

C4, méthodes itératives

Inconvénients des méthodes directes :

1. Coût de la factorisation de la matrice A
2. Taille du stockage de la matrice factorisée (malgré la conservation du profil)

Dans une méthode itérative :

1. Pas de factorisation (uniquement des produits matrice-vecteur)
2. Stockage uniquement des termes non nuls de A

Solution “ultime” : mélange de méthodes directes et itératives (méthode itérative avec une factorisation incomplète)

Principe général

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$

on veut calculer x solution de $Ax = b$

Principe : Construire une suite $(x^{(k)})_{k \in \mathbb{N}}$ t.q. $\lim_{k \rightarrow +\infty} x^{(k)} = x$

Idée principale :

Choisir P et N t.q. $A = P - N$ (et P facile à “inverser”)

Pour $x^{(0)}$ donné la suite $(x^{(k)})_{k \in \mathbb{N}}$ est donnée par

$$Px^{(k+1)} = Nx^{(k)} + b$$

Remarques :

1. Si la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge, $\lim_{k \rightarrow +\infty} x^{(k)} = x$ ($Ax = b$)

2. $x^{(k+1)} = Bx^{(k)} + c$, $B = P^{-1}N$, $c = P^{-1}b$

En pratique, on ne calcule jamais B

3. $e^{(k)} = x^{(k)} - x$, $e^{(k+1)} = Be^{(k)}$

Théorème de convergence

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

Méthode itérative :

$$x^{(k+1)} = Bx^{(k)} + c, \quad B = P^{-1}N, \quad c = P^{-1}b, \quad A = P - N$$

Définition : La méthode est dite convergente si pour tout $x^{(0)} \in \mathbb{R}^n$, $\lim_{k \rightarrow +\infty} x^{(k)} = x$

Théorème :

La méthode est convergente si et seulement si $\rho(B) < 1$

Démonstration: On a vu (cours 3) que la suite $(e^{(k)})_{k \in \mathbb{N}}$ converge vers 0 (pour tout $e^{(0)}$) si et seulement si $\rho(B) < 1$

Vitesse de convergence

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

Méthode itérative :

$$x^{(k+1)} = Bx^{(k)} + c, \quad B = P^{-1}N, \quad c = P^{-1}b, \quad A = P - N, \quad \rho(B) < 1$$

Question : Quelle est la vitesse de convergence ?

Réponse : Pour la plupart des $x^{(0)}$, $\lim_{k \rightarrow \infty} \frac{\|e^{(k+1)}\|}{\|e^{(k)}\|} = \rho(B)$

Démonstration dans un cas simple :

B symétrique. $\{u_1, \dots, u_n\}$ base de $\mathbb{R}^n$, $Bu_i = \lambda_i u_i$.

On suppose $\rho(B) = \lambda_n$, $|\lambda_i| < \rho(B)$ pour $i < n$

$$e^{(0)} = \sum_{i=1}^n \alpha_i u_i, \quad e^{(k)} = \sum_{i=1}^n \alpha_i \lambda_i^k u_i = \rho(B)^k \sum_{i=1}^n \frac{\lambda_i^k}{\rho(B)^k} \alpha_i u_i$$

$$\frac{\|e^{(k+1)}\|}{\|e^{(k)}\|} = \rho(B) \frac{\|\alpha_n u_n + \sum_{i=1}^{n-1} \frac{\lambda_i^{k+1}}{\rho(B)^{k+1}} \alpha_i u_i\|}{\|\alpha_n u_n + \sum_{i=1}^{n-1} \frac{\lambda_i^k}{\rho(B)^k} \alpha_i u_i\|} \rightarrow \rho(B),$$

si $\alpha_n \neq 0$

Méthode de Richardson

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

Méthode de Richardson : $\alpha > 0$ donné

$$x^{(k+1)} = x^{(k)} + \alpha(b - Ax^{(k)}) \quad (P = \frac{1}{\alpha}I, N = P - A)$$

$$B = I - \alpha A = P^{-1}N = I - P^{-1}A$$

La méthode est convergente si et seulement si $\rho(I - \alpha A) < 1$

On retrouvera cette méthode dans la partie “algorithmes pour l’optimisation” du cours

Méthode de Richardson, exemple :

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$x^{(k+1)} = Bx^{(k)} + c, \quad B = (1 - \alpha A), \quad c = \alpha b$$

La méthode est convergente si et seulement si $\rho(B) < 1$

$$B = \begin{bmatrix} 1 - 2\alpha & \alpha \\ \alpha & 1 - 2\alpha \end{bmatrix},$$

$$\det(B - \lambda I) = \lambda^2 - 2\lambda(1 - 2\alpha) + 1 - 4\alpha + 3\alpha^2$$

$$\lambda = 1 - 2\alpha \pm \sqrt{\alpha^2}$$

Si $\alpha \leq 0$, $\rho(B) \geq 1$. La méthode est non convergente

Si $\alpha > 0$, $\rho(B) = \max\{|1 - 3\alpha|, |1 - \alpha|\}$

La méthode est convergente si et seulement si $0 < \alpha < 2/3$

La valeur de α qui minimise $\rho(B)$ est $1/2$

Méthode de Jacobi

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

On suppose $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$

$$A = D - E - F,$$

- ▶ D est diagonale, (partie diagonale de A)
- ▶ E est triangulaire inférieure avec 0 en diagonale (partie inférieure stricte de $-A$)
- ▶ F est triangulaire supérieure avec 0 en diagonale (partie supérieure stricte de $-A$)

Méthode de Jacobi : $Dx^{(k+1)} = (E + F)x^{(k)} + b$

La matrice des itérations est $B_J = D^{-1}(E + F)$

$$(L_i(B_J) = \frac{1}{a_{i,i}} L_i(E + F)$$

La méthode est convergente si et seulement si $\rho(B_J) < 1$

Méthode de Gauss-Seidel

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

On suppose $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$

$$A = D - E - F,$$

- ▶ D est diagonale, (partie diagonale de A)
- ▶ E est triangulaire inférieure avec 0 en diagonale (partie inférieure stricte de $-A$)
- ▶ F est triangulaire supérieure avec 0 en diagonale (partie supérieure stricte de $-A$)

Méthode de Gauss-Seidel : $(D - E)x^{(k+1)} = Fx^{(k)} + b$

Ecriture équivalente : $Dx^{(k+1)} = Ex^{(k+1)} + Fx^{(k)} + b$

La matrice des itérations est $B_{GS} = (D - E)^{-1}F$

La méthode est convergente si et seulement si $\rho(B_{GS}) < 1$

Le calcul des composantes de $x^{(k+1)}$, notées $x_i^{(k+1)}$, se fait dans l'ordre croissant $i = 1, \dots, n$

Jacobi, Gauss-Seidel, exemple

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$B_J = D^{-1}(E + F) = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}$$

$$\rho(B_J) = 1/2$$

$$B_{GS} = (D - E)^{-1}F = \begin{bmatrix} 1/2 & 0 \\ 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/2 \\ 0 & 1/4 \end{bmatrix}$$

$$\rho(B_{GS}) = 1/4$$

Donc, pour cet exemple, GS va converger (sauf cas particuliers)
plus vite que Jacobi

Question : ce résultat (GS meilleur que Jacobi, sauf cas
particuliers) est il général ?

Comparaison Jacobi/Gauss-Seidel

$A \in \mathcal{M}_n(\mathbb{R})$ inversible. On suppose

- ▶ $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$
- ▶ A est tridiagonale, i.e $a_{i,j} = 0$ si $|i - j| > 1$

Alors $\rho(B_{GS}) = \rho(B_J)^2$

Démonstration dans le td4

Conséquences

1. La méthode de Jacobi converge si et seulement si la méthode de Gauss-Seidel converge
2. La méthode de Gauss-Seidel converge plus vite (sauf cas particulier) que la méthode de Jacobi

Méthode de Gauss-Seidel inversée

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

On suppose $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$

$$A = D - E - F,$$

- ▶ D est diagonale, (partie diagonale de A)
- ▶ E est triangulaire inférieure avec 0 en diagonale (partie inférieure stricte de $-A$)
- ▶ F est triangulaire supérieure avec 0 en diagonale (partie supérieure stricte de $-A$)

Méthode de Gauss-Seidel inversée: $(D - F)x^{(k+1)} = Ex^{(k)} + b$

Ecriture équivalente : $Dx^{(k+1)} = Fx^{(k+1)} + Ex^{(k)} + b$

La matrice des itérations est $B_{GSR} = (D - F)^{-1}E$

La méthode est convergente si et seulement si $\rho(B_{GSR}) < 1$

Le calcul des composantes de $x^{(k+1)}$, notées $x_i^{(k+1)}$, se fait dans l'ordre décroissant $i = n, \dots, 1$

Méthode SOR

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

On suppose $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$

$$A = D - E - F,$$

Méthode SOR (Successive Over Relaxation) $0 < \omega < 2$

pour $i = 1, \dots, n$,

$$a_{i,i} \tilde{x}_i^{(k+1)} = \sum_{j < i} a_{i,j} x_j^{(k+1)} + \sum_{j > i} a_{i,j} x_j^{(k)} + b_i \text{ (comme GS)}$$

$$x_i^{(k+1)} = \omega \tilde{x}_i^{(k+1)} + (1 - \omega) x_i^{(k)}$$

Écriture équivalente :

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

La matrice des itérations est $B_\omega = (D - \omega E)^{-1}(\omega F + (1 - \omega)D)$

$\omega = 1$ donne GS. En pratique $\omega \geq 1$, le cas $\omega < 1$ est plutôt pour des problèmes non linéaires

Convergence pour SOR

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

On suppose $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$

$$A = D - E - F,$$

Méthode SOR (Successive Over Relaxation) $0 < \omega < 2$

pour $i = 1, \dots, n$,

$$a_{i,i} \tilde{x}_i^{(k+1)} = \sum_{j < i} a_{i,j} x_j^{(k+1)} + \sum_{j > i} a_{i,j} x_j^{(k)} + b_i \text{ (comme GS)}$$

$$x_i^{(k+1)} = \omega \tilde{x}_i^{(k+1)} + (1 - \omega) x_i^{(k)}$$

On va montrer :

1. CN. Si la méthode est convergente alors $0 < \omega < 2$
2. CS. Si A est s.d.p. et $0 < \omega < 2$, alors la méthode SOR est convergente

Condition nécessaire pour la convergence pour SOR

$A \in \mathcal{M}_n(\mathbb{R})$ inversible

On suppose $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$

La matrice des itérations est $B_\omega = (D - \omega E)^{-1}(\omega F + (1 - \omega)D)$

La méthode est convergente (si et) seulement si $\rho(B_\omega) < 1$

- ▶ D : partie diagonale de A
- ▶ E : partie inférieure stricte de $-A$
- ▶ F : partie supérieure stricte de $-A$

$$\det(B_\omega) = \frac{1}{\det(D)} (1 - \omega)^n \det(D) = (1 - \omega)^n$$

$$\rho(B_\omega) < 1 \implies |\det(B_\omega)| < 1 \implies 0 < \omega < 2$$

A s.d.p., CNS pour la convergence de SOR

$A \in \mathcal{M}_n(\mathbb{R})$ s.d.p. (en particulier $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$)

La matrice des itérations est $B_\omega = (D - \omega E)^{-1}(\omega F + (1 - \omega)D)$

La méthode est convergente si et seulement si $\rho(B_\omega) < 1$

On sait déjà que $0 < \omega < 2$ est une condition nécessaire

Est-elle suffisante ?

On suppose que $0 < \omega < 2$

On va montrer qu'il existe une norme sur $\mathbb{R}^n$, notée $\|\cdot\|_\star$ t.q. pour la norme induite sur $\mathcal{M}_n(\mathbb{R})$, encore notée $\|\cdot\|_\star$, $\|B_\omega\|_\star < 1$ (et donc $\rho(B_\omega) \leq \|B_\omega\|_\star < 1$)

Choix de la norme sur $\mathbb{R}^n$: $\|x\|_\star = (Ax \cdot x)^{1/2}$ (norme induite par un produit scalaire)

On va montrer que $\|B_\omega\|_\star < 1$

A s.d.p., CNS pour la convergence de SOR, fin

$A \in \mathcal{M}_n(\mathbb{R})$ s.d.p. (en particulier $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$)
 $0 < \omega < 2$

on veut montrer que $\|B_\omega\|_\star < 1$ ($\|x\|_\star = (Ax \cdot x)^{1/2}$)

La matrice des itérations est

$$B_\omega = (D - \omega E)^{-1}(\omega F + (1 - \omega)D) = (\frac{D}{\omega} - E)^{-1}(F + \frac{1-\omega}{\omega}D)$$

$$P = \frac{D}{\omega} - E, N = F + \frac{1-\omega}{\omega}D, B_\omega = P^{-1}N, A = P - N$$

Soit $x \in \mathbb{R}^n$ t.q. $\|x\|_\star = 1$, $\|B_\omega\|_\star = \|B_\omega x\|_\star$.

On pose $P^{-1}Ax = y$ (noter que $y \neq 0$ et $Ax = Py$)

$$\begin{aligned}\|B_\omega\|_\star &= \|B_\omega x\|_\star^2 = AB_\omega x \cdot B_\omega x = AP^{-1}Nx \cdot P^{-1}Nx \\ &= A(I - P^{-1}A)x \cdot (I - P^{-1}A)x = A(x - y) \cdot (x - y) \\ &= Ax \cdot x - 2Ax \cdot y + Ay \cdot y = \|x\|_\star^2 + (-2P + A)y \cdot y \\ &= 1 - (N + P)y \cdot y = 1 - (N + P^t)y \cdot y < 1,\end{aligned}$$

car $N + P^t = \frac{2-\omega}{\omega}D$ est s.d.p. et $y \neq 0$

Méthode SSOR

$A \in \mathcal{M}_n(\mathbb{R})$ inversible, $b \in \mathbb{R}^n$, x solution de $Ax = b$

On suppose $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$

$$A = D - E - F,$$

- ▶ D : partie diagonale de A
- ▶ E : partie inférieure stricte de $-A$
- ▶ F : partie supérieure stricte de $-A$

Méthode SSOR (Symmetric Successive Over Relaxation)

$$0 < \omega < 2$$

si k est pair, on utilise SOR, c'est-à-dire GS + relaxation

si k est impair, on remplace GS par GSR

La matrice des itérations est

$$B_{s\omega} = (D - \omega F)^{-1}(\omega E + (1 - \omega)D)(D - \omega E)^{-1}(\omega F + (1 - \omega)D)$$

SOR, recherche du ω optimal

$A \in \mathcal{M}_n(\mathbb{R})$ inversible. On suppose

- ▶ $a_{i,i} \neq 0$ pour tout $1 \leq i \leq n$
- ▶ A est tridiagonale, i.e $a_{i,j} = 0$ si $|i - j| > 1$
- ▶ $Sp(B_J) \subset \mathbb{R}$ et $\rho(B_J) < 1$

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

Alors $\rho(B_{\omega_0}) \leq \rho(B_\omega)$ pour tout $0 < \omega < 2$

De plus $\rho(B_{\omega_0}) = \omega_0 - 1$

Nous ne démontrons pas cette formule mais nous verrons un exemple en tp

Remarques complémentaires

1. Si A est s.d.p., GS converge (car c'est SOR avec $\omega = 1$) mais Jacobi ne converge pas toujours (voir td4)
2. Les mêmes méthodes existent dans une version “décomposition de la matrice par blocs” (voir le polycopié)
3. Les méthodes les plus performantes actuellement sont plutôt du type “gradient conjugué avec préconditionnement” (voir la partie “optimisation” du cours)