

Chapitre 3

MÉTHODES ITÉRATIVES POUR LA RÉOLUTION DES SYSTÈMES LINÉAIRES

Chapitre 3

MÉTHODES ITÉRATIVES POUR LA RÉOLUTION DES SYSTÈMES LINÉAIRES

3.1 Introduction

On va étudier les méthodes itératives pour résoudre un système linéaire :

$$(3.1) \quad Ax = b$$

Une méthode itérative engendre une suite de vecteurs qui doit tendre vers la solution de l'équation (3.1). Dans les méthodes itératives qu'on va étudier, le passage d'un vecteur x^k de la suite au suivant se fait en corrigeant une ou plusieurs composantes de ce vecteur.

3.2 Étude générale

Soit $A = (a_{ij})_{1 \leq i, j \leq n}$ une matrice à coefficients complexes inversible d'ordre n et b un vecteur donné de $\mathbf{C}^n$. Les méthodes qu'on va étudier peuvent se formuler de la manière générale suivante :

-On décompose A sous la forme $A = M - N$ où M est une matrice "facilement" inversible, c'est-à-dire telle que la résolution numérique d'un système linéaire, dont la matrice est M , nécessite relativement peu d'opérations à effectuer. On a

$$Ax = b \Leftrightarrow Mx = Nx + b$$

-La matrice M étant inversible, on définit la méthode itérative permettant de calculer la solution de (3.1) par

$$(3.2) \quad \begin{cases} x^0 & \text{donné dans } \mathbf{C}^n \\ Mx^{k+1} = Nx^k + b \end{cases}$$

Soit, en posant $B = M^{-1}N$, l'écriture équivalente :

$$(3.3) \quad x^{k+1} = Bx^k + c$$

où $c = M^{-1}b$. On dit que B est la matrice de la méthode itérative (3.2). Dans la suite, on commence par étudier la convergence de la méthode itérative (3.2).

Remarque

Dans le cas des matrices creuses, c'est-à-dire celles comportant un grand nombre de coefficients nuls, les méthodes itératives nécessitent en général beaucoup moins de place mémoire dans l'ordinateur que les méthodes directes.

Lemme 3.1

Soit $B \in \mathcal{M}_n(\mathbf{C})$, alors les quatre propriétés suivantes sont équivalentes.

- i) B^k tend vers 0, lorsque k tend vers $+\infty$.
- ii) $B^k v$ tend vers 0 pour tout choix de v dans $\mathbf{C}^n$, lorsque k tend vers $+\infty$.
- iii) $\rho(B) < 1$.
- iv) $\|B\| < 1$ pour au moins une norme matricielle subordonnée.

Démonstration :

i) $\Rightarrow$ ii)

Cette implication est immédiate puisque l'on a pour une norme matricielle subordonnée $\|\cdot\|$:

$$\forall v \in \mathbf{C}^n, \forall k \geq 1, \quad \left\| B^k v \right\| \leq \left\| B^k \right\| \|v\|$$

ii) $\implies$ iii)

Montrons cette implication par l'absurde. Supposons que $\rho(B) \geq 1$, alors ils existent $\lambda \in \mathbf{C}$ tel que $|\lambda| \geq 1$ et $p \in \mathbf{C}^n \setminus \{0\}$ pour lesquels on a

$$Bp = \lambda p$$

d'où

$$\left\| B^k p \right\| = |\lambda|^k \|p\| \geq \|p\| > 0$$

par conséquent, puisque $|\lambda| \geq 1$, on obtient

$$\lim_{k \rightarrow +\infty} \left\| B^k p \right\| \neq 0$$

iii) $\implies$ iv)

On sait d'après le théorème 1.1 du chapitre 1, que pour tout $\varepsilon > 0$, il existe une norme matricielle subordonnée $\|\cdot\|$ telle que

$$\|B\| \leq \rho(B) + \varepsilon$$

utilisant le fait que $\rho(B) < 1$, alors on peut prendre $\varepsilon = \frac{1 - \rho(B)}{2}$, et on a

$$\|B\| \leq \frac{1 + \rho(B)}{2} < 1$$

iv) $\implies$ i)

On a

$$0 \leq \left\| B^k \right\| \leq \|B\|^k$$

On en déduit puisque $\|B\| < 1$, que

$$\lim_{k \rightarrow +\infty} \left\| B^k \right\| = 0$$

c'est-à-dire que B^k tend vers 0 dans $\mathcal{M}_n(\mathbf{C})$, lorsque k tend vers $+\infty$. ■

Théorème 3.1

Soient $B \in \mathcal{M}_n(\mathbf{C})$, $H = I - B$ et $c \in \mathbf{C}^n$. La suite définie par

$$\begin{cases} x^0 & \text{donné} \\ x^{k+1} & = Bx^k + c \end{cases}$$

converge, pour tout choix x^0 vers un élément $\tilde{x} \in \mathbf{C}^n$, solution du système linéaire :

$$H \tilde{x} = c$$

si et seulement si $\rho(B) < 1$.

Démonstration :

Supposons que $\rho(B) < 1$, alors d'après le lemme 3.1, $\|B\| < 1$ pour au moins une norme matricielle subordonnée et la matrice $H = I - B$ est donc inversible d'après le théorème 1.2 du chapitre 1. D'où en prenant $\tilde{x} = H^{-1}c$, on a $H\tilde{x} = c$ ou encore

$$\tilde{x} = B\tilde{x} + c$$

On en déduit que

$$x^k - \tilde{x} = B(x^{k-1} - \tilde{x}) = \dots = B^k(x^0 - \tilde{x})$$

d'où, d'après le lemme 3.1 :

$$\lim_{k \rightarrow +\infty} (x^k - \tilde{x}) = 0$$

pour tout choix $x^0 \in \mathbf{C}^n$.

Réciproquement, si la suite (x^k) converge pour tout choix (x^0) vers $\tilde{x}$ solution de $H\tilde{x} = c$, le même raisonnement entraîne que

$$\forall x^0 \in \mathbf{C}^n \quad \lim_{k \rightarrow +\infty} B^k(x^0 - \tilde{x}) = 0$$

autrement dit

$$\forall v \in \mathbf{C}^n \quad \lim_{k \rightarrow +\infty} B^k v = 0$$

On en déduit en utilisant le lemme 3.1 que $\rho(B) < 1$. ■

3.3 La méthode de Jacobi

Soit $A = (a_{ij})_{1 \leq i, j \leq n} \in \mathcal{M}_n(\mathbf{C}^n)$ inversible. Soit $x \in \mathbf{C}^n$, la solution du système linéaire (3.1), où b est un vecteur donné de $\mathbf{C}^n$.

La méthode de Jacobi n'est utilisable que si $a_{ii} \neq 0$ pour $i = 1$ à n . À partir d'un vecteur x^0 donné dans $\mathbf{C}^n$, la méthode de Jacobi consiste à construire la suite de vecteurs x^k de la manière suivante :

Pour $i = 1$ à n , on calcule

$$(3.4) \quad x_i^{k+1} = \frac{1}{a_{ii}} (b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^k)$$

Pour écrire la relation (3.4) sous forme matricielle, on décompose la matrice A sous la forme

$$(3.5) \quad A = D - E - F$$

où D , E et F sont les matrices définies par

$$(3.6) \quad \begin{aligned} D = (d_{ij}) \quad \text{tel que} \quad & \begin{cases} d_{ij} = 0 & \text{pour } i \neq j \\ d_{ii} = a_{ii} & \text{pour } i = 1 \text{ à } n \end{cases} \\ E = (e_{ij}) \quad \text{tel que} \quad & \begin{cases} e_{ij} = -a_{ij} & \text{pour } i > j \\ e_{ij} = 0 & \text{pour } i \leq j \end{cases} \\ F = (f_{ij}) \quad \text{tel que} \quad & \begin{cases} f_{ij} = 0 & \text{pour } i \geq j \\ f_{ij} = -a_{ij} & \text{pour } i < j \end{cases} \end{aligned}$$

Alors, comme on a supposé que $a_{ii} \neq 0$ pour $i = 1$ à n , D est inversible et la relation (3.4) s'écrit

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

Définition 3.1

Soit $A \in \mathcal{M}_n(\mathbb{C})$, on dit que $A = (a_{ij})_{1 \leq i, j \leq n}$ est à diagonale strictement dominante si et seulement si

$$\forall i \in \{1, 2, \dots, n\}, \quad |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$$

■

Théorème 3.2

Soit A une matrice à diagonale strictement dominante, alors la méthode de Jacobi pour la résolution d'un système linéaire de matrice A est convergente.

Démonstration :

La matrice D définie par (3.6) est inversible puisque les coefficients diagonaux de la matrice A sont non nuls. En effet A est à diagonale strictement dominante par hypothèse d'où d'après la définition 3.1

$$\forall i \in \{1, 2, \dots, n\}, \quad |d_{ii}| = |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \geq 0$$

La matrice J de la méthode de Jacobi est donnée par

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

où D, E et F sont définies par (3.6). Les coefficients sont donc donnés par

$$\forall i, j \in \{1, 2, \dots, n\}, \quad J_{ij} = \begin{cases} -\frac{a_{ij}}{a_{ii}} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases}$$

On a

$$(3.7) \quad \|J\|_{\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |J_{ij}| = \max_{1 \leq i \leq n} \sum_{\substack{j=1 \\ j \neq i}}^n \left| \frac{a_{ij}}{a_{ii}} \right| = \max_{1 \leq i \leq n} \frac{1}{|a_{ii}|} \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$$

Comme A est à diagonale strictement dominante, alors

$$\forall i \in \{1, 2, \dots, n\}, \quad |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$$

D'où d'après (3.7), $\|J\|_{\infty} < 1$. On en déduit d'après le lemme 3.1 que $\rho(J) < 1$ et par conséquent, d'après le théorème 3.1, la méthode de Jacobi est convergente. ■

3.4 La méthode de Gauss-Seidel

Au lieu d'attendre une itération entière pour corriger les composantes de x^k , on peut le faire "au fur et à mesure". Supposons qu'à l'intérieur de la $(k+1)^{\text{ième}}$ 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 calcule

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

ce qui s'écrit vectoriellement $Dx^{k+1} = b + Ex^{k+1} + Fx^k$, ou encore

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

Théorème 3.3

Soit A une matrice à diagonale strictement dominante, alors la méthode de Gauss-Seidel pour la résolution d'un système linéaire de matrice A est convergente.

Démonstration :

Les coefficients diagonaux de la matrice A sont non nuls car A est à diagonale strictement dominante. Il en résulte que $D-E$ où D et E sont les matrices définies par (3.6), est inversible. La matrice $\mathcal{L}_1$ de la matrice de Gauss-Seidel est donnée par

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

Raisonnons par l'absurde en supposant qu'il existe une valeur propre λ de $\mathcal{L}_1$ avec $|\lambda| \geq 1$, on a

$$(3.8) \quad \det(\lambda I - (D - E)^{-1}F) = \lambda^n \det((D - E)^{-1}) \det((D - E) - \frac{1}{\lambda}F)$$

On va montrer que pour $|\lambda| \geq 1$, la matrice $D - E - \frac{1}{\lambda}F$ est à diagonale strictement dominante. En effet, on a d'après (3.6)

$$(3.9) \quad (D - E - \frac{1}{\lambda}F)_{ij} = \begin{cases} a_{ij} & \text{si } i \geq j \\ \frac{a_{ij}}{\lambda} & \text{si } i < j \end{cases}$$

et par conséquent, puisque la matrice A est à diagonale strictement dominante et $|\lambda| \geq 1$:

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

On en déduit d'après (3.9) que la matrice $D - E - \frac{1}{\lambda}F$ est strictement dominante. La matrice $D - E - \frac{1}{\lambda}F$ est donc inversible pour $|\lambda| \geq 1$, on en déduit d'après (3.8) que

$$\det(\lambda I - (D - E)^{-1}F) \neq 0$$

Ce qui est en contradiction avec le fait que λ est valeur propre de $\mathcal{L}_1$. En conclusion, on ne peut avoir une valeur propre λ telle que $|\lambda| \geq 1$, donc nécessairement $\rho(\mathcal{L}_1) < 1$ et la méthode de Gauss-Seidel converge. ■

Lemme 3.2

Soit A une matrice tridiagonale. Soit μ un nombre complexe non nul et soit $A(\mu) = D - \mu E - \frac{1}{\mu}F$ alors

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

Démonstration :

On considère la matrice diagonale S telle que

$$\forall i \in \{1, 2, \dots, n\}, \quad S_{ii} = \mu^i$$

On remarque que l'on a

$$(S^{-1})_{ij} = \frac{1}{\mu^i} \delta_{ij}$$

En effectuant les produits :

$$(SA)_{il} = \sum_{k=1}^p S_{ik}A_{kl} = S_{ii}A_{il} = \mu^i A_{il}$$

$$(SAS^{-1})_{ij} = \sum_{l=1}^p (SA)_{il}(S^{-1})_{lj} = \sum_{l=1}^p \mu^i A_{il} \cdot \frac{1}{\mu^l} \delta_{lj} = \frac{\mu^i}{\mu^j} A_{ij}$$

On obtient

$$(SAS^{-1})_{ii} = A_{ii}$$

$$(SAS^{-1})_{i,i-1} = \mu A_{i,i-1}$$

$$(SAS^{-1})_{i,i+1} = \frac{1}{\mu} A_{i,i+1}$$

$$(SAS^{-1})_{i,j} = 0 \quad j \notin \{i-1, i, i+1\}$$

On en déduit que $A(\mu) = SAS^{-1}$, donc $\det(A(\mu)) = \det(SAS^{-1}) = \det A$. ■

Théorème 3.4

Soit A une matrice tridiagonale. Alors les rayons spectraux des matrices de Jacobi et de Gauss-Seidel sont liés par la relation

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

où $\mathcal{L}_1 = (D - E)^{-1}F$ et $J = D^{-1}(E + F)$ sont respectivement les matrices des méthodes de Gauss-Seidel et de Jacobi. Donc les deux méthodes convergent ou divergent simultanément. Lorsqu'elles convergent, la méthode de Gauss-Seidel converge plus rapidement que la méthode de Jacobi.

Démonstration :

Soient $\text{Sp}(J)$ et $\text{Sp}(\mathcal{L}_1)$ les ensembles respectifs des valeurs propres de J et de $\mathcal{L}_1$. On a

$$(3.10) \quad (\rho(J))^2 = \left(\sup_{\lambda \in \text{Sp}(J)} |\lambda| \right)^2 = \sup_{\lambda \in \text{Sp}(J)} |\lambda|^2 = \sup_{\alpha \in S} |\alpha|$$

où S est l'ensemble défini par

$$S \equiv \{\alpha \in \mathbf{C} / \exists \lambda \in \text{Sp}(J) : \alpha = \lambda^2\}$$

et en désignant par $\alpha^{\frac{1}{2}}$ l'une quelconque des deux racines du nombre complexe α , on peut écrire

$$(3.11) \quad S = \{\alpha \in \mathbf{C} / \alpha^{\frac{1}{2}} \in \text{Sp}(J) \text{ ou } -\alpha^{\frac{1}{2}} \in \text{Sp}(J)\}$$

On va montrer que

$$(3.12) \quad \lambda \in \text{Sp}(J) \iff -\lambda \in \text{Sp}(J)$$

En effet, soit P_J le polynôme caractéristique de la matrice de Jacobi, on va montrer que $P_J(\lambda) = (-1)^n P_J(-\lambda)$. On a

$$(3.13) \quad P_J(\lambda) = \det(\lambda I - J) = \det(\lambda I - D^{-1}E - D^{-1}F) = \det(\lambda D - E - F) \det(D^{-1})$$

On en déduit d'après le lemme 3.2 puisque la matrice $A_\lambda = \lambda D - E - F$ est tridiagonale et donc $\det(A_\lambda) = \det(A_\lambda(-1)) = \det(\lambda D + E + F)$, que

$$P_J(\lambda) = \det(\lambda D - E - F) \det(D^{-1}) = (-1)^n \det(-\lambda D - E - F) \det(D^{-1}) = (-1)^n P_J(-\lambda)$$

D'où l'équivalence (3.12). On en déduit d'après (3.11) et (3.12) que

$$(3.14) \quad S = \{\alpha \in \mathbf{C} / \quad \alpha^{\frac{1}{2}} \in Sp(J)\}$$

On va montrer que $S = Sp(\mathcal{L}_1)$ où $Sp(\mathcal{L}_1)$ est l'ensemble des valeurs propres de $\mathcal{L}_1$. On a pour $\lambda \neq 0$

$$(3.15) \quad \begin{aligned} \det(D - E)P_{\mathcal{L}_1}(\lambda) &= \det(D - E)\det(\lambda I - (D - E)^{-1}F) \\ &= \det(\lambda(D - E) - F) \\ &= \det(\lambda^{\frac{1}{2}}(\lambda^{\frac{1}{2}}D - \lambda^{\frac{1}{2}}E - \lambda^{-\frac{1}{2}}F)) \\ &= \lambda^{\frac{n}{2}}\det(\lambda^{\frac{1}{2}}D - \lambda^{\frac{1}{2}}E - \lambda^{-\frac{1}{2}}F) \end{aligned}$$

La matrice $A_{\lambda^{\frac{1}{2}}} = \lambda^{\frac{1}{2}}D - E - F$ est tridiagonale, on en déduit en utilisant le lemme 3.2 que

$$\det(A_{\lambda^{\frac{1}{2}}}) = \det(A_{\lambda^{\frac{1}{2}}}(\lambda^{\frac{1}{2}})) = \det(\lambda^{\frac{1}{2}}D - \lambda^{\frac{1}{2}}E - \lambda^{-\frac{1}{2}}F)$$

On en déduit d'après (3.13) et (3.15), que

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

Cette relation est aussi valable pour $\lambda = 0$ par continuité. De cette relation, on déduit en utilisant (3.14) que $S = Sp(\mathcal{L}_1)$. D'où d'après (3.10), la relation

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

Cette relation entraîne, que les méthodes de Jacobi et de Gauss-Seidel convergent ou divergent simultanément. Lorsqu'elles convergent, la méthode de Gauss-Seidel converge plus rapidement que la méthode de Jacobi. ■

3.5 La méthode de relaxation

La méthode de Jacobi correspond à la décomposition $A = M - N$ avec $M = D$ et $N = E + F$. La méthode de Gauss-Seidel revient à prendre $M = D - E$ et $N = F$. Dans la méthode de relaxation, on introduit un paramètre réel ω non nul et on prend

$$M = \frac{1}{\omega}D - E$$

$$N = \frac{1 - \omega}{\omega}D + F$$

où D, E et F sont définies par (3.6). La matrice de la méthode itérative ainsi définie est alors donnée par

$$(3.16) \quad \mathcal{L}_\omega = \left(\frac{1}{\omega}D - E \right)^{-1} \left(\frac{1 - \omega}{\omega}D + F \right) = (I - \omega L)^{-1} ((1 - \omega)I + \omega U)$$

où $L = D^{-1}E$ et $U = D^{-1}F$.

Remarques :

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

2) On va chercher les valeurs de ω pour lesquelles la méthode de relaxation converge. On examinera ensuite si pour certaines valeurs de ω , la convergence est plus rapide que pour d'autres valeurs de ce paramètre.

Proposition 3.1

La méthode de relaxation diverge pour tout ω dans $\mathbb{R} \setminus [0, 2]$.

Démonstration :

On peut facilement calculer le déterminant de la matrice $\mathcal{L}_\omega$ définie par (3.16) :

$$\det \mathcal{L}_\omega = \det \left[(I - \omega L)^{-1} \right] \det [(1 - \omega)I + \omega U]$$

On peut facilement vérifier que la matrice $(I - \omega L)^{-1}$ est triangulaire inférieure à diagonale unité et $(1 - \omega)I + \omega U$ est une matrice triangulaire supérieure dont la diagonale est identique à celle de la matrice $(1 - \omega)I$. D'où

$$\det \mathcal{L}_\omega = (1 - \omega)^n$$

Par ailleurs, pour tout $A \in \mathcal{M}_n(\mathbb{C})$ on a

$$|\det A| \leq (\rho(A))^n,$$

puisque le déterminant d'une matrice est égal au produit de ses valeurs propres comptées avec leur ordre de multiplicité. On en déduit que $\rho(\mathcal{L}_\omega) > 1$ pour $\omega \in \mathbb{R} \setminus [0, 2]$. ■

Théorème 3.5

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive décomposée en $A = M - N$ où M est inversible. Soit $B = M^{-1}N = I - M^{-1}A$. Supposons que la matrice (symétrique) $M^T + N$ est définie positive. Alors $\rho(B) < 1$ et donc, la méthode itérative définie par

$$x^{k+1} = Bx^k + c$$

converge.

Démonstration :

1^{ère} étape - Vérifions que $M^T + N$ est symétrique. On a

$$M^T + N = M^T + M - A$$

d'où

$$(M^T + N)^T = (M^T + M)^T - A^T = M + M^T - A = M^T + N$$

- Soit x un vecteur quelconque. Posons $y = Bx = x - M^{-1}Ax$ et montrons l'identité suivante :

$$(3.17) \quad (x, Ax) - (y, Ay) = ((x - y), (M + M^T - A)(x - y)).$$

On a

$$\begin{aligned} (y, Ay) &= (x - M^{-1}Ax, A(x - M^{-1}Ax)) \\ &= (x, Ax) - (x, AM^{-1}Ax) - (AM^{-1}Ax, x) + (M^{-1}Ax, AM^{-1}Ax) \end{aligned}$$

d'où

$$(3.18) \quad \begin{aligned} (x, Ax) - (y, Ay) &= (x, AM^{-1}Ax) + \\ &+ (AM^{-1}Ax, x) - (M^{-1}Ax, AM^{-1}Ax) \end{aligned}$$

comme $x - y = M^{-1}Ax$ on obtient

$$\begin{aligned} (3.19) \quad & (x - y, (M + M^T - A)(x - y)) = \\ &= (M^{-1}Ax, (M + M^T - A)M^{-1}Ax) \\ &= (M^{-1}Ax, Ax) + (M^{-1}Ax, M^T M^{-1}Ax) - (M^{-1}Ax, AM^{-1}Ax) \\ &= (AM^{-1}Ax, x) + (x, A^T M^{-T} M^T M^{-1}Ax) - (M^{-1}Ax, AM^{-1}Ax) \\ &= (AM^{-1}Ax, x) + (x, AM^{-1}Ax) - (M^{-1}Ax, AM^{-1}Ax) \end{aligned}$$

Soit grâce à (3.18)

$$((x - y), (M + M^T - A)(x - y)) = (x, Ax) - (y, Ay)$$

2^{ème} étape :

Supposons que A est définie positive. Soit $x \neq 0$ un vecteur propre associé à une valeur propre λ de B , alors

$$y = x - M^{-1}Ax = x - (I - B)x = Bx = \lambda x$$

et en utilisant (3.17), on obtient

$$\begin{aligned} (x, Ax)(1 - |\lambda|^2) &= (x, Ax) - (y, Ay) = ((1 - \lambda)x, (M + M^T - A)(1 - \lambda)x) \\ &= |1 - \lambda|^2 (x, (M + M^T - A)x) \end{aligned}$$

λ ne peut pas être égale à 1, car cela implique que $Bx = x$ donc $M^{-1}Ax = 0$, ce qui est impossible car A est inversible et $x \neq 0$. Comme A et $M^T + N = M + M^T - A$ sont définies positives par hypothèse, on en déduit que $1 - |\lambda|^2 > 0$ d'où $|\lambda| < 1$, cela étant vrai pour toute valeur propre λ , on a $\rho(B) < 1$ et la méthode converge. ■

Remarque

Le théorème 3.5 reste valable pour $A \in \mathcal{M}_n(\mathbf{C})$ à condition de supposer la matrice A hermitienne définie positive et la matrice hermitienne $M^* + N$ où M^* est l'adjointe de M , définie positive. La démonstration du théorème 3.5 reste la même à condition de remplacer M^T par M^* dans celle-ci. ■

Théorème 3.6

Soit A une matrice symétrique définie positive décomposée selon (3.5) et (3.6) sous la forme $A = D - E - F$. Alors la méthode de relaxation, de matrice

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

est convergente pour $\omega \in]0, 2[$.

Démonstration :

Soient $M = \frac{1}{\omega} D - E$ et $N = \frac{1 - \omega}{\omega} D + F$. La matrice D est définie positive car A l'est. Un calcul simple donne

$$M^T + N = \frac{2 - \omega}{\omega} D - E^T + F = \frac{2 - \omega}{\omega} D$$

puisque A est symétrique ($E^T = F$). Par conséquent $M^T + N$ est définie positive si et seulement si $\omega \in]0, 2[$. Donc en utilisant le théorème précédent, on conclut que

$$\rho(\mathcal{L}_\omega) < 1$$

et la méthode de relaxation converge. ■

Théorème 3.7

Soit A une matrice tridiagonale. On suppose que toutes les valeurs propres de la matrice de Jacobi J sont réelles et strictement inférieures à 1 en valeur absolue. Alors, la méthode de relaxation converge pour $\omega \in]0, 2[$. De plus, il existe une valeur optimale de ω notée ω_0 et donnée par

$$\omega_0 = \frac{2}{1 + \sqrt{1 - (\rho(J))^2}}$$

pour laquelle $\rho(\mathcal{L}_\omega)$ est minimal et $\rho(\mathcal{L}_\omega) = \omega_0 - 1$:

$$\forall \omega \in]0, 2[\quad \rho(\mathcal{L}_\omega) \geq \rho(\mathcal{L}_{\omega_0}) = \omega_0 - 1$$

Démonstration : voir P.G. Ciarlet [1990]. ■

3.6 Vitesse de convergence d'une méthode itérative

On peut caractériser la convergence d'une méthode itérative par la vitesse avec laquelle l'erreur tend vers 0. Pour une méthode itérative dont la matrice est B , l'erreur à l'itération k est donnée par

$$(3.20) \quad e^k \equiv x^k - x = B^k e^0$$

où x^k est le vecteur calculé à la $k^{\text{ième}}$ itération à l'aide de (3.3) et où x est la solution de (3.1). Pour une norme choisie, on a

$$\|e^k\| = \|B^k e^0\|$$

On va voir dans la suite de cette section qu'une méthode itérative est d'autant plus rapide à converger que le rayon spectral de sa matrice d'itération B est plus petit. Ceci découle du lemme suivant.

Lemme 3.3

Pour $B \in \mathcal{M}_n(\mathbf{C})$, on a pour toute norme matricielle subordonnée :

$$\lim_{k \rightarrow +\infty} \|B^k\|^{\frac{1}{k}} = \rho(B)$$

où $\rho(B)$ est le rayon spectral de la matrice B .

Démonstration :

- Soit $\lambda \in \mathbf{C}$ telle que $|\lambda| = \rho(B)$. Il existe alors $x \neq 0$ tel que $Bx = \lambda x$ et on en déduit que

$$\|B^k x\| = \|\lambda^k x\| = |\lambda|^k \|x\| = (\rho(B))^k \|x\|$$

et par conséquent

$$(3.21) \quad \|B^k\|^{\frac{1}{k}} \geq \left(\frac{\|B^k x\|}{\|x\|} \right)^{\frac{1}{k}} \geq \rho(B)$$

Réciproquement, pour $\varepsilon > 0$, on a

$$\rho\left(\frac{B}{\varepsilon + \rho(B)}\right) = \frac{\rho(B)}{\varepsilon + \rho(B)} < 1$$

d'où d'après le lemme 3.1, on a

$$\lim_{k \rightarrow +\infty} \left\| \left(\frac{B}{\varepsilon + \rho(B)} \right)^k \right\| = 0$$

Par conséquent, il existe $N_0 \in \mathbf{N}$ tel que pour $k \geq N_0$, on a

$$\left\| \left(\frac{B}{\varepsilon + \rho(B)} \right)^k \right\| < 1$$

d'où

$$\frac{1}{\varepsilon + \rho(B)} \|B^k\|^{\frac{1}{k}} = \left\| \left(\frac{B}{\varepsilon + \rho(B)} \right)^k \right\|^{\frac{1}{k}} < 1.$$

On en déduit, grâce à (3.21), que pour $k \geq N_0$:

$$\rho(B) \leq \|B^k\|^{\frac{1}{k}} < \varepsilon + \rho(B)$$

d'où finalement

$$\forall \varepsilon > 0 \quad \exists N_0 \in \mathbb{N} \quad k \geq N_0 \implies \left(\rho(B) \leq \|B^k\|^{\frac{1}{k}} < \varepsilon + \rho(B) \right)$$

Ce qui prouve que

$$\lim_{k \rightarrow +\infty} \|B^k\|^{\frac{1}{k}} = \rho(B)$$

■

Commentaire :

Une deuxième méthode itérative dont la matrice de l'itération $\tilde{B}$ a un rayon spectral $\rho(\tilde{B}) < 1$ et tel que $\rho(\tilde{B}) < \rho(B)$ converge plus rapidement que la méthode itérative de matrice B .