

Les Méthodes Itératives

Prof. Bouras mohammed Chérif

Table des matières

1 Objectifs	1
2 Motivation	2
3 Notions générales	5
3.1 Modèle général d'un schéma itératif	5
3.2 Méthode de Jacobi	6
3.3 Méthode de Gauss-Seidel	7
3.4 Les Méthodes de Relaxation	8
3.5 Les Méthodes par blocs	8
3.6 Suites de vecteurs et de matrices	9
3.7 Convergence des méthodes étudiées	10
3.8 Les matrices de l'itération	10
3.9 Taux de convergence	12
4 Les matrices à diagonales dominantes	14
5 Recherche du paramètre optimal de la méthode de Relaxation (cas des matrices tridiagonales par blocs)	15

1 Objectifs

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice inversible et $b \in \mathcal{M}_{n,1}(\mathbb{C})$. Notre objectif est de résoudre le système linéaire $Ax = b$, c'est-à-dire de trouver x solution

de :

$$\begin{cases} x \in \mathbb{R}^n \\ Ax = b \end{cases} \quad (2.1)$$

et comme A est inversible, alors il existe un unique vecteur $x \in \mathcal{M}_{n,1}(\mathbb{R})$ solution de (2.1). Nous allons donc étudier dans la suite de cette section des méthodes de calcul de ce vecteur x dites méthodes “itératives”.

Un des points essentiels dans l’efficacité des méthodes envisagées concerne la taille des systèmes à résoudre. La taille de la mémoire des ordinateurs a augmenté de façon drastique de 1980 à nos jours.

Le développement des méthodes de résolution de systèmes linéaires est liée à l’évolution des machines informatiques qui est un domaine de recherche très actif que de concevoir des méthodes qui permettent de profiter au mieux de l’architecture des machines (méthodes de décomposition en sous domaines pour profiter des architectures parallèles, par exemple).

La méthode des approximations successives pour la résolution de l’équation de point fixe $x = f(x)$ est donnée par le schéma itératif

$$x_{k+1} = f(x_k)$$

où le point initial x_0 est donné. Si la suite (x_k) converge vers x et si f est continue en ce point alors $f(x) = x$ et les x_k constituent autant d’approximations du point fixe x . On transforme un système linéaire $Ax = b$ en une équation de point fixe en écrivant la matrice A sous la forme $A = M - N$ où M est inversible et «facile à inverser ». Le système devient

$$x = M^{-1}Nx + M^{-1}b$$

qui conduit au schéma itératif

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b$$

que sera l’objet de notre cours.

2 Motivation

Les méthodes itératives deviennent indispensables dès que la taille n du système est très grande. En effet, les méthodes directes exigent un nombre d’opérations à virgule flottante de l’ordre de n^3 lorsque n tend vers l’infini

ce qui les rend lentes pour de grandes valeurs de n . De tels systèmes apparaissent par exemple dans les techniques de résolution numérique d'équations aux dérivées partielles. Les matrices des systèmes obtenus sont en général « creuses » (c'est-à-dire qu'elles ont beaucoup de 0) et (semi) définies positives. Voici un exemple classique.

Étant donnée une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, on se propose de trouver une solution approchée $u : \Omega \subset \mathbb{R}^2$ du problème suivant :

$$\begin{cases} -\Delta u = f, & \forall (x, y) \in \Omega =]0, 1[\times]0, 1[\\ u = 0 & \forall (x, y) \in \partial\Omega \end{cases}$$

où $\Delta u = \frac{\partial^2 u}{\partial^2 x} + \frac{\partial^2 u}{\partial^2 y}$ désigne le Laplacien de la fonction u et $\partial\Omega$ la frontière de Ω .

Pour ce faire, on se donne un réel $h > 0$, on effectue une discrétisation de $\Omega =]0, 1[\times]0, 1[$ de pas h (i.e., on quadrille Ω à l'aide de « petits » pavés d'aire h^2) et on cherche une fonction étagée \tilde{u} (dépendante de h) telle que \tilde{u} tende vers u lorsque h tend vers 0. On pose $h = 1/(n + 1)$ et on écrit

$$u = \sum_{i,j} u_{i,j} \Phi_{i,j}$$

où, pour $1 \leq i, j \leq n$, $\Phi_{i,j}$ est la fonction caractéristique du pavé $P_{i,j} = [(i - 1/2)h, (i + 1/2)h[\times](j - 1/2)h, (j + 1/2)h[$. On note alors $x_{i,j} = (ih, jh)$ les nœuds du quadrillage et $P_{i,j}$ est donc le pavé de centre $x_{i,j}$.

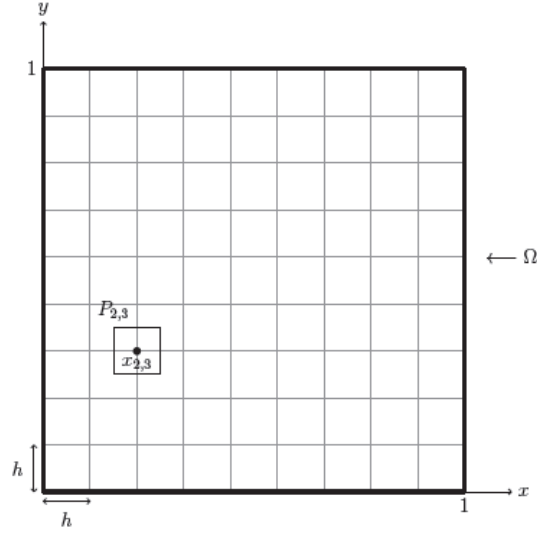
En se basant sur la définition de la dérivée d'une fonction, pour h suffisamment petit, on a les approximations par différences finies suivantes :

$$\begin{aligned} \frac{\partial^2 u}{\partial^2 x} &\approx \frac{u((x + h, y) - 2u(x, y) + u(x - h, y))}{h^2} \\ \frac{\partial^2 u}{\partial^2 y} &\approx \frac{u((x, y + h) - 2u(x, y) + u(x, y - h))}{h^2} \end{aligned}$$

et en décomposant f sous la forme $\sum_{i,j} f_{i,j} \Phi_{i,j}$, notre problème se ramène alors à chercher les $u_{i,j}$; $1 \leq i, j \leq n$ satisfaisants les équations suivantes :

$$\begin{cases} 4u_{i,j} - u_{i-1,j} - u_{i+1,j} - u_{i,j-1} = h^2 f_{i,j}; & 1 \leq i, j \leq n \\ u_{0,j} = u_{n+1,j} = u_{i,0} = u_{i,n+1} & ; \quad 1 \leq i, j \leq n \end{cases}$$

Ce schéma numérique est dit implicite par opposition à un schéma explicite pour lequel il est possible d'ordonner les inconnues de sorte que chacune d'entre elles puisse être déterminée « explicitement » en fonction des



précédentes (système triangulaire). Les schémas numériques implicites ont l'avantage d'être numériquement stables ce qui n'est pas toujours le cas pour les schémas explicites.

On remarque que ces équations forment un système linéaire que nous allons écrire sous forme matricielle. Pour ceci, on pose

$$M = \begin{pmatrix} 4 & -1 & 0 & \dots & 0 \\ -1 & 4 & -1 & \dots & \dots \\ 0 & -1 & 4 & \dots & 0 \\ \dots & \dots & \dots & \dots & -1 \\ 0 & \dots & 0 & -1 & 4 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R})$$

$$X_j = (u_{1,j}, u_{2,j}, \dots, u_{n,j})^T, \text{ pour } 1 \leq j \leq n$$

$$X = (X_1^T, X_2^T, \dots, X_n^T)^T,$$

ainsi que

$$F_j = (f_{1,j}, f_{2,j}, \dots, f_{n,j})^T, \text{ pour } 1 \leq j \leq n$$

$$F = (F_1^T, F_2^T, \dots, F_n^T)^T,$$

En définissant de plus $X_0 = X_n = 0$, le système précédent s'écrit

$$-X_{j-1} + MX_j - X_{j+1} = h^2 F_j, \text{ pour } 1 \leq j \leq n$$

ce qui conduit à

$$AX = h^2 F$$

avec

$$A = \begin{pmatrix} M & -I_n & 0 & \dots & 0 \\ -I_n & M & -I_n & \dots & \dots \\ 0 & -I_n & M & \dots & 0 \\ \dots & \dots & \dots & \dots & -I_n \\ 0 & \dots & 0 & -I_n & M \end{pmatrix} \in \mathcal{M}_{n^2}(\mathbb{R})$$

La matrice A est symétrique réelle et on peut montrer qu'elle est définie positive donc en particulier inversible. En pratique, on a donc à résoudre un système linéaire tridiagonal par blocs de grande taille (notons que faire tendre h vers 0 équivaut à faire tendre n vers l'infini) à résoudre. Notons qu'il existe des méthodes efficaces (e.g., l'algorithme de Thomas qui est une simplification de l'algorithme de Gauss dans le cas particulier des systèmes tridiagonaux) pour résoudre les systèmes linéaires tridiagonaux.

3 Notions générales

On rappelle que les méthodes itératives ne s'appliquent que dans le cas des systèmes à coefficients dans \mathbb{R} ou \mathbb{C}

3.1 Modèle général d'un schéma itératif

On considère une matrice $A \in \mathcal{M}_n(\mathbb{K})$ inversible, un vecteur $b \in \mathcal{M}_{n,1}(\mathbb{K})$ et un système linéaire $(S) : Ax = b$.

Le principe général d'une méthode itérative pour résoudre (S) est de générer une suite de vecteurs qui converge vers la solution $A^{-1}b$. Pour ce faire l'idée est d'écrire le système (S) sous une forme équivalente permettant de voir la solution comme le point fixe d'une certaine fonction, c'est-à-dire

$$(S) \iff x = Bx + c$$

où $B \in \mathcal{M}_n(\mathbb{K})$, $c \in \mathcal{M}_{n,1}(\mathbb{K})$ bien choisis c'est-à-dire $(\mathbb{I} - B)$ inversible et $c = (\mathbb{I} - B)A^{-1}b$.

Pratiquement, on résout les systèmes linéaires successifs : pour tout k ,

$$Mx^{(k+1)} = Nx^{(k)} + b. \tag{1}$$

3.2 Méthode de Jacobi

$A = D - E - F$ avec $a_{kk} \neq 0$ pour $1 \leq k \leq n$,

$$D = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \quad (2)$$

$$E = \begin{pmatrix} 0 & 0 & \dots & 0 \\ a_{21} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{nn-1} & \dots & 0 \end{pmatrix} \quad (1)$$

$$F = \begin{pmatrix} 0 & a_{21} & \dots & a_{n1} \\ 0 & 0 & \dots & \dots \\ \dots & \dots & \dots & a_{n-1n} \\ 0 & \dots & 0 & 0 \end{pmatrix} \quad (2)$$

qui est la décomposition par points de la matrice A .

Définition 1 On appelle méthode itérative de Jacobi par points la méthode itérative :

$$x^0 \in K^n \text{ et } x^{k+1} = D^{-1}(E + F)x^k + D^{-1}b \quad (3)$$

$J = D^{-1}(E + F) = I - D^{-1}A$ est appelée matrice de Jacobi par points. En posant $x^k = (x_1^k, \dots, x_n^k)$, on est conduit à résoudre

$$Dx^{k+1} = (D - A)x^k + b$$

c'est-à-dire

$$\begin{cases} a_{11}x_1^{k+1} = -a_{12}x_2^k - a_{13}x_3^k - \dots a_{1n}x_n^k + b_1 \\ a_{21}x_1^{k+1} = -a_{22}x_2^k - a_{23}x_3^k - \dots a_{2n}x_n^k + b_2 \\ \dots = \dots \\ a_{nn}x_n^{k+1} = -a_{n1}x_1^k - a_{n2}x_2^k - \dots a_{n,n-1}x_{n-1}^k + b_n \end{cases} \quad (4)$$

Exemple 2 Soit à résoudre l'équation :

$$\begin{pmatrix} 10 & 1 \\ 2 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 11 \\ 12 \end{pmatrix}$$

dont la solution est $x_1 = x_2 = 1$.

En initialisant avec le vecteur $x^{(0)} = 0$, on obtient

$$x^{(1)} = \begin{pmatrix} \frac{11}{10} \\ \frac{12}{10} \end{pmatrix}; x^{(2)} = \begin{pmatrix} \frac{98}{100} \\ \frac{98}{100} \end{pmatrix}; x^{(3)} = \begin{pmatrix} \frac{1002}{1000} \\ \frac{1004}{1000} \end{pmatrix}, x^{(4)} = \begin{pmatrix} \frac{9996}{10000} \\ \frac{9992}{10000} \end{pmatrix}$$

Soit à résoudre l'équation :

$$\begin{pmatrix} 1 & 10 \\ 10 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 11 \\ 12 \end{pmatrix}$$

dont la solution est $x_1 = x_2 = 1$.

En initialisant avec le vecteur $x^{(0)} = 0$, on obtient

$$x^{(1)} = \begin{pmatrix} 11 \\ 6 \end{pmatrix}; x^{(2)} = \begin{pmatrix} -49 \\ -49 \end{pmatrix}; x^{(3)} = \begin{pmatrix} 501 \\ 251 \end{pmatrix}, x^{(4)} = \begin{pmatrix} -2499 \\ -2499 \end{pmatrix}$$

On remarque à travers ces deux exemples que la suite engendrée par la méthode de Jacobi peut se "rapprocher" de la solution ou au contraire s'en "éloigner". En conclusion, avant d'entamer les calculs, il faudrait d'abord étudier la convergence de la méthode.

3.3 Méthode de Gauss-Seidel

On peut améliorer la méthode précédente en utilisant les quantités déjà calculées.

Supposons qu'à l'intérieur de la $(k+1)$ ème itération, on ait déjà obtenu $x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_{i-1}^{(k+1)}$, alors en supposant toujours $a_{ii} \neq 0$, on calcul :

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^N a_{ij} x_j^{(k)} \right) \quad (5)$$

et en écrivant (5) sous forme vectorielle, on obtient

$$(D - E) x^{(k+1)} = b + F x^{(k)}$$

ou encore

$$x^{(k+1)} = (D - E)^{-1} F x^{(k)} + (D - E)^{-1} b$$

La matrice $\mathcal{L}_1 = (D - E)^{-1} F$ est appelée matrice de l'itération par point de la méthode de Gauss-Seidel.

Remarque 3 *Très souvent la méthode de Gauss-Seidel convergera plus rapidement que celle de Jacobi car on utilise les nouvelles valeurs des composantes dès qu'elles sont calculées.*

3.4 Les Méthodes de Relaxation

On peut généraliser les deux méthodes précédentes, Jacobi et Gauss-Seidel, en introduisant un paramètre réel ω .

En ajoutant et retranchant la valeur $\frac{1}{\omega}D$ dans $(D - E - F)$, on obtient :

$$\begin{aligned} Ax &= (D - E - F)x = b \\ (D - \frac{1}{\omega}D + \frac{1}{\omega}D - E - F)x &= b \\ x &= (\frac{1}{\omega}D - E)^{-1} (\frac{1-\omega}{\omega}D + F)x + (\frac{1}{\omega}D - E)^{-1} b \end{aligned}$$

Définition 4 *On appelle méthode itérative de relaxation par points la méthode définie pour $\omega \neq 0$ par $x^{(0)} \in \mathcal{M}_{n,1}(\mathbb{K})$ et pour tout $k \geq 0$,*

$$x^{(k+1)} = \left(\frac{1}{\omega}D - E \right)^{-1} \left(\frac{1-\omega}{\omega}D + F \right) x^{(k)} + \left(\frac{1}{\omega}D - E \right)^{-1} b \quad (7)$$

La matrice $\mathcal{L}_\omega = (\frac{1}{\omega}D - E)^{-1} (\frac{1-\omega}{\omega}D + F)$ est appelée matrice de Relaxation par points

Remarque 5 - Pour $\omega = 1$, on retrouve la méthode de Gauss-Seidel.

- Comme $a_{kk} \neq 0$ pour tout k , alors $(\frac{1}{\omega}D - E)$ est inversible.
- Lorsque $\omega > 1$, on parle de sur-relaxation.
- Lorsque $\omega < 1$, on parle de sous-relaxation.

3.5 Les Méthodes par blocs

Toutes les méthodes précédentes peuvent se généraliser en utilisant des décompositions par blocs des matrices.

Supposons que A se décompose par blocs en :

$$\begin{pmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ A_{21} & A_{22} & \dots & A_{2p} \\ \dots & \dots & \dots & \dots \\ A_{p1} & A_{p2} & \dots & A_{pp} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_p \end{pmatrix}$$

où $A_{11}, A_{22}, \dots, A_{pp}$ sont des matrices carrées et où x_i, b_i sont des vecteurs dont la dimension est celle de la matrice A_{ii} .

On définit d'une manière analogue à celle effectuée précédemment la décomposition de A par blocs :

$$A = D - E - F$$

où

$$\begin{aligned} D &= \begin{pmatrix} A_{11} & 0 & \dots & 0 \\ 0 & A_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_{pp} \end{pmatrix} \\ -E &= \begin{pmatrix} 0 & 0 & \dots & 0 \\ A_{21} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ A_{p1} & \dots & A_{p,p-1} & 0 \end{pmatrix} \end{aligned} \tag{3}$$

$$-F = \begin{pmatrix} 0 & A_{21} & \dots & A_{p1} \\ 0 & 0 & \dots & \dots \\ \dots & \dots & \dots & A_{p-1,p} \\ 0 & \dots & 0 & 0 \end{pmatrix} \tag{4}$$

3.6 Suites de vecteurs et de matrices

Théorème 6 Soit $B \in \mathcal{M}_n(\mathbb{K})$. Les conditions suivantes sont équivalentes :

- 1- $\lim_{k \rightarrow +\infty} B^k = 0$
- 2- $\lim_{k \rightarrow +\infty} B^k v = 0$ pour tout $v \in \mathbb{K}^n$
- 3- $\rho(B) < 1$
- 4- $\|B\| < 1$ pour au moins une norme matricielle

Théorème 7 Soit $B \in \mathcal{M}_n(\mathbb{K})$ et $\|\cdot\|$ une norme matricielle. Alors $\lim_{k \rightarrow +\infty} \|B^k\|^{1/k} = \rho(B)$.

3.7 Convergence des méthodes étudiées

Définition 8 La méthode itérative $x^{(0)} \in \mathbb{K}^n$ et pour tout $k \geq 0$,

$$x^{(k+1)} = Bx^{(k)} + c \quad (*)$$

où est dite convergente si pour tout $x^{(0)} \in \mathbb{K}^n$,

$$\lim_{k \rightarrow +\infty} x^{(k)} = x^*$$

où x^* est la solution de $Ax = b$.

Remarque 9 On a alors $x^* = Bx^* + c$

Théorème 10 Critère de convergence des méthodes itératives

Les propositions suivantes sont équivalentes :

- 1– La méthode itérative $(*)$ est convergente vers x^* tel que $x^* = Bx^* + c$;
- 2– $\rho(B) < 1$;
- 3– $\|B\| < 1$ pour au moins une norme matricielle $\|\cdot\|$.

3.8 Les matrices de l'itération

Dans les exemples précédents : Jacobi, Gauss-Seidel et relaxation par points ou par blocs, on a :

– Jacobi :

$$\begin{aligned} M &= D \\ N &= E + F \end{aligned}$$

et la matrice de l'itération de Jacobi est donnée par

$$J = M^{-1}N = D^{-1}(E + F)$$

– Gauss-Seidel :

$$\begin{aligned} M &= D - E \\ N &= F \end{aligned}$$

et la matrice de l'itération de Gauss-Seidel est donnée par

$$= M^{-1}N = D^{-1}(E + F)$$

– Relaxation :

$$\begin{aligned} M &= D/\omega - E \\ N &= \left(\frac{1}{\omega} - 1\right) D + F \end{aligned}$$

et la matrice de l'itération de Jaccobi est donnée par

$$\mathcal{L}_\omega = M^{-1}N = (D/\omega - E)^{-1} \left(\left(\frac{1}{\omega} - 1\right) D + F \right)$$

Ces méthodes s'expriment en fonction de $L = D^{-1}E$ et $U = D^{-1}F$ qui sont respectivement deux matrices strictement triangulaires inférieures ou supérieures.

Pour les méthodes par points on a

$$\begin{aligned} L &= \begin{pmatrix} 0 & 0 & \dots & \dots & 0 \\ \frac{-a_{21}}{a_{22}} & 0 & \dots & \dots & 0 \\ \frac{-a_{31}}{a_{33}} & \frac{-a_{32}}{a_{33}} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \frac{-a_{n1}}{a_{nn}} & \frac{-a_{n2}}{a_{nn}} & \dots & \frac{-a_{n,n-1}}{a_{nn}} & 0 \end{pmatrix} \\ U &= \begin{pmatrix} 0 & \frac{-a_{12}}{a_{11}} & \frac{-a_{13}}{a_{11}} & \dots & \frac{-a_{1n}}{a_{11}} \\ 0 & 0 & \frac{-a_{23}}{a_{22}} & \dots & \frac{-a_{2n}}{a_{22}} \\ 0 & 0 & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & \frac{-a_{n-1,n}}{a_{n-1,n-1}} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \end{aligned}$$

Il faut noter que les matrices de l'itérations seront

$$\text{Jacobi : } J = L + U$$

$$\text{Gauss-Seidel : } \mathcal{L}_1 = (I - L)^{-1} U$$

$$\text{Relaxation : } \mathcal{L}_\omega = (I - \omega L)^{-1} ((1 - \omega) I + \omega U)$$

Exemple 11 On reprend les exemples déjà traités au début du chapitre, à savoir

$$\begin{pmatrix} 10 & 1 \\ 2 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 11 \\ 12 \end{pmatrix}$$

La matrice de l'itération de Jacobi est

$$J = \begin{pmatrix} 0 & -1/10 \\ -2/10 & 0 \end{pmatrix}$$

dont les valeurs propres sont $\lambda = \pm\sqrt{2}/10$, et par suite le rayon spectral sera $\rho(J) < 1$, donc la méthode de Jacobi est convergente.

Par contre dans le deuxième exemple,

$$\begin{pmatrix} 1 & 10 \\ 10 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 11 \\ 12 \end{pmatrix}$$

la matrice de l'itération de Jacobi est donnée par

$$J = \begin{pmatrix} 0 & -10 \\ -5 & 0 \end{pmatrix},$$

dont les valeurs propres sont $-5\sqrt{2}, 5\sqrt{2}$, c'est à dire $\rho(J) = 5\sqrt{2}$, donc la méthode de Jacobi diverge.

Théorème 12 Pour toute matrice A , le rayon spectral de la matrice de l'itération de la méthode de Relaxation \mathcal{L}_ω est supérieur ou égal à $|\omega - 1|$, c'est à dire

$$\rho(\mathcal{L}_\omega) \geq |\omega - 1|$$

Corollaire 13 Pour toute matrice A , une condition nécessaire de convergence de la méthode de Relaxation est que

$$0 < \omega < 2$$

3.9 Taux de convergence

On peut caractériser la convergence d'une méthode par la vitesse à laquelle l'erreur $e^{(k)}$ tend vers 0.

Pour une méthode dont la matrice d'itération est B , on a :

$$e^{(k)} = B^{(k)}e^{(0)}$$

et pour une norme choisie, on a

$$\|e^{(k)}\| = \|B^{(k)}e^{(0)}\| \leq \|B^{(k)}\| \|e^{(0)}\|$$

et par suite

$$\frac{\|e^{(k)}\|}{\|e^{(0)}\|} \leq \|B^{(k)}\|$$

On définit σ le facteur moyen de réduction de l'erreur par itération par la formule :

$$\sigma^k = \frac{\|e^{(k)}\|}{\|e^{(0)}\|}$$

ce qui nous donne

$$\sigma \leq \|B^{(k)}\|^{1/k}$$

Définition 14 On appelle *taux moyen de convergence* pour k itérations le nombre

$$R_k(B) = -\ln \|B^{(k)}\|^{1/k}$$

Remarque 15 Comme $\|B^{(k)}\|^{1/k}$ alors $-\ln \|B^{(k)}\|^{1/k} > 0$. Le *taux moyen de convergence* est inversement proportionnel au nombre d'itérations. En effet, pour déterminer K tel que :

$$\frac{\|e^{(k)}\|}{\|e^{(0)}\|} \leq \varepsilon, \text{ impose } \|B^{(k)}\| \leq \varepsilon$$

d'où

$$-\ln \|B^{(k)}\|^{1/k} \geq \frac{1}{k} \ln \varepsilon$$

donc

$$k \geq \frac{-\ln \varepsilon}{-\ln \|B^{(k)}\|^{1/k}} = \frac{-\ln \varepsilon}{R_k(B)}$$

Cette définition présente néanmoins quelques inconvénients. Le *taux de convergence* dépend de k et aussi de la norme choisie.

On ne calcul jamais $\|B^{(k)}\|^{1/k}$, c'est trop volumineux en nombre d'opérations. On va donc définir un *taux de convergence asymptotique*, c'est-à-dire la limite, si elle existe, de *taux moyen* pour k itérations.

Dans ce cas, on va choisir la norme $\|\cdot\|_2$, ceci nous donnera

$$\|B^{(k)}\|^{1/k} = \rho(B^k)^{1/k} = (\rho^k(B))^{1/k} = \rho(B)$$

Définition 16 On appelle *taux de convergence asymptotique* de la matrice de l'itération B le nombre

$$R(B) = -\ln \rho(B)$$

Proposition 17 Le nombre d'itérations k pour réduire l'erreur e vérifie

$$k \geq \frac{-\ln \varepsilon}{R(B)}$$

4 Les matrices à diagonales dominantes

En algèbre linéaire, une matrice carrée à coefficients réels ou complexes est dite à diagonale dominante lorsque le module de chaque terme diagonal est supérieur ou égal à la somme des modules des autres termes de sa ligne, c'est-à-dire Si $A \in \mathcal{M}_n(\mathbb{K})$, alors pour tout $1 \leq i \leq n$ on a

$$|a_{ii}| \geq \sum_{\substack{i,j=1,n \\ i \neq j}} |a_{ij}|$$

et de la même manière, A est dite à diagonale strictement dominante lorsque pour tout $1 \leq i \leq n$ on a

$$|a_{ii}| > \sum_{\substack{i,j=1,n \\ i \neq j}} |a_{ij}|$$

Théorème 18 *A étant une matrice à diagonale strictement dominante, alors la méthode de Jacobi est convergente.*

Preuve. La matrice de l'itération de la méthode de Jacobi est donnée par

$$J = D^{-1}(E + F)$$

où $J_{ii} = 0$ et $J_{ij} = \frac{-a_{ij}}{a_{ii}}$ pour $i \neq j$. Et par suite

$$\sum_{j=1}^n |J_{ij}| = \sum_{\substack{j=1 \\ i \neq j}}^n \left| \frac{a_{ij}}{a_{ii}} \right| < 1$$

car

$$|a_{ii}| > \sum_{\substack{i,j=1,n \\ i \neq j}} |a_{ij}|$$

Donc

$$\max_{1 \leq i \leq n} \sum_{j=1}^n |J_{ij}| = \|J\|_{\infty} < 1$$

et par suite

$$\rho(J) \leq \|J\|_{\infty} < 1$$

ce qui nous donne la convergence de la méthode de Jacobi sous l'hypothèse que la matrice A est à diagonale strictement dominante. ■

Théorème 19 *A étant une matrice à diagonale strictement dominante, si $0 < \omega \leq 1$ alors la méthode de Relaxation par points converge.*

Preuve. A faire comme devoir à la maison qui sera comptabilisé dans la note de TD ■

Remarque 20 *Ce théorème n'est qu'une condition suffisante de convergence.*

Il ne prouve pas que si $\omega > 1$, la méthode de Relaxation diverge. Au contraire, on verra que dans certains cas, le meilleur choix de ω correspond à $\omega > 1$.

Par exemple, soit

$$A = \begin{pmatrix} 2 & 5 \\ 0 & 1 \end{pmatrix}$$

est une matrice à diagonale strictement dominante, la matrice de l'itération de Relaxation

$$\mathcal{L}_\omega = \begin{pmatrix} 1 - \omega & -\omega/2 \\ 0 & 1 - \omega \end{pmatrix}$$

a pour valeurs propres $\lambda = 1 - \omega$. Donc la méthode de Relaxation converge pour $|1 - \omega| < 1$, c'est-à-dire $0 < \omega < 2$

5 Recherche du paramètre optimal de la méthode de Relaxation (cas des matrices tridiagonales par blocs)

Définition 21 *A est une matrice tridiagonale par blocs si elle s'écrit sous la forme*

$$A = \begin{pmatrix} A_{11} & A_{12} & 0 & \dots & 0 \\ A_{21} & A_{22} & \dots & 0 & 0 \\ 0 & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & A_{p-1,p} \\ 0 & \dots & 0 & A_{p,p-1} & A_{p,p} \end{pmatrix}$$

où chaque A_{ii} est sous matrice carrée d'ordre n_i avec $\sum_{i=1}^n n_i = n$.

Dans tous ce qui suit, on utilise la décomposition par blocs $A = D - E - F$ où D est la matrice diagonale par blocs formée des éléments A_{ii} , E est la matrice triangulaire inférieure et F la matrice triangulaire supérieure.

Lemme 22 Soit A une matrice tridiagonale par blocs. Soit μ un nombre complexe non nul et soit

$$A(\mu) = D - \mu E - \frac{1}{\mu} F$$

alors

$$\det A(\mu) = \det A$$

Théorème 23 Soit A une matrice tridiagonale par blocs, dont les blocs diagonaux sont inversibles.

Si λ est valeur propre de la matrice de l'itération de Jacobi par blocs, alors $-\lambda$ est aussi valeur propre. De plus les méthodes de Jacobi et de Gauss-Seidel par blocs convergent simultanément.

Dans le cas de convergence, la méthode de Gauss-Seidel par blocs converge asymptotiquement deux fois plus vite que celle de Jacobi par blocs.

Preuve. Soit $D_J(\lambda)$ le polynôme caractéristique de la matrice $J = D^{-1}(E + F)$, alors

$$D_J(\lambda) = \det(\lambda I - J) = \det(\lambda I - D^{-1}E - D^{-1}F)$$

et en multipliant cette valeur par $\det D$, on obtient

$$\det D.D_J(\lambda) = \det(\lambda D - E - F)$$

Appliquons le lemme précédent pour $\mu = -1$, on aura

$$\begin{aligned} \det D.D_J(\lambda) &= \det(\lambda D + E + F) = (-1)^n \det(-\lambda D - E - F) \\ &= (-1)^n \det D \det(-\lambda I - D^{-1}(E + F)) \\ &= (-1)^n \det D.D_J(-\lambda) \end{aligned}$$

et comme par hypothèse $\det D \neq 0$, on aura la relation

$$D_J(\lambda) = (-1)^n D_J(-\lambda)$$

ce qui signifie que si λ est valeur propre de J alors $-\lambda$ est aussi valeur propre de J .

Etudions maintenant le polynôme caractéristique $D_{\mathcal{L}_1}(\lambda)$ associé à la matrice de l'itération de Gauss-Seidel $\mathcal{L}_1 = (D - E)^{-1}F$. On a

$$D_{\mathcal{L}_1}(\lambda) = \det(\lambda I - \mathcal{L}_1) = \det(\lambda I - (D - E)^{-1}F)$$

et en multipliant cette valeur par $\det(D - E)$, on aura pour $\lambda \neq 0$

$$\begin{aligned}\det(D - E)D_{\mathcal{L}_1}(\lambda) &= \det(\lambda(D - E) - F) \\ &= \det(\lambda D - \lambda E - F) \\ &= \lambda^{n/2} \det(\lambda^{1/2} D - \lambda^{1/2} E - \frac{1}{\lambda^{1/2}} F)\end{aligned}$$

Appliquons le lemme précédent pour $\mu = \lambda^{1/2}$, on aura

$$\begin{aligned}\det(D - E)D_{\mathcal{L}_1}(\lambda) &= \lambda^{n/2} \det(\lambda^{1/2} D - E - F) \\ &= \lambda^{n/2} \det D \cdot \det(\lambda^{1/2} I - D^{-1}(E + F)) \\ &= \lambda^{n/2} \det D \cdot D_J(\lambda^{1/2})\end{aligned}$$

et comme $\det D = \prod_{i=1}^n \det A_{ii} = \det(D - E)$ alors

$$D_{\mathcal{L}_1}(\lambda) = \lambda^{n/2} D_J(\lambda^{1/2})$$

Donc sauf pour $\lambda = 0$, si λ est valeur propre de \mathcal{L}_1 alors $\lambda^{1/2}$ est aussi valeur propre de J et par suite

$$\rho(\mathcal{L}_1) = \rho^2(J)$$

Si la méthode de Gauss-Seidel converge, alors

$$\rho(\mathcal{L}_1) = \rho^2(J) < 1 \Leftrightarrow \rho(J) < 1$$

donc la méthode de Jacobi converge aussi. Et de la même manière, si la méthode de Jacobi converge, alors

$$\rho(J) = \sqrt{\rho(\mathcal{L}_1)} < 1 \Leftrightarrow \rho(\mathcal{L}_1) < 1$$

donc la méthode de Gauss-Seidel converge aussi. Il en est de même pour la divergence. En conclusion les deux méthodes convergent ou divergent simultanément.

De plus

$$R(\mathcal{L}_1) = -\ln \rho(\mathcal{L}_1) = -\ln \rho^2(J) = -2 \ln \rho(J) = 2R(J)$$

c'est-à-dire la méthode de Gauss-Seidel converge asymptotiquement deux fois plus vite que celle de Jacobi. ■

Théorème 24 Soit A une matrice tridiagonale par blocs, avec $\det A_{ii} \neq 0$. Si λ est une valeur propre de J et si η vérifie la relation :

$$(\eta + \omega - 1)^2 = \eta \omega^2 \lambda^2 \quad (**)$$

alors η est une valeur propre de \mathcal{L}_ω .

Théorème 25 Réciproquement, si η est une valeur propre non nulle de \mathcal{L}_ω et si λ vérifie la relation (**), alors λ est une valeur propre de J .

Preuve. Les valeurs propres de \mathcal{L}_ω sont les racines du polynôme caractéristique

$$D_{\mathcal{L}_\omega}(\eta) = \det(\eta I - \mathcal{L}_\omega) = \det(\eta I - (D - \omega E)^{-1}((1 - \omega)D + \omega F))$$

et en multipliant cette valeur par $\det(D - \omega E)$, on aura

$$\begin{aligned} \det(D - \omega E)D_{\mathcal{L}_\omega}(\eta) &= \det(\eta(D - \omega E) - ((1 - \omega)D - \omega F)) \\ &= \det((\eta + \omega - 1)D - \eta\omega E - \omega F) \end{aligned}$$

Sauf pour $\omega = 1$ (le cas de Gauss-Seidel), 0 n'est pas valeur propre de \mathcal{L}_ω , car

$$\det((1 - \omega)D + \omega F) = (1 - \omega)^n \left(\prod_{i=1}^n \det A_{ii} \right) \neq 0$$

Donc :

$$\det(D - \omega E)D_{\mathcal{L}_\omega}(\eta) = \eta^{n/2} \omega^n \det \left(\frac{(\eta + \omega - 1)}{\eta^{1/2} \omega} D - \eta^{1/2} E - \frac{1}{\eta^{1/2}} F \right)$$

Appliquons le lemme précédent pour $\mu = \eta^{1/2}$, on aura

$$\begin{aligned} \det(D - \omega E).D_{\mathcal{L}_\omega}(\eta) &= \eta^{n/2} \omega^n \det \left(\frac{(\eta + \omega - 1)}{\eta^{1/2} \omega} D - E - F \right) \\ &= \eta^{n/2} \omega^n \det D \det \left(\frac{(\eta + \omega - 1)}{\eta^{1/2} \omega} I - D^{-1}(E + F) \right) \\ &= \eta^{n/2} \omega^n \det D.D_J \left(\frac{\eta + \omega - 1}{\eta^{1/2} \omega} \right) \end{aligned}$$

Comme $\det(D - \omega E) = \det D$, on a donc

$$D_{\mathcal{L}_\omega}(\eta) = \eta^{n/2} \omega^n D_J \left(\frac{\eta + \omega - 1}{\eta^{1/2} \omega} \right)$$

Donc si η est valeur propre de $\mathcal{L}_\omega, \eta \neq 0 (\omega \neq 1)$ et $\frac{\eta + \omega - 1}{\eta^{1/2}\omega}$ est une valeur propre de J . Réciproquement, si λ est une valeur propre de J et si η vérifie

$$\frac{(\eta + \omega - 1)^2}{\eta} = \omega^2 \lambda^2$$

alors η est une valeur propre de \mathcal{L}_ω . ■

Théorème 26 *Si A une matrice tridiagonale par blocs, avec $\det A_{ii} \neq 0$. Si toutes les valeurs propres de la matrice de Jacobi sont réelles, alors la méthode de Jacobi par blocs et la méthode de Relaxation par blocs ($0 < \omega < 2$) convergent ou divergent simultanément.*

Dans le cas de la convergence, il existe une valeur ω^ telle que*

$$\omega^* = \frac{2}{1 + \sqrt{1 - \rho^2(J)}}$$

pour laquelle $\rho(\mathcal{L}_\omega)$ est minimale et vaut $\omega^ - 1$.*