# Chapitre 2

# Différences finies pour les problèmes aux limites

## 2.1 Principe de la méthode

La méthode des différences finies est celle, parmi les méthodes d'approximation des problèmes aux limites, qui ressemble le plus aux schémas numériques utilisés pour approcher les solutions des équations différentielles ordinaires. L'idée de base est identique : il s'agit de remplacer les dérivées qui apparaissent dans l'équation par des quotients différentiels appropriés, d'où le nom de la méthode, *différences finies* par opposition à des différences "infinitésimales" qui correspondraient aux dérivées elles-mêmes. Dans cette méthode, ce que l'on calcule effectivement n'est pas une fonction définie sur l'intervalle sur lequel on travaille, mais des approximations des valeurs que prend la solution du problème aux limites en un nombre *fini* de points de cet intervalle (ce qui est nécessaire si on veut pouvoir implémenter les algorithmes correspondants). On est bien entendu libre ensuite d'interpoler les valeurs approchées ainsi obtenues pour construire une fonction (on peut construire par exemple une fonction continue linéaire par morceaux ou une fonction plus régulière en utilisant des splines cubiques) et dessiner un beau graphe par exemple, mais il ne s'agit plus de la méthode elle-même, tout au plus d'un "post-processing" du résultat de celle-ci.

On commence donc par introduire une *grille de discrétisation uniforme* en se donnant un entier $N \geq 1$ et en posant $h = \frac{1}{N+1}$ et $x_i = ih$ pour $i = 0, 1, \ldots, N+1$, de telle sorte que les points $x_i$ sont uniformément espacés entre eux du *pas* $h$, *i.e.*, $x_{i+1} - x_i = h$, avec $x_0 = 0$ et $x_{N+1} = 1$. Il y a donc $N$ points de la grille dans l'intérieur de l'intervalle, correspondant aux indices $i = 1, \ldots, N$, voir la figure 2.1 ci-dessous. Dans la suite, on fera tendre $N$ vers l'infini, ce qui est équivalent à faire tendre $h$ vers 0. On va calculer des valeurs numériques notées $u_i$, $i = 1, \ldots, N$, qui seront des approximations des valeurs exactes $u(x_i)$ d'autant meilleures que $N$ est grand ou encore $h$ est petit, ce que l'on démontrera ultérieurement, en supposant $u_0 = \alpha, u_{N+1} = \beta$ (dans la figure 2.1 on a illustré le cas $\alpha = \beta = 0$). L'idée heuristique est que, les dérivées étant par définition des limites de quotients différentiels, on ne devrait pas commettre une trop grande erreur en remplaçant celles-ci par de tels quotients, appelés traditionnellement *différences finies*.

Si $\varphi$ est une fonction assez régulière sur $[0, 1]$, on peut approcher la dérivée de $\varphi$ en $x_i$, en

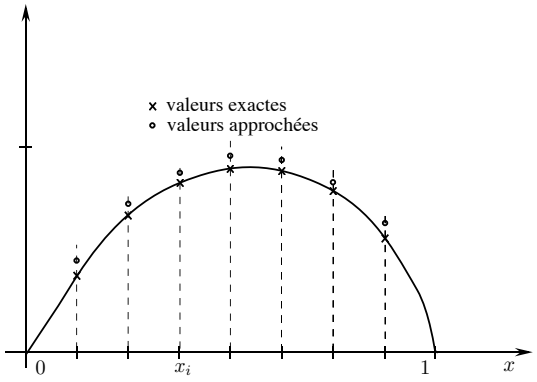



FIGURE 2.1 – Idée de la méthode

supposant $x_{i\pm1} - x_i = \pm h$ "assez petit", par la *différence finie* décentrée à droite

$$\varphi'(x_i) \approx \frac{\varphi(x_{i+1}) - \varphi(x_i)}{x_{i+1} - x_i} = \frac{\varphi(x_{i+1}) - \varphi(x_i)}{h}$$

ou bien par la différence finie décentrée à gauche

$$\varphi'(x_i) \approx \frac{\varphi(x_i) - \varphi(x_{i-1})}{x_i - x_{i-1}} = \frac{\varphi(x_i) - \varphi(x_{i-1})}{h}.$$

Combinant ces deux approximations entre elles, on voit apparaître pour la dérivée seconde

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

Naturellement, ici le signe $\approx$ n'a aucun sens précis. Il indique simplement une façon *a priori* raisonnable d'approcher la dérivée seconde d'une fonction en un point de la grille quand on connaît ses valeurs ponctuelles, ou des approximations de celles-ci, aux points voisins de cette même grille. On peut préciser les choses.

**Théorème 2.1.1** *Soit $\varphi \in \mathcal{C}^4([0,1])$. Pour tout $i \in \{1,\dots,N\}$, il existe un nombre $\theta_i$, avec $|\theta_i| < 1$ tel que*

$$-\varphi''(x_i) = \frac{-\varphi(x_{i+1}) + 2\varphi(x_i) - \varphi(x_{i-1})}{h^2} + \frac{h^2}{12}\varphi^{(4)}(x_i + \theta_i h).$$

*Démonstration.* Comme toujours pour ce type de résultats, la démonstration utilise la formule de Taylor-Lagrange. Comme $\varphi$ est supposée de classe $\mathcal{C}^4$ sur $[0,1]$, on peut utiliser cette dernière jusqu'à l'ordre 4 en tout point de la grille. En particulier, pour tout $i \in \{1,\dots,N\}$, il existe un nombre $\theta_i^+ \in ]0,1[$ tel que

$$\varphi(x_{i+1}) = \varphi(x_i) + h\varphi'(x_i) + \frac{h^2}{2}\varphi''(x_i) + \frac{h^3}{6}\varphi'''(x_i) + \frac{h^4}{24}\varphi^{(4)}(x_i + \theta_i^+ h).$$

De même, il existe $\theta_i^- \in ]0,1[$ tel que

$$\varphi(x_{i-1}) = \varphi(x_i) - h\varphi'(x_i) + \frac{h^2}{2}\varphi''(x_i) - \frac{h^3}{6}\varphi'''(x_i) + \frac{h^4}{24}\varphi^{(4)}(x_i - \theta_i^- h).$$

Additionnant ces deux relations entre elles, il vient

$$\varphi(x_{i+1})+\varphi(x_{i-1}) = 2\varphi(x_i)+h^2\varphi''(x_i)+\frac{h^4}{24}\big(\varphi^{(4)}(x_i+\theta_i^+h)+\varphi^{(4)}(x_i-\theta_i^-h)\big).$$

Comme $\varphi^{(4)}$ est continue par hypothèse, le théorème des valeurs intermédiaires nous dit qu'il existe $y_i \in [x_i-\theta_i^-h, x_i+\theta_i^+h]$ tel que

$$\frac{1}{2}\big(\varphi^{(4)}(x_i+\theta_i^+h)+\varphi^{(4)}(x_i-\theta_i^-h)\big) = \varphi^{(4)}(y_i).$$

En effet, le terme de gauche est la moyenne des valeurs prises par $\varphi^{(4)}$ aux extrémités de l'intervalle $[x_i-\theta_i^-h, x_i+\theta_i^+h]$. Par conséquent, on voit que

$$-\varphi''(x_i) = \frac{-\varphi(x_{i+1})+2\varphi(x_i)-\varphi(x_{i-1})}{h^2}+\frac{h^2}{12}\varphi^{(4)}(y_i).$$

Pour conclure, on remarque que $y_i \in [x_i-\theta_i^-h, x_i+\theta_i^+h] \subset ]x_{i-1}, x_{i+1}[$, donc $\theta_i = \frac{y_i-x_i}{h}$ est tel que $|\theta_i|<1$ et trivialement $y_i = x_i+\theta_i h$. □

**Corollaire 2.1.2** *Sous les hypothèses du théorème 2.1.1, on a*

$$\max_{1\le i\le N}\left|-\varphi''(x_i)-\frac{-\varphi(x_{i+1})+2\varphi(x_i)-\varphi(x_{i-1})}{h^2}\right| \le \frac{h^2}{12}\|\varphi^{(4)}\|_{L^\infty}. \tag{2.1}$$

*Démonstration.* Immédiat d'après le théorème 2.1.1. □

**Remarque 2.1.1** *i) La quantité au membre de gauche de (2.1) s'appelle erreur de consistance de la méthode (vecteur de $\mathbb{R}^N$ évalué en norme $\|.\|_\infty$), comme pour les schémas d'approximation des EDO.*

*ii) Si $\varphi$ est un polynôme de degré inférieur ou égal à 3, l'erreur de consistance est nulle.*

*iii) Si l'on suppose seulement $\varphi$ de classe $\mathcal{C}^3$, on peut uniquement dire que l'erreur de consistance est en $O(h)$ car alors la formule de Taylor-Lagrange n'est valable que jusqu'à l'ordre trois. De même, si $\varphi$ est seulement $\mathcal{C}^2$, alors l'erreur de consistance tend vers 0 quand $h$ tend vers 0, mais pas plus a priori.*

Appliquons ces résultats à la solution $u$ du problème aux limites (P). On note $\overline{U}_h$ le vecteur $(u(x_1),u(x_2),\dots,u(x_N))^T \in \mathbb{R}^N$ (attention à cette notation traditionnelle qui manque un peu de cohérence ; $h$ et $N$ sont liés par la relation $(N+1)h=1$ donc en particulier la dimension du vecteur $\overline{U}_h$ dépend de $h$).

**Corollaire 2.1.3** *Supposons que la solution $u$ du problème aux limites soit de classe $\mathcal{C}^4$ sur $[0,1]$. Il existe alors des points $y_i \in ]x_{i-1}, x_{i+1}[$ tels que le vecteur $\overline{U}_h$ est solution du système suivant :*

$$A_h\overline{U}_h = F_h + K_h, \tag{2.2}$$

*avec*

$$A_h = \frac{1}{h^2}\begin{pmatrix} 2+c(x_1)h^2 & -1 & 0 & \dots & \dots & \dots & 0 \\ -1 & 2+c(x_2)h^2 & -1 & 0 & & & \\ 0 & -1 & \ddots & \ddots & & & \\ \vdots & & \ddots & \ddots & \ddots & & \\ \vdots & & 0 & -1 & 2+c(x_i)h^2 & -1 & 0 \\ \vdots & & & & \ddots & \ddots & \ddots \\ \vdots & & & 0 & -1 & 2+c(x_{N-1})h^2 & -1 \\ 0 & & \dots & & 0 & -1 & 2+c(x_N)h^2 \end{pmatrix} \tag{2.3}$$

*où* $c_i = c(x_i)$,

$$F_h = \begin{pmatrix} f(x_1)+\frac{\alpha}{h^2} \\ f(x_2) \\ \vdots \\ f(x_{N-1}) \\ f(x_N)+\frac{\beta}{h^2} \end{pmatrix} \tag{2.4}$$

*où* $f_i = f(x_i)$ *et*

$$K_h = K_h(u) = -\frac{h^2}{12}\begin{pmatrix} u^{(4)}(y_1) \\ u^{(4)}(y_2) \\ \vdots \\ u^{(4)}(y_N) \end{pmatrix}. \tag{2.5}$$

*Démonstration.* C'est presque immédiat. En effet, en chaque point $x_i$, $1 \le i \le N$, on a par l'équation différentielle

$$-u''(x_i) + c(x_i)u(x_i) = f(x_i).$$

Il suffit de remplacer $-u''(x_i)$ par les expressions déduites du théorème 2.1.1. Il faut distinguer trois cas, suivant que $i = 1$, $2 \le i \le N-1$ ou $i = N$.

- Le cas $i = 1$. Dans ce cas, il vient

$$f(x_1) = -u''(x_1) + c(x_1)u(x_1) = \frac{-u(x_2)+2u(x_1)-u(x_0)}{h^2} + c(x_1)u(x_1) + \frac{h^2}{12}u^{(4)}(y_1).$$

Comme $u(x_0) = u(0) = \alpha$ est connu par la condition aux limites, on le passe au second membre et l'on obtient donc

$$\frac{-u(x_2)+2u(x_1)}{h^2} + c(x_1)u(x_1) = f(x_1) + \frac{u(x_0)}{h^2} - \frac{h^2}{12}u^{(4)}(y_1) = f(x_1) + \frac{\alpha}{h^2} - \frac{h^2}{12}u^{(4)}(y_1),$$

ou encore

$$\frac{1}{h^2}[(2+c(x_1)h^2)u(x_1) - u(x_2)] = f(x_1) + \frac{\alpha}{h^2} - \frac{h^2}{12}u^{(4)}(y_1).$$

- Le cas $i = N$. De façon analogue

$$f(x_N) = \frac{-u(x_{N+1})+2u(x_N)-u(x_{N-1})}{h^2} + c(x_N)u(x_N) + \frac{h^2}{12}u^{(4)}(y_N).$$

Comme $u(x_{N+1}) = u(1) = \beta$, on obtient donc

$$\frac{1}{h^2}[-u(x_{N-1}) + (2 + c(x_N)h^2)u(x_N)] = f(x_N) + \frac{\beta}{h^2} - \frac{h^2}{12}u^{(4)}(y_N).$$

- Le cas $2 \leq i \leq N-1$. Ici rien de spécial,

$$f(x_i) = \frac{-u(x_{i+1}) + 2u(x_i) - u(x_{i-1})}{h^2} + c(x_i)u(x_i) + \frac{h^2}{12}u^{(4)}(y_i),$$

est exactement la $i$ème ligne du système linéaire annoncé. □

**Remarque 2.1.2** *i) Les valeurs de la solution aux points de la grille satisfont exactement le système (2.2) : il n'y a aucune approximation ici. Bien sûr, il n'y a pas de miracle : on ne peut pas résoudre ce système pour calculer effectivement ces valeurs, puisque le second membre contient un terme $K_h = K_h(u)$ inconnu ! Il sera néanmoins utile dans l'analyse de la convergence de la méthode.*

*ii) Par contre, la matrice $A_h$ de dimension $N \times N$ et le vecteur $b_h$ sont connus.*

*iii) On note $\|\cdot\|_\infty$ la norme $\ell^\infty$ sur $\mathbb{R}^N$ (définie dans la proposition 1.5.3). La quantité $\|K_h\|_\infty$ est l'erreur de consistance calculée sur u. Elle satisfait donc, en appliquant le corollaire 2.1.2,*

$$\|K_h\|_\infty \leq \frac{h^2}{12}\|u^{(4)}\|_{L^\infty} \to 0 \text{ quand } N \to +\infty, \tag{2.6}$$

*puisque $h = \frac{1}{N+1}$. Notons ici que l'espace $\mathbb{R}^N$ change lorsqu'on varie N et si on voulait être précis, il conviendrait d'écrire $\|\cdot\|_{\infty,N}$ pour préciser qu'il s'agit de la norme $\ell^\infty$ sur l'espace $\mathbb{R}^N$, mais on ne le fera pas ici pour ne pas alourdir les notations.*

**Idée de la méthode des différences finies :** Puisque $K_h(u)$ est de toutes façons petit quand $h$ est petit (supposant $u$ suffisamment régulière, bien sûr), on décide de l'enlever du second membre du système linéaire et on considère alors le problème *discret* suivant

$$(S_h) \quad \begin{cases} \text{Trouver } U_h \in \mathbb{R}^N \text{ tel que} \\ \\ A_h U_h = F_h. \end{cases}$$

Il s'agit donc de résoudre un système de $N$ équations linéaires à $N$ inconnues qui sont les composantes $(u_1, \ldots, u_n)$ du vecteur $U_h$, de matrice $A_h$ et de second membre $F_h$ connus. Plusieurs questions se posent :

1) La matrice $A_h$ est elle inversible ? Si ce n'est pas le cas, on n'a aucune chance de calculer ce nouveau vecteur $U_h$.

2) En supposant que ce soit le cas, en général on aura $U_h \neq \overline{U}_h$, *i.e.*, $u_i \neq u(x_i)$. On commet ainsi une *erreur* qu'il faut pouvoir estimer. A-t-on alors $U_h - \overline{U}_h \to 0$ en un sens raisonnable et à quelle vitesse en fonction de $h \to 0$ (ou de $N \to +\infty$) ? En d'autres termes, a-t-on ainsi construit des approximations des valeurs de la solution aux points de la grille, et quelle est la qualité de ces approximations ?

**Définition 2.1.1** *On dit que la méthode est*
*i) convergente si*

$$\max_{1\le i\le N}|u_i-u(x_i)|\to 0 \text{ quand } N\to+\infty.$$

*ii) d'ordre p si*

$$\max_{1\le i\le N}|u_i-u(x_i)|\le C(u)h^p,$$

*où $C(u)$ est une constante qui ne dépend que de u.*

Donc, si la méthode est convergente, alors on a bien obtenu des approximations (ici uniformes, mais on pourrait utiliser d'autres normes que la norme $\|\cdot\|_\infty$) des valeurs exactes et ces approximations convergent d'autant plus vite que la méthode est d'ordre élevé.

En résumé, il s'agit maintenant d'effectuer ce que l'on appelle l'*analyse numérique* de la méthode : est-elle bien définie, est-elle convergente et de quel ordre ?

Commençons par traiter la première question, à savoir est-il bien raisonnable de vouloir calculer le vecteur $U_h$.

**Théorème 2.1.4** *Si $c(x)\ge 0$ sur $[0,1]$, la matrice $A_h$ est symétrique, définie positive, donc inversible.*

*Démonstration.* Il est évident que $A_h$ est tridiagonale symétrique, quel que soit le signe de $c$. Supposons maintenant que $c\ge 0$. Soit $V\in\mathbb{R}^N\setminus\{0\}$. Nous devons évaluer le produit scalaire $V^TA_hV=\langle A_hV,V\rangle$. On calcule

$$h^2V^TA_hV=h^2\sum_{i,j}a_{ij}v_iv_j=\sum_{i=1}^{N}(2+c(x_i)h^2)v_i^2-2\sum_{i=1}^{N-1}v_iv_{i+1}\ge 2\sum_{i=1}^{N}v_i^2-2\sum_{i=1}^{N-1}v_iv_{i+1},$$

puisque $c_i=c(x_i)\ge 0$. Par conséquent, en réarrangeant ces deux dernières sommes, il vient

$$\begin{aligned}h^2V^TA_hV &\ge v_1^2+(v_1^2-2v_1v_2+v_2^2)+(v_2^2-2v_2v_3+v_3^2)\\ &\qquad +\cdots+(v_{N-1}^2-2v_{N-1}v_N+v_N^2)+v_N^2\\ &=v_1^2+(v_1-v_2)^2+(v_2-v_3)^2+\cdots+(v_{N-1}-v_N)^2+v_N^2\ge 0.\end{aligned}$$

La matrice $A_h$ est donc positive. De plus, si $V^TA_hV=0$, on voit que nécessairement

$$v_1=0, v_1-v_2=0, v_2-v_3=0,\ldots,v_{N-1}-v_N=0 \text{ et } v_N=0,$$

c'est-à-dire en fait $V=0$. Elle est donc définie positive. On en déduit immédiatement qu'elle est inversible, car

$$V\in\ker A_h\Leftrightarrow A_hV=0\Rightarrow V^TA_hV=0\Leftrightarrow V=0,$$

donc $\ker A_h=\{0\}$. □

**Corollaire 2.1.5** *Si $c(x)\ge 0$ sur $[0,1]$, alors pour tout $f$ et pour tout N, il existe un unique vecteur $U_h\in\mathbb{R}^N$ solution du problème aux différences finies $(S_h)$.*

**Remarque 2.1.3** *Il est intéressant de noter que c'est la même hypothèse de signe sur c qui assure l'existence de la solution du problème discret et l'existence de la solution du problème aux limites lui-même.*

## 2.2 Étude de la convergence

Dans la suite, on suppose toujours que la solution $u$ est de classe $\mathcal{C}^4$ sur $[0,1]$. On va mesurer l'écart entre la solution discrète $U_h$ et les valeurs de la solution continue aux points de la grille $\overline{U}_h$ en utilisant la norme uniforme $\|U_h - \overline{U}_h\|_\infty$ que nous avons utilisé dans la définition 2.1.1 de la convergence. Toutes les normes sur $\mathbb{R}^N$ sont équivalentes mais elles ne sont pas toutes bien adaptées à l'étude de la convergence, puisque $N$ est amené à tendre vers l'infini. Notons que la norme $\|\overline{U}_h\|_\infty$ garde bien un sens alors que, par exemple, $\|\overline{U}_h\|_1 \to \infty$ quand $N \to \infty$. De plus il est possible d'évaluer la norme $\|A_h^{-1}\|_\infty$ comme nous allons le voir, et c'est elle qui va nous permettre d'estimer l'erreur $U_h - \overline{U}_h$.

**Proposition 2.2.1** *On a*

$$\|U_h - \overline{U}_h\|_\infty \leq \|A_h^{-1}\|_\infty \left(\frac{h^2}{12}\|u^{(4)}\|_{L^\infty}\right).$$

*Démonstration.* Écrivons les systèmes linéaires respectivement satisfaits par $U_h$ et $\overline{U}_h$ : par (2.2)

$$\begin{aligned} A_h U_h &= F_h, \\ A_h \overline{U}_h &= F_h + K_h. \end{aligned}$$

Soustrayant ces deux relations entre vecteurs, on obtient

$$A_h(U_h - \overline{U}_h) = -K_h \Longleftrightarrow U_h - \overline{U}_h = -A_h^{-1} K_h$$

puisque $A_h$ est inversible. Prenant les normes $\|\cdot\|_\infty$ de ces vecteurs, on obtient par définition des normes subordonnées

$$\|U_h - \overline{U}_h\|_\infty \leq \|A_h^{-1}\|_\infty \|K_h\|_\infty.$$

Le résultat se déduit alors immédiatement de (2.6). □

L'étude de la convergence se ramène donc maintenant à étudier le comportement de la quantité $\|A_h^{-1}\|_\infty$ (qui ne dépend plus de $u$) en fonction de $N$ ou $h$.

Notons que de façon semblable à ce que l'on fait dans l'étude des schémas d'approximation numérique pour les équations différentielles ordinaires, l'erreur est constituée de deux morceaux, d'une part l'erreur de consistance $\|K_h\|_\infty$, que l'on a déjà estimée, et d'autre part cette quantité $\|A_h^{-1}\|_\infty$ qui ne dépend pas de la solution et que l'on peut appeler constante de *stabilité*.

L'estimation de $\|A_h^{-1}\|_\infty$ est liée aux propriétés de la matrice $A_h$ qui est issue de la modélisation d'un problème physique. La démonstration demande un certain nombre d'étapes. On commence par définir la notion de matrice *monotone*.

**Définition 2.2.1** *i) On introduit une relation d'ordre partiel sur $\mathbb{R}^N$ en posant*

$$V \geq W \text{ si et seulement si } \forall i,\ v_i \geq w_i.$$

*ii) De même pour les matrices, on dit que*

$$A \geq B \text{ si et seulement si } \forall i, j,\ a_{ij} \geq b_{i,j}.$$

*iii) On dit qu'une matrice $A$ est monotone si elle est inversible et si $A^{-1} \geq 0$.*

Attention, la relation d'ordre sur les matrices ainsi définie n'a rien à voir avec celle définie sur les matrices symétriques à l'aide des formes quadratiques : il existe des matrices qui sont positives au sens des formes quadratiques mais pas positives au sens présent et inversement (en trouver des exemples).

Donnons une autre caractérisation de la monotonie d'une matrice, plus pratique que la définition.

**Lemme 1** *Soit $A$ une matrice $N \times N$. Elle est monotone si et seulement si quand on a un vecteur $V$ tel que $AV \geq 0$, alors cela implique que $V \geq 0$.*

*Démonstration.* On procède par condition nécessaire et condition suffisante.

• Condition nécessaire. Soit $A$ une matrice monotone. On se donne un vecteur $V \in \mathbb{R}^N$ tel que $AV \geq 0$, c'est-à-dire $(AV)_i \geq 0$ pour tout indice $1 \leq i \leq N$. Naturellement, $V = A^{-1}(AV)$, ce qui se lit en composantes sous la forme

$$v_i = \sum_{j=1}^{N} (A^{-1})_{ij} (AV)_j.$$

Or $(A^{-1})_{ij} \geq 0$ puisque $A$ est monotone, il vient donc $v_i \geq 0$, c'est-à-dire $V \geq 0$.

• Condition suffisante. Soit $A$ une matrice telle que $Av \geq 0$ implique $V \geq 0$. Montrons tout d'abord qu'elle est inversible. Pour cela, soit $W$ un élément du noyau de $A$, donc tel que $AW = 0$. Comme évidemment, $AW = 0 \geq 0$, on en déduit que $W \geq 0$. De même, $-W$ appartient au noyau et donc $-W \geq 0$. Par conséquent, $w_i = 0$ pour tout $i$ et le noyau est réduit au vecteur nul.

Notons $b_j$ le $j$ème vecteur-colonne de la matrice $A^{-1}$. Ceci signifie que $A^{-1}e_j = b_j$ où $e_j$ est le $j$ème vecteur de base. En d'autres termes, $Ab_j = e_j$, avec bien sûr $e_j \geq 0$. Par conséquent, on en déduit que $b_j \geq 0$ pour tout $j$, ce qui implique immédiatement que $A^{-1} \geq 0$. □

Continuons par une propriété qui nous sera utile dans une démonstration plus loin.

**Lemme 2** *Soit $A$ et $B$ deux matrices monotones, avec $B \geq A$. Alors*

$$A^{-1} \geq B^{-1}$$

*et*

$$\|A^{-1}\|_\infty \geq \|B^{-1}\|_\infty.$$

*Démonstration.* On remarque que pour tout couple de matrices $A$ et $B$ inversibles, on a l'identité :

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}.$$

En effet, $A^{-1} - B^{-1} = A^{-1}(BB^{-1}) - (A^{-1}A)B^{-1}$. Si les matrices sont monotones, et si de plus $B - A \geq 0$ ceci implique $A^{-1} - B^{-1} \geq 0$ car un produit de matrices positives est visiblement positif (il suffit d'écrire la définition). On a donc montré que

$$A^{-1} \geq B^{-1}$$

Considérons maintenant deux matrices telles que $B \geq A \geq 0$. Ceci signifie simplement que $b_{ij} \geq a_{ij} \geq 0$ pour tous $i, j$. Par conséquent,

$$\|B\|_\infty = \max_i \sum_j |b_{ij}| = \max_i \sum_j b_{ij} \geq \max_i \sum_j a_{ij} = \max_i \sum_j |a_{ij}| = \|A\|_\infty.$$

On applique alors ce résultat à $A^{-1} \geq B^{-1} \geq 0$. □

**Remarque 2.2.1** *Si A est une matrice positive, alors sa norme $\ell^\infty$ est donnée par*

$$\|A\|_\infty = \max_i \sum_j |a_{ij}| = \max_i \sum_j a_{ij} = \|Ae\|_\infty,$$

*où $e = (1,\dots,1)^T$ est le vecteur constant. De même, si A est monotone, puisque $A^{-1}$ est positive, alors sa norme $\ell^\infty$ est donnée par $\|A^{-1}e\|_\infty$, c'est à dire*

$$\|A^{-1}\|_\infty = \|V\|_\infty,$$

*où $V \in \mathbb{R}^N$ est l'unique solution de $AV = e$.*

Une classe importante de matrices monotones est donnée par la définiition suivante.

**Définition 2.2.2** *Une matrice A est appellée une M-matrice si et seulement si elle vérifie les trois propriétés suivantes :*
*i) Diagonale principale positive : $a_{i,i} > 0, \quad i = 1,\dots,N$.*
*ii) Autres coefficients négatifs ou nuls : $a_{i,j} \leq 0, \quad i \neq j$.*
*iii) Diagonale strictement dominante : il existe $\mu > 0$ tel que $\sum_j a_{i,j} \geq \mu, \quad i = 1,\dots,N$.*

**Proposition 2.2.2** *Une M-matrice A est monotone, et vérifie en outre $\|A^{-1}\|_\infty \leq \mu^{-1}$.*

*Démonstration.* On utilise la caractérisation donnée dans le lemme 1. Soit $V \in \mathbb{R}^N$ tel que $AV \geq 0$. Considérons $i^*$ tel que $v_{i^*} = \min v_i$. On peut donc écrire

$$0 \leq (AV)_{i^*} = \sum_{j=1}^N a_{i^*,j} v_j \leq (\sum_{j=1}^N a_{i^*,j}) v_i^*,$$

puisque par les propriétés i) et ii)

$$\sum_{j=1}^n a_{i^*,j}(v_j - v_{i^*}) \leq 0,$$

Par la propriété iii), il vient donc

$$\mu v_i^* \geq 0,$$

ce qui montre que $V \geq 0$. Pour le contrôle de la norme $\|A^{-1}\|_\infty$, on peut utiliser la remarque 2.2.1 : si $V$ est tel que $AV = e$, et si $i_0$ est tel que $v_{i_0} = \max v_i = \|V\|_\infty$, alors on a en particulier,

$$1 = (AV)_{i_0} = \sum_{j=1}^N a_{i_0,j} v_j \geq (\sum_{j=1}^N a_{i_0,j}) v_{i_0},$$

puisque par les propriétés i) et ii)

$$\sum_{j=1}^n a_{i_0,j}(v_j - v_{i_0}) \geq 0,$$

et par la propriété iii), il vient donc $v_{i_0} \leq \mu^{-1}$, soit

$$\|A^{-1}\|_\infty = \|V\|_\infty \leq \mu^{-1}.$$

□

Si on en revient à la matrice $A_h$, on constate qu'il s'agit "presque" d'une M-matrice, au sens où les propriétés de signe i) et ii) sont vérifiées mais la propriété de dominance diagonale iii) n'est pas stricte excepté pourr la premiere et la derniere ligne $i = 1$ et $i = N$. Pour les autres valeurs de $i$ on a

$$\sum_j (A_h)_{i,j} = \frac{1}{h^2}(-1-1+2+c(x_i)h^2) = c(x_i) \geq 0$$

puisque $c \geq 0$ mais on est pas assuré d'avoir $c(x_i) \geq \mu > 0$ pour tout $i$. On peut cependant prouver que $A_h$ est monotone grâce à la proposition suivante.

**Proposition 2.2.3** *Si $A$ est inversible et vérifie les propriétés i), ii), ainsi que*

$$\sum_j a_{i,j} \geq 0, \quad i = 1, \ldots, N,$$

*alors $A$ est monotone.*

*Démonstration.* On remarque que pour tout $\varepsilon > 0$ la matrice

$$A_\varepsilon = A + \varepsilon I$$

est une $M$-matrice en prenant $\mu = \varepsilon$. Par conséquence $A_\varepsilon$ est inversible et $A_\varepsilon^{-1} \geq 0$. Comme $A_\varepsilon \to A$ lorsque $\varepsilon \to 0$ et que $A$ est inversible, on est assuré que $A_\varepsilon^{-1} \to A^{-1}$. L'ensemble des matrice $M_N(\mathbb{R})$ étant un espace de dimension finie $N \times N$, cette convergence peut s'exprimer dans n'importe quelle norme et en particulier elle signifie que chaque coefficient $(A_\varepsilon^{-1})_{i,j} \geq 0$ de $A_\varepsilon$ tend vers le coefficient $(A^{-1})_{i,j}$ de $A^{-1}$ qui est donc positif. Ceci nous montre que $A$ est monotone. □

On peut ainsi appliquer ce résultat à la matrice $A_h$ et on a ainsi obtenu le résultat suivant.

**Proposition 2.2.4** *Si $c \geq 0$ alors, pour tout $h > 0$, la matrice $A_h$ est monotone.*

**Remarque 2.2.2** *La monotonie de la matrice $A_h$ est l'analogue discret de la monotonie du problème aux limites (cf. théorème 1.4.4). En effet, par définition (2.4) de $F_h$, si les hypothèses du théorème 1.4.4 sont satisfaites, alors $F_h \geq 0$. Si $F_h \geq 0$ et $A_h U_h = F_h$, on en déduit que $U_h \geq 0$. On parle donc de principe du maximum discret.*

On peut maintenant établir une estimation sur la norme $\ell^\infty$ de $A_h^{-1}$.

**Proposition 2.2.5** *Pour tout $h > 0$, l'inverse de la matrice $A_h$ vérifie l'estimation*

$$\|A_h^{-1}\|_\infty \leq \frac{1}{8}.$$

*Démonstration.* Introduisons la matrice

$$A_{0h} = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & & \\ 0 & -1 & \ddots & \ddots & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & -1 & 2 & -1 \\ 0 & \dots & & 0 & -1 & 2 \end{pmatrix} \tag{2.7}$$

qui correspond au cas où $c = 0$. On sait déjà que $A_h^{-1} \geq 0$ et que $A_{0h}^{-1} \geq 0$. De plus,

$$A_h - A_{0h} = \begin{pmatrix} c(x_1) & 0 & \dots & 0 \\ 0 & c(x_2) & \dots & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & c(x_N) \end{pmatrix} \geq 0,$$

car $c(x_i) \geq 0$. En appliquant maintenant le lemme 2, on obtient alors

$$A_{0h}^{-1} \geq A_h^{-1} \geq 0.$$

et aussi

$$\|A_h^{-1}\|_\infty \leq \|A_{0h}^{-1}\|_\infty,$$

et l'on s'est ramené à traiter le cas $c = 0$.

Il s'agit donc d'estimer la norme $\ell^\infty$ d'une matrice $A_{0h}^{-1}$ positive. Soit donc

$$U_{0h} = A_{0h}^{-1} e \Longleftrightarrow A_{0h} U_{0h} = e.$$

Le vecteur $U_{0h}$ n'est donc autre que la solution du problème aux différences finies associé au problème aux limites particulier

$$\begin{cases} -u_0''(x) = 1 \text{ dans } ]0,1[, \\ u_0(0) = u_0(1) = 0. \end{cases}$$

Or il est facile de calculer la solution exacte de ce problème : c'est la fonction $u_0(x) = \frac{1}{2}x(1-x)$. Il se trouve que c'est un polynôme du second degré et, par conséquent $u_0^{(4)} = 0$, son erreur de consistance est nulle, $K_h(u_0) = 0$. On déduit alors du corollaire 2.1.3 que le vecteur $\overline{U}_{0h}$ des valeurs $u_0(x_i)$ est solution *du même* système linéaire que le vecteur $U_{0h}$. D'où l'expression de celui-ci :

$$(U_{0h})_i = u_0(x_i) = \frac{1}{2} ih(1-ih).$$

On a ainsi

$$\|A_{0h}^{-1}\|_\infty = \|U_{0h}\|_\infty \leq \max_{x\in[0,1]} \left\{ \frac{1}{2}x(1-x) \right\} = \frac{1}{8},$$

ce qui nous donne le résultat annoncé. □

En combinant ceci avec l'estimation d'erreur dans la proposition 2.2.1, on obtient ainsi le résultat de convergence suivant.

**Théorème 2.2.1** *Supposons que c et f sont de classe $\mathcal{C}^2$. Alors on a la majoration d'erreur :*

$$\|U_h - \overline{U}_h\|_\infty = \max_{1\le i\le N} |u_i - u(x_i)| \le \frac{h^2}{96}\|u^{(4)}\|_{L^\infty}.$$

**Remarque 2.2.3** *i) On a montré que la méthode des différences finies est convergente et d'ordre 2 (pour la norme $\ell^\infty$). Les valeurs calculées se placent donc dans un voisinage du graphe de la solution exacte dont l'épaisseur est en $O(h^2)$.*

*ii) On peut montrer que la convergence n'est pas plus rapide en général, i.e., il existe une donnée $f \in \mathcal{C}^2$ telle que les solutions exactes et approchées correspondantes vérifient l'estimation $\|U_h - \overline{U}_h\|_\infty = Ch^2(1+\delta(h))$ avec $\delta(h) \to 0$ quand $h \to 0$ et $C > 0$.*

*iii) L'estimation d'erreur dépend de u, qui est inconnue, par l'intermédiaire de sa dérivée quatrième. Elle n'est donc pas fonction explicite des données du problème, et ne donne pas d'indication quantitative sur l'erreur commise. On dit qu'il s'agit d'une estimation a priori.*

On a obtenu une estimation en norme $\|.\|_\infty$. On peut obtenir un résultat dans une autre norme. Rappelons que les normes sur un espace de dimension finie sont équivalentes, mais les constantes qui interviennent dans les inégalités entre les normes peuvent dépendre de la dimension $N$ de l'espace. Par exemple, la norme euclidienne de $\mathbb{R}^N$ n'est pas intéressante telle quelle car quand $h \to 0$, alors $N \to \infty$ et le nombre de termes de la somme augmente ce qui fait tendre typiquement tendre cette norme vers $+\infty$ par exemple si on l'applique à $\overline{U}_h$. Pour obtenir un résultat de convergence qu'on interprète en terme de norme $L^2$, nous allons d'abord définir une *norme $L^2$ discrète* qui elle a un sens quand $h = \frac{1}{N+1} \to 0$.

**Définition 2.2.3** *Pour tout $V \in \mathbb{R}^{N+2}$, $V = (v_0, ..., v_{N+1})^T$, on définit la norme $L^2$ discrète $\|.\|_{2,\Delta}$ par*

$$\|V\|_{2,\Delta}^2 = h\Big(\frac{1}{2}v_0^2 + \sum_{i=1}^{N} v_i^2 + \frac{1}{2}v_{N+1}^2\Big)$$

On vérife facilement que c'est bien une norme sur $\mathbb{R}^{N+2}$. C'est la norme dans $L^2(0,1)$ de la fonction constante par morceaux $v_\Delta$ égale à $v_i$ sur l'intervalle $](i-1/2)h, i+1/2)h]$, $i = 1,..N$, et $v_0$ sur $[0, h/2]$, $v_N$ sur $]1-h/2, 1]$. La notation avec le symbole $\Delta$ traduit simplement le côté *discret*, on aurait pu noter avec l'indice $h$. Une autre interpretation est que $\|V\|_{2,\Delta}^2$ est l'intégrale de la fonction affine par morceaux qui vaut $v_i^2$ aux points $x_i$. En d'autre terme si les $v_i$ sont les valeurs d'une fonction $\varphi$ aux points $x_i$ alors $\|V\|_{2,\Delta}^2$ est l'approximation de $\int_0^1 |\varphi|^2$ par la formule des trapezes.

Une remarque immédiate est que l'on a

$$\|V\|_{2,\Delta}^2 \le h(N+1) \max_{i=0,...,N+1} |v_i|^2 = \|V\|_\infty^2,$$

autrement dit

$$\|V\|_{2,\Delta} \le \|V\|_\infty,$$

ce qui est consistant avec la propriété $\|\varphi\|_{L^2(0,1)} \le \|\varphi\|_{L^\infty(0,1)}$ sur les fonctions définies sur $]0,1[$. Jusqu'ici, $\overline{U}_h$ et $U_h$ étaient des vecteurs de $\mathbb{R}^N$, on peut aussi les étendre en vecteurs de $\mathbb{R}^{N+2}$ en ajoutant les valeurs imposées en $x_0 = 0$ et $x_{N+1} = 1$, et en posant avec un abus de notation :

$$\overline{U}_h := (\alpha, u(x_1), .., u(x_N), \beta)^T \quad \text{et} \quad U_h := (\alpha, u_1, .., u_N, \beta)^T.$$

On déduit ainsi du théorème 2.2.1 un résultat de convergence en norme $L^2$ discrète.

**Théorème 2.2.2** *Supposons que c et f sont de classe $\mathcal{C}^2$, $c \geq 0$, et soit u la solution du problème aux limites (P), et $U_h \in \mathbb{R}^{N+2}$, la solution du problème discret $(S_h)$ étendue aux valeurs en $x_0$ et $x_{N+1}$. Alors on a la majoration d'erreur :*

$$\|U_h - \overline{U}_h\|_{2,\Delta} \leq \max_{1\leq i\leq N} |u_i - u(x_i)| \leq \frac{h^2}{96}\|u^{(4)}\|_{L^\infty}.$$

Le résultat du théorème 2.2.2 implique alors que $\bar{u}_\Delta$ (fonction constante par morceaux construite à partir des valeurs $u_i$) tend vers $u$ dans $L^2(0,1)$. En effet, si on note $u_\Delta$ la fonction constante par morceaux construite à partir des valeurs $u(x_i)$, on peut écrire

$$\|u - \bar{u}_\Delta\|_{L^2} \leq \|u - u_\Delta\|_{L^2} + \|u_\Delta - \bar{u}_\Delta\|_{L^2}.$$

le dernier terme du membre de droite est exactement $\|\overline{U}_h - U_h\|_{2,\Delta}$ et il tend vers 0 avec $h$ d'après le résultat précédent, le premier tend également vers 0 (on approche une fonction continue par une fonction en escalier sur $[0,1]$). Bien entendu on peut généraliser tout ceci aux norme de type $L^p$ discrètes qu'on définit par

$$\|V\|_{p,\Delta}^p = h(\frac{1}{2}v_0^p + \sum_{i=1}^{N} v_i^p + \frac{1}{2}v_{N+1}^p)$$

## 2.3 Une excursion en dimension 2

À ce niveau du cours, on ne peut pas dire grand-chose des problèmes aux limites en dimension supérieure à 1. Néanmoins, un certain nombre de propriétés peuvent être démontrées de façon élémentaire dans des cas particuliers.

On considérera donc ici le carré ouvert du plan $\Omega = ]0,1[\times]0,1[$. On note les coordonnées $x$ et $y$. Soient $f\colon \Omega \to \mathbb{R}$ et $g\colon \partial\Omega \to \mathbb{R}$ deux fonctions continues. Le problème aux limites va consister à chercher une fonction $u\colon \bar{\Omega} \to \mathbb{R}$ appartenant à $\mathcal{C}^0(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$ telle que :

$$\begin{cases} -\Delta u = f \ \text{ dans } \Omega, \\ u = g \ \text{ sur } \partial\Omega. \end{cases} \tag{2.8}$$

**Théorème 2.3.1** *(Principe du maximum). Soit $u \in \mathcal{C}^0(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$ telle que $\Delta u \leq 0$. Alors u atteint son minimum sur le bord $\partial\Omega$. De même, si $\Delta u \geq 0$, alors u atteint son maximum sur le bord $\partial\Omega$.*

*Démonstration.* Soit $v \in \mathcal{C}^0(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$ une fonction qui atteint un minimum relatif en un point $(x_0,y_0) \in \Omega$ *intérieur*. L'application $v_{y_0}\colon ]0,1[\to \mathbb{R}, t \mapsto v(t,y_0)$ admet donc en particulier un minimum relatif en $t = x_0$, qui est point intérieur d'un intervalle où cette fonction est de classe $\mathcal{C}^2$. Par la formule de Taylor-Lagrange, on en déduit que $\frac{d^2 v_{y_0}}{dt^2}(x_0) \geq 0$ (raisonner par l'absurde). Par définition des dérivées partielles, ceci n'est autre que $\frac{\partial^2 v}{\partial x^2}(x_0,y_0) \geq 0$. De même, on montre que $\frac{\partial^2 v}{\partial y^2}(x_0,y_0) \geq 0$. Additionnant ces deux inégalités, on obtient

$$\Delta v(x_0,y_0) \geq 0$$

en tout point de minimum relatif intérieur de $v$.

Soit maintenant $u \in \mathcal{C}^0(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$ telle que $-\Delta u \geq 0$ dans $\Omega$. On raisonne par l'absurde et l'on suppose que $u$ n'atteint pas son minimum sur $\partial\Omega$. Elle atteint donc son minimum en un point intérieur $(x_0, y_0) \in \Omega$, et on a

$$M = u(x_0, y_0) < N = \min_{(x,y)\in\partial\Omega} u(x,y)$$

On introduit alors une fonction auxiliaire

$$u_\varepsilon(x,y) = u(x,y) - \varepsilon(x^2+y^2) \quad \text{avec} \quad 0 < \varepsilon < \frac{(N-M)}{2}.$$

On voit ainsi que

$$u_\varepsilon(x_0, y_0) \leq M,$$

et que pour tout $(x,y) \in \partial\Omega$ on a

$$u_\varepsilon(x,y) \geq N - 2\varepsilon > N - (N-M) = M.$$

Ceci nous montre que $u_\varepsilon$ atteint son minimum en un point $(x_1, y_1) \in \Omega$ intérieur, et par conséquent $\Delta u_\varepsilon(x_1, y_1) \geq 0$

D'un autre côté, comme $-\Delta u \geq 0$ dans $\Omega$,

$$-\Delta u_\varepsilon(x,y) = -\Delta u(x,y) + \varepsilon\Delta(x^2+y^2) = -\Delta u(x,y) + 4\varepsilon > 0.$$

En particulier, en $(x_1, y_1) \in \Omega$, on obtient

$$-\Delta u_\varepsilon(x_1, y_1) > 0,$$

qui est une contradiction. On montre de la même manière que si $\Delta u \geq 0$, alors $u$ atteint son maximum sur le bord $\partial\Omega$. □

**Remarque 2.3.1** *On ne peut pas faire le raisonnement par l'absurde directement sur u. En effet, celui-ci conduit alors à* $-\Delta u(x_0, y_0) \leq 0$ *et* $-\Delta u(x_0, y_0) \geq 0$*, ce qui n'est pas une contradiction mais implique seulement que* $-\Delta u(x_0, y_0) = 0$*, et on ne va pas plus loin.*

On déduit immédiatement du principe du maximum les propriétés suivantes.

**Corollaire 2.3.2** *Soit* $u \in \mathcal{C}^0(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$ *telle que* $-\Delta u \geq 0$ *dans* $\Omega$ *et* $u \geq 0$ *sur* $\partial\Omega$*. Alors* $u \geq 0$ *dans tout* $\bar{\Omega}$*. De même si* $-\Delta u \leq 0$ *dans* $\Omega$ *et* $u \leq 0$ *sur* $\partial\Omega$*, alors* $u \leq 0$ *dans tout* $\bar{\Omega}$*. En particulier, si* $u \in \mathcal{C}^0(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$ *est telle que* $-\Delta u = 0$ *dans* $\Omega$ *et* $u = 0$ *sur* $\partial\Omega$*, alors* $u = 0$ *dans* $\bar{\Omega}$*.*

On déduit immédiatement du corollaire un résultat d'*unicité*.

**Théorème 2.3.3** *Le problème aux limites (2.8) admet au plus une solution.*

*Démonstration.* Soient $u_1$ et $u_2$ deux solutions du problème aux limites (2.8). On pose $v = u_1 - u_2$. Il vient $\Delta v = \Delta u_1 - \Delta u_2 = f - f = 0$ dans $\Omega$ d'une part, et $v = g - g = 0$ sur $\partial\Omega$ d'autre part. D'après le corollaire 2.3.2, on a donc $v = 0$. □

**Remarque 2.3.2** *Les raisonnements précédents sont encore valables quand $\Omega$ est un ouvert borné quelconque de $\mathbb{R}^n$, pour toute valeur de n, et pas seulement pour un carré en dimension 2 (le vérifier).*

La question de l'*existence* d'une solution au problème (2.8) est nettement plus délicate. Grace à la géométrie simple du domaine, on peut ici l'apréhender à l'aide des séries de Fourier. On remarque tout d'abord qu'on peut prolonger toute fonction $\varphi$ définie sur $[0,1]$ à l'intervalle $[-1,1]$ par imparité, puis à $\mathbb{R}$ tout entier par 2-périodicité. La fonction $\tilde{\varphi}$ ainsi construite peut être développée en série de Fourier de la forme

$$\varphi(x) = \sum_{k=1}^{\infty} b_k \sin(k\pi x),$$

En effet, les coefficients $a_k$ des fonctions $\cos(k\pi x)$ qu'on trouve dans la forme générale des séries de Fourier sont nuls du fait de l'imparité de $\varphi$.

Il faut faire attention ici au sens de la convergence de cette série. On note que toutes les fonctions $\sin(k\pi x)$ s'annulent en 0 et 1 ce qui rend impossible la convergence uniforme si $\varphi$ ne s'annule pas elle même en ces points. En utilisant le théorème de Dirichlet, on peut montrer (exercice) la convergence ponctuelle sur $]0,1[$ si $\varphi$ est de classe $\mathcal{C}^1$ sur $]0,1[$. En étudiant plus finement les coeficients $b_k$, on peut aussi montrer la convergence normale donc uniforme si $\varphi$ est de classe $\mathcal{C}^1$ et s'annule en 0 et 1.

Pour une fonction $\varphi$ plus générale, la convergence a lieu dans $L^2(0,1)$ dès lors que $\varphi$ appartient à cet espace, et les coefficients $(b_k)_{k\geq 1}$ appartiennent à l'espace $\ell^2$ des suites de carré sommable. Ceci qui traduit le fait que la famille des fonctions $x \mapsto \sin(k\pi x)$ constitue une base orthogonale de $L^2(0,1)$. Nous renvoyons le lecteur au chapitre 6 pour plus de détails sur ces concepts fondamentaux.

De la même manière, en dimension 2 on peut décomposer une fonction $\varphi \in L^2(\Omega)$ avec $\Omega = ]0,1[^2$ suivant

$$\varphi(x,y) = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} b_{k,l} \sin(k\pi x)\sin(l\pi y),$$

où la série converge en norme $L^2$. On peut aussi démontrer la convergence ponctuelle sur $]0,1[^2$ sous des hypothèse de régularité sur $\varphi$, par exemple en supposant $\varphi \in \mathcal{C}^1(\bar{\Omega})$. Les fonctions

$$s_{k,l}(x,y) = \sin(k\pi x)\sin(l\pi y),$$

forment une base orthogonale, s'annulent au bord de $\Omega$ et sont des fonctions propres de l'opérateur $-\Delta$ puisqu'elles vérifient

$$-\Delta s_{k,l} = \pi^2(k^2+l^2)s_{k,l}$$

En décomposant le second membre de l'équation (2.8) suivant

$$f = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} f_{k,l} s_{k,l},$$

on voit qu'il est naturel de proposer une solution sous la forme

$$u = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} u_{k,l} s_{k,l}, \tag{2.9}$$

où les coefficients sont donnés par

$$u_{k,l} = \frac{1}{\pi^2(k^2+l^2)} f_{k,l}.$$

Puisque formellement, on voit que $-\Delta u = f$ en dérivant chaque terme et que $u = 0$ sur $\partial\Omega$ si la série converge uniformément. Afin d'énoncer un théorème, il est nécessaire de faire des hypothèses permettant de justifier ce résultat. On obtient par exemple aisément le résultat suivant.

**Proposition 2.3.1** *On suppose que $f \in \mathcal{C}^0(\bar{\Omega})$ est telle que les $(f_{k,l})_{k,l\geq 1}$ forment une suite de $\ell^1$, c'est à dire*

$$\sum_{k,l\geq 1} |f_{k,l}| < \infty.$$

*Alors la fonction u donnée par* (2.9) *appartient à $\mathcal{C}^2(\bar{\Omega})$ et est solution de l'équation* (2.8).

**Remarque 2.3.3** *Bien entendu, des hypothèses de régularité sur $f$ sont cachées dans l'hypothèse de sommabilité des coefficients $f_{k,l}$ puisque la continuité de $f$ ne suffit pas à l'assurer. En particulier il ne suffit pas que $f$ appartienne à $\mathcal{C}^0(\bar{\Omega})$ pour que la série de Fourier considérée converge normalement. On voit aussi que l'hypothèse de cette convergence implique que $f$ s'annule au bord, chose qu'on a pas nécessairement envie d'imposer (contrairement à la solution u). On ne peut pas faire l'économie de la démonstration de la convergence de la série et de celle de ses dérivées. En effet, c'est par des séries de fonctions aussi régulières que l'on veut que l'on construit des exemples de fonctions continues mais dérivables nulle part. Donc la dérivabilité de la somme d'une série de fonctions n'a rien d'évident, pas plus que la dérivation terme à terme de cette série.*

## 2.4 La méthode des différences finies en dimension 2

Donnons un aperçu de ce que l'on peut faire en différences finies sur le problème précédent. On se donne un entier $N \geq 1$, on pose encore $h = \frac{1}{N+1}$ et l'on construit la grille de discrétisation composée des points ou *nœuds* $z_{ij} = (ih, jh)$ pour $0 \leq i \leq N+1$ et $0 \leq j \leq N+1$. Il y a donc au total $(N+2)^2$ points dans cette grille, dont $N^2$ situés à l'intérieur ($1 \leq i \leq N$ et $1 \leq j \leq N$) et $4N+4$ situés sur le bord ($i = 0$ ou $i = N+1$ ou $j = 0$ ou $j = N+1$). Si $\varphi$ est une fonction définie sur $\bar{\Omega}$, on écrira indifféremment $\varphi(x,y) = \varphi(z)$ avec $z = (x,y)$.

De façon analogue à la dimension un, on introduit une discrétisation du Laplacien.

**Définition 2.4.1** *Soit $\varphi \in \mathcal{C}^0(\bar{\Omega})$. On appelle Laplacien discret à 5 points de $\varphi$ la quantité*

$$\Delta_h \varphi(z_{ij}) = \frac{1}{h^2}[-4\varphi(z_{ij}) + \varphi(z_{i-1,j}) + \varphi(z_{i+1,j}) + \varphi(z_{i,j-1}) + \varphi(z_{i,j+1})]$$

*définie en tout point $z_{ij}$ ($1 \leq i \leq N$ et $1 \leq j \leq N$) intérieur à la grille.*

**Remarque 2.4.1** *La terminologie est claire, le Laplacien discret à 5 points utilise les quatre plus proches voisins du point considéré sur la grille. Il existe des variantes à 9 points, etc. Noter la distinction entre $\Delta\varphi$ qui est une fonction définie sur $\Omega$ et $\Delta_h\varphi$ qui est défini seulement sur la grille.*

Estimons l'erreur de consistance.

**Théorème 2.4.1** *Soit $\varphi \in \mathcal{C}^4(\bar{\Omega})$. Alors pour tout $i,j \in \{1,\dots,N\}$,*

$$|\Delta\varphi(z_{ij}) - \Delta_h\varphi(z_{ij})| \leq \frac{h^2}{12}\Big(\|\frac{\partial^4\varphi}{\partial x^4}\|_{L^\infty} + \|\frac{\partial^4\varphi}{\partial y^4}\|_{L^\infty}\Big),$$

*où pour toute fonction v continue sur $\bar{\Omega}$, on a noté $\|v\|_{L^\infty} = \|v\|_{L^\infty(\bar{\Omega})} = \max_{x\in\bar{\Omega}}|v(x)|$.*

*Démonstration.* Par définition de ce qu'est une dérivée partielle, si l'on introduit la fonction $\theta_1^{ij}(t) = \varphi(ih+t, jh)$, qui est définie et de classe $\mathcal{C}^4$ sur un voisinage de zéro en $t$ (lequel contient au moins l'intervalle $]-h,h[$, vues les valeurs de $i$ et $j$), on a $\frac{\partial^k\varphi}{\partial x^k}(ih+t,jh) = \frac{d^k\theta_1^{ij}}{dt^k}(t)$ pour $0 \leq k \leq 4$. De même avec $\theta_2^{ij}(s) = \varphi(ih, jh+s)$, $\frac{\partial^k\varphi}{\partial y^k}(ih,jh+s) = \frac{d^k\theta_2^{ij}}{ds^k}(s)$. En particulier,

$$\Delta\varphi(z_{ij}) = (\theta_1^{ij})''(0) + (\theta_2^{ij})''(0).$$

Or, les mêmes calculs de développements de Taylor-Lagrange qu'en dimension 1 montrent que

$$h^2(\theta_1^{ij})''(0) = \theta_1^{ij}(-h) - 2\theta_1^{ij}(0) + \theta_1^{ij}(h) + \frac{h^4}{12}(\theta_1^{ij})^{(4)}(t_{ij})$$

pour un certain $t_{ij} \in ]-h,h[$ et

$$h^2(\theta_2^{ij})''(0) = \theta_2^{ij}(-h) - 2\theta_2^{ij}(0) + \theta_2^{ij}(h) + \frac{h^4}{12}(\theta^{ij})^{(4)}(s_{ij})$$

pour un certain $s_{ij} \in ]-h,h[$. Par définition des fonctions $\theta_1^{ij}$ et $\theta_2^{ij}$, il est facile de voir que $\theta_1^{ij}(0) = \varphi(z_{ij})$, $\theta_1^{ij}(-h) = \varphi(z_{i-1,j})$, $\theta_1^{ij}(h) = \varphi(z_{i+1,j})$, $\theta_2^{ij}(0) = \varphi(z_{ij})$, $\theta_2^{ij}(-h) = \varphi(z_{i,j-1})$ et $\theta_1^{ij}(h) = \varphi(z_{i,j+1})$. En remplaçant et en sommant, on obtient donc :

$$\Delta\varphi(z_{ij}) = \Delta_h\varphi(z_{ij}) + \frac{h^2}{12}\Big(\frac{\partial^4\varphi}{\partial x^4}(ih+t_{ij},jh) + \frac{\partial^4\varphi}{\partial y^4}(ih,jh+s_{ij})\Big),$$

et l'on conclut immédiatement à l'aide des valeurs absolues. □

On voit donc apparaître une erreur de consistance en $O(h^2)$. Par le même cheminement qu'en dimension 1, on introduit la méthode des différences finies de la façon suivante. Notons $\Omega_h = \{z_{ij}, 1 \leq i \leq N, 1 \leq j \leq N\}$ l'ensemble des points de grille intérieurs et $\bar{\Omega}_h = \{z_{ij}, 0 \leq i \leq N+1, 0 \leq j \leq N+1\}$ l'ensemble de tous les points de la grille, y compris ceux du bord. On note finalement $\partial\Omega_h := \bar{\Omega}_h \setminus \Omega_h$ l'ensemble des noeuds sur le bord.

On va donc chercher des valeurs $u_{i,j}$ qui approchent les valeurs exactes $u(x_{i,j})$ en résolvant les équations

$$\begin{cases} \frac{1}{h^2}(4u_{ij} - u_{i-1,j} - u_{i+1,j} - u_{i,j-1} - u_{i,j+1})) = f(z_{ij}), & z_{i,j} \in \Omega_h, \\ u_{ij} = g(z_{ij}), & z_{i,j} \in \partial\Omega_h. \end{cases} \tag{2.10}$$

Comme dans la méthode en dimension 1, on note $\overline{U}_h$ le vecteur de coordonnées $u(z_{i,j})$ et $U_h$ le vecteur de coordonnées $u_{i,j}$, pour $i,j = 1,\dots,N$ et on étudie l'erreur en les comparant.

On peut aussi identifier $U_h$ à une fonction $u_h$ *définie sur la grille* de sorte que les équations précèdentes peuvent aussi s'écrire sous la forme

$$\begin{cases} -\Delta_h u_h(z_{ij}) = f(z_{ij}) \text{ sur } \Omega_h, \\ u_h(z_{ij}) = g(z_{ij}) \text{ sur } \bar{\Omega}_h \setminus \Omega_h. \end{cases} \tag{2.11}$$

Il n'y a bien sûr aucune difficulté à appliquer le Laplacien discret, défini initialement pour une fonction continue sur $\bar{\Omega}$, à une fonction *discrète* définie seulement sur $\bar{\Omega}_h$. On vérifie aisément en regardant les indices que toutes les valeurs dont on a besoin pour calculer le Laplacien discret sont bien à notre disposition dans ce cas. On voit qu'en fait seules les valeurs de $U_h$ aux points intérieurs sont inconnues, ses valeurs sur les points du bord étant données par la condition aux limites (comme en dimension 1). Il s'agit donc d'un problème à $N^2$ inconnues (les valeurs de $u_h$ sur $\Omega_h$) et $N^2$ équations linéaires (les valeurs de $-\Delta_h u_h$ sur $\Omega_h$).

À la différence de la dimension 1, ces inconnues et équations ne sont pas arrangées naturellement comme les composantes d'un vecteur — en fait, elles sont arrangées naturellement comme les composantes d'une matrice avec deux indices — et la forme matricielle du problème aux différences finies n'apparaît pas directement à l'œil nu comme précédemment.

Pour faire apparaître cette forme matricielle, il faut réarranger abstraitement les valeurs de $U_h$, c'est-à-dire en fait les points de la grille, en une seule colonne. En d'autres termes, on doit *numéroter* les nœuds. En dimension 1, la question ne se posait pas puisqu'une numérotation naturelle s'imposait, donnée par les indices des points de la grille. Si l'on avait eu l'esprit compliqué, on aurait pu tout aussi bien choisir alors une autre numérotation, c'est-à-dire effectuer une permutation des indices, c'est-à-dire un changement de base qui permute les vecteurs de base. La matrice qui en aurait résulté aurait été obtenue à partir de la matrice initiale par changement de base par une matrice de permutation. Elle aurait en particulier perdu ses agréables propriétés d'être tridiagonale et symétrique (en passant, que peut-on dire de la monotonie ?), propriétés qui sont très utiles lorsqu'il est temps d'appliquer effectivement des méthodes de résolution de systèmes linéaires. Ici, on n'a pas le choix. Il n'y a pas de numérotation naturelle qui saute aux yeux. On va en parachuter une qui marche bien au sens où elle fournit une matrice symétrique et dont les éléments non nuls sont relativement concentrés autour de la diagonale, deux conditions favorables à l'utilisation, dans une étape ultérieure, d'algorithmes efficaces de résolution de systèmes linéaires.

On convient donc de numéroter les nœuds de gauche à droite et de bas en haut, comme sur la figure 2.2. On se convainc sans grand peine que le numéro du point $z_{ij}$ est donné par $(j-1)N+i$. On définit donc le vecteur $U_h$ par ses composantes :

$$u_k = (U_h)_k = u_h(z_{ij}) = u_{i,j} \quad \text{pour } k = (j-1)N+i.$$

*Attention à l'abus de notation :* on identifie la fonction $u_h$ définie sur la grille du vecteur $U_h$, bien que leur identification passe par une numérotation des nœuds largement arbitraire.

Les indices $k$ ainsi définis varient entre 1 et $N^2$, donc $U_h \in \mathbb{R}^{N^2}$. La bijection entre $\{1,\dots,N\} \times \{1,\dots,N\}$ et $\{1,\dots,N^2\}$ que l'on vient juste d'introduire admet pour inverse l'application $k \mapsto \left(k - \left[\frac{k-1}{N}\right]N, \left[\frac{k-1}{N}\right]+1\right)$ où $[t]$ désigne la partie entière de $t$ (le vérifier).

Nous pouvons maintenant écrire la forme matricielle de la méthode associée à la numérota-

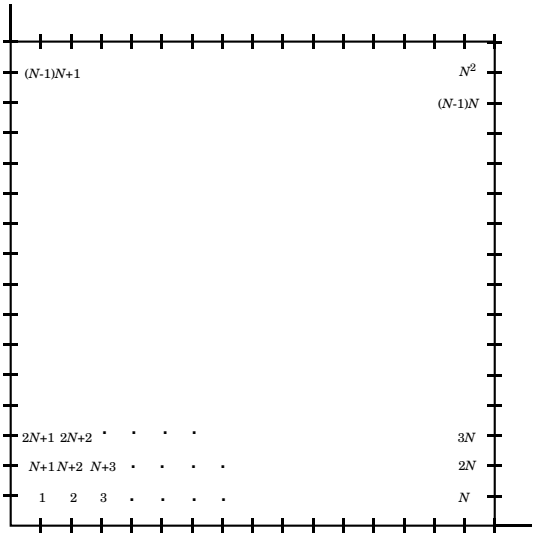



FIGURE 2.2 – Numérotation des noeuds

tion choisie. Pour cela, on introduit la matrice $N \times N$

$$T_4 = \begin{pmatrix} 4 & -1 & 0 & \dots & \dots & 0 \\ -1 & 4 & -1 & 0 & & \vdots \\ 0 & -1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & 0 & -1 & 4 & -1 \\ 0 & \dots & \dots & 0 & -1 & 4 \end{pmatrix}$$

et l'on note $I$ la matrice identité $N \times N$.

On va définir la matrice $A_h$ du système linéaire, c'est une matrice $N^2 \times N^2$ que l'on peut écrire *par blocs* $N \times N$, en utilisant les blocs $T_4$ et $I$, ainsi que le bloc 0 correspondant à la matrice $N \times N$ nulle. De même, on peut écrire le second membre $F_h$ par blocs, on fait apparaître ci-dessous les deux premiers blocs (sur $N$ blocs) et le début du troisième.

**Proposition 2.4.1** *Le vecteur $U_h \in \mathbb{R}^{N^2}$ est solution du système linéaire suivant :*

$$A_h U_h = F_h,$$

*avec*

$$A_h = \frac{1}{h^2} \begin{pmatrix} T_4 & -I & 0 & \dots & \dots & 0 \\ -I & T_4 & -I & 0 & & \vdots \\ 0 & -I & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & 0 & -I & T_4 & -I \\ 0 & \dots & \dots & 0 & -I & T_4 \end{pmatrix}$$

*et*

$$F_h = \begin{pmatrix} f_1 + \frac{1}{h^2}(g(z_{1,0}) + g(z_{0,1})) \\ f_2 + \frac{1}{h^2} g(z_{2,0}) \\ f_3 + \frac{1}{h^2} g(z_{3,0}) \\ \vdots \\ f_N + \frac{1}{h^2}(g(z_{N,0}) + g(z_{N+1,1})) \\ f_{N+1} + \frac{1}{h^2} g(z_{0,2}) \\ f_{N+2} \\ \vdots \\ f_{2N} + \frac{1}{h^2} g(z_{N+1,2}) \\ f_{2N+1} + \frac{1}{h^2} g(z_{0,3}) \\ f_{2N+2} \\ \vdots \end{pmatrix}$$

*où* $f_k = f(z_{ij})$ *pour* $k = (j-1)N + i$.

*Démonstration.* Il faut distinguer plusieurs cas, suivant que l'indice de ligne $k$ considéré correspond à un point dont les quatre voisins sont dans $\Omega_h$, un point dont seuls trois voisins sont dans $\Omega_h$ ou un point dont seuls deux voisins sont dans $\Omega_h$.

♣ Quatre voisins dans $\Omega_h$. Ce sont les points $z_{ij}$ avec $2 \le i \le N-2$ et $2 \le j \le N-2$. Ils correspondent aux indices $k \in \cup_{l=1}^{N-2} \{lN+2, lN+3, \ldots, (l+1)N-1\}$. (Par exemple, pour $l = 1$, ceci donne $k = N+2, N+3, \ldots, 2N-1$. Ensuite on saute directement à $2N+2$ et ainsi de suite)

En un tel point $z_{ij}$, de numéro $k = (j-1)N + i$, les points voisins sont tous numérotés et ont pour numéro respectif

$$\begin{aligned} \#z_{i-1,j} &= (j-1)N + (i-1) = k-1, \\ \#z_{i+1,j} &= (j-1)N + (i+1) = k+1, \\ \#z_{i,j-1} &= ((j-1)-1)N + i = k-N, \\ \#z_{i,j+1} &= ((j+1)-1)N + i = k+N. \end{aligned}$$

On obtient donc pour le problème discret la ligne

$$\frac{1}{h^2}(-u_{k-N} - u_{k-1} + 4u_k - u_{k+1} - u_{k+N}) = f(z_{ij}) = f_k.$$

♠ Trois voisins dans $\Omega_h$. Ceci peut se produire de quatre façons : $j = 1, i = 2, 3, \ldots, N-1$ ; $j = N, i = 2, 3, \ldots, N-1$ ; $i = 1, j = 2, 3, \ldots, N-1$ ou $i = N, j = 2, 3, \ldots, N-1$. Ces quatre façons correspondent respectivement à $k = 2, 3, \ldots, N-1$ ; $k = (N-1)N+2, (N-1)N+3, \ldots, N^2 - 1$ ; $k = N+1, 2N+1, \ldots, (N-2)N+1$ et $k = 2N, 3N, \ldots, (N-1)N$. Regardons la première possibilité (les autres sont laissées en exercice). Le point $z_{i,j-1} = z_{i,0}$ est situé sur le bord. On doit donc passer la valeur de $U_h$ correspondante, qui est donnée par la condition aux limites, dans le membre de droite, ce qui donne

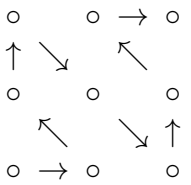
$$\frac{1}{h^2}(-u_{k-1} + 4u_k - u_{k+1} - u_{k+N}) = f_k + \frac{1}{h^2} g(z_{k,0}).$$

♢ Deux voisins dans $\Omega_h$. Ceci se produit aux quatre coins : $z_{1,1}$, $z_{N,1}$, $z_{1,N}$ et $z_{N,N}$, donc pour $k = 1$, $k = N$, $k = (N-1)N+1$ et $k = N^2$. Regardons le cas $k = 1$. Il vient

$$\frac{1}{h^2}(4u_1 - u_2 - u_{1+N}) = f_1 + \frac{1}{h^2}(g(z_{1,0}) + g(z_{0,1})).$$

Il faut enfin recoller tous les morceaux. C'est un peu fastidieux, il faut utiliser une grande feuille de papier, mais on arrive à la forme matricielle annoncée. □

**Remarque 2.4.2** *Pour se convaincre de l'importance du choix de la numérotation, il peut être amusant d'écrire entièrement la matrice pour $N = 3$ par exemple (donc une matrice $9 \times 9$) pour la numérotation proposée, puis pour une autre numérotation, a priori tout aussi naturelle, celle que l'on utilise pour montrer que $\mathbb{N}^2$ est en bijection avec $\mathbb{N}$ :*



En suivant exactement les mêmes procédés que pour la matrice $A_h$ en dimension 1, on obtient le résultat suivant.

**Proposition 2.4.2** *La matrice $A_h$ est symétrique, définie positive et monotone.*

Une conséquence immédiate est que la méthode des différences finies est encore bien posée. Il nous reste à contrôler la norme de son inverse. Pour cela on va utiliser une version discrète du principe du maximum établi dans le Théorème 2.3.1.

**Théorème 2.4.2** *(Principe du maximum discret). Soit $v_h$ une fonction définie sur la grille $\bar{\Omega}_h$ telle que $\Delta_h v_h \leq 0$ dans $\Omega_h$. Alors $v_h$ atteint son minimum sur $\partial\Omega_h$. De même, si $v_h$ est telle que $\Delta_h v_h \geq 0$ dans $\Omega_h$, alors $v_h$ atteint son maximum sur $\partial\Omega_h$*

*Démonstration.* La preuve est tout à fait similaire à celle du Théorème 2.3.1 alors on en indique seulement les grandes lignes. On commence à montrer que si une fonction $w_h$ définie sur la grille admet un minimum en un point $z_{i,j}$ intérieur, on a alors

$$\Delta_h w_h(z_{i,j}) \geq 0.$$

Ensuite, on raisonne par l'absude en sypposant que $v_h$ prend une valeur minimale en un point de grille $z_{i_0,j_0} \in \Omega_h$ intérieur, strictement inférieure à la valeur minimale sur le bord $\partial\Omega_h$. On introduit la fonction auxilliaire discrete

$$v_{h,\varepsilon}(z_{i,j}) = v_h(z_{i,j}) - \varepsilon(i^2 + j^2),$$

avec $\varepsilon > 0$ suffisament petit de façon que $v_{h,\varepsilon}$ atteint un minimum en un point de grille $z_{i_1,j_1} \in \Omega_h$ intérieur et on obtient une contradiction car $\Delta_h v_{h,\varepsilon} < 0$. □

**Théorème 2.4.3** *On a $\|A_h^{-1}\|_\infty \leq \frac{1}{2}$.*

*Démonstration.* Soit $F_h \in \mathbb{R}^{N^2}$ et $V_h \in \mathbb{R}^{N^2}$ non nul tels que $A_h V_h = F_h$. Identifiant $V_h$ avec une fonction $v_h$ définie sur $\bar{\Omega}_h$ et nulle au bord et $F_h$ avec une fonction $f_h$ définie sur $\Omega_h$, ceci est équivalent à $-\Delta_h v_h = f_h$ dans $\Omega_h$.

Introduisons la fonction discrète $w_h(z_{ij}) = \frac{h^2}{4}(i^2+j^2)$. On vérifie par un petit calcul que $-\Delta_h w_h = -1$ dans $\Omega_h$. Posons alors

$$w_h^+ = \|-\Delta_h v_h\|_\infty w_h - v_h,$$

où on laisse en évidence le signe pour se souvenir que $-\Delta_h v_h = f_h$. Prenant le Laplacien discret de $w_h^+$, il vient, pour tout $z_{i,j} \in \Omega_h$

$$-\Delta_h w_h^+(z_{i,j}) = -\|-\Delta_h v_h\|_\infty \Delta_h w_h(z_{i,j}) + \Delta_h v_h(z_{i,j}) = \Delta_h v_h(z_{i,j}) - \|-\Delta_h v_h\|_\infty \leq 0.$$

Le principe du maximum discret s'applique et montre que $w_h^+$ atteint son maximum sur $\partial\Omega_h$. Mais $v_h = 0$ sur $\partial\Omega_h$, donc pour tout $z_{i,j} \in \Omega_h$

$$w_h^+(z_{ij}) \leq \|-\Delta_h v_h\|_\infty \max_{\partial\Omega_h} w_h = \frac{1}{2}\|-\Delta_h v_h\|_\infty.$$

Par conséquent,

$$v_h(z_{i,j}) = \|-\Delta_h v_h\|_\infty w_h(z_{i,j}) - w_h^+(z_{i,j}) \geq -w_h^+(z_{i,j}) \geq -\frac{1}{2}\|-\Delta_h v_h\|_\infty.$$

Utilisant de la même façon la fonction $w_h^- = \|-\Delta_h v_h\|_\infty w_h + v_h$, on montre que

$$v_h(z_{i,j}) \leq \frac{1}{2}\|-\Delta_h v_h\|_\infty.$$

Par conséquent,

$$\|v_h\|_\infty = \max_{z_{i,j}\in\Omega_h} |v_h(z_{i,j})| \leq \frac{1}{2}\|-\Delta_h v_h\|_\infty.$$

En termes matriciels, ceci s'écrit encore

$$\|A_h^{-1} F_h\|_\infty \leq \frac{1}{2}\|F_h\|_\infty,$$

d'où le résultat en divisant cette inégalité par $\|F_h\|_\infty$. □

Nous avons étudié la consistance grâce au théorème 2.4.1 et la stabilité par l'estimation de $\|A_h^{-1}\|_\infty$. Nous pouvons maintenant énoncer le théorème de convergence.

**Théorème 2.4.4** *Supposons que $u \in \mathcal{C}^4(\bar{\Omega})$, alors*

$$\max_{i,j} |u(z_{ij}) - u_{ij}| \leq \frac{2h^2}{24}\Big(\|\frac{\partial^4 u}{\partial x^4}\|_{L^\infty} + \|\frac{\partial^4 u}{\partial y^4}\|_{L^\infty}\Big).$$

*Démonstration.* On note $\overline{U}_h$ (resp. $F_h$) le vecteur de composantes $u(z_{ij}), 1 \leq i,j \leq N$ (resp. $f(z_{ij})$), et $K_h = A_h\overline{U}_h - F_h$ l'erreur de consistance. En leur associant les fonctions $\overline{u}_h$, $f_h$ et $\kappa_h$ définies sur $\Omega_h$, ceci s'écrit aussi

$$\kappa_h = -\Delta_h \overline{u}_h - f_h.$$

On prolonge $\overline{u}_h$ et $u_h$ sur le bord $\partial\Omega_h$ par les conditions aux limites discrétisées $g(z_{i,j})$. On sait déjà que

$$\|\kappa_h\|_\infty \leq \frac{h^2}{12}\Big(\|\frac{\partial^4 u}{\partial x^4}\|_{L^\infty} + \|\frac{\partial^4 u}{\partial y^4}\|_{L^\infty}\Big).$$

Mais

$$\begin{aligned} -\Delta_h u_h &= f_h, \\ -\Delta_h \overline{u}_h &= f_h + \kappa_h, \end{aligned}$$

d'où

$$\begin{cases} -\Delta_h(\overline{u}_h - u_h) = \kappa_h \text{ dans } \Omega_h, \\ \overline{u}_h - u_h = 0 \text{ sur } \partial\Omega_h. \end{cases}$$

Comme $\overline{u}_h - u_h$ s'annule au bord, on peut donc réécrire ceci matriciellement sous la forme

$$A_h(\overline{U}_h - U_h) = K_h \Longrightarrow \|\overline{U}_h - U_h\|_\infty \leq \|A_h^{-1}\|_\infty \|K_h\|_\infty.$$

Ceci termine la démonstration. □

**Remarque 2.4.3** *i) On peut étendre ces idées en dimension 3, 4 ou plus sans difficulté de principe. Toutefois, les matrices se compliquent et surtout leur taille est de l'ordre de $N^3$, $N^4$ et ainsi de suite. Cette augmentation de taille exponentielle par rapport à la dimension devient vite prohibitive du point de vue pratique.*

*ii) Attention à l'hypothèse de régularité $u \in \mathcal{C}^4(\bar{\Omega})$. Comme on l'a déjà indiqué en passant, elle ne va pas de soi en dimension 2.*