

Méthodes numériques pour la résolution de systèmes linéaires

1 Motivation

La modélisation mathématique de phénomènes physiques débouche généralement sur des équations différentielles et/ou aux dérivées partielles dont la résolution analytique n'est pas toujours possible. En effet, seulement dans certaines situations particulières qu'il est possible d'obtenir la solution exacte de ces équations. Par exemple, la distribution de la température u dans une barre métallique de longueur $\ell > 0$ chauffée de l'intérieur par un flux de chaleur f est régie par l'équation de la chaleur stationnaire en 1D:

$$\begin{cases} -u''(x) = f(x), & \forall x \in (0, \ell), \\ u(0) = u(\ell) = 0. \end{cases} \quad (1)$$

Les extrémités de la barre sont maintenues à une température nulle. Si la fonction f n'est pas simple à intégrer ou n'est connue qu'en certains points $x_j \in (0, \ell)$, le calcul de la température exacte u solution du problème (1) en tout point x de $(0, \ell)$ n'est pas facile/possible. Dans ce cas, nous cherchons plutôt à déterminer une approximation de $u(x)$. Pour cela, on subdivise le domaine $(0, \ell)$ en un nombre fini de points: Etant donné $N \in \mathbb{N}^*$, on considère le pas $\Delta x = \ell/(N+1)$ et les points $x_i = i\Delta x$, pour $i = 0, \dots, N+1$.

Afin de trouver une approximation à $u''(x_i)$, nous utilisons le Développement Limité (DL) de u au point x_i à l'ordre 2:

$$u(x) = u(x_i) + u'(x_i)(x - x_i) + \frac{u''(x_i)}{2}(x - x_i)^2 + o((x - x_i)^2), \quad \forall x.$$

En considérant le changement de variable $h = x - x_i$, il vient que

$$u(x_i + h) = u(x_i) + u'(x_i)h + \frac{h^2}{2}u''(x_i) + h^2\varepsilon(h), \quad \text{où } \lim_{h \rightarrow 0} \varepsilon(h) = 0.$$

Pour $h = \pm\Delta x$ assez petit, en négligeant le reste, nous obtenons

$$\begin{cases} u(x_{i+1}) \approx u(x_i) + u'(x_i)\Delta x + \frac{\Delta^2 x}{2}u''(x_i), \\ u(x_{i-1}) \approx u(x_i) - u'(x_i)\Delta x + \frac{\Delta^2 x}{2}u''(x_i), \end{cases} \implies u''(x_i) \approx \frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1}))}{\Delta^2 x}.$$

Par conséquent, afin de déterminer $u_i \approx u(x_i)$ en tout point x_i , pour $i = 1, \dots, N$, nous résolvons le système linéaire de N équations à N inconnus suivant:

$$\begin{cases} -\frac{1}{\Delta^2 x}(u_{i+1} - 2u_i + u_{i-1}) = f(x_i), & \text{pour } i = 1, \dots, N \\ u_0 = u_{N+1} = 0. \end{cases} \quad (2)$$


qui peut se mettre sous la forme matricielle suivante:

$$\overbrace{\begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & \dots & & 0 & -1 & 2 \end{pmatrix}}^A \overbrace{\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{N-1} \\ u_N \end{pmatrix}}^X = (\Delta x)^2 \overbrace{\begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_{N-1}) \\ f(x_N) \end{pmatrix}}^b. \quad (3)$$

Ceci est un exemple de problèmes physiques dont l'approximation de la solution est obtenue à partir de la résolution d'un système linéaire $AX = b$, où A est une matrice connue de taille $N \times N$, $b \in \mathbb{R}^N$ donné et $X \in \mathbb{R}^N$ le vecteur inconnu. En outre, l'approximation de $u''(x_i)$ est d'autant plus précise que le pas $\Delta x = \ell/(N+1)$ est petit ce qui est équivalent à prendre N grand. Donc, dans la pratique les systèmes linéaires les plus intéressants sont ceux où la taille N est assez grand.

Si la matrice A est inversible, alors la solution X de $AX = b$ est unique et définie par $X = A^{-1}b$. En revanche, le calcul de la matrice inverse A^{-1} n'est plus faisable pour N grand. La méthode la plus connue pour calculer A^{-1} est celle de Cramer. La complexité (le nombre d'opérations nécessaires pour calculer la solution) de cette dernière est de $(N+1)(NN! - 1)$. En se référant à la formule de Stirling $N! \sim \sqrt{2N\pi}(N/e)^N$ et utilisant un ordinateur fonctionnant à 100 megaflops (flops=opération à virgule flottante par seconde), il faudrait environ 3×10^{146} années pour résoudre avec la méthode de Cramer le système linéaire $AX = b$, où A est une matrice inversible de taille 100×100 .

Pour résoudre $AX = b$, nous utilisons des méthodes numériques qui se déclinent en deux catégories: Méthodes directes ont généralement une complexité d'ordre N^3 mais assez précises: Méthode de Gauss $O(2N^3/3)$, méthode LU $O(N^3/2)$, méthode de Cholesky $O(N^3/6)$ et méthode de Householder $O(4N^3/3)$. Méthodes itératives ont une complexité d'ordre N^2 mais moins précises: Méthode Jacobi et méthode de Gauss-Seidel.

 **But du projet:** Etudier des méthodes numériques directes et itératives permettant de résoudre des systèmes linéaires $AX = b$ **SANS** calculer A^{-1} .

2 Exercice

Nous considérons les éléments suivants:

$$A = \begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 3.0001 \end{bmatrix} \quad \text{et} \quad \delta b = \begin{bmatrix} -0.0001 \\ 0.0001 \end{bmatrix}$$

1. Calculer A^{-1} , $\|A\|_\infty$ et $\|A^{-1}\|_\infty$. En déduire le conditionnement de la matrice A .
2. Montrer que $X = (1, 1)^\top$ résout $Ax = b$ et que $\|A\|_\infty \|X\|_\infty = \|b\|_\infty$.

3. Montrer que $\delta X = (2, -1 - 10^{-4}/2)^\top$ est solution de $A\delta X = \delta b$.
4. Résoudre $AX = b + \delta b$. Montrer que $\|A^{-1}\|_\infty \|\delta b\|_\infty = \|\delta x\|_\infty$.
5. Dans le cas général, rappeler l'inégalité entre $\frac{\|\delta X\|_\infty}{\|X\|_\infty}$ et $\frac{\|\delta b\|_\infty}{\|b\|_\infty}$. En utilisant les calculs précédents, montrer qu'il n'est pas possible d'obtenir une majoration plus fine.

3 Projet

3.1 Première partie du projet: Méthodes directes

Le but de cette première partie est d'étudier des méthodes directes **économiques** pour la résolution des systèmes linéaires $AX = b$. Le mot économique est pour dire que nous ne cherchons pas à calculer explicitement la matrice inverse A^{-1} de A afin de résoudre $AX = b$. Le calcul de A^{-1} est connu dans la littérature par le fait qu'il est très coûteux en particulier pour des systèmes linéaires de grande taille.

🚩 Quand est-ce que la résolution de $AX = b$ par des méthodes directes est possible?

- 1. Ecrire une fonction $X = \text{Solinf}(L, b)$ qui étant donné une matrice inversible **triangulaire inférieure** L de taille $N \times N$ et un vecteur colonne b de taille N , calcule la solution X du problème: $LX = b$.
- 2. Ecrire une fonction $X = \text{Solsup}(U, b)$ qui étant donné une matrice inversible **triangulaire supérieure** U de taille $N \times N$ et un vecteur colonne b de taille N , calcule la solution X du système linéaire: $UX = b$.

► Application: Utiliser les deux fonctions *Solinf* et *Solsup* pour résoudre:

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 1 & 4 & -1 \end{bmatrix} x = \begin{bmatrix} 1 \\ 8 \\ 10 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 8 \\ 0 & 0 & 5 \end{bmatrix} x = \begin{bmatrix} 6 \\ 16 \\ 15 \end{bmatrix}.$$

- 3. Ecrire une fonction $[U, e] = \text{Trisup}(A, b)$ qui étant donné une matrice carrée A inversible de taille $N \times N$ et un vecteur colonne b de taille N , détermine par **la méthode d'élimination de Gauss**, quand c'est possible, la matrice triangulaire supérieure U et le vecteur colonne associé e tels que: $AX = b \Leftrightarrow UX = e$.

► Application: Utiliser la fonction *Trisup* pour déterminer le système linéaire $Ux = e$ équivalent au système suivant:

$$\begin{bmatrix} 3 & 1 & 2 \\ 3 & 2 & 6 \\ 6 & 1 & -1 \end{bmatrix} x = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}.$$

- 4. Ecrire une fonction $X = \text{ResolG}(A, b)$ qui étant donné une matrice carrée A inversible de taille $N \times N$ et un vecteur colonne b de taille N , calcule quand c'est possible la solution

du problème $AX = b$ en utilisant la **méthode de Gauss**. Cette fonction doit utiliser les deux fonctions précédentes.

► **Application:** Utiliser la fonction *ResolG* pour résoudre

$$\begin{bmatrix} 1 & 2 & 3 \\ 5 & 2 & 1 \\ 3 & -1 & 1 \end{bmatrix} x = \begin{bmatrix} 5 \\ 5 \\ 6 \end{bmatrix} \quad \text{puis} \quad \begin{bmatrix} 2 & 1 & 5 \\ 1 & 2 & 4 \\ 3 & 4 & 10 \end{bmatrix} x = \begin{bmatrix} 5 \\ 5 \\ 6 \end{bmatrix}.$$

- **5.** Ecrire une fonction $[L,U]=LU(A)$ qui étant donnée une matrice carrée A inversible de taille $N \times N$, détermine en appliquant la méthode de **factorisation LU** la matrice triangulaire supérieure U et la matrice triangulaire inférieure L telles que $A = LU$. On peut en particulier reprendre la fonction précédente *Trisup* que l'on modifiera pour obtenir la fonction *LU*.

► **Application:** Utiliser la fonction *LU* pour déterminer les deux matrices L et U dans le cas où la matrice A est la suivante:

$$A = \begin{bmatrix} 3 & 1 & 2 \\ 3 & 2 & 6 \\ 6 & 1 & -1 \end{bmatrix}.$$

- **6.** Ecrire une fonction $[x]=ResolLU(A,b)$ qui étant donnés une matrice carrée A inversible de taille $N \times N$ et un vecteur colonne b de taille N , calcule quand c'est possible la solution X de $AX = b$ en utilisant la factorisation $A = LU$.

► **Application:** Utiliser la fonction *ResolLU* pour déterminer la solution de

$$\begin{bmatrix} 1 & 2 & 3 \\ 5 & 2 & 1 \\ 3 & -1 & 1 \end{bmatrix} X = \begin{bmatrix} 5 \\ 5 \\ 6 \end{bmatrix}.$$

- **7. Décomposition de Cholesky :** Montrer que toute matrice carrée, symétrique et définie positive A admet une unique décomposition du type

$$A = LL^T.$$

où L est une matrice triangulaire inférieure dont tous les éléments diagonaux sont strictement positifs.

2) Ecrire une fonction $[L]=Cholesk(A)$ qui étant donnée une matrice carrée A symétrique définie positive, calcule la matrice L associée. Que vaut le déterminant de A ?

► **Application:** Ecrire une fonction $[X]=CholeskResol(A,b)$ qui utilise les trois fonctions précédentes *Cholesk*, *solsup* et *solinf* pour résoudre

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 5 & 5 & 5 \\ 1 & 5 & 14 & 14 \\ 1 & 5 & 14 & 15 \end{bmatrix} x = \begin{bmatrix} 5 \\ 1 \\ 3 \\ 1 \end{bmatrix} \quad \text{vérifier que l'on a } L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 2 & 3 & 0 \\ 1 & 2 & 3 & 1 \end{bmatrix}.$$

3.2 Deuxième partie du projet: Méthodes itératives

► **But:** Construire une suite (X_k) qui, étant donné un itéré initial $X_0 \in \mathbb{R}^N$, converge après un nombre fini K d'itérations vers $X \in \mathbb{R}^N$ la solution de $AX = b$:

$$\lim_{k \rightarrow K} X_k = X \quad \text{tel que} \quad AX = b.$$

► **Comment construire la suite (X_k) ?** On décompose la matrice $A = M - N$, où M est une matrice "simple" et inversible. Par conséquent, il vient que

$$AX = b \Leftrightarrow X = M^{-1}(NX + b).$$

L'idée générale consiste à définir les éléments de la suite (X_k) comme les solutions du schéma numérique suivant:

$$\begin{cases} X_{k+1} = M^{-1}(NX_k + b), & \forall k \geq 0 \\ X_0 \text{ donné} \end{cases}$$

• **Interprétation:** Si la suite (X_k) converge vers X_K après un nombre fini K d'itérations c-à-d $X_{K+1} = X_K = X_k, \forall k \geq K$ alors, on aura

$$X_K = M^{-1}(NX_K + b) \Leftrightarrow MX_K = NX_K + b \Leftrightarrow (M - N)X_K = b \Leftrightarrow AX_K = b.$$

☞ La limite X_K est bien la solution unique du système linéaire $AX = b$.

• **Condition pour la convergence:** La solution X de $AX = b$ vérifie donc $X = M^{-1}(NX + b)$. Par conséquent, on a

$$\|X_k - X\| = \|M^{-1}N(X_{k-1} - X)\| \leq \|M^{-1}N\| \|X_{k-1} - X\|, \quad \forall k \geq 1.$$

Ensuite, par récurrence, on obtient

$$\|X_k - X\| \leq \|M^{-1}N\|^k \|X_0 - X\|, \quad \forall k \geq 1.$$

☞ Les itérations (X_k) convergent si $\|M^{-1}N\| < 1$. La convergence est d'autant plus rapide que $\|M^{-1}N\|$ est petite.

★ **Première méthode:** Méthode de Jacobi

L'idée consiste à écrire $A = D - (E + F)$, où

- * D est la matrice diagonale définie par les mêmes éléments diagonaux de A .
- * E est la matrice dont les éléments en dessous de la diagonale sont les mêmes que ceux de A multipliés par -1 tandis que tous ses autres éléments sont égaux à zéro.
- * F est la matrice dont les éléments au-dessus de la diagonale sont les mêmes que ceux de A multipliés par -1 tandis que tous ses autres éléments sont égaux à zéro.

• **Itérations de Jacobi:** On note par X_k^i , pour $i = 1, \dots, N$ les composantes du vecteur $X_k \in \mathbb{R}^N$. De la même manière, on note par b^i et A_{ij} les éléments définissant A et b .

$$\begin{cases} X_{k+1} = D^{-1}((E + F)X_k + b), & \forall k \geq 0 \\ X_0 \text{ donné} \end{cases}$$

Ceci implique que pour tout $k \geq 1$,

$$X_{k+1}^i = \frac{b^i - \sum_{j=1, j \neq i}^N A_{ij} X_k^j}{A_{ii}}, \quad \text{pour } i = 1, \dots, N$$

★ Deuxième méthode: Méthode de Gauss-Seidel

$$X_{k+1}^i = \frac{b^i - \sum_{j=1}^{i-1} A_{ij} X_{k+1}^j - \sum_{j=i+1}^N A_{ij} X_k^j}{A_{ii}}, \quad \text{pour } i = 1, \dots, N$$