Résolution des systèmes linéaires

Méthodes directes pour résoudre $Ax = b$

Cadre. On cherche les vecteurs $x \in \mathbb{K}^n$ solutions de $Ax = b$, avec $A \in \mathcal{M}_n(\mathbb{K})$ et $b \in \mathbb{K}^n$ (avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$).

— Si A est inversible, alors, il existe une unique solution $x = A^{-1}b$

— Si A est non inversible et si $b \in Im(A)$ alors, il existe une infinité de solutions (qui diffèrent d'un élément de $ker(A)$)

— Si A est non inversible et si $b \notin Im(A)$ alors, il n'y a pas de solution.

Nous nous restreindrons par la suite au cas où A est inversible.

Proposition 1. Formules de Cramer. (p.80 [Cia], Prop.5.1.1 [AK]) *Soit $A \in \mathcal{M}_n(\mathbb{K})$ inversible et $a_1, ..., a_n \in \mathbb{K}^n$ les colonnes de A :*

$$A = (a_1|a_2|...|a_n).$$

Soit $b \in \mathbb{K}^n$. Alors l'unique solution $x \in \mathbb{K}^n$ du système $Ax = b$ est donnée par $x = (x_1, ..., x_n)$ avec

$$x_i = \frac{det((a_1|...|a_{i-1}|b|a_{i+1}|...|a_n))}{det(A)}, \quad \forall i \in \{1, ..., n\}.$$

Remarque 1. *Les formules de Cramer sont très peu adaptées au calcul sur ordinateur puisque calculer un déterminant est très coûteux (le coût d'un déterminant $n \times n$ est $c_n = n(1 + c_{n-1}) \geq nc_{n-1}$ donc $c_n \geq n!$ et nous devons évaluer $n + 1$ déterminants donc le coût des formules de Cramer est de l'ordre de $(n + 1)!$). (Ch.5 [AK]).*

Remarque 2. *Les méthodes directes reposent sur une décomposition de la matrice A en $A = MN$ où M est facile à inverser (triangulaire ou orthonormale) et N est triangulaire. On aura donc*

$$Ax = b \Leftrightarrow MNx = b \Leftrightarrow \begin{cases} My = b, \\ Nx = y. \end{cases}$$

Il faut donc savoir résoudre le système $Ax = b$ lorsque A est triangulaire inférieure (ou supérieure) : c'est l'algorithme de descente (ou de remontée).

Proposition 2. Algorithme de descente. (p. 89 [Ser]) *Soit A triangulaire inférieure et inversible (i.e $a_{i,i} \neq 0, \forall i \in \{1, ..., n\}$)*

$$A = \begin{pmatrix} a_{1,1} & 0 & & 0 \\ a_{2,1} & a_{2,2} & 0 & \\ \vdots & & \ddots & 0 \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix}.$$

La solution x de $Ax = b$ est alors donnée par

$$\begin{cases} x_1 = \dfrac{b_1}{a_{1,1}}, \\ x_i = \dfrac{b_i - \sum_{j=1}^{i-1} a_{i,j}x_j}{a_{i,i}}, \quad \textit{pour } i \in \{2, ..., n\}. \end{cases}$$

Remarque 3. *Si A est triangulaire supérieure, on trouve x_n en premier puis on en déduit $x_{n-1}, x_{n-2}, ..., x_1$, il s'agit de l'algorithme de remontée. (§4.1 [Cia]).*

1 Méthode de Gauss

1.1 Opérations élémentaires sur les lignes

Remarque 4. Matrices de dilatation. *Pour $\lambda \in \mathbb{K}^*$ et $p \in \{1, ..., n\}$, la matrice de dilatation $\mathcal{D}_p(\lambda)$ est définie par*

$$\mathcal{D}_p(\lambda) = \begin{array}{cc} \begin{array}{ccccccc} & & & p & & & \\ & & & \downarrow & & & \end{array} & \\ \begin{pmatrix} 1 & & & & 0 & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \lambda & & & \\ & & & & 1 & & \\ & 0 & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} & \begin{array}{c} \\ \\ \\ \leftarrow p \\ \\ \\ \\ \end{array} \end{array}.$$

Pour toute matrice $M \in \mathcal{M}_n(\mathbb{K})$, la multiplication à gauche par la matrice $\mathcal{D}_p(\lambda)$ multiplie la ligne p de M par λ.

Matrices de transposition. *Pour $(p,q) \in \{1, ..., n\}^2$, la matrice de transposition $\mathcal{P}_{p,q}$ (cas particulier d'une matrice de permutation) est définie par*

$$\mathcal{P}_{p,q} = \begin{array}{cc} \begin{array}{cccccccccc} & & & p & & & & q & & \\ & & & \downarrow & & & & \downarrow & & \end{array} & \\ \begin{pmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 0 & & & & 1 & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & 1 & & & & 0 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{pmatrix} & \begin{array}{c} \\ \\ \\ \leftarrow p \\ \\ \\ \\ \leftarrow q \\ \\ \\ \end{array} \end{array}.$$

Pour toute matrice $M \in \mathcal{M}_n(\mathbb{K})$, la multiplication à gauche par la matrice de transposition $\mathcal{P}_{p,q}$ inverse les lignes p et q de la matrice.

Matrices de transvection. *Pour $(p,q) \in \{1, ..., n\}^2$ et $\lambda \in \mathbb{K}^*$ on appelle la matrice de transvection, la matrice suivante*

$$\mathcal{T}_{p,q}(\lambda) = \begin{array}{cc} \begin{array}{ccccc} & & & q & \\ & & & \downarrow & \end{array} & \\ \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \ddots & \lambda & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} & \begin{array}{c} \\ \\ \leftarrow p \\ \\ \\ \end{array} \end{array}.$$

Pour toute matrice $M \in \mathcal{M}_n(\mathbb{K})$, la multiplication à gauche par la matrice $\mathcal{T}_{p,q}(\lambda)$ change L_p, la p-ième ligne de M, en $L_p + \lambda L_q$.

Proposition 3. *La matrice de transvection $\mathcal{T}_{p,q}(\lambda)$ est inversible, d'inverse $\mathcal{T}_{p,q}(\lambda)^{-1} = \mathcal{T}_{p,q}(-\lambda)$.*

Définition 1. *On appelle opération élémentaire sur les lignes une application de la forme $\mathcal{M}_n(\mathbb{K}) \ni M \mapsto UM$ où U est une matrice de transvection, transposition ou dilatation.*

1.2 La méthode de Gauss

Théorème 1 (Thm 4.2-1 [Cia], Thm 6.1.1 [AK])**.** *Soit $A \in \mathcal{M}_n(\mathbb{K})$ (inversible ou non) alors il existe (au moins) une matrice $M \in \mathcal{G}l_n(\mathbb{K})$, produit de matrices de transposition et de transvection, telle que MA soit triangulaire supérieure.*

Remarque 5. *Pour obtenir que MA soit une matrice triangulaire inférieure, il faudrait raisonner sur des opérations élémentaires sur les colonnes.*

Démonstration du Théorème 1. L'idée est de construire en $n-1$ étapes une suite de matrices A^i ayant toutes des zéros en dessous de la diagonale dans les i premières colonnes.
Initialisation : On pose $A^0 = A$.
Étape i, $i \geq 1$: On suppose qu'à la fin de l'étape $i-1$, on a construit une matrice A^{i-1} telle que

$$A^{i-1}_{k,j} = 0, \text{ si } k > j, \text{ pour tout } j \in \{1, ..., i-1\}.$$

Les coefficients de A^{i-1} sont bien nuls au dessous de la diagonale dans les $i-1$ premières colonnes.

$$A^{i-1} = \begin{pmatrix} * & & & & \\ & \ddots & & * & \\ & & * & & \\ & & & A^{i-1}_{i,i} & \\ & 0 & & A^{i-1}_{i+1,i} & * \\ & & & \vdots & \\ & & & A^{i-1}_{in,i} & \end{pmatrix}$$

- Si A est inversible.

1. Si $A^{i-1}_{i,i} = 0$, on choisit $j > i$ tel que $A^{i-1}_{j,i} \neq 0$ (un tel j existe puisque sinon, la i-ème ligne serait combinaison linéaire des $i-1$ premières lignes et la matrice A ne serait pas inversible). On choisit ce $A^{i-1}_{j,i}$ comme pivôt et on échange les lignes i et j de A^{i-1} : $L_i \leftrightarrow L_j$.
   Du point de vue matriciel, cela revient à multiplier la matrice A^{i-1} par la matrice de transposition $\mathcal{P}_{i,j}$. On notera P^i cette matrice : $P^i = \mathcal{P}_{i,j}$.
   (par la suite L_j, L_i, $A^{i-1}_{i,i}$, $A^{i-1}_{j,i}$ seront les lignes et les coefficients de la matrice A^{i-1} après cette étape de permutation).
2. On fait la somme de chaque ligne L_j de A^{i-1} avec $-\frac{A^{i-1}_{j,i}}{A^{i-1}_{i,i}} L_i$ pour $j > i$

   $$L_j \to L_j - \frac{A^{i-1}_{j,i}}{A^{i-1}_{i,i}} L_i, \quad \forall j > i.$$

   Du point de vue matriciel, cela revient à multiplier A^{i-1} à gauche par une somme de matrices de transvection $\sum_{j=i+1}^n \mathcal{T}_{j,i}(-\frac{A^{i-1}_{j,i}}{A^{i-1}_{i,i}})$. Notons E^i cette matrice

   $$E^i = \sum_{j=i+1}^n \mathcal{T}_{j,i}(-\frac{A^{i-1}_{j,i}}{A^{i-1}_{i,i}}) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & 0 \\ & & 1 & & \\ & & -\frac{A^{i-1}_{i+1,i}}{A^{i-1}_{i,i}} & 1 & \\ 0 & & \vdots & & \ddots & \\ & & -\frac{A^{i-1}_{n,i}}{A^{i-1}_{i,i}} & & & 1 \end{pmatrix}.$$

   À la fin de l'étape i, nous obtenons une matrice

   $$A^i = E^i P^i A^{i-1} = E^i P^i ... E^1 P^1 A.$$

- Si A est non inversible, alors les éléments $A^{i-1}_{j,i}$, $j > i$, sont tous égaux à 0 pour au moins un i. Dans ce cas, la matrice A^{i-1} est déjà sous la forme souhaitée et on pose $P^i = E^i = I_n$, la matrice identité.

A la fin des $n-1$ étapes : on aura obtenu une matrice

$$A^{n-1} = E^{n-1}P^{n-1}...E^1P^1A$$

triangulaire supérieure.

On pose $M = E^{n-1}P^{n-1}...E^1P^1$ et on a

- MA triangulaire supérieure par construction
- M inversible et $\det(M) = \pm 1$ selon le nombre de permutations de deux lignes effectuées ($\det(M) = 1$ si le nombre de permutations est pair et $\det(M) = -1$ si ce nombre est impair). □

Remarque 6. *Attention, en pratique, on ne calcule pas les E^i et P^i (et on ne les stocke pas même si ce sont des matrices creuses !), on implémente directement la transformation sur les lignes L_i.*
Pour résoudre $Ax = b$, il suffit de résoudre $A^{n-1}x = Mb$ par l'algorithme de remontée (pour calculer Mb, on effectue les mêmes opérations d'échanges de lignes et de multiplications par $-\frac{A^{i-1}_{j,i}}{A^{i-1}_{i,i}}$ sur le vecteur b). Si A n'est pas inversible, on pourra toujours écrire la décomposition $A^{n-1}x = Mb$, mais il n'y aura pas forcément de solution.

2 Décomposition LU

Objectif. Décomposer la matrice A en produit de deux matrices triangulaires, et ainsi ramener le système $Ax = b$ à la résolution de deux systèmes triangulaires.
Nous verrons que cela revient à appliquer la méthode de Gauss sans l'étape de permutation des lignes (puisqu'on suppose des hypothèses supplémentaires sur la matrice A).

Définition 2. *Soit $A \in \mathcal{M}_n(\mathbb{K})$. On appelle sous-matrice d'ordre k de A la matrice $\Delta^k \in \mathcal{M}_k(\mathbb{K})$ construite à partir de A par*

$$\Delta^k_{i,j} = A_{i,j}, \qquad \forall (i,j) \in \{1, ..., k\}^2.$$

On appelle «mineur principal» d'ordre k de A le déterminant $\det(\Delta^k)$.

Théorème 2 (Thm 4.3-1 du [Cia], Thm 8.1.1 [Ser], Thm 6.2.1 [AK])**.** *Soit $A \in \mathcal{M}_n(\mathbb{K})$. Si toutes les sous-matrices Δ^k d'ordre k de A sont inversibles alors il existe un unique couple (L, U) de matrices avec $L \in \mathcal{M}_n(\mathbb{K})$ triangulaire inférieure à diagonale unitaire (i.e $L_{i,i} = 1$) et $U \in \mathcal{M}_n(\mathbb{K})$ triangulaire supérieure tels que $A = LU$.*

Remarque 7. *L'appellation «décomposition LU» vient de la forme des matrices*

$$L = \text{"lower triangular matrix" et } U = \text{"upper triangular matrix"}.$$

Remarque 8. *En particulier, si A est une matrice symétrique définie positive (i.e $x^TAx > 0$ pour $x \neq 0$) alors A vérifie bien les hypothèses du théorème de décomposition LU : $\det(\Delta^k) \neq 0$ pour tout $k \in \{1, ..., n\}$ (on a même $\det(\Delta^k) > 0$ dans ce cas particulier, la preuve se trouve à l'exercice 4 p.247 [Gou]).*

Démonstration du Théorème 2. Unicité : Supposons (L_1, U_1) et (L_2, U_2) comme dans le théorème 2. On a donc

$$A = L_1U_1 = L_2U_2. \tag{1}$$

Pour $j \in \{1, 2\}$, $(L_j)_{i,i} = 1$ et L_j est triangulaire (inférieure) donc L_j est bien inversible. De même, $A = \Delta^n$ qui est bien supposé inversible donc $U_j = L_j^{-1}A$ est bien inversible pour $j \in \{1, 2\}$. L'équation (1) implique donc

$$L_2^{-1}L_1 = U_2U_1^{-1}.$$

Comme $L_2^{-1}L_1$ est triangulaire inférieure et $U_2U_1^{-1}$ est triangulaire supérieure, on a que $L_2^{-1}L_1$ et $U_2U_1^{-1}$ sont diagonales.

Mais L_1 et L_2 sont à diagonale unitaire donc $L_2^{-1}L_1$ est, elle aussi, à diagonale unitaire. Donc

$$U_2U_1^{-1} = L_2^{-1}L_1 = I_n.$$

Ainsi $L_1 = L_2$ et $U_1 = U_2$, d'où l'unicité de la décomposition.

Existence : Supposons que, au cours de l'algorithme de Gauss, il n'y ait pas besoin de changer de pivôt (i.e $A^{i-1}_{i,i} \neq 0$, tous les pivots naturels sont non nuls). On montre alors avec la méthode de Gauss que la matrice $A^{n-1} = E^{n-1}...E^1 A$ (pas besoin des matrices de permutation P^i) est une matrice triangulaire supérieure, avec la matrice E^i définie précédemment à l'algorithme de Gauss.

Soit

$$U = A^{n-1} \text{ et } L = \left(E^{n-1}...E^1\right)^{-1}.$$

Alors, U est triangulaire supérieure et $A = LU$. Il nous reste à vérifier que L est une matrice triangulaire inférieure à diagonale unitaire.

Par définition de E^i, on a $E^i = \sum_{j=i+1}^n \mathcal{T}_{j,i}(-\frac{A^{i-1}_{j,i}}{A^{i-1}_{i,i}})$. D'après la proposition 3, on montre facilement que E^i est inversible, d'inverse

$$\begin{aligned}(E^i)^{-1} &= \sum_{j=i+1}^n \mathcal{T}_{j,i}\left(\frac{A^{i-1}_{j,i}}{A^{i-1}_{i,i}}\right)\\ &= \begin{pmatrix} 1 & & & & \\ & \ddots & & & 0 \\ & & 1 & & \\ & & \frac{A^{i-1}_{i+1,i}}{A^{i-1}_{i,i}} & 1 & \\ 0 & & \vdots & & \ddots \\ & & \frac{A^{i-1}_{n,i}}{A^{i-1}_{i,i}} & & & 1 \end{pmatrix}.\end{aligned}$$

Ainsi, en notant C^i_k la k-ième colonne de la matrice $(E^i)^{-1}$, nous avons

$$\left(E^{n-1}...E^1\right)^{-1} = (E^1)^{-1}(E^2)^{-1}...(E^{n-1})^{-1} = \begin{pmatrix} 1 & & & & \\ * & 1 & & 0 & \\ C^1_1 & C^2_2 & \ddots & & \\ * & * & * & 1 & \\ * & * & * & C^{n-1}_{n-1} & 1 \end{pmatrix}.$$

La matrice L est donc triangulaire inférieure et $L_{i,i} = 1, \forall i \in \{1, ..., n\}$.

Montrons que tous les pivôts «naturels» sont effectivement non nuls : Il faut montrer (par récurrence sur k) que l'hypothèse $\det(\Delta^k) \neq 0$ pour tout $k \in \{1, ..., n\}$ implique que les pivôts de chaque étape sont non nuls.

• Initialisation : Pour $k = 1$, $\det(\Delta^1) \neq 0$ signifie que $A^0_{1,1} = A_{1,1} \neq 0$. Donc le pivôt de la première étape $A^0_{1,1}$ est non nul.

• Soit $i \in [\![1, n]\!]$. Supposons l'hypothèse de récurrence : «$\det(\Delta^k) \neq 0$, $\forall k = 1, ..., i-1$ implique que tous les pivôts jusqu'à l'étape $i-1$ sont non nuls.» Alors montrons que «si $\det(\Delta^k) \neq 0, \forall k = 1, ...i$ alors le pivôt de l'étape i est non nul.»

Puisque $\det(\Delta^k) \neq 0$ pour tout $k \in \{1, ..., i-1\}$ (entre autre), on a pu construire par hypothèse de récurrence une matrice A^{i-1} telle que

$$A^{i-1} = E^{i-1}...E^1 A.$$

C'est-à-dire

$$\begin{pmatrix} * & & & & \\ & \ddots & & * & \\ & & * & & \\ & & & A^{i-1}_{i,i} & \\ 0 & & & A^{i-1}_{i+1,i} & * \\ & & & \vdots & \\ & & & A^{i-1}_{in,i} & \end{pmatrix} = \begin{pmatrix} 1 & & & & & & \\ * & \ddots & & 0 & & & \\ C^1_1 & * & 1 & & & & \\ * & * & * & 1 & & & \\ * & * & C^{i-1}_{i-1} & & 1 & & \\ * & * & * & 0 & & 1 & \\ * & * & * & & & & 1 \end{pmatrix} \left(\begin{array}{cc|cc} & \Delta^i & & * \\ - & - & A_{i,i} & \\ & & & \\ & * & & * \end{array}\right).$$

En posant $M^{i-1} = E^{i-1}...E^1$, on a donc en considérant les matrices précédentes par blocs

$$\begin{pmatrix} (A^{i-1})_{1,...,i} & * \\ * & * \end{pmatrix} = \begin{pmatrix} (M^{i-1})_{1,...,i} & 0 \\ * & I_{n-i} \end{pmatrix} \begin{pmatrix} \Delta^i & * \\ * & * \end{pmatrix}.$$

On a donc : $\det\left(\Delta^i\right)\det(M^{i-1})_{1,...,i} = \det(A^{i-1})_{1,...,i}$. Puisque $\det\left(\Delta^i\right) \neq 0$ par hypothèse et $\det(M^{i-1})_{1,...,i} = 1$, alors $\det(A^{i-1})_{1,...,i}$ est lui aussi non nul.

Or $\det(A^{i-1})_{1,...,i} = \prod_{j=1}^{i} A^{i-1}_{j,j} = \prod_{j=1}^{i} A^{j-1}_{j,j} \neq 0$ Donc $A^{i-1}_{i,i} \neq 0$ et ainsi, le pivôt de l'étape i est bien non nul.

□

Remarque 9. *Lorsque la matrice A est tridiagonale (et qu'elle vérifie les hypothèses du théorème 2) alors les matrices L et U sont des matrices bandes. (p.85 [Cia]).*

3 Décomposition de Cholesky

Dans cette section, $\mathbb{K} = \mathbb{R}$ et on se restreindra à des matrices A symétriques définies positives. Pour de telles matrices, on a vu précédemment qu'une décomposition LU existait, mais il existe une décomposition «plus simple» dans le sens où elle ne fait intervenir qu'une seule matrice : la décomposition de Cholesky.

Rappel. *Une matrice A est dite symétrique définie positive si et seulement si $x^T Ax > 0$ pour tout $x \neq 0$ (la forme quadratique définie par A est strictement positive).*

Théorème 3 (Thm 4.4-1 [Cia], Thm 8.2.1 [Ser], Thm 6.3.1 [AK])**.** *Soit $A \in \mathcal{M}_n(\mathbb{R})$ symétrique définie positive, alors il existe une unique matrice $B \in \mathcal{M}_n(\mathbb{R})$ triangulaire inférieure avec des éléments diagonaux strictement positifs telle que $A = BB^T$.*

Remarque 10. *Attention, ne marche que si A est symétrique définie positive !*

Démonstration du Théorème 3. Existence : La matrice A étant symétrique définie positive, ses mineurs principaux sont tous strictement positifs (voir remarque 8), elle admet donc une décomposition LU.

On a donc

$$\left(\begin{array}{ccc|c} & \Delta^i & | & * \\ - & - & A_{i,i} & \\ & & & \\ & * & & * \end{array}\right) = \begin{pmatrix} 1 & & & & \\ & 1 & & 0 & \\ & & \ddots & & \\ & * & & \ddots & \\ & & & 1 & \\ & & & & 1 \end{pmatrix} \begin{pmatrix} U_{1,1} & & & \\ & U_{2,2} & & * \\ & & \ddots & \\ & 0 & & \ddots \\ & & & & U_{n,n} \end{pmatrix}.$$

En regardant les blocs en haut à gauche de chaque matrice (argument déjà utilisé dans la preuve de la décomposition LU) nous avons l'égalité des déterminants

$$\prod_{j=1}^{i} U_{j,j} = \det(\Delta^i) > 0.$$

Donc par récurrence, chaque $U_{i,i}$ est strictement positif, on peut donc définir la matrice (inversible) D suivante :

$$D = \text{diag}(\sqrt{U_{1,1}}, \sqrt{U_{2,2}}, ..., \sqrt{U_{n,n}}).$$

On a $A = \underbrace{LD}_{=B}\underbrace{D^{-1}U}_{=C}$.

Montrons que $C = B^T$: Par hypothèse sur A, on a $A = A^T$ et en même temps $A = BC$. Donc

$$BC = C^T B^T \Longrightarrow (C^T)^{-1}B = B^T C^{-1}.$$

Or $(C^T)^{-1}B$ est une matrice triangulaire inférieure et $B^T C^{-1}$ est une matrice triangulaire supérieure. Ainsi $(C^T)^{-1}B$ et $B^T C^{-1}$ sont toutes deux des matrices diagonales.

Or nous avons, pour tout $i \in [\![1, n]\!]$,

$$B_{i,i} = \sqrt{U_{i,i}} \text{ (puisque } L_{i,i} = 1)$$

et

$$C_{i,i} = D_{i,i}^{-1} U_{i,i} = \frac{U_{i,i}}{\sqrt{U_{i,i}}} = \sqrt{U_{i,i}}.$$

Donc $(C^T)_{i,i} = \sqrt{U_{i,i}}$ et $(C^T)_{i,i}^{-1} = \frac{1}{\sqrt{U_{i,i}}}$, pour tout $i \in \{1, ..., n\}$.

Ainsi nous obtenons

$$\begin{aligned}((C^T)^{-1}B)_{i,i} &= \sum_{k=1}^{n} (C^T)_{i,k}^{-1} B_{k,i} \\ &= \sum_{k=i}^{n} (C^T)_{i,k}^{-1} B_{k,i} \text{ (car } B \text{ est triangulaire inférieure)} \\ &= (C^T)_{i,i}^{-1} B_{i,i} \text{ (car } (C^T)^{-1} \text{ est triangulaire supérieure)} \\ &= 1.\end{aligned}$$

Donc $C^T = B$ et ainsi la matrice A se décompose en $A = BB^T$.

Unicité : Supposons qu'il existe deux décompositions de Cholesky de la matrice : $A = B_1 B_1^T = B_2 B_2^T$.

On obtient ainsi $B_2^{-1} B_1 = B_2^T (B_1^T)^{-1}$ avec $B_2^{-1} B_1$ une matrice triangulaire inférieure et $B_2^T (B_1^T)^{-1}$ une matrice triangulaire supérieure. Donc ces deux matrices sont en réalité diagonales et on note D cette matrice diagonale : $B_2^{-1} B_1 = D \Longrightarrow B_1 = B_2 D$.

La décomposition de Cholesky donne $B_2 B_2^T = A = B_1 B_1^T = (B_2 D)(B_2 D)^T = B_2 (DD^T) B_2^T$. Ainsi $DD^T = I_n$ (donc $D_{i,i} = \pm 1$) et l'hypothèse que les coefficients diagonaux de B_1 et de B_2 sont strictement positifs implique que $D = I_n$.

Nous avons donc prouvé que $B_1 = B_2 D = B_2$, d'où l'unicité de la décomposition de Cholesky. $\square$

Remarque 11. *En pratique, on ne calcule pas la factorisation de Cholesky à partir de la décomposition LU comme fait dans la preuve précédente, mais on la calcule directement à partir de la matrice A (voir p.88 [Cia] ou §6.3.1 [AK]).*

Remarque 12. *Il existe une décomposition de Cholesky lorsque A est une matrice hermitienne définie positive ($\mathbb{K} = \mathbb{C}$). Les coefficients de la matrice B sont alors complexes sauf les coefficients diagonaux (qui restent réels positifs). (p. 96 [Ser]).*

4 Factorisation QR

Principe. C'est la décomposition de A en une matrice unitaire Q et une matrice triangulaire supérieure R. L'avantage de la décomposition QR c'est qu'elle s'applique à toutes matrices (pas de contrainte sur les mineurs principaux comme pour LU ni sur le caractère symétrique défini positif comme pour Cholesky). Cette méthode peut même s'appliquer sur des matrices non carrées. L'inconvénient de cette méthode est qu'elle est «moins efficace» que les décompositions LU ou de Cholesky (car Q est une matrice unitaire et non triangulaire).

Définition 3. *Une matrice $Q \in \mathcal{M}_n(\mathbb{R})$ est dite orthogonale (ou unitaire) si et seulement si $Q^{-1} = Q^T$. Une matrice $Q \in \mathcal{M}_n(\mathbb{C})$ est dite unitaire si et seulement si $Q^{-1} = Q^* = \bar{Q}^T$.*

Théorème 4 (Thm 4.5-2 [Cia], Prop 8.3.1 [Ser], Thm 6.4.1 [AK])**.** *Soit $A \in \mathcal{M}_n(\mathbb{K})$ inversible. Alors il existe un unique couple (Q, R) avec Q unitaire, R triangulaire supérieure à diagonale positive, tel que $A = QR$.*

Remarque 13. *Résoudre le système $Ax = b$ revient alors à résoudre $Rx = Q^T b$ au moyen d'un algorithme de remontée.*

Démonstration du Théorème 4. Il existe une preuve de la décomposition QR se basant sur le procédé d'orthonormalisation de Gramm-Schmidt dans $\mathbb{C}^n$ (p. 97 [Ser], p.117 [AK]). Mais nous lui préfèrerons la méthode de Householder développée ci-après (puisque c'est elle qu'on codera en pratique). Cette méthode de Householder consiste à multiplier la matrice A par une succession de matrices unitaires (les matrices de Householder) pour rendre A progressivement triangulaire supérieure. □

Remarque 14. *Lorsque la matrice A n'est pas inversible, il existe une décomposition QR mais elle n'est pas unique (p.98 [Ser]).*

4.1 Algorithme de Householder

Pour simplifier, on prend $\mathbb{K} = \mathbb{R}$ dans ce paragraphe mais l'algorithme marche aussi dans $\mathbb{C}$.

Définition 4. *Soit $v \in \mathbb{R}^n \backslash \{0\}$ (vecteur colonne), la matrice de Householder associée à v est définie par*

$$H_v = I_n - \frac{2vv^T}{||v||^2} \in \mathcal{M}_n(\mathbb{R}).$$

Remarque 15. *Le vecteur v étant un vecteur colonne, v^T est un vecteur ligne. Le produit vv^T est donc une matrice $n \times n$ dont le coefficient d'indice (i, j) s'écrit $v_i v_j$.*

Remarque 16. *Si $v = 0 \in \mathbb{R}^n$, alors $H_0 = I_n$ (la matrice identité est donc considérée comme une matrice de Householder).*

Proposition 4 (Lem 7.3.1 [AK], Thm 4.5-1 [Cia])**.** *Soit $v \in \mathbb{R}^n$,*
- *la matrice de Householder H_v est symétrique et orthogonale.*
- *Soit e_i le $i^{\text{ème}}$ vecteur de la base canonique de $\mathbb{R}^n$, on a $H_{v\pm||v||e_i} v = \mp||v||e_i$ (attention, au changement de signe).*

Démonstration de la proposition 4. • Le coefficient d'indice (i, j) de la matrice H_v vaut

$$(H_v)_{i,j} = \delta_{i,j} - \frac{2v_i v_j}{||v||^2} = (H_v)_{j,i}.$$

Ainsi la matrice H_v est bien symétrique.
• Puisque H_v est symétrique, nous avons

$$\begin{aligned} HH^T = HH &= \left(I_n - \frac{2vv^T}{||v||^2}\right)\left(I_n - \frac{2vv^T}{||v||^2}\right) \\ &= I_n - \frac{4vv^T}{||v||^2} + \frac{4(vv^T)(vv^T)}{||v||^4} \\ &= I_n - \frac{4vv^T}{||v||^2} + \frac{4v(v^T v)v^T}{||v||^4}. \end{aligned}$$

Puisque $v^T v = ||v||^2$, nous obtenons

$$HH^T = I_n - \frac{4vv^T}{||v||^2} + \frac{4v||v||^2 v^T}{||v||^4}.$$

Ainsi, $HH^T = I_n$ et H est bien une matrice orthogonale.
• Soit e_i le $i^{\text{ème}}$ vecteur de la base canonique de $\mathbb{R}^n$, nous posons $w = v \pm ||v||e_i$. Alors soit w est le vecteur nul et dans ce cas $H_w v = v$ (d'après la remarque 16). On a alors $H_w v = v = w \mp ||v||e_i = \mp||v||e_i$.
Soit w est un vecteur non nul, dans ce cas,

$$H_w v = v - \frac{2ww^T}{||w||^2} v = v - \frac{2(v \pm ||v||e_i)(v^T v \pm ||v||v)}{||v \pm ||v||e_i||^2} = v - \frac{2(v \pm ||v||e_i)(||v||^2 \pm ||v||v_i)}{||v \pm ||v||e_i||^2}.$$

Or $||v \pm ||v||e_i||^2 = 2||v||^2 \pm 2||v||v_i$, donc

$$H_w v = v - (v \pm ||v||e_i) = \mp ||v||e_i.$$

$\square$

Algorithme de Householder. (p.138 [AK], p.91-92 [Cia])
Étape 1 : On pose $A^1 = A$. Soit $a_1 \in \mathbb{R}^n$ la première colonne de A^1.

— Si $a_1 = \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, on pose $H^1 = I_n$.

— Sinon, on pose $H^1 = H_{a_1 + ||a_1||e_1}$ (où e_1 est le premier vecteur de la base canonique de $\mathbb{R}^n$).

On définit $A^2 = H^1 A^1$ qui est donc de la forme $\begin{pmatrix} -||a_1|| & * & \\ 0 & * & \\ \vdots & * & * \\ 0 & * & \end{pmatrix}$ d'après la proposition 4.

Étape k : On suppose qu'à la fin de l'étape $k-1$, on a obtenu A^k où les $k-1$ premières colonnes sont nulles sous la diagonale

$$A^k = \begin{pmatrix} * & & & \\ & \ddots & & * \\ & & A^k_{k,k} & \\ & & & \\ & 0 & \vdots & * \\ & & & \\ & & * & \end{pmatrix}.$$

On considère $a_k = \begin{pmatrix} A^k_{k,k} \\ \vdots \\ A^k_{n,k} \end{pmatrix} \in \mathbb{R}^{n-(k-1)}$.

— Si $a_k = \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, on pose $H^k = I_n$.

— Sinon, on pose $H^k = \begin{pmatrix} I_{k-1} & 0 \\ 0 & H_{a_k + ||a_k||e_1} \end{pmatrix}$, où e_1 est le premier vecteur de la base canonique de $\mathbb{R}^{n-(k-1)}$.

On définit $A^{k+1} = H^k A^k$, qui est bien de la forme souhaitée : la colonne k est nulle sous la diagonale.

En effet, la k ième-colonne de A^{k+1} vaut $\begin{pmatrix} A^k_{1,k} \\ A^k_{2,k} \\ \vdots \\ A^k_{k-1,k} \\ H_{a_k + ||a_k||e_1} a_k \end{pmatrix} = \begin{pmatrix} A^k_{1,k} \\ A^k_{2,k} \\ \vdots \\ A^k_{k-1,k} \\ -||a_k|| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, (et les $k-1$ premières colonnes sont inchangées).

À la fin de $n-1$ étapes, on a une matrice

$$A^n = H^n ... H^1 A$$

triangulaire supérieure, donc

$$A = \underbrace{\left(H^n ... H^1\right)^{-1}}_{=Q} \underbrace{A^n}_{=R},$$

avec Q orthogonale car chaque H^k l'est (voir proposition 4) et R triangulaire supérieure.

Pour imposer aux coefficients diagonaux de R d'être tous positifs ou nuls on pose D la matrice diagonale telle que $D_{i,i} = \pm 1$ pour que $\tilde{R} = D^{-1}R$ avec $\tilde{R}$ une matrice triangulaire supérieure à diagonale positive. Alors $\tilde{Q} = QD$ est encore une matrice orthogonale et $A = \tilde{Q}\tilde{R}$.

Remarque 17. *Nous n'avons pas encore utilisé le caractère inversible de la matrice A (il sera primordial pour l'uncité des matrices Q et R). La décomposition QR existe donc pour une matrice A non inversible (et non carrée).*

Démonstration de l'unicité de R et Q dans le cas où A est inversible. Si la matrice A est inversible, on peut imposer à la matrice R d'être à diagonale strictement positive et dans ce cas, R sera inversible. Soit deux telles décompositions Q_1R_1 et Q_2R_2 (avec R_1 et R_2 inversibles à diagonale sctrictement positive) :

$$A = Q_1R_1 = Q_2R_2.$$

On a $Q_2^TQ_1 = R_2R_1^{-1}$. Notons T cette matrice, on a

$$T^TT = Q_1^TQ_2Q_2^TQ_1 = I_n \text{ puisque } Q_i \text{ est orthogonale.}$$

Or R_i est triangulaire supérieure, donc T l'est aussi. De plus on a $(R_1)_{i,i} > 0$ et $(R_2)_{i,i} > 0$ donc $T_{i,i} = \frac{(R_2)_{i,i}}{(R_1)_{i,i}} > 0$. L'égalité précédente n'est donc rien d'autre qu'une décomposition de Cholesky de la matrice identité. Par unicité d'une telle décomposition, on a $T = I_n$ et donc $R_1 = R_2$. On en déduit alors facilement que $Q_1 = Q_2$.

$\square$

Résolution des systèmes linéaires

Méthodes itératives pour résoudre $Ax = b$

Cadre. Les méthodes directes donnent des solutions exactes de $Ax = b$ mais le calcul est souvent assez coûteux (par exemple si la matrice est pleine ou si elle est de grande taille). L'idée des méthodes itératives est de construire une suite de vecteurs $(x_k)_k \in \mathbb{K}^n$ qui converge vers la solution x de $Ax = b$.

Remarque 18. *Les méthodes itératives que nous étudierons ici sont toutes basées sur une décomposition de la matrice A en $A = M - N$ avec M une matrice inversible (et facilement inversible, par exemple M diagonalisable ou M triangulaire). On a alors*

$$\begin{aligned} Ax = b &\Leftrightarrow (M - N)x = b \\ &\Leftrightarrow Mx = Nx + b \\ &\Leftrightarrow x = M^{-1}Nx + M^{-1}b. \end{aligned}$$

Cette forme suggère d'étudier les points fixes de la fonction

$$y \in \mathbb{K}^n \mapsto M^{-1}Ny + M^{-1}b.$$

5 Principe des méthodes itératives

Définition 5 (Def 8.1.1 [AK]). *Soit $A \in \mathbb{M}_n(\mathbb{K})$ une matrice inversible (pour que la matrice M soit elle-aussi inversible).*
- *Une paire de matrices (M, N) avec M facilement inversible satisfaisant*

$$A = M - N$$

est appellée une décomposition régulière de A.
- *La méthode itérative basée sur la décomposition régulière $A = M - N$, $M, N \in \mathcal{M}_n(\mathbb{K})$ est donnée par la suite*

$$\begin{cases} x_0 \in \mathbb{K}^n \text{ quelconque}, \\ x_{k+1} = M^{-1}Nx_k + M^{-1}b. \end{cases} \tag{2}$$

- *La matrice $B = M^{-1}N$ est appellée matrice de la méthode itérative.*

Remarque 19. *Si la suite $(x_k)_k$ admet une limite x alors cette limite sera forcément solution de $Ax = b$ (par l'unicité de la solution puisque A est inversible).*

Définition 6. • *La méthode itérative est dite convergente si $\forall x_0 \in \mathbb{K}^n$, la suite $(x_k)_k$ définie par (2) converge.*
- *On appelle l'erreur à l'itération k la différence $e_k = ||x_k - x||$.*
- *On appelle le résidu à l'itération k la valeur $r_k = ||Ax_k - b||$.*

Une question à se poser en pratique est de savoir quand arrêter la méthode itérative (on parle de critère d'arrêt). En toute généralité, il faudrait arrêter la méthode lorsque l'erreur e_k est petite, mais on ne connait pas x ce qui rend ce critère d'arrêt inutilisable. La convergence en pratique est déterminée grâce au critère d'arrêt sur le résidu r_k.

Théorème 5. Critères de convergence. (Thm 8.1.1 [AK], Thm 5.1-1 [Cia], Prop 9.1.1 [Ser]). *Soit la méthode itérative donnée par la suite (2), de matrice $B = M^{-1}N$. Alors, les trois propositions suivantes sont équivalentes :*
- *la méthode itérative converge,*
- *le rayon spectral de la matrice B vérifie*

$$\rho(B) < 1,$$

- *on a $|||B||| < 1$ pour au moins une norme matricielle $|||\cdot|||$.*

Rappel. Le rayon spectral $\rho(A)$ d'une matrice A est défini par $\rho(A) = \max_{i \in [\![1,n]\!]} |\lambda_i|$ où les λ_i sont les valeurs propres complexes (comptées avec multiplicité) de la matrice A.

De plus le rayon spectral vérifie les deux inégalités suivantes : (Thm 1.4-3 [Cia], Prop 3.1.4 [AK], Cor 4.4.1 [Ser])

1. Pour toute matrice $A \in \mathcal{M}_n(\mathbb{K})$ et pour toute norme matricielle $|||\cdot|||$ sur $\mathcal{M}_n(\mathbb{K})$, on a $\rho(A) \leq |||A|||$.
2. Pour tout $\epsilon > 0$ et pour toute matrice $A \in \mathcal{M}_n(\mathbb{K})$, il existe une norme matricielle $|||\cdot|||$ (qui dépend de A) telle que $|||A||| \leq \rho(A) + \epsilon$.

Démonstration du rappel. • Soit $A \in \mathcal{M}_n(\mathbb{K})$ et $|||\cdot|||$ une norme matricielle. Si la norme est subordonnée à une norme vectorielle, alors $|||A||| = \sup_{x \in \mathbb{K}^n \setminus \{0\}} \frac{||Ax||}{||x||}$. Soit $\lambda \in \mathbb{K}^n$ tel que $|\lambda| = \rho(A)$ et soit u un vecteur propre associé. Alors on a $\rho(A)||u|| = ||\lambda u|| = ||Au|| \leq |||A|||.||u||$.
Si la norme est non subordonnée à une norme vectorielle. On prend $\lambda \in \mathbb{K}^n$ et u comme précédemment. On choisit aussi un vecteur $v \in \mathbb{K}^n$ tel que $uv^* \neq 0 \in \mathcal{M}_n(\mathbb{K})$. Alors $|\lambda||||uv^*||| = |||\lambda uv^*||| = |||Auv^*||| \leq |||A|||.|||uv^*|||$ (car $|||\cdot|||$ est une norme matricielle).

• Soit $A \in \mathcal{M}_n(\mathbb{K})$. Il existe U inversible tel que $U^{-1}AU = T$ où T est une matrice triangulaire avec $T_{i,i}$ les valeurs propres de A, pour $i \in \{1, ..., n\}$.

Soit $\delta > 0$, on définie la matrice D_δ diagonale suivante $D_\delta = \mathrm{diag}(1, \delta, \delta^2, ..., \delta^{n-1})$ et on montre que

$$(UD_\delta)^{-1}A(UD_\delta) = \begin{pmatrix} T_{1,1} & \delta T_{1,2} & & \dots & \delta^{n-1}T_{1,n} \\ & T_{2,2} & \delta T_{2,3} & \dots & \delta^{n-2}T_{2,n} \\ & & \ddots & & \vdots \\ & 0 & & \ddots & \vdots \\ & & & & T_{n,n} \end{pmatrix}.$$

Définissions la norme suivante :

$$|||A|||_\delta = ||(UD_\delta)^{-1}A(UD_\delta)||_\infty = \max_{i \in [\![1,n]\!]} \sum_{j=1}^{n} |(UD_\delta)^{-1}A(UD_\delta)|_{i,j}.$$

Soit $\epsilon > 0$ fixé, il ne nous reste plus qu'à choisir $\delta > 0$ tel que $\sum_{j=1}^{n} |\delta^{j-1}T_{i,j}| \leq \epsilon$ et on aura dans ce cas $|||A|||_\delta \leq \rho(A) + \epsilon$.

□

Démonstration du Théorème 5. Par définition, la méthode itérative converge si et seulement si, pour tout $x_0 \in \mathbb{K}^n$, nous avons $x_k \underset{k \to +\infty}{\longrightarrow} x$, avec nécessairement $Ax = b$ (soit $x = M^{-1}Nx + M^{-1}b$). C'est-à-dire, si et seulement si, pour tout $e_0 \in \mathbb{K}^n$ donné, $e_k = x_k - x \underset{k \to +\infty}{\longrightarrow} 0$.

Or $x_{k+1} = M^{-1}Nx_k + M^{-1}b$ ce qui implique que

$$x_{k+1} - x = M^{-1}Nx_k + M^{-1}b - x = M^{-1}Nx_k + M^{-1}b - M^{-1}Nx - M^{-1}b = M^{-1}N(x_k - x).$$

Ainsi, $e_k \underset{k \to +\infty}{\longrightarrow} 0$ si et seulement si $(M^{-1}N)^{k+1} \underset{k \to +\infty}{\longrightarrow} 0$ (pour une certaine norme matricielle) ce qui est équivalent aux deux propositions suivantes :
• $\rho(M^{-1}N) < 1$,
• $|||M^{-1}N||| < 1$ pour au moins une norme matricielle $|||\cdot|||$, d'après le rappel précédent.

□

Remarque 20. *Les méthode itératives reposent sur la recherche des points fixes de*

$$F : y \in \mathbb{K}^n \mapsto M^{-1}Ny + M^{-1}b$$

par la méthode des approximations successives (ou méthode de Picard). Nous savons que cette méthode de Picard converge si F est contractante ce qui revient à supposer $|||M^{-1}N||| < 1$ comme dans le théorème 5. Il n'est donc pas étonnant que la méthode itérative converge si et seulement si cette condition est respectée (voir p.96 [Cia]).

5.1 Condition suffisante de convergence pour les matrices hermitiennes définies positives

Dans certains cas, il n'est pas nécessaire de calculer le rayon spectral de la matrice B pour savoir s'il y a convergence ou pas. C'est le cas notamment des matrices hermitiennes définies positives, comme expliqué dans le théorème suivant.

Théorème 6 (Thm 8.1.2 [AK], Thm 5.3-1 [Cia], Lem 9.3.1 [Ser])**.** *Soit A une matrice hermitienne définie positive et M, N telles que $A = M - N$ avec M inversible. Alors, si la matrice $M^* + N$, qui est hermitienne, est aussi définie positive, la méthode itérative de matrice $M^{-1}N$ converge.*

Démonstration du théorème 6. Matrice hermitienne : La matrice $M^* + N$ est bien hermitienne car A est hermitienne, donc $A = A^*$ soit encore $(M - N)^* = M - N$. Donc $M^* - N^* = M - N$ donc $M^* + N = M + N^* = (M^* + N)^*$.

Condition du théorème 5 vérifiée : On montre que si $M^* + N$ est définie positive alors, il existe une norme matricielle $|||\cdot|||$ telle que $|||M^{-1}N||| < 1$.

Puisque A est hermitienne définie positive, on peut définir une norme sur $\mathbb{K}^n$ par $||x||_A = \sqrt{\langle Ax, x\rangle}$.

Soit $|||\cdot|||_A$ la norme matricielle subordonnée à $||\cdot||_A$. Par définition, on a donc $|||M^{-1}N|||_A = \sup_{||x||_A=1} ||M^{-1}Nx||_A$. Soit $x \in \mathbb{K}^n$ tel que $||x||_A = 1$ alors $\langle Ax, x\rangle = 1$ et on a

$$\begin{aligned}
||M^{-1}Nx||_A^2 &= \left\langle A(M^{-1}Nx), M^{-1}Nx\right\rangle \\
&= \left\langle A(M^{-1}(M-A)x), M^{-1}(M-A)x\right\rangle \quad \text{(définition de } N\text{)} \\
&= \left\langle Ax - AM^{-1}Ax, x - M^{-1}Ax\right\rangle \\
&= \langle Ax, x\rangle - \left\langle AM^{-1}Ax, x\right\rangle - \left\langle Ax, M^{-1}Ax\right\rangle + \left\langle AM^{-1}Ax, M^{-1}Ax\right\rangle.
\end{aligned}$$

En notant $v = M^{-1}Ax$ on a $Mv = Ax$ et donc les égalités précédentes deviennent

$$\begin{aligned}
||M^{-1}Nx||_A^2 &= 1 - \langle Av, x\rangle - \langle Mv, v\rangle + \langle Av, v\rangle \quad \text{(car } x \text{ est choisi tel que } \langle Ax, x\rangle = 1\text{)} \\
&= 1 - \langle v, Mv\rangle - \langle Mv, v\rangle + \langle Av, v\rangle \quad \text{(car } A \text{ est hermitienne donc } \langle Av, x\rangle = \langle v, Ax\rangle = \langle v, Mv\rangle\text{)} \\
&= 1 - \langle (M^* + M - A)v, v\rangle \\
&= 1 - \langle (M^* + N)v, v\rangle
\end{aligned}$$

Or $M^* + N$ est supposé définie positive donc pour $v \neq 0$ on a $\langle (M^* + N)v, v\rangle > 0$ donc

$$||M^{-1}Nx||_A^2 = 1 - \langle (M^* + N)v, v\rangle < 1.$$

Soit $|||M^{-1}N|||_A < 1$. Le théorème 5 sur les critères de convergence permet de conclure quant à la convergence de la méthode itérative. □

6 Méthodes itératives de Jacobi, Gauss-Seidel et relaxation

Remarque 21. *Ces trois méthodes se basent sur la décomposition de A en les trois matrices suivantes : Soit $A \in \mathcal{M}_n(\mathbb{K})$. On considère*

- *D la diagonale de A,*
- *$-E$ sa partie strictement triangulaire inférieure et*
- *$-F$ sa partie triangulaire strictement supérieure*

$$A = \begin{pmatrix} \ddots & & -F \\ & D & \\ -E & & \ddots \end{pmatrix}$$

telles que $A = D - E - F = D - (E + F)$

Définition 7 (Def 8.2.1, 8.2.2, 8.2.3 [AK], §9.2 [Ser], §5.2 [Cia]). *Soit $A \in \mathcal{M}_n(\mathbb{K})$ avec une décomposition en les matrices $D, -E, -F$ comme précédemment. Supposons de plus D inversible (par exemple, A inversible suffit).*

- *La méhode de Jacobi se base sur la décomposition*

$$A = D - (D - A) = \underbrace{D}_{M} - \underbrace{(E + F)}_{N}$$

(i.e on ne distingue pas séparément les matrices $-E$ et $-F$ ici). La matrice de la méthode est $J = D^{-1}(D - A) = I_n - D^{-1}A$.

- *La méthode de Gauss-Seidel est basée sur la décomposition*

$$A = \underbrace{(D - E)}_{M} - \underbrace{F}_{N}.$$

La matrice de la méthode est $G = (D - E)^{-1}F$.

- *La méthode de relaxation de paramètre $\omega \neq 0$ se base sur la décomposition*

$$A = \underbrace{\left(\frac{D}{\omega} - E\right)}_{M} - \underbrace{\left(\frac{1-\omega}{\omega} D + F\right)}_{N}.$$

La matrice de la méthode est donnée par $G_\omega = \left(\frac{D}{\omega} - E\right)^{-1} \left(\frac{1-\omega}{\omega} D + F\right)$.

Calcul pratique. Une itération pour calculer x_{k+1} à partide x_k dans ces méthodes est donnée par

$$\begin{aligned}
Dx_{k+1} &= (D - A)x_k + b \quad \text{(méthode de Jacobi)},\\
(D - E)x_{k+1} &= Fx_k + b \quad \text{(méthode de Gauss-Seidel)},\\
(\frac{D}{\omega} - E)x_{k+1} &= (\frac{1-\omega}{\omega} D + F)x_k + b \quad \text{(méthode de relaxation)}.
\end{aligned}$$

Remarque 22. *Si $\omega = 1$ dans la méthode de relaxation, on retrouve la méthode de Gauss-Seidel. Le choix de ω dans la méthode de relaxation est important puisque qu'il permet de minimiser $\rho(G_\omega)$ et ainsi d'avoir une méthode qui converge plus rapidement.*

Remarque 23. *Ces trois méthodes sont bien définies si la matrice diagonale D est inversible.*

Remarque 24. *La méthode de relaxation est aussi appelée méthode SOR pour "Successive OverRelaxation method".*

6.1 Cas d'une matrice A à diagonale strictement dominante

Proposition 5 (Prop 9.3.1 [Ser]). *Si A est à diagonale strictement dominante alors la méthode de Jacobi converge, de même que la méthode de relaxation pour $\omega \in]0,1]$ (ce qui inclut la méthode de Gauss-Seidel).*

Rappel. Une matrice A est dite à diagonale strictement dominante si et seulement si $|A_{i,i}| > \sum_{j \neq i} |A_{i,j}|$.

Démonstration de la proposition 5. Méthode de Jacobi : D'après le théorème 5 des critères de convergence, il suffit de montrer que

$$|||I_n - D^{-1}A|||_\infty = \max_{i \in [\![1,n]\!]} \sum_{j=1}^{n} |(I_n - D^{-1}A)_{i,j}| < 1.$$

Or, pour tout $(i,j) \in \{1, ..., n\}^2$, on a

$$(I_n - D^{-1}A)_{i,j} = \begin{cases} 1 - \sum_{k=1}^{n} D^{-1}_{i,k} A_{k,i} = 1 - D^{-1}_{i,i} A_{i,i} = 1 - \dfrac{A_{i,i}}{A_{i,i}} = 0, & \text{si } i = j, \\ -\sum_{k=1}^{n} D^{-1}_{i,k} A_{k,j} = -\dfrac{A_{i,j}}{A_{i,i}}, & \text{si } i \neq j. \end{cases}$$

On a donc
$$\sum_{j=1}^{n} |(I_n - D^{-1}A)_{i,j}| = \sum_{j=1, j\neq i}^{n} \left|\frac{A_{i,j}}{A_{i,i}}\right| = \frac{1}{|A_{i,i}|} \sum_{j=1, j\neq i}^{n} |A_{i,j}| < 1.$$
La dernière inégalité vient du fait que A soit à diagonale strictement dominante.

Méthode de Gauss-Seidel : Nous rappelons au préalable le théorème de Gerschgörin (Prop 4.5.1 [Ser])

Théorème 7. *Le spectre d'une matrice B est inclus dans la réunion des disques de Gerschgörin, qui sont les disques $\mathscr{D}_i$ de centre $b_{i,i}$ et de rayon $\sum_{j\neq i} |b_{i,j}|$. C'est-à-dire, que pour tout λ valeur propre de B, il existe $i_0 \in \{1, ..., n\}$ tel que*
$$|\lambda - b_{i_0,i_0}| \leq \sum_{j\neq i_0} |b_{i_0,j}|.$$

Pour démontrer la proposition 5 dans le cas de la décomposition de Gauss-Seidel, il suffit de montrer, d'après le théorème 5 des critères de convergence, que
$$\rho((D-E)^{-1}F) < 1.$$

Soit λ une valeur propre de $(D-E)^{-1}F$, alors par définition $\det(\lambda(D-E) - F) = 0$, soit encore $\det(\lambda D - (\lambda E + F)) = 0$. Ainsi, λ apparaît comme une valeur propre de la matrice $D^{-1}(\lambda E + F)$. Le coefficient d'indice (i,j) de cette matrice $D^{-1}(\lambda E + F)$ vaut
$$\begin{aligned}
\left(D^{-1}(\lambda E + F)\right)_{i,j} &= \sum_{k=1}^{n} D^{-1}_{i,k}(\lambda E_{k,j} + F_{k,j}) \\
&= D^{-1}_{i,i}(\lambda E_{i,j} + F_{i,j}) \quad \text{(puisque } D \text{ est une matrice diagonale)} \\
&= \begin{cases} \frac{1}{A_{i,i}}\left(-\lambda A_{i,j}\right), & \text{si } i > j, \\ \frac{1}{A_{i,i}}\left(-A_{i,j}\right), & \text{si } i < j, \\ 0, & \text{si } i = j. \end{cases} \quad \text{d'après les définitions des matrices } D, E, F.
\end{aligned}$$

D'après le théorème de Gerschgörin, rappelé précédemment, il existe donc $i_0 \in \{1, ..., n\}$ tel que
$$|\lambda - 0| \leq \sum_{j<i_0} \left| -\lambda \frac{A_{i_0,j}}{A_{i_0,i_0}} \right| + \sum_{j>i_0} \left| -\frac{A_{i_0,j}}{A_{i_0,i_0}} \right|.$$
Soit encore, en considérant tous les indices i des lignes
$$|\lambda| \leq \max_{i\in\{1,\dots,n\}} \left\{ \sum_{j<i} |\lambda| \left|\frac{A_{i,j}}{A_{i,i}}\right| + \sum_{j>i} \left|\frac{A_{i,j}}{A_{i,i}}\right| \right\}.$$

L'inégalité précédente fournit $|\lambda| \leq \max(|\lambda|, 1) \max_{i\in\{1,\dots,n\}} \left\{ \frac{1}{|A_{i,i}|} \sum_{j\neq i} |A_{i,j}| \right\}$.

Puisque A est supposée à diagonale strictement dominante, nous obtenons $|\lambda| < \max(|\lambda|, 1)$ ce qui n'est vrai que si $|\lambda| < 1$. Nous avons donc montré que toutes les valeurs propres de $(D-E)^{-1}F$ étaient de module strictement inférieur à 1, on a donc bien $\rho((D-E)^{-1}F) < 1$. □

Remarque 25. *La preuve de la proposition 5 dans le cas de la méthode de relaxation se démontre de la même manière que pour la méthode de Gauss-Seidel en introduisant le paramètre ω en plus (une preuve se trouve p.103 [Ser]).*

6.2 Cas d'une matrice A hermitienne définie positive

Proposition 6. *Si A est hermitienne, la méthode de Jacobi converge si A et $2D - A$ sont définies positives.*

Démonstration de la proposition 6. Si A est définie positive alors $D_{ii} > 0$ donc D est bien inversible (donc la méthode de Jacobi est bien définie). De plus, $M^* + N = D^* + D - A = 2D - A$. Le théorème 6 permet de conclure quant à la convergence de cette méthode. □

Proposition 7. *Si A est hermitienne définie positive, la méthode de Gauss-Seidel converge.*

Démonstration de la proposition 7. Nous allons utiliser le théorème 6.

Montrons tout d'abord que $D - E$ (qui est la matrice M de la méthode de Gauss-Seidel) est bien inversible. On a $(D - E)_{i,i} = D_{i,i} = A_{i,i} > 0$ car A est définie positive, donc $D - E$ est inversible (car triangulaire à diagonale sans 0).

Ensuite, montrons que $M^* + N = (D - E)^* + F$ est hermitienne définie positive. La matrice A est hermitienne donc $A^* = \bar{A}^T = A$ donc $(-E)^* = -F$ donc $E^* = F$. Ainsi

$$M^* + N = (D - E)^* + F = D^* - E^* + F = D$$

qui est définie positive car $D_{ii} = A_{ii} > 0$ et D est diagonale.

Les hypothèses du théorème 6 sont donc vérifiées et la méthode de Gauss-Seidel converge donc.

□

Remarque 26. *Il existe aussi un critère de convergence pour la méthode de relaxation sur les matrices hermitiennes définies positives : (Thm 8.2.1 [AK], Thm 5.3-2 [Cia], Thm 9.3.1 [Ser])*
Si A est hermitienne définie positive et si $\omega \in]0, 2[$ alors la méthode de relaxation converge.

Remarque 27. *Notons également le cas particulier des matrices tridiagonales (§9.4 [Ser], §8.3 [AK], Thm 5.3-4, Thm 5.3-5 [Cia])...*

Références

[AK] G. Allaire and S. M. Kaber. *Numerical Linear Algebra*, volume 55 of *Texts in Applied Mathematics*. Springer, 2008.

[Cia] P. G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimsation*. Dunod, 2006.

[Gou] X. Gourdon. *Algèbre*. Les maths en tête. Ellipses, 2009.

[Ser] D. Serre. *Les Matrices*. Dunod, 2001.