233 ANALYSE NUMÉRIQUE MATRICIELLE : RÉSOLUTION APPROCHÉE DE SYSTÈMES LINÉAIRES, RECHERCHE DE VECTEURS PROPRES, EXEMPLES.

Soit $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, $n \in \mathbb{N}$. Pour $A \in \mathcal{M}_n(\mathbb{K})$, on désigne par A^* la matrice $\overline{A^\top}$.

I. Préliminaires théoriques

I. A. Normes matricielles [AK02, ChI.3, p51]

Définition norme matricielle : norme sous-multiplicative sur $\mathcal{M}_n(\mathbb{K})$
Contre-exemple : $\|A\| = \max_{i,j} |a_{i\,j}|$
Définition norme subordonnée : dérive d'une autre norme
Une norme subordonnée est une norme matricielle
Exemples des normes p subordonnées : $|||A|||_1 = \sup_{1\leq j\leq n} \sum_{1\leq i\leq n} |a_{i\,j}|$ et $|||A|||_\infty = \sup_{1\leq i\leq n} \sum_{1\leq j\leq n} |a_{i\,j}|$
Exemple de norme matricielle non subordonnée : la norme de Frobenius : $\|A\|_F = \sqrt{\mathrm{Tr}(A^\top A)}$

I. B. Rayon spectral [AK02, ChI.3, p51]

Définition du rayon spectral. Lien avec la norme 2 subordonnée : $|||A|||_2 = |||A^*|||_2 = \sqrt{\rho(AA^*)}$
Proposition : toute norme matricielle vérifie $\rho(.) \leq \|.\|$. Réciproquement pour toute matrice A, on a $\rho(A) = \inf\{\|A\| \mid \|.\| \text{ est une norme subordonnée}\}$ (Householder)
Remarque : c'est faux pour une norme quelconque (considérer $\|A\| = \max_{i,j} |a_{i\,j}|$)
Équivalence entre $A^n \longrightarrow 0$, $A^n x \longrightarrow 0 \quad \forall x$, $\rho(A) < 1$ et il existe une norme subordonnée telle que $\|A\| < 1$
Application : lorsque $\rho(A) < 1$, $I_n - A$ est inversible, d'inverse $\sum_{i\geq 0} A^i$

I. C. Conditionnement [All05, §II.5.5, p89]

Soit A une matrice inversible.
Exemple de matrice A et vecteur b pour laquelle les résolution de $Ax = b$ et $Ax = b + \delta b$ donnent des solutions très différentes. (petite erreur sur les données implique grosse erreur sur la solution.
Contrôle de cette erreur par la notion de conditionnement
Définition du conditionnement par rapport à une norme subordonnée
Exemples de cond_p
Proposition : si $Ax = b$ et $A(x + \delta x) = b + \delta b$, alors $\frac{\|\delta x\|}{\|x\|} \leq \mathrm{cond}(A)\frac{\|\delta b\|}{\|b\|}$ et l'égalité est atteinte pour certains choix de x et δb. Si de plus $(A + \delta A)(x + \delta x) = b$, alors on a $\frac{\|\delta x\|}{\|x+\delta x\|} \leq \mathrm{cond}(A)\frac{\|\delta A\|}{\|A\|}$ et l'égalité est atteinte pour certains choix de δa et b
Propriétés du conditionnement : $\mathrm{cond}(A) = \mathrm{cond}(A^{-1}) = \mathrm{cond}(\alpha A)$ $(\alpha \neq 0)$
Pour la norme 2 : si U est unitaire, $\mathrm{cond}_2(AU) = \mathrm{cond}_2(UA) = \mathrm{cond}_2(A)$, et si A, B sont symétriques définies positives, on a $\mathrm{cond}_2(A + B) \leq \max(\mathrm{cond}_2(A), \mathrm{cond}_2(B))$. Liens avec les valeurs singulières ou même les valeurs propres dans le cas particulier où A est normale
Exemple : problème du laplacien

II. Résolutions de systèmes linéaires $Ax = b$

Soit A inversible.

II. A. Méthodes directes [AK02, PIII, p105]

Formules de Cramer. Dans la pratique le coût en $O(n!)$ est très mauvais ! On ne l'utilise donc jamais (sauf en dimensions 2 ou 3)
Remarque : pour une matrice triangulaire, le coût est en $n^2/2$ (voir algorithme en annexe), on se ramène donc à l'étude de systèmes linéaires !
Théorème d'élimination de Gauss : il existe T triangulaire supérieure telle que $T = M_1 \dots M_r A$ où les M_i sont des matrices de transvection ou de permutation.
Méthode : en pratique on ne calcule pas $M_1 \dots M_r$: on met à jour le second membre $b' = M_1 \dots M_r b$ à chaque étape puis on résout $Tx = b'$. On a une complexité en $n^3/3$ mais on peut récupérer $\det(A)$.

Théorème 1. [Factorisation LU]
Soit $A \in \mathcal{M}_n(\mathbb{K})$ une matrice dont tous les mineurs principaux sont non nuls. Alors il existe un unique couple (L, U) de matrices tel que L est triangulaire inférieure avec diagonale de 1, U est triangulaire supérieure et $A = LU$.

Théorème 2. [Factorisation de Cholesky]
Soit A une matrice symétrique réelle définie positive. Alors il existe une unique matrice B triangulaire inférieure telle que tous ses éléments diagonaux soient positifs et $A = BB^*$.

Théorème : factorisation LU, lorsque les mineurs principaux de A sont tous non nuls
Méthode : on résout $Ly = b$ puis $Ux = y$
Théorème : méthode de Cholesky : si A est symétrique réelle définie positive, il existe B triangulaire inférieure réelle à diagonale positive telle que $A = BB^\top$
Complexité en $n^3/6$, exemple

II. B. Méthodes itératives [AK02, PIV, p153]

Idée : décomposition régulière $A = M - N$ avec M facile à inverser
On définit $x_0 \in \mathbb{K}^n$ et $x_{n+1} = M^{-1}(Nx_n + b)$. Si cette suite converge, la limite est la solution de $Ax = b$
Théorème : la méthode itérative converge pour tout x_0 si et seulement si $\rho(M^{-1}N) < 1$. Dans

ce cas, la vitesse de convergence est géométrique.
Exemple : méthode de RICHARDSON
Pour une matrice A hermitienne définie positive, si M, N est une décomposition satisfaisante, on a $M^* + N$ est hermitienne et si elle est définie positive, alors $\rho(M^{-1}N) < 1$
Méthode de JACOBI : on prend pour D la diagonale de A. On a $M^{-1}N = I_n - D^{-1}A$ (on peut supposer D inversible quitte à permuter les colonnes de A
Si A est hermitienne, on a convergence si A et $2D - A$ sont définies positives
Méthode de GAUSS-SEIDEL : on prend pour M la matrice triangulaire inférieure issue de A, et pour N l'opposé des coefficients restants. Si A est hermitienne définie positive, on a la convergence de la méthode car $M^* + N$ l'est aussi
Remarque : les itérations correspondent à un calcul matriciel, on a donc n^2 opérations. Les inversions de M nécessitent n opérations pour JACOBI, $n^2/2$ pour GAUSS-SEIDEL, donc pour un nombre d'itérations $\ll n$, ces méthodes sont plus rapides que les méthodes directes

II. C. Minimisation de fonctionnelle [FGN12, §1.21, p39–41]

Minimisation de $x \longmapsto \frac{1}{2}\langle Ax \mid x\rangle - \langle b \mid x\rangle$

THÉORÈME 3. [MÉTHODE DE GRADIENT À PAS OPTIMAL]
Soit $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ elliptique, c'est-à-dire f est $\mathcal{C}^1$ et telle qu'il existe $\alpha > 0$ satisfaisant

$$\forall x, y \in \mathbb{R}^p, \langle \nabla f(x) - \nabla f(y) \mid x - y\rangle \geq \alpha \|x - y\|^2$$

Alors la suite $(x_n)_{n\in\mathbb{N}}$ définie par $x_0 \in \mathbb{R}^p$ et $x_{n+1} = x_n - \rho_n \nabla f(x_n)$ où $\rho_n = \operatorname{argmin}_{\rho>0} f(x_n - \rho\nabla f(x_n))$ converge vers l'unique minimum global de f.

III. Recherche de valeurs propres

III. A. Localisation [FGN07, §2.8, p80]

Proposition : matrice à diagonale dominante
Disques de GERSHGÖRIN.

III. B. Méthode de la puissance [AK02, Ch15, p215]

Méthode de la puissance pour trouver la plus grande valeur propre en module, ainsi qu'un vecteur propre, vitesse géométrique

III. C. Méthode de JACOBI [AK02, Ch15, p215]

Matrices de GIVENS
Méthode de JACOBI : $A_k \underset{k\to+\infty}{\longrightarrow} \operatorname{diag}(\lambda_{\sigma(i)})$ pour un $\sigma \in \mathfrak{S}_n$

ANNEXE

Quelques algorithmes vus dans la leçon

SPEECH

On s'intéresse à deux problèmes : résoudre l'équation $Ax = b$ et rechercher des valeurs propres de matrices données.

COMMENTAIRES

Si [AK02] n'est pas disponible, la dernière annexe de [All05] peut permettre de s'en sortir quand même, mais reste beaucoup moins détaillée.

QUESTIONS

Q Montrer que $\operatorname{cond}_2(A + B) \leq \max(\operatorname{cond}_2(A), \operatorname{cond}_2(B))$ pour $A, B \in S_n^{++}(\mathbb{R})$.

R Comme $A + B \in S_n^{++}(\mathbb{R})$, écrivons $\operatorname{Sp}(A) = \{\lambda_1 \leq \cdots \leq \lambda_n\}$, $\operatorname{Sp}(B) = \{\mu_1 \leq \cdots \leq \mu_n\}$ et $\operatorname{Sp}(A + B) = \{\gamma_1 \leq \cdots \leq \gamma_n\}$. Alors $\|A\|_2 = \lambda_n$ et $\|A^{-1}\|_2 = 1/\lambda_1$. Donc

$$\operatorname{cond}_2(A) = \lambda_n/\lambda_1 \qquad \operatorname{cond}_2(B) = \mu_n/\mu_1 \qquad \operatorname{cond}_2(A + B) = \gamma_n/\gamma_1$$

Or il est clair que $\gamma_n \leq \lambda_n + \mu_n$ puisque si x est un vecteur propre associé à γ_n, on a $\|(A + B)x\| \leq (\lambda_n + \mu_n)\|x\|$.
De plus $\gamma_1 \geq \lambda_1 + \mu_1$ puisque l'on a :

$$\gamma_1 = \inf_{\|x\|=1} \langle x \mid (A + B)x\rangle \geq \inf_{\|x\|=1} \langle x \mid Ax\rangle + \inf_{\|x\|=1} \langle x \mid Bx\rangle = \lambda_1 + \mu_1$$

On en déduit le résultat.

Q Soit $A \in \mathcal{M}_n(\mathbb{R})$ telle que $A_{ij} = 0$ lorsque $|i - j| \geq p$ (A est une matrice bande). Soit L, U sa décomposition. Montrer que L, U sont des matrices bandes. [AK02, p119]

BIBLIOGRAPHIE

[AK02] G. ALLAIRE et S.-M. KABER : *Algèbre linéaire numérique*. Ellipses, 2002.

[All05] G. ALLAIRE : *Analyse numérique et optimisation*. Les éditions de l'École Polytechnique, 2005.

[FGN07] S. FRANCINOU, H. GIANELLA et S. NICOLAS : *Oraux X-ENS - Algèbre 2*. Cassini, 2007.

[FGN12] S. FRANCINOU, H. GIANELLA et S. NICOLAS : *Oraux X-ENS - Analyse 4*. Cassini, 2012.