

# Chapitre 3

## Approximation des Équations aux Dérivées partielles

### 3.1 Classification des EDP linéaires du second ordre

#### 3.1.1 Opérateur aux dérivées partielles

Soit  $u$  une fonction définie dans  $\mathbb{R}^n$  et  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$  un multi-indice. La longueur de  $\alpha$  est définie par

$$|\alpha| = \alpha_1 + \dots + \alpha_n.$$

On définit l'opérateur

$$D^\alpha u = \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}.$$

**Exemple**  $\alpha = (1, 1)$ ,  $x = (x_1, x_2)$ , dans ce cas  $|\alpha| = 2$  d'où

$$D^\alpha = \frac{\partial^2 u}{\partial x_1 \partial x_2}.$$

**Définition 3.1** On appelle équation aux dérivées partielles ou EDP une équation de la forme

$$F(x, u, Du, \dots, D^\alpha u) = 0 \tag{3.1}$$

où  $u$  est la fonction inconnue de la variable  $x = (x_1, \dots, x_n)$  et  $\alpha = (\alpha_1, \dots, \alpha_n)$  est un multi-indice.

L'équation (3.1) est dite du premier ordre si  $|\alpha| = 1$  et du second ordre si  $|\alpha| = 2$ .

**Définition 3.2** L'EDP (3.1) est dite :

- linéaire si  $F$  est linéaire en  $u, Du, \dots, d^\alpha u$ .
- semi-linéaire si  $F$  est linéaire en  $Du, \dots, D^\alpha u$ .
- quasi-linéaire si  $F$  est linéaire en  $D^\alpha u$ .
- non linéaire si  $F$  n'est pas linéaire en au moins une dérivée.

**Définition 3.3** On dit qu'un problème est bien posé (au sens de Hadamard) si la solution existe et est unique et dépend continument des données initiales et au bord.

### 3.1.2 Classification des EDP linéaires du second ordre

Soit  $\Omega \subset \mathbb{R}^n$  un domaine (ouvert et connexe) et considérons l'EDP linéaire du second ordre suivante

$$\sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial u}{\partial x_i} + c(x)u = f(x), \quad (3.2)$$

où  $x \in \Omega$ . Les coefficients  $a_{ij}, b_i, c$  sont des fonctions définies sur  $\Omega$ . Soit  $x_0 \in \Omega$  et considérons la matrice définie par

$$A(x_0) = (a_{ij}(x_0))_{1 \leq i,j \leq n}$$

et notons par  $N_+(x_0)$  le nombre de valeurs propres de  $A(x_0)$  strictement positives, par  $N_-(x_0)$  celles qui sont strictement négatives et  $N_0(x_0)$  celle qui sont nulles. Notons par

$$N(x_0) = N_+(x_0) + N_-(x_0) + N_0(x_0)$$

le nombre total de valeurs propres.

**Définition 3.4** L'équation (3.2) est dite

- elliptique en  $x_0$  si  $N(x_0) = N_+(x_0)$  ou  $N(x_0) = N_-(x_0)$ .
- hyperbolique en  $x_0$  si  $N(x_0) = N_0(x_0) - 1$  et  $N_+(x_0) = 1$  ou bien si  $N(x_0) = N_+(x_0) - 1$  et  $N_0(x_0) = 1$
- parabolique en  $x_0$  si  $N_0(x_0) > 0$

L'équation (3.2) est dite elliptique, parabolique ou hyperbolique dans  $\Omega$  si elle possède cette propriété en tout point  $x \in \Omega$ .

**Exemples***L'équation de Poisson*Soit  $\Omega \subset \mathbb{R}^n$  un domaine et  $f : \Omega \rightarrow \mathbb{R}$  une fonction donnée.

Considérons l'équation suivante dite de Poisson

$$-\Delta u(x) = f(x), \quad x \in \Omega$$

où  $\Delta$  est l'opérateur Laplacien défini par

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \dots + \frac{\partial^2}{\partial x_n^2}.$$

On a en tout point de  $x \in \Omega$ 

$$a_{ij}(x) = \begin{cases} -1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

d'où

$$A(x) = \begin{pmatrix} -1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -1 \end{pmatrix}_{n \times n}$$

dont l'unique valeur propre est  $\lambda = -1$ . Donc  $N_0 = N_+ = 0$  et  $N_- = 1$ . L'équation de Poisson est donc de type elliptique sur  $\Omega$ .*L'équation de la chaleur*Soit  $\Omega \subset \mathbb{R}^n$  un ouvert,  $I \subset \mathbb{R}$  un intervalle et  $f : I \times \Omega \rightarrow \mathbb{R}$  une fonction donnée. On appelle équation de la chaleur l'équation suivante

$$\frac{\partial u}{\partial t}(t, x) - \Delta u(t, x) = f(t, x), \quad t \in I, \quad x \in \Omega.$$

On a en tout point  $(t, x) \in I \times \Omega$ 

$$a_{ij}(t, x) = \begin{cases} 0, & \text{si } i = j = 1, \\ -1, & \text{si } i = j \geq 2 \\ 0, & \text{si } i \neq j \end{cases}$$

d'où la matrice  $A(x)$  suivante

$$A(x) = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & -1 & \ddots & 0 \\ \vdots & \ddots & -1 & 0 \\ 0 & \dots & 0 & -1 \end{pmatrix}_{(n+1) \times (n+1)}$$

dont les valeurs propres sont  $\lambda_1 = 0$ ,  $\lambda_2 = -1$  d'où  $N_+ = 0$ ,  $N_- = 1$ ,  $N_0 = 1$ . L'équation de la chaleur est donc de type parabolique.

*L'équation des ondes*

Considérons maintenant l'équation suivante dite équation des ondes

$$\frac{\partial^2 u}{\partial t^2}(t, x) - \Delta u(t, x) = f(t, x), \quad t \in I, \quad x \in \Omega.$$

La matrice  $A(x)$  s'écrit dans ce cas

$$A(x) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 \end{pmatrix}_{(n+1) \times (n+1)}$$

qui possède les valeurs propres suivantes  $\lambda_1 = 1$ ,  $\lambda_2 = -1$  d'où  $N_+ = 1$ ,  $N_- = 1$ ,  $N_0 = 0$ . L'équation des ondes est donc de type hyperbolique.

### 3.1.3 Équation de Poisson

Soit  $\Omega \subset \mathbb{R}^n$  un ouvert borné et  $f : \Omega \rightarrow \mathbb{R}$  une fonction continue. Considérons l'équation de Poisson

$$-\Delta u(x) = f(x), \quad x \in \Omega, \tag{3.3}$$

qui modélise des phénomènes d'équilibre, potentiel électrostatique, membrane en équilibre, champ gravitationnel, etc.

Si  $f = 0$  on obtient l'équation dite de Laplace

$$-\Delta u(x) = 0,$$

et on dit dans ce cas que  $u$  est une fonction harmonique. Associons à l'équation (3.3) l'une des données aux bord suivantes

$$u(x) = g(x), \quad x \in \partial\Omega \tag{3.4}$$

$$\frac{\partial u}{\partial n}(x) = g(x), \quad x \in \partial\Omega \tag{3.5}$$

où  $g : \partial\Omega \rightarrow \mathbb{R}$  est une fonction continue et  $\frac{\partial u}{\partial n}(x)$  désigne la dérivée normale de  $u$  en  $x$ . Le problème (3.3)-(3.4) est appelé problème de Dirichlet alors que le problème (3.3)-(3.5) s'appelle problème de Neumann.

### Principe du maximum

**Théorème 3.1** (*Principe du maximum faible*)

Soit  $u \in C^1(\Omega) \cap C(\overline{\Omega})$  tel que  $\Delta u \geq 0$  sur  $\Omega$ . Alors

$$\max_{\overline{\Omega}} u = \max_{\partial\Omega} u$$

**Théorème 3.2** (*Principe du maximum fort*)

Soit  $u \in C^1(\Omega) \cap C(\overline{\Omega})$  tel que  $\Delta u = 0$  sur un domaine  $\Omega$  de  $\mathbb{R}^n$ , alors ou bien  $u$  est constante ou bien

$$\min_{\partial\Omega} u < u(x) < \max_{\partial\Omega} u, \quad \forall x \in \Omega.$$

### 3.1.4 Équation de la chaleur

Soit  $\Omega \subset \mathbb{R}^n$  un domaine borné de frontière  $\partial\Omega$  suffisamment régulière et soit  $f : (0, T) \times \Omega \rightarrow \mathbb{R}$  une fonction continue. On définit l'équation de la chaleur par

$$\frac{\partial u}{\partial t}(t, x) - \Delta u(t, x) = f(t, x), \quad t \in (0, T], \quad x \in \Omega \quad (3.6)$$

à laquelle on associe l'une des deux conditions aux bord

$$u(x, t) = g(x, t), \quad x \in \partial\Omega, \quad t \in (0, T] \quad (3.7)$$

$$\frac{\partial u}{\partial n}(x, t) = g(x, t), \quad x \in \partial\Omega, \quad t \in (0, T] \quad (3.8)$$

ainsi que la donnée initiale

$$u(0, x) = u_0(x), \quad x \in \Omega. \quad (3.9)$$

L'équation de la chaleur (3.6) modélise des phénomènes de diffusion de la chaleur, transfert de particules dans l'air etc.

### Principe du maximum

Définissons  $D \subset \mathbb{R}^{n+1}$  par  $D = (0, T) \times \Omega$  et posons

$$\partial_D \Omega = \partial\Omega \times [0, T] \cup \Omega \times \{0\}$$

$\partial_D \Omega$  est appelé frontière parabolique de  $D$ .

**Théorème 3.3** Soit  $u \in C^{1,2}((0, T) \times \Omega) \cap C([0, T] \times \overline{\Omega})$  une fonction telle que

$$u_t - \Delta u \leq 0, \quad \text{sur } (0, T) \times \Omega$$

alors

$$\max_{[0, T] \times \overline{\Omega}} u(t, x) = \max_{\partial_D \Omega} u(t, x)$$

où  $u_t = \frac{\partial u}{\partial t}$ . Autrement dit le maximum de  $u$  est atteint sur la frontière parabolique de  $D$ .

### 3.1.5 L'équation des ondes

Soit  $\Omega \subset \mathbb{R}^n$  un ouvert borné de frontière  $\partial\Omega$  supposée de classe  $C^1$  et  $f : (0, T) \times \Omega \rightarrow \mathbb{R}$  une fonction continue. L'équation

$$\frac{\partial^2 u}{\partial t^2}(t, x) - \Delta u(t, x) = f(t, x), \quad t \in (0, T] \quad x \in \Omega \quad (3.10)$$

est appelée équation des ondes et modélise des phénomènes de vibration en acoustique, champ électromagnétique, ect. On lui associe l'une des deux conditions aux bord suivantes

$$u(t, x) = g(t, x), \quad t \in (0, T], \quad x \in \partial\Omega, \quad (3.11)$$

$$\frac{\partial u}{\partial n}(t, x) = g(t, x), \quad t \in (0, T], \quad x \in \partial\Omega, \quad (3.12)$$

ainsi que les deux conditions initiales

$$\begin{aligned} u(0, x) &= h_1(x), & x \in \Omega \\ \frac{\partial u}{\partial t}(0, x) &= h_2(x), & x \in \partial\Omega. \end{aligned} \quad (3.13)$$

## 3.2 Discrétisation de l'équation de Poisson 1d

### 3.2.1 Équation de Poisson 1d avec conditions aux bord de type Dirichlet

Soit  $\Omega = (a, b)$  un intervalle borné de  $\mathbb{R}$  et notons par  $\partial\Omega = \{a, b\}$  la frontière de  $\Omega$ . Considérons le problème suivant

$$\begin{cases} -u_{xx}(x) = f(x), & x \in \Omega \\ u(a) = \alpha, \quad u(b) = \beta, & x \in \partial\Omega \end{cases} \quad (3.14)$$

Considérons une subdivision uniforme de pas  $h$  de l'intervalle  $(a, b)$

$$a = x_0 < x_1 < \dots < x_n = b,$$

où les  $x_i$  sont donnés par  $x_i = a + ih$ ,  $i = 0, \dots, n$ . On a en chaque noeud  $x_i$  de  $\Omega$

$$-u_{xx}(x_i) = f(x_i), \quad i = 1, \dots, n-1. \quad (3.15)$$

La méthode des différences finies consiste à approcher la dérivée seconde  $u_{xx}(x_i)$  par un quotient de manière à en déduire un système linéaire d'équations dont la solution représente l'approximation de la solution exacte  $u$  en  $x_i$ . En utilisant la relation (A.16) de l'annexe on a

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

Posons  $f_i = f(x_i)$  et notons par  $u_i$  une valeur approchée de  $u(x_i)$ , on obtient alors le schéma numérique suivant

$$\begin{cases} \frac{2u_i - u_{i+1} - u_{i-1}}{h^2} = f_i, & i = 1, \dots, n-1 \\ u_0 = \alpha, \quad u_n = \beta \end{cases} \quad (3.16)$$

qui représente un système linéaire de  $(n-1)$  équations à  $(n-1)$  inconnues.

**Remarque** Il ne faut pas confondre  $u_i$  et  $u(x_i)$ , le système (3.16) fait intervenir les inconnues discrètes  $u_i$  et non pas les valeurs  $u(x_i)$  de la solution exacte  $u$ .

Écrivons le système (3.16) sous forme matricielle. La première équation  $i = 1$  s'écrit

$$\frac{1}{h^2}(2u_1 - u_0 - u_2) = f_1$$

or puisque  $u_0 = \alpha$  on obtient

$$\frac{1}{h^2}(2u_1 - u_2) = \frac{\alpha}{h^2} + f_1.$$

L'équation  $i$ ,  $(2 < i < n-1)$  s'écrit

$$\frac{1}{h^2}(2u_i - u_{i+1} - u_{i-1}) = f_i,$$

et la dernière équation  $i = n-1$  s'écrit à son tour

$$\frac{1}{h^2}(2u_{n-1} - u_n - u_{n-2}) = f_{n-1}$$

or comme  $u_n = \beta$  la dernière équation prend la forme

$$\frac{1}{h^2}(2u_{n-1} - u_{n-2}) = \frac{\beta}{h^2} + f_{n-1}.$$

Le système (3.16) s'écrit alors sous forme matricielle

$$A_h U_h = b_h, \quad (3.17)$$

avec

$$A_h = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}, \quad b_h = \begin{pmatrix} f_1 + \frac{\alpha}{h^2} \\ f_2 \\ \vdots \\ f_{n-2} \\ f_{n-1} + \frac{\beta}{h^2} \end{pmatrix}. \quad (3.18)$$

On a déjà montré (voir la relation (1.1)) que la matrice  $A_h$  est monotone, elle donc inversible.

**Proposition 3.1** Pour tout  $h > 0$ , Le système (3.17) admet une solution unique  $U_h = (u_1, \dots, u_{n-1})^\top$  qui n'est autre que la solution approchée du problème (3.14).

**Définition 3.5** On appelle erreur de consistance  $R_i$  en  $x_i$  la quantité obtenue en remplaçant la solution approchée par la solution exacte dans le schéma numérique (3.16).

On dira que le schéma (3.16) est consistant d'ordre  $p \geq 1$  si

$$\max_{1 \leq i \leq n-1} |R_i| = o(h^p). \quad (3.19)$$

En particulier si  $p = 1$  alors le schéma est dit tout simplement consistant.

D'après la définition précédente l'erreur de consistance s'écrit

$$R_i = \frac{2u(x_i) - u(x_{i+1}) - u(x_{i-1}))}{h^2} - f_i \quad (3.20)$$

or d'après le schéma (3.16)

$$f_i = \frac{2u_i - u_{i+1} - u_{i-1}}{h^2},$$

d'où

$$R_i = \frac{2(u(x_i) - u_i) - (u(x_{i+1}) - u_{i+1}) - (u(x_{i-1}) - u_{i-1}))}{h^2}. \quad (3.21)$$

Posons  $R = (R_1, \dots, R_{n-1})^\top$ , la relation (3.20) s'écrit alors sous forme matricielle

$$R = A_h(U - U_h) \quad (3.22)$$

avec  $U = (u(x_1), \dots, u(x_{n-1}))^\top$ .

**Lemme 3.1** Supposons que  $u \in C^4([a, b])$ , alors

$$\max_{1 \leq i \leq n-1} |R_i| \leq \frac{h^2}{12} \sup_{x \in [a, b]} |u^{(4)}(x)|. \quad (3.23)$$

**Preuve** Écrivons les développements limités de  $u$  à l'ordre 4

$$\begin{aligned} u(x_{i+1}) &= u(x_i) + hu'(x_i) + \frac{h^2}{2}u''(x_i) + \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\xi_i), \\ u(x_{i-1}) &= u(x_i) - hu'(x_i) + \frac{h^2}{2}u''(x_i) - \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\eta_i), \end{aligned}$$

avec  $\xi_i, \eta_i \in (x_{i-1}, x_{i+1})$ . En reportant les deux expressions précédentes dans (3.19) et puisque  $-u''(x_i) = f_i$  on obtient

$$R_i = -\frac{h^2}{24} (u^{(4)}(\xi_i) + u^{(4)}(\eta_i))$$

soit

$$|R_i| \leq \frac{h^2}{12} \sup_{x \in [a, b]} |u^{(4)}(x)|.$$

□

La relation (3.23) entraîne que  $|R_i| \rightarrow 0$  lorsque  $h \rightarrow 0$  pour tout  $1 \leq i \leq n-1$ . Le schéma (3.16) est donc consistant d'ordre 2.

**Définition 3.6** On appelle erreur d'approximation  $e_i$  en  $x_i$  la différence entre la solution exacte et la solution approchée en  $x_i$

$$e_i = u(x_i) - u_i, \quad 0 \leq i \leq n.$$

Le lemme suivant est l'équivalent du principe du maximum dans le cas discret.

**Lemme 3.2** (*Principe du maximum discret*)

Soit  $v = (v_1, \dots, v_{n-1}) \in \mathbb{R}^{n-1}$  un vecteur satisfaisant le problème suivant

$$\begin{cases} 2v_i - v_{i+1} - v_{i-1} \leq 0, & 1 \leq i \leq n-1, \\ v_0 = \alpha, \quad v_n = \beta \end{cases} \quad (3.24)$$

alors

$$\max_{1 \leq i \leq n-1} v_i \leq \max(\alpha, \beta),$$

autrement dit le maximum est atteint aux extrémités  $\alpha$  ou  $\beta$ .

**Preuve** Définissons  $i_0$  comme suit

$$i_0 = \min\{i : v_i = \max_{1 \leq j \leq n-1} v_j\},$$

et supposons que  $1 \leq i_0 \leq n-1$ . L'inégalité (3.24) entraîne alors

$$(v_{i_0} - v_{i_0+1}) + (v_{i_0} - v_{i_0-1}) \leq 0 \quad (3.25)$$

or  $v_{i_0} \geq v_{i_0+1}$  et par construction  $v_{i_0} > v_{i_0-1}$  d'où

$$(v_{i_0} - v_{i_0+1}) + (v_{i_0} - v_{i_0-1}) > 0$$

ce qui contredit (3.25). Donc  $i_0 = 0$  ou bien  $i_0 = n$  c'est à dire que

$$\max_{1 \leq i \leq n-1} v_i \leq \max(\alpha, \beta). \quad \square$$

On a le résultat de convergence suivant.

**Théorème 3.4** Supposons que  $u \in C^4([a, b])$ . Alors on a l'estimation d'erreur suivante

$$\max_{1 \leq i \leq n-1} |e_i| \leq \frac{h^2}{24} \max(a^2, b^2) \sup_{x \in [a, b]} |u^{(4)}(x)|. \quad (3.26)$$

**Preuve** Posons  $M = \max_{1 \leq i \leq n-1} |R_i|$  et définissons  $e_i^+$  et  $e_i^-$  par

$$\begin{aligned} e_i^+ &= e_i + \frac{1}{2} M x_i^2, & 0 \leq i \leq n \\ e_i^- &= -e_i + \frac{1}{2} M x_i^2, & 0 \leq i \leq n \end{aligned}$$

on a alors

$$2e_i^+ - e_{i+1}^+ - e_{i-1}^+ = 2e_i - e_{i+1} - e_{i-1} + \frac{1}{2} M (2x_i^2 - x_{i+1}^2 - x_{i-1}^2)$$

or d'après (3.21)  $2e_i - e_{i+1} - e_{i-1} = h^2 R_i$  et puisque

$$2x_i^2 - x_{i+1}^2 - x_{i-1}^2 = -2h^2 \text{ il s'en suit que}$$

$$2e_i^+ - e_{i+1}^+ - e_{i-1}^+ = h^2 (R_i - M) \leq 0. \quad (3.27)$$

De même

$$2e_i^- - e_{i+1}^- - e_{i-1}^- = -(R_i + M)h^2 \leq 0. \quad (3.28)$$

Le Lemme 3.2 appliqué à (3.27) et (3.28) entraîne

$$\begin{aligned} \max_{1 \leq i \leq n-1} e_i^+ &\leq \max(e_0^+, e_n^+), \\ \max_{1 \leq i \leq n-1} e_i^- &\leq \max(e_0^-, e_n^-). \end{aligned} \quad (3.29)$$

Or

$$\begin{aligned} e_0^+ &= e_0 + \frac{1}{2}Mx_0^2 = \frac{1}{2}Ma^2, \\ e_n^+ &= e_n + \frac{1}{2}Mx_n^2 = \frac{1}{2}Mb^2, \\ e_0^- &= -e_0 + \frac{1}{2}Mx_0^2 = \frac{1}{2}Ma^2, \\ e_n^- &= -e_n + \frac{1}{2}Mx_n^2 = \frac{1}{2}Mb^2, \end{aligned}$$

(3.29) entraîne alors

$$\max_{1 \leq i \leq n-1} |e_i| \leq \frac{1}{2}M \max(a^2, b^2). \quad (3.30)$$

D'un autre coté d'après le Lemme 3.1

$$M = \max_{1 \leq i \leq n-1} |R_i| \leq \frac{h^2}{12} \sup_{x \in [a,b]} |u^{(4)}(x)|$$

la relation (3.30) devient donc

$$\max_{1 \leq i \leq n-1} |e_i| \leq \frac{h^2}{24} \max(a^2, b^2) \sup_{x \in [a,b]} |u^{(4)}(x)|.$$

□

**Remarque** L'estimation d'erreur (3.26) montre que la solution approchée  $U_h$  converge vers la solution exacte  $u$  lorsque  $h \rightarrow 0$ . On dit que le schéma (3.16) est convergent d'ordre 2.

**Exemple** Calculer une solution approchée du problème de Dirichlet suivant

$$\begin{cases} -u''(x) = x^2, & x \in (0, 1) \\ u(0) = 1, u(1) = 2, \end{cases}$$

en prenant les points équidistant  $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$  de l'intervalle  $(0, 1)$ . Comme  $h = \frac{1}{4}$  la matrice  $A_h$  s'écrit

$$A_h = 16 \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 32 & -16 & 0 \\ -16 & 32 & -16 \\ 0 & -16 & 32 \end{pmatrix},$$

on calcule ensuite le second membre  $b_h$ . Calculons d'abord le vecteur  $f = (f_1, f_2, f_3)^\top$ , on a

$$\begin{aligned} f_1 &= f(x_1) = x_1^2 = \frac{1}{16} \\ f_2 &= f(x_2) = x_2^2 = \frac{1}{4} \\ f_3 &= f(x_3) = x_3^2 = \frac{9}{16} \end{aligned}$$

d'où  $b_h = (16.0625, 0.25, 32.5625)^\top$ . Le système linéaire s'écrit alors

$$\begin{pmatrix} 32 & -16 & 0 \\ -16 & 32 & -16 \\ 0 & -16 & 32 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 16.0625 \\ 0.25 \\ 32.5625 \end{pmatrix}.$$

L'algorithme de Thomas (1.7) donne

$u_1 = 1.2696$ ,  $u_2 = 1.5355$ ,  $u_3 = 1.7851$ , ce qui signifie

$$u(0.25) \simeq 1.2696, \quad u(0.5) \simeq 1.5355, \quad u(0.75) \simeq 1.7851.$$

La solution exacte est  $u(x) = -\frac{1}{12}x^4 + \frac{13}{12}x + 1$ , d'où les erreurs commises aux noeuds  $\{0.25, 0.5, 0.75\}$

$$e_1 = 0.0011, \quad e_2 = 0.0013, \quad e_3 = 0.001.$$

### 3.2.2 Équation de Poisson 1d avec conditions aux bord de type de Neumann

Soit  $c : [a, b] \rightarrow \mathbb{R}$  une fonction donnée et considérons le problème modèle suivant

$$\begin{cases} -u''(x) + c(x)u(x) = f(x), & x \in (a, b) \\ u'(a) = \alpha, \quad u'(b) = \beta. \end{cases} \quad (3.31)$$

Si  $c \equiv 0$  le problème (3.31) est mal posé car si  $u$  est une solution alors  $u + k$  est également une solution  $\forall k \in \mathbb{R}$ . On supposera donc que  $c > 0$ . Soit  $a = x_0 < \dots < x_n = b$  une grille uniforme de pas  $h = x_{i+1} - x_i$ . On a aux noeuds internes de la grille

$$-u''(x_i) + c(x_i)u(x_i) = f(x_i), \quad 1 \leq i \leq n-1.$$

Posons  $c_i = c(x_i)$  et  $f_i = f(x_i)$ . En approchant  $u''(x_i)$  à l'aide du schéma à trois points et en notant par  $u_i$  une valeur approchée de  $u(x_i)$  on obtient l'équation aux différences

$$\frac{2u_i - u_{i+1} - u_{i-1}}{h^2} + c_i u_i = f_i, \quad i = 1, \dots, n-1.$$

D'autre part l'approximation des données aux bord donne

$$\begin{aligned} u'(a) &\simeq \frac{u(x_1) - u(a)}{h}, \\ u'(b) &\simeq \frac{u(b) - u(x_{n-1})}{h} \end{aligned}$$

d'ou finalement le schéma suivant

$$\begin{cases} \left(\frac{2}{h^2} + c_i\right) u_i - \frac{1}{h^2} u_{i-1} - \frac{1}{h^2} u_{i+1} = f_i, & 1 \leq i \leq n-1 \\ u_n = h\beta + u_{n-1}, \quad u_0 = u_1 - \alpha h \end{cases} \quad (3.32)$$

**Remarque** Contrairement au cas Dirichlet les valeurs  $u_0$  et  $u_n$  ne sont pas connues, on les calcule donc en fonction respectivement de  $u_1$  et  $u_{n-1}$ .

Écrivons la forme matricielle du schéma (3.32). Pour cela on remplace  $u_0$  et  $u_n$  dans la première et dernière équations respectivement par leurs valeurs tirées de la deuxième équation de (3.32). On a pour  $i = 1$

$$\left(\frac{2}{h^2} + c_1\right) u_1 - \frac{1}{h^2} u_2 = f_1 - \frac{\alpha}{h},$$

et en  $i = n-1$

$$\left(\frac{2}{h^2} + c_{n-1}\right) u_{n-1} - \frac{1}{h^2} u_{n-2} = f_{n-1} + \frac{\beta}{h},$$

d'où la forme matricielle  $A_h U^h = b_h$  du schéma (3.32) avec

$$A_h = \begin{pmatrix} \frac{1}{h^2} + c_1 & -\frac{1}{h^2} & 0 & \cdots & 0 \\ -\frac{1}{h^2} & \frac{2}{h^2} + c_2 & -\frac{1}{h^2} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \frac{2}{h^2} + c_{n-2} & -\frac{1}{h^2} \\ 0 & \cdots & 0 & -\frac{1}{h^2} & \frac{1}{h^2} + c_{n-1} \end{pmatrix}_{(n-1) \times (n-1)}$$

$$b_h = \begin{pmatrix} f_1 - \frac{\alpha}{h} \\ f_2 \\ \vdots \\ f_{n-2} \\ f_{n-1} + \frac{\beta}{h} \end{pmatrix} \quad (3.33)$$

$A_h$  est encore une matrice tridiagonale et symétrique. Montrons qu'elle est inversible. Pour cela on montre que  $A_h$  est monotone. Écrivons que

$$A_h x \geq 0$$

et montrons que  $x \geq 0$ . L'inégalité précédente s'écrit

$$\begin{cases} \left(\frac{2}{h^2} + c_1\right) x_1 - \frac{1}{h^2} x_2 \geq 0, \\ \left(\frac{2}{h^2} + c_i\right) x_i - \frac{1}{h^2} x_{i-1} - \frac{1}{h^2} x_{i+1} \geq 0, \quad 2 \leq i \leq n-2, \\ \left(\frac{2}{h^2} + c_{n-1}\right) x_{n-1} - \frac{1}{h^2} x_{n-2} \geq 0. \end{cases} \quad (3.34)$$

Posons  $i_0 = \min\{i : x_i = \min_{1 \leq j \leq n-1} x_j\}$ . Il suffit alors de montrer que  $x_{i_0} \geq 0$ .

Supposons que  $2 \leq i \leq n-2$ , dans ce cas

$$\left(\frac{2}{h^2} + c_{i_0}\right) x_{i_0} - \frac{1}{h^2} x_{i_0-1} - \frac{1}{h^2} x_{i_0+1} \geq 0,$$

d'où l'on tire puisque  $(x_{i_0-1} - x_{i_0}) > 0$  et  $(x_{i_0+1} - x_{i_0}) \geq 0$

$$c_{i_0} x_{i_0} \geq \frac{1}{h^2} (x_{i_0-1} - x_{i_0}) + \frac{1}{h^2} (x_{i_0+1} - x_{i_0}) > 0$$

Comme  $c_{i_0} > 0$  alors  $x_{i_0} > 0$ .

Dans le cas  $i_0 = 1$  on a alors de la première inégalité (3.34) et puisque  $(x_2 - x_1) \geq 0$

$$\left(\frac{1}{h^2} + c_1\right) x_1 \geq \frac{1}{h^2} (x_2 - x_1) \geq 0$$

d'où l'on déduit que  $x_1 \geq 0$ .

Dans le cas  $i_0 = n-1$  on a de la dernière inégalité (3.34) et puisque  $(x_{n-2} - x_{n-1}) \geq 0$

$$\left(\frac{1}{h^2} + c_{n-1}\right) x_{n-1} \geq \frac{1}{h^2} (x_{n-2} - x_{n-1}) \geq 0$$

ce qui donne  $x_{n-1} \geq 0$ . dans tous les cas on a montré que  $x \geq 0$ , la matrice  $A_h$  est donc monotone. On a montré le résultat suivant.

**Proposition 3.2** Le schéma (3.32) admet pour tout  $h > 0$  une solution unique  $U_h = (u_1, \dots, u_{n-1})^\top$ .

Comme dans le cas Dirichlet on peut montrer le résultat de convergence suivant. La preuve est laissée en exercice.

**Théorème 3.5** Sous les mêmes hypothèses que celle du Théorème 3.4 il existe une constante  $C > 0$  dépendant de  $a, b$  et de la solution exacte  $u$  telle que

$$\max_{1 \leq i \leq n-1} |e_i| \leq Ch^2, \quad (3.35)$$

où  $e_i = u(x_i) - u_i$  est l'erreur d'approximation en  $x_i$ .

**Exemple** Calculer une valeur approchée en  $x = 0$  et  $x = 1$  de la solution du problème suivant

$$\begin{cases} -u''(x) + xu(x) = x^2, & x \in (0, 1) \\ u'(0) = 0, \quad u'(1) = 1, \end{cases}$$

en prenant les noeuds  $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ , donc  $h = \frac{1}{4}$ . Écrivons le système matriciel avec  $c_1 = x_1 = 0.25$ ,  $c_2 = x_2 = 0.5$ ,  $c_3 = x_3 = 0.75$

$$A_h = \begin{pmatrix} 16.25 & -16 & 0 \\ -16 & 32.5 & -16 \\ 0 & -16 & 16.25 \end{pmatrix}, \quad b_h = \begin{pmatrix} 0.0625 \\ 0.25 \\ -3.4375 \end{pmatrix}$$

La résolution du système  $A_h U_h = b_h$  donne

$$U_h = (-0.1494, -0.1557, -0.1824)^\top$$

Pour calculer une valeur approchée de  $u(0)$  et  $u(1)$  on utilise les dernières relations du schéma (3.32) d'où  $u_0 = u_1 - \alpha h = -0.1494$  et  $u_4 = u_3 + \beta h = 0.0676$ , on a donc les approximations

$$u(0) \simeq -0.1494, \quad u(1) \simeq 0.0676.$$

### 3.3 Discrétisation de l'équation de Poisson 2d

On considère l'équation de Poisson sur un rectangle  $\Omega = (a, b) \times (c, d) \subset \mathbb{R}^2$  avec condition aux bord de type Dirichlet

$$\begin{cases} -\Delta u(x, y) = f(x, y), & (x, y) \in \Omega, \\ u(x, y) = g(x, y), & (x, y) \in \partial\Omega \end{cases} \quad (3.36)$$

où  $f : \Omega \rightarrow \mathbb{R}$  et  $g : \partial\Omega \rightarrow \mathbb{R}$  sont des fonctions données supposées continues et  $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$  est l'opérateur Laplacien.

L'idée pour approcher la solution du problème (3.36) consiste à approcher chacune des dérivées  $\frac{\partial^2 u}{\partial x^2}$ ,  $\frac{\partial^2 u}{\partial y^2}$  à l'aide d'un schéma à trois points sur une grille sur chaque axe. Considérons pour cela une grille uniforme de l'intervalle  $[a, b]$  sur l'axe des  $x$

$$a = x_0 < x_1 < \dots < x_n = b$$

de pas  $h_x = x_{i+1} - x_i$  et une seconde grille uniforme de l'intervalle  $[c, d]$  sur l'axe des  $y$

$$c = y_0 < y_1 < \dots < y_m = d$$

de pas  $h_y = y_{i+1} - y_i$ . Considérons la grille de calcul constituée des indices des noeuds situés à l'intérieur du rectangle  $(a, b) \times (c, d)$  ainsi que celle constituée des noeuds sur le bord de  $\Omega$

$$\begin{aligned}\Delta_h &= \{(i, j) : 1 \leq i \leq n-1, 1 \leq j \leq m-1\}, \\ \partial\Delta_h &= \{(i, 0), (i, m), (0, j), (n, j) : 0 \leq i \leq n, 0 \leq j \leq m\}\end{aligned}$$

Écrivons la première équation de (3.36) aux noeuds de  $\Delta_h$

$$-\frac{\partial^2 u}{\partial x^2}(x_i, y_j) - \frac{\partial^2 u}{\partial y^2}(x_i, y_j) = f(x_i, y_j),$$

puis on approche chacune des dérivées partielles par le schéma à trois points (A.16)

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2}(x_i, y_j) &\simeq \frac{u(x_{i+1}, y_j) + u(x_{i-1}, y_j) - 2u(x_i, y_j)}{h_x^2} \\ \frac{\partial^2 u}{\partial y^2}(x_i, y_j) &\simeq \frac{u(x_i, y_{j+1}) + u(x_i, y_{j-1}) - 2u(x_i, y_j)}{h_y^2}\end{aligned}$$

posons  $f_{ij} = f(x_i, y_j)$ , on obtient alors le schéma

$$\frac{2u_{ij} - u_{i+1j} - u_{i-1j}}{h_x^2} + \frac{2u_{ij} - u_{ij+1} - u_{ij-1}}{h_y^2} = f_{ij}$$

où  $u_{ij}$  dénote une valeur approchée de  $u(x_i, y_j)$ . Supposons pour simplifier que  $h_x = h_y = h$ , le schéma précédent devient

$$\frac{4u_{ij} - u_{i+1j} - u_{i-1j} - u_{ij+1} - u_{ij-1}}{h^2} = f_{ij} \quad (3.37)$$

avec les données aux bord

$$u_{ij} = g_{ij}, \quad (i, j) \in \partial\Delta_h, \quad (3.38)$$

où  $g_{ij} = g(x_i, y_j)$ . Le schéma (3.37) est un schéma à cinq points car pour calculer  $u_{ij}$  on a besoin de  $u_{i+1j}$ ,  $u_{i-1j}$ ,  $u_{ij+1}$  et  $u_{ij-1}$ .

Écrivons la forme matricielle du schéma (3.37)-(3.38). Pour cela posons

$$U^h = (u_{11}, u_{21}, \dots, u_{n-11}, u_{12}, u_{22}, \dots, u_{n-12}, \dots, u_{1m-1}, \dots, u_{n-1m-1})$$

ou de manière plus condensée

$$U^h = (u_1, \dots, u_{m-1}) \in \mathbb{R}^{(n-1) \times (m-1)}$$

avec

$$u_j = (u_{1j}, \dots, u_{n-1j}) \in \mathbb{R}^{n-1}, \quad 1 \leq j \leq m-1.$$

Pour construire la matrice  $P_h$  du schéma (3.37) regardons la matrice  $P_h$  comme une matrice bloc dans laquelle l'indice  $j$  désigne la position du bloc dans la matrice  $P_h$  et l'indice  $i$  désigne la position du terme  $u_{ij}$  au sein du même bloc. Le schéma (3.37) montre que la matrice  $P_h$  est tridiagonale par bloc, les blocs d'indices  $j$  sont ceux situés sur la diagonale et ceux ayant les indices  $j-1$  et  $j+1$  sont situés sur la sous-diagonale et la sur-diagonale respectivement. Le bloc d'indice  $j$  est lui même une matrice tridiagonale notée  $T$  dont les éléments de la diagonale sont égaux à 4 et ceux de la sous-diagonale et la sur-diagonale sont égaux à  $-1$ . La matrice  $P_h$  a donc la forme suivante

$$P_h = \frac{1}{h^2} \begin{pmatrix} T & \boxed{S_1} & \boxed{0} & \cdots & \boxed{0} \\ \boxed{S_2} & T & \boxed{S_1} & \ddots & \vdots \\ \boxed{0} & \ddots & \ddots & \ddots & \boxed{0} \\ \vdots & \ddots & \ddots & \ddots & \boxed{S_1} \\ \boxed{0} & \cdots & \boxed{0} & \boxed{S_2} & T \end{pmatrix}$$

où  $\boxed{0}$  désigne un bloc nul avec

$$T = \begin{pmatrix} 4 & -1 & 0 & \cdots & 0 \\ -1 & 4 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 4 \end{pmatrix} \quad (3.39)$$

qui est de taille  $(n-1) \times (n-1)$ . Le bloc  $S_1$  situé sur la sur-diagonale a pour indice  $j+1$  et d'après le schéma (3.37) le coefficient du terme diagonal  $u_{ij+1}$  est  $-1$  les autres coefficients étant nuls. De même pour le bloc  $S_2$  d'indice  $j-1$ , le coefficient de  $u_{ij-1}$  au sein de ce bloc est  $-1$  les autres étant nuls.  $S_1$  et  $S_2$  sont donc des matrices diagonales de la forme

$$S_1 = S_2 = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 \end{pmatrix}$$

de taille également  $(n-1) \times (n-1)$ . Finalement la matrice  $P_h$  prend la forme suivante

$$P_h = \frac{1}{h^2} \begin{pmatrix} T & -I & \boxed{0} & \cdots & \boxed{0} \\ -I & T & \ddots & \ddots & \vdots \\ \boxed{0} & \ddots & \ddots & \ddots & \boxed{0} \\ \vdots & \ddots & \ddots & \ddots & -I \\ \boxed{0} & \cdots & \boxed{0} & -I & T \end{pmatrix} \quad (3.40)$$

où  $I$  est la matrice identité de taille  $(n-1) \times (n-1)$ . La matrice  $P_h$  est donc de taille  $q \times q = (n-1)(m-1) \times (n-1)(m-1)$ .

Calculons maintenant le second membre  $B_h$  du système linéaire. Écrivons  $B_h$  sous la forme

$$B_h = (b_1, \dots, b_{m-1})^\top \quad (3.41)$$

où  $b_j = (b_{1,j}, \dots, b_{n-1,j})^\top$ ,  $1 \leq j \leq m-1$ . Chaque  $b_j$  est un vecteur correspondant au bloc  $j$  et est la somme du second membre  $f_{ij}$  du schéma (3.37) ainsi que des termes situés sur le bord  $\frac{u_{0,j}}{h^2}, \frac{u_{n,j}}{h^2}, \frac{u_{i,0}}{h^2}, \frac{u_{i,m}}{h^2}$ . Comme  $u_{0,j} = g_{0,j}, u_{n,j} = g_{n,j}, u_{i,0} = g_{i,0}, u_{i,m} = g_{i,m}$  on déduit donc que

Si  $1 < j < m-1$  dans ce cas les seuls termes  $u_{ij}$  situés sur le bord sont  $u_{0,j}$  et  $u_{n,j}$  qui correspondent aux valeurs  $i = 1$  et  $i = n-1$  de l'indice  $i$  respectivement. D'où

$$b_j = \left( \frac{g_{0,j}}{h^2} + f_{1,j}, f_{2,j}, \dots, f_{n-2,j}, \frac{g_{n,j}}{h^2} + f_{n-1,j} \right). \quad (3.42)$$

Si  $j = 1$ , dans ce cas il y a, en plus des termes précédents, le terme  $u_{i,0} = g_{i,0}$  qui également situé sur le bord, donc

$$b_1 = \left( \frac{g_{0,1}}{h^2} + f_{1,1} + \frac{g_{1,0}}{h^2}, f_{2,1} + \frac{g_{1,0}}{h^2}, \dots, f_{n-2,1} + \frac{g_{n-2,0}}{h^2}, \frac{g_{n,1}}{h^2} + f_{n-1,1} + \frac{g_{n-1,0}}{h^2} \right). \quad (3.43)$$

Dans la cas  $j = m-1$  il y a le terme  $u_{i,m} = g_{i,m}$  qui est situé sur le bord, d'où

$$b_{m-1} = \left( \frac{g_{0,m-1}}{h^2} + f_{1,m-1} + \frac{g_{1,m}}{h^2}, f_{2,m-1} + \frac{g_{2,m}}{h^2}, \dots, f_{n-2,m-1} + \frac{g_{n-2,m}}{h^2}, \frac{g_{n,m-1}}{h^2} + f_{n-1,m-1} + \frac{g_{n-1,m}}{h^2} \right). \quad (3.44)$$

On obtient finalement la forme matricielle suivante

$$P_h U^h = B_h, \quad (3.45)$$

où  $P_h$  est donnée par (3.40) et  $B_h$  par (3.41) – (3.44).

Montrons maintenant que la matrice  $P_h$  est inversible. On a le résultat suivant dont la preuve est laissée en exercice.

**Lemme 3.3** La matrice  $T$  est définie positive, plus précisément on a

$$\forall x \in \mathbb{R}^{n-1}, \quad x^t T x \geq 2\|x\|_2^2.$$

**Proposition 3.3** La matrice  $P_h$  est inversible. Le système (3.45) admet donc une solution unique  $U^h$  pour tout  $h > 0$ .

**Preuve** On montre que si  $P_h U = 0$  alors  $U = 0$ . Écrivons l'égalité par bloc

$$\begin{pmatrix} T & -I & \boxed{0} & \cdots & \boxed{0} \\ -I & T & \ddots & \ddots & \vdots \\ \boxed{0} & \ddots & \ddots & \ddots & \boxed{0} \\ \vdots & \ddots & \ddots & \ddots & -I \\ \boxed{0} & \cdots & \boxed{0} & -I & T \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_{m-1} \end{pmatrix} = \begin{pmatrix} \boxed{0} \\ \boxed{0} \\ \vdots \\ \vdots \\ \boxed{0} \end{pmatrix}$$

d'où le système

$$\begin{aligned} T u_1 &= u_2 \\ T u_2 &= u_1 + u_3 \\ &\vdots \\ T u_{m-1} &= u_{m-2}. \end{aligned}$$

En multipliant chaque équation du système précédent par  $u_i$  et en sommant on obtient

$$u_1 T u_1 + u_2 T u_2 + \dots + u_{m-1} T u_{m-1} = 2(u_1 \cdot u_2) + \dots + 2(u_{m-2} \cdot u_{m-1}),$$

où la notation  $\cdot$  désigne le produit scalaire de deux vecteurs. D'autre part le Lemme 3.3 donne

$$2(u_1 \cdot u_2) + \dots + 2(u_{m-2} \cdot u_{m-1}) \geq 2\|u_1\|_2^2 + \dots + 2\|u_{m-1}\|_2^2. \quad (3.46)$$

En appliquant l'inégalité de Cauchy-Schwarz à chaque terme du coté gauche de (3.46) on arrive à

$$\|u_1\|_2^2 + (\|u_1\|_2 - \|u_2\|_2)^2 + \dots + (\|u_{m-1}\|_2 - \|u_{m-2}\|_2)^2 + \|u_{m-1}\|_2^2 \leq 0$$

d'où l'on déduit que  $u_1 = u_2 = \dots = u_{m-1} = 0$  et donc  $U = 0$ .  $\square$

Montrons maintenant la consistance du schéma (3.37). Notons par  $\tau_{ij}(h)$  l'erreur de consistance au noeud  $(x_i, y_j)$  définie par

$$\tau_{ij}(h) = \frac{4u(x_i, y_j) - u(x_{i+1}, y_j) - u(x_{i-1}, y_j) - u(x_i, y_{j+1}) - u(x_i, y_{j-1})}{h^2} - f_{ij}. \quad (3.47)$$

**Proposition 3.4** Supposons que la solution  $u$  du problème (3.36) vérifie  $u \in C^4([a, b] \times [c, d])$ . Alors

$$\max_{(i,j) \in \Delta_h} |\tau_{ij}(h)| = o(h^2) \quad (3.48)$$

c'est à dire que le schéma est consistant d'ordre 2.

**Preuve** On a les développements limités suivants

$$\begin{aligned} u(x_{i+1}, y_j) &= u(x_i, y_j) + hu_x(x_i, y_j) + \frac{h^2}{2}u_{xx}(x_i, y_j) + \frac{h^3}{6}u_{xxx}(x_i, y_j) + o(h^4), \\ u(x_{i-1}, y_j) &= u(x_i, y_j) - hu_x(x_i, y_j) + \frac{h^2}{2}u_{xx}(x_i, y_j) - \frac{h^3}{6}u_{xxx}(x_i, y_j) + o(h^4), \\ u(x_i, y_{j+1}) &= u(x_i, y_j) + hu_y(x_i, y_j) + \frac{h^2}{2}u_{yy}(x_i, y_j) + \frac{h^3}{6}u_{yyy}(x_i, y_j) + o(h^4), \\ u(x_i, y_{j-1}) &= u(x_i, y_j) - hu_y(x_i, y_j) + \frac{h^2}{2}u_{yy}(x_i, y_j) - \frac{h^3}{6}u_{yyy}(x_i, y_j) + o(h^4), \end{aligned}$$

en reportant ces égalités dans (3.47) et puisque  $-\Delta u(x_i, y_j) = f_{ij}$  on obtient

$$\tau_{ij}(h) = o(h^2).$$

□

**Remarque** La relation (3.48) signifie qu'il existe une constante  $C > 0$  indépendante de  $h$  telle que

$$\max_{(i,j) \in \Delta_h} |\tau_{ij}(h)| \leq Ch^2. \quad (3.49)$$

Montrons maintenant la convergence du schéma. Définissons d'abord l'erreur d'approximation  $e_{ij}$  au noeud  $(x_i, y_j)$  par

$$e_{ij} = u(x_i, y_j) - u_{ij},$$

en retranchant (3.37) de (3.47) on obtient la forme suivante de l'erreur de consistance

$$\tau_{ij}(h) = \frac{4e_{ij} - e_{i+1j} - e_{i-1j} - e_{ij+1} - e_{ij-1}}{h^2}. \quad (3.50)$$

Posons  $M = \sup_{(i,j) \in \Delta_h} |\tau_{ij}(h)|$ . On a alors le résultat de convergence suivant.

**Théorème 3.6** Supposons que la solution exacte  $u \in C^4([a, b] \times [c, d])$ .

Alors

$$\max_{(i,j) \in \Delta_h} |e_{ij}| \leq \frac{1}{2} C \max(a^2, b^2) h^2, \quad (3.51)$$

le schéma (3.37) est donc convergent d'ordre 2.

Avant de donner la preuve de ce théorème on aura besoin du principe du maximum discret suivant.

**Lemme 3.4** (*Principe du maximum discret*)

Soit  $U^h = (u_{00}, u_{10}, \dots, u_{n0}, u_{01}, u_{11}, \dots, u_{n1}, \dots, u_{0m}, \dots, u_{nm})$  un vecteur de  $\mathbb{R}^{(n+1) \times (m+1)}$  satisfaisant

$$4u_{ij} - u_{i+1j} - u_{i-1j} - u_{ij+1} - u_{ij-1} \leq 0, \quad \forall (i, j) \in \Delta_h \quad (3.52)$$

alors

$$\max_{(i,j) \in \Delta_h} u_{ij} \leq \max_{(i,j) \in \partial \Delta_h} u_{ij}$$

**Preuve** Définissons l'ensemble

$$S = \left\{ (i_0, j_0) \in \Delta_h : \max_{(i,j) \in \Delta_h} u_{ij} = u_{i_0 j_0} \right\}$$

et

$$i_1 = \min \{i_0 : \exists j_0 \text{ satisfaisant } (i_0, j_0) \in S\}.$$

A ce  $i_1$  correspond un  $j_1$  telle que  $(i_1, j_1) \in S$ . On a partir de (3.52)

$$4u_{i_1 j_1} \leq u_{i_1+1 j_1} + u_{i_1-1 j_1} + u_{i_1 j_1+1} + u_{i_1 j_1-1}. \quad (3.53)$$

Étudions les différents cas possibles :

1.  $2 \leq i_1 \leq n-2$  : comme  $u_{i_1-1 j_1} < u_{i_1 j_1}$  et  $u_{i_1+1 j_1} \leq u_{i_1 j_1}$  alors

$$2u_{i_1 j_1} < u_{i_1 j_1-1} + u_{i_1 j_1+1} \quad (3.54)$$

- Si  $2 \leq j_1 \leq m-2$  alors il s'ensuit de (2.54) et puisque  $u_{i_1 j_1-1} < u_{i_1 j_1}$  et  $u_{i_1 j_1+1} < u_{i_1 j_1}$  que  $u_{i_1 j_1} < u_{i_1 j_1}$ , ce qui est une contradiction. Ce cas est donc impossible.
- Si  $j_1 = 1$ , alors on a dans le premier cas

$$\begin{aligned} 2u_{i_1 j_1} &< u_{i_1 0} + u_{i_1 j_1+1} \\ &\leq \max_{(i,j) \in \partial \Delta_h} u_{ij} + u_{i_1 j_1} \end{aligned}$$

$$\text{d'où } u_{i_1 j_1} \leq \max_{(i,j) \in \partial \Delta_h} u_{ij}.$$

— Si  $j_1 = m - 1$ , dans cas (3.54) implique

$$\begin{aligned} 2u_{i_1 j_1} &< u_{i_1 j_1 - 1} + u_{i_1 m} \\ &\leq u_{i_1 j_1} + \max_{(i,j) \in \partial \Delta_h} u_{ij} \end{aligned}$$

d'où l'on tire  $u_{i_1 j_1} \leq \max_{(i,j) \in \partial \Delta_h} u_{ij}$ .

2.  $i_1 = 1$  : dans ce cas on a de (3.53)

$$4u_{i_1 j_1} \leq u_{i_1 + 1 j_1} + u_{0 j_1} + u_{i_1 j_1 + 1} + u_{i_1 j_1 - 1} \quad (3.55)$$

— Si  $j_1 = 1$  alors (3.55) entraîne

$$\begin{aligned} 4u_{i_1 j_1} &\leq u_{i_1 + 1 j_1} + u_{0 j_1} + u_{i_1 j_1 + 1} + u_{i_1 0} \\ &\leq 2u_{i_1 j_1} + 2 \max_{(i,j) \in \partial \Delta_h} u_{ij} \end{aligned}$$

d'où il s'ensuit que  $u_{i_1 j_1} \leq \max_{(i,j) \in \partial \Delta_h} u_{ij}$ .

— De même si  $j_1 = m - 1$ , alors (3.55) donne

$$\begin{aligned} 4u_{i_1 j_1} &\leq u_{i_1 + 1 j_1} + u_{0 j_1} + u_{i_1 m} + u_{i_1 j_1 - 1} \\ &\leq 2u_{i_1 j_1} + 2 \max_{(i,j) \in \partial \Delta_h} u_{ij} \end{aligned}$$

ce qui implique que  $u_{i_1 j_1} \leq \max_{(i,j) \in \partial \Delta_h} u_{ij}$ .

3.  $i_1 = n - 1$  : d'après (3.53)

$$4u_{i_1 j_1} \leq u_{n j_1} + u_{i_1 - 1 j_1} + u_{i_1 j_1 + 1} + u_{i_1 j_1 - 1}. \quad (3.56)$$

— Si  $2 \leq j_1 \leq m - 2$  : dans ce cas et puisque  $u_{i_1 - 1 j_1} \leq u_{i_1 j_1}$ ,  $u_{i_1 j_1 + 1} \leq u_{i_1 j_1}$  et  $u_{i_1 j_1 - 1} \leq u_{i_1 j_1}$ , (3.56) entraîne

$$u_{i_1 j_1} \leq \max_{(i,j) \in \partial \Delta_h} u_{ij}$$

— Si  $j_1 = 1$  : On a a partir de (3.56)

$$\begin{aligned} 4u_{i_1 j_1} &\leq u_{n j_1} + u_{i_1 - 1 j_1} + u_{i_1 j_1 + 1} + u_{i_1 0} \\ &\leq u_{i_1 - 1 j_1} + u_{i_1 j_1 + 1} + 2 \max_{(i,j) \in \partial \Delta_h} u_{ij} \end{aligned}$$

d'où l'on tire  $u_{i_1 j_1} \leq \max_{(i,j) \in \partial \Delta_h} u_{ij}$ .

— Si  $j_1 = m - 1$  : dans ce cas (3.56) donne

$$\begin{aligned} 4u_{i_1 j_1} &\leq u_{n j_1} + u_{i_1 - 1 j_1} + u_{i_1 m} + u_{i_1 j_1 - 1} \\ &\leq u_{i_1 - 1 j_1} + u_{i_1 j_1 - 1} + 2 \max_{(i,j) \in \partial \Delta_h} u_{ij} \end{aligned}$$

ce qui entraîne que  $u_{i_1 j_1} \leq \max_{(i,j) \in \partial \Delta_h} u_{ij}$ .

Dans tous les cas on a montré que  $u_{i_1j_1} \leq \max_{(i,j) \in \partial\Delta_h} u_{ij}$  ce qui implique que  $\max_{(i,j) \in \Delta_h} u_{ij} \leq \max_{(i,j) \in \partial\Delta_h} u_{ij}$ .  $\square$

On est maintenant en mesure de démontrer le Théorème 3.6.

**Preuve du Théorème 3.6** La preuve est similaire à celle du Théorème 3.4.

Posons  $e_{ij}^- = -e_{ij} + \frac{1}{2}Mx_i^2$  et  $e_{ij}^+ = e_{ij} + \frac{1}{2}Mx_i^2$ , où  $e_{ij}$  désigne l'erreur d'approximation au noeud  $(x_i, y_j)$ . On a alors de (3.50)

$$\begin{aligned} 4e_{ij}^- - e_{i+1j}^- - e_{i-1j}^- - e_{ij+1}^- - e_{ij-1}^- &= h^2\tau_{ij}(h) + \frac{1}{2}M(2x_i^2 - x_{i-1}^2 - x_{i+1}^2) \\ &= h^2(-\tau_{ij}(h) - M) \\ &\leq 0, \end{aligned}$$

pour tout  $(i, j) \in \Delta_h$ . De même

$$\begin{aligned} 4e_{ij}^+ - e_{i+1j}^+ - e_{i-1j}^+ - e_{ij+1}^+ - e_{ij-1}^+ &= h^2(\tau_{ij}(h) - M) \\ &\leq 0 \end{aligned}$$

pour tout  $(i, j) \in \Delta_h$ . D'après le Lemme 3.4

$$\begin{aligned} \max_{(i,j) \in \Delta_h} e_{ij}^+ &\leq \max_{(i,j) \in \partial\Delta_h} e_{ij}^+ \\ \max_{(i,j) \in \Delta_h} e_{ij}^- &\leq \max_{(i,j) \in \partial\Delta_h} e_{ij}^-. \end{aligned}$$

Or  $e_{ij} = g_{ij} - g_{ij} = 0$  pour tout  $(i, j) \in \partial\Delta_h$ , il s'ensuit alors de la définition de  $e_{ij}^-$  et  $e_{ij}^+$  que

$$\begin{aligned} \max_{(i,j) \in \partial\Delta_h} e_{ij}^+ &\leq \frac{1}{2}M \max_i(x_i^2) \\ &\leq \frac{1}{2}M \max(a^2, b^2) \\ \max_{(i,j) \in \partial\Delta_h} e_{ij}^- &\leq \frac{1}{2}M \max(a^2, b^2) \end{aligned} \quad (3.57)$$

Les inégalités (3.57) entraînent

$$\max_{(i,j) \in \Delta_h} (\pm e_{ij}) = \max_{(i,j) \in \Delta_h} |e_{ij}| \leq \frac{1}{2}M \max(a^2, b^2). \quad (3.58)$$

D'autre part d'après (3.49) il existe une constante  $C > 0$  indépendante de  $h$  telle que

$$M = \max_{(i,j) \in \Delta_h} |\tau_{ij}(h)| \leq Ch^2,$$

qui donne avec l'inégalité (3.58)

$$\max_{(i,j) \in \Delta_h} |e_{ij}| \leq \frac{1}{2}C \max(a^2, b^2)h^2.$$

$\square$

### Simulation numérique de l'équation de Poisson 2d dans un rectangle

Considérons le script Matlab suivant qui calcule une solution approchée de l'équation de Poisson (3.36) dans le rectangle  $[a, b] \times [c, d]$ .

```

% poisson2 est une fonction Matlab qui calcule la solution
% approchée  $z_{\{i,j\}}$  du problème de Dirichlet  $-\Delta u=f$ 
% sur un rectangle  $\Omega$  avec la donnée au bord  $u=g$ .
% les arguments d'entrées sont:
% les extrémités du rectangle  $a, b, c, d$ 
% le nombre de noeuds  $n$  sur l'axe des  $x$ 
% le nombre de noeuds  $m$  sur l'axe des  $y$ 
% On suppose dans ce programme que le pas  $h$  sur chaque
% axe est le même c'est à dire que  $(b-a)/n=(d-c)/m$ 
% Les autres arguments sont:
% la fonction  $f$ 
% la fonction sur le bord  $g$ 
% Le programme renvoie la solution approchée  $z$ 
function z=poisson2(a,b,c,d,n,m,f,g)
h=(b-a)/n;
x=linspace(a+h,b-h,n);
y=linspace(c+h,d-h,m);
% calcul de la matrice ah (tapez help gallery dans Matlab)
ah=gallery('poisson',n);
% calcul du second membre bh (voir relations
% (3.41)–(3.44) du cours)
for i=1:n;
    if i==1;
        bh(i)=(h^2)*f(x(i),y(i))+g(a,y(i))+g(x(i),c);
    elseif i==n;
        bh(i)=(h^2)*f(x(i),y(1))+g(b,y(1))+g(x(i),c);
    else
        bh(i)=(h^2)*f(x(i),y(1))+g(x(i),c);
    end
end
for k=2:m;
    if k<m;
        for j=1:n;
            if j==1;

```

```

        bh=[bh, (h^2)*f(x(j),y(k))+g(a,y(k))];
    elseif and(j>1,j<n);
        bh=[bh,(h^2)*f(x(j),y(k))];
    else j==n;
        bh=[bh,(h^2)*f(x(j),y(k))+g(a,y(k))];
    end
end
end
else
    k==m;
    for j=1:n;
        if j==1;
            bh=[bh, (h^2)*f(x(j),y(k))+g(a,y(k))+g(x(j),d)];
        elseif and(1<j,j<m);
            bh=[bh,(h^2)*f(x(j),y(k))+g(x(j),d)];
        else j==n;
            bh=[bh,(h^2)*f(x(j),y(k))+g(b,y(k))+g(x(j),d)];
        end
    end
end
end
end
% resolution du systeme discret ah*uh=bh
uh=ah\bh.';
% calcul de la cote z(i,j) correspondant au point (x_i,y_j)
for j=1:m;
    for i=1:n;
        z(i,j)=uh((j-1)*n+i);
    end
end
end
% graphe de la solution
surf(x,y,z)

```

**Exemple** Considérons le problème suivant [10]

$$\begin{cases} -\Delta u(x,y) = 8\pi^2 \sin(2\pi x) \cos(2\pi y), & (x,y) \in (0,1) \times (0,1) \\ u(x,y) = \begin{cases} 0, & \text{si } x = 0, \text{ ou } x = 1, \\ \sin 2\pi x, & \text{si } y = 0, \text{ ou } y = 1 \end{cases} \end{cases} \quad (3.59)$$

On peut montrer que la solution exacte est donnée par  
 $u(x,y) = \sin(2\pi x) \cos(2\pi y)$ .

Les figures 3.1 et 3.2, qui représentent les graphiques de la solution

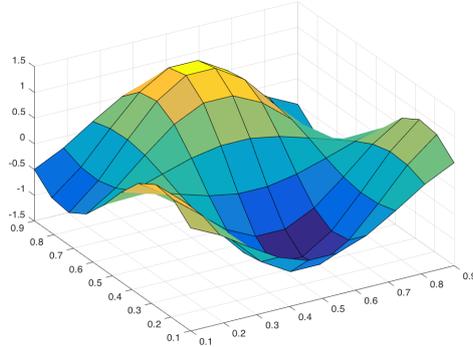


FIGURE 3.1 – Graphe de la solution du problème (3.59) dans le cas  $n = m = 10$

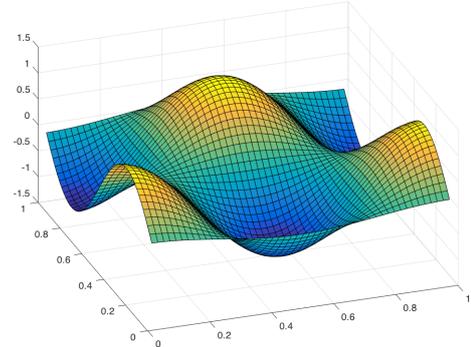


FIGURE 3.2 – Graphe de la solution du problème (3.59) dans le cas  $n = m = 50$

approchée dans les cas  $n = m = 10$  et  $n = m = 50$  respectivement, montrent que la solution s'affine lorsque  $n$  et  $m$  augmentent. En effet dans le cas  $n = m = 10$  l'erreur  $\max_{ij \in \Delta_h} e_{ij} = 0.2932$  alors que dans les cas  $n = m = 50$  on obtient l'erreur  $\max_{ij \in \Delta_h} e_{ij} = 0.0465$ .

## Exercices

1. Démontrer le Théorème 3.5.
2. (*Equation du transport non conservative*)  
Soient  $v \in C([0, 1], \mathbb{R}_+)$  et  $a_0, a_1 \in \mathbb{R}$ . On considère l'équation du transport suivante qui modélise l'écoulement de fluides compressibles polyphasiques :

$$\begin{aligned} -u''(x) + v(x)u'(x) &= 0, \quad x \in (0, 1), \\ u(0) &= a_0, \quad u(1) = a_1. \end{aligned} \quad (2)$$

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

$$\begin{aligned} \frac{1}{h^2}(2u_i - u_{i-1} - u_{i+1}) + \frac{1}{h}v_i(u_i - u_{i-1}) &= 0, \quad i = 1, \dots, n, \\ u_0 &= a_0, \quad u_{n+1} = a_1, \end{aligned} \quad (3)$$

où  $v_i = v(x_i)$  pour  $i = 1, \dots, n$ .

— Montrer que le système (3) s'écrit sous la forme  $MU = b$  avec

$U = (u_1, \dots, u_n)$ ,  $b \in \mathbb{R}^n$  et  $M$  est une matrice telle que :

- (a)  $MU \geq 0$  entraîne  $U \geq 0$  (les inégalités s'entendent composantes par composantes),
- (b)  $M$  est inversible,
- (c) (*Principe du maximum discret*) Si  $U$  est solution de  $MU = b$  alors  $\min(a_0, a_1) \leq u_i \leq \max(a_0, a_1)$ , où  $U = (u_1, \dots, u_n)$ .

**TP :** Écrire une fonction Matlab d'arguments  $n, a_0, a_1, v$  qui retourne la solution approchée  $U$  du problème (2). Calculer une solution approchée pour  $v(x) = 1 + x$ ,  $a_0 = 1$ ,  $a_1 = 2$ ,  $n = 100$ . Comparer graphiquement la solution exacte et la solution approchée trouvée.

## 3.4 Discrétisation de l'équation de la chaleur

Dans ce chapitre on s'intéresse à l'approximation par la méthode des différences finies de l'équation de la chaleur en dimension 1. Le problème modèle étant l'équation de la chaleur avec conditions aux bord de type Dirichlet, Neumann ou Robin.

### 3.4.1 Le problème modèle

Beaucoup de problèmes d'ingénierie sont modélisés par une équation parabolique du type

$$\frac{\partial u}{\partial t}(t, x) = \frac{\partial}{\partial x} \left( b(t, x) \frac{\partial u}{\partial x}(t, x) \right) + c(t, x)u(t, x) + d(t, x), \quad t \in (0, T),$$

$$x \in (a, b). \quad (3.60)$$

où  $u(t, x)$  désigne la solution. La condition initiale est donnée par

$$u(0, x) = u_0(x), \quad x \in (a, b), \quad (3.61)$$

où  $u_0$  est une fonction donnée. La solution  $u(t, x)$  est définie sur  $(0, T) \times (a, b)$ . On associe à l'équation (3.60) la condition au bord suivante

$$\begin{cases} \alpha_0(t)u(t, a) + \alpha_1(t)\frac{\partial u}{\partial t}(t, a) = \alpha_2(t), & t \in (0, T) \\ \beta_0(t)u(t, b) + \beta_1(t)\frac{\partial u}{\partial t}(t, b) = \beta_2(t), & t \in (0, T) \end{cases} \quad (3.62)$$