233 Analyse numérique matricielle : résolution approchée de systèmes linéaires, recherche de vecteurs propres

$\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, $A \in \mathcal{M}_n(\mathbb{K})$, $b \in \mathbb{K}^n$ et $n \in \mathbb{N}^*$

I. Normes matricielles et rayon spectral.

[A-K] p 52

Def ① : Soit $\|.\|$ une norme sur $\mathcal{M}_n(\mathbb{R})$. On dit que c'est une norme matricielle si elle vérifie $\forall A, B \in \mathcal{M}_n(\mathbb{K})$ $\|AB\| \leq \|A\| \|B\|$.

Ex ② : • La norme de Frobénius : $\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{1/2}$

• $\|A\| = \max_{1 \leq i,j \leq n} |a_{ij}|$ n'est pas matricielle

Def ③ : Soit $\|.\|$ une norme sur $\mathbb{K}^n$. On lui associe une norme matricielle, dite subordonnée définie par :

$$\|A\| = \sup_{x \in \mathbb{K}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}.$$

Rq ④ : $\|.\|_F$ n'est pas subordonnée car $\|I_n\|_F = \sqrt{n} \neq 1$.

prop ⑤ : • soit $\|.\|_1$ la norme ℓ^1 sur $\mathbb{K}^n$, alors :

$$\|A\|_1 = \sup_{1 \leq j \leq n} \left(\sum_{i=1}^{n} |a_{ij}|\right)$$

• soit $\|.\|_\infty$ la norme ℓ^∞ sur $\mathbb{K}^n$, alors : $\|A\|_\infty = \sup_{1 \leq i \leq n} \left(\sum_{j=1}^{n} |a_{ij}|\right)$

Def ⑥ : le rayon spectral de $A \in \mathcal{M}_n(\mathbb{K})$ est $\rho(A) = \max_{\lambda \text{ v.p de } A} |\lambda|$.

prop ⑦ : soit $\|.\|_2$ la norme subordonnée à la norme euclidienne dans $\mathbb{K}^n$, on a : $\|A\|_2 = \|A^*\| = \rho(AA^*)^{1/2}$ où $A^* = {}^t\overline{A}$.

p 57

prop ⑧ : Soit $\|.\|$ une norme matricielle sur $\mathcal{M}_n(\mathbb{C})$, on a $\rho(A) \leq \|A\|$

prop ⑨ : $\forall A \in \mathcal{M}_n(\mathbb{C})$, $\forall \varepsilon > 0$, $\exists \|.\|_{A,\varepsilon}$ norme subordonnée telle que $\|A\| \leq \rho(A) + \varepsilon$.

prop ⑩ : Soit $A \in \mathcal{M}_n(\mathbb{C})$. on a équivalence entre :

(i) $\lim_{i \to \infty} A^i = 0$

(ii) $\lim_{i \to \infty} A^i x = 0$

(iii) $\rho(A) < 1$

(iv) $\exists$ une norme matricielle tq $\|A\| < 1$

II. Conditionnement

But : résoudre $Ax = b$ de manière rapides (en temps, espace) importance de la précision.

p 88

Ex ⑪ : $\begin{pmatrix} 8 & 6 & 4 & 1 \\ 1 & 4 & 5 & 1 \\ 8 & 4 & 1 & 1 \\ 1 & 4 & 3 & 6 \end{pmatrix} x = \begin{pmatrix} 19 \\ 11 \\ 14 \\ 14 \end{pmatrix} \Leftrightarrow x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ mais :

$Ax = \begin{pmatrix} 19{,}01 \\ 11{,}05 \\ 14{,}07 \\ 14{,}05 \end{pmatrix} \Rightarrow x = \begin{pmatrix} -2{,}34 \\ 9{,}745 \\ -4{,}85 \\ -1{,}34 \end{pmatrix}$ (pbm d'approximation). de petites erreurs entrainent de grosses erreurs de solutions

Def ⑫ : soit $\|.\|$ norme subordonnée, on appelle conditionnement de $A \in \mathcal{M}_n(\mathbb{C})$ relatif à $\|.\|$ la quantité $\mathrm{cond}(A) = \|A\| \|A^{-1}\| \geq 1$

prop ⑬ : Soit A inversible et $b \neq 0$.

(i) si x et $x + \delta x$ les solutions de $Ax = b$ et $A(x + \delta x) = b + \delta b$

on a : $\dfrac{\|\delta x\|}{\|x\|} \leq \mathrm{cond}(A) \dfrac{\|\delta b\|}{\|b\|}$

(ii) si x et $x + \delta x$ les solutions de $Ax = b$ et $(A + \delta A)(x + \delta x) = b$

p 90

on a : $\dfrac{\|\delta x\|}{\|x + \delta x\|} \leq \mathrm{cond}(A) \dfrac{\|\delta A\|}{\|A\|}$.

on note $\mathrm{cond}_p(A) = \|A\|_p \|A^{-1}\|_p$.

prop ⑭ : 1) $\mathrm{cond}(A) = \mathrm{cond}(A^{-1}) = \mathrm{cond}(\alpha A)$ $\forall \alpha \neq 0$,

2) si $AA^* = A^*A$, $\mathrm{cond}_2(A) = \dfrac{\rho(A)}{\lambda_{min}(A)} = \rho(A)\rho(A^{-1})$

si A est symétrique définie positive alors $\mathrm{cond}_2(A) = \dfrac{\max_{\|x\|=1} \langle Ax, x\rangle}{\min_{\|x\|=1} \langle Ax, x\rangle}$

3) $\forall U \in \mathcal{U}_n(\mathbb{C})$, $\mathrm{cond}_2(AU) = \mathrm{cond}_2(UA) = \mathrm{cond}(A)$.

III. Résolution de $Ax = b$, Méthodes directes

p 77

prop ⑮ (Formule de Cramer) soit $A = (a_1 | \dots | a_n)$ inversible de colonnes $a_i \in \mathbb{R}^n$, alors $Ax = b \Leftrightarrow x_i = \dfrac{\det(a_1 | \dots | b | a_{i+1} | \dots)}{\det(A)}$

Rq (16): le nombre de multiplications pour calculer le déterminant est ici $c_n = n(1+c_{n-1}) \geq n\,c_{n-1}$ d'où $c_n \geq n!$
pour $n=50$, un calcul par cette méthode prendrait $\simeq 4{,}8 \cdot 10^{49}$ années.
• si A est triangulaire supérieure, le calcul est effectué en n^2 opérations.

Thm (17) (Élimination de Gauss) : Soit A une matrice (inversible ou non). Il existe au moins une matrice inversible M
p102 tq $T = MA$ soit triangulaire supérieure où M est un produit de matrices de transvections et permutations

Thm (18) (Factorisation LU) Soit $A = (a_{ij})_{ij}$ une matrice dont toutes les sous-matrices diagonales $A_k = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}$ sont inversibles. Il existe un unique couple
p113 de matrices (L, U) avec U triangulaire supérieure, et L triangulaire inférieure avec des 1 sur la diagonale

Rq (19) : • si $A = LU$, toutes ses sous-matrices diagonales sont inversibles
• si A est définie positive, elle vérifie les conditions du Thm.
• La méthode LU s'effectue en $O(n^3)$ opérations.

Thm (20) : (Factorisation QR) : soit $A \in GL_n(\mathbb{R})$. Il existe un unique
p125 (Q, R) tel que $Q \in O_n(\mathbb{R})$ et R triangulaire supérieure dont tous les éléments diagonaux sont positifs tel que $A = QR$. $(Ax = b \iff Rx = {}^tQb)$

Rq (21) : • le nbr d'opérations est $O(n^3)$ (dû à Gram-Schmidt)
• Cette méthode est assez instable numériquement.

IV. Méthodes itératives

$A \in GL_n(\mathbb{R})$ $b \in \mathbb{R}^n$

Def (22) : On appelle décomposition régulière de A un couple (M, N) avec M inversible, et "facile" à inverser tel que $A = M - N$. Une méthode itérative basée sur la décomposition régulière (M, N) est définie par :

$$\left\{ \begin{array}{l} x_0 \in \mathbb{R}^n \\ M x_{k+1} = N x_k + b \quad \forall k \geq 1. \end{array} \right. \quad * \qquad (\text{si } x_k \xrightarrow[k \to \infty]{} x \text{, on a } Ax = b)$$

Def (23) : Une méthode itérative est convergente si $\forall x_0 \in \mathbb{R}^n$ x_k converge vers la solution x.

Thm (24) : La méthode itérative $*$ converge SSI $\rho(M^{-1}N) < 1$.

Thm (25) : Si A est symétrique définie positive, de décomposition régulière $A = M - N$. Si $M^* + N$ est définie positive, alors $\rho(M^{-1}N) < 1$.

1) Lien avec l'optimisation

Thm (26) : La fonctionnelle quadratique $J : x \mapsto \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ où A est symétrique définie positive est coercive, et strictement convexe donc admet un unique minimum global en $\bar{x}$. Or $J \in \mathcal{C}^\infty(\mathbb{R}^n)$ et $\nabla J(\bar{x}) = 0 \iff A\bar{x} - b = 0$.

La méthode du gradient à pas fixe définie par

$$\left\{ \begin{array}{l} x_0 \in \mathbb{R}^n \\ x_{k+1} = x_k - \rho \nabla J(x_k) \quad \forall k \geq 0 \end{array} \right. \quad \left(M = \frac{I_n}{\rho},\ N = \frac{I_n}{\rho} - A\right) \text{ converge}$$

[C] p 193

p 159

Lorsque $p \in]0, \frac{\lambda_1}{\rho(A)}[$, le p optimal est $p_{opt} = \frac{2}{\lambda_1 + \lambda_n}$ où $0 < \lambda_1 \leq \dots \leq \lambda_n$ sont les vp de A.

prop 27 : (Méthode de Jacobi) : décomposition $M = D$ et $N = D - A$ où $D = \text{diag}(a_{11}, \dots, a_{nn})$ où l'on suppose $D \in GL_n(\mathbb{R})$. Si A est symétrique, alors la méthode de Jacobi converge si A et $2D - A$ sont définies positives.

prop 28 : (Gauss-Seidel) : on décompose $A = D - E - F$ où D est la partie diagonale, $-E$ la partie inférieure et $-F$ supérieure.

$A =$ 

on pose $M = \underbrace{D - E}_{\text{triangulaire}}$ et $N = F$.

si A est symétrique définie positive alors $M^* + N = D$ est symétrique définie positive, la méthode converge.

Ex 29 : Une discrétisation du problème $-u''(x) = f(x)$ conduit un système $Ax = b$ où $A = n^2 \begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix} \in \mathcal{M}_{n-1}(\mathbb{R})$

A_n est inversible est ses v.p sont $\lambda_k = 4n^2 \sin^2\left(k\frac{\pi}{2n}\right)$ $k \in \{1 \dots n-1\}$.

Jacobi converge car $\rho(I_n - D^{-1}A) = \cos\left(\frac{\pi}{n}\right) < 1$

Gauss-Seidel aussi car $\rho(M^{-1}N) = \cos^2\left(\frac{\pi}{n}\right) < 1$.

V. Calculs de vecteurs propres et valeurs propres

p 176

Thm 30 (méthode de la puissance) : soit $A \in \mathcal{M}_n(\mathbb{R})$ diagonalisable sur $\mathbb{R}$, de v.p $(\lambda_1 \dots \lambda_n)$ et de $\vec{v.p}$ associés $(e_1 \dots e_n)$. On suppose que $|\lambda_1| \leq \dots \leq |\lambda_{n-1}| < \lambda_n$.

On note $x_0 = \sum_{i=1}^{n} \beta_i e_i$ avec $\beta_n \neq 0$, $\|x_0\| = 1$. On définit $x_{k+1} = \frac{Ax_k}{\|Ax_k\|}$ $\forall k \geq 0$. Alors on a : $\lim_{k \to +\infty} \|y_k\| = \lambda_n$ où $y_k = Ax_{k-1}$ $\forall k \geq 1$ et $\lim_{k \to +\infty} x_k = x_\infty = \pm e_n$.

De plus $\left| \|y_k\| - \lambda_n \right| \leq C \left| \frac{\lambda_{n-1}}{\lambda_n} \right|^k$, $\|x_k - x_\infty\| \leq C \left| \frac{\lambda_{n-1}}{\lambda_n} \right|^k$

Références:

Allaire-Kaber (Algèbre liné numérique) 99%

Carlet (pour le pas fixe [26])