CS
CM 3.1

Résolution de systèmes linéaires : sensibilité et erreurs.

I/ Motivations

A - Exemples

ex1 : Soit le syst lin.

$$\underbrace{\begin{bmatrix} 0.78 & 0.563 \\ & 0.689 \end{bmatrix}}_{A} x = \begin{bmatrix} 0{,}07 \\ 0.258 \end{bmatrix}.$$

On a à notre disposition 2 algo, fournissant les solutions

$$x_1 = \begin{bmatrix} -20.568 \\ 28.881 \end{bmatrix} \qquad x_2 = \begin{bmatrix} 0.999 \\ -1 \end{bmatrix}$$

quelle s° choisir ?

1) On compare les résidus

$$r_1 = b - Ax_1 = \begin{bmatrix} 37 \\ 5 \end{bmatrix} \cdot 10^{-6}$$

$$r_2 = b - Ax_2 = \begin{bmatrix} 78 \\ 91 \end{bmatrix} \cdot 10^{-5}$$

$|r_1| \leq |r_2|$: x_1 semble être la "meilleure" solution.

2) $x^* = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ est la sol° exacte

$\Rightarrow x_2$ est la plus proche de x^*

Pb : Comment quantifier l'erreur d'une solution numérique ne connaissant pas la solution théorique ?

ex2 : Ayant la solution de $Ax = b$ (théorique ou par un algo), quelles seront l'amplitude des changements de la solution à des perturbations de A et/ou b ?

Avec l'exemple 1, des perturbations des entrées de la matrice de l'ordre de 10^{-6} peuvent rendre la matrice non inversible.

B - Représentation normalisée des nombres en virgule flottante.

Rappel : $x \in \mathbb{R}$ $x = (-1)^{\circ} m b^{e}$ b : base
e : exposant $m = a_1 + \frac{a_2}{b} + \dots + \frac{a_p}{b^{p-1}} \in [1, b[$ mantisse

avec $a_1 \in [1, b-1]$
$a_i \in [0, b-1] \ \forall i \neq 1$.

Tous les nombres ne sont pas exactement représentables en machine

$\Longrightarrow$ Erreurs d'arrondis

Définition : Précision machine ε

On appelle précision machine, notée ε, le plus petit nombre positif représentable en machine tq $1 + \varepsilon > 1$.

On a $\varepsilon = \frac{1}{b^{p-1}}$.

Application : • Erreur relative, critères d'arrêt :

$x \in \mathbb{R}_+^*$.

Soit y le nombre positif représentable en machine juste au dessus de x $\frac{|y - x|}{|x|} \geqslant \frac{\varepsilon}{b}$

• Optimisation :

$\frac{\| x_{k+1} - x_k \|}{\| x_k \|} \leqslant tol$ $\Longrightarrow$ Choix de tol ?

II / Sensibilité de la solution aux perturbations des données du système.

On cherche à résoudre $Ax = b$. En pratique, les données A et b du système sont incertaines : issues de mesures entachées d'erreur(s).

Autrement dit, en pratique on résoud plutôt $(A + \Delta A)(x + \Delta x) = b + \Delta b$

Q ? : $\frac{\| \Delta x \|}{x}$?

CS
CH3.2

A - Conditionnement d'une matrice

Définition : Soit $A \in \mathcal{M}_n(\mathbb{K})$
Soit $\|.\|$ une norme sur $\mathbb{K}^n$
On note également $\|.\|$ sa norme subordonnée sur $\mathcal{M}_n(\mathbb{K})$.

On appelle conditionnement de A, noté $K(A)$, la quantité $K(A) = \|A\| \|A^{-1}\|$

Propriété : Soit $A \in \mathcal{M}_n(\mathbb{K})$ inversible

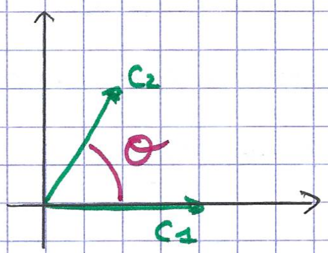
- i) $\forall \alpha \in \mathbb{R}^* \quad K(\alpha A) = K(A)$
- ii) $K(A) \geq 1$
- iii) $K_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$

avec $\sigma_1 \geq \dots \geq \sigma_n > 0$ les valeurs singulières de A.

- iv) $K_2(A) = 1 \iff A = \alpha V$ avec $\alpha \neq 0$ V orthogonale.

Illustration : $A = \begin{bmatrix} 1 & \cos\theta \\ 0 & \sin\theta \end{bmatrix}$ $\quad C_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\quad C_2 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$

On a $K_2(A) = \frac{1}{|\sin\theta|}$



et • $\theta = \frac{\pi}{2}$ $K_2(A) = 1$
si les col. de A st $\perp$

• $\theta \ll 1$: $K_2(A) \sim \frac{1}{|\theta|} \xrightarrow[\theta \to 0]{} +\infty$ les col. tendent à être colinéaires

$A = \begin{bmatrix} \varepsilon & 0 \\ 0 & 1 \end{bmatrix}$ avec $0 < \varepsilon \ll 1$ $\quad K_2(A) = \frac{1}{\varepsilon} \xrightarrow[\varepsilon \to 0]{} +\infty$ "Facteur d'échelle".

Remarque :
- On dira que le conditionnement est bon (resp. mauvais) si $K(A)$ est petit (resp. grand)
- La valeur de $K(A)$ dépend de la norme

relative.

Définition : Distance relative à la singularité

Soit $A \in \mathcal{M}_n(K)$ $A \neq 0$

Soit $\|.\|$ une norme sur K^n, on note $\|.\|$ sa norme subordonnée sur $\mathcal{M}_n(K)$

$\Rightarrow$ On appelle distance relative à la singularité, notée $dist_S(A)$, la quantité :

$$dist_S(A) = \min\left\{\frac{\|\Delta A\|}{\|A\|} \text{ tq } A+\Delta A \text{ non inversible}\right\}$$

$$dist_S(A) \in [0,1]$$

Théorème : Kation, Gastinel

Soit $A \in \mathcal{M}_n(\mathbb{C})$
Soit $\|.\|$ une norme sur $\mathbb{C}^N$, on note également $\|.\|$ sa norme subordonnée sur $\mathcal{M}_n(\mathbb{C})$

On a $dist_S(A) = (\|A\|\|A^{-1}\|)^{-1} = K(A)^{-1} = \dfrac{1}{K(A)}$

Remarque : Le conditionnement fournit des informations sur l'amplitude des perturbations à apporter à une matrice pour la rendre non inversible.

B- Sensibilité de la solution du système aux perturbations de A et/ou b

Propriété : Soit $A \in \mathcal{M}_n(K)$ inversible
Soit $b \in K^n$
Soit x la solution de $Ax = b$

- i) Perturbation du second membre :

  Soit $y = x + \Delta x$ la s° de $Ay = b + \Delta b$

  On a $\dfrac{\|\Delta x\|}{\|x\|} \leq K(A) \dfrac{\|\Delta b\|}{\|b\|}$

- ii) Perturbation de la matrice du système

  Soit $y = x + \Delta x$ la s° de $(A+\Delta A)y = b$

  On a $\dfrac{\|\Delta x\|}{\|y\|} \leq K(A) \dfrac{\|\Delta A\|}{\|A\|}$

CS
CH 3.3

Si de plus, $k(A)\frac{\|\Delta A\|}{\|A\|} < 1$, alors $\frac{\|\Delta x\|}{\|x\|} \leq \frac{k(A)\frac{\|\Delta A\|}{\|A\|}}{1-k(A)\frac{\|\Delta A\|}{\|A\|}}$

preuve : Soit x tq $Ax = b$

i) $A(x+\Delta x) = b+\Delta b$ or $Ax = b \Rightarrow A\Delta x = \Delta b$

$\Rightarrow \Delta x = A^{-1}\Delta b$

D'où $\|\Delta x\| \leq \|A^{-1}\| \, \|\Delta b\|$

$\Rightarrow \frac{\|\Delta x\|}{\|x\|} \leq \|A^{-1}\| \frac{\|\Delta b\|}{\|x\|} \leq k(A)\frac{\|\Delta b\|}{\|A\|\,\|x\|}$

Or $\|Ax\| = \|b\|$ $(Ax = b)$

$\Rightarrow \|b\| \leq \|A\|\,\|x\|$

D'où $\frac{\|\Delta x\|}{\|x\|} \leq k(A)\frac{\|\Delta b\|}{\|b\|}$

ii) $(A+\Delta A)(x+\Delta x) = b$

$Ax = b \Rightarrow A\Delta x + \Delta A(x+\Delta x) = 0$

$\Rightarrow \Delta x = -A^{-1}\Delta A(x+\Delta x)$

$\Rightarrow \frac{\|\Delta x\|}{\|x+\Delta x\|} \leq \|A^{-1}\|\,\|\Delta A\|$

$\Rightarrow \frac{\|\Delta x\|}{\|x+\Delta x\|} \leq k(A)\frac{\|\Delta A\|}{\|A\|}$

De plus si $k(A)\frac{\|\Delta A\|}{\|A\|} < 1$ $\quad \|\Delta x\| \leq k(A)\frac{\|\Delta A\|}{\|A\|}(\|x\|+\|\Delta x\|)$

D'où $\left(1-k(A)\frac{\|\Delta A\|}{\|A\|}\right)\|\Delta x\| \leq k(A)\frac{\|\Delta A\|}{\|A\|}\|x\|$

et $k(A)\frac{\|\Delta A\|}{\|A\|} < 1 \Rightarrow \frac{\|\Delta x\|}{\|x\|} \leq \frac{k(A)\frac{\|\Delta A\|}{\|A\|}}{1-k(A)\frac{\|\Delta A\|}{\|A\|}}$

Remarque : $k(A)\frac{\|\Delta A\|}{\|A\|} < 1 \iff \frac{\|\Delta A\|}{\|A\|} < k(A)^{-1} = \text{dist}_S(A)$

On a ainsi que $A+\Delta A$ inversible.

III / Erreurs directes et inverse d'un algorithme pour la résolution de systèmes linéaires

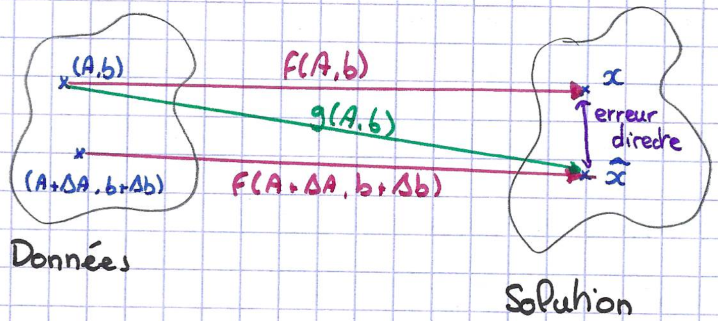
A - Principes

Soit $(A,b) \in \mathcal{M}_n(\mathbb{R}) \times \mathbb{R}^n$, le couple de données du système.

Soit f la fonction "théorique" (précision exacte) qui au couple (A,b) renvoie x la solution du système : $f(A,b) = x$ avec x tq $Ax = b$.

En pratique, on a un algorithme g implantant f et travaillant en précision finie.

Appliquer au couple (A,b), on obtient la solution $\hat{x}$ tq $g(A,b) = \hat{x}$



Idées : On considère que $\hat{x}$ est la solution exacte, obtenue par f, du système perturbé.

$$\hat{x} = g(A,b) = f(A+\Delta A, b+\Delta b)$$

Q? : 1) Peut-on quantifier $\|\Delta A\|$ et $\|\Delta b\|$?

2) Ces perturbations sont-elles acceptables ?

B - Erreur inverse ("Backward Error")

Théorème : R/Migal et Gaches

Soient $(A,b) \in \mathcal{M}_n(\mathbb{R}) \times \mathbb{R}^n$ les données du système
Soit $\hat{x}$ la solution obtenue par un algorithme numérique.

On définit l'erreur inverse par :

$$B_{er} = \min \left\{ \begin{array}{l} \varepsilon > 0 \text{ tq } \exists (\Delta A, \Delta b) \in \mathcal{M}_n(\mathbb{R}) \times \mathbb{R}^n \text{ tq} \\ \|\Delta A\| \le \varepsilon \|A\|_2 ,\ \|\Delta b\|_2 \le \varepsilon \|b\|_2 \\ \text{et } (A + \Delta A)\hat{x} = b + \Delta b \end{array} \right\}$$

On a $B_{er} = \dfrac{\|A\hat{x} - b\|_2}{\|A\|_2 \|x\|_2 + \|b\|_2}$

preuve : 1) On montre que $Ber \geq \frac{\|A\hat{x} - b\|_2}{\|A\|_2 \|x\|_2 + \|b\|_2}$

2) On vérifie que ε_0 est atteinte.

$$\Delta A_0 = \frac{-\|A\|_2}{\|A\|_2 \|\hat{x}\|_2 + \|b\|_2} (A\hat{x} - b)\hat{x}^T$$

$$\Delta b_0 = \frac{\|b\|_2}{\|A\|_2 \|\hat{x}\|_2 + \|b\|_2} (A\hat{x} - b)$$

Définition : Stabilité pour l'erreur inverse

Un algorithme est stable pour l'erreur inverse si pour toutes données (A, b) il produit une solution avec une erreur inverse "faible".

C- Majoration de l'erreur directe

Propriété : Soient (A, b) les données du système
Soit $\hat{x}$ une solution obtenue par un algo. $\hat{x} = x + \Delta x$ tq
$(A + \Delta A)(x + \Delta x) = b + \Delta b$

avec $\|\Delta A\| \leq Ber \|A\|$
$\|\Delta b\| \leq Ber \|b\|$

Si $Ber\, K(A) < 1$

alors $$\frac{\|\Delta x\|}{\|x\|} \leq \frac{Ber\, K(A)}{1 - Ber\, K(A)} \left(1 + \frac{\|b\|}{\|A\| \|x\|}\right)$$
$$\leq 2 \frac{Ber\, K(A)}{1 - Ber\, K(A)}$$

preuve : $(A + \Delta A)(x + \Delta x) = b + \Delta b$ or $Ax = b$

D'où $A \Delta x + \Delta A (x + \Delta x) = \Delta b$

$\Rightarrow \Delta x = -A^{-1}(\Delta A (x + \Delta x) - \Delta b)$.

et $\|\Delta x\| \leq \|A^{-1}\| (\|\Delta b\| + \underbrace{\|\Delta A\|}_{\leq Ber \|A\|} (\|x\| + \|\Delta x\|))$

$$(1 - K(A) Ber) \|\Delta x\| \leq Ber\, K(A) \left(\frac{\|b\|}{\|A\|} + \|x\|\right)$$

Or $K(A) Ber < 1 \Rightarrow \frac{\|\Delta x\|}{\|x\|} \leq \frac{K(A) Ber}{1 - K(A) Ber} \left(\frac{\|b\|}{\|A\| \|x\|} + 1\right)$

$$\leq 2 \frac{K(A) Ber}{1 - K(A) Ber}.$$

Remarque : i) $\kappa(A)\,\mathrm{Ber} < 1$: les perturbations de A associées à la résolution de $Ax = b$ par l'algo ne rendent pas $A + \Delta A$ inv.

ii) On dit par abus de langage

$$\text{erreur directe} \leq \kappa(A) \times \text{erreur inverse}.$$