_{x,i,0} + \frac{1}{6} u_{x,i+1,0} = \frac{u_{i+1,0} - u_{i-1,0}}{2h}, \quad 1 \leq i \leq N-1. \quad (3.67)$$

Matriciellement, (3.67) se réécrit

$$\frac{1}{6} e_1 U_{x,LB} + P_s U_{x,B} + \frac{1}{6} e_{N-1} U_{x,RB} = \frac{1}{2h} K U_B - \frac{1}{2h} e_1 U_{LB} + \frac{1}{2h} e_{N-1} U_{RB}. \quad (3.68)$$

On reconnaît (3.60)<sub>a</sub>, on obtient la relation

$$(P_s - \frac{1}{6}\mathcal{B})U_{x,B} = \frac{1}{2h}(K - \frac{1}{3}\mathcal{A})U_B + \frac{1}{2h}(F_2 - \frac{1}{3}\mathcal{C})(e_1U_{LB} + e_{N-1}U_{RB}). \quad (3.69)$$

En utilisant les matrices  $\mathcal{D}$  et  $\mathcal{E}$  en (2.113), (3.69) s'écrit

$$U_{x,B} = \frac{1}{h}\mathcal{D}U_B + \frac{1}{h}\mathcal{E}(e_1U_{LB} + e_{N-1}U_{RB}), \quad (3.70)$$

d'où (3.64)<sub>a</sub>. De la même manière on démontre (3.64)<sub>b</sub> et (3.65)<sub>a,b</sub>. ■

**Proposition 3.3.** *Les vecteurs  $U_{x,L}$ ,  $U_{x,R}$ ,  $U_{y,B}$ ,  $U_{y,T} \in \mathbb{R}^{(N-1)}$ , dérivées normales par rapport à  $x$  le long des quatres arêtes "L", "R", "B", "T" du carré  $\Omega = (a, b)^2$ , vérifient*

$$\begin{cases} e_1 \otimes U_{x,L} + e_{N-1} \otimes U_{x,R} = \frac{1}{h}(\mathcal{A} \otimes I)U - (\mathcal{B} \otimes I)U_x + \frac{1}{h}(\mathcal{C} \otimes I)(e_1 \otimes U_L + e_{N-1} \otimes U_R), & (a) \\ U_{y,B} \otimes e_1 + U_{y,T} \otimes e_{N-1} = \frac{1}{h}(I \otimes \mathcal{A})U - (I \otimes \mathcal{B})U_y + \frac{1}{h}(I \otimes \mathcal{C})(U_B \otimes e_1 + U_T \otimes e_{N-1}). & (b) \end{cases} \quad (3.71)$$

**Preuve.** D'après (3.59)<sub>a</sub>, pour tout  $1 \leq j \leq N-1$  on a

$$e_1U_{x,L,j} + e_{N-1}U_{x,R,j} = \frac{1}{h}\mathcal{A}U_{:,j} - \mathcal{B}U_{x,;,j} + \frac{1}{h}\mathcal{C}(e_1U_{L,j} + e_{N-1}U_{R,j}). \quad (3.72)$$

Pour tout  $W \in \mathbb{R}^{(N-1)}$ , on peut écrire

$$W = \sum_{i=1}^{N-1} e_i W_i. \quad (3.73)$$

Pour tout  $V \in \mathbb{R}^{(N-1)^2}$ , on peut écrire

$$V = \sum_{i=1}^{N-1} (e_i \otimes V_{i,:}) = \sum_{j=1}^{N-1} (V_{:,j} \otimes e_j). \quad (3.74)$$

avec

$$\begin{cases} V_{i,:} = [V_{i,1}, V_{i,2}, \dots, V_{i,N-1}]^T, \\ V_{:,j} = [V_{1,j}, V_{2,j}, \dots, V_{N-1,j}]^T. \end{cases} \quad (3.75)$$

En utilisant (3.72), (3.73) et (3.75), on obtient

$$\begin{aligned}
e_1 \otimes U_{x,L} + e_{N-1} \otimes U_{x,R} &= e_1 \otimes \left( \sum_{j=1}^{N-1} e_j U_{x,L,j} \right) + e_{N-1} \otimes \left( \sum_{j=1}^{N-1} e_j U_{x,R,j} \right) \\
&= \sum_{j=1}^{N-1} (e_1 \otimes e_j U_{x,L,j}) + \sum_{j=1}^{N-1} (e_{N-1} \otimes e_j U_{x,R,j}) \\
&= \sum_{j=1}^{N-1} (e_1 U_{x,L,j} + e_{N-1} U_{x,R,j}) \otimes e_j \\
&= \sum_{j=1}^{N-1} \left( \frac{1}{h} \mathcal{A} U_{:,j} - \mathcal{B} U_{x, :, j} + \mathcal{C} (e_1 U_{L,j} + e_{N-1} U_{R,j}) \right) \otimes e_j \\
&= \frac{1}{h} \sum_{j=1}^{N-1} (\mathcal{A} U_{:,j} \otimes e_j) - \sum_{j=1}^{N-1} (\mathcal{B} U_{x, :, j} \otimes e_j) \\
&\quad - \frac{1}{h} \sum_{j=1}^{N-1} \left( \mathcal{C} (e_1 U_{L,j} + e_{N-1} U_{R,j}) \otimes e_j \right) \\
&= \frac{1}{h} \sum_{j=1}^{N-1} (\mathcal{A} \otimes I) (U_{:,j} \otimes e_j) - \sum_{j=1}^{N-1} (\mathcal{B} \otimes I) (U_{x, :, j} \otimes e_j) \\
&\quad - \frac{1}{h} \sum_{j=1}^{N-1} (\mathcal{C} \otimes I) (e_1 U_{L,j} + e_{N-1} U_{R,j}) \\
&= \frac{1}{h} (\mathcal{A} \otimes I) U - (\mathcal{B} \otimes I) U_x + \frac{1}{h} (\mathcal{C} \otimes I) (e_1 U_L + e_{N-1} U_R),
\end{aligned}$$

d'où (3.71)<sub>a</sub>. De la même manière on démontre (3.71)<sub>b</sub>. ■

**Corollaire 3.6.3.** *Les vecteurs des dérivées normales par rapport à  $x$ ,  $U_{x,L}$  et  $U_{x,R} \in \mathbb{R}^{(N-1)}$  vérifient*

$$\begin{cases} U_{x,L} = \frac{1}{h} (e_1^T \mathcal{A} \otimes I) U - (e_1^T \mathcal{B} \otimes I) U_x + \frac{1}{h} (e_1^T \mathcal{C} \otimes I) (e_1 \otimes U_L + e_{N-1} \otimes U_R), & (a) \\ U_{x,R} = \frac{1}{h} (e_{N-1}^T \mathcal{A} \otimes I) U - (e_{N-1}^T \mathcal{B} \otimes I) U_x + \frac{1}{h} (e_{N-1}^T \mathcal{C} \otimes I) (e_1 \otimes U_L + e_{N-1} \otimes U_R). & (b) \end{cases} \quad (3.76)$$

**Preuve.** Multiplions (3.71)<sub>a</sub> de deux côtés par  $e_1^T \otimes I$ , on obtient

$$e_1^T e_1 \otimes U_{x,L} + e_1^T e_{N-1} \otimes U_{x,R} = \frac{1}{h} (e_1^T \mathcal{A} \otimes I) U - (e_1^T \mathcal{B} \otimes I) U_x + \frac{1}{h} (e_1^T \mathcal{C} \otimes I) (e_1 \otimes U_L + e_{N-1} \otimes U_R). \quad (3.77)$$

Comme  $e_1^T e_{N-1} = 0$ , (3.77) devient

$$e_1^T e_1 \otimes U_{x,L} = \frac{1}{h}(e_1^T \mathcal{A} \otimes I)U - (e_1^T \mathcal{B} \otimes I)U_x + \frac{1}{h}(e_1^T \mathcal{C} \otimes I)(e_1 \otimes U_L + e_{N-1} \otimes U_R), \quad (3.78)$$

or  $e_1^T e_1 \otimes U_{x,L} = 1 \otimes U_{x,L} = U_{x,L}$ , d'où (3.76)<sub>a</sub>. De la même manière on démontre (3.76)<sub>b</sub>. ■

### 3.6.2 Forme matricielle générale du schéma boîte hermitien

Dans cette section, nous établissons l'un des résultats principaux de ce travail qui est la forme matricielle des relations (3.49)<sub>a,b,c</sub>. Rappelons que (3.49)<sub>a,b,c</sub> consiste en la forme mixte du schéma boîte hermitien aux points intérieurs du maillage cartésien. Ce système constitue la base du schéma HB en dimension 2, il consiste en :

- L'équation de conservation discrète

$$-(\sigma_y \circ \delta_x)u_{x,i,j} - (\sigma_x \circ \delta_y)u_{y,i,j} = \Pi^0 f_{i,j}, \quad 1 \leq i, j \leq N-1. \quad (3.79)$$

- Les deux équations pour le gradient discret  $(u_x, u_y)$

$$\begin{cases} \sigma_x u_{x,i,j} = \delta_x u_{i,j}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, & (a) \\ \sigma_y u_{y,i,j} = \delta_y u_{i,j}, & 1 \leq j \leq N-1, \quad 0 \leq i \leq N. & (b) \end{cases} \quad (3.80)$$

Comme déjà mentionné à la section 3.5.2, les relations (3.79) et (3.80) représentent  $(N-1)^2 + 2(N-1)(N+1)$  équations pour  $3(N+1)^2$  inconnues. Pour avoir le même nombre d'équations et d'inconnues on a ajouté le système formé par les conditions limites naturelles et l'un de deux exemples d'approximations des dérivées normales au bord (3.53) et (3.55).

Pour obtenir une forme matricielle générale du schéma boîte on va considérer les approximations des dérivées au bord données par leur forme matricielle générale dans la section 3.6.1.

**Proposition 3.4.** *Les vecteurs dérivés  $U_x = \text{vec}_2(u_x)$  et  $U_y = \text{vec}_2(u_y)$  s'expriment en fonction de  $U = \text{vec}_2(u)$  sous la forme*

$$\begin{cases} U_x = \frac{1}{h}(\mathcal{D} \otimes I)U + \frac{1}{h}(\mathcal{E} \otimes I)(e_1 \otimes U_L + e_{N-1} \otimes U_R), & (a) \\ U_y = \frac{1}{h}(I \otimes \mathcal{D})U + \frac{1}{h}(I \otimes \mathcal{E})(U_B \otimes e_1 + U_T \otimes e_{N-1}), & (b) \end{cases} \quad (3.81)$$

où les matrices  $\mathcal{D} = \mathcal{D}(\mathcal{A}, \mathcal{B})$  et  $\mathcal{E} = \mathcal{E}(\mathcal{B}, \mathcal{C})$  sont données en (2.113).

**Preuve.** On a

$$U_x = \sum_{j=1}^{N-1} (U_{x,;j} \otimes e_j), \quad (3.82)$$

avec  $U_{x,;j}$  est la dérivée hermitienne par rapport à  $x$  sur la droite  $y = jh$ . (3.80)<sub>a</sub> donne

$$U_{x,;j} = \frac{1}{h} \mathcal{D}U_{:,j} + \frac{1}{h} \mathcal{E}(e_1 U_{L,j} + e_{N-1} U_{R,j}), \quad (3.83)$$

donc

$$\begin{aligned} U_x &= \sum_{j=1}^{N-1} \left( \frac{1}{h} \mathcal{D}U_{:,j} + \frac{1}{h} \mathcal{E}(e_1 U_{L,j} + e_{N-1} U_{R,j}) \right) \otimes e_j \\ &= \frac{1}{h} (\mathcal{D} \otimes I) \sum_{j=1}^{N-1} (U_{:,j} \otimes e_j) + \frac{1}{h} (\mathcal{E} \otimes I) \sum_{j=1}^{N-1} (e_1 U_{L,j} + e_{N-1} U_{R,j}) \otimes e_j. \end{aligned}$$

La décomposition de  $U$  à l'aide des vecteurs  $U_{:,j}$  est

$$\sum_{j=1}^{N-1} (U_{:,j} \otimes e_j) = U \quad (3.84)$$

et

$$\begin{aligned} \sum_{j=1}^{N-1} (e_1 U_{L,j} + e_{N-1} U_{R,j}) \otimes e_j &= \sum_{j=1}^{N-1} (e_1 \otimes I)(U_{L,j} \otimes e_j) + \sum_{j=1}^{N-1} (e_{N-1} \otimes I)(U_{R,j} \otimes e_j) \\ &= (e_1 \otimes I) \sum_{j=1}^{N-1} U_{L,j} e_j + (e_{N-1} \otimes I) \sum_{j=1}^{N-1} U_{R,j} e_j \\ &= (e_1 \otimes I) U_L + (e_{N-1} \otimes I) U_R \\ &= (e_1 \otimes U_L) + (e_{N-1} \otimes U_R). \end{aligned}$$

Donc

$$U_x = \frac{1}{h} (\mathcal{D} \otimes I) U + \frac{1}{h} (\mathcal{E} \otimes I) \left( (e_1 \otimes U_L) + (e_{N-1} \otimes U_R) \right), \quad (3.85)$$

d'où (3.81)<sub>a</sub>. De la même manière on démontre (3.81)<sub>b</sub>. ■

**Proposition 3.5** (Forme matricielle générale du schéma boîte en dimension 2). *Le schéma boîte hermitien formé des équations (3.49) et des approximations (3.59) des dérivées au bord, possède la forme matricielle suivante,*

$$\frac{1}{h^2} (\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H}) U = F - G_x - G_y, \quad (3.86)$$

où

- Le vecteur  $F$  correspond au terme source moyenné en (3.40),

$$F = \text{vec}_2(\Pi^0 f). \quad (3.87)$$

- Les vecteurs  $G_x, G_y \in \mathbb{R}^{(N-1)^2}$  correspondent aux conditions au bord non homogènes,

$$\left\{ \begin{array}{l} G_x = \frac{1}{h^2}(\mathcal{G} \otimes P_s)(e_1 \otimes U_L + e_{N-1} \otimes U_R) \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}U_B + \mathcal{G}(e_1 U_{LB} + e_{N-1} U_{RB}) \right) \otimes e_1 \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}U_T + \mathcal{G}(e_1 U_{LT} + e_{N-1} U_{RT}) \right) \otimes e_{N-1}, \end{array} \right. \quad (3.88)$$

et

$$\left\{ \begin{array}{l} G_y = \frac{1}{h^2}(P_s \otimes \mathcal{G})(U_B \otimes e_1 + U_T \otimes e_{N-1}) \\ \quad + \frac{1}{6h^2} e_1 \otimes \left( \mathcal{H}U_L + \mathcal{G}(e_1 U_{LB} + e_{N-1} U_{LT}) \right) \\ \quad + \frac{1}{6h^2} e_{N-1} \otimes \left( \mathcal{H}U_R + \mathcal{G}(e_1 U_{RB} + e_{N-1} U_{RT}) \right). \end{array} \right. \quad (3.89)$$

- $\mathcal{H} = \mathcal{H}(\mathcal{A}, \mathcal{B})$  et  $\mathcal{G} = \mathcal{G}(\mathcal{B}, \mathcal{C})$  les matrices données en (2.117).

Pour démontrer la proposition 3.5, on a besoin de deux lemmes suivants.

**Lemme 3.6.4.** • La fonction de grille  $(\sigma_y \circ \delta_x)u_x$  dans (3.79) s'écrit matriciellement en fonction de  $U_x$  sous la forme

$$\left\{ \begin{array}{l} \text{vec}_2((\sigma_y \circ \delta_x)u_x) = \frac{1}{2h}(K \otimes P_s)U_x \quad (a) \\ \quad + \frac{1}{2h}(F_2 \otimes P_s)(e_1 \otimes U_{x,L} + e_{N-1} \otimes U_{x,R}) \quad (b) \\ \quad + \frac{1}{12h}(I \otimes e_1)KU_{x,B} + \frac{1}{12h}(I \otimes e_1)F_2(e_1 U_{x,LB} + e_{N-1} U_{x,RB}) \quad (c) \\ \quad + \frac{1}{12h}(I \otimes e_{N-1})KU_{x,T} + \frac{1}{12h}(I \otimes e_{N-1})F_2(e_1 U_{x,LT} + e_{N-1} U_{x,RT}). \quad (d) \end{array} \right. \quad (3.90)$$



• La fonction de grille  $(\sigma_x \circ \delta_y)u_y$  dans (3.79) s'écrit matriciellement en fonction de  $U_y$  sous la forme

$$\left\{ \begin{aligned} \text{vec}_2((\sigma_x \circ \delta_y)u_y) &= \frac{1}{2h}(P_s \otimes K)U_y \\ &+ \frac{1}{2h}(P_s \otimes F_2)(U_{y,B} \otimes e_1 + U_{y,T} \otimes e_{N-1}) \\ &+ \frac{1}{12h}(e_1 \otimes I)KU_{y,L} + \frac{1}{12h}(e_1 \otimes I)F_2(e_1U_{y,LB} + e_{N-1}U_{y,LT}) \\ &+ \frac{1}{12h}(e_{N-1} \otimes I)KU_{y,R} + \frac{1}{12h}(e_{N-1} \otimes I)F_2(e_1U_{y,RB} + e_{N-1}U_{y,RT}). \end{aligned} \right. \quad (3.91)$$

**Preuve.** On montre (i). Pour  $v = \delta_x u_x$ , (3.35)<sub>b</sub> donne

$$\text{vec}_2((\sigma_y \circ \delta_x)u_x) = (I \otimes P_s)V + \frac{1}{6}(I \otimes e_1)V_B + \frac{1}{6}(I \otimes e_{N-1})V_T. \quad (3.92)$$

avec  $V = \text{vec}_2(v) = \text{vec}_2(\delta_x u_x)$ ,  $V_B = \text{vec}_1[\delta_x u_{x,i,0}]_{1 \leq i \leq N-1}$  et  $V_T = \text{vec}_1[\delta_x u_{x,i,N}]_{1 \leq i \leq N-1}$ .

On peut facilement démontrer que

$$\left\{ \begin{aligned} V_B &= \frac{1}{2h}KU_{x,B} + \frac{1}{2h}F_2(e_1U_{x,LB} + e_{N-1}U_{x,RB}), \\ V_T &= \frac{1}{2h}KU_{x,T} + \frac{1}{2h}F_2(e_1U_{x,LT} + e_{N-1}U_{x,RT}). \end{aligned} \right. \quad (3.93)$$

De plus, d'après le lemme 3.4.1, la forme matricielle de  $\text{vec}_2(\delta_x u_x) \in \mathbb{R}^{(N-1)^2}$  est

$$\begin{aligned} \text{vec}_2(\delta_x u_x) &= \frac{1}{2h}(K \otimes I)U_x - \frac{1}{2h}(e_1 \otimes I)U_{x,L} + \frac{1}{2h}(e_{N-1} \otimes I)U_{x,R} \\ &= \frac{1}{2h}(K \otimes I)U_x + \frac{1}{2h}(F_2 \otimes I)(e_1 \otimes U_{x,L} + e_{N-1} \otimes U_{x,R}). \end{aligned} \quad (3.94)$$

En utilisant (3.93) et (3.94) dans (3.92), on obtient

$$\begin{aligned} \text{vec}_2((\sigma_y \circ \delta_x)u_x) &= \frac{1}{2h}(I \otimes P_s) \left( (K \otimes I)U_x + (F_2 \otimes I)(e_1 \otimes U_{x,L} + e_{N-1} \otimes U_{x,R}) \right) \\ &+ \frac{1}{12h}(I \otimes e_1) \left( KU_{x,B} + F_2(e_1U_{x,LB} + e_{N-1}U_{x,RB}) \right) \\ &+ \frac{1}{12h}(I \otimes e_{N-1}) \left( KU_{x,T} + F_2(e_1U_{x,LT} + e_{N-1}U_{x,RT}) \right). \end{aligned}$$

Ce qui est équivalent à

$$\begin{aligned}
\text{vec}_2\left((\sigma_y \circ \delta_x)u_x\right) &= \frac{1}{2h}(K \otimes P_s)U_x \\
&+ \frac{1}{2h}(F_2 \otimes P_s)(e_1 \otimes U_{x,L} + e_{N-1} \otimes U_{x,R}) \\
&+ \frac{1}{12h}(I \otimes e_1)KU_{x,B} + \frac{1}{12h}(I \otimes e_1)F_2(e_1U_{x,LB} + e_{N-1}U_{x,RB}) \\
&+ \frac{1}{12h}(I \otimes e_{N-1})KU_{x,T} + \frac{1}{12h}(I \otimes e_{N-1})F_2(e_1U_{x,LT} + e_{N-1}U_{x,RT}).
\end{aligned}$$

d'où (3.90). De la même manière on démontre (3.91).  $\blacksquare$

**Lemme 3.6.5.** *En considérant les approximations (3.59) des dérivées au bord de  $\Omega = (a,b)^2$ . Les équations (3.80) permettent d'écrire les fonctions de grille  $(\sigma_y \circ \delta_x)u_x$  et  $(\sigma_x \circ \delta_y)u_y$  sous les formes matricielles suivantes :*

$$\left\{ \begin{aligned}
\text{vec}_2\left((\sigma_y \circ \delta_x)u_x\right) &= -\frac{1}{h^2}(\mathcal{H} \otimes P_s)U - \frac{1}{h^2}(\mathcal{G} \otimes P_s)(e_1 \otimes U_L + e_{N-1} \otimes U_R) \\
&- \frac{1}{6h^2}\mathcal{H}U_B \otimes e_1 - \frac{1}{6h^2}\mathcal{G}(e_1U_{LB} + e_{N-1}U_{RB}) \otimes e_1 \\
&- \frac{1}{6h^2}\mathcal{H}U_T \otimes e_{N-1} - \frac{1}{6h^2}\mathcal{G}(e_1U_{LT} + e_{N-1}U_{RT}) \otimes e_{N-1}.
\end{aligned} \right. \quad (3.95)$$

$$\left\{ \begin{aligned}
\text{vec}_2\left((\sigma_x \circ \delta_y)u_y\right) &= -\frac{1}{h^2}(P_s \otimes \mathcal{H})U - \frac{1}{h^2}(P_s \otimes \mathcal{G})(U_B \otimes e_1 + U_T \otimes e_{N-1}) \\
&- \frac{1}{6h^2}e_1 \otimes \mathcal{H}U_L - \frac{1}{6h^2}e_1 \otimes \mathcal{G}(e_1U_{LB} + e_{N-1}U_{LT}) \\
&- \frac{1}{6h^2}e_{N-1} \otimes \mathcal{H}U_R - \frac{1}{6h^2}e_{N-1} \otimes \mathcal{G}(e_1U_{RB} + e_{N-1}U_{RT}).
\end{aligned} \right. \quad (3.96)$$

**Preuve.** On détaille chacun des quatre termes (a), (b), (c), (d) dans (3.90) et on les exprime en terme de l'inconnue primale  $U$  et des données au bord.

- *Terme (a)*

En remplaçant  $U_x$  par sa forme (3.81)<sub>a</sub>, le terme (a) s'écrit

$$\begin{aligned}
(a) &= \frac{1}{2h}(K \otimes P_s)U_x \\
&= \frac{1}{2h^2}(K \otimes P_s)\left((\mathcal{D} \otimes I)U + (\mathcal{E} \otimes I)(e_1 \otimes U_L + e_{N-1} \otimes U_R)\right).
\end{aligned}$$

- *Terme (b)*

Le terme (b) correspond à la dérivée normale le long de bord. D'après (3.71) et (3.81)<sub>a</sub>,

il s'écrit sous la forme

$$\begin{aligned}
(b) &= \frac{1}{2h}(F_2 \otimes P_s)(e_1 \otimes U_{x,L} + e_{N-1} \otimes U_{x,R}) \\
&= \frac{1}{2h^2}(F_2 \otimes P_s) \left\{ (\mathcal{A} \otimes I)U - h(\mathcal{B} \otimes I)U_x + (\mathcal{C} \otimes I)(e_1 \otimes U_L + e_{N-1} \otimes U_R) \right\} \\
&= \frac{1}{2h^2}(F_2 \otimes P_s) \left\{ (\mathcal{A} \otimes I)U - (\mathcal{B} \otimes I) \left( (\mathcal{D} \otimes I)U + (\mathcal{E} \otimes I)(e_1 \otimes U_L + e_{N-1} \otimes U_R) \right) \right. \\
&\quad \left. + (\mathcal{C} \otimes I)(e_1 \otimes U_L + e_{N-1} \otimes U_R) \right\}.
\end{aligned}$$

En regroupant les termes en  $U$  et  $e_1 \otimes U_L + e_{N-1} \otimes U_R$  on obtient

$$(b) = \frac{1}{2h^2} \left\{ \left( F_2(\mathcal{A} - \mathcal{B}\mathcal{D}) \otimes P_s \right) U + \left( F_2(\mathcal{C} - \mathcal{B}\mathcal{E}) \otimes P_s \right) (e_1 \otimes U_L + e_{N-1} \otimes U_R) \right\}. \quad (3.97)$$

• Terme (c)

Le terme (c) correspond aux dérivées tangentielles le long du côté "B". En utilisant (3.60) et (3.64), le terme (c) s'écrit

$$\begin{aligned}
(c) &= \frac{1}{12h^2}(I \otimes e_1)K \left( \mathcal{D}U_B + \mathcal{E}(e_1U_{LB} + e_{N-1}U_{RB}) \right) \\
&\quad + \frac{1}{12h^2}(I \otimes e_1)F_2 \left( \mathcal{A}U_B - h\mathcal{B}U_{x,B} + \mathcal{C}(e_1U_{LB} + e_{N-1}U_{RB}) \right).
\end{aligned}$$

En utilisant à nouveau (3.64), le terme (c) s'exprime sous la forme

$$\begin{aligned}
(c) &= \frac{1}{12h^2}(I \otimes e_1)(K\mathcal{D} + F_2\mathcal{A})U_B - \frac{1}{12h^2}(I \otimes e_1)F_2\mathcal{B} \left( \mathcal{D}U_B + \mathcal{E}(e_1U_{LB} + e_{N-1}U_{RB}) \right) \\
&\quad + \frac{1}{12h^2}(I \otimes e_1) \left( K\mathcal{E} + F_2\mathcal{C} \right) \left( e_1U_{LB} + e_{N-1}U_{RB} \right) \\
&= \frac{1}{12h^2}(I \otimes e_1) \left( K\mathcal{D} + F_2(\mathcal{A} - \mathcal{B}\mathcal{D}) \right) U_B \\
&\quad + \frac{1}{12h^2}(I \otimes e_1) \left( K\mathcal{E} + F_2(\mathcal{C} - \mathcal{B}\mathcal{E}) \right) \left( e_1U_{LB} + e_{N-1}U_{RB} \right).
\end{aligned}$$

En utilisant l'expression des matrices  $\mathcal{H}$  et  $\mathcal{G}$  dans (2.116), on obtient

$$(c) = -\frac{1}{6h^2}\mathcal{H}U_B \otimes e_1 - \frac{1}{6h^2}\mathcal{G} \left( e_1U_{LB} + e_{N-1}U_{RB} \right) \otimes e_1. \quad (3.98)$$

- Terme (d)

De manière analogue, le terme (d) s'écrit

$$(d) = -\frac{1}{6h^2}\mathcal{H}U_T \otimes e_{N-1} - \frac{1}{6h^2}\mathcal{G}\left(e_1U_{LT} + e_{N-1}U_{RT}\right) \otimes e_{N-1}. \quad (3.99)$$

En sommant les quatre termes (a), (b), (c), (d) on obtient (3.95). De la même manière on démontre (3.96). ■

**Preuve de la proposition 3.5.** L'équation de conservation (3.49)<sub>a</sub> donne

$$-\text{vec}_2\left((\sigma_y \circ \delta_x)u_x\right) - \text{vec}_2\left((\sigma_x \circ \delta_y)u_y\right) = \text{vec}_2(\Pi^0 f). \quad (3.100)$$

En remplaçant  $\text{vec}_2\left((\sigma_y \circ \delta_x)u_x\right)$  et  $\text{vec}_2\left((\sigma_x \circ \delta_y)u_y\right)$  par leurs formes matricielles (3.95) et (3.96), on obtient

$$-\frac{1}{h^2}(\mathcal{H} \otimes P_s)U - \frac{1}{h^2}(P_s \otimes \mathcal{H})U = F - G_x - G_y, \quad (3.101)$$

avec  $F = \text{vec}_2(\Pi^0 f)$ .  $G_x$  et  $G_y$  sont donnés en (3.88) et (3.89). ■

Le paramètre laissé à l'utilisateur en (3.95) consiste en le choix de l'approximation discrète de  $\frac{\partial u}{\partial n}$  sur  $\partial\Omega$ . Ceci se traduit matriciellement par le choix des matrices  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$  en (3.59) qui à leur tour déterminent les matrices  $\mathcal{D}$ ,  $\mathcal{E}$  en (2.113) et  $\mathcal{H}$ ,  $\mathcal{G}$  en (2.117).

## 3.7 Le schéma boîte hermitien HB1 en dimension 2

### 3.7.1 Forme matricielle du schéma HB1

Le Schéma HB1 a été établi en (3.54). Il résulte de la proposition 3.5 qu'il possède la forme matricielle suivante

$$\frac{1}{h^2}(\mathcal{H}_1 \otimes P_s + P_s \otimes \mathcal{H}_1)U = F - G_{x,1} - G_{y,1}, \quad (3.102)$$

où

- Le vecteur  $F$  correspond au terme source moyenné en (3.40),

$$F = \text{vec}_2(\Pi^0 f). \quad (3.103)$$

- Les vecteurs  $G_{x,1}, G_{y,1} \in \mathbb{R}^{(N-1)^2}$  correspondent aux conditions de bord non homogènes,

$$\left\{ \begin{array}{l} G_{x,1} = \frac{1}{h^2}(\mathcal{G}_1 \otimes P_s)(e_1 \otimes U_L + e_{N-1} \otimes U_R) \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}_1 U_B + \mathcal{G}_1(e_1 U_{LB} + e_{N-1} U_{RB}) \right) \otimes e_1 \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}_1 U_T + \mathcal{G}_1(e_1 U_{LT} + e_{N-1} U_{RT}) \right) \otimes e_{N-1}, \end{array} \right. \quad (3.104)$$

et

$$\left\{ \begin{array}{l} G_{y,1} = \frac{1}{h^2}(P_s \otimes \mathcal{G}_1)(U_B \otimes e_1 + U_T \otimes e_{N-1}) \\ \quad + \frac{1}{6h^2} e_1 \otimes \left( \mathcal{H}_1 U_L + \mathcal{G}_1(e_1 U_{LB} + e_{N-1} U_{LT}) \right) \\ \quad + \frac{1}{6h^2} e_{N-1} \otimes \left( \mathcal{H}_1 U_R + \mathcal{G}_1(e_1 U_{RB} + e_{N-1} U_{RT}) \right). \end{array} \right. \quad (3.105)$$

- Les matrices  $\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1$  sont celles données en (2.43). Elles traduisent la forme particulières des approximations des dérivées au bord.  $\mathcal{H}_1 = \mathcal{H}_1(\mathcal{A}_1, \mathcal{B}_1)$  et  $\mathcal{G}_1 = \mathcal{G}_1(\mathcal{B}_1, \mathcal{C}_1)$  les matrices données en (2.48).

On définit la matrice  $H_1$  par

$$H_1 = \mathcal{H}_1 \otimes P_s + P_s \otimes \mathcal{H}_1. \quad (3.106)$$

Le système (3.102) devient

$$H_1 U = h^2(F - G_{x,1} - G_{y,1}). \quad (3.107)$$

Les matrices  $P_s$  et  $\mathcal{H}_1$  sont symétriques définies positives diagonales dans la base  $Z^k$ , avec les valeurs propres  $\alpha_k = 1 - \lambda_k/6$ ,  $\mu_k$  respectivement. On en déduit que la matrice  $H_1$  est symétrique définie positive, également diagonalisable dans la base  $Z^k \otimes Z^l$  avec la décomposition spectrale

$$H_1 = (Z \otimes Z) \left( \mathcal{M} \otimes (I - \Lambda/6) + (I - \Lambda/6) \otimes \mathcal{M} \right) (Z^T \otimes Z^T), \quad (3.108)$$

où

$$\left\{ \begin{array}{l} \mathcal{M} = \text{diag}(\mu_1, \mu_2, \dots, \mu_{N-1}), \\ \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{N-1}). \end{array} \right. \quad (3.109)$$

Les valeurs propres de la matrice  $H_1$  sont

$$\beta_{k,l} = \mu_k \alpha_l + \alpha_k \mu_l, \quad 1 \leq k, l \leq N-1. \quad (3.110)$$

On montre dans la section suivante comment tirer parti de (3.108) pour définir un algorithme rapide pour la résolution de (3.102).

### 3.7.2 Algorithme rapide pour le schéma HB1

Nous présentons un algorithme de résolution rapide du système linéaire (3.102). On tire parti pour cela de la forme diagonale de la matrice  $H_1$  en (3.106). La résolution repose sur la décomposition du vecteur  $h^2(F - G_{x,1} - G_{y,1}) \in \mathbb{R}^{(N-1)^2}$  dans la base orthonormale  $Z^k \otimes Z^l$  de la matrice  $\mathcal{H}_1 \otimes P_s + P_s \otimes \mathcal{H}_1$ . Elle est effectuée par la transformée en sinus rapide.

**Algorithme 2. – Etape 1 :** *Ecrire le vecteur  $\bar{F} = h^2(F - G_{x,1} - G_{y,1})$  dans la base  $Z^k \otimes Z^l$ ,  $1 \leq k, l \leq N-1$ ,*

$$\bar{F}^Z = \sum_{k=1}^{N-1} \sum_{l=1}^{N-1} \bar{F}_{k,l}^Z (Z^k \otimes Z^l), \quad (3.111)$$

avec

$$\bar{F}_{k,l}^Z = (\bar{F}, Z^k \otimes Z^l), \quad 1 \leq k, l \leq N-1. \quad (3.112)$$

*Cette étape est effectuée par la transformation en sinus rapide [27] (voir l'annexe).*

– **Etape 2 :** *Calcul des composantes  $U_{k,l}^Z$  de la solution dans l'espace de Fourier par*

$$U_{k,l}^Z = \frac{\bar{F}_{k,l}^Z}{\beta_{k,l}}, \quad 1 \leq k, l \leq N-1, \quad (3.113)$$

avec  $\beta_{k,l}$  les valeurs propres de la matrice  $H_1$  données en (3.110).

– **Etape 3 :** *Calcul des composantes de  $U = (U_{i,j})_{1 \leq i, j \leq N-1} \in \mathbb{R}^{(N-1)^2}$  dans la base canonique*

$$U_{i,j} = \sum_{k,l=1}^{N-1} U_{k,l}^Z Z_i^k Z_j^l. \quad (3.114)$$

*Cette étape est effectuée par une transformation inverse en sinus rapide (voir l'annexe).*

Les étapes 1 et 3 sont  $O(N^2 \log_2(N))$  parcequ'elles sont effectuées par une transformation en sinus rapide et l'étape 2 est  $O(N^2)$ , ce qui donne une complexité algorithmique globale de  $O(N^2 \log_2(N))$ .

## 3.8 Mise en oeuvre du schéma HB1

### 3.8.1 Introduction

Dans ce qui précède, on a vu que le schéma boîte hermitien se compose de deux types d'approximations :

- L'approximation interne constituée par les trois relations (3.49)<sub>a,b,c</sub>. On a vu que l'ordre de consistance de ces relations est quatre, cela est dû à l'approximation par la formule de Simpson qui est d'ordre 4 (cf proposition 1.2).
- L'approximation au bord de  $\frac{\partial u}{\partial n}$ . On a donné deux exemples d'approximation de  $\frac{\partial u}{\partial n}$ . Le premier exemple (3.50), celui du schéma HB1, est d'ordre 1. Le deuxième exemple (3.51), celui du schéma HB2, est d'ordre 3. Dans la section 3.10 on verra que le schéma HB2 est d'ordre 4 en pratique pour  $u$  et  $\nabla u$ . Autrement dit, le schéma global a une erreur d'ordre toujours supérieur à ce que l'on pourrait déduire de l'ordre d'approximation au bord. Noter que l'analyse numérique théorique du schéma HB2 semble à priori difficile.

Dans la partie 3.8.2, on donne des détails de l'implémentation des algorithmes 2 et 3. Les parties 3.8.3 et 3.10.3 sont consacrées à la présentation d'une série de résultats numériques obtenus avec les schémas HB1 et HB2 pour le problème de Poisson (3.37).

- Pour le schéma HB1, on observe en général un ordre de convergence 2 pour  $u$  et  $\nabla u$ . Un ordre de convergence 4 est observé pour  $u$  et  $\nabla u$  dans certains cas particuliers.
- Pour le schéma HB2, on observe un ordre de convergence 4 pour  $u$  et  $\nabla u$ .

On présente également les courbes de convergence de chacun des schémas boîte HB1, HB2 ainsi que le schéma différence finie standard à cinq points (DF2).

### 3.8.2 Détails de l'implémentation

On présente les détails de l'implémentation de la librairie HBoxLib développée dans le cadre de ce travail. Cette librairie contient des fonctions de calcul rapide de l'approximation par l'opérateur de Simpson et des fonctions de calcul des produits kroneckeriens

avec des matrices particulières comme la matrice identité. Le développement a été effectué en Fortran 90. Les calculs sont effectués sur une machine de bureau.

Les ingrédients pour appliquer l'algorithme 2 sont les suivants :

1- *Formule de Simpson en dimension 2*

Pour le second membre  $F = \text{vec}_2(\Pi^0 f)$  en (3.87),  $f(x, y)$  est connue par  $-\Delta u(x, y) = f(x, y)$ . Soit  $f = (f_{i,j})_{0 \leq i,j \leq N} \in L_h^2$  la fonction de grille associée à  $f(x, y)$ ,

$$f_{i,j} = f(x_i, y_j), \quad 0 \leq i, j \leq N. \quad (3.115)$$

On évalue  $\Pi^0 f = (\Pi^0 f_{i,j})_{1 \leq i,j \leq N-1} \in L_{h,0}^2$  en approchant l'opérateur de moyen  $\Pi^0$  en (3.40) par la formule de Simpson (tensorielle),

$$\left\{ \begin{array}{l} \Pi^0 f_{i,j} \simeq (\sigma_y \circ \sigma_x) f_{i,j} \\ = \frac{1}{36} f_{i-1,j-1} + \frac{2}{18} f_{i-1,j} + \frac{1}{36} f_{i-1,j+1} \\ + \frac{2}{18} f_{i,j-1} + \frac{4}{9} f_{i,j} + \frac{2}{18} f_{i,j+1} \\ + \frac{1}{36} f_{i+1,j-1} + \frac{2}{18} f_{i+1,j} + \frac{1}{36} f_{i+1,j+1}, \quad 1 \leq i, j \leq N-1. \end{array} \right. \quad (3.116)$$

2- *Produit kroneckerien*

Les vecteurs  $U_L, U_R, U_B, U_T \in \mathbb{R}^{(N-1)}$  des données de Dirichlet aux quatres arêtes du carré ainsi que les valeurs de  $u$  aux quatres sommets du carré,  $U_{LB}, U_{LT}, U_{RB}, U_{RT}$ , sont connus par la condition de Dirichlet  $u = g$  au bord. Numériquement, le produit kroneckerien de deux matrices quelconques d'ordre  $N-1$  nécessite normalement  $(N-1)^4$  opérations, ce qui est prohibitif. Pour évaluer un tel produit, on calcule les vecteurs  $G_x, G_y$  en (3.88) en utilisant (3.22). Cela donne :

$$\left\{ \begin{array}{l} G_x = \frac{1}{h^2} (\mathcal{G}e_1 \otimes P_s U_L + \mathcal{G}e_{N-1} \otimes P_s U_R) \\ + \frac{1}{6h^2} \left( \mathcal{H}U_B + \mathcal{G}(e_1 U_{LB} + e_{N-1} U_{RB}) \right) \otimes e_1 \\ + \frac{1}{6h^2} \left( \mathcal{H}U_T + \mathcal{G}(e_1 U_{LT} + e_{N-1} U_{RT}) \right) \otimes e_{N-1}, \end{array} \right. \quad (3.117)$$

et

$$\left\{ \begin{array}{l} G_y = \frac{1}{h^2} (P_s U_B \otimes \mathcal{G}e_1 + P_s U_T \otimes \mathcal{G}e_{N-1}) \\ + \frac{1}{6h^2} e_1 \otimes \left( \mathcal{H}U_L + \mathcal{G}(e_1 U_{LB} + e_{N-1} U_{LT}) \right) \\ + \frac{1}{6h^2} e_{N-1} \otimes \left( \mathcal{H}U_R + \mathcal{G}(e_1 U_{RB} + e_{N-1} U_{RT}) \right). \end{array} \right. \quad (3.118)$$



(3.117) et (3.118) contiennent seulement des produits kroneckeriens de vecteurs, ce qui est numériquement intéressant car ils nécessitent seulement  $(N-1)^2$  opérations.

L'implémentation des vecteurs dérivés  $U_x$  et  $U_y$  donnés dans la proposition 3.4 est effectuée en utilisant également la propriété (3.22) dans les produits kroneckeriens  $(\mathcal{E} \otimes I)(e_1 \otimes U_L + e_{N-1} \otimes U_R)$  et  $(I \otimes \mathcal{E})(U_B \otimes e_1 + U_T \otimes e_{N-1}) \in \mathbb{R}^{(N-1)^2}$ . C'est-à-dire

$$\begin{cases} U_x = \frac{1}{h} \left( (\mathcal{D} \otimes I)U + (\mathcal{E}e_1 \otimes U_L) + (\mathcal{E}e_{N-1} \otimes U_R) \right), & (a) \\ U_y = \frac{1}{h} \left( (I \otimes \mathcal{D})U + (U_B \otimes \mathcal{E}e_1) + (U_T \otimes \mathcal{E}e_{N-1}) \right), & (b) \end{cases} \quad (3.119)$$

où les matrices  $\mathcal{D} = \mathcal{D}(\mathcal{A}, \mathcal{B})$  et  $\mathcal{E} = \mathcal{E}(\mathcal{B}, \mathcal{C})$  sont données en (2.113).

Pour les vecteurs  $(\mathcal{D} \otimes I)U$  et  $(I \otimes \mathcal{D})U \in \mathbb{R}^{(N-1)^2}$ , on a implémenté des fonctions dans la librairie HBoxLib qui nous permet de les calculer avec un coût de  $O(N^2)$ .

### 3- Transformée en sinus (dimension 1)

Soit TS1 l'application suivante

$$\begin{aligned} \mathbb{R}^{(N-1)} &\longrightarrow \mathbb{R}^{(N-1)} \\ x = (x_i)_{1 \leq i \leq N-1} &\longrightarrow y = \text{TS1}(x) \end{aligned} \quad (3.120)$$

tel que  $y = (y_j)_{1 \leq j \leq N-1}$  avec  $y_j = \sum_{i=1}^{N-1} x_i \sin\left(\frac{ij\pi}{N}\right)$ ,  $1 \leq j \leq N-1$ .

Numériquement,  $y$  peut être calculé par un algorithme rapide avec un coût  $O(N \log_2(N))$ .

Ici, cf annexe pour l'algorithme rapide de calcul de  $\text{TS1}(x)$ .

### 4- Transformée en sinus (dimension 2)

Soit TS2 l'application suivante

$$\begin{cases} \mathbb{R}^{(N-1)} \times \mathbb{R}^{(N-1)} &\longrightarrow \mathbb{R}^{(N-1)^2} \\ x = (x_{i,j})_{1 \leq i,j \leq N-1} &\longrightarrow y = \text{TS2}(x) \end{cases} \quad (3.121)$$

tel que  $y = (y_{k,l})_{1 \leq k,l \leq N-1}$  avec

$$y_{k,l} = \sum_{i,j=1}^{N-1} x_{i,j} \sin\left(\frac{ik\pi}{N}\right) \sin\left(\frac{jl\pi}{N}\right), \quad 1 \leq k, l \leq N-1. \quad (3.122)$$

Numériquement,  $y$  peut être calculé par un algorithme rapide avec un coût  $O(N^2 \log_2(N))$ , cf annexe pour l'algorithme rapide de calcul de  $\text{TS2}(x)$ .

### 4- Transformée en sinus (dimension 3)

Soit TS3 l'application suivante

$$\begin{aligned} \mathbb{R}^{(N-1)}\mathbb{R}^{(N-1)} \times \mathbb{R}^{(N-1)} &\longrightarrow \mathbb{R}^{(N-1)^2} \\ x = (x_{i,j,k})_{1 \leq i,j,k \leq N-1} &\longrightarrow y = TS3(x) \end{aligned} \quad (3.123)$$

tel que  $y = (y_{l,m,n})_{1 \leq l,m,n \leq N-1}$  avec

$$y_{l,m,n} = \sum_{i,j,k=1}^{N-1} x_{i,j,k} \sin\left(\frac{il\pi}{N}\right) \sin\left(\frac{jm\pi}{N}\right) \sin\left(\frac{kn\pi}{N}\right), \quad 1 \leq l,m,n \leq N-1. \quad (3.124)$$

Numériquement,  $y$  peut être calculé par un algorithme rapide avec un coût  $O(N^3 \log_2(N))$ , cf annexe pour l'algorithme rapide de calcul de TS3( $x$ ).

On présente l'implémentation de l'algorithme 2.

**Algorithme 2** (Implémentation).

– **Etape 1** : On a  $\bar{F} = h^2(F - G_{x,1} - G_{y,1}) \in \mathbb{R}^{(N-1)^2}$ . On calcule  $F = \text{vec}_2(\Pi^0 f)$  comme indiqué en (3.116).  $G_{x,1}$  et  $G_{y,1}$  sont calculés comme indiqué en (3.117) et (3.118). On obtient  $\bar{F}$ . Le calcul des  $\bar{F}_{k,l}^Z$  coefficients de  $\bar{F}$  dans la base  $Z^k \otimes Z^l$  s'obtient de la façon suivante. Le vecteur  $\bar{F} \in \mathbb{R}^{(N-1)^2}$  s'écrit

$$\bar{F} = \sum_{i,j=1}^{N-1} \bar{F}_{i,j} (e_i \otimes e_j). \quad (3.125)$$

Or

$$\begin{aligned} \bar{F}_{k,l}^Z &= (\bar{F}, Z^k \otimes Z^l) \\ &= \sum_{i,j=1}^{N-1} \bar{F}_{i,j} (e_i \otimes e_j, Z^k \otimes Z^l), \end{aligned} \quad (3.126)$$

et

$$\begin{aligned} (e_i \otimes e_j, Z^k \otimes Z^l) &= (Z^k \otimes Z^l)^T (e_i \otimes e_j) \\ &= ((Z^k)^T \otimes (Z^l)^T) (e_i \otimes e_j) \\ &= (Z^k)^T e_i \otimes (Z^l)^T e_j \\ &= Z_i^k \otimes Z_j^l \\ &= Z_i^k Z_j^l \\ &= \frac{2}{N} \sin\left(\frac{ik\pi}{N}\right) \sin\left(\frac{jl\pi}{N}\right), \end{aligned} \quad (3.127)$$

donc les composantes  $\bar{F}_{k,l}^Z$  sont

$$\bar{F}_{k,l}^Z = \sum_{i,j=1}^{N-1} \frac{2}{N} \bar{F}_{i,j} \sin\left(\frac{ik\pi}{N}\right) \sin\left(\frac{jl\pi}{N}\right). \quad (3.128)$$

- **Etape 2** : Le calcul des coefficients  $U_{k,l}^Z$  du vecteur  $U$  dans la base  $Z^k \otimes Z^l$  s'obtient par

$$H_1 U = \bar{F}^Z. \quad (3.129)$$

(3.129) est équivalent à

$$H_1 \sum_{k,l=1}^{N-1} U_{k,l}^Z (Z^k \otimes Z^l) = \sum_{k,l=1}^{N-1} \bar{F}_{k,l}^Z (Z^k \otimes Z^l), \quad (3.130)$$

ce qui donne

$$\sum_{k,l=1}^{N-1} U_{k,l}^Z \beta_{k,l} (Z^k \otimes Z^l) = \sum_{k,l=1}^{N-1} \bar{F}_{k,l}^Z (Z^k \otimes Z^l), \quad (3.131)$$

avec  $\beta_{k,l}$  les valeurs de  $H_1 \cdot Z^k \otimes Z^l$  étant une base orthonormale on obtient

$$U_{k,l}^Z = \frac{\bar{F}_{k,l}^Z}{\beta_{k,l}}, \quad (3.132)$$

d'où (3.113). Numériquement, les coefficients  $U_{k,l}^Z$  sont calculés par  $(N-1)^2$  divisions.

- **Etape 3** : A l'étape 2, on a calculé les coefficients  $U_{k,l}^Z$  de vecteur solution  $U$  dans la base  $Z^k \otimes Z^l$ . Dans cette étape, on calcule les coefficients  $U_{i,j}$  de  $U$  dans la base canonique  $e_i \otimes e_j$ . Ceci est effectué par la *transformée inverse en sinus rapide* (cf annexe).

### 3.8.3 Résultats numériques pour le schéma HB1

On montre quelques résultats numériques obtenus par le schéma HB1 en utilisant l'algorithme 2. Le code de calcul est effectué en Matlab et Fortran 90. Le problème résolu est (3.37) sur  $\Omega = (0,1)^2$ . Pour mesurer les erreurs, on utilise la norme  $L^2$  discrète définie par

$$\|u_{ex} - u\|_h = \left( h^2 \sum_{i,j} (u_{ex}(x_i, y_j) - u_{i,j})^2 \right)^{\frac{1}{2}}. \quad (3.133)$$

Les vecteurs dérivés  $U_x$  et  $U_y$  sont calculés par (3.119) en remplaçant les matrices  $\mathcal{D}$  et  $\mathcal{E}$  par les matrices  $\mathcal{D}_1$  et  $\mathcal{E}_1$ . Les matrices  $\mathcal{D}_1 = \mathcal{D}_1(\mathcal{A}_1, \mathcal{B}_1)$  et  $\mathcal{E}_1 = \mathcal{E}_1(\mathcal{B}_1, \mathcal{C}_1)$  sont obtenues à partir de (2.113) en remplaçant les matrices  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$  par les matrices

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$T = T_{\text{Matlab}}$	$T_{\text{F90}}$	$T/(N^2 \log_2(N))$
128 × 128	5.873(-6)	1.617(-5)	0.29 s.	0.01 s.	2.528(-6)
Taux de conv.	1.99	1.87			
256 × 256	1.468(-6)	4.420(-6)	0.57 s.	0.09	1.087 s.(-6)
Taux de conv.	2.00	1.89			
512 × 512	3.670(-7)	1.191(-6)	3.19 s.	0.53 s.	1.352(-6)
Taux de conv.	2.00	1.90			
1024 × 1024	9.177(-8)	3.179(-7)	20.46 s.	2.43 s.	1.951(-6)

TAB. 3.1: Erreur et taux de convergence pour le schéma HB1 avec  $u_{ex}(x, y) = \exp(-((x - 0.5)^2 + (y - 0.5)^2))$

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$T = T_{\text{Matlab}}$	$T_{\text{F90}}$	$T/(N^2 \log_2(N))$
128 × 128	5.674(-6)	1.318(-4)	0.33 s.	0.03 s.	2.877(-6)
Taux de conv.	4.02	4.03			
256 × 256	3.474(-7)	8.038(-6)	0.64 s.	0.17 s.	1.220(-6)
Taux de conv.	4.00	4.00			
512 × 512	2.160(-8)	4.993(-7)	3.21 s.	0.65 s.	1.360(-6)
Taux de conv.	4.00	4.00			
1024 × 1024	1.348(-9)	3.116(-8)	20.62 s.	2.34 s.	1.966(-6)

TAB. 3.2: Erreur et taux de convergence du schéma HB1 avec  $u_{ex}(x, y) = \exp(-400((x - 0.5)^2 + (y - 0.5)^2))$

$\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1$  en (2.43) respectivement.

**Test 1 :** Ce cas correspond à la gaussienne

$$u_{ex}(x, y) = \exp(-\alpha((x - 0.5)^2 + (y - 0.5)^2)). \quad (3.134)$$

**Test 1.1 :**  $\alpha = 1$ . Ce cas correspond à la gaussienne avec des gradients modérés. Les résultats numériques sont donnés dans la table 3.1. Observons la précision d'ordre 2 du schéma HB1 pour les trois inconnues. La taille du problème pour  $N = 1024$  est d'environ trois millions d'inconnues.

**Test 1.2 :**  $\alpha = 400$ . Ce cas correspond à une gaussienne avec fort gradient. Les résultats sont donnés dans la table 3.2 jusqu'à un maillage de taille  $N = 1024$  dans les directions de  $x$  et  $y$ . On voit que le schéma HB1 est d'ordre 4 pour ce tel cas, car cette fonction est nulle au voisinage des points de bord  $x_0$  et  $x_N$ . Dans ce cas le schéma HB1 est beaucoup plus précis que le schéma différence finie standard cinq points (DF2)(figure 3.6).

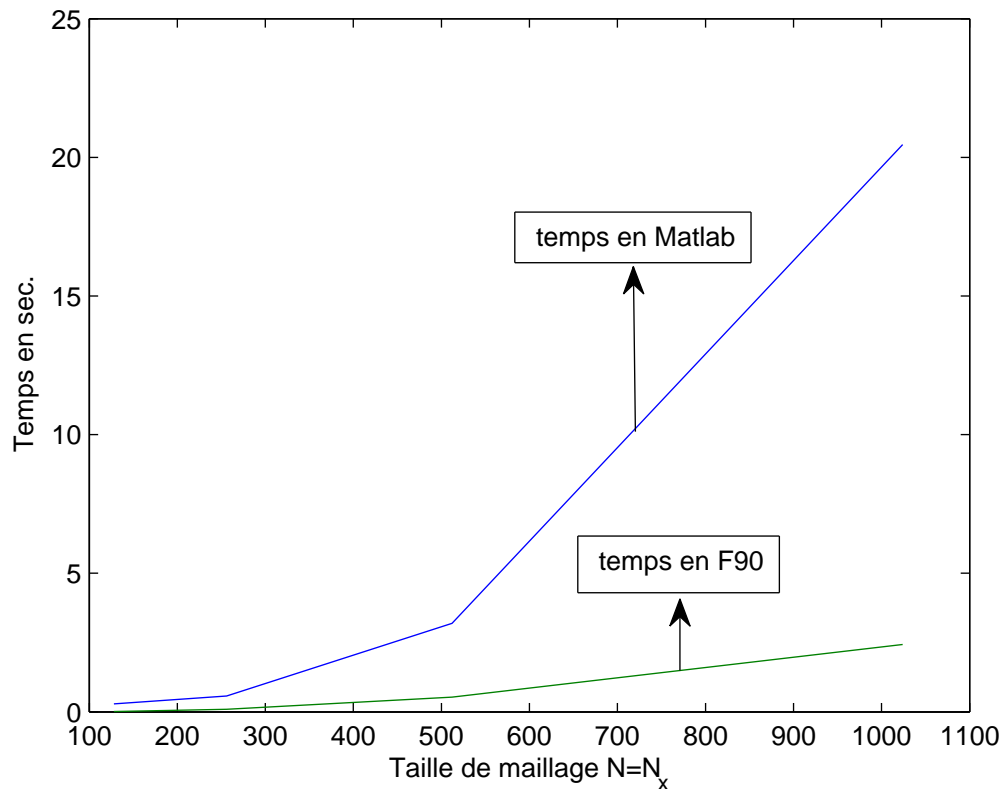


FIG. 3.3: Temps calcul en Matlab et F90 pour le test 1.1

**Test 2 :** Dans ce cas, la solution exacte est  $u_{ex}(x, y) = \ln(x + y^2 + 1)$ . Les résultats sont donnés dans la table 3.3. Le schéma HB1 est d'ordre 2 pour ce cas.

**Test 3 :** On considère ici le cas le plus difficile avec isolignes sur la première diagonale. La solution exacte est  $u_{ex}(x, y) = \cos(5\pi(x - y)^3)$ . Dans la table 3.4 le taux de convergence du schéma HB1 est deux pour les trois inconnues. Une bonne précision est atteinte sur la grille finale.

**Test 4 :** On présente dans la Table 3.5 des résultats numériques d'un cas présenté dans [8] avec la solution exacte

$$u_{ex}(x, y) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2)) \cos(20(x + y - 1)). \quad (3.135)$$

Ce cas est difficile à calculer. Une bonne précision est atteinte sur la grille finale, moins précise que celle obtenue en [8]. Noter que le schéma HB1 calcule les dérivées ce qui

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{y,ex} - u_y\ _h$	$T = T_{\text{Matlab}}$	$T_{\text{F90}}$	$T/(N^2 \log_2(N))$
$128 \times 128$	4.820(-6)	2.596(-5)	3.058(-5)	0.29 s.	0.07 s.	2.528(-6)
T. C.	1.99	1.85	1.88			
$256 \times 256$	1.205(-6)	7.224(-6)	8.296(-6)	0.58 s.	0.10 s.	1.106(-6)
T. C.	1.99	1.87	1.89			
$512 \times 512$	3.014(-7)	1.975(-6)	2.225(-6)	3.37 s.	0.65 s.	1.428(-6)
T. C.	1.99	1.89	1.91			
$1024 \times 1024$	7.536(-8)	5.331(-7)	5.917(-7)	21.91 s.	2.35 s.	2.089(-6)

TAB. 3.3: Erreur et taux de convergence du schéma HB1 avec  $u_{ex}(x, y) = \ln(x + y^2 + 1)$

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$T = T_{\text{Matlab}}$	$T_{\text{F90}}$	$T/(N^2 \log_2(N))$
$128 \times 128$	1.200(-3)	4.310(-2)	0.29 s.	0.03 s.	2.528(-6)
Taux de conv.	2.00	1.85			
$256 \times 256$	2.987(-4)	1.190(-2)	0.80 s.	0.10 s.	1.525(-6)
Taux de conv.	1.99	1.89			
$512 \times 512$	7.489(-5)	3.200(-3)	3.75 s.	0.625 s.	1.589(-6)
Taux de conv.	1.99	1.91			
$1024 \times 1024$	1.873(-5)	8.506(-4)	23.58 s.	2.35 s.	2.248(-6)

TAB. 3.4: Erreur et taux de convergence du schéma HB1 avec  $u_{ex}(x, y) = \cos(5\pi(x - y)^3)$

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$T = T_{\text{Matlab}}$	$T_{\text{F90}}$	$T/(N^2 \log_2(N))$
$128 \times 128$	2.572(-6)	3.619(-5)	0.29 s.	0.03 s.	2.528(-6)
Taux de conv.	3.34	3.07			
$256 \times 256$	2.534(-7)	4.285(-6)	0.67 s.	0.14 s.	1.277(-6)
Taux de conv.	2.29	2.16			
$512 \times 512$	5.163(-8)	9.592(-7)	3.59 s.	0.67 s.	1.521(-6)
Taux de conv.	2.02	2.00			
$1024 \times 1024$	1.270(-8)	2.390(-7)	22.45 s.	2.56 s.	2.141(-6)

TAB. 3.5: Erreur et taux de convergence du schéma HB1 avec  $u_{ex}(x, y) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2)) \cos(20(x + y - 1))$

n'est pas le cas dans [8].

### 3.9 Le schéma boîte hermitien HB2 en dimension 2

#### 3.9.1 Forme matricielle du schéma HB2

Il résulte de la proposition 3.5 que le schéma HB2 (3.56) possède la forme matricielle suivante

$$\frac{1}{h^2} (\mathcal{H}_2 \otimes P_s + P_s \otimes \mathcal{H}_2) = F - G_{x,2} - G_{y,2}, \quad (3.136)$$

où

- Le vecteur  $F$  correspond au terme source moyenné en (3.40),

$$F = \text{vec}_2(\Pi^0 f). \quad (3.137)$$

- Les vecteurs  $G_{x,2}, G_{y,2} \in \mathbb{R}^{(N-1)^2}$  correspondent aux conditions de bord non homogènes,

$$\left\{ \begin{array}{l} G_{x,2} = \frac{1}{h^2} (\mathcal{G}_2 \otimes P_s) (e_1 \otimes U_L + e_{N-1} \otimes U_R) \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}_2 U_B + \mathcal{G}_2 (e_1 U_{LB} + e_{N-1} U_{RB}) \right) \otimes e_1 \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}_2 U_T + \mathcal{G}_2 (e_1 U_{LT} + e_{N-1} U_{RT}) \right) \otimes e_{N-1}, \end{array} \right. \quad (3.138)$$

et

$$\left\{ \begin{array}{l} G_{y,2} = \frac{1}{h^2} (P_s \otimes \mathcal{G}_2) (U_B \otimes e_1 + U_T \otimes e_{N-1}) \\ \quad + \frac{1}{6h^2} e_1 \otimes \left( \mathcal{H}_2 U_L + \mathcal{G}_2 (e_1 U_{LB} + e_{N-1} U_{LT}) \right) \\ \quad + \frac{1}{6h^2} e_{N-1} \otimes \left( \mathcal{H}_2 U_R + \mathcal{G}_2 (e_1 U_{RB} + e_{N-1} U_{RT}) \right), \end{array} \right. \quad (3.139)$$

où les matrices  $\mathcal{A}_2, \mathcal{B}_2, \mathcal{C}_2$  sont les matrices données en (2.122) et  $\mathcal{H}_2 = \mathcal{H}_2(\mathcal{A}_2, \mathcal{B}_2)$ ,  $\mathcal{G}_2 = \mathcal{G}_2(\mathcal{B}_2, \mathcal{C}_2)$  sont les matrices données en (2.127). On remarque que la structure matricielle du schéma HB2 (3.136) conserve exactement la même forme que celle du schéma HB1 (3.102).

En outre, en utilisant la structure de la matrice  $\mathcal{H}_2$  (2.173), on obtient

$$\underbrace{\mathcal{H}_2 \otimes P_s + P_s \otimes \mathcal{H}_2}_{H_2} = \underbrace{(\mathcal{H}_1 \otimes P_s + P_s \otimes \mathcal{H}_1)}_{H_1} + \underbrace{(RS^T \otimes P_s + P_s \otimes RS^T)}_{\delta H}. \quad (3.140)$$

Le système linéaire à résoudre est donc

$$H_2 U = h^2 (F - G_{x,2} - G_{y,2}). \quad (3.141)$$

Donnons maintenant la structure détaillée de la matrice  $\delta H$ .

La matrice  $P_s$  s'exprime sous la forme

$$P_s = Z \operatorname{diag}(\alpha_1, \dots, \alpha_{N_1}) Z^T. \quad (3.142)$$

On définit le vecteur  $Z'^k \in \mathbb{R}^{(N-1)}$  par  $Z'^k = \alpha_k^{1/2} Z^k$ . La matrice  $Z' \in \mathbb{M}_{N-1}(\mathbb{R})$  est définie par

$$Z' = [Z'^1, \dots, Z'^{N-1}]. \quad (3.143)$$

(3.142) se réécrit

$$P_s = Z' Z'^T. \quad (3.144)$$

Donc la matrice  $\delta H$  en (3.140) peut s'écrire sous la forme

$$\begin{aligned} \delta H &= RS^T \otimes P_s + P_s \otimes RS^T \\ &= \begin{bmatrix} R_1 \otimes Z', \dots, R_6 \otimes Z', Z' \otimes R_1, \dots, Z' \otimes R_6 \end{bmatrix} \begin{bmatrix} S_1^T \otimes Z' \\ \vdots \\ S_6^T \otimes Z' \\ Z'^T \otimes S_1 \\ \vdots \\ Z'^T \otimes S_6 \end{bmatrix} \\ &= LM^T. \end{aligned}$$

Les matrices rectangulaires  $L, M \in \mathbb{M}_{(N-1)^2, 12(N-1)}(\mathbb{R})$  sont

$$L = [L_1, \dots, L_{12}], \quad M = [M_1, \dots, M_{12}], \quad (3.145)$$

où les matrices  $L_i, M_i \in \mathbb{M}_{(N-1)^2, (N-1)}(\mathbb{R})$  sont, pour  $1 \leq i \leq 6$ ,

$$\begin{cases} L_i = R_i \otimes Z', & L_{i+6} = Z' \otimes R_i, \\ M_i = S_i \otimes Z', & M_{i+6} = Z' \otimes S_i. \end{cases} \quad (3.146)$$

La base de notre algorithme rapide est la formule de Sherman-Morrison-Woodbury (2.155). En appliquant cette formule à (3.140) on obtient l'inverse de la matrice  $H_2 = H_1 + LM^T$ ,

$$H_2^{-1} = H_1^{-1} - H_1^{-1} L \left( I_{12(N-1)} + M^T H_1^{-1} L \right)^{-1} M^T H_1^{-1}. \quad (3.147)$$



(3.147) est la base de l'algorithme rapide détaillé dans la section suivante.

### 3.9.2 Algorithme rapide pour le schéma HB2

Cette section est consacrée à l'un des résultats les plus importants de ce travail. Il consiste en une procédure rapide de résolution du système linéaire (3.141) en utilisant (3.147). L'efficacité de cette procédure s'appuie sur deux éléments.

- La première est la formule de Sherman-Morrison-Woodbury (2.155). Elle permet de ramener la résolution d'un système de la forme  $H_2U = b$  à celles de systèmes linéaires de la forme  $H_1U = b$ .
- La transformée en sinus rapide. Ceci nous permet de résoudre à un coût de calcul de  $O(N^2 \log_2(N))$  les systèmes linéaires qui interviennent. L'algorithme correspondant à (3.147) est le suivant.

**Algorithme 3. – Etape 1 :** Calcul de  $\bar{F} = h^2(F - G_{x,2} - G_{y,2}) \in \mathbb{R}^{(N-1)^2}$ . Résolution de système linéaire

$$H_1 g = \bar{F}, \quad (3.148)$$

par l'algorithme 2. Le vecteur  $g \in \mathbb{R}^{(N-1)^2}$  est stocké pour l'étape 7.

– **Etape 2 :** Calcul de vecteur  $M^T g \in \mathbb{R}^{12(N-1)}$ ,

$$M^T g = \left[ M_1^T g, M_2^T g, \dots, M_{12}^T g \right]^T. \quad (3.149)$$

Par exemple, la première composante  $M_1^T g \in \mathbb{R}^{(N-1)}$  est

$$M_1^T g = \left[ (S_1 \otimes Z'^{,1})^T g, \dots, (S_1 \otimes Z'^{,N-1})^T g \right]^T. \quad (3.150)$$

Pour  $1 \leq l \leq N-1$ , la composante  $n^{\circ l}$  du vecteur  $M_1^T g$  en (3.150) est

$$\left\{ \begin{aligned} (S_1 \otimes Z'^{,l})^T g &= \sum_{i=1}^{N-1} (S_1)_i \sum_{j=1}^{N-1} g_{i,j} Z_j'^{,l} \\ &= \alpha_l^{1/2} \sum_{i=1}^{N-1} (S_1)_i \sum_{j=1}^{N-1} g_{i,j} Z_j^l. \end{aligned} \right. \quad (3.151)$$

Les termes

$$\sum_{j=1}^{N-1} g_{i,j} Z_j^l, \quad 1 \leq i \leq N-1, \quad (3.152)$$

sont calculés par la transformée en sinus rapide en utilisant

$$\sum_{j=1}^{N-1} g_{i,j} Z_j^l = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{j=1}^{N-1} g_{i,j} \sin\left(\frac{jl\pi}{N}\right). \quad (3.153)$$

Cette méthode est utilisée dans (3.149) pour calculer  $M_2^T g, \dots, M_6^T g$ . Considérons ensuite les composantes de  $M_7^T g$ ,

$$\begin{cases} (Z^{l,k} \otimes S_1)^T g &= \sum_{j=1}^{N-1} (S_1)_j \sum_{i=1}^{N-1} g_{i,j} Z_i^{l,k} \\ &= \alpha_k^{1/2} \sum_{j=1}^{N-1} (S_1)_j \sum_{i=1}^{N-1} g_{i,j} Z_i^k. \end{cases} \quad (3.154)$$

A nouveau, la transformée en sinus rapide est utilisée pour calculer

$$\sum_{i=1}^{N-1} g_{i,j} Z_i^k = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{i=1}^{N-1} g_{i,j} \sin\left(\frac{ik\pi}{N}\right). \quad (3.155)$$

Les vecteurs  $M_8^T g, \dots, M_{12}^T g$  sont calculés de la même manière. Les  $N-1$  transformées en sinus rapide en (3.153) et (3.155) sont effectuées à un coût de  $O(2N^2 \log_2(N))$ . Pour chaque  $1 \leq l \leq N-1$ , le produit scalaire en (3.151) ou (3.154) ajoute un coût additionnel de  $2N$ . Le coût des produits scalaires est donc  $24N^2$ . En total, le coût de l'étape 2 est en  $O(12N^2 \log_2(N)) + 24N^2$ .

– **Étape 3 :** Résoudre le système linéaire auxiliaire dans  $\mathbb{R}^{12(N-1)}$ .

$$(I_{12(N-1)} + M^T H_1^{-1} L)w = M^T g. \quad (3.156)$$

La complexité algorithmique de l'algorithme repose sur l'efficacité de cette résolution.

On utilise l'algorithme GMRES sans préconditionnement. Des détails sur l'implémentation et le coût de calcul de cette étape sont donnés dans la section 3.10.2.

– **Étape 4 :** La solution  $w \in \mathbb{R}^{12(N-1)}$  de l'étape 3 est décomposée en

$$w = [w_1, w_2, \dots, w_{12}]^T, w_1, w_2, \dots, w_{12} \in \mathbb{R}^{N-1}. \quad (3.157)$$

On note  $t \in \mathbb{R}^{12(N-1)}$  le produit  $t = Lw$ .

$$t = t_1 + t_2 + t_3 + \dots + t_{12}, \quad (3.158)$$

avec

$$\left\{ \begin{array}{l} t_1 = (R_1 \otimes [Z'^1, \dots, Z'^{N-1}])w_1, \quad t_2 = (R_2 \otimes [Z'^1, \dots, Z'^{N-1}])w_2, \\ \dots \\ t_5 = (R_5 \otimes [Z'^1, \dots, Z'^{N-1}])w_5, \quad t_6 = (R_6 \otimes [Z'^1, \dots, Z'^{N-1}])w_6, \\ t_7 = ([Z'^1, \dots, Z'^{N-1}] \otimes R_1)w_7, \quad t_8 = ([Z'^1, \dots, Z'^{N-1}] \otimes R_2)w_8, \\ \dots \\ t_{11} = ([Z'^1, \dots, Z'^{N-1}] \otimes R_5)w_{11}, \quad t_{12} = ([Z'^1, \dots, Z'^{N-1}] \otimes R_6)w_{12}. \end{array} \right. \quad (3.159)$$

Par exemple, on a pour  $1 \leq i, j \leq N-1$ ,

$$\left\{ \begin{array}{l} (t_1)_{i,j} = (R_1)_i \sum_{l=1}^{N-1} (w_1)_l Z_j'^l, \quad (t_2)_{i,j} = (R_2)_i \sum_{l=1}^{N-1} (w_2)_l Z_j'^l, \\ (t_{11})_{i,j} = (R_5)_j \sum_{k=1}^{N-1} (w_{11})_k Z_i'^k, \quad (t_{12})_{i,j} = (R_6)_j \sum_{k=1}^{N-1} (w_{12})_k Z_i'^k. \end{array} \right. \quad (3.160)$$

Chaque somme dans le membre de droite est calculée par sinus rapide, ce qui donne un coût de calcul de  $O(12N^2 \log_2(N))$ . Chacun des 12 vecteurs  $t_1, \dots, t_{12}$  nécessite également  $N^2$  multiplications. La somme en (5.149) demande  $12N^2$  additions. Cela donne donc un coût total de  $O(12N^2 \log_2(N)) + 24N^2$ .

– **Étape 5** : Résolution du système linéaire dans  $\mathbb{R}^{(N-1)^2}$

$$v = H_1^{-1}t, \quad (3.161)$$

par l'algorithme 2. Le coût de calcul est  $O(2N^2 \log_2(N)) + O(N^2)$ .

– **Étape 6** : Assemblage de la solution  $U \in \mathbb{R}^{(N-1)^2}$  du système linéaire (3.136) par

$$U = g - v, \quad (3.162)$$

où  $g$  et  $v \in \mathbb{R}^{(N-1)^2}$  sont donnés en (3.148) et (3.161) respectivement. Le coût de calcul est  $O(N^2)$ .

– **Étape 7** : Calcul si nécessaire des gradients hermitiens  $U_x$  et  $U_y$  en remplaçant dans (3.119) les matrices  $\mathcal{D}$  et  $\mathcal{E}$  par les matrices  $\mathcal{D}_2$  et  $\mathcal{E}_2$ . Les matrices  $\mathcal{D}_2 = \mathcal{D}_2(\mathcal{A}_2, \mathcal{B}_2)$  et  $\mathcal{E}_2 = \mathcal{E}_2(\mathcal{B}_2, \mathcal{C}_2)$  sont obtenues à partir de (2.113) en remplaçant les matrices  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$  par les matrices  $\mathcal{A}_2$ ,  $\mathcal{B}_2$ ,  $\mathcal{C}_2$  (2.122) respectivement. Ce calcul est décrit en détail dans la section 3.8.2.

Noter qu'il n'est à priori pas possible d'utiliser une méthode itérative classique utilisant la matrice d'itération  $H_1^{-1}RS^T$ . En effet, on vérifie numériquement que le rayon spectral de  $H_1^{-1}RS^T$  est 1.48. Ceci justifie l'intérêt de l'algorithme ci-dessus.

## 3.10 Mise en oeuvre du schéma HB2

### 3.10.1 Introduction

Dans cette section, on présente des détails pratiques de l'implémentation de l'étape 3 de l'algorithme 3 pour le schéma HB2. Les résultats principaux sont les suivants :

- Mise en évidence de l'efficacité algorithmique de l'algorithme 3 sur un PC de bureau. On peut traiter une taille  $4096 \times 4096$  (16 millions d'inconnues, 48 millions avec les gradients), [1].
- Mise en évidence du conditionnement modéré de la matrice auxiliaire.

### 3.10.2 Détail de l'implémentation de l'étape 3 (résolution du système auxiliaire)

On détaille l'implémentation de l'étape 3 de l'algorithme 3 qu'on utilise pour résoudre le système linéaire du schéma HB2 (3.136).

- Résolution de (3.156)

La matrice de ce système linéaire est

$$I_{12(N-1)} + M^T H_1^{-1} L, \quad (3.163)$$

avec

$$M^T H_1^{-1} L = \begin{bmatrix} M_1^T H_1^{-1} L_1 & M_1^T H_1^{-1} L_2 & \dots & M_1^T H_1^{-1} L_{12} \\ M_2^T H_1^{-1} L_1 & M_2^T H_1^{-1} L_2 & \dots & M_2^T H_1^{-1} L_{12} \\ \dots & \dots & \dots & \dots \\ M_{12}^T H_1^{-1} L_1 & M_{12}^T H_1^{-1} L_2 & \dots & M_{12}^T H_1^{-1} L_{12} \end{bmatrix}. \quad (3.164)$$

Les matrices  $M_k$  et  $L_k \in \mathbb{M}_{(N-1)^2, N-1}(\mathbb{R})$ ,  $1 \leq k \leq 12$ , sont données en (3.146).

Les matrices  $M_i^T H_1^{-1} L_j$  s'écrivent

$$\begin{cases} M_i^T H_1^{-1} L_j = (S_i \otimes Z')^T H_1^{-1} (R_j \otimes Z'), & 1 \leq i, j \leq 6, & (a) \\ M_i^T H_1^{-1} L_j = (S_i \otimes Z')^T H_1^{-1} (Z' \otimes R_j), & 1 \leq i \leq 6, \quad 7 \leq j \leq 12, & (b) \\ M_i^T H_1^{-1} L_j = (Z' \otimes S_i)^T H_1^{-1} (R_j \otimes Z'), & 7 \leq i \leq 12, \quad 1 \leq j \leq 6, & (c) \\ M_i^T H_1^{-1} L_j = (Z' \otimes S_i)^T H_1^{-1} (Z' \otimes R_j), & 7 \leq i, j \leq 12. & (d) \end{cases} \quad (3.165)$$

En écrivant les vecteurs  $S_i, R_j \in \mathbb{R}^{(N-1)}$  dans la base orthonormale  $(Z^k)_{1 \leq k \leq N-1}$ , on obtient

$$\begin{cases} S_i = \sum_{k=1}^{N-1} s_{i,k} Z^k, \\ R_j = \sum_{l=1}^{N-1} r_{j,l} Z^l. \end{cases} \quad (3.166)$$

On a donc

$$\begin{aligned} R_j \otimes Z' &= R_j \otimes [Z'^1, Z'^2, \dots, Z'^{N-1}] \\ &= [R_j \otimes Z'^1, R_j \otimes Z'^2, \dots, R_j \otimes Z'^{N-1}], \end{aligned}$$

et

$$\begin{aligned} S_i \otimes Z' &= S_i \otimes [Z'^1, Z'^2, \dots, Z'^{N-1}] \\ &= [S_i \otimes Z'^1, S_i \otimes Z'^2, \dots, S_i \otimes Z'^{N-1}]. \end{aligned}$$

• Cas (a) dans (3.165). Pour  $1 \leq i, j \leq 6$ , on a

$$M_i^T H_1^{-1} L_j = (S_i \otimes Z')^T H_1^{-1} (R_j \otimes Z'). \quad (3.167)$$

Donc

$$(S_i \otimes Z')^T H_1^{-1} (R_j \otimes Z') = \begin{bmatrix} (S_i \otimes Z'^1)^T H_1^{-1} (R_j \otimes Z'^1) & \dots & \dots & (S_i \otimes Z'^1)^T H_1^{-1} (R_j \otimes Z'^{N-1}) \\ (S_i \otimes Z'^2)^T H_1^{-1} (R_j \otimes Z'^1) & \dots & \dots & (S_i \otimes Z'^2)^T H_1^{-1} (R_j \otimes Z'^{N-1}) \\ \dots & \dots & \dots & \dots \\ (S_i \otimes Z'^{N-1})^T H_1^{-1} (R_j \otimes Z'^1) & \dots & \dots & (S_i \otimes Z'^{N-1})^T H_1^{-1} (R_j \otimes Z'^{N-1}) \end{bmatrix}. \quad (3.168)$$

Or  $(S_i \otimes Z'^l)^T H_1^{-1}(R_j \otimes Z'^m) = 0$  si  $l \neq m$  car la base  $Z^k \otimes Z^l$  est orthonormale. Si  $l = m$ , on a

$$\left\{ \begin{aligned} (S_i \otimes Z'^m)^T H_1^{-1}(R_j \otimes Z'^m) &= \left( \sum_{k=1}^{N-1} s_{i,k} Z^k \otimes \alpha_m^{1/2} Z^m \right)^T H_1^{-1} \left( \sum_{l=1}^{N-1} r_{j,l} Z^l \otimes \alpha_m^{1/2} Z^m \right) \\ &= \left( \sum_{k=1}^{N-1} s_{i,k} \alpha_m^{1/2} Z^k \otimes Z^m \right)^T \left( \sum_{l=1}^{N-1} r_{j,l} \alpha_m^{1/2} \beta_{l,m}^{-1} Z^l \otimes Z^m \right) \\ &= \sum_{k=1}^{N-1} \alpha_m s_{i,k} r_{j,k} \beta_{k,m}^{-1}, \end{aligned} \right. \quad (3.169)$$

avec  $\beta_{k,m}$  les valeurs propres de la matrice  $H_1$  données en (3.110). L'équation (3.168) donne

$$(S_i \otimes Z')^T H_1^{-1}(R_j \otimes Z') = \begin{bmatrix} \sum_{k=1}^{N-1} s_{i,k} r_{j,k} \alpha_1 \beta_{k,1}^{-1} & 0 & \dots & 0 \\ 0 & \sum_{k=1}^{N-1} s_{i,k} r_{j,k} \alpha_2 \beta_{k,2}^{-1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sum_{k=1}^{N-1} s_{i,k} r_{j,k} \alpha_{N-1} \beta_{k,N-1}^{-1} \end{bmatrix}. \quad (3.170)$$

Le coût de calcul de (3.170) est  $O(N^2)$ .

• Cas (c) dans (3.165). Pour  $7 \leq i \leq 12$ ,  $1 \leq j \leq 6$ , on a

$$M_i^T H_1^{-1} L_j = (Z' \otimes S_i) H_1^{-1} (Z' \otimes R_j), \quad (3.171)$$

alors  $(Z'^l \otimes S_i)^T H_1^{-1}(Z'^m \otimes R_j) = 0$  si  $l \neq m$  car la base  $Z^k \otimes Z^l$  est orthonormale. Si  $l = m$

$$(Z'^m \otimes S_i)^T H_1^{-1}(Z'^m \otimes R_j) = \sum_{k=1}^{N-1} \alpha_m s_{i,k} \beta_{m,k}^{-1} r_{j,k}. \quad (3.172)$$

D'où l'égalité entre les blocs (3.165)<sub>a</sub> et (3.165)<sub>d</sub>.

• Cas (b) dans (3.165). Pour  $1 \leq i \leq 6$ ,  $7 \leq j \leq 12$ ,

$$M_i^T H_1^{-1} L_j = (S_i \otimes Z')^T H_1^{-1} (Z' \otimes R_j). \quad (3.173)$$

$N$	8	16	32	64	128	256	512
Cond.	11.3376	11.2784	11.2299	11.2015	11.1851	11.1757	11.1703

TAB. 3.6: Conditionnement de la matrice  $I_{12(N-1)} + M^T H_1^{-1} L$ 

On a

$$(S_i \otimes Z')^T H_1^{-1} (Z' \otimes R_j) = \begin{bmatrix} (S_i \otimes Z'^{,1})^T H_1^{-1} (Z'^{,1} \otimes R_j) & \dots & \dots & (S_i \otimes Z'^{,1})^T H_1^{-1} (Z'^{,N-1} \otimes R_j) \\ (S_i \otimes Z'^{,2})^T H_1^{-1} (Z'^{,1} \otimes R_j) & \dots & \dots & (S_i \otimes Z'^{,2})^T H_1^{-1} (Z'^{,N-1} \otimes R_j) \\ \dots & \dots & \dots & \dots \\ (S_i \otimes Z'^{,N-1})^T H_1^{-1} (Z'^{,1} \otimes R_j) & \dots & \dots & (S_i \otimes Z'^{,N-1})^T H_1^{-1} (Z'^{,N-1} \otimes R_j) \end{bmatrix}. \quad (3.174)$$

Or

$$(S_i \otimes Z'^{,l})^T H_1^{-1} (Z'^{,m} \otimes R_j) = \alpha_l^{1/2} s_{i,m} \beta_{m,l}^{-1} \alpha_m^{1/2} r_{j,l}, \quad (3.175)$$

alors

$$(S_i \otimes Z')^T H_1^{-1} (Z' \otimes R_j) = \begin{bmatrix} \alpha_1^{1/2} s_{i,1} \beta_{1,1}^{-1} \alpha_1^{1/2} r_{j,1} & \dots & \dots & \alpha_1^{1/2} s_{i,N-1} \beta_{N-1,1}^{-1} \alpha_{N-1}^{1/2} r_{j,1} \\ \alpha_2^{1/2} s_{i,1} \beta_{1,2}^{-1} \alpha_1^{1/2} r_{j,2} & \dots & \dots & \alpha_2^{1/2} s_{i,N-1} \beta_{N-1,2}^{-1} \alpha_{N-1}^{1/2} r_{j,2} \\ \dots & \dots & \dots & \dots \\ \alpha_{N-1}^{1/2} s_{i,1} \beta_{1,N-1}^{-1} \alpha_1^{1/2} r_{j,N-1} & \dots & \dots & \alpha_{N-1}^{1/2} s_{i,N-1} \beta_{N-1,N-1}^{-1} \alpha_{N-1}^{1/2} r_{j,N-1} \end{bmatrix}. \quad (3.176)$$

Le coût de calcul de (3.168) est  $O(N^2)$ .

- Cas (d) dans (3.165). Pour  $7 \leq i, j \leq 12$ , on a

$$M_i^T H_1^{-1} L_j = (Z' \otimes S_i) H_1^{-1} (R_j \otimes Z'), \quad (3.177)$$

alors

$$(Z'^{,l} \otimes S_i)^T H_1^{-1} (R_j \otimes Z'^{,m}) = \alpha_l^{1/2} s_{i,m} \beta_{l,m}^{-1} \alpha_m^{1/2} r_{j,l} \quad (3.178)$$

D'où l'égalité entre les blocs (3.165)<sub>b</sub> et (3.165)<sub>c</sub>.

Finalement, le coût de calcul de la matrice (3.163) est  $72O(N^2)$ . On présente dans la table 3.6 le conditionnement de la matrice (3.163). Il apparaît clairement indépendant de la taille de maillage  $N$ .

### 3.10.3 Résultats numériques pour le schéma HB2

On donne quelques résultats numériques obtenus par le schéma HB2 pour le problème de Poisson (3.37) sur  $\Omega = (0, 1)^2$  en utilisant l'algorithme 3. Deux codes de calcul ont été réalisés, le premier en `matlab`, le second en Fortran 90, sur un PC de processeur Intel Core2 Duo, 2.8GHz, 3.48 Go de RAM. Le temps calcul en Matlab  $T = T_{\text{Matlab}}$  est calculé à l'aide des fonctions `tic` et `toc`. Le quotient  $T/(N^2 \log_2 N)$  apparaît constant ce qui indique une complexité arithmétique  $O(N^2 \log_2 N)$ .

En Fortran 90, le code de calcul est séquentiel et utilise le compilateur `g95` avec niveau d'optimisation  $-O_3$ . On utilise la librairie `fftpack5.f90` pour la transformée en sinus rapide, [35]. Le temps calcul  $T_{\text{F90}}$  est calculé en utilisant la fonction `cpu_time`.

On utilise la norme  $L^2$  discrète pour mesurer l'erreur

$$\|u_{ex} - u\|_h = \left( h^2 \sum_{i,j} (u_{ex}(x_i, y_j) - u_{i,j})^2 \right)^{\frac{1}{2}}. \quad (3.179)$$

Rappelons que les conditions au bord non homogènes, qui sont données dans les vecteurs  $G_{x,2}$  et  $G_{y,2}$  (3.138, 3.139), donnent une contribution additionnelle dans le membre de droite. Les cas tests proposés sont les mêmes que pour le schéma HB1 (voir la section 3.8.3).

**Test 1 :** On considère la fonction gaussienne  $u_{ex}(x, y) = \exp(-\alpha((x-0.5)^2 + (y-0.5)^2))$  comme solution exacte où  $\alpha$  est une constante. Ce cas est introduit dans [8].

**Test 1.1 :**  $\alpha = 1$ . Ce cas correspond à une gaussienne avec des gradients modérés (figure 3.5). Les résultats sont présentés dans la table 3.7 jusqu'à une grille de taille  $N = 1024$  dans les directions  $x$  et  $y$ . Dans ce dernier cas, la taille du problème est d'environ un million d'inconnues. La précision est d'ordre 4 pour les 3 inconnues  $u$ ,  $u_x$ ,  $u_y$ . La figure 3.5 montre que le schéma HB2 donne une solution  $u$  plus précise que celles obtenues par le schéma HB1 et le schéma à 5 points (DF2).

**Test 1.2 :**  $\alpha = 400$ . Ce cas correspond à celui d'une gaussienne avec fort gradient (figure 3.6). Les résultats sont présentés dans la table (3.8) jusqu'à un maillage de taille  $N = 1024$  dans les directions de  $x$  et  $y$ . La figure 3.6 montre que les schémas HB1 et HB2 possèdent la même précision, alors que le schéma DF2 est moins précis dans ce cas.

**Test 2 :** On considère la solution exacte  $u_{ex}(x, y) = \ln(x + y^2 + 1)$ . Les résultats sont donnés dans la table 3.9. Le nombre d'itérations de GMRES est égal à 15. Notons que



$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$T = T_{\text{Matlab}}$	$T_{\text{F90}}$	$T/(N^2 \log_2(N))$	GMRES it.
$128 \times 128$	2.385(-10)	5.793(-10)	0.82 s.	0.17 s.	7.149(-6)	15
Taux de conv.	3.98	4.01				
$256 \times 256$	1.504(-11)	3.585(-11)	1.78 s.	0.60 s.	3.395(-6)	15
Taux de conv.	3.99	3.97				
$512 \times 512$	9,444(-13)	2.284(-12)	7.23 s.	2.78 s.	3.064(-6)	15
Taux de conv.	4.02	0.79				
$1024 \times 1024$	5.796(-14)	1.319(-12)	55.84 s.	11.62 s.	5.325(-6)	15

TAB. 3.7: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y) = \exp(-((x - 0.5)^2 + (y - 0.5)^2))$

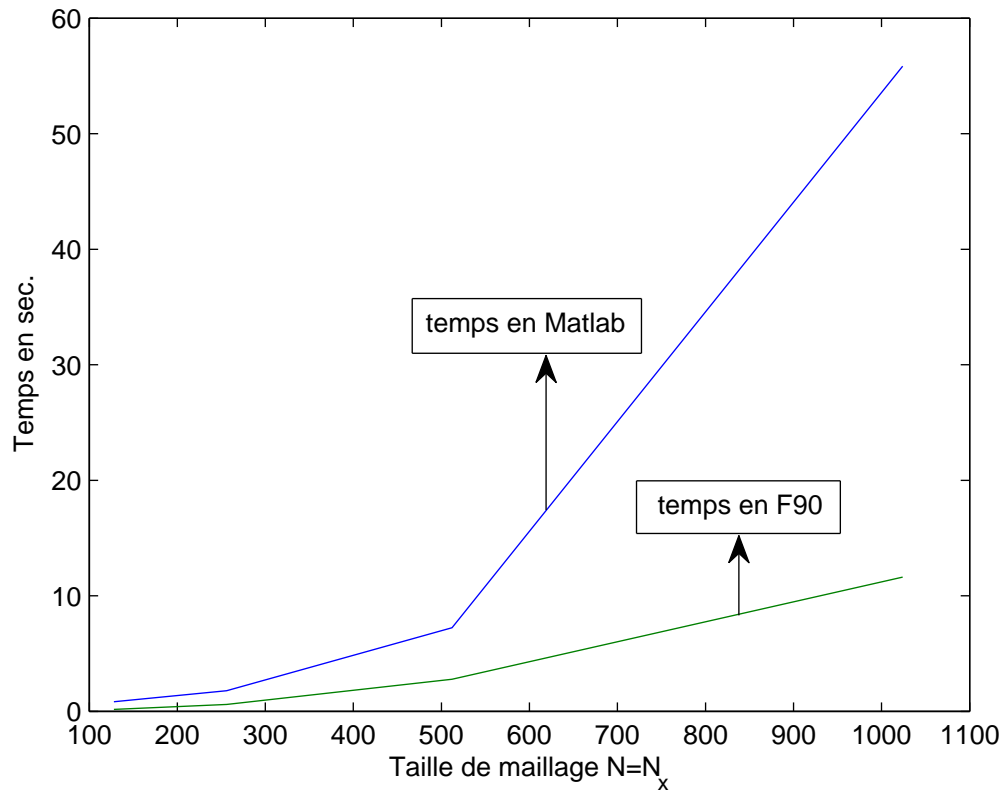


FIG. 3.4: Temps calcul en Matlab et F90 pour le test 1.1

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$T = T_{\text{Matlab}}$	$T_{\text{F90}}$	$T/(N^2 \log_2(N))$	GMRES it.
128 × 128	5.674(-6)	1.318(-4)	0.96 s.	0.12 s.	1.207(-5)	12
Taux de conv.	4.02	4.03				
256 × 256	3.474(-7)	8.038(-6)	2.79 s.	0.68 s.	7.677(-6)	14
Taux de conv.	4.00	4.00				
512 × 512	2.160(-8)	4.993(-7)	12 s.	3.14 s.	7.337(-6)	16
Taux de conv.	4.00	4.00				
1024 × 1024	1.348(-9)	3.116(-8)	64 s.	13.78 s.	8.805(-6)	17

TAB. 3.8: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y) = \exp(-400((x - 0.5)^2 + (y - 0.5)^2))$

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{y,ex} - u_y\ _h$	$T = T_{\text{Matlab}}$	$T_{\text{F90}}$	$T/(N^2 \log_2(N))$
128 × 128	3.478(-10)	5.509(-9)	9.421(-9)	1 s.	0.14 s.	1.257(-5)
T. C.	4.00	3.96	3.99			
256 × 256	2.167(-11)	3.526(-10)	5.917(-10)	2.94 s.	0.54 s.	8.090(-6)
T. C.	4.00	3.97	3.99			
512 × 512	1.351(-12)	2.250(-11)	3.719(-11)	12.5 s.	2.56 s.	7.643(-6)
T. C.	4.00	3.41	3.74			
1024 × 1024	8.420(-14)	2.109(-12)	2.782(-12)	71 s.	11.64 s.	9.768(-6)

TAB. 3.9: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y) = \ln(x + y^2 + 1)$

la précision de l'ordinateur  $10^{-12}$  est atteinte pour  $N = 1024$ , comme c'est déjà observé dans la décroissance du taux de convergence pour  $u_x, u_y$ .

**Test 3 :** On considère le cas où  $u_{ex}(x, y) = \cos(5\pi(x - y)^3)$ . Ce cas est plus difficile avec des isolignes le long de la première diagonale. Le taux de convergence est à peu près 4 pour les trois inconnues  $u, u_x, u_y$ . Les résultats sont donnés dans la table 3.10. Une très bonne précision est atteinte sur la grille finale avec  $N = 1024$ .

**Test 4 :** On présente dans la table (3.11) les résultats numériques d'un cas présenté dans [8] avec la solution exacte

$$u_{ex}(x, y) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2)) \cos(20(x + y - 1)). \quad (3.180)$$

La fonction (3.180) présente de très forts gradients et des oscillations (figure 3.7). Une très bonne précision est atteinte sur la grille finale, moins précise cependant que celle

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$T = T_{\text{Matlab}}$	$T_{F90}$	$T/(N^2 \log_2(N))$	GMRES it.
$128 \times 128$	2.985(-5)	1.169(-3)	1 s.	0.26 s.	1.257(-5)	16
T. C.	3.99	3.78	3.79			
$256 \times 256$	1.869(-6)	8.402(-5)	3.1 s.	0.70 s.	8.530(-6)	16
T. C.	4.00	3.84				
$512 \times 512$	1.163(-7)	5.834(-6)	13.23 s.	2.60 s.	8.090(-6)	15
T. C.	4.00	3.95	3.85			
$1024 \times 1024$	7.240(-9)	3.967(-7)	76.23 s.	11.53 s.	1.048(-5)	15

TAB. 3.10: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y) = \cos(5\pi(x - y)^3)$

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$T = T_{\text{Matlab}}$	$T_{F90}$	$T/(N^2 \log_2(N))$	GMRES it.
$128 \times 128$	2.440(-6)	3.308(-5)	0.99 s.	0.12 s.	1.245(-5)	13
Taux de conv.	4.00	4.00				
$256 \times 256$	1.518(-7)	2.056(-6)	2.8 s.	0.48 s.	7.704(-6)	12
Taux de conv.	4.00	4.00				
$512 \times 512$	9.480(-9)	1.283(-7)	11.69 s.	2.26 s.	7.148(-6)	12
Taux de conv.	4.00	4.00				
$1024 \times 1024$	5.923(-10)	8.021(-9)	62 s.	10.28 s.	8.530(-6)	12

TAB. 3.11: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2)) \cos(20(x + y - 1))$

obtenue en [8]. Noter toutefois que le schéma HB fournit une approximation de l'inconnue et du gradient, ce qui n'est pas le cas dans [8]. De plus, [8] utilise une méthode de soustraction de singularité, ce qui n'est pas le cas ici. <sup>1</sup>

<sup>1</sup>Le travail de ce chapitre et le chapitre précédent a été publié au *Journal of Scientific Computing* dans l'article [1]

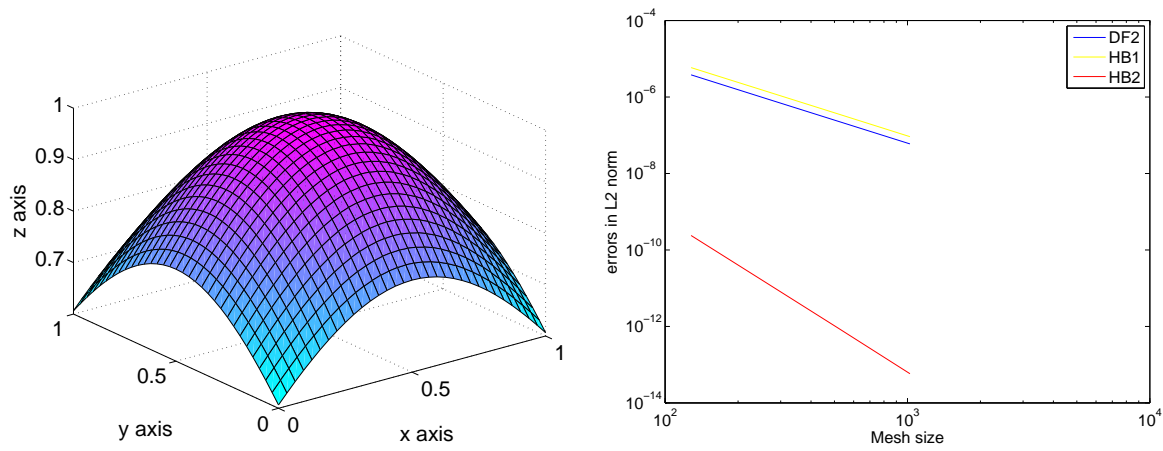


FIG. 3.5: A gauche, la solution exacte  $u_{ex}(x, y) = \exp(-(x - 0.5)^2 - (y - 0.5)^2)$ . A droite, des courbes de convergence.

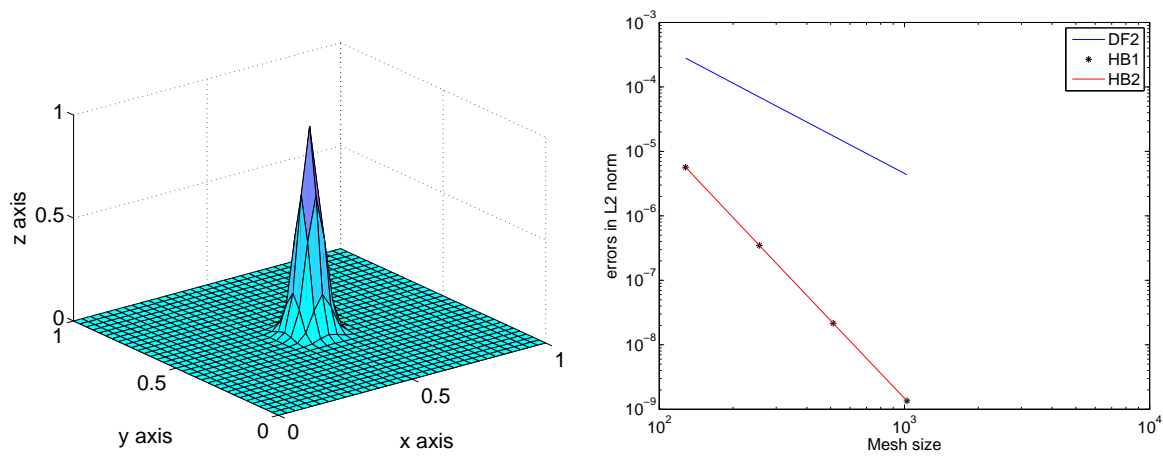


FIG. 3.6: A gauche, la solution exacte  $u_{ex}(x, y) = \exp(-400(x - 0.5)^2 - 400(y - 0.5)^2)$ . A droite, des courbes de convergence.

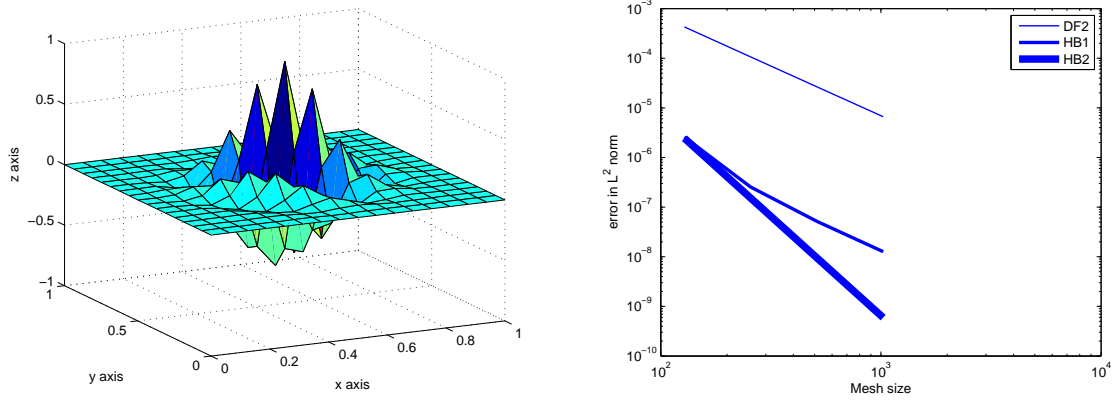


FIG. 3.7:  $u_{ex}(x, y) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2)) \cos(20(x + y - 1))$ . A droite, des courbes de convergence.



## Chapitre 4

# Algorithmes de type boîte hermitien

### 4.1 Introduction

Dans le chapitre 3 on a introduit une famille de schémas boîte hermitiens pour le problème de Poisson en géométrie cartésienne. Nous voulons à présent utiliser le schéma pour des problèmes plus généraux.

- Une première application concerne des modèles elliptiques plus complexes que le problème de Poisson. On veut résoudre des équations elliptiques non séparables du type

$$-\operatorname{div}(a(x, y)\nabla u) = f, \quad \text{sur } \Omega, \quad (4.1)$$

avec conditions au bord de Dirichlet. On introduit d’abord un schéma boîte hermitien pour de tels problèmes. Ensuite on résout le schéma obtenu en adaptant l’algorithme de Concus [11]. Le principe est d’utiliser une méthode de relaxation et d’adopter l’algorithme 3 comme solveur de base.

- Une seconde application est la définition d’un schéma boîte hermitien multiéchelle pour le problème de Poisson (3.37). La méthode utilisée est de type “local defect correction”. Elle consiste en la décomposition du domaine  $\Omega$  en deux grilles, l’une “grossière” et l’autre “fine”. La résolution sur la grille fine est supposée plus précise que celle sur la grille grossière. On applique cette méthode pour résoudre des problèmes de Poisson dans un carré où la solution possède une région de “forte activité”, c’est-à-dire à fort gradient.

## 4.2 Définition du schéma (HB) pour certains problèmes elliptiques non séparables

On considère le problème de Dirichlet pour l'équation elliptique à coefficients variables suivante

$$\begin{cases} \mathcal{L}u(x, y) = f(x, y), & (x, y) \in \Omega, & (a) \\ u(x, y) = g(x, y), & (x, y) \in \partial\Omega = \Gamma. & (b) \end{cases} \quad (4.2)$$

Dans (4.2),  $\Omega = (a, b)^2$  est un carré et  $\mathcal{L}$  désigne l'opérateur elliptique

$$\mathcal{L}u(x, y) = -\frac{\partial}{\partial x} \left( a(x, y) \frac{\partial u}{\partial x}(x, y) \right) - \frac{\partial}{\partial y} \left( a(x, y) \frac{\partial u}{\partial y}(x, y) \right). \quad (4.3)$$

La diffusion est isotrope donnée par la fonction scalaire  $a(x, y)$  qui vérifie

$$0 < a_{\min} \leq a(x, y) \leq a_{\max}. \quad (4.4)$$

On note également  $\mathcal{L}_0 = -\Delta$ ,

$$\mathcal{L}_0 u(x, y) = -\frac{\partial}{\partial x} \left( \frac{\partial u}{\partial x}(x, y) \right) - \frac{\partial}{\partial y} \left( \frac{\partial u}{\partial y}(x, y) \right). \quad (4.5)$$

Dans la suite on suppose que  $a(x, y)$ ,  $f(x, y)$  et  $g(x, y)$  sont telles que la solution  $u(x, y)$  est suffisamment régulière. La positivité de  $a(x, y)$  implique que l'opérateur  $\mathcal{L}$  est défini positif. La méthode de Concus [11] consiste à effectuer le changement de variables suivant

$$w(x, y) = [a(x, y)]^{1/2} u(x, y), \quad (4.6)$$

on suppose que  $a^{1/2}$  est deux fois différentiable. Dans ce cas, l'équation (4.2)<sub>a</sub> devient

$$a^{-1/2} \mathcal{L} \equiv -\Delta w + pw = q, \quad (4.7)$$

avec

$$\begin{cases} p = a^{-1/2} \Delta(a^{1/2}), \\ q = a^{-1/2} f. \end{cases} \quad (4.8)$$

Les conditions au bord pour  $w(x, y)$  sont

$$w(x, y) = a^{1/2}(x, y)g(x, y), \quad (x, y) \in \Gamma. \quad (4.9)$$



Le problème (4.2) est équivalent à

$$\begin{cases} \mathcal{M}w(x, y) = q(x, y), & (x, y) \in \Omega, \\ w(x, y) = a^{1/2}(x, y)g(x, y), & (x, y) \in \partial\Omega, \end{cases} \quad (4.10)$$

avec

$$\mathcal{M} = -\Delta + p. \quad (4.11)$$

Donc la relation (4.6) transforme l'opérateur  $\mathcal{L}$  en un autre opérateur dont la partie différentielle d'ordre 2 est  $-\Delta$ . On introduit les deux variables  $v^1 = \partial_x u$  et  $v^2 = \partial_y u$ . La forme mixte de (4.10) est

$$\begin{cases} -\partial_x v^1 - \partial_y v^2 - pw = q, & \text{dans } \Omega, & (a) \\ v^1 = \partial_x w, & \text{dans } \bar{\Omega}, & (b) \\ v^2 = \partial_y w, & \text{dans } \bar{\Omega}, & (c) \\ w = a^{1/2}g. & \text{sur } \partial\Omega. & (d) \end{cases} \quad (4.12)$$

En mettant  $pw$  à droite dans (4.12)<sub>a</sub> on obtient

$$\begin{cases} -\partial_x v^1 - \partial_y v^2 = pw + q, & \text{dans } \Omega, & (a) \\ v^1 = \partial_x w, & \text{dans } \bar{\Omega}, & (b) \\ v^2 = \partial_y w, & \text{dans } \bar{\Omega}, & (c) \\ w = a^{1/2}g. & \text{sur } \partial\Omega. & (d) \end{cases} \quad (4.13)$$

Noter que (4.13) est analogue à la forme mixte (3.38) du problème de Poisson (3.37). Soient  $w$ ,  $w_x$ ,  $w_y \in L_h^2$  les fonctions de grille approchant  $w(x, y)$ ,  $\partial_x w(x, y)$ ,  $\partial_y w(x, y)$  respectivement. En suivant la même procédure que pour le schéma HB1 en

(3.54), on déduit de (3.54) le système discret

$$\left\{ \begin{array}{ll}
 -(\sigma_y \circ \delta_x)w_{x,i,j} - (\sigma_x \circ \delta_y)w_{y,i,j} - \Pi^0(pw)_{i,j} = \Pi^0 q_{i,j}, & 1 \leq i, j \leq N-1, \quad (a) \\
 \sigma_x w_{x,i,j} = \delta_x w_{i,j}, & 1 \leq i, j \leq N-1, \quad 0 \leq j \leq N, \quad (b) \\
 \sigma_y w_{y,i,j} = \delta_y w_{i,j}, & 1 \leq i, j \leq N-1, \quad 0 \leq i \leq N, \quad (c) \\
 \frac{2}{3}w_{x,0,j} + \frac{1}{3}w_{x,1,j} = \frac{w_{1,j} - w_{0,j}}{h}, & 0 \leq j \leq N, \quad (d) \\
 \frac{2}{3}w_{x,N,j} + \frac{1}{3}w_{x,N-1,j} = \frac{w_{N,j} - w_{N-1,j}}{h}, & 0 \leq j \leq N, \quad (e) \\
 \frac{2}{3}w_{y,i,0} + \frac{1}{3}w_{y,i,1} = \frac{w_{i,1} - w_{i,0}}{h}, & 0 \leq i \leq N, \quad (f) \\
 \frac{2}{3}w_{y,i,N} + \frac{1}{3}w_{y,i,N-1} = \frac{w_{i,N} - w_{i,N-1}}{h}, & 0 \leq i \leq N, \quad (g) \\
 w_{i,0} = a^{1/2}(x_i, y_0)g(x_i, y_0), & 0 \leq i \leq N, \quad (h) \\
 w_{i,N} = a^{1/2}(x_i, y_N)g(x_i, y_N), & 0 \leq i \leq N, \quad (l) \\
 w_{0,j} = a^{1/2}(x_0, y_j)g(x_0, y_j), & 1 \leq j \leq N-1, \quad (m) \\
 w_{N,j} = a^{1/2}(x_N, y_j)g(x_N, y_j), & 1 \leq j \leq N-1. \quad (n)
 \end{array} \right. \quad (4.14)$$

Le système (4.14) est le schéma HB1 pour le problème (4.10). De même, en suivant la même procédure que pour le schéma HB2 en (3.56), on déduit de (3.56) le système discret

$$\left\{ \begin{array}{ll}
 -(\sigma_y \circ \delta_x)w_{x,i,j} - (\sigma_x \circ \delta_y)w_{y,i,j} - \Pi^0(pw)_{i,j} = \Pi^0 q_{i,j}, & 1 \leq i, j \leq N-1, \quad (a) \\
 \sigma_x w_{x,i,j} = \delta_x w_{i,j}, & 1 \leq i \leq N-1, \quad 0 \leq j \leq N, \quad (b) \\
 \sigma_y w_{y,i,j} = \delta_y w_{i,j}, & 1 \leq j \leq N-1, \quad 0 \leq i \leq N, \quad (c) \\
 \frac{1}{3}w_{x,0,j} + \frac{2}{3}w_{x,1,j} = \frac{1}{h} \left( \frac{1}{6}w_{2,j} + \frac{2}{3}w_{1,j} - \frac{5}{6}w_{0,j} \right), & 0 \leq j \leq N, \quad (d) \\
 \frac{1}{3}w_{x,N,j} + \frac{2}{3}w_{x,N-1,j} = \frac{1}{h} \left( \frac{5}{6}w_{N,j} - \frac{1}{6}w_{N-2,j} - \frac{2}{3}w_{N-1,j} \right), & 0 \leq j \leq N, \quad (e) \\
 \frac{1}{3}w_{y,i,0} + \frac{2}{3}w_{y,i,1} = \frac{1}{h} \left( \frac{1}{6}w_{i,2} + \frac{2}{3}w_{i,1} - \frac{5}{6}w_{i,0} \right), & 0 \leq i \leq N, \quad (f) \\
 \frac{1}{3}w_{y,i,N} + \frac{2}{3}w_{y,i,N-1} = \frac{1}{h} \left( \frac{5}{6}w_{i,N} - \frac{1}{6}w_{i,N-2} - \frac{2}{3}w_{i,N-1} \right), & 0 \leq i \leq N, \quad (g) \\
 w_{i,0} = a^{1/2}(x_i, y_0)g(x_i, y_0), & 0 \leq i \leq N, \quad (h) \\
 w_{i,N} = a^{1/2}(x_i, y_N)g(x_i, y_N), & 0 \leq i \leq N, \quad (l) \\
 w_{0,j} = a^{1/2}(x_0, y_j)g(x_0, y_j), & 1 \leq j \leq N-1, \quad (m) \\
 w_{N,j} = a^{1/2}(x_N, y_j)g(x_N, y_j), & 1 \leq j \leq N-1. \quad (n)
 \end{array} \right. \quad (4.15)$$

Le système (4.15) constitue le schéma HB2 pour le problème (4.10). Les deux systèmes (4.14) et (4.15) contiennent  $3(N+1)^2$  équations qui correspondent aux trois inconnues  $w = (w_{i,j})$ ,  $w_x = (w_{x,i,j})$ ,  $w_y = (w_{y,i,j})$ ,  $0 \leq i, j \leq N$ .

### 4.3 Forme matricielle du schéma boîte hermitien

Dans cette section, on présente la forme matricielle du problème (4.10) en utilisant le schéma boîte hermitien. Le vecteur inconnu solution est noté par  $W \in \mathbb{R}^{(N-1)^2}$  et les vecteurs dérivées sont notés par  $W_x, W_y \in \mathbb{R}^{(N-1)^2}$ . Concernant les conditions au bord, on garde les mêmes notations qu'à la section 3.3.1. Par exemple,  $W_L$  représente le vecteur des données au bord à l'intérieur de l'arête "gauche" (ou "L") du domaine carré, et  $W_{RB}$  représente la donnée au bord au sommet "droite-sud" (ou "RB") du carré (figure 3.1).

Dans la suite de cette section, pour toute fonction de grille  $v = (v_{i,j})_{0 \leq i,j \leq N} \in L_h^2$ , on utilise l'approximation suivante de la moyenne

$$\Pi^0 v \simeq (\sigma_y \circ \sigma_x) v, \quad (4.16)$$

avec  $\Pi^0 v_{i,j} = \frac{1}{4h^2} \int_{K_{i,j}} v(x) dx$ ,  $1 \leq i, j \leq N-1$ . C'est-à-dire on utilise la formule de quadrature d'ordre 4 de Simpson (voir figure 3.2). La proposition ci-dessous fournit la forme matricielle de  $(\sigma_y \circ \sigma_x) v$  en fonction des vecteurs aux points intérieurs  $V = \text{vec}_2(v)$  et des données du Dirichlet au bord.

**Proposition 4.1 (Forme matricielle de  $(\sigma_y \circ \sigma_x) v$ ).** *Pour toute fonction de grille  $v = (v_{i,j})_{0 \leq i,j \leq N} \in L_h^2$ , on a*

$$\left\{ \begin{array}{l} \text{vec}_2((\sigma_y \circ \sigma_x) v) = (P_s \otimes P_s) V \\ \quad + \frac{1}{6} (e_1 \otimes P_s) V_L + \frac{1}{6} (e_{N-1} \otimes P_s) V_R + \frac{1}{6} (P_s \otimes e_1) V_B + \frac{1}{6} (P_s \otimes e_{N-1}) V_T \\ \quad + \frac{1}{36} (e_1 \otimes e_1) V_{LB} + \frac{1}{36} (e_{N-1} \otimes e_1) V_{RB} \\ \quad + \frac{1}{36} (e_1 \otimes e_{N-1}) V_{LT} + \frac{1}{36} (e_{N-1} \otimes e_{N-1}) V_{RT}. \end{array} \right. \quad (4.17)$$

**Preuve.** Pour  $v = \sigma_x v$ , la relation (3.35)<sub>a</sub> donne

$$\begin{cases} \text{vec}_2((\sigma_y \circ \sigma_x)v) &= \text{vec}_2(\sigma_y(\sigma_x v)) \\ &= (I \otimes P_s)\text{vec}_2(\sigma_x v) + \frac{1}{6}(I \otimes e_1)\text{vec}_1(\sigma_x v)_B + \frac{1}{6}(I \otimes e_{N-1})\text{vec}_1(\sigma_x v)_T. \end{cases} \quad (4.18)$$

On peut aussi démontrer facilement les relations matricielles suivantes

$$\begin{cases} \text{vec}_1(\sigma_x v)_B = P_s V_B + \frac{1}{6}V_{LB} + \frac{1}{6}V_{RB}, \\ \text{vec}_1(\sigma_x v)_T = P_s V_T + \frac{1}{6}V_{LT} + \frac{1}{6}V_{RT}. \end{cases} \quad (4.19)$$

Donc en utilisant les relations suivantes déduites de la proposition 3.2

$$\begin{cases} (I \otimes e_1)P_s V_B = (P_s \otimes e_1)V_B, \\ (I \otimes e_{N-1})P_s V_T = (P_s \otimes e_{N-1})V_T, \\ (I \otimes e_1)e_1 V_{LB} = (e_1 \otimes e_1)V_{LB}, \\ (I \otimes e_1)e_{N-1} V_{RB} = (e_{N-1} \otimes e_1)V_{RB}, \\ (I \otimes e_{N-1})e_1 V_{LT} = (e_1 \otimes e_{N-1})V_{LT}, \\ (I \otimes e_{N-1})e_{N-1} V_{RT} = (e_{N-1} \otimes e_{N-1})V_{RT}, \end{cases} \quad (4.20)$$

et en remplaçant  $\text{vec}_2(\sigma_x v)$ ,  $\text{vec}_1(\sigma_x v)_B$  et  $\text{vec}_1(\sigma_x v)_T$  dans (4.18) par ses expressions données en (3.35)<sub>a</sub> et (4.19)<sub>a,b</sub>, on obtient (4.17). ■

**Corollaire 4.3.1.** *Pour toute fonction de grille  $v = (v_{i,j})_{0 \leq i,j \leq N}$  nulle au bord (c'est-à-dire  $v \in L_{h,0}^2$ ), on a*

$$\text{vec}_2((\sigma_y \circ \sigma_x)v) = (P_s \otimes P_s)V. \quad (4.21)$$

**Preuve.** (4.21) se déduit de (4.17) car  $v \in L_{h,0}^2$  alors les vecteurs  $W_L$ ,  $W_R$ ,  $W_B$ ,  $W_T \in \mathbb{R}^{(N-1)}$ , ainsi que, les scalaires  $W_{LB}$ ,  $W_{LT}$ ,  $W_{RB}$  et  $W_{RT}$  sont tous nuls. ■

L'intérêt principal de la proposition 4.1 et du corollaire 4.3.1 réside dans le calcul explicite des matrices associées aux opérateurs  $\mathcal{M}$  et  $-\Delta + K$ .

**Proposition 4.2.** *En appliquant le schéma boîte hermitien, (4.10) s'écrit sous forme matricielle*

$$\underbrace{\frac{1}{h^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})W + \text{vec}_2(\Pi^0(pw))}_{M_{p,h}^{\mathcal{H}}W} + G_x + G_y = Q, \quad (4.22)$$

avec  $W = \text{vec}_2(w)$ ,  $Q = \text{vec}_2(\Pi^0 q) \in \mathbb{R}^{(N-1)^2}$  et  $G_x, G_y \in \mathbb{R}^{(N-1)^2}$  sont les vecteurs correspondant aux données de bord

$$\begin{cases} G_x = \frac{1}{h^2}(\mathcal{G} \otimes P_s)(e_1 \otimes W_L + e_{N-1} \otimes W_R) \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}W_B + \mathcal{G}(e_1 W_{LB} + e_{N-1} W_{RB}) \right) \otimes e_1 \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}W_T + \mathcal{G}(e_1 W_{LT} + e_{N-1} W_{RT}) \right) \otimes e_{N-1}, \end{cases} \quad (4.23)$$

et

$$\begin{cases} G_y = \frac{1}{h^2}(P_s \otimes \mathcal{G})(W_B \otimes e_1 + W_T \otimes e_{N-1}) \\ \quad + \frac{1}{6h^2} e_1 \otimes \left( \mathcal{H}W_L + \mathcal{G}(e_1 W_{LB} + e_{N-1} W_{LT}) \right) \\ \quad + \frac{1}{6h^2} e_{N-1} \otimes \left( \mathcal{H}W_R + \mathcal{G}(e_1 W_{RB} + e_{N-1} W_{RT}) \right). \end{cases} \quad (4.24)$$

Les matrices  $\mathcal{H} = \mathcal{H}(\mathcal{A}, \mathcal{B})$  et  $\mathcal{G} = \mathcal{G}(\mathcal{B}, \mathcal{C})$  sont données en (2.117).

**Preuve.** Le problème (4.10) se réécrit

$$\begin{cases} -\Delta w = q - pw, & \text{dans } \Omega, \quad (a) \\ w = a^{1/2}g. & \text{sur } \partial\Omega. \quad (b) \end{cases} \quad (4.25)$$

Le problème (4.25) peut être formellement considéré comme un problème de Poisson avec second membre  $q - pw$ . Donc la forme matricielle de (4.25) se déduit de la proposition 3.5. Elle s'écrit

$$\frac{1}{h^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})W = \text{vec}_2(\Pi^0(q - pw)) - G_x - G_y. \quad (4.26)$$

Or

$$\text{vec}_2(\Pi^0(q - pw)) = \text{vec}_2(\Pi^0 q) - \text{vec}_2(\Pi^0(pw)). \quad (4.27)$$

Donc (4.26) devient

$$\frac{1}{h^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})W = \text{vec}_2(\Pi^0 q) - \text{vec}_2(\Pi^0(pw)) - G_x - G_y. \quad (4.28)$$

En posant  $Q = \text{vec}_2(\Pi^0 q)$  on déduit (4.22). ■

**Corollaire 4.3.2.** Les vecteurs dérivées  $W_x, W_y \in \mathbb{R}^{(N-1)^2}$  du problème (4.10) sont

$$\begin{cases} W_x = \frac{1}{h}(\mathcal{D} \otimes I)W + \frac{1}{h}(\mathcal{E} \otimes I)(e_1 \otimes W_L + e_{N-1} \otimes W_R), & (a) \\ W_y = \frac{1}{h}(I \otimes \mathcal{D})W + \frac{1}{h}(I \otimes \mathcal{E})(W_B \otimes e_1 + W_T \otimes e_{N-1}), & (b) \end{cases} \quad (4.29)$$

où  $\mathcal{D} = \mathcal{D}(\mathcal{A}, \mathcal{B})$  et  $\mathcal{E} = \mathcal{E}(\mathcal{B}, \mathcal{C})$  sont les matrices données en (2.113).

**Preuve.** Les équations (4.13)<sub>b,c</sub> de la forme mixte (4.13) sont exactement les mêmes que celles de la forme mixte du problème de Poisson (3.38) (en prenant  $w$  au lieu de  $u$ ). Donc on déduit de la proposition 3.4 les formes matricielles des vecteurs dérivées  $W_x$  et  $W_y$  données en (4.29). ■

**Corollaire 4.3.3.** En utilisant le schéma boîte hermitien, le problème

$$\begin{cases} (-\Delta + K)w = q, & \text{dans } \Omega, & (a) \\ w = 0, & \text{sur } \partial\Omega, & (b) \end{cases} \quad (4.30)$$

où  $K$  est une constante, possède la forme matricielle

$$\underbrace{\frac{1}{h^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H}) + K(P_s \otimes P_s)}_{\Delta_{K,h}^{\mathcal{H}} W} = Q, \quad (4.31)$$

avec  $W = \text{vec}_2(w)$ ,  $Q = \text{vec}_2(\Pi^0 q) \in \mathbb{R}^{(N-1)^2}$ .

**Preuve.** En supposant  $p(x, y) = K$ , on déduit de la proposition 4.2 que le problème (4.30) possède la forme matricielle

$$\frac{1}{h^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})W + K \text{vec}_2((\sigma_y \circ \sigma_x)w) + G_x + G_y = Q. \quad (4.32)$$

Les conditions de Dirichlet homogènes (4.30)<sub>b</sub> entraînent

$$G_x = G_y = 0. \quad (4.33)$$

De plus, le corollaire 4.3.1 donne

$$\text{vec}_2((\sigma_y \circ \sigma_x)w) = (P_s \otimes P_s)W, \quad (4.34)$$

d'où (4.31). ■

## 4.4 Algorithme de Concus et Golub pour le schéma boîte hermitien

### 4.4.1 Introduction

On propose une méthode itérative pour la résolution numérique d'équations elliptiques non séparables sous forme divergence en dimension deux dans un carré avec conditions de Dirichlet au bord. Le problème doit être au préalable convenablement adimensionné. La méthode est due à Concus et Golub [11]. Elle utilise entre autres le principe de relaxation suivant. Supposons que le problème non séparable original soit

$$\mathcal{L}u = f. \quad (4.35)$$

Après changement de variables, (4.35) s'écrit

$$\mathcal{M}w = q. \quad (4.36)$$

Afin de résoudre (4.36), on considère une procédure itérative formulée par Concus et Golub [11] pour les équations elliptiques du type (4.2) sur un carré. La méthode est basée sur la forme modifiée de la méthode itérative (relaxation)

$$-\Delta u^{n+1} = -\Delta u^n - \tau(\mathcal{L}u^n - f), \quad (4.37)$$

où  $\mathcal{L}u = f$  est le problème original à résoudre et  $\tau$  est un paramètre  $> 0$ . La difficulté est que (4.37) peut avoir une convergence très lente dans le cas où le rayon spectral de la matrice d'itérations est proche de 1. L'idée principale introduite par Concus et Golub est d'utiliser un "shift" dans (4.37). Considérons le problème (4.10) (problème obtenu après "scaling"). Le schéma itératif utilisé est de la forme

$$(-\Delta + K)w^{n+1} = (-\Delta + K)w^n - \tau(\mathcal{M}w^n - q), \quad (4.38)$$

où  $K$  est une constante,  $\tau$  est un paramètre  $> 0$  et  $\Delta$  est l'opérateur de Laplace. La résolution de (4.38) s'effectue en utilisant un solveur rapide de Poisson pour  $-\Delta + K$ . Dans notre cas, ce solveur est celui d'un schéma boîte hermitien. En utilisant (4.11), (4.38) devient

$$(-\Delta + K)w^{n+1} = (-\Delta + K)w^n - \tau(-\Delta + p)w^n + \tau q. \quad (4.39)$$

Dans [11], Concus et Golub utilisent le Laplacien à cinq points comme solveur de base de  $-\Delta + K$ . Ils ont ainsi montré que si la constante  $K$  est donnée par

$$K = \frac{1}{2}(\min p + \max p), \quad (4.40)$$

où  $\min p$  et  $\max p$  dénotent les valeurs minimum et maximum de  $p(x, y)$  sur le carré fermé  $\bar{\Omega}$ , alors le choix optimal de  $\tau$  qui rend minimal le rayon de convergence de la matrice d'itération associée à la méthode itérative considérée est

$$\tau = 1. \quad (4.41)$$

La relation (4.39) se simplifie sous la forme

$$(-\Delta + K)w^{n+1} = (K - p)w^n + q. \quad (4.42)$$

Dans la suite on présente la résolution de problème (4.10) par le schéma boîte hermitien en utilisant la méthode itérative (4.38).

#### 4.4.2 Schéma itératif utilisant le schéma boîte hermitien

Le schéma itératif utilisé pour résoudre (4.10) est

$$(-\Delta + K)w^{n+1} = (-\Delta + K)w^n - \tau(\mathcal{M}w^n - q). \quad (4.43)$$

La forme discrète des opérateurs  $\mathcal{M}$  et  $-\Delta + K$  sont respectivement  $M_{p,h}^{\mathcal{H}}$  et  $\Delta_{K,h}^{\mathcal{H}}$ . D'après la proposition 4.2, pour  $w$  et  $q$  deux fonctions de grilles données, la forme discrète de  $(w, q) \rightarrow \mathcal{M}w - q$  est

$$M_{p,h}^{\mathcal{H}}W - Q + G_x + G_y, \quad (4.44)$$

avec

$$M_{p,h}^{\mathcal{H}}W = \frac{1}{h^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})W + \text{vec}_2(\Pi^0(pw)). \quad (4.45)$$

Noter que la matrice associée à l'opérateur discret  $M_{p,h}^{\mathcal{H}}$  est donnée implicitement dans (4.45). Pour obtenir sa forme explicite on doit donner explicitement la forme matricielle du vecteur  $\text{vec}_2(\Pi^0(pw)) \in \mathbb{R}^{(N-1)^2}$ . En pratique ceci n'est pas utilisé dans la méthode itérative (4.43) car  $M_{p,h}^{\mathcal{H}}$  intervient seulement dans le membre de droite de la relation (4.43).



Pour l'opérateur  $-\Delta + K$ , la présence de cet opérateur des deux côtés de (4.43) a pour effet d'annuler les vecteurs de conditions au bord de sa forme discrète. C'est pour cette raison qu'on a imposé de conditions de Dirichlet homogènes en (4.30). On a vu au corollaire 4.3.3 que l'opérateur discret associé à  $-\Delta + K$  est

$$\Delta_{K,h}^{\mathcal{H}} = \frac{1}{h^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H}) + K(P_s \otimes P_s). \quad (4.46)$$

Donc la forme discrète de (4.43) est

$$\Delta_{K,h}^{\mathcal{H}} W^{n+1} = \Delta_{K,h}^{\mathcal{H}} W^n - \tau(M_{p,h}^{\mathcal{H}} W^n + G_x + G_y - Q). \quad (4.47)$$

Finalement, (4.47) est une méthode itérative pour la résolution du problème (4.10). Dans la suite on présente les étapes principales de l'itération (4.47).

### Schéma itératif utilisant HB1

Le schéma itératif utilisant HB1 s'obtient à partir de (4.47) en remplaçant les matrices  $\mathcal{H}$  et  $\mathcal{G}$  dans  $M_{p,h}^{\mathcal{H}}$ ,  $\Delta_{K,h}^{\mathcal{H}}$ ,  $G_x$ ,  $G_y$  par les matrices  $\mathcal{H}_1$  et  $\mathcal{G}_1$  en (2.48). On obtient

$$M_{p,h}^{\mathcal{H}_1} W = \frac{1}{h^2}(\mathcal{H}_1 \otimes P_s + P_s \otimes \mathcal{H}_1)W + \text{vec}_2(\Pi^0 pw). \quad (4.48)$$

Les vecteurs  $G_{x,1}$  et  $G_{y,1}$  correspondant aux conditions de bord sont

$$\left\{ \begin{array}{l} G_{x,1} = \frac{1}{h^2}(\mathcal{G}_1 \otimes P_s)(e_1 \otimes W_L + e_{N-1} \otimes W_R) \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}_1 W_B + \mathcal{G}_1(e_1 W_{LB} + e_{N-1} W_{RB}) \right) \otimes e_1 \\ \quad + \frac{1}{6h^2} \left( \mathcal{H}_1 W_T + \mathcal{G}_1(e_1 W_{LT} + e_{N-1} W_{RT}) \right) \otimes e_{N-1}, \end{array} \right. \quad (4.49)$$

et

$$\left\{ \begin{array}{l} G_{y,1} = \frac{1}{h^2}(P_s \otimes \mathcal{G}_1)(W_B \otimes e_1 + W_T \otimes e_{N-1}) \\ \quad + \frac{1}{6h^2} e_1 \otimes \left( \mathcal{H}_1 W_L + \mathcal{G}_1(e_1 W_{LB} + e_{N-1} W_{LT}) \right) \\ \quad + \frac{1}{6h^2} e_{N-1} \otimes \left( \mathcal{H}_1 W_R + \mathcal{G}_1(e_1 W_{RB} + e_{N-1} W_{RT}) \right). \end{array} \right. \quad (4.50)$$

La matrice  $\Delta_{K,h}^{\mathcal{H}_1}$  devient

$$\Delta_{K,h}^{\mathcal{H}_1} = \frac{1}{h^2}(\mathcal{H}_1 \otimes P_s + P_s \otimes \mathcal{H}_1) + K(P_s \otimes P_s). \quad (4.51)$$

**Algorithme 4.** • *Initialisation* : On choisit une fonction de grille quelconque  $w = (w_{i,j})_{0 \leq i,j \leq N}$  telle que  $w|_{\partial\Omega} = a^{1/2}g$ . On obtient alors le vecteur initial  $W^0 = \text{vec}_2(w) \in \mathbb{R}^{(N-1)^2}$ , ainsi que les vecteurs des conditions au bord  $W_L, W_R, W_B, W_T \in \mathbb{R}^{(N-1)}$  et les valeurs aux sommets  $W_{LB}, W_{LT}, W_{RB}, W_{RT} \in \mathbb{R}$ . Ensuite, on choisit deux valeurs pour  $\tau$  et  $K$  ( $\tau_0 = 1, K_0 = 0$ , par exemple).

• *Itération* :  $n = 0, 1, 2, \dots, \text{iter1}$

$$W^{n+1} = W^n - \tau_0(\Delta_{K,h}^{\mathcal{H}_1})^{-1}(M_{p,h}^{\mathcal{H}_1}W^n + G_{x,1} + G_{y,1} - Q). \quad (4.52)$$

- **Etape 1** : Calcul des vecteurs  $G_{x,1}, G_{y,1}$  et  $Q$ . On procède de la même façon qu'en (3.117), (3.118) et (3.116).

- **Etape 2** : Cette étape consiste en le calcul du produit matrice-vecteur  $M_h^p W^n$  avec

$$M_{p,h}^{\mathcal{H}_1}W^n = \frac{1}{h^2}(\mathcal{H}_1 \otimes P_s + P_s \otimes \mathcal{H}_1)W^n + \text{vec}_2(\Pi^0(pw^n)). \quad (4.53)$$

Dans (4.53)  $pw^n$  est la fonction de grille vérifiant  $(pw^n)_{i,j} = p(x_i, y_j)w_{i,j}^n, 0 \leq i, j \leq N$ . Le calcul de vecteur  $\text{vec}_2(\Pi^0(pw^n))$  est analogue à celui du  $Q = \text{vec}_2(\Pi^0 q)$  donné en (3.116).

Le produit matrice-vecteur  $(\mathcal{H}_1 \otimes P_s + P_s \otimes \mathcal{H}_1)W^n$  s'effectue en utilisant la transformée en sinus rapide. (en utilisant la forme diagonale de la matrice  $H_1$  en (3.108)).

- **Etape 3** : Le calcul de  $(\Delta_{K,h}^{\mathcal{H}_1})^{-1}(M_{p,h}^{\mathcal{H}_1}W^n + G_{x,1} + G_{y,1} - Q)$  s'effectue en utilisant l'algorithme 2.

### Schéma itératif utilisant HB2

Le schéma itératif utilisant le schéma HB2 s'obtient à partir de (4.47) en remplaçant les matrices  $\mathcal{H}$  et  $\mathcal{G}$  dans  $M_{p,h}^{\mathcal{H}}, \Delta_{K,h}^{\mathcal{H}}, G_x, G_y$  par les matrices  $\mathcal{H}_2$  et  $\mathcal{G}_2$  données en (2.127).

**Algorithme 5.** • *Initialisation* : On choisit une fonction de grille quelconque  $w = (w_{i,j})_{0 \leq i,j \leq N}$  telle que  $w|_{\partial\Omega} = a^{1/2}g$ . On obtient alors le vecteur initial  $W^0 = \text{vec}_2(w) \in \mathbb{R}^{(N-1)^2}$ , ainsi, les vecteurs des conditions au bord  $W_L, W_R, W_B, W_T \in \mathbb{R}^{(N-1)}$  et les valeurs aux sommets  $W_{LB}, W_{LT}, W_{RB}, W_{RT}$ . Ensuite, on choisit deux valeurs pour  $\tau$  et  $K$  ( $\tau_0 = 1, K_0 = 0$ , par exemple).

• *Itération* :  $n = 0, 1, 2, \dots, \text{iter2}$

$$W^{n+1} = W^n - \tau_0(\Delta_{Kx,h}^{\mathcal{H}_2})^{-1}(M_{p,h}^{\mathcal{H}_2}W^n + G_{x,2} + G_{y,2} - Q). \quad (4.54)$$

- **Etape 1** : Calcul des vecteurs  $G_{x,2}$ ,  $G_{y,2}$  et  $Q$ . On procède de la même façon qu'en (3.117), (3.118) et (3.116).

- **Etape 2** : Cette étape consiste au calcul du produit matrice-vecteur  $M_{p,h}^{\mathcal{H}_2} W^n$

$$M_{p,h}^{\mathcal{H}_2} W^n = \frac{1}{h^2} (\mathcal{H}_2 \otimes P_s + P_s \otimes \mathcal{H}_2) W^n + \text{vec}_2(\Pi^0(pw^n)), \quad (4.55)$$

avec  $pw^n$  la fonction de grille vérifiant  $(pw^n)_{i,j} = p(x_i, y_j)w_{i,j}^n$ ,  $0 \leq i, j \leq N$ . Le calcul de vecteur  $\text{vec}_2(\Pi^0(pw^n))$  est analogue à celui de  $Q = \text{vec}_2(q)$  en (3.116).

Le produit matrice-vecteur  $(\mathcal{H}_2 \otimes P_s + P_s \otimes \mathcal{H}_2) W^n$  s'effectue en utilisant l'égalité (3.140).

- **Etape 3** : Le calcul de  $(\Delta_{K,h}^{\mathcal{H}_2})^{-1}(M_{p,h}^{\mathcal{H}_2} W^n + G_{x,2} + G_{y,2} - Q)$  s'effectue en utilisant l'algorithme 3.

### Schéma itératif utilisant HB1 et HB2

On présente un algorithme de résolution de problème (4.10) combinant les algorithmes 4 et 5.

**Algorithme 6.** • *Initialisation* : On utilise l'algorithme 4 pour l'initialisation avec un critère d'arrêt. En effet, on choisit une fonction de grille quelconque  $v = (v_{i,j})_{0 \leq i,j \leq N}$  telle que  $v|_{\partial\Omega} = a^{1/2}g$ . On obtient le vecteur  $V^0 = \text{vec}_2(v) \in \mathbb{R}^{(N-1)^2}$  et les vecteurs des conditions au bord  $W_L = \text{vec}_1[v_{0,j}]_{1 \leq j \leq N-1}$ ,  $W_R = \text{vec}_1[v_{N,j}]_{1 \leq j \leq N-1}$ ,  $W_B = \text{vec}_1[v_{i,0}]_{1 \leq i \leq N-1}$ ,  $W_T = \text{vec}_1[v_{i,N}]_{1 \leq i \leq N-1} \in \mathbb{R}^{(N-1)}$  ainsi que les quatres valeurs aux sommets  $W_{LB} = v_{0,0}$ ,  $W_{LT} = v_{0,N}$ ,  $W_{RB} = v_{N,0}$ ,  $W_{RT} = v_{N,N}$ . Ensuite, on choisit deux valeurs pour  $\tau$  et  $K$  ( $\tau_0 = 1$ ,  $K_0 = 0$ , par exemple) et une tolérance ( $\text{tol} = 10^{-4}$  par exemple). On effectue les itérations suivantes en utilisant le schéma HB1.

• Pour  $\|V^{n+1} - V^n\|_2 \leq \text{tol}$ , faire

$$V^{n+1} = V^n - \tau_0 (\Delta_{K,h}^{\mathcal{H}_1})^{-1} (M_{p,h}^{\mathcal{H}_1} V^n + G_{x,1} + G_{y,1} - Q). \quad (4.56)$$

Soit  $\text{iter1}$  le nombre d'itérations de cette initialisation. Soit  $W^0$  le vecteur initial  $W^0 = V^{\text{iter1}}$ .

• *Itération* : On utilise le schéma HB2 pour obtenir le vecteur solution. Soit  $\text{iter2}$  le nombre d'itérations choisi a priori.

Pour  $n = 0, 1, 2, \dots, \text{iter2}$

$$W^{n+1} = W^n - \tau_0 (\Delta_{K,h}^{\mathcal{H}_2})^{-1} (M_{p,h}^{\mathcal{H}_2} W^n + G_{x,2} + G_{y,2} - Q). \quad (4.57)$$

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	iter1	iter2	temps(s.)
$8 \times 8$	1.055(-12)	2.943(-12)	2.163(-12)	6.003(-12)	5	8	0.38
$16 \times 16$	2.664(-13)	7.940(-13)	5.478(-13)	1.624(-12)	6	8	0.54
$32 \times 32$	3.793(-13)	1.164(-12)	7.789(-13)	2.872(-12)	6	8	0.90
$64 \times 64$	9.646(-14)	3.614(-13)	2.004(-13)	2.629(-12)	6	8	1.77

TAB. 4.1: Erreur et taux de convergence de Schéma HB2 avec

$$u_{ex}(x, y) = 2[(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2] \text{ et } a(x, y) = [1 + \sin(\frac{\pi}{2}(x + y))]^2$$

### 4.4.3 Résultats numériques

Dans cette partie nous donnons quelques résultats numériques pour le problème elliptique (4.10) en utilisant les schémas donnés dans les algorithmes 5 et 6 pour résoudre le problème (4.10) sur le carré  $\Omega = (0, 1)^2$ . Dans les tables de résultats numériques,  $\text{iter}_2$  représente le nombre d'itérations indiqué dans chacun des algorithmes 5 et 6 et  $\text{iter}_1$  représente le nombre d'itérations de l'initialisation indiqué dans l'algorithme 6. Les conditions limites sont données par la solution exacte. Les erreurs pour la solution exacte et sa dérivée sont calculées par les normes  $L^2$  et  $L^\infty$  discrètes

$$\begin{cases} \|u_{ex} - u\|_h = \left( h^2 \sum_{i,j=1}^{N-1} (u_{ex}(x_i, y_j) - u_{i,j})^2 \right)^{\frac{1}{2}}, \\ \|u_{ex} - u\|_\infty = \max_{i,j=1,\dots,N-1} |u_{ex}(x_i, y_j) - u_{i,j}|. \end{cases} \quad (4.58)$$

Les vecteurs dérivés  $W_x, W_y \in \mathbb{R}^{(N-1)^2}$  sont calculés à partir de (4.29). Dans chaque table des résultats numériques on précise également le nombre d'itérations et le temps calcul.

**Test 1 :** La solution exacte est  $u_{ex}(x, y) = 2[(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2]$  avec  $a(x, y) = [1 + \sin(\frac{\pi}{2}(x + y))]^2$ . On utilise l'algorithme 6. Les résultats donnés dans la table 4.1 montrent une convergence vers la solution exacte en 8 itérations.

**Test 2 :** On suppose  $u_{ex}(x, y) = x^2 + y^2$  et  $a(x, y) = \delta + x + y$  (figures 4.1).

**Test 2.1 :** On suppose  $\delta = 1$ . Les résultats obtenus par l'algorithme 5, sont donnés dans la table 4.2.

**Test 2.2 :** On suppose  $\delta = 0.1$ . Les résultats obtenus par l'algorithme 5, sont donnés dans la table 4.3.

Dans les tests 2.1 et 2.2, une très bonne précision est atteinte sur la grille finale  $128 \times 128$  en cinq itérations.

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	iter2	temps(s.)
$16 \times 16$	4.484(-7)	3.843(-6)	2.302(-6)	2.191(-5)	5	0.22
T.C.	3.97	3.32	3.79	2.70		
$32 \times 32$	3.083(-8)	3.851(-7)	1.653(-7)	3.370(-6)	5	0.39
T.C.	3.99	3.39	3.89	2.78		
$64 \times 64$	1.939(-9)	3.662(-8)	1.113(-8)	4.901(-7)	5	0.91
T.C.	4.02	3.44	3.94	2.88		
$128 \times 128$	1.192(-10)	3.374(-9)	7.239(-10)	6.657(-8)	5	2.52
T.C.	4.06	3.46	3.97	2.93		
$256 \times 256$	7.122(-12)	3.062(-10)	4.616(-11)	8.686(-9)	5	8.21

TAB. 4.2: Erreur et taux de convergence de Schéma HB2 avec  $u_{ex}(x, y) = x^2 + y^2$  et  $a(x, y) = 1 + x + y$  après 5 itérations.

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	iter2	temps(s.)
$16 \times 16$	5.443(-6)	4.570(-5)	5.357(-5)	4.597(-4)	5	0.23
T.C.	3.64	2.84	3.05	2.06		
$32 \times 32$	4.364(-7)	6.360(-6)	6.443(-6)	1.101(-4)	5	0.46
T.C.	3.27	3.05	3.34	2.36		
$64 \times 64$	4.518(-8)	7.678(-7)	6.364(-7)	2.136(-5)	5	1
T.C.	0.95	2.67	3.30	2.53		
$128 \times 128$	2.335(-8)	1.201(-7)	6.421(-8)	3.677(-6)	5	2.7
T.C.	0.05	0.26	0.13	2.33		
$256 \times 256$	2.242(-8)	9.992(-8)	5.860(-8)	7.304(-7)	5	9.22

TAB. 4.3: Erreur et taux de convergence de schéma HB2 avec  $u_{ex}(x, y) = x^2 + y^2$  et  $a(x, y) = 0.1 + x + y$  après 5 itérations.

### Test 3 : Cas d'un coefficient de diffusion oscillant.

Dans cet exemple, on considère un coefficient de diffusion de la forme

$$a(x, y) = (4 + \sin(4\pi x) \sin(4\pi y))^2, \quad (4.59)$$

avec solution exacte  $u_{ex}(x, y) = \sin(2\pi x) \sin(2\pi y)$ .

Ce cas est difficile car  $a(x, y)$  et sa dérivée sont oscillantes (figures 4.2). La résolution de (4.10) est effectuée par l'algorithme 6. On constate que les taux de convergence sont compris entre 2 et 4 pour iter2 = 5 itérations (table 4.4) et entre 3 et 4 pour iter2 = 10 itérations (table 4.5). On renvoie à [8] pour davantage de commentaires sur la difficulté de ce cas. Les résultats obtenus dans [8] sont un peu plus précis. Cependant, le schéma

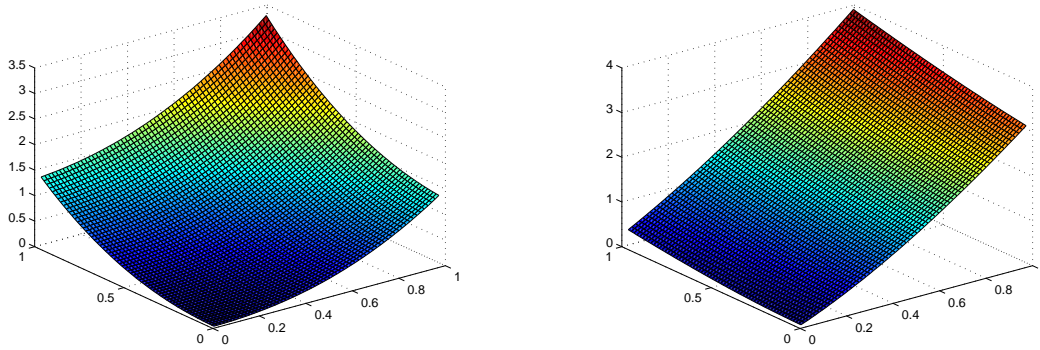


FIG. 4.1: Solution exacte  $u_{ex}(x, y) = x^2 + y^2$  (à gauche) et sa dérivée  $\partial_x u(x, y) = 2x$  (à droite)

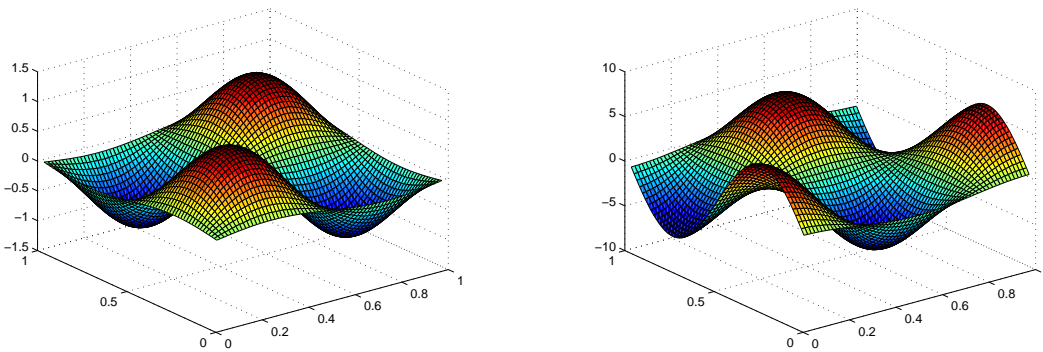


FIG. 4.2: solutions exacte  $u_{ex}(x, y) = \sin(2\pi x)\sin(2\pi y)$  (à gauche) et sa dérivée (à droite)

HB2 calcule les dérivées ce qui n'est pas le cas dans [8].

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	iter1	iter2	temps(s.)
$32 \times 32$	1.622(-4)	2.800(-3)	4.655(-4)	1.620(-2)	27	5	0.79
T.C.	3.82	3.79	3.72	2.81			
$64 \times 64$	1.144(-5)	2.689(-4)	3.484(-5)	2.300(-3)	28	5	1.82
T.C.	2.72	3.45	2.97	2.96			
$128 \times 128$	1.727(-6)	2.458(-5)	4.445(-6)	2.935(-4)	30	5	4.63
T.C.	2.00	3.23	2.19	2.96			
$256 \times 256$	4.451(-7)	2.608(-6)	9.715(-7)	3.768(-5)	31	5	13.86

TAB. 4.4: Erreur et taux de convergence du Schéma HB2  $u_{ex}(x, y) = \sin(2\pi x) \sin(2\pi y)$  et  $a(x, y) = (4 + \sin(4\pi x) \sin(4\pi y))^2$  après 5 itérations

$N_x \times N_y$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	iter1	iter2	temps(s.)
$32 \times 32$	1.684(-4)	2.782(-3)	4.514(-4)	1.610(-2)	33	10	1.10
T.C.	3.91	3.38	3.80	2.80			
$64 \times 64$	1.119(-5)	2.671(-4)	3.241(-5)	2.300(-3)	33	10	2.57
T.C.	4.01	3.49	3.91	3.00			
$128 \times 128$	6.915(-7)	2.360(-5)	2.141(-6)	2.889(-4)	33	10	7.18
T.C.	3.99	3.50	3.56	3.00			
$256 \times 256$	4.350(-8)	2.079(-6)	1.814(-7)	3.545(-5)	33	10	21.69

TAB. 4.5: Erreur et taux de convergence du Schéma HB2 avec  $u_{ex}(x, y) = \sin(2\pi x) \sin(2\pi y)$  et  $a(x, y) = (4 + \sin(4\pi x) \sin(4\pi y))^2$  après 10 itérations

## 4.5 Algorithmes de type “Local Defect Correction” pour les schémas boîte hermitiens

### 4.5.1 Introduction

La méthode Local Defect Correction (LDC) a été introduite par W. Hackbusch [22] pour la résolution de problèmes elliptiques. Il s’agit d’un algorithme générique itératif de type multiniveau pour la résolution d’un problème discret. Deux ou plusieurs grilles de calcul sont utilisées.

Une analyse de l’algorithme LDC dans le contexte différences finies a été effectuée par Ferket et al. [14] et Anthonissen [2] dans le contexte de problèmes en combustion. Dans [2], l’algorithme LDC est combiné avec des techniques de grilles adaptatives, de raffinement et de décomposition de domaines.

Le principe de la méthode LDC est le suivant. On considère le problème elliptique

$$\begin{cases} Lu = f, & \text{dans } \Omega, & (a) \\ u = g, & \text{sur } \partial\Omega, & (b) \end{cases} \quad (4.60)$$

où  $L$  est un opérateur différentiel linéaire elliptique,  $f$  et  $g$  sont respectivement le terme source et la condition de Dirichlet au bord. Pour discrétiser (4.60), on choisit d'abord un maillage grossier global (de pas de discrétisation  $H$ ), qu'on dénote par  $\Omega^H$ . Soit  $L^H$  un opérateur discret approchant l'opérateur continu  $L$ . Une approximation initiale  $u_i^H$ ,  $i = 0$  sur  $\Omega^H$  est obtenue par résolution du système

$$L^H u_i^H = f^H. \quad (4.61)$$

Dans (4.61), le vecteur  $f^H$  contient le terme source  $f$  ainsi que les conditions de Dirichlet  $g$ . On suppose que  $L^H$  est inversible.

Supposons de plus que la solution exacte  $u(x, y)$  de (4.60) possède une région de grande "activité". Ceci signifie par exemple qu'elle possède une région de fort gradient dans une partie du domaine (figures 4.5, 4.6). On choisit un sous-domaine  $\Omega_l \subset \Omega$  tel que cette région de haute activité de  $u(x, y)$  soit contenue dans  $\Omega_l$ . Le sous domaine  $\Omega_l$  est ensuite discrétisé par une grille fine locale (de pas de discrétisation  $h$ ). On la dénote  $\Omega_l^h$ . La grille fine  $\Omega_l^h$  est choisie telle que  $\Omega_l^H \cap \Omega_l \subset \Omega_l^h$ . C'est-à-dire les points grossiers qui sont dans la région du raffinement sont aussi des points de la grille fine  $\Omega_l^h$ . Pour formuler un problème discret sur  $\Omega_l^h$ , on doit définir des conditions au bord artificielles sur  $\Gamma$ , l'interface entre  $\Omega_l$  et  $\Omega \setminus \Omega_l$  (figure 4.3). Dans la suite, on utilise un opérateur d'interpolation  $P^{h,H}$  pour obtenir des conditions au bord artificielles sur  $\Gamma$ . L'opérateur  $P^{h,H}$  donne des valeurs aux points de la grille fine qui sont sur l'interface  $\Gamma^h$  en interpolant des valeurs aux points de la grille grossière qui sont sur l'interface  $\Gamma^H$ .

Sur la partie  $\partial\Omega_l \setminus \Gamma$ , on obtient des conditions au bord en utilisant la condition de Dirichlet  $g$  en (4.60)<sub>b</sub>.

Dans la simulation numérique, l'opérateur  $P^{h,H}$  est basé sur une interpolation spline. Sur la grille fine  $\Omega_l^h$  on considère le problème discret suivant, à l'itération  $i = 0$ ,

$$L_l^h u_{l,i}^h = f_l^h - B_{l,\Gamma}^h u_{l,i}^h|_{\Gamma}, \quad (4.62)$$

où

$$u_{l,i}^h|_{\Gamma} = P^{h,H} u_i^H|_{\Gamma}. \quad (4.63)$$



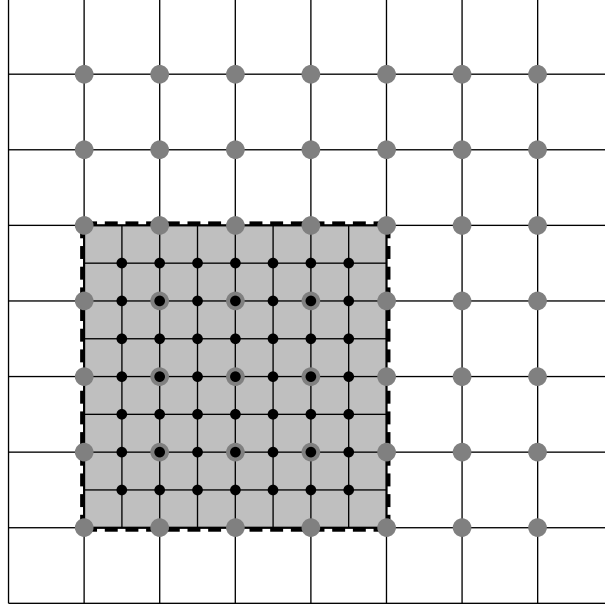


FIG. 4.3: Une grille grossière  $\Omega^H$  et une grille fine  $\Omega_l^h$ . La région sombre est celle du grande activité  $\Omega_l$ . L'interface  $\Gamma$  est pointillé. Les grands points gris sont les noeuds intérieurs de la grille grossière. Les points noirs sont les noeuds intérieurs de la grille fine. Noter que  $\Gamma = \partial\Omega_l$  dans ce cas.

Dans (4.62),  $i$  représente un indice d'itération et  $B_{l,\Gamma}^h$  est une matrice carrée qui représente la dépendance de la solution sur la grille fine par rapport aux conditions au bord artificielles sur  $\Gamma$ . Dans (4.62), l'opérateur  $L_l^h$  (supposé inversible) est une approximation discrète de l'opérateur  $L$  sur le sous-domaine  $\Omega_l$ . Dans le terme de droite de (4.62),  $f_l^h$  contient le terme source  $f$  ainsi que la condition au bord de Dirichlet  $g$  sur  $\partial\Omega_l \setminus \Gamma$  donnée en (4.60). Noter que  $\Gamma \cap \partial\Omega$  peut être vide comme sur la figure 4.3. On utilise ensuite la solution obtenue sur la grille fine  $u_{l,0}^h$  pour reconstruire une approximation sur la grille grossière. En effet, supposons que l'on connaisse la solution exacte  $u(x, y)$  de (4.60). On définit le "défaut" de  $u(x, y)$  sur la grille  $\Omega^H$  par

$$d^H = L^H(u|_{\Omega^H}) - f^H. \quad (4.64)$$

Si l'on connaît les valeurs du défaut  $d^H$ , on les utilise pour trouver une approximation plus précise sur la grille grossière. Ceci s'obtient en plaçant  $d^H$  au membre de droite de (4.61). Cependant, comme on ne connaît pas la solution exacte du problème continu, on ne peut pas calculer  $d^H$ . On utilise alors l'approximation  $u_{l,0}^h$  calculée sur la grille fine pour approcher  $d^H$ . En effet, pour tout  $i$ , indice d'itération, la fonction  $w_i^H, i = 0$ ,

est la fonction sur la grille grossière définie par :

$$w_i^H(x, y) = \begin{cases} u_{l,i}^h(x, y), & (x, y) \in \Omega_l^H, \\ u_i^H(x, y), & (x, y) \in \Omega^H \setminus \Omega_l^H. \end{cases} \quad (4.65)$$

Pour chaque indice d'itération  $i$  on définit  $d_i^H$  par

$$d_i^H = L^H w_i^H - f^H. \quad (4.66)$$

$d_i^H$  fournit une estimation de l'erreur de la discrétisation locale aux points de la grille

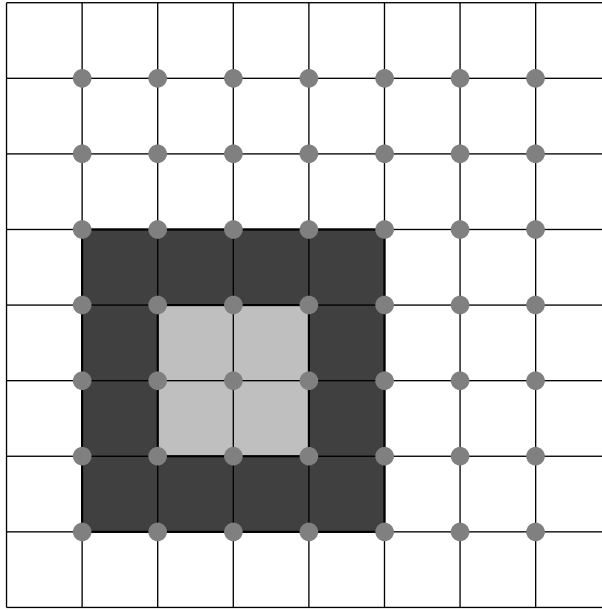


FIG. 4.4: La région de raffinement est la réunion de la partie grise et la partie noire. Le “safety region” est la partie noire. Le défaut est calculé seulement dans la partie grise. Dans ce cas,  $\Omega_l^H$  et  $\Omega_\epsilon^H$  sont formés de 9 et 1 points respectivement

grossière sur tout  $\Omega_l^H$ . On a observé qu’il est préférable de ne pas utiliser l’approximation (4.66) sur tous les points de  $\Omega_l^H$  mais sur un sous-ensemble de  $\Omega_l^H$  seulement. En particulier, les points de la grille grossière situés près de  $\partial\Omega_l$  doivent être exclus. Ceci amène à introduire un “safety region” noté par  $\Omega_\epsilon^H$  (figure 4.4). L’estimation de l’erreur locale de la discrétisation de la grille grossière est placée dans le membre de droite des équations correspondant aux points grossiers appartenant à l’ensemble  $\Omega_\epsilon^H$  seulement. On applique ensuite l’étape de correction sur la grille grossière pour trouver  $u_i^H, i = 1,$

$$L^H u_i^H = \begin{cases} f^H(x, y) + d_{i-1}^H(x, y), & (x, y) \in \Omega_\epsilon^H, \\ f^H(x, y), & (x, y) \in \Omega^H \setminus \Omega_\epsilon^H. \end{cases} \quad (4.67)$$

Comme (4.67) contient des estimations de l'erreur local de la discrétisation sur la grille grossière, on s'attend à ce que  $u_{i+1}^H$  soit une approximation plus précise que  $u_i^H$ . On obtient alors une meilleur condition au bord sur  $\Gamma$  ainsi qu'une solution sur la grille fine locale par (4.62) avec  $i = 1$ . En remplaçant  $i + 1$  par la valeur  $i$  on obtient une méthode itérative.

En effectuant des itérations sur l'indice  $i$ , on obtient l'algorithme suivant :

**Algorithme 7** (LDC avec le domaine de raffinement  $\Omega_l$  choisi à priori). • *Initialisation*

- Résoudre le problème (4.61) sur la grille grossière  $\Omega^H$  avec  $i = 0$ . On obtient le vecteur  $u_0^H$ .

- Résoudre le problème (4.62) sur la grille fine locale  $\Omega_l^h$  avec  $i = 0$ . On obtient le vecteur  $u_{l,0}^h$ .

• *Itération  $i = 1, 2, \dots$*

- Calculer  $W_{i-1}^H$  par (4.65).

- Calculer  $d_{i-1}^H$  par (4.66).

- Résoudre le nouveau problème (4.67) sur la grille grossière  $\Omega^H$ . On obtient le vecteur  $u_i^H$ .

- Résoudre le problème (4.62) sur la grille fine locale  $\Omega_l^h$ . On obtient le vecteur  $u_{l,i}^h$ .

En pratique, une ou deux itérations suffisent souvent pour obtenir une approximation satisfaisante sur la grille composée  $\Omega^{H,h}$ , c'est-à-dire la grille formée de la réunion de la grille grossière et la grille fine locale,  $\Omega^{H,h} = \Omega^H \cup \Omega_l^h$ .

L'algorithme 7 est une version élémentaire de la méthode LDC. Plusieurs généralisations sont possibles :

- Usage de plusieurs grilles fines locales. Noter que les problèmes locaux sont indépendants les uns des autres et peuvent être résolus simultanément. Raffinement récursif où une grille fine locale peut être une grille grossière pour une autre grille fine locale.
- Usage de discrétisations différentes. On renvoie à [2] pour une analyse détaillée du comportement de la convergence pour le problème du Poisson résolu avec le schéma aux différences finies classique à cinq points.

### 4.5.2 Schéma boîte hermitien et algorithme LDC en dimension 1

Dans cette section, on utilise la méthode LDC présentée à la section 4.5.1 combinée avec le schéma boîte hermitien. On considère le problème

$$\begin{cases} -u''(x) = f(x), & x \in \Omega = (a, b), \\ u(a) = g_a, u(b) = g_b. \end{cases} \quad (4.68)$$

On discrétise  $\Omega$  par la grille (1.4-1.5) avec pas de maillage  $H = 1/N$ . On note les points discrets par  $x_i^H = a + iH$ ,  $0 \leq i \leq N$ . La grille grossière est  $\Omega^H$ . Les vecteurs inconnus sont  $U^H$  et  $U_x^H \in \mathbb{R}^{(N-1)}$  tels que  $U_i^H \approx u(x_i^H)$  et  $U_{x,i}^H \approx u'(x_i^H)$ . Pour éviter toute confusion entre les variables, on effectue un changement des notations pour la suite

$$\begin{cases} H = h_c, & (\text{Pas de maillage grossier}) \\ h = h_f. & (\text{Pas de maillage fin}) \end{cases} \quad (4.69)$$

D'après ce qu'on a vu au chapitre 2 dans la proposition 2.7, le schéma boîte hermitien associé au problème (4.68) possède la forme matricielle

$$\frac{1}{h_c^2} \mathcal{H} U_i^{h_c} = F - \frac{1}{h_c^2} \mathcal{G}(e_1 g_a + e_{N-1} g_b), \quad (4.70)$$

avec  $F = \text{vec}_1(\Pi^0 f)$ . En comparant avec (4.61) on trouve

$$\begin{cases} L^{h_c} = \frac{1}{h_c^2} \mathcal{H}, \\ f^{h_c} = F - \frac{1}{h_c^2} \mathcal{G}(e_1 g_a + e_{N-1} g_b). \end{cases} \quad (4.71)$$

En résolvant (4.70) on obtient le vecteur solution  $U_i^{h_c}$ ,  $i = 0$ , sur  $\Omega^{h_c}$  et le vecteur solution globale  $\tilde{U}_i^{h_c} \in \mathbb{R}^{(N+1)}$

$$\tilde{U}_i^{h_c} = [g_a, U_{i,1}^{h_c}, U_{i,2}^{h_c}, \dots, U_{i,N-1}^{h_c}, g_b]^T. \quad (4.72)$$

Le domaine  $\Omega_l$  a été choisi a priori tel que  $\Omega_l = (a_1, b_1)$ , avec  $a_1, b_1$  deux scalaires fixés entre  $a$  et  $b$ ,  $a \leq a_1 < b_1 \leq b$ . Le choix des  $a_1$  et  $b_1$  est adapté tel que pour tout  $N$  il existe  $i_0$  et  $j_0$  entre 1 et  $N - 1$  tels que

$$a_1 = x^{h_c}(i_0), \quad b_1 = x^{h_c}(j_0), \quad (4.73)$$

c'est-à-dire  $a_1$  et  $b_1$  sont deux points de la grille grossière  $\Omega^{h_c}$ . Pour que  $\Omega_l^h$  possède le même nombre de points que  $\Omega^H$ , on choisit  $a_1$  et  $b_1$  tels que  $b_1 - a_1 = \frac{b-a}{2}$ . Supposons que l'on connaisse les valeurs  $g_{a_1} = u(a_1)$  et  $g_{b_1} = u(b_1)$  où  $u(x)$  est la solution exacte. Alors sur le domaine  $\Omega_l$  on peut écrire

$$\begin{cases} -u_l'' = f_l, & \text{dans } \Omega_l, \\ u_l(a_1) = g_{a_1}, \quad u_l(b_1) = g_{b_1}, \end{cases} \quad (4.74)$$

avec  $u_l = u|_{\Omega_l}$  et  $f_l = f|_{\Omega_l}$ . On discrétise à nouveau  $\Omega_l$  par la grille (1.4-1.5) avec  $h_f = \frac{h_c}{2}$ . On obtient les points discrets  $x_i^{h_f}$ ,  $x_i^{h_f} = a_1 + ih_f$ ,  $0 \leq i \leq N_f$ , avec  $N_f = \frac{b_1 - a_1}{h_f}$ . Noter l'égalité entre  $N$  et  $N_f$  quand  $b_1 - a_1 = \frac{b-a}{2}$ . La grille fine est  $\Omega_l^{h_f}$ . Les vecteurs inconnus sont  $U_l^{h_f}$  et  $U_{l,x}^{h_f} \in \mathbb{R}^{(N-1)}$  tels que  $U_{l,i}^{h_f} \approx u_l(x_i^{h_f})$  et  $U_{l,x,i}^{h_f} \approx u_l'(x_i^{h_f})$ ,  $1 \leq i \leq N-1$ . La frontière entre  $\Omega_l$  et  $\Omega$  est  $\partial\Omega_l = \{a_1, b_1\}$ . D'après ce qu'on a vu au chapitre 2 dans la proposition 2.7, le schéma boîte hermitien associé à (4.74) est

$$\frac{1}{h_f^2} \mathcal{H} U_i^{h_f} = F_l - \frac{1}{h_f^2} \mathcal{G}(e_1 g_{a_1} + e_{N-1} g_{b_1}), \quad (4.75)$$

avec  $F_l = \text{vec}_1(\Pi^0 f_l)$ . Le problème est que l'on ne connaît pas en fait les valeurs  $g_{a_1}$  et  $g_{b_1}$  sur  $\partial\Omega_l$  dans (4.74). Pour les approcher, on les remplace par les conditions au bord artificielles

$$\begin{cases} g_{a_1} \approx \tilde{U}_{0,i_0}^{h_c}, \\ g_{b_1} \approx \tilde{U}_{0,j_0}^{h_c}, \end{cases} \quad (4.76)$$

avec  $i_0$  et  $j_0$  sont les indices des points  $a_1$  et  $b_1$  dans la grille grossière définis en (4.73). Ici, en particulier, la frontière  $\partial\Omega_l = \{a_1, b_1\}$  est formée des deux points de la grille grossière  $\Omega^{h_c}$  donc il n'y a pas d'interpolation alors la matrice  $B_{l,\Gamma}^{h_c}$  est nulle. En comparant (4.75) et (4.62) on obtient

$$\begin{cases} L_l^{h_f} = \frac{1}{h_f^2} \mathcal{H}, \\ f_l^{h_f} = F_l - \frac{1}{h_f^2} \mathcal{G}(e_1 \tilde{U}_{i_0}^{h_c} + e_{N-1} \tilde{U}_{j_0}^{h_c}). \end{cases} \quad (4.77)$$

En résolvant (4.75) avec  $i = 0$

$$\frac{1}{h_f^2} \mathcal{H} U_{l,i}^{h_f} = F_l - \frac{1}{h_f^2} \mathcal{G}(e_1 \tilde{U}_{i,i_0}^{h_c} + e_{N-1} \tilde{U}_{i,j_0}^{h_c}), \quad (4.78)$$

on obtient l'approximation  $U_{l,0}^{h_f}$ . D'après (4.65) le vecteur  $W_i^{h_c} = (W_{i,j}^{h_c})_{1 \leq j \leq N-1}$  vérifie

$$W_i^{h_c} = \begin{cases} U_{l,i}^{h_c}, & \text{dans } \Omega_l^{h_c}, \\ U_i^{h_c}, & \text{dans } \Omega^{h_c} \setminus \Omega_l^{h_c}. \end{cases} \quad (4.79)$$

Le vecteur de défaut  $d_i^{h_c} \in \mathbb{R}^{(N-1)}$  est calculé par

$$d_i^{h_c} = \frac{1}{h_c^2} \mathcal{H} W_i^{h_c} - f^{h_c}. \quad (4.80)$$

Finalement, l'algorithme utilisant LDC et le schéma boîte hermitien est le suivant :

**Algorithme 8.** • *Initialisation*

- Résoudre le problème (4.70) sur la grille grossière  $\Omega^{h_c}$ . Le vecteur solution est noté  $U_0^{h_c} \in \mathbb{R}^{(N-1)}$ .

- Résoudre le problème (4.78) sur la grille fine locale  $\Omega_l^{h_f}$ . Le vecteur solution est noté  $U_{l,0}^{h_f} \in \mathbb{R}^{(N-1)}$ .

• Itération  $i = 1, 2, \dots$

- Calculer le vecteur  $W_{i-1}^{h_c}$  par (4.79).

- Calculer le vecteur de défaut  $d_{i-1}^{h_c}$  par (4.80).

- Résoudre sur la grille grossière

$$\frac{1}{h_c^2} \mathcal{H}(U_i^{h_c})_j = \begin{cases} f_j^{h_c} + d_{i-1,j}^{h_c}, & x_j^{h_c} \in \Omega_\epsilon^{h_c}, \\ f_j^{h_c}, & x_j^{h_c} \in \Omega^{h_c} \setminus \Omega_\epsilon^{h_c}, \end{cases} \quad (4.81)$$

avec  $\epsilon > 0$ . Le vecteur solution est noté  $U_i^{h_c} \in \mathbb{R}^{(N-1)}$ .

- Résoudre le problème (4.78) sur la grille fine. Le vecteur solution est noté  $U_{l,i}^{h_f} \in \mathbb{R}^{(N-1)}$ .

### 4.5.3 Résultats numériques

Dans la section 4.5.2 on a donné la description générale de l'algorithme LDC couplé au schéma boîte hermitien. L'algorithme utilisé ici se déduit de l'algorithme 8 en remplaçant les matrices  $\mathcal{H}$  et  $\mathcal{G}$  utilisées dans (4.75) et (4.78) par les matrices  $\mathcal{H}_2$  et  $\mathcal{G}_2$  données en (2.127).

Dans cette section on donne quelques résultats numériques prouvant l'intérêt de schéma boîte HB2 d'ordre 4 combiné avec l'algorithme LDC. On utilise la norme  $L^\infty$  discrète

$h_f$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	points de grille
1/128	1.289(-4)	3.778(-3)	128
Taux de conv.	4.00	4.00	
1/256	7.832(-6)	2.391(-4)	256
Taux de conv.	4.00	4.00	
1/512	4.861(-7)	1.482(-5)	512
Taux de conv.	4.00	4.00	
1/1024	3.033(-8)	9.246(-7)	1024
Taux de conv.	4.00	4.00	
1/2048	1.887(-9)	5.775(-8)	2048

TAB. 4.6: Erreur et taux de convergence pour le test 1 sur une grille uniforme

$h_c$	$\ u_{ex} - u^{h_f, h_c}\ _\infty$	$\ u_{x,ex} - u_x^{h_f, h_c}\ _\infty$	points de grille
1/64	1.289(-4)	3.778(-3)	96
Taux de conv.	4.00	4.00	
1/128	7.832(-6)	2.391(-4)	192
Taux de conv.	4.00	4.00	
1/256	4.861(-7)	1.482(-5)	384
Taux de conv.	4.00	4.00	
1/512	3.033(-8)	9.246(-7)	768
Taux de conv.	4.00	4.00	
1/1024	1.887(-9)	5.775(-8)	1536

TAB. 4.7: Erreur et taux de convergence pour le test 1 sur une grille composée

pour estimer l'erreur

$$\|u_{ex} - u\|_\infty = \max_{i=1, \dots, N-1} |u_{ex}(x_i) - u_i|. \quad (4.82)$$

**Test 1 :** On considère la gaussienne  $u_{ex}(x) = \exp(-500(x - \frac{1}{2})^2)$ . Cette fonction a une région de haute activité autour de  $x = \frac{1}{2}$ . Les résultats obtenus sur la grille composée sont comparés avec ceux calculés par le schéma HB2 sur une grille uniforme avec pas de maille fin  $h_f$ . Observons qu'une précision identique est obtenue par les deux approches avec moins de points sur la grille composée.

**Test 2 :** Ici, on suppose que la solution exacte est  $u_{ex}(x) = \frac{1}{2}[\tanh 50((x - \frac{1}{2})) + 1]$  sur  $\Omega = (0, 1)$ . Cette fonction possède un fort gradient autour de  $x = \frac{1}{2}$ , donc on choisit  $\Omega_l = (\frac{1}{4}, \frac{3}{4})$  et  $\Omega_\epsilon = (\frac{1}{4} + \epsilon, \frac{3}{4} - \epsilon)$  avec  $\epsilon = \frac{1}{16}$ .

Dans les deux tests précédents, on voit constate qu'une grille de taille  $N = 1024$  donne

$h_f$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	points de grille
1/256	3.481(-5)	3.500(-3)	256
Taux de conv.	4.02	4.08	
1/512	2.082(-6)	2.061(-4)	512
Taux de conv.	4.01	4.02	
1/1024	1.291(-7)	1.269(-5)	1024
Taux de conv.	4.00	4.00	
1/2048	8.062(-9)	7.904(-7)	2048
Taux de conv.	3.99	4.00	
1/4096	5.0406(-10)	4.935(-8)	4096

TAB. 4.8: Erreur et taux de convergence pour le test 2 sur une grille uniforme

$h_c$	$\ u_{ex} - u^{h_f, h_c}\ _\infty$	$\ u_{x,ex} - u_x^{h_f, h_x}\ _\infty$	points de grille
1/128	3.481(-5)	3.500(-3)	192
Taux de conv.	4.02	4.08	
1/256	2.082(-6)	2.061(-4)	384
Taux de conv.	4.01	4.02	
1/512	1.291(-7)	1.269(-5)	768
Taux de conv.	4.00	4.00	
1/1024	8.062(-9)	7.904(-7)	1536
Taux de conv.	3.99	4.00	
1/2048	5.0406(-10)	4.935(-8)	3072

TAB. 4.9: Erreur et taux de convergence pour test 2 sur une grille composée

la même précision qu'un calcul direct sur une grille de taille  $N = 2048$ .



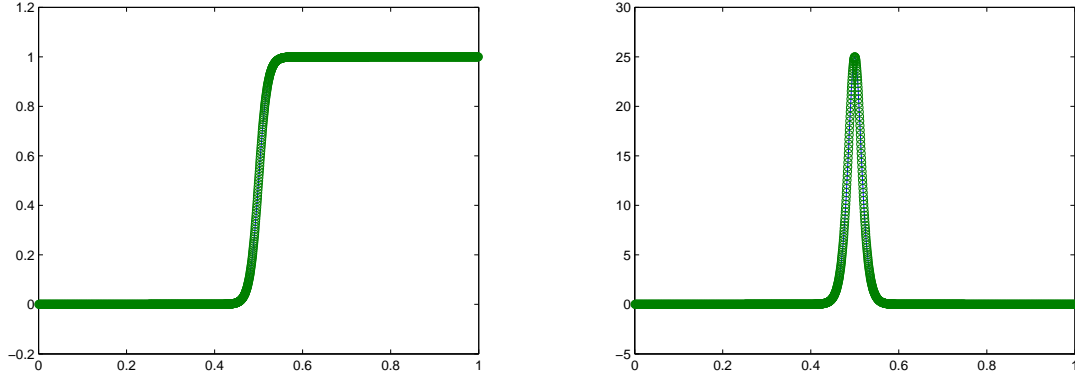


FIG. 4.5: La solution exacte du test 2 (à gauche) et sa dérivée (à droite)

#### 4.5.4 Schéma boîte hermitien avec LDC en dimension 2

Dans cette partie, on utilise la méthode LDC avec le schéma boîte hermitien pour résoudre le problème de Poisson dans le carré  $\Omega = (a, b)^2$

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega = (a, b)^2, \\ u = g, & \text{sur } \partial\Omega. \end{cases} \quad (4.83)$$

Le domaine  $\Omega$  est discrétisé par la grille (3.1) avec un pas de maillage  $h_c = \frac{1}{N}$ . Les points de discrétisation sont  $x_i^{h_c} = a + ih_c$ ,  $0 \leq i \leq N$ , et  $y_j^{h_c} = a + jh_c$ ,  $0 \leq j \leq N$ . La grille grossière est  $\Omega^{h_c}$ . Les vecteurs inconnus sont  $U^{h_c}$ ,  $U_x^{h_c}$ ,  $U_y^{h_c} \in \mathbb{R}^{(N-1)^2}$ . Comme on l'a déjà vu au chapitre 3 dans la proposition 3.5, le schéma boîte hermitien associé au problème (4.83) possède la forme matricielle

$$\frac{1}{h_c^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})U^{h_c} = F - G_x^{h_c} - G_y^{h_c}, \quad (4.84)$$

avec  $F$ ,  $G_x^{h_c}$  et  $G_y^{h_c} \in \mathbb{R}^{(N-1)^2}$  donnés en (3.87), (3.88) et (3.89) respectivement. En comparant avec (4.61) on trouve

$$\begin{cases} L^{h_c} = \frac{1}{h_c^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H}), \\ f^{h_c} = F - G_x^{h_c} - G_y^{h_c}. \end{cases} \quad (4.85)$$

La résolution de (4.84) donne le vecteur solution initial  $U_0^{h_c} \in \mathbb{R}^{(N-1)^2}$ . Le domaine  $\Omega_l$  a été choisi a priori tel que  $\Omega_l = (a_1, b_1) \times (c_1, d_1)$  où  $a_1$ ,  $b_1$ ,  $c_1$ ,  $d_1$  sont des scalaires

fixés tels que  $a \leq a_1 < b_1 \leq b$  et  $a \leq c_1 < d_1 \leq b$ . Le choix des  $a_1, b_1, c_1, d_1$  est tel que les points  $(a_1, c_1), (a_1, d_1), (b_1, c_1), (b_1, d_1)$  sont des points de la grille grossière  $\Omega^{h_c}$ . C'est-à-dire il existe  $i_0, i_1, j_0, j_1$  entre 1 et  $N - 1$  vérifiant

$$\begin{cases} a_1 = x^{h_c}(i_0), & b_1 = x^{h_c}(i_1), \\ c_1 = y^{h_c}(j_0), & d_1 = y^{h_c}(j_1). \end{cases} \quad (4.86)$$

$i_0, i_1, j_0, j_1$  sont les indices des points  $a_1, b_1, c_1, d_1$  dans la grille grossière. On veut que la grille grossière et la grille fine possèdent le même nombre de points, pour cela on choisit  $a_1, b_1, c_1, d_1$  vérifiant

$$\begin{cases} b_1 - a_1 = \frac{b - a}{2}, \\ d_1 - c_1 = \frac{b - a}{2}. \end{cases} \quad (4.87)$$

L'interface  $\partial\Omega_l$  entre le domaine  $\Omega_l$  et  $\Omega$  est le carré donné par

$$\partial\Omega_l = \underbrace{[a_1, b_1] \times \{c_1\}}_{\partial\Omega_{l,B}} \cup \underbrace{[a_1, b_1] \times \{d_1\}}_{\partial\Omega_{l,T}} \cup \underbrace{\{a_1\} \times [c_1, d_1]}_{\partial\Omega_{l,L}} \cup \underbrace{\{b_1\} \times [c_1, d_1]}_{\partial\Omega_{l,R}}. \quad (4.88)$$

On note  $u_l = u|_{\Omega_l}$  et  $f_l = f|_{\Omega_l}$ . Supposons pour l'instant que l'on connaisse les conditions de Dirichlet  $g_l$  sur  $\partial\Omega_l$ . On considère sur  $\Omega_l$  le problème de Poisson

$$\begin{cases} -\Delta u_l = f_l, & \text{dans } \Omega_l, \\ u_l = g_l, & \text{sur } \partial\Omega_l. \end{cases} \quad (4.89)$$

Le domaine carré  $\Omega_l$  est discrétisé par (3.1) avec pas de maillage  $h_f = \frac{h_c}{2}$ . On obtient les points  $x_i^{h_f} = a_1 + ih_f, 0 \leq i \leq N$  et  $y_j^{h_f} = a_1 + jh_f, 0 \leq j \leq N$ . Le domaine discret local est  $\Omega_l^{h_f}$ . Les vecteurs inconnus discrets sont  $U_l^{h_f}, U_{l,x}^{h_f}, U_{l,y}^{h_f} \in \mathbb{R}^{(N-1)^2}$ . D'après ce qu'on a vu au chapitre 3 dans la proposition 3.5, le schéma boîte hermitien associé au problème (4.89) possède la forme matricielle suivante

$$\frac{1}{h_f^2} (\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H}) U_l^{h_f} = F_l - G_x^{h_f} - G_y^{h_f}, \quad (4.90)$$

avec  $F_l, G_x^{h_f}, G_y^{h_f} \in \mathbb{R}^{(N-1)^2}$  donnés en (3.87), (3.88) et (3.89). En general, la condition limite  $g_l$  n'est pas connue. De la même façon qu'en (4.76), on approche  $g_l$  à l'aide de  $U^{h_c}|_{\Gamma}$ , la restriction de la solution grossière sur  $\Gamma$ .

Ici, en deux dimensions, pour avoir des données pour les points de la frontière artificielle

on utilise l'opérateur d'interpolation spline en deux dimensions  $P^{h_f, h_c}$  décrit à la section 4.5.5. Les vecteurs de données au bord artificielles sont notés  $U_{\Gamma_L}^{h_f}, U_{\Gamma_R}^{h_f}, U_{\Gamma_B}^{h_f}, U_{\Gamma_T}^{h_f} \in \mathbb{R}^{(N-1)}$ , où  $\Gamma_L, \Gamma_R, \Gamma_B, \Gamma_T$  représentent les quatres arêtes du  $\Gamma$ . Les données sur les quatres sommets de  $\Gamma$  sont notées par  $U_{\Gamma_{LB}}^{h_f}, U_{\Gamma_{LT}}^{h_f}, U_{\Gamma_{RB}}^{h_f}, U_{\Gamma_{RT}}^{h_f}$ . On obtient les vecteurs  $G_x^{h_f}, G_y^{h_f}$ . En comparant (4.90) avec (4.62) on obtient

$$\begin{cases} L^{h_f} = \frac{1}{h_f^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H}), \\ f_{l,\Gamma}^{h_f} - B_{l,\Gamma}^{h_f} u^{h_f}|_{\Gamma} = F_l - G_x^{h_f} - G_y^{h_f}. \end{cases} \quad (4.91)$$

La résolution de (4.90) donne le vecteur solution initial  $U_0^{h_f} \in \mathbb{R}^{(N-1)^2}$  sur la grille fine locale  $\Omega_l^{h_f}$ . Le “safety region” est  $\Omega_\epsilon = (a_1 + \epsilon, b_1 - \epsilon) \times (c_1 + \epsilon, d_1 - \epsilon)$  avec  $\epsilon > 0$ . D'après (4.65) le vecteur  $W^{h_c} \in \mathbb{R}^{(N-1)^2}$  vérifie

$$W^{h_c} = \begin{cases} U_l^{h_c}, & \text{dans } \Omega_l^{h_c}, \\ U^{h_c}, & \text{dans } \Omega^{h_c} \setminus \Omega_l^{h_c}. \end{cases} \quad (4.92)$$

Le vecteur de défaut  $d^{h_c} \in \mathbb{R}^{(N-1)^2}$  est calculé par

$$d^{h_c} = \frac{1}{h_c^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})W^{h_c} - f^{h_c}. \quad (4.93)$$

#### 4.5.5 Interpolation de bord par spline cubique

Dans cette partie, on détaille la construction des vecteurs de conditions au bord artificielles sur  $\Gamma = \partial\Omega_l \setminus (\partial\Omega \cap \partial\Omega_l)$  en utilisant une interpolation par spline cubique des valeurs de  $\Gamma^{h_c} = \Gamma \cap \Omega^{h_c}$ .

##### • Fonctions de grille sur la grille grossière

Au vecteur  $U^{h_c} \in \mathbb{R}^{(N-1)^2}$  calculé sur la grille grossière correspond la fonction de grille  $u^{h_c} = (u_{i,j}^{h_c})_{0 \leq i,j \leq N}$  dans  $L_{h_c}^2$ . Numériquement ceci consiste à construire une matrice  $u^{h_c} \in \mathbb{M}_{N+1}(\mathbb{R})$  à partir du vecteur  $U^{h_c} \in \mathbb{R}^{(N-1)^2}$  et des données au bord  $\partial\Omega^{h_c}$ .

$$(U^{h_c}, U_L^{h_c}, U_R^{h_c}, U_B^{h_c}, U_T^{h_c}, U_{LB}^{h_c}, U_{LT}^{h_c}, U_{RB}^{h_c}, U_{RT}^{h_c}) \longrightarrow u^{h_c}, \quad (4.94)$$

telle que

$$u_{i,j}^{h_c} = U_{(N-1)(i-1)+j}^{h_c}, \quad 1 \leq i, j \leq N-1. \quad (4.95)$$

Les valeurs de  $u^{h_c}$  sur les quatres arêtes de  $\partial\Omega^{h_c}$  sont

$$\begin{cases} u^{h_c}(x_0^{h_c}, y_j^{h_c}) = U_{L,j}^{h_c}, & 1 \leq j \leq N-1, \\ u^{h_c}(x_N^{h_c}, y_j^{h_c}) = U_{R,j}^{h_c}, & 1 \leq j \leq N-1, \\ u^{h_c}(x_i^{h_c}, y_0^{h_c}) = U_{B,i}^{h_c}, & 1 \leq i \leq N-1, \\ u^{h_c}(x_i^{h_c}, y_N^{h_c}) = U_{T,i}^{h_c}, & 1 \leq i \leq N-1. \end{cases} \quad (4.96)$$

De plus, les valeurs de  $u^{h_c}$  aux quatres sommets de  $\partial\Omega^{h_c}$  sont

$$\begin{cases} u^{h_c}(x_0^{h_c}, y_0^{h_c}) = U_{LB}^{h_c}, \\ u^{h_c}(x_0^{h_c}, y_N^{h_c}) = U_{LT}^{h_c}, \\ u^{h_c}(x_N^{h_c}, y_0^{h_c}) = U_{RB}^{h_c}, \\ u^{h_c}(x_N^{h_c}, y_N^{h_c}) = U_{RT}^{h_c}. \end{cases} \quad (4.97)$$

On procède de la même manière pour les vecteurs  $U_x^{h_c}, U_y^{h_c} \in \mathbb{R}^{(N-1)^2}$ . On obtient ainsi les fonctions de grille  $u_x^{h_c}, u_y^{h_c}$  avec  $u_{x,i,j}^{h_c} \approx \partial_x u(x_i^{h_c}, y_j^{h_c})$  et  $u_{y,i,j}^{h_c} \approx \partial_y u(x_i^{h_c}, y_j^{h_c})$ .

• *Fonctions de grille sur la grille fine*

Au vecteur  $U_l^{h_f} \in \mathbb{R}^{(N-1)^2}$  calculé sur la grille fine correspond la fonction de grille  $u_l^{h_f} = (u_{l,i,j}^{h_f})_{0 \leq i,j \leq N} \in L_{h_f}^2$ . On a

$$u_{i,j}^{h_f} = U_{(N-1)(i-1)+j}^{h_f}, \quad 1 \leq i, j \leq N-1. \quad (4.98)$$

Les valeurs de  $u^{h_f}$  sur les quatres arêtes de  $\partial\Omega_l^{h_f}$  sont

$$\begin{cases} u_l^{h_f}(x_0^{h_f}, y_j^{h_f}) = U_{l,L,j}^{h_f}, & 1 \leq j \leq N-1, \\ u_l^{h_f}(x_N^{h_f}, y_j^{h_f}) = U_{l,R,j}^{h_f}, & 1 \leq j \leq N-1, \\ u_l^{h_f}(x_i^{h_f}, y_0^{h_f}) = U_{l,B,i}^{h_f}, & 1 \leq i \leq N-1, \\ u_l^{h_f}(x_i^{h_f}, y_N^{h_f}) = U_{l,T,i}^{h_f}, & 1 \leq i \leq N-1. \end{cases} \quad (4.99)$$

De plus, les valeurs de  $u^{h_f}$  aux quatres sommets de  $\partial\Omega_l^{h_f}$  sont

$$\begin{cases} u_l^{h_f}(x_0^{h_f}, y_0^{h_f}) = U_{l,LB}^{h_f}, \\ u_l^{h_f}(x_0^{h_f}, y_N^{h_f}) = U_{l,LT}^{h_f}, \\ u_l^{h_f}(x_N^{h_f}, y_0^{h_f}) = U_{l,RB}^{h_f}, \\ u_l^{h_f}(x_N^{h_f}, y_N^{h_f}) = U_{l,RT}^{h_f}. \end{cases} \quad (4.100)$$

De la même manière on obtient à partir des vecteurs  $U_x^{h_f}$  et  $U_y^{h_f}$  les deux fonctions de grille  $u_{l,x}^{h_f}$  et  $u_{l,y}^{h_f}$  avec  $u_{x,i,j}^{h_f} \approx \partial_x u(x_i^{h_f}, y_j^{h_f})$  et  $u_{y,i,j}^{h_f} \approx \partial_y u(x_i^{h_f}, y_j^{h_f})$ .

• *Interpolation spline*

Il est plus efficace de réaliser une interpolation polynômiale par morceaux, qu'une interpolation globale. Soit  $s(x)$ ,  $a \leq x \leq b$ , un polynôme défini par morceaux par

$$s(x_j) = u(x_j), \quad j = 0, 1, \dots, N, \quad (4.101)$$

où les valeurs  $x_0 = a < x_1 < \dots < x_N = b$ . Pour  $\bar{x}$ ,  $s(\bar{x})$  est déterminé par quelques valeurs  $(x_i, u(x_i))$  avec  $x_i$  proche de  $\bar{x}$ .

**Spline cubique :**  $s(x)$  est un polynôme cubique par morceaux.

$$s(x) = s_j(x)|_{[x_j, x_{j+1}]}, \quad j = 0, 1, \dots, N-1. \quad (4.102)$$

Sur chaque intervalle, on a les relations de collocation

$$s_j(x_j) = u(x_j), \quad s_j(x_{j+1}) = u(x_{j+1}). \quad (4.103)$$

Il faut deux autres conditions pour déterminer un polynôme cubique

$$s(x) = a_0 + a_1x + a_2x^2 + a_3x^3. \quad (4.104)$$

On interpole de plus les dérivées en  $(x_j, u'(x_j)), (x_{j+1}, u'(x_{j+1}))$ . On réalise donc une interpolation de type Hermite. Les valeurs  $u'(x_j)$  peuvent être remplacées par des valeurs approchées de la dérivée, évaluées par ailleurs. On a

$$s_j(x) = u(x_j) + u'(x_j)(x - x_j) + c_2(x - x_j)^2 + c_3(x - x_j)^3. \quad (4.105)$$

Le système linéaire de deux équations à deux inconnues

$$\begin{cases} s_j(x_{j+1}) = u(x_{j+1}), \\ s'_j(x_{j+1}) = u'(x_{j+1}), \end{cases} \quad (4.106)$$

donne la solution

$$\begin{cases} c_2 = \frac{3(u(x_{j+1}) - u(x_j))}{(x_{j+1} - x_j)^2} - \frac{2u'(x_j) + u'(x_{j+1})}{x_{j+1} - x_j}, \\ c_3 = \frac{2(u(x_j) - u(x_{j+1}))}{(x_{j+1} - x_j)^3} + \frac{u'(x_j) + u'(x_{j+1})}{(x_{j+1} - x_j)^2}. \end{cases} \quad (4.107)$$

Soit  $u(x, y) \in C^1((a, b)^2)$ . On déduit de (4.105) les polynômes d'interpolation  $S_{x_i}^j$ ,  $0 \leq i \leq N-1$ ,  $0 \leq j \leq N$ , et  $S_{y_j}^i$ ,  $0 \leq i \leq N$ ,  $0 \leq j \leq N-1$ , définis par

$$\begin{cases} S_{x_i}^j(x) = S_x^j(x_i; x) = u(x_i, y_j) + \partial_x u(x_i, y_j)(x - x_i) + c_2^j(x - x_i)^2 + c_3^j(x - x_i)^3, & x \in [x_i, x_{i+1}], \\ S_{y_j}^i(y) = S_y^i(y_j; y) = u(x_i, y_j) + \partial_y u(x_i, y_j)(y - y_j) + d_2^i(y - y_j)^2 + d_3^i(y - y_j)^3, & y \in [y_j, y_{j+1}]. \end{cases} \quad (4.108)$$

Les coefficients  $c_2^j$ ,  $c_3^j$ ,  $d_2^i$ ,  $d_3^i$  sont déduites de (4.107),

$$\begin{cases} c_2^j = \frac{3(u(x_{i+1}, y_j) - u(x_i, y_j))}{(x_{i+1} - x_i)^2} - \frac{2\partial_x u(x_i, y_j) + \partial_x u(x_{i+1}, y_j)}{x_{i+1} - x_i}, \\ c_3^j = \frac{2(u(x_i, y_j) - u(x_{i+1}, y_j))}{(x_{i+1} - x_i)^3} + \frac{\partial_x u(x_i, y_j) + \partial_x u(x_{i+1}, y_j)}{(x_{i+1} - x_i)^2}, \end{cases} \quad (4.109)$$

et

$$\begin{cases} d_2^i = \frac{3(u(x_i, y_{j+1}) - u(x_i, y_j))}{(y_{j+1} - y_j)^2} - \frac{2\partial_y u(x_i, y_j) + \partial_y u(x_i, y_{j+1})}{y_{j+1} - y_j}, \\ d_3^i = \frac{2(u(x_i, y_j) - u(x_i, y_{j+1}))}{(y_{j+1} - y_j)^3} + \frac{\partial_y u(x_i, y_j) + \partial_y u(x_i, y_{j+1})}{(y_{j+1} - y_j)^2}. \end{cases} \quad (4.110)$$

Sous forme discrète sur la grille grossière  $\Omega^{h_c}$ , on remplace les fonctions exactes  $u$ ,  $\partial_x u$ ,  $\partial_y u$  par les fonctions de grille  $u^{h_c}$ ,  $u_x^{h_c}$ ,  $u_y^{h_c}$  respectivement.

Le but de cette interpolation est d'évaluer les vecteurs des conditions au bord  $U_{\Gamma_L}^{h_f}$ ,  $U_{\Gamma_R}^{h_f}$ ,  $U_{\Gamma_B}^{h_f}$ ,  $U_{\Gamma_T}^{h_f} \in \mathbb{R}^{(N-1)}$  ainsi que les quatres valeurs  $U_{\Gamma_{LB}}^{h_f}$ ,  $U_{\Gamma_{LT}}^{h_f}$ ,  $U_{\Gamma_{RB}}^{h_f}$ ,  $U_{\Gamma_{RT}}^{h_f}$ . Par exemple, on évalue

le vecteur  $U_{\Gamma_B}^{h_f} = (U_{\Gamma_B,i}^{h_f})_{1 \leq i \leq N-1} \in \mathbb{R}^{(N-1)}$  par

$$U_{\Gamma_B,i}^{h_f} = u_l^{h_f}(x_i^{h_f}, c_1) = \begin{cases} u^{h_c}(x_i^{h_f}, c_1), & \text{si } x_i^{h_f} \in \Omega_l^{h_c}, \\ S^{j_0}(x_{i-1}^{h_f}; x_i^{h_f}), & \text{sinon,} \end{cases} \quad (4.111)$$

avec  $u_l^{h_f}$  la fonction de grille fine correspondante.

#### 4.5.6 Mise en oeuvre de l'algorithme LDC

Dans cette partie, on décrit un exemple de mise en oeuvre de l'algorithme précédent.

On utilise :

- 1- Domaine de raffinement choisi a priori.
- 2- Couplage avec HB2, ce qui signifie que les matrices  $\mathcal{H}$  et  $\mathcal{G}$  sont remplacées par les matrices  $\mathcal{H}_2$  et  $\mathcal{G}_2$  en (2.127). L'algorithme est le suivant :

##### Algorithme 9. •Initialisation

- Résoudre le problème (4.84) sur la grille grossière  $\Omega^{h_c}$ . Le vecteur solution est noté  $U_0^{h_c} \in \mathbb{R}^{(N-1)^2}$ .

- Résoudre le problème (4.90) sur la grille fine  $\Omega_l^{h_f}$ . Le vecteur solution est noté  $U_{l,0}^{h_f} \in \mathbb{R}^{(N-1)^2}$ .

• Itération  $i = 1, 2, \dots$

- Calculer le vecteur  $W_{i-1}^{h_c} \in \mathbb{R}^{(N-1)^2}$  par (4.92).

- Calculer le vecteur  $d_{i-1}^{h_c} \in \mathbb{R}^{(N-1)^2}$  par (4.93).

- Résoudre sur la grille grossière

$$\frac{1}{h_c^2}(\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})U_i^{h_c} = \begin{cases} f^{h_c} + d_{i-1}^{h_c}, & \text{dans } \Omega_\epsilon^{h_c}, \\ f^{h_c}, & \text{dans } \Omega^{h_c} \setminus \Omega_\epsilon^{h_c}, \end{cases} \quad (4.112)$$

avec  $\epsilon > 0$ . Le vecteur solution est noté  $U_i^{h_c} \in \mathbb{R}^{(N-1)^2}$ .

- Résoudre le problème (4.90) sur la grille fine. Le vecteur solution est noté  $U_{l,i}^{h_f} \in \mathbb{R}^{(N-1)^2}$ .

#### 4.5.7 Résultats numériques

Dans cette section, on présente des résultats numériques préliminaires obtenus par le schéma HB2 pour le problème de Poisson sur le domaine  $\Omega = (0, 1)^2$ . On choisit

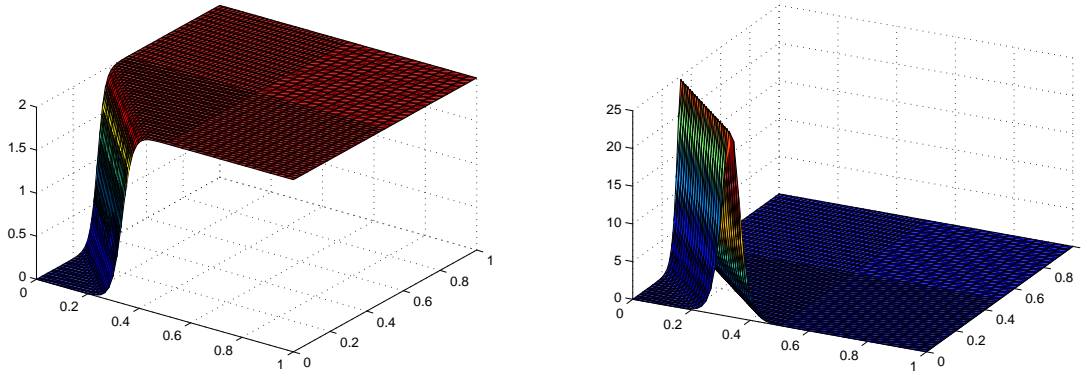


FIG. 4.6: La solution exacte du test 1 (à gauche) et sa dérivée (à droite).

une grille grossière uniforme  $\Omega^{h_c}$  dans  $\Omega$  avec pas de discrétisation  $h_c$ . On choisit une grille uniforme  $\Omega_l^{h_f}$  dans  $\Omega_l$  avec pas de discrétisation  $h_f$ . Le “safety region” est tel que  $\Omega_\epsilon = (\epsilon, \frac{1}{2} - \epsilon)^2$  avec  $\epsilon = 1/16$ . L’opérateur  $P^{h_f, h_c}$  est l’interpolation cubique par spline. La raison de ce choix est que l’interpolation cubique par spline préserve la précision d’ordre 4 du schéma.

Les erreurs sont estimées en utilisant les normes  $L^2$  et  $L^\infty$  discrètes

$$\begin{cases} \|u_{ex} - u\|_h = \left( h^2 \sum_{i,j} (u_{ex}(x_i, y_j) - u_{i,j})^2 \right)^{\frac{1}{2}}, \\ \|u_{ex} - u\|_\infty = \max_{i,j=1,\dots,N-1} |u_{ex}(x_i, y_j) - u_{i,j}|. \end{cases} \quad (4.113)$$

**Test 1 :** Le terme source  $f$  et la condition de bord  $g$  sont choisis tel que la solution exacte soit

$$u_{ex}(x, y) = \tanh[25(x + y - \frac{1}{3})] + 1. \quad (4.114)$$

Cette fonction possède un fort gradient dans la région  $\Omega_l = (0, \frac{1}{2})^2$  (voir figure 4.6). Les résultats numériques sont présentés dans les tables 4.10 et 4.11.

**Test 2 :** On considère la solution exacte

$$u_{ex}(x, y) = \exp\left(-400\left(\left(x - \frac{1}{4}\right)^2 - \left(y - \frac{1}{4}\right)^2\right)\right). \quad (4.115)$$

Cette fonction possède un fort gradient dans la région  $\Omega_l = ]0, \frac{1}{2}[ \times ]0, \frac{1}{2}[$  (voir figure 4.7). Les résultats numériques sont donnés dans les tables 4.12 et 4.13. Dans les deux tests précédents, on obtient avec le calcul direct sur une grille  $N = 512$  la même précision



$h_f$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	nbre de points	CPU(s.)
1/64	2.236(-4)	9.100(-3)	2.600(-3)	1.406(-1)	4096	0.50
T. C.	4.37	4.13	4.03	3.01		
1/128	1.018(-5)	5.180(-4)	1.558(-4)	1.740(-2)	16384	0.80
T. C.	4.11	3.97	3.64	3.63		
1/256	6.255(-7)	3.302(-5)	1.271(-5)	1.400(-3)	65536	1.97
T. C.	4.01	3.98	3.85	3.85		
1/512	3.872(-8)	2.081(-6)	8.793(-7)	9.694(-5)	262144	7.80

TAB. 4.10: Erreur et taux de convergence du test 1 sur une grille uniforme

$h_c$	$\ u_{ex} - u^{h_f, h_c}\ _h$	$\ u_{x,ex} - u_x^{h_f, h_c}\ _h$	$\ u_{ex} - u^{h_f, h_c}\ _\infty$	$\ u_{x,ex} - u_x^{h_f, h_c}\ _\infty$	n. points	CPU(s.)
1/64	1.053(-5)	5.692(-4)	1.584(-4)	1.740(-2)	7168	0.66
T. C.	4.07	4.10	3.64	3.63		
1/128	6.265(-7)	3.304(-5)	1.271(-5)	1.400(-3)	28672	1.62
T. C.	4.01	3.98	3.85	3.85		
1/256	3.879(-8)	2.084(-6)	8.794(-7)	9.693(-5)	114688	5.05
T. C.	4.00	3.99	3.93	3.92		
1/512	2.412(-9)	1.308(-7)	5.768(-8)	6376(-6)	458752	21

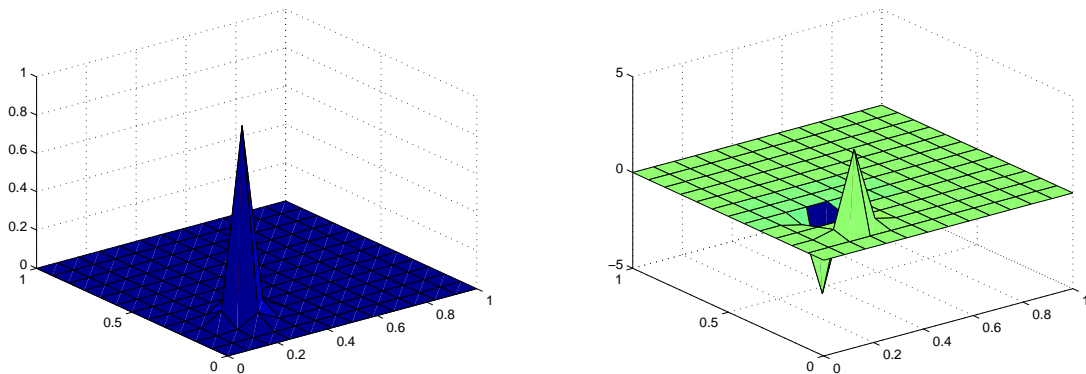
TAB. 4.11: Erreur et taux de convergence du test 1 sur une grille composée avec  $\epsilon = 1/16$ 

FIG. 4.7: A gauche la solution exacte du test 2 et à droite sa dérivée.

$h_f$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	nbre de points	CPU(s.)
1/128	5.674(-6)	1.318(-4)	1.366(-4)	2.4(-3)	16384	0.76
T. C.	4.02	4.03	4.03	4.03		
1/256	3.474(-7)	8.038(-6)	8.342(-6)	1.467(-4)	65536	1.81
T. C.	4.00	4.00	4.00	4.00		
1/512	2.160(-8)	4.993(-7)	5.183(-7)	9.113(-6)	262144	6.39

TAB. 4.12: Erreur et taux de convergence du test 2 sur une grille uniforme

$h_c$	$\ u_{ex} - u^{h_f, h_c}\ _h$	$\ u_{x,ex} - u_x^{h_f, h_c}\ _h$	$\ u_{ex} - u^{h_f, h_c}\ _\infty$	$\ u_{x,ex} - u_x^{h_f, h_c}\ _\infty$	n. points	CPU(s.)
1/64	5.674(-6)	1.318(-4)	1.366(-4)	2.4(-3)	7168	0.68
T. C.	4.02	4.03	4.03	4.03		
1/128	3.474(-7)	8.038(-6)	8.342(-6)	1.467(-4)	28672	1.67
T. C.	4.00	4.00	4.00	4.00		
1/256	2.160(-8)	4.993(-7)	5.183(-7)	9.113(-6)	114688	5.04
T. C.	4.00	4.00	4.00	4.00		
1/512	1.348(-9)	3.116(-8)	3.235(-8)	5.686(-7)	458752	19.7

TAB. 4.13: Erreur et taux de convergence du test 2 sur une grille composée avec  $\epsilon = 1/16$ 

qu'une grille composée de taille  $N = 258$ .

## Chapitre 5

# Schémas boîte hermitiens en dimension 3

### 5.1 Introduction

Le but de ce chapitre est de présenter un nouveau schéma boîte hermitien en trois dimensions. Les schémas aux différences en dimension 3 constituent un enjeu très important pour les applications. Ici, on se limite au problème de Poisson avec conditions aux limites de Dirichlet comme exemple prototype. Le principe du schéma présenté est une généralisation du schéma en deux dimensions. On se concentre ici sur la façon de généraliser le schéma  $\mathbb{R}^2$  en  $\mathbb{R}^3$ , en particulier sur les aspects algébriques.

### 5.2 Opérateurs aux différences en trois dimensions

On considère le maillage du cube  $\Omega = (a, b)^3$  par la grille de différences finies régulière

$$\begin{cases} x_i = a + ih, & 0 \leq i \leq N, \\ y_j = a + jh, & 0 \leq j \leq N, \\ z_k = a + kh, & 0 \leq k \leq N. \end{cases} \quad (5.1)$$

Le pas du maillage est  $h = \frac{b-a}{N}$ . Les inconnues discrètes sont localisées aux points  $(x_i, y_j, z_k)_{0 \leq i, j, k \leq N}$ . Les points intérieurs sont les points  $(x_i, y_j, z_k)$  avec  $1 \leq i, j, k \leq N-1$  et les points frontières sont les points  $(x_i, y_j, z_k)$  avec  $i \in \{0, N\}$  ou  $j \in \{0, N\}$

ou  $k \in \{0, N\}$ . Toutes les notions introduites en deux dimensions au chapitre 3 se généralisent en trois dimensions. On précise en détail cette généralisation dans ce qui suit.

- *Fonctions de grille*

On note  $L_h^2$  l'espace des fonctions de grille  $(v_{i,j,k})_{0 \leq i,j,k \leq N}$  et par  $L_{h,0}^2$  le sous-espace de  $L_h^2$  telles que  $v_{i,j,k} = 0$  aux points frontières.

- Le produit scalaire et la norme sur  $L_h^2$  sont définis pour  $u, v \in L_h^2$  par

$$(u, v)_h = h^3 \sum_{i,j,k=0}^N u_{i,j,k} v_{i,j,k}, \quad |u|_h^2 = h^3 \sum_{i,j,k=0}^N u_{i,j,k}^2. \quad (5.2)$$

- *Dérivées centrées*

Les opérateurs aux dérivées centrées  $\delta_x, \delta_y, \delta_z$  sont définis par

$$\begin{cases} \delta_x u_{i,j,k} = \frac{u_{i+1,j,k} - u_{i-1,j,k}}{2h}, & 1 \leq i \leq N-1, \quad 0 \leq j, k \leq N, \\ \delta_y u_{i,j,k} = \frac{u_{i,j+1,k} - u_{i,j-1,k}}{2h}, & 1 \leq j \leq N-1, \quad 0 \leq i, k \leq N, \\ \delta_z u_{i,j,k} = \frac{u_{i,j,k+1} - u_{i,j,k-1}}{2h}, & 1 \leq k \leq N-1, \quad 0 \leq i, j \leq N. \end{cases} \quad (5.3)$$

On prolonge  $\delta_x, \delta_y, \delta_z$  à tous les points de bord par

$$\begin{cases} \delta_x u_{0,j,k} = \delta_x u_{N,j,k} = 0, & 0 \leq j, k \leq N, \\ \delta_y u_{i,0,k} = \delta_y u_{i,N,k} = 0, & 0 \leq i, k \leq N, \\ \delta_z u_{i,j,0} = \delta_z u_{i,j,N} = 0, & 0 \leq i, j \leq N. \end{cases} \quad (5.4)$$

- *Dérivées décentrées*

Les opérateurs aux dérivées décentrées  $\delta_x^+, \delta_x^-, \delta_y^+, \delta_y^-, \delta_z^+, \delta_z^-$  sont définis par

$$\begin{cases} \delta_x^+ u_{i,j,k} = \frac{u_{i+1,j,k} - u_{i,j,k}}{h}, & 0 \leq i \leq N-1, \quad 0 \leq j, k \leq N, & (a) \\ \delta_x^- u_{i,j,k} = \frac{u_{i,j,k} - u_{i-1,j,k}}{h}, & 1 \leq i \leq N, \quad 0 \leq j, k \leq N, & (b) \\ \delta_y^+ u_{i,j,k} = \frac{u_{i,j+1,k} - u_{i,j,k}}{h}, & 0 \leq j \leq N-1, \quad 0 \leq i, k \leq N, & (c) \\ \delta_y^- u_{i,j,k} = \frac{u_{i,j-1,k} - u_{i,j,k}}{h}, & 1 \leq j \leq N, \quad 0 \leq i, k \leq N, & (d) \\ \delta_z^+ u_{i,j,k} = \frac{u_{i,j,k+1} - u_{i,j,k}}{h}, & 0 \leq k \leq N-1, \quad 0 \leq i, j \leq N, & (e) \\ \delta_z^- u_{i,j,k} = \frac{u_{i,j,k-1} - u_{i,j,k}}{h}, & 1 \leq k \leq N, \quad 0 \leq i, j \leq N. & (f) \end{cases} \quad (5.5)$$

On prolonge  $\delta_x^+$ ,  $\delta_x^-$ ,  $\delta_y^+$ ,  $\delta_y^-$ ,  $\delta_z^+$ ,  $\delta_z^-$  à tous les points de bord par

$$\begin{cases} \delta_x^+ u_{N,j,k} = 0, & 0 \leq j, k \leq N, \\ \delta_x^- u_{0,j,k} = 0, & 0 \leq j, k \leq N, \\ \delta_y^+ u_{i,N,k} = 0, & 0 \leq i, k \leq N, \\ \delta_y^- u_{i,0,k} = 0, & 0 \leq i, k \leq N, \\ \delta_z^+ u_{i,j,N} = 0, & 0 \leq i, j \leq N, \\ \delta_z^- u_{i,k,0} = 0, & 0 \leq i, j \leq N. \end{cases} \quad (5.6)$$

• *Laplacien à trois points*

Les opérateurs aux dérivées secondes centrées  $\delta_x^2$ ,  $\delta_y^2$ ,  $\delta_z^2$  sont définis par

$$\begin{cases} \delta_x^2 u_{i,j,k} = \frac{u_{i+1,j,k} + u_{i-1,j,k} - 2u_{i,j,k}}{h^2}, & 0 \leq i \leq N, \quad 1 \leq j, k \leq N-1, \\ \delta_y^2 u_{i,j,k} = \frac{u_{i,j+1,k} + u_{i,j-1,k} - 2u_{i,j,k}}{h^2}, & 0 \leq j \leq N, \quad 1 \leq i, k \leq N-1, \\ \delta_z^2 u_{i,j,k} = \frac{u_{i,j,k+1} + u_{i,j,k-1} - 2u_{i,j,k}}{h^2}, & 0 \leq k \leq N, \quad 1 \leq i, j \leq N-1. \end{cases} \quad (5.7)$$

On prolonge  $\delta_x^2$ ,  $\delta_y^2$ ,  $\delta_z^2$  à tous les points de bord par

$$\begin{cases} \delta_x^2 u_{0,j,k} = \delta_x^2 u_{N,j,k} = 0, & 0 \leq j, k \leq N, \\ \delta_y^2 u_{i,0,k} = \delta_y^2 u_{i,N,k} = 0, & 0 \leq i, k \leq N, \\ \delta_z^2 u_{i,j,0} = \delta_z^2 u_{i,j,N} = 0, & 0 \leq i, j \leq N. \end{cases} \quad (5.8)$$

• *Opérateurs de Simpson*

Les opérateurs de Simpson  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_z$  sont définis par

$$\begin{cases} \sigma_x u_{i,j,k} = \frac{1}{6} u_{i-1,j,k} + \frac{2}{3} u_{i,j,k} + \frac{1}{6} u_{i+1,j,k}, & 1 \leq i \leq N-1, \quad 0 \leq j, k \leq N, \\ \sigma_y u_{i,j,k} = \frac{1}{6} u_{i,j-1,k} + \frac{2}{3} u_{i,j,k} + \frac{1}{6} u_{i,j+1,k}, & 1 \leq j \leq N-1, \quad 0 \leq i, k \leq N, \\ \sigma_z u_{i,j,k} = \frac{1}{6} u_{i,j,k-1} + \frac{2}{3} u_{i,j,k} + \frac{1}{6} u_{i,j,k+1}, & 1 \leq k \leq N-1, \quad 0 \leq i, j \leq N. \end{cases} \quad (5.9)$$

On prolonge  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_z$  à tous les points de bord par

$$\begin{cases} \sigma_x u_{0,j,k} = \sigma_x u_{N,j,k} = 0, & 0 \leq j, k \leq N, \\ \sigma_y u_{i,0,k} = \sigma_y u_{i,N,k} = 0, & 0 \leq i, k \leq N, \\ \sigma_z u_{i,j,0} = \sigma_z u_{i,j,N} = 0, & 0 \leq i, j \leq N. \end{cases} \quad (5.10)$$

### 5.3 Notations algébriques

Dans cette section, on donne les notations matricielles et vectorielles utiles dans la suite. Elles généralisent de façon naturelle les notations vues dans la section 3.3.

• *Opérateur  $\text{vec}_3$*

**Définition 5.** Soient  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$  une fonction de grille et  $(e_i)_{1 \leq i \leq N-1}$  la base canonique de  $\mathbb{R}^{(N-1)}$ . On définit l'opérateur  $\text{vec}_3$  par

$$\begin{aligned} \text{vec}_3 : L_h^2 &\longrightarrow \mathbb{R}^{(N-1)^3} \\ v &\longrightarrow V = \text{vec}_3(v) \end{aligned}$$

avec

$$\text{vec}_3(v) = \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i,j,k}. \quad (5.11)$$

Matriciellement, l'opérateur  $\text{vec}_3$  crée à partir d'une fonction de grille  $(v_{i,j,k})_{0 \leq i,j,k \leq N}$ , un vecteur de taille  $(N-1)^3$  formé des points intérieurs de cette fonction de grille (les points  $(v_{i,j,k})_{1 \leq i,j,k \leq N-1}$ ) en utilisant l'ordre lexicographique sur  $(i, j, k)$ .

**Exemple :** Soit  $\Omega = (0, 1)^3$ . On considère la grille de différences finies

$$\begin{cases} x_i = ih, & 0 \leq i \leq 3, \\ y_j = jh, & 0 \leq j \leq 3, \\ z_k = kh, & 0 \leq k \leq 3. \end{cases} \quad (5.12)$$

avec  $h = 1/N = 1/3$ . Soit  $f(x, y, z) = \cos(x + y^2 + 3z + 1)$ , on introduit la fonction de grille  $f^* = (f_{i,j,k}^*)$  avec  $f_{i,j,k}^* = \cos(x_i + y_j^2 + 3z_k + 1)$ ,  $i, j, k \in 0, 1, 2, 3$ . On suppose  $u = f^*$ . Alors on a

$$U = \text{vec}_3(u) = [\cos(x_1 + y_1^2 + 3z_1 + 1), \cos(x_1 + y_1^2 + 3z_2 + 1), \cos(x_1 + y_2^2 + 3z_1 + 1), \cos(x_1 + y_2^2 + 3z_2 + 1), \\ \cos(x_2 + y_1^2 + 3z_1 + 1), \cos(x_2 + y_1^2 + 3z_2 + 1), \cos(x_2 + y_2^2 + 3z_1 + 1), \cos(x_2 + y_2^2 + 3z_2 + 1)]^T.$$

**Définition 6.** Soit  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$ . On définit les vecteurs  $V_{i,:}, V_{:,j}, V_{:,k} \in \mathbb{R}^{(N-1)^2}$  par

$$\begin{cases} V_{i,:} = \text{vec}_2[v_{i,j,k}]_{1 \leq j,k \leq N-1} = \sum_{j,k=1}^{N-1} (e_j \otimes e_k) v_{i,j,k}, & (a) \\ V_{:,j} = \text{vec}_2[v_{i,j,k}]_{1 \leq i,k \leq N-1} = \sum_{i,k=1}^{N-1} (e_i \otimes e_k) v_{i,j,k}, & (b) \\ V_{:,k} = \text{vec}_2[v_{i,j,k}]_{1 \leq i,j \leq N-1} = \sum_{i,j=1}^{N-1} (e_i \otimes e_j) v_{i,j,k}, & (c) \end{cases} \quad (5.13)$$

où  $\text{vec}_2$  est l'opérateur donné en (3.12).

**Définition 7.** Soit  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$  une fonction de grille. On définit les vecteurs  $V_{i,j}, V_{i,:k}, V_{:,j,k} \in \mathbb{R}^{(N-1)}$  par

$$\begin{cases} V_{i,j} = \text{vec}_1[v_{i,j,k}]_{1 \leq k \leq N-1} = \sum_{k=1}^{N-1} e_k v_{i,j,k}, & (a) \\ V_{i,:k} = \text{vec}_1[v_{i,j,k}]_{1 \leq j \leq N-1} = \sum_{j=1}^{N-1} e_j v_{i,j,k}, & (b) \\ V_{:,j,k} = \text{vec}_1[v_{i,j,k}]_{1 \leq i \leq N-1} = \sum_{i=1}^{N-1} e_i v_{i,j,k}, & (c) \end{cases} \quad (5.14)$$

où  $\text{vec}_1$  est l'opérateur donné en (1.36).

**Lemme 5.3.1.** Soient  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$  une fonction de grille et  $V = \text{vec}_3(v)$  alors on a

$$\begin{cases} V = \sum_{j,k=1}^{N-1} (V_{:,j,k} \otimes e_j \otimes e_k), & (a) \\ V = \sum_{i,k=1}^{N-1} (e_i \otimes V_{i,:k} \otimes e_k), & (b) \\ V = \sum_{i,j=1}^{N-1} (e_i \otimes e_j \otimes V_{i,j,:}). & (c) \end{cases} \quad (5.15)$$

**Preuve.** On démontre par exemple (5.15)<sub>a</sub>. Par définition de l'opérateur  $\text{vec}_3$  (5.11) on a

$$\begin{aligned} V &= \text{vec}_3(v) \\ &= \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i,j,k}. \end{aligned}$$

On en déduit

$$\begin{aligned}
V &= \sum_{j,k=1}^{N-1} \sum_{i=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i,j,k} \\
&= \sum_{j,k=1}^{N-1} \left( \sum_{i=1}^{N-1} e_i v_{i,j,k} \otimes e_j \otimes e_k \right) \\
&= \sum_{j,k=1}^{N-1} (V_{:,j,k} \otimes e_j \otimes e_k),
\end{aligned}$$

d'où (5.15)<sub>a</sub>. De la même manière on démontre (5.15)<sub>b,c</sub>. ■

• *Notations vectorielles*

De la même façon qu'à la section 3.3, les lettres capitales désignent des vecteurs et les lettres minuscules désignent des fonctions de grille.

• On note  $U, U_x, U_y, U_z \in \mathbb{R}^{(N-1)^3}$  les vecteurs associés aux points intérieurs des fonctions de grille  $u, u_x, u_y, u_z$ , respectivement,

$$\begin{cases}
U = \text{vec}_3(u) = \text{vec}_3[u_{i,j,k}]_{1 \leq i,j,k \leq N-1}, \\
U_x = \text{vec}_3(u_x) = \text{vec}_3[u_{x,i,j,k}]_{1 \leq i,j,k \leq N-1}, \\
U_y = \text{vec}_3(u_y) = \text{vec}_3[u_{y,i,j,k}]_{1 \leq i,j,k \leq N-1}, \\
U_z = \text{vec}_3(u_z) = \text{vec}_3[u_{z,i,j,k}]_{1 \leq i,j,k \leq N-1}.
\end{cases} \quad (5.16)$$

• Soient  $U_L, U_R \in \mathbb{R}^{(N-1)^2}$  les vecteurs associés aux conditions limites de Dirichlet aux points intérieurs des faces gauche "L" et droite "R" du cube,

$$\begin{cases}
U_L = \text{vec}_2[u_{0,j,k}]_{1 \leq j,k \leq N-1} = [u_{0,1,1}, \dots, u_{0,1,N-1}, u_{0,2,1}, \dots, u_{0,2,N-1}, \dots, u_{0,N-1,1}, \dots, u_{0,N-1,N-1}]^T, \\
U_R = \text{vec}_2[u_{N,j,k}]_{1 \leq j,k \leq N-1} = [u_{N,1,1}, \dots, u_{N,1,N-1}, u_{N,2,1}, \dots, u_{N,2,N-1}, \dots, u_{N,N-1,1}, \dots, u_{N,N-1,N-1}]^T.
\end{cases} \quad (5.17)$$

• De même,  $U_B, U_T \in \mathbb{R}^{(N-1)^2}$  sont les vecteurs associés aux conditions limites de Dirichlet aux points intérieurs des faces inférieure "B" et supérieure "T" du cube

$$\begin{cases}
U_B = \text{vec}_2[u_{i,0,k}]_{1 \leq i,k \leq N-1} = [u_{1,0,1}, \dots, u_{1,0,N-1}, u_{2,0,1}, \dots, u_{2,0,N-1}, \dots, u_{N-1,0,1}, \dots, u_{N-1,0,N-1}]^T, \\
U_T = \text{vec}_2[u_{i,N,k}]_{1 \leq i,k \leq N-1} = [u_{1,N,1}, \dots, u_{1,N,N-1}, u_{2,N,1}, \dots, u_{2,N,N-1}, \dots, u_{N-1,N,1}, \dots, u_{N-1,N,N-1}]^T.
\end{cases} \quad (5.18)$$



- De plus  $U_F, U_C \in \mathbb{R}^{(N-1)^2}$  sont les vecteurs associés aux conditions limites de Dirichlet à l'intérieur des faces arrière "F" et avant "C" du cube (figure 5.1),

$$\begin{cases} U_F = \text{vec}_2[u_{i,j,0}]_{1 \leq i,j \leq N-1} = [u_{1,1,0}, \dots, u_{1,N-1,0}, u_{2,1,0}, \dots, u_{2,N-1,0}, \dots, u_{N-1,1,0}, \dots, u_{N-1,N-1,0}]^T. \\ U_C = \text{vec}_2[u_{i,j,N}]_{1 \leq i,j \leq N-1} = [u_{1,1,N}, \dots, u_{1,N-1,N}, u_{2,1,N}, \dots, u_{2,N-1,N}, \dots, u_{N-1,1,N}, \dots, u_{N-1,N-1,N}]^T. \end{cases} \quad (5.19)$$

Les vecteurs associés aux valeurs de Dirichlet aux points "intérieurs" des douzes arêtes

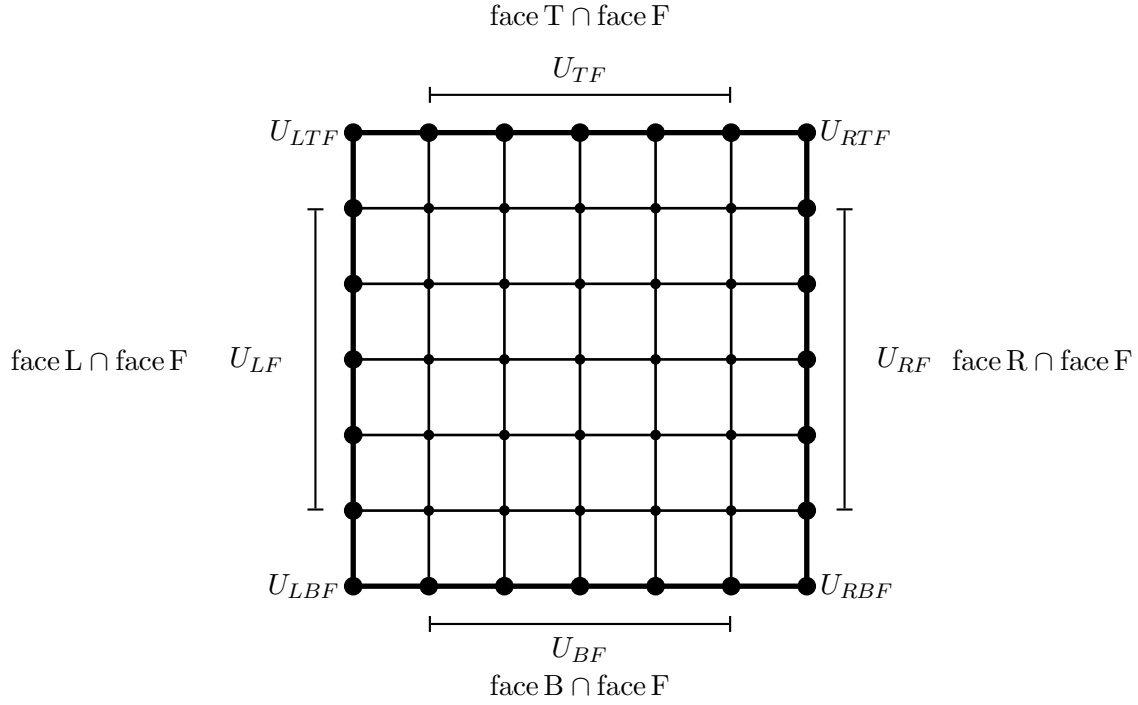


FIG. 5.1: Notations correspondent à la face "arrière" ou "F" du cube  $\Omega$ . Le vecteur  $U_F$  est formé de petites boules. Les grands boules désignent l'intersection entre cette face et les autres faces du  $\Omega$ .

du cube sont notés

$$U_{LB}, U_{RB}, U_{LT}, U_{RT}, U_{LF}, U_{RF}, U_{LC}, U_{RC}, U_{BF}, U_{TF}, U_{BC}, U_{TC} \in \mathbb{R}^{(N-1)}. \quad (5.20)$$

La figure 5.1 illustre graphiquement ces notations. Par exemple, on a

$$\begin{cases} U_{LB} = \text{vec}_1[u_{0,0,k}]_{1 \leq k \leq N-1} = [u_{0,0,1}, \dots, u_{0,0,N-1}]^T, \\ U_{LF} = \text{vec}_1[u_{0,j,0}]_{1 \leq j \leq N-1} = [u_{0,1,0}, \dots, u_{0,N-1,0}]^T. \end{cases} \quad (5.21)$$

- Les valeurs de Dirichlet sur les huit sommets du cube sont (figure 5.1) notées

$$\begin{cases} U_{LBF} = u_{0,0,0}, U_{LBC} = u_{0,0,N}, U_{LTF} = u_{0,N,0}, U_{LTC} = u_{0,N,N}, \\ U_{RBF} = u_{N,0,0}, U_{RBC} = u_{N,0,N}, U_{RTF} = u_{N,N,0}, U_{RTC} = u_{N,N,N}. \end{cases} \quad (5.22)$$

## 5.4 Fonctions de grille et algèbre de Kronecker

Dans cette section, on introduit des relations algébriques utiles dans la suite. Les lemmes suivants sont utilisés pour établir la forme matricielle du schéma boîte hermitien en dimension 3 dans la section 5.6. Dans la suite,  $e_0$  et  $e_N$  sont deux vecteurs nuls de  $\mathbb{R}^{N-1}$ .

**Lemme 5.4.1.** *Pour toute fonction de grille  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$  on a*

$$\begin{cases} \text{vec}_3(\delta_x v) = \frac{1}{2h}(K \otimes I \otimes I)\text{vec}_3(v) - \frac{1}{2h}(e_1 \otimes I \otimes I)V_L + \frac{1}{2h}(e_{N-1} \otimes I \otimes I)V_R, & (a) \\ \text{vec}_3(\delta_y v) = \frac{1}{2h}(I \otimes K \otimes I)\text{vec}_3(v) - \frac{1}{2h}(I \otimes e_1 \otimes I)V_B + \frac{1}{2h}(I \otimes e_{N-1} \otimes I)V_T, & (b) \\ \text{vec}_3(\delta_z v) = \frac{1}{2h}(I \otimes I \otimes K)\text{vec}_3(v) - \frac{1}{2h}(I \otimes I \otimes e_1)V_F + \frac{1}{2h}(I \otimes I \otimes e_{N-1})V_C. & (c) \end{cases} \quad (5.23)$$

**Preuve** Par définition de l'opérateur  $\text{vec}_3$  (5.11) on a

$$\text{vec}_3(\delta_x v) = \sum_{i,j,k=1}^{N-1} (e_1 \otimes e_j \otimes e_k)(\delta_x v_{i,j,k})_{1 \leq i,j,k \leq N-1}.$$

Donc

$$\begin{aligned} \text{vec}_3(\delta_x v) &= \frac{1}{2h} \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k)(v_{i+1,j,k} - v_{i-1,j,k}) \\ &= \frac{1}{2h} \left( \sum_{j,k=1}^{N-1} (e_{N-1} \otimes e_j \otimes e_k)v_{N,j,k} + \sum_{i=1}^{N-2} \sum_{j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k)v_{i+1,j,k} \right. \\ &\quad \left. - \sum_{i=2}^{N-1} \sum_{j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k)v_{i-1,j,k} - \sum_{j,k=1}^{N-1} (e_1 \otimes e_j \otimes e_k)v_{0,j,k} \right) \\ &= \frac{1}{2h} \left( \sum_{i,j,k=1}^{N-1} ((e_{i-1} - e_{i+1}) \otimes e_j \otimes e_k)v_{i,j,k} \right) - \frac{1}{2h}(e_1 \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k)v_{0,j,k} \\ &\quad + \frac{1}{2h}(e_{N-1} \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k)v_{N,j,k}. \end{aligned}$$

Or  $e_{i-1} - e_{i+1} = Ke_i$ ,  $1 \leq i \leq N-1$ , donc

$$\begin{aligned}
\text{vec}_3(\delta_x v) &= \frac{1}{2h} \left( \sum_{i,j,k=1}^{N-1} (Ke_i \otimes e_j \otimes e_k) v_{i,j,k} \right) - \frac{1}{2h} (e_1 \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k) v_{0,j,k} \\
&+ \frac{1}{2h} (e_{N-1} \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k) v_{N,j,k} \\
&= \frac{1}{2h} \left( (K \otimes I \otimes I) \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i,j,k} \right) - \frac{1}{2h} (e_1 \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k) v_{0,j,k} \\
&+ \frac{1}{2h} (e_{N-1} \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k) v_{N,j,k} \\
&= \frac{1}{2h} (K \otimes I \otimes I) \text{vec}_3(v) - \frac{1}{2h} (e_1 \otimes I \otimes I) V_L + \frac{1}{2h} (e_{N-1} \otimes I \otimes I) V_R,
\end{aligned}$$

d'où (5.23)<sub>a</sub>. De la même manière on démontre (5.23)<sub>b,c</sub>. ■

**Lemme 5.4.2.** Pour toute fonction de grille  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$ , on a

$$(i) \left\{ \begin{array}{l} \text{vec}_2(\delta_x v)_B = \frac{1}{2h} (K \otimes I) V_B - \frac{1}{2h} (e_1 \otimes I) V_{LB} + \frac{1}{2h} (e_{N-1} \otimes I) V_{RB}, \quad (a) \\ \text{vec}_2(\delta_x v)_T = \frac{1}{2h} (K \otimes I) V_T - \frac{1}{2h} (e_1 \otimes I) V_{LT} + \frac{1}{2h} (e_{N-1} \otimes I) V_{RT}, \quad (b) \\ \text{vec}_2(\delta_x v)_F = \frac{1}{2h} (K \otimes I) V_F - \frac{1}{2h} (e_1 \otimes I) V_{LF} + \frac{1}{2h} (e_{N-1} \otimes I) V_{RF}, \quad (c) \\ \text{vec}_2(\delta_x v)_C = \frac{1}{2h} (K \otimes I) V_C - \frac{1}{2h} (e_1 \otimes I) V_{LC} + \frac{1}{2h} (e_{N-1} \otimes I) V_{RC}. \quad (d) \end{array} \right. \quad (5.24)$$

$$(ii) \left\{ \begin{array}{l} \text{vec}_2(\delta_y v)_L = \frac{1}{2h} (K \otimes I) U_L - \frac{1}{2h} (e_1 \otimes I) U_{LB} + \frac{1}{2h} (e_{N-1} \otimes I) U_{LT}, \quad (a) \\ \text{vec}_2(\delta_y v)_R = \frac{1}{2h} (K \otimes I) U_R - \frac{1}{2h} (e_1 \otimes I) U_{RB} + \frac{1}{2h} (e_{N-1} \otimes I) U_{RT}, \quad (b) \\ \text{vec}_2(\delta_y v)_F = \frac{1}{2h} (I \otimes K) U_F - \frac{1}{2h} (I \otimes e_1) U_{BF} + \frac{1}{2h} (I \otimes e_{N-1}) U_{TF}, \quad (c) \\ \text{vec}_2(\delta_y v)_C = \frac{1}{2h} (I \otimes K) U_C - \frac{1}{2h} (I \otimes e_1) U_{BC} + \frac{1}{2h} (I \otimes e_{N-1}) U_{TC}. \quad (d) \end{array} \right. \quad (5.25)$$

$$(iii) \left\{ \begin{array}{l} \text{vec}_2(\delta_z v)_L = \frac{1}{2h} (I \otimes K) U_L - \frac{1}{2h} (I \otimes e_1) U_{LF} + \frac{1}{2h} (I \otimes e_{N-1}) U_{LC}, \quad (a) \\ \text{vec}_2(\delta_z v)_R = \frac{1}{2h} (I \otimes K) U_R - \frac{1}{2h} (I \otimes e_1) U_{RF} + \frac{1}{2h} (I \otimes e_{N-1}) U_{RC}, \quad (b) \\ \text{vec}_2(\delta_z v)_B = \frac{1}{2h} (I \otimes K) U_B - \frac{1}{2h} (I \otimes e_1) U_{BF} + \frac{1}{2h} (I \otimes e_{N-1}) U_{BC}, \quad (c) \\ \text{vec}_2(\delta_z v)_T = \frac{1}{2h} (I \otimes K) U_T - \frac{1}{2h} (I \otimes e_1) U_{TF} + \frac{1}{2h} (I \otimes e_{N-1}) U_{TC}. \quad (d) \end{array} \right. \quad (5.26)$$

**Preuve.** On a par définition de l'opérateur  $\text{vec}_2$  (3.12)

$$\text{vec}_2(\delta_x v)_B = \text{vec}_2[(\delta_x v)_{i,0,k}]_{1 \leq i, k \leq N-1} = \sum_{i,k=1}^{N-1} (e_i \otimes e_k)(\delta_x v)_{i,0,k}.$$

Donc

$$\begin{aligned} \text{vec}_2(\delta_x v)_B &= \frac{1}{2h} \sum_{i,k=1}^{N-1} (e_i \otimes e_k)(v_{i+1,0,k} - v_{i-1,0,k}) \\ &= \frac{1}{2h} \left( \sum_{i=1}^{N-2} \sum_{k=1}^{N-1} (e_i \otimes e_k)v_{i+1,0,k} - \sum_{i=2}^{N-1} (e_i \otimes e_k)v_{i-1,0,k} \right. \\ &\quad \left. + \sum_{k=1}^{N-1} (e_{N-1} \otimes e_k)v_{N,0,k} - \sum_{k=1}^{N-1} (e_1 \otimes e_k)v_{0,0,k} \right) \\ &= \frac{1}{2h} \sum_{i,k=1}^{N-1} \left( (e_{i-1} - e_{i+1}) \otimes e_k \right) v_{i,0,k} - \frac{1}{2h} (e_1 \otimes I) \sum_{k=1}^{N-1} e_k v_{0,0,k} \\ &\quad + \frac{1}{2h} (e_{N-1} \otimes I) \sum_{k=1}^{N-1} (e_k \otimes v_{N,0,k}). \end{aligned}$$

Or  $e_{i-1} - e_{i+1} = Ke_i$ ,  $1 \leq i \leq N-1$ , donc

$$\text{vec}_2(\delta_x v)_B = \frac{1}{2h} (K \otimes I)V_B - \frac{1}{2h} (e_1 \otimes I)V_{LB} + \frac{1}{2h} (e_{N-1} \otimes V_{RB}),$$

d'où (5.24)<sub>a</sub>. De la même manière on démontre (5.24)<sub>b,c,d</sub>, (5.25) et (5.26).  $\blacksquare$

**Lemme 5.4.3.** Pour toute fonction de grille  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$  on a

$$(i) \left\{ \begin{array}{l} \text{vec}_1(\delta_x v)_{BF} = \frac{1}{2h} KV_{BF} - \frac{1}{2h} e_1 V_{LBF} + \frac{1}{2h} e_{N-1} V_{RBF}, \quad (a) \\ \text{vec}_1(\delta_x v)_{BC} = \frac{1}{2h} KV_{BC} - \frac{1}{2h} e_1 V_{LBC} + \frac{1}{2h} e_{N-1} V_{RBC}, \quad (b) \\ \text{vec}_1(\delta_x v)_{TF} = \frac{1}{2h} KV_{TF} - \frac{1}{2h} e_1 V_{LTF} + \frac{1}{2h} e_{N-1} V_{RTF}, \quad (c) \\ \text{vec}_1(\delta_x v)_{TC} = \frac{1}{2h} KV_{TC} - \frac{1}{2h} e_1 V_{LTC} + \frac{1}{2h} e_{N-1} V_{RTC}. \quad (d) \end{array} \right. \quad (5.27)$$

$$(ii) \left\{ \begin{array}{l} \text{vec}_1(\delta_y v)_{LF} = \frac{1}{2h} KU_{LF} - \frac{1}{2h} e_1 U_{LBF} + \frac{1}{2h} e_{N-1} U_{LTF}, \quad (a) \\ \text{vec}_1(\delta_y v)_{LC} = \frac{1}{2h} KU_{LC} - \frac{1}{2h} e_1 U_{LBC} + \frac{1}{2h} e_{N-1} U_{LTC}, \quad (b) \\ \text{vec}_1(\delta_y v)_{RF} = \frac{1}{2h} KU_{RF} - \frac{1}{2h} e_1 U_{RBF} + \frac{1}{2h} e_{N-1} U_{RTF}, \quad (c) \\ \text{vec}_1(\delta_y v)_{RC} = \frac{1}{2h} KU_{RC} - \frac{1}{2h} e_1 U_{RBC} + \frac{1}{2h} e_{N-1} U_{RTC}. \quad (d) \end{array} \right. \quad (5.28)$$

$$(iii) \left\{ \begin{array}{l} \text{vec}_1(\delta_z v)_{LB} = \frac{1}{2h} K U_{LB} - \frac{1}{2h} e_1 U_{LBF} + \frac{1}{2h} e_{N-1} U_{LBC}, \quad (a) \\ \text{vec}_1(\delta_z v)_{LT} = \frac{1}{2h} K U_{LT} - \frac{1}{2h} e_1 U_{LTF} + \frac{1}{2h} e_{N-1} U_{LTC}, \quad (b) \\ \text{vec}_1(\delta_z v)_{RB} = \frac{1}{2h} K U_{RB} - \frac{1}{2h} e_1 U_{RBF} + \frac{1}{2h} e_{N-1} U_{RBC}, \quad (c) \\ \text{vec}_1(\delta_z v)_{RT} = \frac{1}{2h} K U_{RT} - \frac{1}{2h} e_1 U_{RTF} + \frac{1}{2h} e_{N-1} U_{RTC}. \quad (d) \end{array} \right. \quad (5.29)$$

**Preuve.** D'après la définition de  $\text{vec}_1$  (1.36) on a

$$\text{vec}_1(\delta_x v)_{BF} = \text{vec}_1[(\delta_x v)_{i,0,0}]_{1 \leq i \leq N-1} = \sum_{i=1}^{N-1} e_i (\delta_x v)_{i,0,0}.$$

Donc

$$\begin{aligned} \text{vec}_1(\delta_x v)_{BF} &= \frac{1}{2h} \sum_{i=1}^{N-1} e_1 (v_{i+1,0,0} - v_{i-1,0,0}) \\ &= \frac{1}{2h} \left( \sum_{i=1}^{N-1} e_i v_{i+1,0,0} - \sum_{i=1}^{N-1} e_i v_{i-1,0,0} \right) \\ &= \frac{1}{2h} \sum_{i=1}^{N-2} e_i v_{i+1,0,0} + e_{N-1} v_{N,0,0} - e_1 v_{0,0,0} - \sum_{i=2}^{N-1} e_i v_{i-1,0,0} \\ &= \frac{1}{2h} e_{N-1} v_{N,0,0} - \frac{1}{2h} e_1 v_{0,0,0} + \frac{1}{2h} \sum_{i=1}^{N-1} (e_{i-1} - e_{i+1}) v_{i,0,0}. \end{aligned}$$

Or on a  $e_{i-1} - e_{i+1} = K e_i$ ,  $1 \leq i \leq N-1$ , donc

$$\begin{aligned} \text{vec}_1(\delta_x v)_{BF} &= \frac{1}{2h} e_{N-1} V_{RBF} - \frac{1}{2h} e_1 V_{LBF} + \frac{1}{2h} \sum_{i=1}^{N-1} V_{i,BF} \\ &= \frac{1}{2h} K V_{BF} - \frac{1}{2h} e_1 V_{LBF} + \frac{1}{2h} e_{N-1} V_{RBF}, \end{aligned}$$

d'où (5.27)<sub>a</sub>. De la même manière on démontre (5.27)<sub>b,c,d</sub>, (5.28) et (5.29). ■

**Lemme 5.4.4.** Pour toute fonction de grille  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$  on a

$$\left\{ \begin{array}{l} \text{vec}_3(\sigma_x v) = (P_s \otimes I \otimes I)V + \frac{1}{6}(e_1 \otimes I \otimes I)V_L + \frac{1}{6}(e_{N-1} \otimes I \otimes I)V_R, \quad (a) \\ \text{vec}_3(\sigma_y v) = (I \otimes P_s \otimes I)V + \frac{1}{6}(I \otimes e_1 \otimes I)V_B + \frac{1}{6}(I \otimes e_{N-1} \otimes I)V_T, \quad (b) \\ \text{vec}_3(\sigma_z v) = (I \otimes I \otimes P_s)V + \frac{1}{6}(I \otimes I \otimes e_1)V_F + \frac{1}{6}(I \otimes I \otimes e_{N-1})V_C. \quad (c) \end{array} \right. \quad (5.30)$$

**Preuve.** Par définition de l'opérateur  $\text{vec}_3$  (5.11) on a

$$\text{vec}_3(\sigma_x v) = \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k)(\sigma_x v)_{i,j,k}.$$

Donc

$$\begin{aligned} \text{vec}_3(\sigma_x v) &= \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) \left( \frac{1}{6} v_{i-1,j,k} + \frac{2}{3} v_{i,j,k} + \frac{1}{6} v_{i+1,j,k} \right) \\ &= \frac{1}{6} \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i-1,j,k} + \frac{2}{3} \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i,j,k} + \frac{1}{6} \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i+1,j,k} \\ &= \frac{1}{6} \sum_{j,k=1}^{N-1} (e_1 \otimes e_j \otimes e_k) v_{0,j,k} + \frac{1}{6} \sum_{i=2}^{N-1} \sum_{j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i-1,j,k} \\ &+ \frac{2}{3} \sum_{i,j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i,j,k} + \frac{1}{6} \sum_{i=1}^{N-2} \sum_{j,k=1}^{N-1} (e_i \otimes e_j \otimes e_k) v_{i+1,j,k} + \frac{1}{6} \sum_{j,k=1}^{N-1} (e_{N-1} \otimes e_j \otimes e_k) v_{N,j,k} \\ &= \frac{1}{6} (e_1 \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k) v_{0,j,k} + \sum_{i,j,k=1}^{N-1} \left( \left( \frac{1}{6} e_{i-1} + \frac{2}{3} e_i + \frac{1}{6} e_{i+1} \right) \otimes e_j \otimes e_k \right) v_{i,j,k} \\ &+ \frac{1}{6} (e_{N-1} \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k) v_{N,j,k}. \end{aligned}$$

Or  $\frac{1}{6} e_{i-1} + \frac{2}{3} e_i + \frac{1}{6} e_{i+1} = P_s e_i$ ,  $1 \leq i \leq N-1$ , donc

$$\text{vec}_3(\sigma_x v) = \frac{1}{6} (e_1 \otimes I \otimes I) V_L + (P_s \otimes I \otimes I) V + \frac{1}{6} (e_{N-1} \otimes I \otimes I) V_R,$$

d'où (5.30). De la même manière on démontre (5.30)<sub>b,c</sub>. ■

**Lemme 5.4.5.** Pour toute fonction de grille  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N-1} \in L_h^2$  on a

$$(i) \begin{cases} \text{vec}_2(\sigma_x v)_B = (P_s \otimes I) V_B + \frac{1}{6} (e_1 \otimes I) V_{LB} + \frac{1}{6} (e_{N-1} \otimes I) V_{RB}, & (a) \\ \text{vec}_2(\sigma_x v)_T = (P_s \otimes I) V_T + \frac{1}{6} (e_1 \otimes I) V_{LT} + \frac{1}{6} (e_{N-1} \otimes I) V_{RT}, & (b) \\ \text{vec}_2(\sigma_x v)_F = (P_s \otimes I) V_F + \frac{1}{6} (e_1 \otimes I) V_{LF} + \frac{1}{6} (e_{N-1} \otimes I) V_{RF}, & (c) \\ \text{vec}_2(\sigma_x v)_C = (P_s \otimes I) V_C + \frac{1}{6} (e_1 \otimes I) V_{LC} + \frac{1}{6} (e_{N-1} \otimes I) V_{RC}. & (d) \end{cases} \quad (5.31)$$

$$(ii) \left\{ \begin{array}{l} \text{vec}_2(\sigma_y v)_L = (P_s \otimes I)V_L + \frac{1}{6}(e_1 \otimes I)V_{LB} + \frac{1}{6}(e_{N-1} \otimes I)V_{LT}, \quad (a) \\ \text{vec}_2(\sigma_y v)_R = (P_s \otimes I)V_R + \frac{1}{6}(e_1 \otimes I)V_{RB} + \frac{1}{6}(e_{N-1} \otimes I)V_{RT}, \quad (b) \\ \text{vec}_2(\sigma_y v)_F = (I \otimes P_s)V_F + \frac{1}{6}(I \otimes e_1)V_{BF} + \frac{1}{6}(I \otimes e_{N-1})V_{TF}, \quad (c) \\ \text{vec}_2(\sigma_y v)_C = (I \otimes P_s)V_C + \frac{1}{6}(I \otimes e_1)V_{BC} + \frac{1}{6}(I \otimes e_{N-1})V_{TC}. \quad (d) \end{array} \right. \quad (5.32)$$

$$(iii) \left\{ \begin{array}{l} \text{vec}_2(\sigma_z v)_L = (I \otimes P_s)V_L + \frac{1}{6}(I \otimes e_1)V_{LF} + \frac{1}{6}(I \otimes e_{N-1})V_{LC}, \quad (a) \\ \text{vec}_2(\sigma_z v)_R = (I \otimes P_s)V_R + \frac{1}{6}(I \otimes e_1)V_{RF} + \frac{1}{6}(I \otimes e_{N-1})V_{RC}, \quad (b) \\ \text{vec}_2(\sigma_z v)_B = (I \otimes P_s)V_B + \frac{1}{6}(I \otimes e_1)V_{BF} + \frac{1}{6}(I \otimes e_{N-1})V_{BC}, \quad (c) \\ \text{vec}_2(\sigma_z v)_T = (I \otimes P_s)V_T + \frac{1}{6}(I \otimes e_1)V_{TF} + \frac{1}{6}(I \otimes e_{N-1})V_{TC}. \quad (d) \end{array} \right. \quad (5.33)$$

**Preuve.** Par définition de l'opérateur  $\text{vec}_2$  (3.12) on a

$$\begin{aligned} \text{vec}_2(\sigma_x v)_B &= \sum_{i,k=1}^{N-1} (e_i \otimes e_k)(\sigma_x v)_{i,0,k} \\ &= \sum_{i,k=1}^{N-1} (e_i \otimes e_k) \left( \frac{1}{6}v_{i-1,0,k} + \frac{2}{3}v_{i,0,k} + \frac{1}{6}v_{i+1,0,k} \right) \\ &= \frac{1}{6} \sum_{i,k=1}^{N-1} (e_i \otimes e_k)v_{i-1,0,k} + \frac{2}{3} \sum_{i,k=1}^{N-1} (e_i \otimes e_k)v_{i,0,k} + \frac{1}{6} \sum_{i,k=1}^{N-1} (e_i \otimes e_k)v_{i+1,0,k} \\ &= \frac{1}{6} \sum_{k=1}^{N-1} (e_1 \otimes e_k)v_{0,0,k} + \frac{1}{6} \sum_{i=2}^{N-1} \sum_{j=1}^{N-1} (e_i \otimes e_k)v_{i-1,0,k} + \frac{2}{3} \sum_{i,k=1}^{N-1} (e_i \otimes e_k)v_{i,0,k} \\ &\quad + \frac{1}{6} \sum_{i=1}^{N-2} \sum_{k=1}^{N-1} (e_i \otimes e_k)v_{i+1,0,k} + \frac{1}{6} \sum_{k=1}^{N-1} (e_{N-1} \otimes e_k)v_{N,0,k} \\ &= \frac{1}{6} (e_1 \otimes I) \sum_{k=1}^{N-1} e_k v_{0,0,k} + \sum_{i,k=1}^{N-1} \left( \left( \frac{1}{6}e_{i-1} + \frac{2}{3}e_i + \frac{1}{6}e_{i+1} \right) \otimes e_k \right) v_{i,0,k} \\ &\quad + \frac{1}{6} (e_{N-1} \otimes I) \sum_{k=1}^{N-1} e_k v_{N,0,k}. \end{aligned}$$

Or  $\frac{1}{6}e_{i-1} + \frac{2}{3}e_i + \frac{1}{6}e_{i+1} = P_s e_i$ ,  $1 \leq i \leq N-1$ , donc

$$\begin{aligned} \text{vec}_2(\sigma_x v)_B &= \frac{1}{6} (e_1 \otimes I) \sum_{k=1}^{N-1} e_k v_{0,0,k} + \sum_{i,k=1}^{N-1} (P_s e_i \otimes e_k) v_{i,0,k} + \frac{1}{6} (e_{N-1} \otimes I) \sum_{k=1}^{N-1} e_k v_{N,0,k} \\ &= \frac{1}{6} (e_1 \otimes I)V_{LB} + (P_s \otimes I)V_B + \frac{1}{6} (e_{N-1} \otimes I)V_{RB}, \end{aligned}$$

d'où (5.31)<sub>a</sub>. De la même manière on démontre (5.31)<sub>b,c,d</sub>, (5.32) et (5.33).  $\blacksquare$

**Lemme 5.4.6.** *Pour toute fonction de grille  $v = (v_{i,j,k})_{0 \leq i,j,k \leq N} \in L_h^2$  on a*

$$(i) \begin{cases} \text{vec}_1(\sigma_x v)_{BF} = P_s V_{BF} + \frac{1}{6} V_{LBF} + \frac{1}{6} V_{RBF}, & (a) \\ \text{vec}_1(\sigma_x v)_{BC} = P_s V_{BC} + \frac{1}{6} V_{LBC} + \frac{1}{6} V_{RBC}, & (b) \\ \text{vec}_1(\sigma_x v)_{TF} = P_s V_{TF} + \frac{1}{6} V_{LTF} + \frac{1}{6} V_{RTF}, & (c) \\ \text{vec}_1(\sigma_x v)_{TC} = P_s V_{TC} + \frac{1}{6} V_{LTC} + \frac{1}{6} V_{RTC}. & (d) \end{cases} \quad (5.34)$$

$$(ii) \begin{cases} \text{vec}_1(\sigma_y v)_{LF} = P_s V_{LF} + \frac{1}{6} V_{LBF} + \frac{1}{6} V_{LTF}, & (a) \\ \text{vec}_1(\sigma_y v)_{LC} = P_s V_{LC} + \frac{1}{6} V_{LBC} + \frac{1}{6} V_{LTC}, & (b) \\ \text{vec}_1(\sigma_y v)_{RF} = P_s V_{RF} + \frac{1}{6} V_{RBF} + \frac{1}{6} V_{RTF}, & (c) \\ \text{vec}_1(\sigma_y v)_{RC} = P_s V_{RC} + \frac{1}{6} V_{RBC} + \frac{1}{6} V_{RTC}. & (d) \end{cases} \quad (5.35)$$

$$(iii) \begin{cases} \text{vec}_1(\sigma_z v)_{LB} = P_s V_{LB} + \frac{1}{6} V_{LBF} + \frac{1}{6} V_{LBC}, & (a) \\ \text{vec}_1(\sigma_z v)_{LT} = P_s V_{LT} + \frac{1}{6} V_{LTF} + \frac{1}{6} V_{LTC}, & (b) \\ \text{vec}_1(\sigma_z v)_{RB} = P_s V_{RB} + \frac{1}{6} V_{RBF} + \frac{1}{6} V_{RBC}, & (c) \\ \text{vec}_1(\sigma_z v)_{RT} = P_s V_{RT} + \frac{1}{6} V_{RTF} + \frac{1}{6} V_{RTC}. & (d) \end{cases} \quad (5.36)$$

**Preuve.** On a par définition de l'opérateur  $\text{vec}_1$

$$\begin{aligned} \text{vec}_1(\sigma_x v)_{BF} &= \sum_{i=1}^{N-1} e_i (\sigma_x v)_{i,0,0} \\ &= \sum_{i=1}^{N-1} e_i \left( \frac{1}{6} v_{i-1,0,0} + \frac{2}{3} v_{i,0,0} + \frac{1}{6} v_{i+1,0,0} \right) \\ &= \frac{1}{6} \sum_{i=1}^{N-1} e_i v_{i-1,0,0} + \frac{2}{3} \sum_{i=1}^{N-1} e_i v_{i,0,0} + \frac{1}{6} \sum_{i=1}^{N-1} e_i v_{i+1,0,0} \\ &= \frac{1}{6} e_1 v_{0,0,0} + \frac{1}{6} \sum_{i=2}^{N-1} e_i v_{i-1,0,0} + \frac{2}{3} \sum_{i=1}^{N-1} e_i v_{i,0,0} \\ &\quad + \frac{1}{6} \sum_{i=1}^{N-2} e_i v_{i+1,0,0} + \frac{1}{6} e_{N-1} v_{N,0,0} \\ &= \frac{1}{6} e_1 v_{0,0,0} + \sum_{i=1}^{N-1} \left( \frac{1}{6} e_{i-1} + \frac{2}{3} e_i + \frac{1}{6} e_{i+1} \right) v_{i,0,0} + \frac{1}{6} e_{N-1} v_{N,0,0}. \end{aligned}$$



Or on a  $\frac{1}{6}e_{i-1} + \frac{2}{3}e_i + \frac{1}{6}e_{i+1} = P_s e_i$ ,  $1 \leq i \leq N-1$ , donc

$$\begin{aligned} \text{vec}_1(\sigma_x v)_{BF} &= \frac{1}{6}e_1 v_{0,0,0} + \sum_{i=1}^{N-1} P_s e_i v_{i,0,0} + \frac{1}{6}e_{N-1} v_{N,0,0} \\ &= \frac{1}{6}e_1 V_{LBF} + P_s V_B + \frac{1}{6}V_{RBF}, \end{aligned}$$

d'où (5.34)<sub>a</sub>. De la même manière on démontre (5.34)<sub>b,c,d</sub>, (5.35) et (5.36). ■

## 5.5 Schéma boîte hermitien en dimension 3

### 5.5.1 Introduction

Dans la section 2.2, on a vu le principe du schéma boîte hermitien repose sur :

- La formulation mixte (2.6).
- L'évaluation des moyennes sur  $[x_{j-1}, x_{j+1}]$  de (2.7) et (2.9).

Ce point de vue est généralisé en trois dimensions pour le problème de Poisson dans un cube. En utilisant le théorème de la divergence (3.36), on établit un système d'équations qui sera *la base du schéma HB en trois dimensions*. Pour compléter ce système, on donne dans la section 5.5.3 deux exemples d'approximations des dérivées normales  $\frac{\partial u}{\partial n}$  sur  $\partial\Omega$  déduites des approximations des dérivées au bord utilisées au chapitre 2.

### 5.5.2 Définition du schéma hermitien (HB) pour le problème de Poisson dans $\Omega = (a, b)^3$

On considère le problème de Poisson dans le cube  $\Omega = (a, b)^3$

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega, \\ u = g, & \text{sur } \partial\Omega. \end{cases} \quad (5.37)$$

La forme mixte du (5.37) s'obtient en introduisant  $v^1 = \partial_x u$ ,  $v^2 = \partial_y u$ ,  $v^3 = \partial_z u$ ,

$$\begin{cases} -(\partial_x v^1 + \partial_y v^2 + \partial_z v^3) = f, & \text{dans } \Omega & (a) \\ v^1 - \partial_x u = 0, & \text{dans } \bar{\Omega}, & (b) \\ v^2 - \partial_y u = 0, & \text{dans } \bar{\Omega}, & (c) \\ v^3 - \partial_z u = 0, & \text{dans } \bar{\Omega}, & (d) \\ u = g, & \text{sur } \partial\Omega. & (e) \end{cases} \quad (5.38)$$

Le domaine  $\Omega = (a, b)^3$  est discrétisé par la grille (5.1). Les boîtes sur lesquelles on fait les moyennes (5.38)<sub>b,c,d</sub> sont  $K_{i,j,k}$ , centrées aux points  $(x_i, y_j, z_k)$  de longueur  $2h$  (voir figure 5.2), c'est-à-dire pour  $1 \leq i, j, k \leq N - 1$ ,

$$K_{i,j,k} = [x_i - h, x_i + h] \times [y_j - h, y_j + h] \times [z_k - h, z_k + h]. \quad (5.39)$$

A chaque point  $(x_i, y_j, z_k)$ ,  $0 \leq i, j, k \leq N$ , sont localisées les quatres inconnues

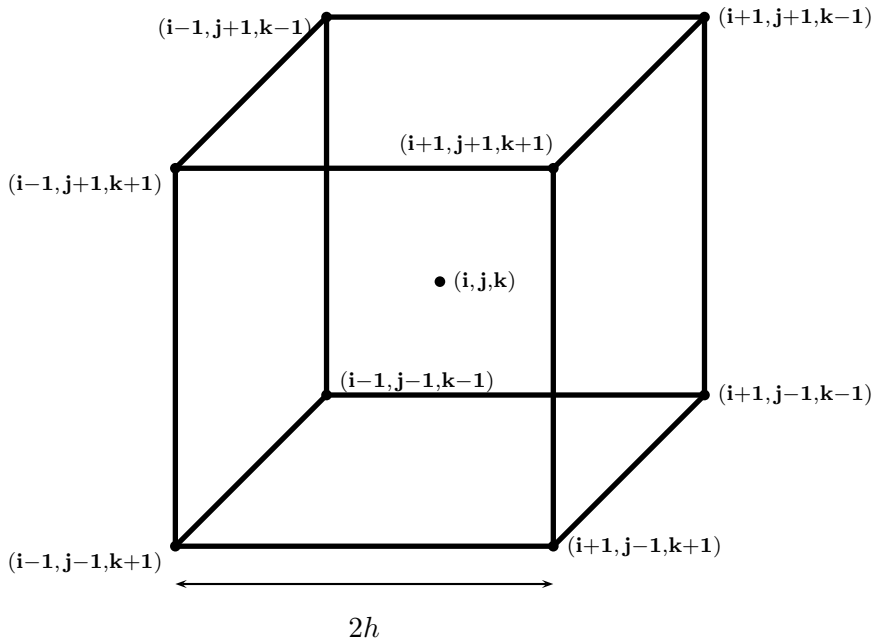


FIG. 5.2: Schéma HB en trois dimensions. Le flux est intégré sur la surface de ce cube

$$K_{i,j,k} = [x_i - h, x_i + h] \times [y_j - h, y_j + h] \times [z_k - h, z_k + h].$$

$$\begin{cases} u_{i,j,k} \simeq u(x_i, y_j, z_k), \\ u_{x,i,j,k} \simeq \partial_x u(x_i, y_j, z_k), \\ u_{y,i,j,k} \simeq \partial_y u(x_i, y_j, z_k), \\ u_{z,i,j,k} \simeq \partial_z u(x_i, y_j, z_k). \end{cases} \quad (5.40)$$

On a donc quatre fonctions de grille  $u, u_x, u_y, u_z \in L_h^2$ .

Établissons d'abord la *forme discrète* de (5.38).

- On intègre l'équation conservative (5.38)<sub>a</sub> sur la boîte  $K_{i,j,k}$  (figure 5.2). Le membre

de droite de (5.38)<sub>a</sub> s'écrit

$$\Pi^0 f_{i,j,k} = \frac{1}{(2h)^3} \int_{K_{i,j,k}} f(x, y, z) dx dy dz. \quad (5.41)$$

En utilisant le théorème de la divergence (3.36), le membre de gauche de (5.38)<sub>a</sub> s'écrit

$$-\frac{1}{4h^2} \int_{\partial K_{i,j,k}} (v^1 n_1 + v^2 n_2 + v^3 n_3) d\sigma = \Pi^0 f_{i,j,k}, \quad 1 \leq i, j, k \leq N-1, \quad (5.42)$$

où  $\vec{n}(n_1, n_2, n_3)$  est le vecteur normal dirigé vers l'extérieur de  $\partial K_{i,j,k}$ . La relation (5.42) comporte six intégrales bidimensionnelles associées aux six faces du cube. Chacune de ces intégrales est approchée par la formule de Simpson bidimensionnelle. Par exemple sur la face arrière du cube  $K_{i,j,k}$  qui correspond à  $z = z_{k-1}$  et  $\vec{n}(0, 0, -1)$  on a

$$\frac{1}{4h^2} \int_{y_{j-1}}^{y_{j+1}} \int_{x_{i-1}}^{x_{i+1}} \partial_z u(x, y, z) dy dx \simeq \left( (\sigma_y \circ \sigma_x) u_z \right)_{i,j,k-1}, \quad 1 \leq i, j \leq N-1. \quad (5.43)$$

En utilisant les opérateurs aux différences  $\delta_x, \delta_y, \delta_z$  en (5.3) et  $\sigma_x, \sigma_y, \sigma_z$  en (5.9), l'équation (5.42) s'écrit

$$-(\sigma_z \circ \sigma_y \circ \delta_x) u_{x,i,j,k} - (\sigma_z \circ \sigma_x \circ \delta_y) u_{y,i,j,k} - (\sigma_y \circ \sigma_x \circ \delta_z) u_{z,i,j,k} = \Pi^0 f_{i,j,k}, \quad (5.44)$$

avec  $1 \leq i, j, k \leq N-1$ .

Le symbole  $\circ$  désigne la composition des opérateurs. L'expression (5.44) donne  $(N-1)^3$  équations.

• La relation suivante entre les fonctions exactes  $u(x, y)$  et  $v^1(x, y)$  est déduite de l'identité (5.38)<sub>b</sub>,

$$v^1(x, y) = \partial_x u(x, y), \text{ sur } \bar{\Omega}. \quad (5.45)$$

En intégrant (5.45) sur le segment  $[x_{i-1}, x_{i+1}] \times \{y_j\} \times \{z_k\}$ , on obtient l'identité

$$\frac{1}{2h} \int_{x_{i-1}}^{x_{i+1}} v^1(x, y_j, z_k) dx = \frac{u(x_{i+1}, y_j, z_k) - u(x_{i-1}, y_j, z_k)}{2h}, \quad 1 \leq i \leq N-1, 0 \leq j, k \leq N. \quad (5.46)$$

La relation suivante entre les fonctions  $u(x, y)$  et  $v^2(x, y)$  est déduite de (5.38)<sub>c</sub>

$$v^2(x, y) = \partial_y u(x, y), \text{ sur } \bar{\Omega}. \quad (5.47)$$

En intégrant (5.47) sur le segment  $\{x_i\} \times [y_{j-1}, y_{j+1}] \times \{z_k\}$ , on obtient l'identité

$$\frac{1}{2h} \int_{y_{j-1}}^{y_{j+1}} v^2 u(x_i, y, z_k) dy = \frac{u(x_i, y_{j+1}, z_k) - u(x_i, y_{j-1}, z_k)}{2h}, \quad 1 \leq j \leq N-1, 0 \leq i, k \leq N. \quad (5.48)$$

De même, la relation suivante entre les fonctions  $u(x, y)$  et  $v^3(x, y)$  est déduite de (5.38)<sub>d</sub>

$$v^3(x, y) = \partial_y u(x, y), \text{ sur } \bar{\Omega}, \quad (5.49)$$

En intégrant (5.49) sur le segment  $\{x_i\} \times \{y_j\} \times [z_{k-1}, z_{k+1}]$ , on obtient l'identité

$$\frac{1}{2h} \int_{z_{k-1}}^{z_{k+1}} v^3(x_i, y_j, z) dz = \frac{u(x_i, y_j, z_{k+1}) - u(x_i, y_j, z_{k-1})}{2h}, \quad 1 \leq k \leq N-1, 0 \leq i, j \leq N. \quad (5.50)$$

On approche à présent  $v^1(x, y)$  par la fonction de grille  $u_x \in L_h^2$ , c'est-à-dire  $u_{x,i,j,k} \simeq v^1(x_i, y_j, z_k)$ . En approchant l'intégrale de droite de (5.46) par l'opérateur de Simpson  $\sigma_x$  (5.9) on obtient l'identité discrète suivante

$$\sigma_x u_{x,i,j,k} = \delta_x u_{i,j,k}, \quad 1 \leq i \leq N-1, \quad 0 \leq j, k \leq N. \quad (5.51)$$

On en déduit que (5.51) est une approximation hermitienne de (5.38)<sub>b</sub>.

De même, en approchant  $v^2$  et  $v^3$  par les fonctions de grille  $u_y$  et  $u_z \in L_h^2$  respectivement, on obtient

$$\begin{cases} \sigma_y u_{y,i,j,k} = \delta_y u_{i,j,k}, & 0 \leq i, k \leq N, \quad 1 \leq j \leq N-1, & (a) \\ \sigma_z u_{z,i,j,k} = \delta_z u_{i,j,k}, & 0 \leq i, j \leq N, \quad 1 \leq k \leq N-1, & (b) \end{cases} \quad (5.52)$$

où  $\sigma_y, \sigma_z$  sont donnés en (5.9). Les identités (5.52)<sub>a,b</sub> sont des approximations hermitiennes de (5.38)<sub>c,d</sub> basées sur l'approximation des intégrales par la formule de Simpson. Le décompte des équations et des inconnues est à ce stade  $3(N-1)(N+1)^2$  équations correspondant aux (5.51-5.52) pour  $4(N+1)^3$  inconnues correspondant aux  $u = (u_{i,j,k}), u_x = (u_{x,i,j,k}), u_y = (u_{y,i,j,k}), u_z = (u_{z,i,j,k}) \in L_h^2, 0 \leq i, j, k \leq N$ .

• Les conditions limites de Dirichlet s'écrivent :

$$\left\{ \begin{array}{lll} u_{i,j,0} = g(x_i, y_j, z_0), & 0 \leq i, j \leq N, & \text{Face "F"} \\ u_{i,j,N} = g(x_i, y_j, z_N), & 0 \leq i, j \leq N, & \text{Face "C"} \\ u_{i,0,k} = g(x_i, y_0, z_k), & 0 \leq i \leq N, \quad 1 \leq k \leq N-1, & \text{Face "B"} \\ u_{i,N,k} = g(x_i, y_N, z_k), & 0 \leq i \leq N, \quad 1 \leq k \leq N-1, & \text{Face "T"} \\ u_{0,j,k} = g(x_0, y_j, z_k), & 1 \leq j, k \leq N-1, & \text{Face "L"} \\ u_{N,j,k} = g(x_N, y_j, z_k), & 1 \leq j, k \leq N-1. & \text{Face "R"} \end{array} \right. \quad (5.53)$$

Ce qui donne  $2(N+1)^2 + 2(N-1)(N+1) + 2(N-1)^2 = 6N^2 + 2$  équations.

En regroupant (5.44), (5.51), (5.52) et (5.53) on obtient

$$\left\{ \begin{array}{ll} -(\sigma_z \circ \sigma_y \circ \delta_x)u_{x,i,j,k} - (\sigma_z \circ \sigma_x \circ \delta_y)u_{y,i,j,k} & \\ -(\sigma_y \circ \sigma_x \circ \delta_z)u_{z,i,j,k} = \Pi^0 f_{i,j,k}, & 1 \leq i, j, k \leq N-1, \quad (a) \\ \sigma_x u_{x,i,j,k} = \delta_x u_{i,j,k}, & 1 \leq i \leq N-1, \quad 0 \leq j, k \leq N, \quad (b) \\ \sigma_y u_{y,i,j,k} = \delta_y u_{i,j,k}, & 0 \leq i, k \leq N, \quad 1 \leq j \leq N-1, \quad (c) \\ \sigma_z u_{z,i,j,k} = \delta_z u_{i,j,k}, & 0 \leq i, j \leq N, \quad 1 \leq k \leq N-1, \quad (d) \\ u_{i,j,0} = g(x_i, y_j, z_0), & 0 \leq i, j \leq N, \quad (e) \\ u_{i,j,N} = g(x_i, y_j, z_N), & 0 \leq i, j \leq N, \quad (f) \\ u_{i,0,k} = g(x_i, y_0, z_k), & 0 \leq i \leq N, \quad 1 \leq k \leq N-1, \quad (g) \\ u_{i,N,k} = g(x_i, y_N, z_k), & 0 \leq i \leq N, \quad 1 \leq k \leq N-1, \quad (l) \\ u_{0,j,k} = g(x_0, y_j, z_k), & 1 \leq j, k \leq N-1, \quad (m) \\ u_{N,j,k} = g(x_N, y_j, z_k), & 1 \leq j, k \leq N-1. \quad (n) \end{array} \right. \quad (5.54)$$

Le système (5.54) constitue les relations de *base du schéma boîte hermitien en dimension 3*. Il consiste en  $(N-1)^3 + 3(N-1)(N+1)^2 + 6N^2 + 2$  équations pour  $4(N+1)^3$  inconnues. Pour obtenir le bon décompte entre le nombre d'équations et le nombre d'inconnues, on complète le système (5.54) par des approximations des dérivées normales au bord  $u_{x,0,j,k}$ ,  $u_{x,N,j,k}$ ,  $0 \leq j, k \leq N$ ,  $u_{y,i,0,k}$ ,  $u_{y,i,N,k}$ ,  $0 \leq i, k \leq N$ ,  $u_{z,i,j,0}$ ,  $u_{z,i,j,N}$ ,  $0 \leq i, j \leq N$ .

Comme on l'a déjà fait en dimension 1 et 2 dans les chapitres 2 et 3, on présente dans la section suivante deux exemples d'approximations de ces dérivées au bord.

### 5.5.3 Exemples d'approximations de $\frac{\partial u}{\partial n}$ sur $\partial\Omega$

#### 5.5.3.1 Schéma HB1

En dimension 1, le premier exemple d'approximations des dérivées au bord en (2.26) est

$$\left\{ \begin{array}{ll} \frac{2}{3}u_{x,0} + \frac{1}{3}u_{x,1} = \frac{u_1 - u_0}{h}, & \text{“L”} \\ \frac{2}{3}u_{x,N} + \frac{1}{3}u_{x,N-1} = \frac{u_N - u_{N-1}}{h}. & \text{“R”} \end{array} \right. \quad (5.55)$$

En dimension 2, ceci est généralisé en (3.53).

En dimension 3, on généralise (5.55), sur les six faces du cube  $\Omega = (a, b)^3$  par les 6

relations

$$\left\{ \begin{array}{l} \frac{2}{3}u_{x,0,j,k} + \frac{1}{3}u_{x,1,j,k} = \frac{u_{1,j,k} - u_{0,j,k}}{h}, \quad 0 \leq j, k \leq N, \text{ Face "L"} \\ \frac{2}{3}u_{x,N,j,k} + \frac{1}{3}u_{x,N-1,j,k} = \frac{u_{N,j,k} - u_{N-1,j,k}}{h}, \quad 0 \leq j, k \leq N, \text{ Face "R"} \\ \frac{2}{3}u_{y,i,0,k} + \frac{1}{3}u_{y,i,1,k} = \frac{u_{i,1,k} - u_{i,0,k}}{h}, \quad 0 \leq i, k \leq N, \text{ Face "B"} \\ \frac{2}{3}u_{y,i,N,k} + \frac{1}{3}u_{y,i,N-1,k} = \frac{u_{i,N,k} - u_{i,N-1,k}}{h}, \quad 0 \leq i, k \leq N, \text{ Face "T"} \\ \frac{2}{3}u_{z,i,j,0} + \frac{1}{3}u_{z,i,j,1} = \frac{u_{i,j,1} - u_{i,j,0}}{h}, \quad 0 \leq i, j \leq N, \text{ Face "F"} \\ \frac{2}{3}u_{z,i,j,N} + \frac{1}{3}u_{z,i,j,N-1} = \frac{u_{i,j,N} - u_{i,j,N-1}}{h}, \quad 0 \leq i, j \leq N, \text{ Face "C"} \end{array} \right. \quad (5.56)$$

(5.56) donne  $6(N+1)^2$  équations.

En complétant (5.54) par (5.56) on obtient le système discret :

$$\left\{ \begin{array}{l} -(\sigma_z \circ \sigma_y \circ \delta_x)u_{x,i,j,k} - (\sigma_z \circ \sigma_x \circ \delta_y)u_{y,i,j,k} \\ -(\sigma_y \circ \sigma_x \circ \delta_z)u_{z,i,j,k} = \Pi^0 f_{i,j,k}, \quad 1 \leq i, j, k \leq N-1, \quad (a) \\ \sigma_x u_{x,i,j,k} = \delta_x u_{i,j,k}, \quad 1 \leq i \leq N-1, 0 \leq j, k \leq N, \quad (b_1) \\ \sigma_y u_{y,i,j,k} = \delta_y u_{i,j,k}, \quad 0 \leq i, k \leq N, 1 \leq j \leq N-1, \quad (b_2) \\ \sigma_z u_{z,i,j,k} = \delta_z u_{i,j,k}, \quad 0 \leq i, j \leq N, 1 \leq k \leq N-1, \quad (b_3) \\ \frac{2}{3}u_{x,0,j,k} + \frac{1}{3}u_{x,1,j,k} = \frac{u_{1,j,k} - u_{0,j,k}}{h}, \quad 0 \leq j, k \leq N, \quad (c_1) \\ \frac{2}{3}u_{x,N,j,k} + \frac{1}{3}u_{x,N-1,j,k} = \frac{u_{N,j,k} - u_{N-1,j,k}}{h}, \quad 0 \leq j, k \leq N, \quad (c_2) \\ \frac{2}{3}u_{y,i,0,k} + \frac{1}{3}u_{y,i,1,k} = \frac{u_{i,1,k} - u_{i,0,k}}{h}, \quad 0 \leq i, k \leq N, \quad (c_3) \\ \frac{2}{3}u_{y,i,N,k} + \frac{1}{3}u_{y,i,N-1,k} = \frac{u_{i,N,k} - u_{i,N-1,k}}{h}, \quad 0 \leq i, k \leq N, \quad (c_4) \\ \frac{2}{3}u_{z,i,j,0} + \frac{1}{3}u_{z,i,j,1} = \frac{u_{i,j,1} - u_{i,j,0}}{h}, \quad 0 \leq i, j \leq N, \quad (c_5) \\ \frac{2}{3}u_{z,i,j,N} + \frac{1}{3}u_{z,i,j,N-1} = \frac{u_{i,j,N} - u_{i,j,N-1}}{h}, \quad 0 \leq i, j \leq N, \quad (c_6) \\ u_{i,j,0} = g(x_i, y_j, z_0), \quad 0 \leq i, j \leq N, \quad (d_1) \\ u_{i,j,N} = g(x_i, y_j, z_N), \quad 0 \leq i, j \leq N, \quad (d_2) \\ u_{i,0,k} = g(x_i, y_0, z_k), \quad 0 \leq i \leq N, 1 \leq k \leq N-1, \quad (d_3) \\ u_{i,N,k} = g(x_i, y_N, z_k), \quad 0 \leq i \leq N, 1 \leq k \leq N-1, \quad (d_4) \\ u_{0,j,k} = g(x_0, y_j, z_k), \quad 1 \leq j, k \leq N-1, \quad (d_5) \\ u_{N,j,k} = g(x_N, y_j, z_k), \quad 1 \leq j, k \leq N-1. \quad (d_6) \end{array} \right. \quad (5.57)$$

On a au total :

- (5.57)<sub>a</sub> est la forme discrète de l'équation conservative (5.38)<sub>a</sub>.
- (5.57)<sub>b<sub>1</sub>,b<sub>2</sub>,b<sub>3</sub></sub> sont les formes discrètes des équations (5.38)<sub>b,c,d</sub>.
- (5.57)<sub>c<sub>1</sub>,c<sub>2</sub>,c<sub>3</sub>,c<sub>4</sub>,c<sub>5</sub>,c<sub>6</sub></sub> sont des approximations des dérivées normales sur les six faces du cube.
- (5.57)<sub>d<sub>1</sub>,d<sub>2</sub>,d<sub>3</sub>,d<sub>4</sub>,d<sub>5</sub>,d<sub>6</sub></sub> sont les conditions de Dirichlet sur les six faces du cube données en (5.38)<sub>e</sub>.

Le système (5.57) consiste en  $(N-1)^3 + 3(N+1)^2(N-1) + 6(N+1)^2 + 6N^2 + 2 = 4(N+1)^3$  équations pour  $4(N+1)^3$  inconnues qui correspondent aux  $u, u_x, u_y, u_z$ . Le système (5.57) est le schéma HB1 en dimension 3.

### 5.5.3.2 Schéma HB2

En dimension 1, les approximations d'ordre 3 en (2.97) sont données par

$$\begin{cases} \frac{1}{3}u_{x,0} + \frac{2}{3}u_{x,1} = \frac{1}{h} \left( \frac{1}{6}u_2 + \frac{2}{3}u_1 - \frac{5}{6}u_0 \right), & \text{“L”} \\ \frac{1}{3}u_{x,N} + \frac{2}{3}u_{x,N-1} = \frac{1}{h} \left( \frac{5}{6}u_N - \frac{1}{6}u_{N-2} - \frac{2}{3}u_{N-1} \right). & \text{“R”} \end{cases} \quad (5.58)$$

En dimension 2, (5.58) est généralisé en (3.55).

En dimension 3, la généralisation de (5.58) sur les six faces du cube  $\Omega = (a, b)^3$  est

$$\begin{cases} \frac{1}{3}u_{x,0,j,k} + \frac{2}{3}u_{x,1,j,k} = \frac{1}{h} \left( \frac{1}{6}u_{2,j,k} + \frac{2}{3}u_{1,j,k} - \frac{5}{6}u_{0,j,k} \right), & 0 \leq j, k \leq N, \text{ Face “L”} \\ \frac{1}{3}u_{x,N,j,k} + \frac{2}{3}u_{x,N-1,j,k} = \frac{1}{h} \left( \frac{5}{6}u_{N,j,k} - \frac{1}{6}u_{N-2,j,k} - \frac{2}{3}u_{N-1,j,k} \right), & 0 \leq j, k \leq N, \text{ Face “R”} \\ \frac{1}{3}u_{y,i,0,k} + \frac{2}{3}u_{y,i,1,k} = \frac{1}{h} \left( \frac{1}{6}u_{i,2,k} + \frac{2}{3}u_{i,1,k} - \frac{5}{6}u_{i,0,k} \right), & 0 \leq i, k \leq N, \text{ Face “B”} \\ \frac{1}{3}u_{y,i,N,k} + \frac{2}{3}u_{y,i,N-1,k} = \frac{1}{h} \left( \frac{5}{6}u_{i,N,k} - \frac{1}{6}u_{i,N-2,k} - \frac{2}{3}u_{i,N-1,k} \right), & 0 \leq i, k \leq N, \text{ Face “T”} \\ \frac{1}{3}u_{z,i,j,0} + \frac{2}{3}u_{z,i,j,1} = \frac{1}{h} \left( \frac{1}{6}u_{i,j,2} + \frac{2}{3}u_{i,j,1} - \frac{5}{6}u_{i,j,0} \right), & 0 \leq i, j \leq N, \text{ Face “F”} \\ \frac{1}{3}u_{z,i,j,N} + \frac{2}{3}u_{z,i,j,N-1} = \frac{1}{h} \left( \frac{5}{6}u_{i,j,N} - \frac{1}{6}u_{i,j,N-2} - \frac{2}{3}u_{i,j,N-1} \right), & 0 \leq i, j \leq N. \text{ Face “C”} \end{cases} \quad (5.59)$$

En complétant le système (5.54) par (5.59) on obtient

$$\left\{ \begin{array}{ll}
 -(\sigma_z \circ \sigma_y \circ \delta_x)u_{x,i,j,k} - (\sigma_z \circ \sigma_x \circ \delta_y)u_{y,i,j,k} & \\
 -(\sigma_y \circ \sigma_x \circ \delta_z)u_{z,i,j,k} = \Pi^0 f_{i,j,k}, & 1 \leq i, j, k \leq N-1, \quad (a) \\
 \sigma_x u_{x,i,j,k} = \delta_x u_{i,j,k}, & 1 \leq i \leq N-1, 0 \leq j, k \leq N, \quad (b_1) \\
 \sigma_y u_{y,i,j,k} = \delta_y u_{i,j,k}, & 0 \leq i, k \leq N, 1 \leq j \leq N-1, \quad (b_2) \\
 \sigma_z u_{z,i,j,k} = \delta_z u_{i,j,k}, & 0 \leq i, j \leq N, 1 \leq k \leq N-1, \quad (b_3) \\
 \frac{1}{3}u_{x,0,j,k} + \frac{2}{3}u_{x,1,j,k} = \frac{1}{h} \left( \frac{1}{6}u_{2,j,k} + \frac{2}{3}u_{1,j,k} - \frac{5}{6}u_{0,j,k} \right), & 0 \leq j, k \leq N, \quad (c_1) \\
 \frac{1}{3}u_{x,N,j,k} + \frac{2}{3}u_{x,N-1,j,k} = \frac{1}{h} \left( \frac{5}{6}u_{N,j,k} - \frac{1}{6}u_{N-2,j,k} - \frac{2}{3}u_{N-1,j,k} \right), & 0 \leq j, k \leq N, \quad (c_2) \\
 \frac{1}{3}u_{y,i,0,k} + \frac{2}{3}u_{y,i,1,k} = \frac{1}{h} \left( \frac{1}{6}u_{i,2,k} + \frac{2}{3}u_{i,1,k} - \frac{5}{6}u_{i,0,k} \right), & 0 \leq i, k \leq N, \quad (c_3) \\
 \frac{1}{3}u_{y,i,N,k} + \frac{2}{3}u_{y,i,N-1,k} = \frac{1}{h} \left( \frac{5}{6}u_{i,N,k} - \frac{1}{6}u_{i,N-2,k} - \frac{2}{3}u_{i,N-1,k} \right), & 0 \leq i, k \leq N. \quad (c_4) \\
 \frac{1}{3}u_{z,i,j,0} + \frac{2}{3}u_{z,i,j,1} = \frac{1}{h} \left( \frac{1}{6}u_{i,j,2} + \frac{2}{3}u_{i,j,1} - \frac{5}{6}u_{i,j,0} \right), & 0 \leq i, j \leq N, \quad (c_5) \\
 \frac{1}{3}u_{z,i,j,N} + \frac{2}{3}u_{z,i,j,N-1} = \frac{1}{h} \left( \frac{5}{6}u_{i,j,N} - \frac{1}{6}u_{i,j,N-2} - \frac{2}{3}u_{i,j,N-1} \right), & 0 \leq i, j \leq N. \quad (c_6) \\
 u_{i,j,0} = g(x_i, y_j, z_0), & 0 \leq i, j \leq N, \quad (d_1) \\
 u_{i,j,N} = g(x_i, y_j, z_N), & 0 \leq i, j \leq N, \quad (d_2) \\
 u_{i,0,k} = g(x_i, y_0, z_k), & 0 \leq i \leq N, 1 \leq k \leq N-1, \quad (d_3) \\
 u_{i,N,k} = g(x_i, y_N, z_k), & 0 \leq i \leq N, 1 \leq k \leq N-1, \quad (d_4) \\
 u_{0,j,k} = g(x_0, y_j, z_k), & 1 \leq j, k \leq N-1, \quad (d_5) \\
 u_{N,j,k} = g(x_N, y_j, z_k), & 1 \leq j, k \leq N-1. \quad (d_6)
 \end{array} \right. \quad (5.60)$$

Au total on a

- (5.60)<sub>a</sub> est la forme discrète de l'équation conservative (5.38)<sub>a</sub>.
- (5.60)<sub>b<sub>1</sub>,b<sub>2</sub>,b<sub>3</sub></sub> sont les formes discrètes des équations (5.38)<sub>b,c,d</sub>.
- (5.60)<sub>c<sub>1</sub>,c<sub>2</sub>,c<sub>3</sub>,c<sub>4</sub>,c<sub>5</sub>,c<sub>6</sub></sub> sont des approximations des dérivées normales sur les six faces du cube.
- (5.60)<sub>d<sub>1</sub>,d<sub>2</sub>,d<sub>3</sub>,d<sub>4</sub>,d<sub>5</sub>,d<sub>6</sub></sub> sont les conditions de Dirichlet sur les six faces du cube données en (5.38)<sub>e</sub>.

Le système (5.60) consiste en  $(N-1)^3 + 3(N+1)^2(N-1) + 6(N+1)^2 + 6N^2 + 2 = 4(N+1)^3$  équations pour  $4(N+1)^3$  inconnues qui correspondent aux  $u, u_x, u_y, u_z$ . Le système (5.60) est le schéma HB2 en dimension 3.



## 5.6 Forme matricielle du schéma boîte hermitien en dimension trois : cas général

### 5.6.1 Introduction

En dimension 1, au chapitre 2, on a vu que le schéma boîte hermitien pour le problème de Poisson (2.1) possède la forme matricielle

$$\frac{1}{h^2} \mathcal{H}U = F - G_x, \quad (5.61)$$

avec  $\mathcal{H} \in \mathbb{M}_{N-1}(\mathbb{R})$  et  $F, G_x \in \mathbb{R}^{(N-1)}$ .

En dimension 2, au chapitre 3, on a vu que le schéma boîte hermitien pour le problème de Poisson (3.37) possède la forme matricielle suivante

$$\frac{1}{h^2} (\mathcal{H} \otimes P_s + P_s \otimes \mathcal{H})U = F - G_x - G_y, \quad (5.62)$$

avec  $\mathcal{H}, P_s \in \mathbb{M}_{N-1}(\mathbb{R})$  et  $F, G_x, G_y \in \mathbb{R}^{(N-1)^2}$ .

Dans ce qui suit, nous cherchons la forme matricielle générale du schéma boîte hermitien pour le problème de Poisson (5.37) en dimension trois.

### 5.6.2 Forme matricielle générale des dérivées au bord

En dimension 1, au chapitre 2 on a vu que toute approximation des dérivées au bord  $U_{x,L}$  et  $U_{x,R}$  en fonction des points de bord  $U_L, U_R$  et des points intérieurs  $U$  vérifient la relation

$$e_1 U_{x,L} + e_{N-1} U_{x,R} = \frac{1}{h} \mathcal{A}U - \mathcal{B}U_x + \frac{1}{h} \mathcal{C}(e_1 U_{x,L} + e_{N-1} U_{x,R}), \quad (5.63)$$

avec  $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathbb{M}_{N-1}(\mathbb{R})$ .

En dimension 2, ceci est généralisé en (3.59).

En dimension 3, la relation (5.63) est généralisée de la façon suivante. Sur chaque segment du maillage du cube  $(a, b)^3$  on applique (5.63). Pour  $j$  et  $k$  fixés entre 0 et  $N$ , on déduit de (5.63) sur le segment  $(a, b) \times \{y_j\} \times \{z_k\}$  la relation suivante

$$e_1 U_{x,L,j,k} + e_{N-1} U_{x,R,j,k} = \frac{1}{h} \mathcal{A}U_{:,j,k} - \mathcal{B}U_{x, :, j, k} + \frac{1}{h} \mathcal{C}(e_1 U_{L,j,k} + e_{N-1} U_{R,j,k}), \quad 0 \leq j, k \leq N. \quad (5.64)$$

Pour  $i$  et  $k$  fixés entre 0 et  $N$ , on déduit de (5.63) sur le segment  $\{x_i\} \times (a, b) \times \{z_k\}$  la relation suivante

$$e_1 U_{y,i,B,k} + e_{N-1} U_{y,i,T,k} = \frac{1}{h} \mathcal{A} U_{i, :, k} - \mathcal{B} U_{y, i, :, k} + \frac{1}{h} \mathcal{C} (e_1 U_{i, B, k} + e_{N-1} U_{i, T, k}), \quad 0 \leq i, k \leq N.. \quad (5.65)$$

Pour  $i$  et  $j$  fixés entre 0 et  $N$ , on déduit de (5.63) sur le segment  $\{x_i\} \times \{y_j\} \times (a, b)$  la relation suivante

$$e_1 U_{z, i, j, F} + e_{N-1} U_{z, i, j, C} = \frac{1}{h} \mathcal{A} U_{i, j, :} - \mathcal{B} U_{z, i, j, :} + \frac{1}{h} \mathcal{C} (e_1 U_{i, j, F} + e_{N-1} U_{i, j, C}), \quad 0 \leq i, j \leq N. \quad (5.66)$$

## 5.7 Dérivées normales

**Lemme 5.7.1.** *Les dérivées normales par rapport à  $x$ ,  $y$ ,  $z$ , sur les huit sommets “LBF”, “RBF”, “LBC”, “RBC”, “LTF”, “RTF”, “LTC”, “RTC” du cube  $\Omega = (a, b)^3$ , vérifient*

$$(i) \left\{ \begin{array}{l} e_1 U_{x, LBF} + e_{N-1} U_{x, RBF} = \frac{1}{h} \mathcal{A} U_{BF} - \mathcal{B} U_{x, BF} + \frac{1}{h} \mathcal{C} (e_1 U_{LBF} + e_{N-1} U_{RBF}), \quad (a) \\ e_1 U_{x, LBC} + e_{N-1} U_{x, RBC} = \frac{1}{h} \mathcal{A} U_{BC} - \mathcal{B} U_{x, BC} + \frac{1}{h} \mathcal{C} (e_1 U_{LBC} + e_{N-1} U_{RBC}), \quad (b) \\ e_1 U_{x, LTF} + e_{N-1} U_{x, RTF} = \frac{1}{h} \mathcal{A} U_{TF} - \mathcal{B} U_{x, TF} + \frac{1}{h} \mathcal{C} (e_1 U_{LTF} + e_{N-1} U_{RTF}), \quad (c) \\ e_1 U_{x, LTC} + e_{N-1} U_{x, RTC} = \frac{1}{h} \mathcal{A} U_{TC} - \mathcal{B} U_{x, TC} + \frac{1}{h} \mathcal{C} (e_1 U_{LTC} + e_{N-1} U_{RTC}). \quad (d) \end{array} \right. \quad (5.67)$$

$$(ii) \left\{ \begin{array}{l} e_1 U_{y, LBF} + e_{N-1} U_{y, LTF} = \frac{1}{h} \mathcal{A} U_{LF} - \mathcal{B} U_{y, LF} + \frac{1}{h} \mathcal{C} (e_1 U_{LBF} + e_{N-1} U_{LTF}), \quad (a) \\ e_1 U_{y, LBC} + e_{N-1} U_{y, LTC} = \frac{1}{h} \mathcal{A} U_{LC} - \mathcal{B} U_{y, LC} + \frac{1}{h} \mathcal{C} (e_1 U_{LBC} + e_{N-1} U_{LTC}), \quad (b) \\ e_1 U_{y, RBF} + e_{N-1} U_{y, RTF} = \frac{1}{h} \mathcal{A} U_{RF} - \mathcal{B} U_{y, RF} + \frac{1}{h} \mathcal{C} (e_1 U_{RBF} + e_{N-1} U_{RTF}), \quad (c) \\ e_1 U_{y, RBC} + e_{N-1} U_{y, RTC} = \frac{1}{h} \mathcal{A} U_{RC} - \mathcal{B} U_{y, RC} + \frac{1}{h} \mathcal{C} (e_1 U_{RBC} + e_{N-1} U_{RTC}). \quad (d) \end{array} \right. \quad (5.68)$$

$$(iii) \left\{ \begin{array}{l} e_1 U_{z, LBF} + e_{N-1} U_{z, LBC} = \frac{1}{h} \mathcal{A} U_{LB} - \mathcal{B} U_{z, LB} + \frac{1}{h} \mathcal{C} (e_1 U_{LBF} + e_{N-1} U_{LBC}), \quad (a) \\ e_1 U_{z, LTF} + e_{N-1} U_{z, LTC} = \frac{1}{h} \mathcal{A} U_{LT} - \mathcal{B} U_{z, LT} + \frac{1}{h} \mathcal{C} (e_1 U_{LTF} + e_{N-1} U_{LTC}), \quad (b) \\ e_1 U_{z, RBF} + e_{N-1} U_{z, RBC} = \frac{1}{h} \mathcal{A} U_{RB} - \mathcal{B} U_{z, RB} + \frac{1}{h} \mathcal{C} (e_1 U_{RBF} + e_{N-1} U_{RBC}), \quad (c) \\ e_1 U_{z, RTF} + e_{N-1} U_{z, RTC} = \frac{1}{h} \mathcal{A} U_{RT} - \mathcal{B} U_{z, RT} + \frac{1}{h} \mathcal{C} (e_1 U_{RTF} + e_{N-1} U_{RTC}). \quad (d) \end{array} \right. \quad (5.69)$$



Les vecteurs  $U_{y,B}, U_{y,T} \in \mathbb{R}^{(N-1)^2}$  des dérivées normales par rapport à  $y$  sur les deux faces “bas” ( $y = a$ ) et “haut” ( $y = b$ ) du cube  $\Omega = (a, b)^3$  vérifient les relations

$$\begin{cases} (I \otimes e_1 \otimes I)U_{y,B} + (I \otimes e_{N-1} \otimes I)U_{y,T} &= \frac{1}{h}(I \otimes \mathcal{A} \otimes I)U - (I \otimes \mathcal{B} \otimes I)U_y \\ &+ \frac{1}{h}(I \otimes \mathcal{C} \otimes I)\left((I \otimes e_1 \otimes I)U_B + (I \otimes e_{N-1} \otimes I)U_T\right), \end{cases} \quad (5.74)$$

Les vecteurs  $U_{z,F}, U_{z,C} \in \mathbb{R}^{(N-1)^2}$  des dérivées normales par rapport à  $z$  sur les deux faces “arrière” ( $z = a$ ) et “avant” ( $z = b$ ) du cube  $\Omega = (a, b)^3$  vérifient les relations

$$\begin{cases} (I \otimes I \otimes e_1)U_{z,F} + (I \otimes I \otimes e_{N-1})U_{z,C} &= \frac{1}{h}(I \otimes I \otimes \mathcal{A})U - (I \otimes I \otimes \mathcal{B})U_z \\ &+ \frac{1}{h}(I \otimes I \otimes \mathcal{C})\left((I \otimes I \otimes e_1)U_F + (I \otimes I \otimes e_{N-1})U_C\right), \end{cases} \quad (5.75)$$

**Preuve.** D’après les notations (5.17) on a

$$\begin{cases} U_{x,L} = \text{vec}_2[u_{x,0,j,k}]_{1 \leq j,k \leq N-1} = \sum_{j,k=1}^{N-1} (e_j \otimes e_k)U_{x,L,j,k}, \\ U_{x,R} = \text{vec}_2[u_{x,N,j,k}]_{1 \leq j,k \leq N-1} = \sum_{j,k=1}^{N-1} (e_j \otimes e_k)U_{x,R,j,k}. \end{cases} \quad (5.76)$$

Donc

$$\begin{aligned} (e_1 \otimes I \otimes I)U_{x,L} + (e_{N-1} \otimes I \otimes I)U_{x,R} &= (e_1 \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k)U_{x,L,j,k} \\ &+ (e_{N-1} \otimes I \otimes I) \sum_{j,k=1}^{N-1} (e_j \otimes e_k)U_{x,R,j,k} \\ &= \sum_{j,k=1}^{N-1} (e_1 \otimes e_j \otimes e_k)U_{x,L,j,k} + \sum_{j,k=1}^{N-1} (e_{N-1} \otimes e_j \otimes e_k)U_{x,R,j,k} \\ &= \sum_{j,k=1}^{N-1} (e_1 U_{x,L,j,k} \otimes e_j \otimes e_k) + \sum_{j,k=1}^{N-1} (e_{N-1} U_{x,R,j,k} \otimes e_j \otimes e_k) \\ &= \sum_{j,k=1}^{N-1} \left( (e_1 U_{x,L,j,k} + e_{N-1} U_{x,R,j,k}) \otimes e_j \otimes e_k \right). \end{aligned}$$

En utilisant (5.64),

$$\begin{aligned}
(e_1 \otimes I \otimes I)U_{x,L} + (e_{N-1} \otimes I \otimes I)U_{x,R} &= \sum_{j,k=1}^{N-1} \left( \frac{1}{h} \mathcal{A} U_{:,j,k} \otimes e_j \otimes e_k \right) - \sum_{j,k=1}^{N-1} \left( \mathcal{B} U_{x,j,k} \otimes e_j \otimes e_k \right) \\
&+ \frac{1}{h} \left( \mathcal{C} (e_1 U_{L,j,k} + e_{N-1} U_{R,j,k}) \otimes e_j \otimes e_k \right) \\
&= \frac{1}{h} (\mathcal{A} \otimes I \otimes I)U - (\mathcal{B} \otimes I \otimes I)U_x \\
&+ \frac{1}{h} (\mathcal{C} \otimes I \otimes I) \left( (e_1 \otimes I \otimes I)U_L + (e_{N-1} \otimes I \otimes I)U_R \right),
\end{aligned}$$

d'où (5.73). De la même manière on démontre (5.74) et (5.75). ■

## 5.8 Dérivées tangentielles

**Lemme 5.8.1.** *Les vecteurs  $U_{x,BF}$ ,  $U_{x,BC}$ ,  $U_{x,TF}$ ,  $U_{x,TC} \in \mathbb{R}^{(N-1)}$ , dérivées tangentielles par rapport à  $x$  sur les quatres arêtes “BF”, “BC”, “TF”, “TC” du cube  $\Omega = (a, b)^3$ , vérifient*

$$\left\{ \begin{array}{l} U_{x,BF} = \frac{1}{h} \mathcal{D}U_{BF} + \frac{1}{h} \mathcal{E}(e_1 U_{LBF} + e_{N-1} U_{RBF}), \quad (a) \\ U_{x,BC} = \frac{1}{h} \mathcal{D}U_{BC} + \frac{1}{h} \mathcal{E}(e_1 U_{LBC} + e_{N-1} U_{RBC}), \quad (b) \\ U_{x,TF} = \frac{1}{h} \mathcal{D}U_{TF} + \frac{1}{h} \mathcal{E}(e_1 U_{LTF} + e_{N-1} U_{RTF}), \quad (c) \\ U_{x,TC} = \frac{1}{h} \mathcal{D}U_{TC} + \frac{1}{h} \mathcal{E}(e_1 U_{LTC} + e_{N-1} U_{RTC}). \quad (d) \end{array} \right. \quad (5.77)$$

Les vecteurs  $U_{y,LF}$ ,  $U_{y,LC}$ ,  $U_{y,RF}$ ,  $U_{y,RC} \in \mathbb{R}^{(N-1)}$ , dérivées tangentielles par rapport à  $y$  sur les quatres arêtes “LF”, “LC”, “RF”, “RC” du cube  $\Omega = (a, b)^3$ , vérifient

$$\left\{ \begin{array}{l} U_{y,LF} = \frac{1}{h} \mathcal{D}U_{LF} + \frac{1}{h} \mathcal{E}(e_1 U_{LBF} + e_{N-1} U_{LTF}), \quad (a) \\ U_{y,LC} = \frac{1}{h} \mathcal{D}U_{LC} + \frac{1}{h} \mathcal{E}(e_1 U_{LBC} + e_{N-1} U_{LTC}), \quad (b) \\ U_{y,RF} = \frac{1}{h} \mathcal{D}U_{RF} + \frac{1}{h} \mathcal{E}(e_1 U_{RBF} + e_{N-1} U_{RTF}), \quad (c) \\ U_{y,RC} = \frac{1}{h} \mathcal{D}U_{RC} + \frac{1}{h} \mathcal{E}(e_1 U_{RBC} + e_{N-1} U_{RTC}). \quad (d) \end{array} \right. \quad (5.78)$$

Les vecteurs  $U_{z,LB}, U_{z,LT}, U_{z,RB}, U_{z,RT} \in \mathbb{R}^{(N-1)}$ , dérivées tangentielles par rapport à  $z$  sur les quatres arêtes “LB”, “LT”, “RB”, “RT” du cube  $\Omega = (a, b)^3$ , vérifient

$$\begin{cases} U_{z,LB} = \frac{1}{h}\mathcal{D}U_{LB} + \frac{1}{h}\mathcal{E}(e_1U_{LBF} + e_{N-1}U_{LBC}), & (a) \\ U_{z,LT} = \frac{1}{h}\mathcal{D}U_{LT} + \frac{1}{h}\mathcal{E}(e_1U_{LTF} + e_{N-1}U_{LTC}), & (b) \\ U_{z,RB} = \frac{1}{h}\mathcal{D}U_{RB} + \frac{1}{h}\mathcal{E}(e_1U_{RBF} + e_{N-1}U_{RBC}), & (c) \\ U_{z,RT} = \frac{1}{h}\mathcal{D}U_{RT} + \frac{1}{h}\mathcal{E}(e_1U_{RTF} + e_{N-1}U_{RTC}). & (d) \end{cases} \quad (5.79)$$

**Preuve.** D’après (5.54)<sub>b</sub> on a

$$\sigma_x u_{x,i,j,k} = \delta_x u_{i,j,k}, \quad (5.80)$$

pour  $j = 0, k = 0$ , on obtient

$$\sigma_x u_{x,i,0,0} = \delta_x u_{i,0,0}. \quad (5.81)$$

Donc

$$\frac{1}{6}u_{x,i-1,0,0} + \frac{2}{3}u_{x,i,0,0} + \frac{1}{6}u_{x,i+1,0,0} = \frac{u_{i+1} - u_{i-1}}{2h}, \quad 1 \leq i \leq N-1. \quad (5.82)$$

Matriciellement, on obtient

$$\frac{1}{6}e_1U_{x,LBF} + P_s U_{x,BF} + \frac{1}{6}e_{N-1}U_{x,RBF} = -\frac{1}{2h}e_1U_{LBF} + \frac{K}{2h}U_{BF} + \frac{1}{2h}e_{N-1}U_{RBF}. \quad (5.83)$$

En utilisant la relation (5.67)<sub>a</sub>, (5.83) devient

$$(P_s - \frac{1}{6}\mathcal{B})U_{x,BF} = \frac{1}{2h}(K - \frac{1}{3}\mathcal{A})U_{BF} + \frac{1}{2h}(F_2 - \frac{1}{3}\mathcal{C})(e_1U_{LBF} + e_{N-1}U_{RBF}). \quad (5.84)$$

En utilisant les matrices  $\mathcal{D}$  et  $\mathcal{E}$  en (2.113) on obtient

$$U_{x,BF} = \frac{1}{2h}\mathcal{D}U_{BF} + \frac{1}{2h}\mathcal{E}(e_1U_{LBF} + e_{N-1}U_{RBF}), \quad (5.85)$$

d’où (5.77)<sub>a</sub>. De la même manière on démontre (5.77)<sub>b,c,d</sub>, (5.78) et (5.79). ■

**Lemme 5.8.2.** Les vecteurs  $U_{x,B}, U_{x,T}, U_{x,F}, U_{x,C} \in \mathbb{R}^{(N-1)^2}$ , dérivées tangentielles par rapport à  $x$  sur les quatres faces “B”, “T”, “F”, “C” du cube  $\Omega = (a, b)^3$ , vérifient

$$\begin{cases} U_{x,B} = \frac{1}{h}(\mathcal{D} \otimes I)U_B + \frac{1}{h}(\mathcal{E} \otimes I)\left((e_1 \otimes I)U_{LB} + (e_{N-1} \otimes I)U_{RB}\right), & (a) \\ U_{x,T} = \frac{1}{h}(\mathcal{D} \otimes I)U_T + \frac{1}{h}(\mathcal{E} \otimes I)\left((e_1 \otimes I)U_{LT} + (e_{N-1} \otimes I)U_{RT}\right), & (b) \\ U_{x,F} = \frac{1}{h}(\mathcal{D} \otimes I)U_F + \frac{1}{h}(\mathcal{E} \otimes I)\left((e_1 \otimes I)U_{LF} + (e_{N-1} \otimes I)U_{RF}\right), & (c) \\ U_{x,C} = \frac{1}{h}(\mathcal{D} \otimes I)U_C + \frac{1}{h}(\mathcal{E} \otimes I)\left((e_1 \otimes I)U_{LC} + (e_{N-1} \otimes I)U_{RC}\right). & (d) \end{cases} \quad (5.86)$$

Les vecteurs  $U_{y,L}, U_{y,R}, U_{y,F}, U_{y,C} \in \mathbb{R}^{(N-1)^2}$ , dérivées tangentielles par rapport à  $y$  sur les quatres faces “L”, “R”, “F”, “C” du cube  $\Omega = (a, b)^3$ , vérifient

$$\begin{cases} U_{y,L} = \frac{1}{h}(\mathcal{D} \otimes I)U_L + \frac{1}{h}(\mathcal{E} \otimes I)\left((e_1 \otimes I)U_{LB} + (e_{N-1} \otimes I)U_{LT}\right), & (a) \\ U_{y,R} = \frac{1}{h}(\mathcal{D} \otimes I)U_R + \frac{1}{h}(\mathcal{E} \otimes I)\left((e_1 \otimes I)U_{RB} + (e_{N-1} \otimes I)U_{RT}\right), & (b) \\ U_{y,F} = \frac{1}{h}(I \otimes \mathcal{D})U_F + \frac{1}{h}(I \otimes \mathcal{E})\left((I \otimes e_1)U_{BF} + (I \otimes e_{N-1})U_{TF}\right), & (c) \\ U_{y,C} = \frac{1}{h}(I \otimes \mathcal{D})U_C + \frac{1}{h}(I \otimes \mathcal{E})\left((I \otimes e_1)U_{BC} + (I \otimes e_{N-1})U_{TC}\right). & (d) \end{cases} \quad (5.87)$$

Les vecteurs  $U_{z,L}, U_{z,R}, U_{z,B}, U_{z,T} \in \mathbb{R}^{(N-1)^2}$ , dérivées tangentielles par rapport à  $z$  sur les quatres faces “L”, “R”, “B”, “T” du cube  $\Omega = (a, b)^3$ , vérifient

$$\begin{cases} U_{z,L} = \frac{1}{h}(I \otimes \mathcal{D})U_L + \frac{1}{h}(I \otimes \mathcal{E})\left((I \otimes e_1)U_{LF} + (I \otimes e_{N-1})U_{LC}\right), & (a) \\ U_{z,R} = \frac{1}{h}(I \otimes \mathcal{D})U_R + \frac{1}{h}(I \otimes \mathcal{E})\left((I \otimes e_1)U_{RF} + (I \otimes e_{N-1})U_{RC}\right), & (b) \\ U_{z,B} = \frac{1}{h}(I \otimes \mathcal{D})U_B + \frac{1}{h}(I \otimes \mathcal{E})\left((I \otimes e_1)U_{BF} + (I \otimes e_{N-1})U_{BC}\right), & (c) \\ U_{z,T} = \frac{1}{h}(I \otimes \mathcal{D})U_T + \frac{1}{h}(I \otimes \mathcal{E})\left((I \otimes e_1)U_{TF} + (I \otimes e_{N-1})U_{TC}\right). & (d) \end{cases} \quad (5.88)$$

**Preuve.** D’après (5.54)<sub>b</sub> on a

$$\sigma_x u_{x,i,j,k} = \delta_x u_{i,j,k}, \quad 1 \leq i \leq N-1, \quad 0 \leq j, k \leq N, \quad (5.89)$$

ceci est équivalent à

$$\frac{1}{6}u_{x,i-1,j,k} + \frac{2}{3}u_{x,i,j,k} + \frac{1}{6}u_{x,i+1,j,k} = \frac{u_{i+1,j,k} - u_{i-1,j,k}}{2h}. \quad (5.90)$$

Matriciellement, l'équation (5.90) s'écrit

$$\frac{1}{6}e_1 U_{x,L,j,k} + P_s U_{x,:,j,k} + \frac{1}{6}e_{N-1} U_{x,R,j,k} = -\frac{1}{2h}e_1 U_{L,j,k} + \frac{1}{2h}U_{:,j,k} + \frac{1}{2h}e_{N-1} U_{R,j,k}, \quad (5.91)$$

ce qui donne

$$U_{x,:,j,k} = \frac{1}{h}\mathcal{D}U_{:,j,k} + \frac{1}{h}\mathcal{E}(e_1 U_{L,j,k} + e_{N-1} U_{R,j,k}). \quad (5.92)$$

On a par définition de l'opérateur  $\text{vec}_2$  en (3.12)

$$\begin{aligned} U_{x,B} &= \text{vec}_2[u_{x,i,0,k}]_{1 \leq i,k \leq N-1} \\ &= \sum_{i,k=1}^{N-1} (e_i \otimes e_k) u_{x,i,0,k} \\ &= \sum_{i,k=1}^{N-1} (U_{x,:,0,k} \otimes e_k). \end{aligned}$$

Or d'après (5.92) pour  $j = 0$  on a

$$U_{x,:,0,k} = \frac{1}{h}\mathcal{D}U_{:,0,k} + \frac{1}{h}\mathcal{E}(e_1 U_{L,0,k} + e_{N-1} U_{R,0,k}). \quad (5.93)$$

Donc

$$\begin{aligned} U_{x,B} &= \sum_{k=1}^{N-1} (U_{x,0,:,k} \otimes e_k) \\ &= \frac{1}{h} \sum_{k=1}^{N-1} (\mathcal{D}U_{:,0,k} \otimes e_k) + \frac{1}{h} \sum_{k=1}^{N-1} (\mathcal{E}(e_1 U_{L,0,k} + e_{N-1} U_{R,0,k}) \otimes e_k) \\ &= \frac{1}{h} (\mathcal{D} \otimes I) \sum_{k=1}^{N-1} (U_{:,0,k} \otimes e_k) + \frac{1}{h} (\mathcal{E} \otimes I) \sum_{k=1}^{N-1} (e_1 U_{L,0,k} + e_{N-1} U_{R,0,k}) \otimes e_k \\ &= \frac{1}{h} (\mathcal{D} \otimes I) U_B + \frac{1}{h} (\mathcal{E} \otimes I) \left( \sum_{k=1}^{N-1} e_1 U_{L,0,k} \otimes e_k + \sum_{k=1}^{N-1} e_{N-1} U_{R,0,k} \otimes e_k \right) \\ &= \frac{1}{h} (\mathcal{D} \otimes I) U_B + \frac{1}{h} (\mathcal{E} \otimes I) \left( (e_1 \otimes I) U_{LB} + (e_{N-1} \otimes I) U_{RB} \right), \end{aligned}$$

d'où (5.86)<sub>a</sub>. De la même manière on démontre (5.86)<sub>b,c,d</sub>, (5.87) et (5.88). ■



## 5.9 Expression du gradient $(U_x, U_y, U_z)$ en fonction de $U$

**Proposition 5.1.** • Le vecteur  $U_x \in \mathbb{R}^{(N-1)^3}$  s'exprime en fonction de  $U, U_L, U_R$  par

$$U_x = \frac{1}{h}(\mathcal{D} \otimes I \otimes I)U + \frac{1}{h}(\mathcal{E} \otimes I \otimes I) \left( (e_1 \otimes I \otimes I)U_L + (e_{N-1} \otimes I \otimes I)U_R \right). \quad (5.94)$$

• Le vecteur  $U_y \in \mathbb{R}^{(N-1)^3}$  s'exprime en fonction de  $U, U_B, U_T$  par

$$U_y = \frac{1}{h}(I \otimes \mathcal{D} \otimes I)U + \frac{1}{h}(I \otimes \mathcal{E} \otimes I) \left( (I \otimes e_1 \otimes I)U_B + (I \otimes e_{N-1} \otimes I)U_T \right). \quad (5.95)$$

• Le vecteur  $U_z \in \mathbb{R}^{(N-1)^3}$  s'exprime en fonction de  $U, U_F, U_C$  par

$$U_z = \frac{1}{h}(I \otimes I \otimes \mathcal{D})U + \frac{1}{h}(I \otimes I \otimes \mathcal{E}) \left( (I \otimes I \otimes e_1)U_F + (I \otimes I \otimes e_{N-1})U_C \right). \quad (5.96)$$

**Preuve.** D'après (5.54)<sub>b</sub> on a

$$\sigma_x u_{x,i,j,k} = \delta_x u_{i,j,k}, \quad 1 \leq i \leq N-1, 0 \leq j, k \leq N. \quad (5.97)$$

Ceci est équivalent à

$$\frac{1}{6}u_{x,i-1,j,k} + \frac{2}{3}u_{x,i,j,k} + \frac{1}{6}u_{x,i+1,j,k} = \frac{u_{i+1,j,k} - u_{i-1,j,k}}{2h}. \quad (5.98)$$

Matriciellement, on obtient

$$\frac{1}{6}e_1 U_{x,L,j,k} + P_s U_{x,:,j,k} + \frac{1}{6}e_{N-1} U_{x,R,j,k} = -\frac{1}{2h}e_1 U_{L,j,k} + \frac{1}{2h}U_{:,j,k} + \frac{1}{2h}e_{N-1} U_{R,j,k}, \quad (5.99)$$

avec  $P_s$  la matrice donnée en (1.67). En utilisant les matrices données en (2.113), la relation (5.99) donne

$$U_{x,:,j,k} = \frac{1}{h}\mathcal{D}U_{:,j,k} + \frac{1}{h}\mathcal{E}(e_1 U_{L,j,k} + e_{N-1} U_{R,j,k}). \quad (5.100)$$

On a

$$\begin{aligned}
U_x &= \text{vec}_3[u_{x,i,j,k}]_{1 \leq i,j,k \leq N-1} \\
&= \sum_{j,k=1}^{N-1} (U_{x,.,j,k} \otimes e_j \otimes e_k) \\
&= \sum_{j,k=1}^{N-1} \left( \frac{1}{h} \mathcal{D}U_{:,j,k} + \frac{1}{h} \mathcal{E}(e_1 U_{L,j,k} + e_{N-1} U_{R,j,k}) \right) (e_j \otimes e_k) \\
&= \frac{1}{h} \sum_{j,k=1}^{N-1} (\mathcal{D}U_{:,j,k} \otimes e_j \otimes e_k) + \frac{1}{h} \sum_{j,k=1}^{N-1} \left( \mathcal{E}(e_1 U_{L,j,k} + e_{N-1} U_{R,j,k}) \right) (e_j \otimes e_k) \\
&= \frac{1}{h} (\mathcal{D} \otimes I \otimes I) \sum_{j,k=1}^{N-1} (U_{:,j,k} \otimes e_j \otimes e_k) \\
&\quad + \frac{1}{h} (\mathcal{E} \otimes I \otimes I) \sum_{j,k=1}^{N-1} \left( (e_1 \otimes e_j \otimes e_k) U_{L,j,k} + (e_{N-1} \otimes e_j \otimes e_k) U_{R,j,k} \right) \\
&= \frac{1}{h} (\mathcal{D} \otimes I \otimes I) U + \frac{1}{h} (\mathcal{E} \otimes I \otimes I) \left( (e_1 \otimes I \otimes I) U_L + (e_{N-1} \otimes I \otimes I) U_R \right),
\end{aligned}$$

d'où (5.94). De la même manière on démontre (5.95) et (5.95).  $\blacksquare$

## 5.10 Forme matricielle générale du schéma boîte hermitien

On a mentionné dans la section 5.6 que le système (5.54) constitue les relations de base du schéma HB en trois dimensions, donc afin d'obtenir une forme matricielle du schéma, on traduit matriciellement les équations de ce système.

**Proposition 5.2.** *La fonction de grille  $(\sigma_z \circ \sigma_y \circ \delta_x)u_x$  s'écrit matriciellement en fonction de  $U = \text{vec}_3(u)$  et des données de Dirichlet au bord de  $\Omega = (a, b)^3$  sous la forme*

$$\text{vec}_3 \left( (\sigma_z \circ \sigma_y \circ \delta_x) u_x \right) = -\frac{1}{h^2} (\mathcal{H} \otimes P_s \otimes P_s) U - G_x, \quad (5.101)$$

où  $G_x \in \mathbb{R}^{(N-1)^3}$  est le vecteur

$$\left\{ \begin{aligned} G_x &= \frac{1}{h^2}(\mathcal{G} \otimes P_s \otimes P_s) \left( (e_1 \otimes I \otimes I)U_L + (e_{N-1} \otimes I \otimes I)U_R \right) \\ &+ \frac{1}{6h^2}(I \otimes e_1 \otimes P_s) \left\{ (\mathcal{H} \otimes I)U_B + (\mathcal{G} \otimes I)(e_1 \otimes U_{LB} + e_{N-1} \otimes U_{RB}) \right\} \\ &+ \frac{1}{6h^2}(I \otimes e_{N-1} \otimes P_s) \left\{ (\mathcal{H} \otimes I)U_T + (\mathcal{G} \otimes I)(e_1 \otimes U_{LT} + e_{N-1} \otimes U_{RT}) \right\} \\ &+ \frac{1}{6h^2}(I \otimes I \otimes e_1) \left\{ (\mathcal{H} \otimes P_s)U_F + (\mathcal{G} \otimes P_s)(e_1 \otimes U_{LF} + e_{N-1} \otimes U_{RF}) \right\} \\ &+ \frac{1}{6h^2}(I \otimes I \otimes e_{N-1}) \left\{ (\mathcal{H} \otimes P_s)U_C + (\mathcal{G} \otimes P_s)(e_1 \otimes U_{LC} + e_{N-1} \otimes U_{RC}) \right\} \\ &+ \frac{1}{36h^2}(I \otimes I \otimes e_1)(I \otimes e_1) \left( \mathcal{H}U_{BF} + \mathcal{G}(e_1U_{LBF} + e_{N-1}U_{RBF}) \right) \\ &+ \frac{1}{36h^2}(I \otimes I \otimes e_1)(I \otimes e_{N-1}) \left( \mathcal{H}U_{TF} + \mathcal{G}(e_1U_{LTF} + e_{N-1}U_{RTF}) \right) \\ &+ \frac{1}{36h^2}(I \otimes I \otimes e_{N-1})(I \otimes e_1) \left( \mathcal{H}U_{BC} + \mathcal{G}(e_1U_{LBC} + e_{N-1}U_{RBC}) \right) \\ &+ \frac{1}{36h^2}(I \otimes I \otimes e_{N-1})(I \otimes e_{N-1}) \left( \mathcal{H}U_{TC} + \mathcal{G}(e_1U_{LTC} + e_{N-1}U_{RTC}) \right). \end{aligned} \right. \quad (5.102)$$

Pour démontrer la proposition 5.2 on utilise la relation suivante

$$\left\{ \begin{aligned} \text{vec}_3 \left( (\sigma_z \circ \sigma_y \circ \delta_x)u_x \right) &= (I \otimes I \otimes P_s) \text{vec}_3 \left( (\sigma_y \circ \delta_x)u_x \right) \\ &+ \frac{1}{6}(I \otimes I \otimes e_1) \text{vec}_2 \left( (\sigma_y \circ \delta_x)u_x \right)_F \\ &+ \frac{1}{6}(I \otimes I \otimes e_{N-1}) \text{vec}_2 \left( (\sigma_z \circ \delta_x)u_x \right)_C. \end{aligned} \right. \quad (5.103)$$

Dans le lemme 5.10.1, on donne la forme matricielle du vecteur  $\text{vec}_3 \left( (\sigma_y \circ \delta_x)u_x \right) \in \mathbb{R}^{(N-1)^3}$ .

Dans le lemme 5.10.2, on donne les formes matricielles des vecteurs  $\text{vec}_2 \left( (\sigma_z \circ \delta_x)u_x \right)_F$  et  $\text{vec}_2 \left( (\sigma_z \circ \delta_x)u_x \right)_C \in \mathbb{R}^{(N-1)^2}$ .

**Lemme 5.10.1.** *Le vecteur  $\text{vec}_3((\sigma_y \circ \delta_x)u_x) \in \mathbb{R}^{(N-1)^3}$  s'écrit matriciellement sous la forme*

$$\left\{ \begin{array}{l} \text{vec}_3((\sigma_y \circ \delta_x)u_x) = (K \otimes P_s \otimes I)U_x \\ \quad + \frac{1}{2h}(F_2 \otimes P_s \otimes I)\left((e_1 \otimes I \otimes I)U_{x,L} + (e_{N-1} \otimes I \otimes I)U_{x,R}\right) \\ \quad - \frac{1}{6h^2}(I \otimes e_1 \otimes I)\left((\mathcal{H} \otimes I)U_B + (\mathcal{G} \otimes I)(e_1 \otimes U_{LB} + e_{N-1} \otimes U_{RB})\right) \\ \quad - \frac{1}{6h^2}(I \otimes e_{N-1} \otimes I)\left((\mathcal{H} \otimes I)U_T + (\mathcal{G} \otimes I)(e_1 \otimes U_{LT} + e_{N-1} \otimes U_{RT})\right). \end{array} \right. \quad (5.104)$$

**Preuve.** En posant  $v = \delta_x u_x$ , on déduit de (5.30)<sub>b</sub> que

$$\begin{aligned} \text{vec}_3(\sigma_y \circ \delta_x)u_x &= (I \otimes P_s \otimes I)\text{vec}_3(\delta_x u_x) \\ &\quad + \frac{1}{6}(I \otimes e_1 \otimes I)\text{vec}_2(\delta_x u_x)_B + \frac{1}{6}(I \otimes e_{N-1} \otimes I)\text{vec}_2(\delta_x u_x)_T. \end{aligned}$$

Pour  $v = u_x$ , la relation (5.23)<sub>a</sub> donne

$$\text{vec}_3(\delta_x u_x) = \frac{1}{2h}(K \otimes I \otimes I)U_x + \frac{1}{2h}(F_2 \otimes I \otimes I)\left((e_1 \otimes I \otimes I)U_{x,L} + (e_{N-1} \otimes I \otimes I)U_{x,R}\right).$$

Donc,

$$\left\{ \begin{array}{l} \text{vec}_3(\sigma_y \circ \delta_x)u_x = \frac{1}{2h}(K \otimes P_s \otimes I)U_x \\ \quad + \frac{1}{2h}(F_2 \otimes P_s \otimes I)\left((e_1 \otimes I \otimes I)U_{x,L} + (e_{N-1} \otimes I \otimes I)U_{x,R}\right) \\ \quad + \frac{1}{6}(I \otimes e_1 \otimes I)\text{vec}_2(\delta_x u_x)_B + \frac{1}{6}(I \otimes e_{N-1} \otimes I)\text{vec}_2(\delta_x u_x)_T. \end{array} \right. \quad (5.105)$$

Or pour  $v = u_x$ , (5.24)<sub>a</sub> donne

$$\text{vec}_2(\delta_x u_x)_B = \frac{1}{2h}(K \otimes I)U_{x,B} + \frac{1}{2h}(F_2 \otimes I)\left((e_1 \otimes I)U_{x,LB} + (e_{N-1} \otimes I)U_{x,RB}\right). \quad (5.106)$$

En remplaçant dans (5.106) les vecteurs  $U_{x,B}$  et  $(e_1 \otimes I)U_{x,LB} + (e_{N-1} \otimes I)U_{x,RB}$  par ses expressions en (5.86)<sub>a</sub> et (5.70)<sub>a</sub> respectivement, on obtient

$$\text{vec}_2(\delta_x u_x)_B = -\frac{1}{h^2}(\mathcal{H} \otimes I)U_B - \frac{1}{h^2}(\mathcal{G} \otimes I)(e_1 \otimes U_{LB} + e_{N-1} \otimes U_{RB}), \quad (5.107)$$

avec  $\mathcal{H}$ ,  $\mathcal{G}$  les matrices données en (2.116). De la même manière on obtient

$$\text{vec}_2(\delta_x u_x)_T = -\frac{1}{h^2}(\mathcal{H} \otimes I)U_T - \frac{1}{h^2}(\mathcal{G} \otimes I)(e_1 \otimes U_{LT} + e_{N-1} \otimes U_{RT}). \quad (5.108)$$

Enfin, en remplaçant dans (5.105) les vecteurs  $\text{vec}_2(\delta_x u_x)_B$  et  $\text{vec}_2(\delta_x u_x)_T$  par ses expressions en (5.107) et (5.108) respectivement, on conclut (5.104). ■

**Lemme 5.10.2.** *Les vecteurs  $\text{vec}_2(\sigma_y \circ \delta_x u_x)_F$  et  $\text{vec}_2(\sigma_y \circ \delta_x u_x)_C \in \mathbb{R}^{(N-1)^2}$  s'écrivent matriciellement sous la forme*

$$\left\{ \begin{array}{l} \text{vec}_2(\sigma_y \circ \delta_x u_x)_F = -\frac{1}{h^2} \left( (\mathcal{H} \otimes P_s) U_F + (\mathcal{G} \otimes I)(e_1 \otimes U_{LF} + e_{N-1} \otimes U_{RF}) \right) \\ - \frac{1}{6h^2} (I \otimes e_1) \left( \mathcal{H} U_{BF} + \mathcal{G}(e_1 U_{LBF} + e_{N-1} U_{RBF}) \right) \\ - \frac{1}{6h^2} (I \otimes e_{N-1}) \left( \mathcal{H} U_{TF} + \mathcal{G}(e_1 U_{LTF} + e_{N-1} U_{RTF}) \right), \end{array} \right. \quad (5.109)$$

et

$$\left\{ \begin{array}{l} \text{vec}_2(\sigma_y \circ \delta_x u_x)_C = -\frac{1}{h^2} \left( (\mathcal{H} \otimes P_s) U_C + (\mathcal{G} \otimes I)(e_1 \otimes U_{LC} + e_{N-1} \otimes U_{RC}) \right) \\ - \frac{1}{6h^2} (I \otimes e_1) \left( \mathcal{H} U_{BC} + \mathcal{G}(e_1 U_{LBC} + e_{N-1} U_{RBC}) \right) \\ - \frac{1}{6h^2} (I \otimes e_{N-1}) \left( \mathcal{H} U_{TC} + \mathcal{G}(e_1 U_{LTC} + e_{N-1} U_{RTC}) \right). \end{array} \right. \quad (5.110)$$

**Preuve.** Pour  $v = \delta_x u_x$ , la relation (5.32)<sub>c</sub> donne

$$\text{vec}_2(\sigma_y \circ \delta_x u_x)_F = (I \otimes P_s) \text{vec}_2(\delta_x u_x)_F + \frac{1}{6} (I \otimes e_1) \text{vec}_1(\delta_x u_x)_{BF} + \frac{1}{6} (I \otimes e_{N-1}) \text{vec}_1(\delta_x u_x)_{TF}. \quad (5.111)$$

Pour  $v = u_x$ , la relation (5.24)<sub>c</sub> donne

$$\text{vec}_2(\delta_x u_x)_F = \frac{1}{2h} (K \otimes I) U_{x,F} + \frac{1}{2h} (F_2 \otimes I)(e_1 \otimes U_{x,LF} + e_{N-1} \otimes U_{x,RF}). \quad (5.112)$$

En remplaçant dans (5.112) les vecteurs  $U_{x,F}$  et  $e_1 \otimes U_{x,LF} + e_{N-1} \otimes U_{x,RF} \in \mathbb{R}^{(N-1)^2}$  par ses formes en (5.86) et (5.70) respectivement, on obtient

$$\text{vec}_2(\delta_x u_x)_F = -\frac{1}{h^2} \left( (\mathcal{H} \otimes P_s) U_F + (\mathcal{G} \otimes P_s)(e_1 \otimes U_{LF} + e_{N-1} \otimes U_{RF}) \right), \quad (5.113)$$

avec  $\mathcal{H}$ ,  $\mathcal{G}$  les matrices données en (2.116). D'autre part, pour  $v = \delta_x u_x$  on déduit de (5.27)<sub>a</sub>

$$\text{vec}_1(\delta_x u_x)_{BF} = \frac{1}{2h} K U_{x,BF} + \frac{1}{2h} (e_1 U_{x,LBF} + e_{N-1} U_{x,RBF}). \quad (5.114)$$

En remplaçant dans (5.114) les vecteurs  $U_{x,BF}$  et  $e_1U_{x,LBF} + e_{N-1}U_{x,RBF} \in \mathbb{R}^{(N-1)}$  par ses formes données en (5.77) et (5.67)<sub>a</sub> respectivement, on obtient

$$\text{vec}_1(\delta_x u_x)_{BF} = -\frac{1}{\hbar^2} \mathcal{H}U_{BF} - \frac{1}{\hbar^2} \mathcal{G}(e_1U_{LBF} + e_{N-1}U_{RBF}), \quad (5.115)$$

avec  $\mathcal{H}$ ,  $\mathcal{G}$  les matrices données en (2.116). De la même manière on obtient

$$\text{vec}_1(\delta_x u_x)_{TF} = -\frac{1}{\hbar^2} \mathcal{H}U_{TF} - \frac{1}{\hbar^2} \mathcal{G}(e_1U_{LTF} + e_{N-1}U_{RTF}). \quad (5.116)$$

En utilisant (5.113), (5.115), (5.116) dans (5.111) on conclut (5.109). De la même manière on démontre (5.110). ■

**Preuve de la proposition 5.2.** D'après (5.30)<sub>c</sub> on a

$$\left\{ \begin{array}{l} \text{vec}_3((\sigma_z \circ \sigma_y \circ \delta_x)u_x) \\ \text{vec}_3((\sigma_y \circ \delta_x)u_x) \\ \text{vec}_2(\sigma_y \circ \delta_x u_x)_F \\ \text{vec}_2(\sigma_y \circ \delta_x u_x)_C \end{array} \right. = \begin{array}{l} (I \otimes I \otimes P_s) \text{vec}_3((\sigma_y \circ \delta_x)u_x) \\ + \frac{1}{6}(I \otimes I \otimes e_1) \text{vec}_2(\sigma_y \circ \delta_x u_x)_F \\ + \frac{1}{6}(I \otimes I \otimes e_{N-1}) \text{vec}_2(\sigma_y \circ \delta_x u_x)_C. \end{array} \quad (5.117)$$

En remplaçant les vecteurs  $(e_1 \otimes I \otimes I)U_{x,L} + (e_{N-1} \otimes I \otimes I)U_{x,R}$  et le vecteur  $U_x$  par les expressions données en (5.73) et (5.94) respectivement et les vecteurs  $\text{vec}_2(\sigma_y \circ \delta_x u_x)_F$  et  $\text{vec}_2(\sigma_y \circ \delta_x u_x)_C$  par les expressions obtenues au lemme 5.10.2 on obtient (5.101). ■

**Proposition 5.3.** Le vecteur  $\text{vec}_3((\sigma_z \circ \sigma_x \circ \delta_y)u_y) \in \mathbb{R}^{(N-1)^3}$  s'écrit en fonction de  $U = \text{vec}_3(u)$  et des données de Dirichlet au bord de  $\Omega = (a, b)^3$  sous la forme

$$\text{vec}_3((\sigma_z \circ \sigma_x \circ \delta_y)u_y) = -\frac{1}{\hbar^2} (P_s \otimes \mathcal{H} \otimes P_s)U - G_y, \quad (5.118)$$

avec

$$\left\{ \begin{aligned}
 G_y &= \frac{1}{h^2}(P_s \otimes \mathcal{G} \otimes P_s) \left( (I \otimes e_1 \otimes I)U_B + (I \otimes e_{N-1} \otimes I)U_T \right) \\
 &+ \frac{1}{6h^2}(e_1 \otimes I \otimes P_s) \left\{ (\mathcal{H} \otimes I)U_L + (\mathcal{G} \otimes I)(e_1 \otimes U_{LB} + e_{N-1} \otimes U_{LT}) \right\} \\
 &+ \frac{1}{6h^2}(e_{N-1} \otimes I \otimes P_s) \left\{ (\mathcal{H} \otimes I)U_R + (\mathcal{G} \otimes I)(e_1 \otimes U_{RB} + e_{N-1} \otimes U_{RT}) \right\} \\
 &+ \frac{1}{6h^2}(I \otimes I \otimes e_1) \left\{ (P_s \otimes \mathcal{H})U_F + (P_s \otimes \mathcal{G})(U_{BF} \otimes e_1 + U_{TF} \otimes e_{N-1}) \right\} \\
 &+ \frac{1}{6h^2}(I \otimes I \otimes e_{N-1}) \left\{ (P_s \otimes \mathcal{H})U_C + (P_s \otimes \mathcal{G})(U_{BC} \otimes e_1 + U_{TC} \otimes e_{N-1}) \right\} \\
 &+ \frac{1}{36h^2}(I \otimes I \otimes e_1)(e_1 \otimes I) \left( \mathcal{H}U_{LF} + \mathcal{G}(e_1U_{LBF} + e_{N-1}U_{LTF}) \right) \\
 &+ \frac{1}{36h^2}(I \otimes I \otimes e_1)(e_{N-1} \otimes I) \left( \mathcal{H}U_{RF} + \mathcal{G}(e_1U_{RBF} + e_{N-1}U_{RTF}) \right) \\
 &+ \frac{1}{36h^2}(I \otimes I \otimes e_{N-1})(e_1 \otimes I) \left( \mathcal{H}U_{LC} + \mathcal{G}(e_1U_{LBC} + e_{N-1}U_{LTC}) \right) \\
 &+ \frac{1}{36h^2}(I \otimes I \otimes e_{N-1})(e_{N-1} \otimes I) \left( \mathcal{H}U_{RC} + \mathcal{G}(e_1U_{RBC} + e_{N-1}U_{RTC}) \right).
 \end{aligned} \right. \tag{5.119}$$

**Proposition 5.4.** *Le vecteur  $\text{vec}_3((\sigma_y \circ \sigma_x \circ \delta_z)u_z) \in \mathbb{R}^{(N-1)^3}$  s'écrit en fonction de  $U = \text{vec}_3(u)$  et des données de Dirichlet au bord de  $\Omega = (a, b)^3$  sous la forme*

$$\text{vec}_3((\sigma_y \circ \sigma_x \circ \delta_z)u_z) = -\frac{1}{h^2}(P_s \otimes P_s \otimes \mathcal{H})U - G_z, \tag{5.120}$$

avec

$$\left\{ \begin{aligned}
 G_z &= \frac{1}{h^2}(P_s \otimes P_s \otimes \mathcal{G}) \left( (I \otimes I \otimes e_1)U_F + (I \otimes I \otimes e_{N-1})U_C \right) \\
 &+ \frac{1}{6h^2}(e_1 \otimes P_s \otimes I) \left\{ (I \otimes \mathcal{H})U_L + (I \otimes \mathcal{G})(U_{LF} \otimes e_1 + U_{LC} \otimes e_{N-1}) \right\} \\
 &+ \frac{1}{6h^2}(e_{N-1} \otimes P_s \otimes I) \left\{ (I \otimes \mathcal{H})U_R + (I \otimes \mathcal{G})(U_{RF} \otimes e_1 + U_{RC} \otimes e_{N-1}) \right\} \\
 &+ \frac{1}{6h^2}(I \otimes e_1 \otimes I) \left\{ (P_s \otimes \mathcal{H})U_B + (P_s \otimes \mathcal{G})(U_{LB} \otimes e_1 + U_{RB} \otimes e_{N-1}) \right\} \\
 &+ \frac{1}{6h^2}(I \otimes e_{N-1} \otimes I) \left\{ (P_s \otimes \mathcal{H})U_T + (P_s \otimes \mathcal{G})(U_{LT} \otimes e_1 + U_{RT} \otimes e_{N-1}) \right\} \\
 &+ \frac{1}{36h^2}(I \otimes e_1 \otimes I)(e_1 \otimes I) \left( \mathcal{H}U_{LB} + \mathcal{G}(e_1U_{LBF} + e_{N-1}U_{LBC}) \right) \\
 &+ \frac{1}{36h^2}(I \otimes e_1 \otimes I)(e_{N-1} \otimes I) \left( \mathcal{H}U_{RB} + \mathcal{G}(e_1U_{RBF} + e_{N-1}U_{RBC}) \right) \\
 &+ \frac{1}{36h^2}(I \otimes e_{N-1} \otimes I)(e_1 \otimes I) \left( \mathcal{H}U_{LT} + \mathcal{G}(e_1U_{LTF} + e_{N-1}U_{LTC}) \right) \\
 &+ \frac{1}{36h^2}(I \otimes e_{N-1} \otimes I)(e_{N-1} \otimes I) \left( \mathcal{H}U_{RT} + \mathcal{G}(e_1U_{RTF} + e_{N-1}U_{RTC}) \right).
 \end{aligned} \right. \quad (5.121)$$

**Proposition 5.5.** *Le schéma boîte hermitien en dimension 3 pour le problème de Poisson (5.37) possède la forme matricielle générale suivante*

$$\frac{1}{h^2}(\mathcal{H} \otimes P_s \otimes P_s + P_s \otimes \mathcal{H} \otimes P_s + P_s \otimes P_s \otimes \mathcal{H})U = F - G_x - G_y - G_z, \quad (5.122)$$

où  $F = \text{vec}_3(\Pi^0 f)$  et  $G_x, G_y, G_z$  sont donnés en (5.102), (5.119), (5.121) respectivement.

**Preuve.** La relation (5.54)<sub>a</sub> s'écrit matriciellement sous la forme

$$-\text{vec}_3(\sigma_z \circ \sigma_y \circ \delta_x u_x) - \text{vec}_3(\sigma_z \circ \sigma_x \circ \delta_y u_y) - \text{vec}_3(\sigma_y \circ \sigma_x \circ \delta_z u_z) = \text{vec}_3(\Pi^0 f). \quad (5.123)$$

En remplaçant dans (5.123) les vecteurs  $\text{vec}_3(\sigma_z \circ \sigma_y \circ \delta_x u_x)$ ,  $\text{vec}_3(\sigma_z \circ \sigma_x \circ \delta_y u_y)$  et  $\text{vec}_3(\sigma_y \circ \sigma_x \circ \delta_z u_z) \in \mathbb{R}^{(N-1)^3}$  par leurs formes matricielles obtenues en (5.101), (5.118) et (5.120) respectivement, on obtient (5.122). ■



## 5.11 Solveurs rapides

### 5.11.1 Solveur rapide pour le schéma HB1

Il résulte de la proposition 5.5 que le schéma HB1 (5.57) possède la forme matricielle suivante

$$\frac{1}{h^2} \underbrace{(\mathcal{H}_1 \otimes P_s \otimes P_s + P_s \otimes \mathcal{H}_1 \otimes P_s + P_s \otimes P_s \otimes \mathcal{H}_1)}_{H_1} U = F - G_{x,1} - G_{y,1} - G_{z,1}, \quad (5.124)$$

où le vecteur  $F$  correspond au terme source moyenné,

$$F = \text{vec}_3(\Pi^0 f). \quad (5.125)$$

Les vecteurs  $G_{x,1}$ ,  $G_{y,1}$ ,  $G_{z,1}$  sont obtenus à partir des vecteurs  $G_x$  (5.102),  $G_y$  (5.119),  $G_z$  (5.121) en remplaçant les matrices  $\mathcal{H}, \mathcal{G}$  par les matrices  $\mathcal{H}_1, \mathcal{G}_1$  en (2.48). Le système linéaire (5.122) est réécrit comme

$$H_1 U = h^2 (F - G_{x,1} - G_{y,1} - G_{z,1}), \quad (5.126)$$

où la matrice  $H_1$  est

$$H_1 = \mathcal{H}_1 \otimes P_s \otimes P_s + P_s \otimes \mathcal{H}_1 \otimes P_s + P_s \otimes P_s \otimes \mathcal{H}_1. \quad (5.127)$$

Rappelons que  $Z^l$  est une base de diagonalisation de  $\mathcal{H}_1$  (avec valeurs propres  $\mu_l$ ), ainsi que de  $P_s$  (de valeurs propres  $\alpha_l = 1 - \lambda_l/6$  avec  $\lambda_l$  donnés en (1.43)<sub>b</sub>). Donc  $Z^l \otimes Z^m \otimes Z^n$  est une base de diagonalisation de  $H_1$  avec valeurs propres  $\beta_{l,m,n}$  données par

$$\beta_{l,m,n} = \mu_l \alpha_m \alpha_n + \alpha_l \mu_m \alpha_n + \alpha_l \alpha_m \mu_n, \quad 1 \leq l, m, n \leq N-1. \quad (5.128)$$

Ceci permet de résoudre (5.126) par l'algorithme rapide suivant.

**Algorithme 10.** (*Algorithme rapide pour le schéma HB1*)

- **Étape 1 :** Décomposer le terme source  $f = h^2 (F - G_{x,1} - G_{y,1} - G_{z,1})$  sur la base orthonormale  $Z^l \otimes Z^m \otimes Z^n$  de  $\mathbb{R}^{(N-1)^3}$  qui est la base de diagonalisation de  $H_1$ . On rappelle que  $\mu_l$  et  $\alpha_l = 1 - \frac{1}{6}\lambda_l$  sont les valeurs propres des matrices  $\mathcal{H}_1$  et  $P_s$  respectivement. Cette étape consiste à calculer les coefficients  $f_{l,m,n}^Z = (f, Z^l \otimes Z^m \otimes Z^n)$  et elle est effectuée par la transformation en sinus rapide, [27].

- **Etape 2** : Calculer les composantes  $U_{l,m,n}^Z$  de la solution dans l'espace de Fourier par

$$U_{l,m,n}^Z = \frac{f_{l,m,n}^Z}{\beta_{l,m,n}}, \quad 1 \leq l, m, n \leq N-1. \quad (5.129)$$

- **Etape 3** : Assembler les composantes de la solution en utilisant la décomposition du vecteur  $U \in \mathbb{R}^{(N-1)^3}$  dans la base  $Z^l \otimes Z^m \otimes Z^n$  par

$$U_{i,j,k} = \sum_{l,m,n=1}^{N-1} U_{l,m,n}^Z Z_i^l Z_j^m Z_k^n. \quad (5.130)$$

Cette étape est effectuée par la transformation inverse en sinus rapide.

Les étapes 1 et 3 sont  $O(N^3 \log_2(N))$  et l'étape 2 est  $O(N^3)$ , ce qui donne un algorithme  $O(N^3 \log_2(N))$  en total.

### 5.11.2 Résultats numériques pour le schéma HB1

Dans cette section, on donne quelques résultats numériques préliminaires obtenus par le schéma HB1 concernant la résolution du problème du Poisson (5.37) sur  $\Omega = (0, 1)^3$ . Les erreurs sont évalués par les normes  $L^2$  et  $L^\infty$  discrètes

$$\begin{cases} \|u_{ex} - u\|_h = \left( h^3 \sum_{i,j,k} (u_{ex}(x_i, y_j, z_k) - u_{i,j,k})^2 \right)^{\frac{1}{2}}, & (a) \\ \|u_{ex} - u\|_\infty = \max_{i,j,k=1,\dots,N-1} |u_{ex}(x_i, y_j, z_k) - u_{i,j,k}|. & (b) \end{cases} \quad (5.131)$$

Le code de calcul est effectué en Matlab. Le maillage varie entre  $N = 8$  ( $N = 2^3$ ) et  $N = 128$  ( $N = 2^7$ ) dans chaque direction. La taille de système linéaire à résoudre pour  $N = 128$  est d'environ deux millions d'inconnues pour la solution  $u$ . Le temps calcul  $T$  est évalué à l'aide des fonctions `tic` et `toc`. Le quotient  $T/(N^3 \log_2(N))$  apparaît clairement constant, ce qui montre une complexité algorithmique  $O(N^3 \log_2(N))$ .

**Test 1** : On considère la solution exacte  $u_{ex}(x, y, z) = \cos(\pi x) \cos(\pi y) \cos(\pi z)$ . Les résultats sont donnés dans la table 5.1. Le schéma HB1 est d'ordre 2 pour les quatres inconnues  $u$ ,  $u_x$ ,  $u_y$ ,  $u_z$ .

**Test 2** : On considère la solution exacte  $u_{ex}(x, y, z) = \ln(1 + x + y^3 + z^2)$ . Les résultats sont donnés dans la table 5.2. On observe l'ordre 2 du schéma HB1 pour les quatres inconnues.

**Test 3** : On considère la Gaussienne  $u_{ex}(x, y, z) = \exp(-((x - 0.5)^2 + (y - 0.5)^2 + (z -$

$N_x \times N_y \times N_z$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	T(s.)
$8 \times 8 \times 8$	5.700(-3)	3.350(-2)	3.260(-2)	1.470(-1)	0.77
T. C.	1.92	1.73	1.66	1.42	
$16 \times 16 \times 16$	1.500(-3)	1.010(-2)	1.030(-2)	5.480(-2)	1.36
T. C.	1.93	1.75	1.93	0.86	
$32 \times 32 \times 32$	3.918(-4)	3.000(-3)	2.700(-3)	3.010(-2)	3.90
T. C.	1.99	1.80	1.96	0.96	
$64 \times 64 \times 64$	9.863(-5)	8.577(-4)	6.920(-4)	1.540(-2)	18.64
T. C.	1.99	1.83	1.99	0.98	
$128 \times 128 \times 128$	2.471(-5)	2.411(-4)	1.738(-4)	7.800(-3)	131

TAB. 5.1: Erreur et taux de convergence du schéma HB1 avec  $u_{ex}(x, y, z) = \cos(\pi x) \cos(\pi y) \cos(\pi z)$

$N_x \times N_y \times N_z$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{y,ex} - u_y\ _h$	$\ u_{z,ex} - u_z\ _h$	T(s.)	$T/(N^3 \log_2(N))$
$8 \times 8 \times 8$	1.100(-3)	2.500(-3)	3.400(-3)	3.600(-3)	0.89	5.794(-4)
T. C.	1.92	1.47	1.62	1.58		
$16 \times 16 \times 16$	2.902(-4)	9.019(-4)	1.100(-3)	1.200(-3)	1.34	8.178(-5)
T. C.	1.97	1.64	1.74	1.70		
$32 \times 32 \times 32$	7.402(-5)	2.895(-4)	3.294(-4)	3.681(-4)	3.93	2.398(-5)
T. C.	1.99	1.73	1.79	1.77		
$64 \times 64 \times 64$	1.862(-5)	8.719(-5)	9.468(-5)	1.075(-4)	18.60	1.182(-5)
T. C.	1.99	1.79	1.83	1.82		
$128 \times 128 \times 128$	4.662(-6)	2.518(-5)	2.649(-5)	3.040(-5)	131	8.923(-6)

TAB. 5.2: Erreur et taux de convergence du schéma HB1 avec  $u_{ex}(x, y, z) = \ln(1 + x + y^2 + z^3)$

0.5)<sup>2</sup>). Les résultats sont donnés dans la table 5.3. Le schéma HB1 est encore d'ordre 2 dans ce cas.

**Test 4 :** On considère la solution exacte  $u_{ex}(x, y, z) = \exp(-30((x-0.5)^2 + (y-0.5)^2) + (z-0.5)^2) \cos(20(x+y+z-1))$ . Les résultats numériques sont donnés dans la table 5.4. Dans ce cas on constate que le schéma HB1 est observé d'ordre 4 pour les quatres inconnues  $u, u_x, u_y, u_z$ .

### 5.11.3 Solveur rapide pour le schéma HB2

Il résulte de la proposition 5.5 que le schéma HB2 (5.57) possède la forme matricielle suivante

$$\frac{1}{h^2}(\mathcal{H}_2 \otimes P_s \otimes P_s + P_s \otimes \mathcal{H}_2 \otimes P_s + P_s \otimes P_s \otimes \mathcal{H}_2)U = F - G_{x,2} - G_{y,2} - G_{z,2}, \quad (5.132)$$

$N_x \times N_y \times N_z$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	T(s.)
$8 \times 8 \times 8$	1.500(-3)	2.100(-3)	3.700(-3)	7.500(-3)	0.77
T. C.	2.03	1.40	1.98	1.05	
$16 \times 16 \times 16$	3.669(-4)	7.923(-4)	9.340(-4)	3.600(-3)	1.36
T. C.	1.99	1.68	2.00	1.08	
$32 \times 32 \times 32$	9.192(-5)	2.471(-4)	2.330(-4)	1.700(-3)	3.86
T. C.	1.99	1.78	2.00	1.11	
$64 \times 64 \times 64$	2.299(-5)	7.196(-5)	5.821(-5)	7.859(-4)	18.57
T. C.	2.00	1.83	2.00	1.03	
$128 \times 128 \times 128$	5.750(-6)	2.022(-5)	1.454(-5)	3.827(-4)	132.30

TAB. 5.3: Erreur et taux de convergence du schéma HB1 avec  $u_{ex}(x, y, z) = \exp(-(x - 0.5)^2 - (y - 0.5)^2 - (z - 0.5)^2)$

$N_x \times N_y \times N_z$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	T(s.)	$T/(N^3 \log_2(N))$
$16 \times 16 \times 16$	8.1(-3)	1.094(-1)	1.234(-1)	1.477	1.37	8.483(-5)
T. C.	4.60	4.60	4.54	4.46		
$32 \times 32 \times 32$	3.339(-4)	4.5(-3)	5.3(-3)	6.70(-2)	3.93	2.398(-5)
T. C.	4.10	4.11	4.09	4.06		
$64 \times 64 \times 64$	1.934(-5)	2.662(-4)	3.105(-4)	4(-3)	18.57	1.180(-5)
T. C.	3.96	3.86	3.98	4.01		
$128 \times 128 \times 128$	1.242(-6)	1.785(-5)	1.961(-5)	2.478(-4)	132.05	8.991(-6)

TAB. 5.4: Erreur et taux de convergence du schéma HB1 avec  $u_{ex}(x, y, z) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2)) \cos(20(x + y + z - 1))$

où le vecteur  $F$  correspond au terme source moyenné,

$$F = \text{vec}_3(\Pi^0 f). \quad (5.133)$$

Les vecteurs  $G_{x,2}$ ,  $G_{y,2}$ ,  $G_{z,2}$  sont obtenus à partir des vecteurs  $G_x$  (5.102),  $G_y$  (5.119),  $G_z$  (5.121) en remplaçant les matrices  $\mathcal{H}$ ,  $\mathcal{G}$  par les matrices  $\mathcal{H}_2$ ,  $\mathcal{G}_2$ . On donne maintenant une procédure de résolution directe du nouveau schéma HB2 d'ordre 4 en dimension trois.

En utilisant la structure de la matrice  $\mathcal{H}_2$  en (2.156), on obtient

$$\underbrace{\mathcal{H}_2 \otimes P_s \otimes P_s + P_s \otimes \mathcal{H}_2 \otimes P_s + P_s \otimes P_s \otimes \mathcal{H}_2}_{H_2} = \underbrace{(\mathcal{H}_1 \otimes P_s \otimes P_s + P_s \otimes \mathcal{H}_1 \otimes P_s + P_s \otimes P_s \otimes \mathcal{H}_1)}_{H_1} + \underbrace{(RS^T \otimes P_s \otimes P_s + P_s \otimes RS^T \otimes P_s + P_s \otimes P_s \otimes RS^T)}_{\delta H}.$$

Le système linéaire (5.134) devient

$$H_2 U = f, \quad (5.134)$$

où  $f = h^2(F - G_{x,2} - G_{y,2} - G_z, 2)$ . La matrice  $\delta H = RS^T \otimes P_s \otimes P_s + P_s \otimes RS^T \otimes P_s + P_s \otimes P_s \otimes RS^T$  est une perturbation de petit rang de  $H_1$ . Sa structure est la suivante. La matrice  $P_s$  est exprimée par

$$P_s = Z \operatorname{diag}(\alpha_1, \dots, \alpha_{N_1}) Z^T. \quad (5.135)$$

Définissant le vecteur  $Z'^l = \alpha_l^{1/2} Z^l$ , la matrice  $Z'$  d'ordre  $(N-1)$  est

$$Z' = [Z'^1, \dots, Z'^{N-1}]. \quad (5.136)$$

De plus,

$$P_s = Z' Z'^T. \quad (5.137)$$

Donc  $\delta H$  peut être écrite sous la forme

$$\delta H = RS^T \otimes P_s \otimes P_s + P_s \otimes RS^T \otimes P_s + P_s \otimes P_s \otimes RS^T.$$

On en déduit que  $\delta H$  se met sous la forme  $OQ^T$  où les matrices rectangulaires  $O, Q \in \mathbb{M}_{(N-1)^3, 18(N-1)^2}(\mathbb{R})$  sont

$$O = [O_1, \dots, O_{18}], \quad Q = [Q_1, \dots, Q_{18}], \quad (5.138)$$

avec les matrices  $O_i, Q_i \in \mathbb{M}_{(N-1)^3, (N-1)^2}(\mathbb{R}), i = 1, \dots, 6$  sont

$$\begin{cases} O_i = R_i \otimes Z' \otimes Z', & O_{i+6} = Z' \otimes R_i \otimes Z', & O_{i+12} = Z' \otimes Z' \otimes R_i, \\ Q_i = S_i \otimes Z' \otimes Z', & Q_{i+6} = Z' \otimes S_i \otimes Z', & Q_{i+12} = Z' \otimes Z' \otimes S_i. \end{cases} \quad (5.139)$$

De même qu'en dimension deux, la base de l'algorithme rapide proposé est la formule de Sherman-Morrison-Woodbury (2.155) appliquée à la matrice (5.134). Rappelons qu'elle consiste en l'écriture de l'inverse de la matrice  $H_2 = H_1 + OQ^T$  sous la forme

$$H_2^{-1} = H_1^{-1} - H_1^{-1} O \left( I_{18(N-1)^2} + Q^T H_1^{-1} O \right)^{-1} Q^T H_1^{-1}. \quad (5.140)$$

L'algorithme suivant effectue la résolution de (5.134) en utilisant (5.140). Les indications de complexité de calcul sont données à chaque étape de l'algorithme. Notons que

l'implémentation de cet algorithme reste simple en utilisant les opérations vectorielles standard dans Matlab ou (Fortran 90).

**Algorithme 11.** (*Algorithme rapide pour le schéma HB2*)

- **Étape 1 :** Soit  $f = h^2(F - G_{x,2} - G_{y,2} - G_{z,2}) \in \mathbb{R}^{(N-1)^3}$  le vecteur terme source. Résoudre le système linéaire

$$H_1 g = f. \quad (5.141)$$

Il est résolu en utilisant l'algorithme 10 à un coût de  $O(N^3 \log_2(N))$ . Le vecteur  $g \in \mathbb{R}^{(N-1)^3}$  est stocké pour l'étape 7.

- **Étape 2 :** Calculer le vecteur  $Q^T g \in \mathbb{R}^{18(N-1)^2}$ ,

$$Q^T g = [Q_1^T g, Q_2^T g, \dots, Q_{18}^T g]^T. \quad (5.142)$$

On a par exemple pour  $Q_1^T g$ ,

$$Q_1^T g = [Q_{1,i,j}^T g]^T, \quad 1 \leq i, j \leq N-1. \quad (5.143)$$

avec

$$\begin{aligned} (Q_{1,i,j}^T g) &= (S_1 \otimes Z^{l,i}, Z^{l,j})^T g \\ &= \sum_{i=1}^{N-1} (S_1)_i \sum_{j,k=1}^{N-1} g_{i,j,k} Z_l^{l,j} Z_m^{l,k} \\ &= \alpha_l^{1/2} \sum_{i=1}^{N-1} (S_1)_i \sum_{j,k=1}^{N-1} g_{i,j,k} Z_l^j Z_m^k. \end{aligned}$$

Les termes

$$\sum_{j,k=1}^{N-1} g_{i,j,k} Z_l^j Z_m^k, \quad 1 \leq i \leq N-1, \quad (5.144)$$

sont calculés par FFT en utilisant

$$\sum_{j,k=1}^{N-1} g_{i,j,k} Z_l^j Z_m^k = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{j,k=1}^{N-1} g_{i,j,k} \sin\left(\frac{l j \pi}{N}\right) \sin\left(\frac{m k \pi}{N}\right). \quad (5.145)$$

- **Étape 3 :** Assembler la matrice d'ordre  $18(N-1)^2$  entre parenthèses dans (5.140). C'est la matrice non-symétrique

$$I_{18(N-1)^2} + Q^T H_1^{-1} O, \quad (5.146)$$

où

$$Q^T H_1^{-1} O = \begin{bmatrix} Q_1^T H_1^{-1} O_1 & Q_1^T H_1^{-1} O_2 & \dots & Q_1^T H_1^{-1} O_{18} \\ Q_2^T H_1^{-1} O_1 & Q_2^T H_1^{-1} O_2 & \dots & Q_2^T H_1^{-1} O_{18} \\ \dots & \dots & \dots & \dots \\ Q_{18}^T H_1^{-1} O_1 & Q_{18}^T H_1^{-1} O_2 & \dots & Q_{18}^T H_1^{-1} O_{18} \end{bmatrix}.$$

Les matrices  $O_k$  et  $Q_k$ ,  $1 \leq k \leq 18$ , d'ordre  $(N-1)^3 \times (N-1)^2$  sont données en (5.139). Chaque matrice  $M_k^T H_1^{-1} L_i$  est de nouveau calculée par FFT. Noter que la matrice (5.146) est calculée une fois pour toute.

– **Étape 4** : Résoudre le système linéaire d'ordre  $18(N-1)^2 \times 18(N-1)^2$

$$(I_{18(N-1)^2} + Q^T H_1^{-1} O)w = Q^T g. \quad (5.147)$$

La complexité de calcul de tout l'algorithme s'appuie sur l'efficacité de cette résolution. Elle est résolue ici par la méthode GMRES (sans préconditionnement). Les résultats numériques montrent un coût de calcul  $O(N^3 \log_2(N))$ . La solution  $w \in \mathbb{R}^{18(N-1)^2}$  est décomposée en blocs

$$w = [w_1, w_2, \dots, w_{18}]^T, w_1, w_2, \dots, w_{18} \in \mathbb{R}^{(N-1)^2}. \quad (5.148)$$

– **Étape 5** : Calcul du produit  $t = Lw$ ,  $w \in \mathbb{R}^{18(N-1)^2}$ ,  $t \in \mathbb{R}^{(N-1)^3}$

$$t = t_1 + t_2 + t_3 + \dots + t_{18}, \quad t_i \in \mathbb{R}^{(N-1)^3}, \quad (5.149)$$

avec

$$\left\{ \begin{array}{l} t_1 = (R_1 \otimes [Z'^1, \dots, Z'^{N-1}] \otimes [Z'^1, \dots, Z'^{N-1}])w_1, \\ \dots \\ t_6 = (R_6 \otimes [Z'^1, \dots, Z'^{N-1}] \otimes [Z'^1, \dots, Z'^{N-1}])w_6, \\ t_7 = ([Z'^1, \dots, Z'^{N-1}] \otimes R_1 \otimes [Z'^1, \dots, Z'^{N-1}])w_7, \\ \dots \\ t_{12} = ([Z'^1, \dots, Z'^{N-1}] \otimes R_6 \otimes [Z'^1, \dots, Z'^{N-1}])w_{12}, \\ t_{13} = ([Z'^1, \dots, Z'^{N-1}] \otimes [Z'^1, \dots, Z'^{N-1}] \otimes R_1)w_{13}, \\ \dots \\ t_{18} = ([Z'^1, \dots, Z'^{N-1}] \otimes [Z'^1, \dots, Z'^{N-1}] \otimes R_6)w_{18}. \end{array} \right. \quad (5.150)$$

Le calcul du vecteur  $t \in \mathbb{R}^{(N-1)^3}$  en (5.149) est donc  $O(N^3 \log_2(N))$ .

- **Etape 6** : Résolution du système linéaire dans  $\mathbb{R}^{(N-1)^3}$

$$v = H_1^{-1}t, \quad (5.151)$$

via le solveur FFT rapide comme à l'étape 1. Le coût de calcul est  $O(N^3 \log_2(N))$ .

- **Etape 8** : Assemblage de la solution  $U \in \mathbb{R}^{(N-1)^3}$  par

$$U = g - v, \quad (5.152)$$

où  $g, v \in \mathbb{R}^{(N-1)^3}$  sont donnés en (5.142-5.151). Le coût est  $O(N^3)$ .

- **Etape 8** : Calcul des gradients hermitiens  $U_x, U_y, U_z \in \mathbb{R}^{(N-1)^3}$  en utilisant (5.94), (5.95), (5.96).

Noter que l'utilisation d'une méthode itérative classique avec la matrice d'itération  $H_1^{-1}RS^T$  n'est pas envisageable car le rayon spectral est autour 1.48. Ceci justifie l'intérêt de l'algorithme 11.

#### 5.11.4 Résultats numériques pour le schéma HB2

Dans cette section, on montre quelques résultats numériques préliminaires qui prouvent l'efficacité du schéma HB2 d'ordre 4 dans différents cas de calcul de résolution de problème de Poisson (5.37) sur  $\Omega = (0,1)^3$ . Le code de calcul final est effectué en Matlab. Le temps calcul  $T$  est évalué à l'aide des fonctions matlab `tic` et `toc`. Noter que le temps  $T$  pourrait être notablement réduit en utilisant un langage compilé comme Fortran 90( en dimension deux, nous avons observé une réduction du temps calcul d'un facteur 5, cf les figures 3.3 et 3.4). Ici, on se concentre sur la complexité arithmétique en fonction de la taille  $N$  de problème. On donne en particulier le quotient  $T/(N^3 \log_2(N))$ . Il apparait constant ce qui indique une complexité algorithmique de  $O(N^3 \log_2(N))$ . Le maillage varie entre  $N = 8 = 2^3$  et  $N = 128 = 2^7$  dans les trois directions. Les erreurs  $L^2$  et  $L^\infty$  discrètes utilisées pour estimer les erreurs sont

$$\begin{cases} \|u_{ex} - u\|_h = \left( h^3 \sum_{i,j,k} (u_{ex}(x_i, y_j, z_k) - u_{i,j,k})^2 \right)^{\frac{1}{2}}, \\ \|u_{ex} - u\|_\infty = \max_{1 \leq i,j,k \leq N-1} |u_{ex}(x_i, y_j, z_k) - u_{i,j,k}|. \end{cases} \quad (b) \quad (5.153)$$

**Test 1** : On considère la fonction  $u_{ex}(x, y, z) = \cos(\pi x) \cos(\pi y) \cos(\pi z)$  comme solution exacte. Les résultats sont donnés dans la table 5.5. On observe la précision d'ordre



$N_x \times N_y \times N_z$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{y,ex} - u_y\ _\infty$	$\ u_{z,ex} - u_z\ _\infty$	T(s.)
$8 \times 8 \times 8$	2.8344(-4)	1.5(-3)	1.5(-3)	1.5(-3)	2.64
T. C.	4	3.8	3.8	3.8	
$16 \times 16 \times 16$	1.7644(-5)	1.147(-4)	1.147(-4)	1.147(-4)	8.49
T. C.	4.08	3.828	3.828	3.828	
$32 \times 32 \times 32$	1.044(-6)	8.0785(-6)	8.0785(-6)	8.0785(-6)	33.40
T. C.	4.07	3.84	3.84	3.84	
$64 \times 64 \times 64$	6.1915(-8)	5.6367(-7)	5.6367(-7)	5.6367(-7)	152.61
T. C.	4.04	3.86	3.86	3.86	
$128 \times 128 \times 128$	3.7541(-9)	3.885(-8)	3.885(-8)	3.885(-8)	911.11

TAB. 5.5: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y, z) = \cos(\pi x) \cos(\pi y) \cos(\pi z)$

$N_x \times N_y \times N_z$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{y,ex} - u_y\ _h$	$\ u_{z,ex} - u_z\ _h$	T(s.)
$8 \times 8 \times 8$	2.0897(-5)	3.8499(-5)	7.0151(-5)	1.0292(-4)	3.23
T. C.	4.26	3.564	4.04	3.82	
$16 \times 16 \times 16$	1.0879(-6)	3.2545(-6)	4.2489(-6)	7.2749(-6)	9.59
T. C.	4.1	3.62	3.84	3.84	
$32 \times 32 \times 32$	6.309(-8)	2.6329(-7)	2.965(-7)	5.05(-7)	38.45
T. C.	4	3.70	3.8	3.86	
$64 \times 64 \times 64$	3.8013(-9)	2.0146(-8)	2.1186(-8)	3.461(-8)	170.40
T. C.	4	3.77	3.83	3.88	
$128 \times 128 \times 128$	2.3308(-10)	1.4738(-9)	1.4896(-9)	2.3439(-9)	864.80

TAB. 5.6: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y, z) = \ln(1 + x + y^3 + z^2)$

quatre pour les quatres inconnues  $u$ ,  $u_x$ ,  $u_y$ ,  $u_z$ .

**Test 2 :** On considère le problème de Poisson avec la solution exacte  $u_{ex}(x, y, z) = \ln(1 + x + y^3 + z^2)$ . Observons dans la table (5.11.4) l'ordre 4 du schéma HB2. Noter la singularité du gradient à proximité du bord. Le schéma HB2 est d'ordre 4 pour les quatres inconnues. Une très bonne précision est atteinte sur la grille finale.

**Test 3 :** On considère la gaussienne  $u_{ex}(x, y, z) = \exp(-((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2))$ . Les résultats sont donnés dans la table 5.7. Le schéma HB2 est d'ordre 4 dans ce cas.

**Test 4 :** On considère dans la table 5.8 des résultats numériques avec la solution exacte  $u_{ex}(x, y, z) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2)) \cos(20(x + y + z - 1))$ . Cette fonction est oscillante et plus difficile à calculer que les exemples précédents. Le schéma

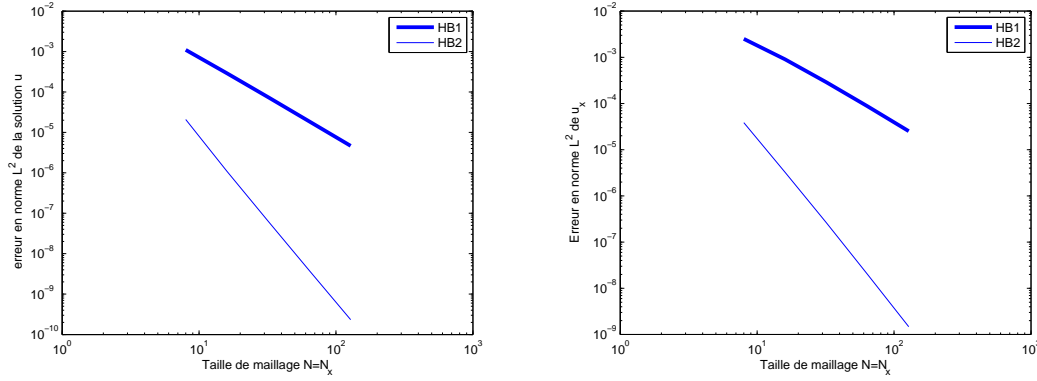


FIG. 5.3: Courbes de convergence pour le test 2 de  $u$  (à gauche) et  $u_x$  (à droite)

$N_x \times N_y \times N_z$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	T(s.)	$T/(N^3 \log_2(N))$
$8 \times 8 \times 8$	3.404(-5)	5.127(-5)	8.653(-5)	1.953(-4)	2.79	2.6(-3)
T. C.	4.76	3.97	4.36	3.79		
$16 \times 16 \times 16$	1.251(-6)	3.263(-6)	4.196(-6)	1.412(-5)	8.95	7.880(-4)
T. C.	4.40	4.18	4.22	3.98		
$32 \times 32 \times 32$	5.924(-8)	1.792(-7)	2.236(-7)	8.925(-7)	32.41	2.853(-4)
T. C.	4.11	4.13	4.12	3.61		
$64 \times 64 \times 64$	3.420(-9)	1.020(-8)	1.279(-8)	7.306(-8)	148.52	1.362(-4)
T. C.	4.01	4.06	4.06	3.41		
$128 \times 128 \times 128$	2.111(-10)	6.090(-10)	7.631(-10)	6.863(-9)	853	8.383(-5)

TAB. 5.7: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y, z) = \exp(-((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2))$

HB2 est encore d'ordre 4 dans ce cas.

$N_x \times N_y \times N_z$	$\ u_{ex} - u\ _h$	$\ u_{x,ex} - u_x\ _h$	$\ u_{ex} - u\ _\infty$	$\ u_{x,ex} - u_x\ _\infty$	T(s.)	$T/(N^3 \log_2(N))$
$8 \times 8 \times 8$	2.533(-1)	3.469	2.538	25.030	3.10	2.9(-3)
T. C.	4.96	4.98	4.36	4.08		
$16 \times 16 \times 16$	8.1(-3)	1.095(-1)	1.236(-1)	1.477	9.02	7.944(-4)
T. C.	4.60	4.60	4.54	4.45		
$32 \times 32 \times 32$	3.339(-4)	4.5(-3)	5.3(-3)	6.71(-2)	29.74	2.618(-4)
T. C.	4.11	4.12	4.09	4.06		
$64 \times 64 \times 64$	1.928(-5)	2.584(-4)	3.101(-4)	4(-3)	130.47	1.196(-4)
T. C.	4.02	4.02	4.00	4.02		
$128 \times 128 \times 128$	1.183(-6)	1.582(-5)	1.934(-5)	2.451(-4)	677	6.660(-5)

TAB. 5.8: Erreur et taux de convergence du schéma HB2 avec  $u_{ex}(x, y, z) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2)) \cos(20(x + y + z - 1))$



# Conclusion

Dans ce mémoire, nous présentons un nouveau schéma aux différences d'ordre 4 pour le problème de Poisson dans un carré et dans un cube avec conditions limites de Dirichlet. Ce schéma généralise l'approche donnée par Croisille [12]. La conception du schéma est basée sur une formulation mixte combinant l'approximation de gradient par une dérivée hermitienne d'ordre 4 avec une formulation conservative discrète sur des boîtes de longueur  $2h$ . L'étude comporte les deux aspects suivants.

- Premièrement, montrer qu'une précision d'ordre 4 est obtenue pour les deux inconnues, la solution et son gradient.
- Deuxièmement, donner un algorithme de résolution directe rapide basé sur la formule de Sherman-Morrison-Woodbury et la transformée en sinus rapide. Plusieurs résultats numériques montrent que la complexité algorithmique est en  $O(N^2 \log_2(N))$  en dimension 2 et  $O(N^3 \log_2(N))$  en dimension 3.

De très bonnes performances de calcul ont été obtenues sur une machine de bureau. A titre d'exemple, un calcul sur un maillage  $1024 \times 1024$  est effectué en moins de cinq secondes sur un PC ordinaire. Ce solveur sert de préconditionnement à des problèmes elliptiques non réguliers. Une autre application concerne les maillages cartésiens multiéchelle. Ce type de maillages permet une résolution locale également d'ordre 4 sur des zones raffinées. Le raffinement est récursif.

Les schémas boîte sont actuellement de plus en plus étudiés. Ils sont étudiés par I. Greff [19] pour différents problèmes elliptiques avec les méthodes d'éléments finis. Ils ont été récemment étudié par J.B. Perot [32].

Les perspectives principales de ce travail sont les suivantes :

- Généralisation des schémas boîte sur grille cartésienne à des problèmes avec obstacles. La méthode envisagée est de type "embedded boundary".
- Utilisation du schéma présenté pour des applications en physique des particules. L'opérateur de moyenne  $\Pi^0 f_{i,j}$  dans (3.40) correspond à la charge électrique dans la boîte  $K_{i,j}$ . Le schéma sert dans ce cas à calculer le potentiel et le champ électromagnétique

aux points de grille.

- Utilisation du schéma pour des modèles complexes d'équations elliptiques, comme MEMS(Microelectromechanical Systems), [13], [26], [31], [37].

## Annexe

### Transformée en sinus rapide

Cet annexe est consacré à donner en détail l'implémentation de la transformée en sinus utilisée dans les chapitres 3 et 5.

En Matlab, les fonctions `dst` et `idst` permettent de calculer la transformée en sinus directe “forward” et la transformée inverse en sinus “bakward”. Elles sont identiques à une constante près. La fonction  $y = \text{dst}(x)$  est définie par

$$x \in \mathbb{R}^{N-1} \longrightarrow y \in \mathbb{R}^{N-1} \quad (5.154)$$

avec

$$y_j = \sum_{i=1}^{N-1} x_i \sin\left(\frac{ij\pi}{N}\right), \quad j = 1, 2, \dots, N-1. \quad (5.155)$$

De même,  $x = \text{idst}(y)$  calcule la transformée inverse en sinus rapide du vecteur  $y \in \mathbb{R}^{(N-1)}$  avec

$$x_j = \frac{2}{N} \sum_{i=1}^{N-1} y_i \sin\left(\frac{ij\pi}{N}\right), \quad j = 1, 2, \dots, N-1. \quad (5.156)$$

Les calculs (5.155) et (5.156) peuvent être effectués à un coût  $O(N \log_2(N))$ . Ces fonctions sont normalisées, c'est-à-dire l'appel de la fonction `dst` suivi de l'appel `idst` (et vice versa) reproduit le même vecteur original (avec erreur d'arrondi).

Dans cette thèse, on a utilisé les fonctions `dst` et `idst` pour calculer les transformées en sinus qui interviennent dans les algorithmes 2 et 3. En effet, les coefficients  $\bar{F}_{k,l}^Z$ ,  $1 \leq$

$k, l \leq N - 1$ , dans l'étape 1 de l'algorithme 2 sont calculés de la façon suivante :

$$\left\{ \begin{aligned} \bar{F}_{k,l}^Z &= \sum_{i,j=1}^{N-1} \frac{2}{N} F_{i,j} \sin\left(\frac{ik\pi}{N}\right) \sin\left(\frac{jl\pi}{N}\right) \\ &= \frac{2}{N} \sum_{i=1}^{N-1} \underbrace{\sum_{j=1}^{N-1} F_{i,j} \sin\left(\frac{jl\pi}{N}\right)}_{a_{i,l}} \sin\left(\frac{ik\pi}{N}\right), \end{aligned} \right. \quad (5.157)$$

où  $F = \text{vec}_2(\Pi^0 f) \in \mathbb{R}^{(N-1)^2}$  et où  $a = (a_{i,l})_{1 \leq i, l \leq N-1}$  est la matrice carrée d'ordre  $N - 1$  donnée par

$$\begin{aligned} a_{i,l} &= \sum_{j=1}^{N-1} F_{i,j} \sin\left(\frac{jl\pi}{N}\right) \\ &= \left( \text{dst}(F((i-1)(N-1)+1 : i(N-1))) \right)_l, \quad 1 \leq i, l \leq N-1. \end{aligned}$$

Donc

$$\begin{aligned} \bar{F}_{k,l}^Z &= \sum_{i=1}^{N-1} a_{i,l} \sin\left(\frac{ik\pi}{N}\right) \\ &= \frac{2}{N} \left( \text{dst}(a(:, l)) \right)_k. \end{aligned}$$

De même, les coefficients  $U_{i,j}$ ,  $1 \leq i, j \leq N - 1$ , de l'étape 3 de l'algorithme 2 sont calculés à partir des coefficients  $U_{k,l}^Z$ ,  $1 \leq k, l \leq N - 1$ , de l'étape 2 de la façon suivante :

$$\begin{aligned} U_{i,j} &= (U, e_i \otimes e_j) \\ &= \left( \sum_{k,l=1}^{N-1} U_{k,l}^Z Z^k \otimes Z^l, e_i \otimes e_j \right) \\ &= \sum_{k,l=1}^{N-1} U_{k,l}^Z (e_i^T Z^k \otimes e_j^T Z^l) \\ &= \sum_{k,l=1}^{N-1} U_{k,l}^Z Z_i^k Z_j^l \\ &= \frac{2}{N} \sum_{k,l=1}^{N-1} U_{k,l}^Z \sin\left(\frac{ki\pi}{N}\right) \sin\left(\frac{lj\pi}{N}\right). \end{aligned}$$



On obtient

$$\begin{aligned} U_{i,j} &= \frac{2}{N} \sum_{l=1}^{N-1} \underbrace{\sum_{k=1}^{N-1} U_{k,l}^Z \sin\left(\frac{ki\pi}{N}\right) \sin\left(\frac{lj\pi}{N}\right)}_{b_{i,l}} \\ &= \frac{2}{N} \sum_{l=1}^{N-1} b_{i,l} \sin\left(\frac{lj\pi}{N}\right). \end{aligned}$$

Donc

$$U_{i,j} = \left( \text{idst}(b(i, :)) \right)_j, \quad i, j = 1, \dots, N-1, \quad (5.158)$$

avec  $b = (b_{i,l})_{1 \leq i, l \leq N-1}$  la matrice carrée d'ordre  $N-1$  donnée par

$$\begin{aligned} b_{i,l} &= \sum_{k=1}^{N-1} U_{k,l}^Z \sin\left(\frac{ki\pi}{N}\right) \\ &= \frac{N}{2} (\text{idst}(U^Z(l : \text{inc} : (N-2)(N-1) + l)))_i, \quad 1 \leq i, l \leq N-1, \end{aligned}$$

où  $U^Z = \text{vec}_2[U_{k,l}^Z]_{1 \leq k, l \leq N-1} \in \mathbb{R}^{(N-1)^2}$  et l'incrément  $\text{inc} = N-1$ . Ce qui donne

$$b(:, l) = \frac{N}{2} \text{idst}(U^Z(l : \text{inc} : (N-2)(N-1) + l)), \quad l = 1, \dots, N-1. \quad (5.159)$$

Par ailleurs, l'implémentation des transformées en sinus en 3D dans le chapitre 5 qui sont utilisées dans les algorithmes 10 et 11 peut être déduite de (5.157) et (5.158).

En Fortran 90, on a utilisé les sous-routines `dsint1f` et `dsint1b` pour les transformées en sinus rapide “forward” et “backward” respectivement. Ces sous-routines sont dans la bibliothèque `fftpack5.f90` de Swarztrauber, [35]. La sous-routine `dsint1i` sert pour l'initialisation des tableaux d'entrée de `dsint1f` et `dsint1b`.



# Table des figures

1.1	Grille en dimension 1. Les deux points de bord sont $x_0, x_N$ (notés par “o”). Les points $x_i, i = 1, \dots, N-1$ sont les points intérieurs (notés par “•”).	6
2.1	Le stencil du schéma boîte de keller : $\mathcal{K}_j = [x_{j-1}, x_j]$	18
2.2	Le stencil du schéma boîte hermitien : $K_j = [x_{j-1}, x_{j+1}]$	19
2.3	Les notations en dimension 1 sur $\Omega = (a, b)$ . La grande boule à gauche désigne le noeud $x_0 = a$ , celle à droite désigne le noeud $x_N = b$ . Les petites boules désignent les noeuds $x_i, 1 \leq i \leq 11$	24
2.4	A gauche, les solutions exacte et numérique avec $u_{ex}(x) = \cos(\pi x^2)$ . A droite, les courbes de convergence de 3 schémas.	55
2.5	A gauche, les solutions exacte et numérique avec $u_{ex}(x) = \sin(\pi x^3)$ . A droite, les courbes de convergence de 3 schémas.	55
2.6	A gauche, les solutions exacte et numérique avec $u_{ex}(x) = \exp(-(x - \frac{1}{2})^2)$ . A droite, les courbes de convergence de 3 schémas.	56
2.7	A gauche, les solutions exacte et numérique avec $u_{ex}(x) = \exp(-1000(x - \frac{1}{2})^2)$ . A droite, les courbes de convergence de 3 schémas.	56
3.1	Notations de la grille en dimension 2. Les lettres $L$ et $R$ correspondent à “Left” et “Right”. Les lettres $B$ et $T$ correspondent à “Bottom” et “Top”.	64
3.2	Schéma HB en dimension deux. Le flux est intégré sur le contour noir de la boîte $K_{i,j} = [x_i - h, x_i + h] \times [y_j - h, y_j + h]$	68
3.3	Temps calcul en Matlab et F90 pour le test 1.1	93
3.4	Temps calcul en Matlab et F90 pour le test 1.1	105
3.5	A gauche, la solution exacte $u_{ex}(x, y) = \exp(-(x - 0.5)^2 - (y - 0.5)^2)$ . A droite, des courbes de convergence.	108
3.6	A gauche, la solution exacte $u_{ex}(x, y) = \exp(-400(x - 0.5)^2 - 400(y - 0.5)^2)$ . A droite, des courbes de convergence.	108
3.7	$u_{ex}(x, y) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2)) \cos(20(x + y - 1))$ . A droite, des courbes de convergence.	109
4.1	Solution exacte $u_{ex}(x, y) = x^2 + y^2$ (à gauche) et sa dérivée $\partial_x u(x, y) = 2x$ (à droite)	126
4.2	solutions exacte $u_{ex}(x, y) = \sin(2\pi x) \sin(2\pi y)$ (à gauche) et sa dérivée (à droite)	126

4.3	Une grille grossière $\Omega^H$ et une grille fine $\Omega_l^h$ . La région sombre est celle du grande activité $\Omega_l$ . L'interface $\Gamma$ est pointillé. Les grands points gris sont les noeuds intérieurs de la grille grossière. Les points noirs sont les noeuds intérieurs de la grille fine. Noter que $\Gamma = \partial\Omega_l$ dans ce cas. . . . .	129
4.4	La région de raffinement est la réunion de la partie grise et la partie noire. Le "safety region" est la partie noire. Le défaut est calculé seulement dans la partie grise. Dans ce cas, $\Omega_l^H$ et $\Omega_\epsilon^H$ sont formés de 9 et 1 points respectivement . . . . .	130
4.5	La solution exacte du test 2 (à gauche) et sa dérivée (à droite) . . . . .	137
4.6	La solution exacte du test 1 (à gauche) et sa dérivée (à droite). . . . .	144
4.7	A gauche la solution exacte du test 2 et à droite sa dérivée. . . . .	145
5.1	Notations correspondent à la face "arrière" ou "F" du cube $\Omega$ . Le vecteur $U_F$ est formé de petites boules. Les grands boules désignent l'intersection entre cette face et les autres faces du $\Omega$ . . . . .	153
5.2	Schéma HB en trois dimensions. Le flux est intégré sur la surface de ce cube $K_{i,j,k} = [x_i - h, x_i + h] \times [y_j - h, y_j + h] \times [z_k - h, z_k + h]$ . . . . .	162
5.3	Courbes de convergence pour le test 2 de $u$ (à gauche) et $u_x$ (à droite) . . . . .	194

# Liste des tableaux

2.1	Erreur et taux de convergence de Schéma HB1 avec $u_{ex}(x) = \cos(\pi x^2)$ et $u_{x,ex}(x) = -2\pi x \sin(\pi x^2)$ . . . . .	37
2.2	Erreur et taux de convergence de Schéma HB1 avec $u_{ex}(x) = \sin(\pi x^3)$ et $u_{x,ex}(x) = 3\pi x^2 \cos(\pi x^3)$ . . . . .	37
2.3	Erreur et taux de convergence de Schéma HB1 avec $u_{ex}(x) = \exp(-(x - \frac{1}{2})^2)$ . . . . .	38
2.4	Erreur et taux de convergence de Schéma HB1 avec $u_{ex}(x) = \exp(-1000(x - \frac{1}{2})^2)$ . . . . .	38
2.5	Erreur et taux de convergence de Schéma HB2 avec $u_{ex}(x) = \cos(\pi x^2)$ .	53
2.6	Erreur et taux de convergence de Schéma HB2 avec $u_{ex}(x) = \sin(\pi x^3)$ .	53
2.7	Erreur et taux de convergence de Schéma HB2 avec $u_{ex}(x) = \exp(-(x - \frac{1}{2})^2)$ . . . . .	54
2.8	Erreur et taux de convergence de Schéma HB2 avec $u_{ex}(x) = \exp(-1000(x - \frac{1}{2})^2)$ . . . . .	54
3.1	Erreur et taux de convergence pour le schéma HB1 avec $u_{ex}(x, y) = \exp(-((x - 0.5)^2 + (y - 0.5)^2))$ . . . . .	92
3.2	Erreur et taux de convergence du schéma HB1 avec $u_{ex}(x, y) = \exp(-400((x - 0.5)^2 + (y - 0.5)^2))$ . . . . .	92
3.3	Erreur et taux de convergence du schéma HB1 avec $u_{ex}(x, y) = \ln(x + y^2 + 1)$ . . . . .	94
3.4	Erreur et taux de convergence du schéma HB1 avec $u_{ex}(x, y) = \cos(5\pi(x - y)^3)$ . . . . .	94
3.5	Erreur et taux de convergence du schéma HB1 avec $u_{ex}(x, y) = \exp(-30((x - 0.5)^2 + (y - 0.5)^2)) \cos(20(x + y - 1))$ . . . . .	94
3.6	Conditionnement de la matrice $I_{12(N-1)} + M^T H_1^{-1} L$ . . . . .	103
3.7	Erreur et taux de convergence du schéma HB2 avec $u_{ex}(x, y) = \exp(-((x - 0.5)^2 + (y - 0.5)^2))$ . . . . .	105
3.8	Erreur et taux de convergence du schéma HB2 avec $u_{ex}(x, y) = \exp(-400((x - 0.5)^2 + (y - 0.5)^2))$ . . . . .	106
3.9	Erreur et taux de convergence du schéma HB2 avec $u_{ex}(x, y) = \ln(x + y^2 + 1)$ . . . . .	106

3.10	Erreur et taux de convergence du schéma HB2 avec $\cos(5\pi(x - y)^3)$ . . . . .	$u_{ex}(x, y) =$	107
3.11	Erreur et taux de convergence du schéma HB2 avec $\exp(-30((x - 0.5)^2 + (y - 0.5)^2)) \cos(20(x + y - 1))$ . . . . .	$u_{ex}(x, y) =$	107
4.1	Erreur et taux de convergence de Schéma HB2 avec $2[(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2]$ et $a(x, y) = [1 + \sin(\frac{\pi}{2}(x + y))]^2$ . . . . .	$u_{ex}(x, y) =$	124
4.2	Erreur et taux de convergence de Schéma HB2 avec $u_{ex}(x, y) = x^2 + y^2$ et $a(x, y) = 1 + x + y$ . . . . .		125
4.3	Erreur et taux de convergence de schéma HB2 avec $u_{ex}(x, y) = x^2 + y^2$ et $a(x, y) = 0.1 + x + y$ . . . . .		125
4.4	Erreur et taux de convergence du Schéma HB2 $u_{ex}(x, y) = \sin(2\pi x) \sin(2\pi y)$ et $a(x, y) = (4 + \sin(4\pi x) \sin(4\pi y))^2$ après 5 itérations . . . . .		127
4.5	Erreur et taux de convergence du Schéma HB2 avec $u_{ex}(x, y) = \sin(2\pi x) \sin(2\pi y)$ et $a(x, y) = (4 + \sin(4\pi x) \sin(4\pi y))^2$ après 10 itérations . . . . .		127
4.6	Erreur et taux de convergence pour le test 1 sur une grille uniforme . . . . .		135
4.7	Erreur et taux de convergence pour le test 1 sur une grille composée . . . . .		135
4.8	Erreur et taux de convergence pour le test 2 sur une grille uniforme . . . . .		136
4.9	Erreur et taux de convergence pour test 2 sur une grille composée . . . . .		136
4.10	Erreur et taux de convergence du test 1 sur une grille uniforme . . . . .		145
4.11	Erreur et taux de convergence du test 1 sur une grille composée $\epsilon = 1/16$ . . . . .		145
4.12	Erreur et taux de convergence du test 2 sur une grille uniforme . . . . .		146
4.13	Erreur et taux de convergence du test 2 sur une grille composée $\epsilon = 1/16$ . . . . .		146
5.1	Erreur et taux de convergence du schéma HB1 avec $\cos(\pi x) \cos(\pi y) \cos(\pi z)$ . . . . .	$u_{ex}(x, y, z) =$	187
5.2	Erreur et taux de convergence du schéma HB1 avec $\ln(1 + x + y^2 + z^3)$ . . . . .	$u_{ex}(x, y, z) =$	187
5.3	Erreur et taux de convergence du schéma HB1 avec $\exp(-(x - 0.5)^2 - (y - 0.5)^2 - (z - 0.5)^2)$ . . . . .	$u_{ex}(x, y, z) =$	188
5.4	Erreur et taux de convergence du schéma HB1 avec $\exp(-30((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2)) \cos(20(x + y + z - 1))$ . . . . .	$u_{ex}(x, y, z) =$	188
5.5	Erreur et taux de convergence du schéma HB2 avec $\cos(\pi x) \cos(\pi y) \cos(\pi z)$ . . . . .	$u_{ex}(x, y, z) =$	193
5.6	Erreur et taux de convergence du schéma HB2 avec $\ln(1 + x + y^3 + z^2)$ . . . . .	$u_{ex}(x, y, z) =$	193
5.7	Erreur et taux de convergence du schéma HB2 avec $\exp(-((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2))$ . . . . .	$u_{ex}(x, y, z) =$	194
5.8	Erreur et taux de convergence du schéma HB2 avec $\exp(-30((x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2)) \cos(20(x + y + z - 1))$ . . . . .	$u_{ex}(x, y, z) =$	195

# Bibliographie

- [1] A. Abbas and J-P. Croisille. A fourth order Hermitian Box-Scheme with fast solver for the Poisson problem in a square. *J. Sci. Comput.*, 2011.
- [2] M.G.H. Anthonissen. *Local Defect correction Techniques : Analysis and Application to combustion*. PhD thesis, Eindhoven University of Technology, 2001.
- [3] M. Ben-Artzi, J-P. Croisille, and D. Fishelov. A fast direct solver for the biharmonic problem in a rectangular grid. *SIAM J. Sci. Comput.*, 31(1) :303–333, 2008.
- [4] M. Ben-Artzi, J-P. Croisille, and D. Fishelov. *Navier-Stokes equations in planar domains*. Imperial College Press, ISBN 9781848162754, 2012, to appear.
- [5] B. Bialecki, G. Fairweather, and A. Karageorghis. Matrix decomposition algorithms for elliptic boundary value problems : a survey. *Numerical Algorithms*, 56(2) :253–295, 2011.
- [6] P. Bjorstad. Fast numerical solution of the biharmonic Dirichlet problem on rectangles. *SIAM J. Numer. Anal.*, 20, No. 1 :59–71, 1983.
- [7] R.F. Boisvert. A fourth order accurate Fourier method for the Helmholtz equation in three dimensions. *ACM Trans. Math. Soft.*, 13 :221–234, 1987.
- [8] E. Braverman, B. Epstein, M. Israeli, and A. Averbuch. A fast spectral subtractive solver for elliptic equations. *J. Sci. Comput.*, 21(1) :91–128, 2004.
- [9] E. Braverman, M. Israeli, and A. Averbuch. A hierarchical 3-D direct Helmholtz solver by domain decomposition and modified Fourier method. *SIAM J. Sci. Comput.*, 26(5) :1504–1524, 2005.
- [10] L. Collatz. *The Numerical Treatment of Differential Equations*. Springer-Verlag, 3-rd edition, 1960.

- 
- [11] P. Concus and G.H. Golub. Use of fast direct methods for the efficient numerical solution of nonseparable elliptic equations. *SIAM Journal on Numerical Analysis*, Vol. 10, 1973.
- [12] J-P. Croisille. A Hermitian Box-Scheme for one-dimensional elliptic equations - Application to problems with high contrasts in the ellipticity. *Computing*, 78 :329–353, 2006.
- [13] P. Esposito, N. Ghoussoub, and Y. Guo. *Mathematical Analysis of Partial Differential Equations Modeling Electrostatic MEMS*, volume 20. Courant Lecture Notes, 2010.
- [14] P.J.J. Ferret and A.A. Reusken. Further Analysis of the Local Defect Correction Method. *Computing*, 56 :117–139, 1996.
- [15] G.E. Forsythe and W.R. Wasow. *Finite Difference Methods for Partial Differential Equations*. Applied Mathematics Series. John Wiley & Sons, 6th edition, 1960.
- [16] G.H. Golub, L.C. Huang, H. Simon, and W-P. Tang. A fast Poisson solver for the finite difference solution of the incompressible Navier-Stokes equations. *SIAM J. Sci. Comput.*, 19(5) :1606–1624, 1998.
- [17] G.H. Golub and C.F. Van Loan. *Matrix computations*. John Hopkins Univ. Press., 1996, 3rd edition.
- [18] L. Grasedyck. Existence and computation of low Kronecker-Rank approximations for large linear systems of tensor product structure. *Computing*, 72 :247–265, 2004.
- [19] I. Greff. *Schémas boîte : Etude théorique et numérique*. PhD thesis, Université de Metz, France, 2003.
- [20] B. Gustafsson. *High Order Difference Methods for Time Dependent PDE*. Springer-Verlag, 2008.
- [21] W. Hackbusch. *Elliptic Differential Equations - Theory and Numerical Treatment*, volume 15 of *Springer Series in Comp. Math.* Springer-Verlag, 1992.
- [22] W. Hackbusch. Local Defect Correction Method and Domain Decomposition Techniques. *Computing.*, Suppl. 5 :89–113, 1984.
- [23] D.A. Harville. *Matrix algebra from a statistician perspective*. Springer, 2008.



- 
- [24] A. Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge Univ. Press, 1996.
- [25] H. B. Keller. A new difference scheme for parabolic problems. In *Numerical Solution of Partial Differential Equations, II (SYNSPADE 1970) (Proc. Sympos., Univ. of Maryland, College Park, Md., 1970)*, pages 327–350. Academic Press, New York, 1971.
- [26] A. E. Lindsay and M. J. Ward. Asymptotics of some nonlinear eigenvalue problems modelling a mems capacitor : Part ii : Multiple solutions and singular asymptotics. *Europ. J. Appl. Math.*, 2010, (33 pages ).
- [27] C. Van Loan. *Computational Frameworks for the Fast Fourier Transform*. SIAM, 1992.
- [28] P. Londrillo. Adaptive grid-based gas-dynamics and Poisson solvers for gravitating systems. *Mem. A.A. It. Suppl.*, 4(69) :69–74, 2004.
- [29] A.R. Mitchell and D.F. Griffiths. *The Finite Difference Method in Partial Differential Equations*. John Wiley & Sons, 1980.
- [30] C. Montijn, W. Hundsdorfer, and U. Ebert. An adaptive grid refinement strategy for the simulation of negative streamers. *Journal of Computational Physics*, pages 801–835, 2006.
- [31] J. A. Pelesko. Mathematical modeling of electrostatic mems with tailored dielectric properties. *SIAM J. Appl. Math.*, pages 888–908, 2002.
- [32] J.B. Perot and V. Subramanian. A discrete calculus analysis of the keller box scheme and a generalisation of the method to arbitrary meshes. *Journal of Computational Physics*, pages 494–508, 2007.
- [33] J.R. Rice and R.F. Boisvert. *Solving elliptic problems using ELLPACK*. Springer-Verlag, 1985.
- [34] K. Shiraishi and T. Matsuoka. Wave propagation simulation using the CIP method of characteristics equations. *Comm. Comput. Physics*, 3(1) :121–135, 2008.
- [35] P. Swarztrauber. Fast Fourier Transform Algorithms for Vector Computers. *Parallel Computing*, pages 45–63, 1984.

- 
- [36] Yin Wang and Jun Zhang. Sixth order compact scheme combined with multigrid method and extrapolation technique for 2d Poisson equation. *J. Comput. Phys.*, 2009.
- [37] Juncheng Wei and Dong Ye. On mems equation with fringing field. *Proc. Amer. Math. Soc.*, 2010, (8 pages).
- [38] J. Zhang. Fast and high accuracy multigrid solution of the three dimensional poisson equation. *Journal of Computational Physics*, pages 449–461, 1998.