



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

Ali Abbas

## ► To cite this version:

Ali Abbas. Schémas boîte hermitiens - Algorithmes rapides pour la discrétisation des équations aux dérivées partielles. Analyse numérique [math.NA]. Université Paul Verlaine Metz, 2011. Français. NNT : 2011METZ024S . tel-01749061v2

**HAL Id: tel-01749061**

**<https://hal.science/tel-01749061v2>**

Submitted on 16 Mar 2016

**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.

**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, Examinateur
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, Examinateur
M. Dong YE,	Professeur à l'Université de Metz. Examinateur



# 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 envoie 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

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

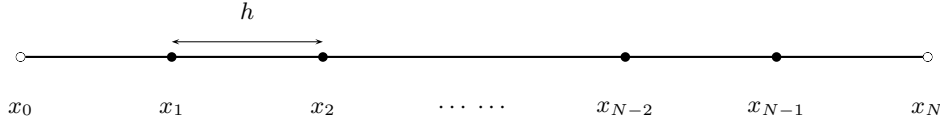


FIG. 1.1: Grille en dimension 1. Les deux points de bord sont  $x_0, x_N$  (notés par “o”). Les points  $x_i, i = 1, \dots, N-1$  sont les points intérieurs (notés par “•”).

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  $e_i(h)$  commis 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*

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

**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)

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

alors

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

d'où (1.56)<sub>a</sub>. En plus,

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

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

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

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

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

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*

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

**Preuve.**

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

d'où (1.71)<sub>a</sub>.

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

d'où (1.71)<sub>b</sub>. ■