Analyse Numériques SV

Systèmes Linéaires

Simone Deparis

EPFL Lausanne – MATH

Printemps 2020

Formulation du problème

On appelle système linéaire d'ordre n (n entier positif), une expression de la forme

$$A\mathbf{x} = \mathbf{b},$$

où $A = (a_{ij})$ est une matrice de taille $n \times n$ donnée, $\mathbf{b} = (b_j)$ est un vecteur colonne également donné et $\mathbf{x} = (x_j)$ est le vecteur des inconnues du système. La relation précédente équivaut aux n équations

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \ \ i = 1, \ldots, n.$$

La matrice A est dite *régulière (non singulière)* si $\det(A) \neq 0$. On a l'existence et l'unicité de la solution $\mathbf{x}$ (pour n'importe quel vecteur $\mathbf{b}$ donné) si et seulement si la matrice associée au système linéaire est régulière.

Sommaire méthodes directes

- Résolution par factorisation LU ou de Choleski
- Coût de la factorisation LU
- Problèmes de Précision

En gros, pour le premier point il faut savoir répondre aux deux prochains transparents

Factorisation LU

La factorisation LU. Si $A = PLU$

- A L et U sont triangulaires
- B $\det P = \pm 1$
- C Rśolution par $L\mathbf{y} = P\mathbf{b}$ et $U\mathbf{x} = \mathbf{y}$
- D sur Python : `x = scipy.linalg.lu(A)`
- E sur Python : `P,L,U = scipy.linalg.lu(A)`

Factorisation de Cholevski

La factorisation de Cholevski A

- A Peut toujours se faire
- B Peut se faire si $A = A^T$
- C Peut se faire si A est spd
- D $A = LL^T$, L triangulaires supérieure
- E $A = LL^T$, L triangulaires inférieure
- F $\det A = (\det L)^2$
- G Rśolution par $L\mathbf{y} = \mathbf{b}$ et $L\mathbf{x} = \mathbf{y}$
- H sur Python : `x = scipy.linalg.choleski(A)`
- I sur Python : `L = scipy.linalg.choleski(A)`

spd = symétrique définie positive

Mineurs principaux I

Les mineurs principaux d'une matrice $A \in \mathbb{R}^{n\times n}$ sont les déterminants des matrices $A_p = (a_{i,j})_{1\leq i,j\leq p}$, $p = 1, ..., n$.
Critère de Sylvester : une matrice symétrique $A \in \mathbb{R}^{n\times n}$ est définie positive si et seulement si les mineurs principaux de A sont tous positifs.

Considérations sur la précision

Definition

On définit le **conditionnement** d'une matrice M symétrique définie positive comme le rapport entre la valeur maximale et minimale de ses valeurs propres, i.e.

$$K(M) = \frac{\lambda_{max}(M)}{\lambda_{min}(M)} \tag{1}$$

On peut montrer la relation suivante :

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq K(A) \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} \tag{2}$$

où $\mathbf{r}$ est le résidu $\mathbf{r} = \mathbf{b} - A\hat{\mathbf{x}}$.

Sommaire méthodes itératives

- Méthodes itératives : définitions
- Méthode de Richardson
- Méthodes de Jacobi et de Gauss-Seidel
  Critères de convergence
- Méthodes du Gradient et du Gradient Conjugués
  Critères de convergence

Méthodes itératives

Résoudre un système linéaire $A\mathbf{x} = \mathbf{b}$ par une méthode itérative consiste à construire une suite de vecteurs $\mathbf{x}^{(k)}$, $k \geq 0$, de $\mathbb{R}^n$ qui converge vers la solution exacte $\mathbf{x}$, c'est-à-dire :

$$\lim_{k \to \infty} \mathbf{x}^{(k)} = \mathbf{x}$$

pour n'importe quelle donnée initiale $\mathbf{x}^{(0)} \in \mathbb{R}^n$.
On peut considérer la relation de récurrence suivante :

$$\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{g}, \quad k \geq 0 \tag{3}$$

où B est une matrice bien choisie (dépendante de A) et $\mathbf{g}$ est un vecteur (dépendant de A et de $\mathbf{b}$), qui vérifient la relation (de consistance)

$$\mathbf{x} = B\mathbf{x} + \mathbf{g}. \tag{4}$$

Étant donné que $\mathbf{x} = A^{-1}\mathbf{b}$, on obtient $\mathbf{g} = (I - B)A^{-1}\mathbf{b}$; la méthode itérative est donc complètement définie par la matrice B qui est appellée *matrice d'itération*.
En définissant l'erreur au pas k comme

$$\boxed{\mathbf{e}^{(k)} = \mathbf{x} - \mathbf{x}^{(k)},}$$

on obtient la relation de récurrence :

$$\boxed{\mathbf{e}^{(k+1)} = B\mathbf{e}^{(k)} \quad \text{et donc} \quad \mathbf{e}^{(k+1)} = B^{k+1}\mathbf{e}^{(0)}, \quad k = 0, 1, \ldots.}$$

On peut montrer que $\lim_{k\to\infty} \mathbf{e}^{(k)} = \mathbf{0}$ pour tout $\mathbf{e}^{(0)}$ (et donc pour tout $\mathbf{x}^{(0)}$) si et seulement si

$$\boxed{\rho(B) < 1,}$$

où $\rho(B)$ est le *rayon spectral* de la matrice B, défini par

$$\boxed{\rho(B) = \max |\lambda_i(B)|}$$

et $\lambda_i(B)$ sont les valeurs propres de la matrice B.

Plus la valeur de $\rho(B)$ est petite, moins il est nécessaire d'effectuer d'itérations pour réduire l'erreur initiale d'un facteur donné.

Construction d'une méthode itérative

Une méthode générale pour construire une méthode itérative est basée sur la décomposition de la matrice A :

$$A = P - (P - A)$$

où P est une matrice inversible appelée *préconditionneur* de A.
Alors,

$$A\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad P\mathbf{x} = (P - A)\mathbf{x} + \mathbf{b}$$

qui est de la forme (4) en posant

$$B = P^{-1}(P - A) = I - P^{-1}A \quad \text{et} \quad \mathbf{g} = P^{-1}\mathbf{b}.$$

On peut définir la méthode itérative correspondante

$$P(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) = \mathbf{r}^{(k)} \qquad k \geq 0$$

où $\mathbf{r}^{(k)}$ désigne le *résidu* à l'itération k : $\boxed{\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}}$
On peut généraliser cette méthode de la manière suivante :

$$P(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) = \alpha_k \mathbf{r}^{(k)} \qquad k \geq 0 \tag{5}$$

où $\alpha_k \neq 0$ est un paramètre pour améliorer la convergence de la suite $\mathbf{x}^{(k)}$.
La méthode (5) est appelée *méthode de Richardson*.

La matrice P doit être choisie de telle manière que le coût de la résolution de (5) soit assez faible. Par exemple, une matrice P diagonale ou triangulaire vérifierait ce critère.

La méthode de Jacobi

Si les éléments diagonaux de A sont non nuls, on peut poser

$$P = D = diag(a_{11}, a_{22}, \ldots, a_{nn})$$

D étant la partie diagonale de A :

$$D_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ a_{ij} & \text{si } i = j. \end{cases}$$

La méthode de Jacobi correspond à ce choix avec $\alpha_k = 1$ pout tout k.
On déduit alors :

$$D\mathbf{x}^{(k+1)} = \mathbf{b} - (A - D)\mathbf{x}^{(k)} \qquad k \geq 0.$$

Par composantes :

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j\neq i}^{n} a_{ij} x_j^{(k)} \right), \quad i = 1, \ldots, n. \tag{6}$$

La méthode de Jacobi peut s'écrire sous la forme générale

$$\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{g},$$

avec

$$B = B_J = D^{-1}(D - A) = I - D^{-1}A, \qquad \mathbf{g} = \mathbf{g}_J = D^{-1}\mathbf{b}.$$

La méthode de Gauss-Seidel

Cette méthode est définie par la formule suivante :

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

Cette méthode correspond à (3) avec $P = D - E$ et $\alpha_k = 1$ ($\forall k \geq 0$) où E est la matrice triangulaire inférieure

$$\begin{cases} E_{ij} = -a_{ij} & \text{si } i > j \\ E_{ij} = 0 & \text{si } i \leq j \end{cases}$$

(partie triangulaire inférieure de A sans la diagonale et avec les éléments changés de signe).

On peut écrire cette méthode sous la forme (5), avec la matrice d'itération $B = B_{GS}$ donnée par

$$\boxed{B_{GS} = (D - E)^{-1}(D - E - A)}$$

et

$$\mathbf{g}_{GS} = (D - E)^{-1}\mathbf{b}.$$

Convergence

On a les résultats de convergence suivants :

- *Si A est une matrice à diagonale dominante stricte par ligne, c'est -à-dire*

$$|a_{ii}| > \sum_{j=1,\dots,n; j\neq i} |a_{ij}|, \quad i = 1, \dots, n.$$

  alors les méthodes de Jacobi et de Gauss-Seidel sont convergentes
- Soit A régulière, tridiagonale et dont les coefficients diagonaux sont tous non-nuls. Alors les méthodes de Jacobi et de Gauss-Seidel sont toutes les deux soit divergentes soit convergentes. Dans le deuxième cas, $\rho(B_{GS}) = \rho(B_J)^2$
- *Si A est une matrice symétrique définie positive, alors la méthode de Gauss-Seidel converge (la méthode de Jacobi pas forcément).*

La méthode de Richardson

Considérons la méthode itérative générale :

$$P(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) = \alpha_k \mathbf{r}^{(k)}, \qquad k \geq 0. \tag{7}$$

Cette méthode est appelée méthode de Richardson stationnaire préconditionné si $\alpha_k = \alpha$ (une constante donnée) ; autrement elle est dite méthode de Richardson dynamique préconditionné quand α_k peut varier au cours des itérations.

La matrice inversible P est appelée *préconditionneur* de A.

Si A et P sont symétriques définies positives, alors on a deux critères optimaux pour le choix de α_k :

1. *Cas stationnaire* :

$$\alpha_k = \alpha_{opt} = \frac{2}{\lambda_{min}(P^{-1}A) + \lambda_{max}(P^{-1}A)}, \qquad k \geq 0,$$

où λ_{min} et λ_{max} désignent respectivement la plus petite et la plus grande valeur propre de la matrice $P^{-1}A$.

2. *Cas dynamique* :

$$\alpha_k = \frac{(\mathbf{z}^{(k)})^T \mathbf{r}^{(k)}}{(\mathbf{z}^{(k)})^T A \mathbf{z}^{(k)}}, \qquad k \geq 0,$$

où $\mathbf{z}^{(k)} = P^{-1}\mathbf{r}^{(k)}$ est le résidu préconditionné.
Cette méthode est aussi appelée méthode du gradient préconditionné.

Si $P = I$ et A est symétrique définie positive, on trouve les méthodes :

- de Richardson stationnaire si on choisit :

$$\alpha_k = \alpha_{opt} = \frac{2}{\lambda_{min}(A) + \lambda_{max}(A)}. \tag{8}$$

- du gradient si :

$$\alpha_k = \frac{(\mathbf{r}^{(k)})^T \mathbf{r}^{(k)}}{(\mathbf{r}^{(k)})^T A \mathbf{r}^{(k)}}, \qquad k \geq 0. \tag{9}$$

On peut récrire plus efficacement la méthode du gradient préconditionné de la manière suivante : soit $\mathbf{x}^{(0)}$, poser $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$, puis pour $k \geq 0$,

$$
\begin{aligned}
P\mathbf{z}^{(k)} &= \mathbf{r}^{(k)} \\
\alpha_k &= \frac{(\mathbf{z}^{(k)})^T \mathbf{r}^{(k)}}{(\mathbf{z}^{(k)})^T A\mathbf{z}^{(k)}} \\
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k \mathbf{z}^{(k)} \\
\mathbf{r}^{(k+1)} &= \mathbf{r}^{(k)} - \alpha_k A\mathbf{z}^{(k)}.
\end{aligned}
$$

On observe qu'on doit résoudre un système linéaire pour la matrice P à chaque itération ; donc P doit être telle que la résolution du système associé soit facile (c'est-à-dire avec un coût raisonnable). Par exemple, on pourra choisir P diagonale (comme dans le cas du gradient ou de Richardson stationnaire) ou triangulaire.

Convergence de la méth. de Richardson

Considérons tout d'abord les méthodes de Richardson stationnaires ; on a le résultat de convergence suivant :

Theorem (Cas stationnaire)

On suppose la matrice P inversible et les valeurs propres de $P^{-1}A$ strictement positives et telles que $\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n = \lambda_{min} > 0$. Alors la méthode de Richardson stationnaire est convergente si et seulement si $0 < \alpha < 2/\lambda_1$. De plus, le rayon spectral de la matrice d'itération R_α est minimal si $\alpha = \alpha_{opt}$

$$\alpha_{opt} = \frac{2}{\lambda_{min} + \lambda_{max}},$$

avec

$$\rho_{opt} = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{min} + \lambda_{max}}$$

Dans le cas dynamique, on a un résultat qui permet de choisir de façon optimale le paramètre d'accélération à chaque étape, si la matrice A est symétrique définie positive :

Theorem (Cas dynamique)

Si A est symétrique définie positive, le choix optimal de α_k est donné par

$$\alpha_k = \frac{(\mathbf{r}^{(k)}, \mathbf{z}^{(k)})}{(A\mathbf{z}^{(k)}, \mathbf{z}^{(k)})}, \qquad k \geq 0 \tag{10}$$

où

$$\mathbf{z}^{(k)} = P^{-1}\mathbf{r}^{(k)}. \tag{11}$$

Pour le cas stationnaire et pour le cas dynamique on peut démontrer que, si A et P sont symétriques définies positives, la suite $\{\mathbf{x}^{(k)}\}$ donnée par la méthode de Richardson (stationnaire et dynamique) converge vers $\mathbf{x}$ lorsque $k \to \infty$, et

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_A \leq \left(\frac{K(P^{-1}A) - 1}{K(P^{-1}A) + 1} \right)^k \|\mathbf{x}^{(0)} - \mathbf{x}\|_A, \quad k \geq 0, \tag{12}$$

où $\|\mathbf{v}\|_A = \sqrt{\mathbf{v}^T A \mathbf{v}}$ et $K(P^{-1}A)$ est le conditionnement de la matrice $P^{-1}A$.

Remarque. Dans le cas de la méthode du gradient ou de Richardson stationnaire l'estimation de l'erreur devient

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_A \leq \left(\frac{K(A)-1}{K(A)+1}\right)^k \|\mathbf{x}^{(0)} - \mathbf{x}\|_A, \quad k \geq 0. \tag{13}$$

Remarque. Si A et P sont symétriques définies positives, on a

$$K(P^{-1}A) = \frac{\lambda_{max}(P^{-1}A)}{\lambda_{min}(P^{-1}A)}.$$

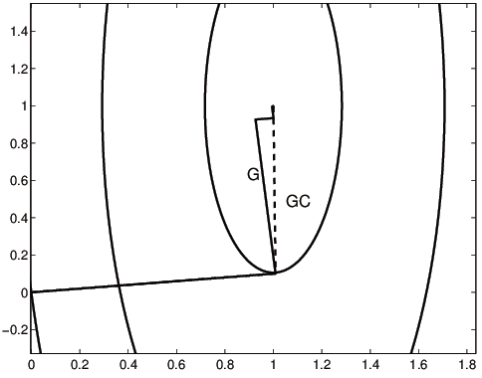
La méthode du gradient conjugué

Une méthode encore plus rapide dans le cas où P et A sont symétriques définies positives est celle du *gradient conjugué préconditionné* qui s'exprime ainsi :
soit $\mathbf{x}^{(0)}$ une donnée initiale ; on calcule $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$, $\mathbf{z}^{(0)} = P^{-1}\mathbf{r}^{(0)}$, $\mathbf{p}^{(0)} = \mathbf{z}^{(0)}$, puis pour $k \geq 0$,

$$
\begin{aligned}
\alpha_k &= \frac{{\mathbf{p}^{(k)}}^T \mathbf{r}^{(k)}}{{\mathbf{p}^{(k)}}^T A\mathbf{p}^{(k)}} \\
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)} \\
\mathbf{r}^{(k+1)} &= \mathbf{r}^{(k)} - \alpha_k A\mathbf{p}^{(k)} \\
P\mathbf{z}^{(k+1)} &= \mathbf{r}^{(k+1)} \\
\beta_k &= \frac{\left(A\mathbf{p}^{(k)}\right)^T \mathbf{z}^{(k+1)}}{\left(A\mathbf{p}^{(k)}\right)^T \mathbf{p}^{(k)}} \\
\mathbf{p}^{(k+1)} &= \mathbf{z}^{(k+1)} - \beta_k \mathbf{p}^{(k)} .
\end{aligned}
$$

Dans ce cas, l'estimation de l'erreur est donnée par

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_A \leq \frac{2c^k}{1+c^{2k}} \|\mathbf{x}^{(0)} - \mathbf{x}\|_A , \quad k \geq 0 \quad \text{où} \quad c = \frac{\sqrt{K_2(P^{-1}A)} - 1}{\sqrt{K_2(P^{-1}A)} + 1} . \tag{14}$$



Critères de convergence

On a la relation suivante :
Si A est une matrice symétrique définie positive, alors

$$\frac{\|\mathbf{x}^{(k)} - \mathbf{x}\|}{\|\mathbf{x}\|} \leq K(A) \frac{\|\mathbf{r}^{(k)}\|}{\|\mathbf{b}\|}. \tag{15}$$

L'erreur relative à la k-ième itération peut être majorée par le résidu relatif multiplié par le conditionnement de A.
En particulier, si $K(A) \approx 1$, une petite valeur de la norme du résidu correspond à une petite valeur de la norme de l'erreur ; si $K(A) \gg 1$, cette relation peut être fausse.
On a également une estimation (utilisée si $P \neq I$) :

$$\frac{\|\mathbf{x}^{(k)} - \mathbf{x}\|}{\|\mathbf{x}\|} \leq K(P^{-1}A) \frac{\|P^{-1}\mathbf{r}^{(k)}\|}{\|P^{-1}\mathbf{b}\|}.$$