Séance V : Analyse numérique matricielle

A) Objectifs de la séance

A la fin de cette séance,

- Je sais caractériser une matrice symétrique définie positive.
- Je sais déterminer une région du plan contenant toutes les valeurs propres d'une matrice.
- Je sais programmer une méthode qui me donne le rayon spectral d'une matrice.
- Je connais la différence entre une méthode directe et une méthode itérative de résolution de système linéaire, et comment les utiliser.
- Je sais utiliser la notion de conditionnement d'une matrice.
- Je sais comment évaluer la complexité d'une méthode numérique.

B) Pour se familiariser avec les concepts (à traiter avant les séances de TD)

Les questions V.1 et V.2 sont à traiter avant la séance de TD 8. Les corrigés sont disponibles sur internet.

Question V.1 (Applications du théorème de Schur)
Soit $A \in \mathcal{M}_q(\mathbb{C})$, où $q \geq 1$.

Q. V.1.1 Rappeler le théorème de décomposition de Schur vu en cours.

Q. V.1.2 Montrer que A est normale, c'est-à-dire que $AA^* = A^*A$, si et seulement si il existe une matrice unitaire U et une matrice diagonale D contenant les valeurs propres de A telles que $A = UDU^*$.
INDICATION : On pourra calculer les éléments diagonaux de TT^* et T^*T où T est la matrice triangulaire supérieure telle que $A = UTU^*$ avec U unitaire.

Q. V.1.3 Montrer que si A est hermitienne, elle est diagonalisable dans une base de vecteurs propres orthogonaux et que ses valeurs propres sont réelles.

Q. V.1.4 Montrer que si A est unitaire, elle est diagonalisable dans une base de vecteurs propres orthogonaux et que ses valeurs propres ont pour module 1.

Question V.2 (Résolution de systèmes triangulaires)
Soit $b \in \mathbb{K}^q$, où $q \geq 1$.

Q. V.2.1 Soit $L \in T_{q,inf}(\mathbb{K})$. Ecrire l'algorithme de résolution du système linéaire $Lx = b$.

Q. V.2.2 Soit $U \in T_{q,sup(\mathbb{K})}$. Ecrire l'algorithme de résolution du système linéaire $Ux = b$.

Q. V.2.3 Calculer le nombre d'opérations nécessitées par ces résolutions.

C) Exercices

Exercice V.1 (Approche intuitive de la résolution de systèmes linéaires)
Soit $A \in GL_q(\mathbb{R})$. Soit $b \in \mathbb{R}^q$. Quelle est la complexité de la résolution du système linéaire $A^2x = b$ par une méthode directe ? Proposer une méthode moins coûteuse.

Exercice V.2 (Décomposition LU d'une matrice particulière)

Soit $A \in \mathcal{M}_q(\mathbb{R})$ une matrice tridiagonale, définie par :

$$A = \begin{pmatrix} 2+c_1 & -1 & & & & & \\ -1 & 2+c_2 & -1 & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2+c_{q-1} & -1 & \\ & & & & -1 & 2+c_q & \end{pmatrix},$$

avec $c_i \geq 0, \quad \forall 1 \leq i \leq q$.

E. V.2.1 Soit $v \in \mathbb{R}^q$. Montrer que :

$$v^T A v = \sum_{i=1}^{q} c_i v_i^2 + \left\{ v_1^2 + v_q^2 + \sum_{i=2}^{q} (v_i - v_{i-1})^2 \right\}.$$

E. V.2.2 En déduire que A admet une décomposition de Cholesky.

E. V.2.3 Dans le cas $c_i = 0$, $(1 \leq i \leq q)$, montrer que la matrice B telle que $A = BB^T$ est bidiagonale et donnée par :

$$\begin{cases} B_{i,i} &= \sqrt{\dfrac{i+1}{i}} \quad (1 \leq i \leq q); \\ B_{i+1,i} &= -\sqrt{\dfrac{i}{i+1}} \quad (1 \leq i \leq q-1). \end{cases}$$

Exercice V.3 (Décomposition QR)
Soit $A \in GL_q(\mathbb{R})$.

E. V.3.1 Montrer qu'il existe une matrice R triangulaire supérieure à diagonale strictement positive telle que :

$$A^T A = R^T R.$$

E. V.3.2 En déduire qu'il existe une matrice orthogonale $Q \in O_q(\mathbb{R})$ telle que : $A = QR$.

E. V.3.3 Montrer que cette décomposition $A = QR$ avec R triangulaire supérieure à diagonale strictement positive et Q orthogonale est unique.

Exercice V.4
Soit $\omega \in \mathbb{R} \setminus \{-1\}$. Soit $A \in \mathcal{M}_q(\mathbb{R})$ une matrice inversible telle que

$$A = (1+\omega)P - (N + \omega P),$$

avec P inversible et $P^{-1}N$ de valeurs propres réelles $1 > \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_q$.

E. V.4.1 Trouver les valeurs de ω pour lesquelles la méthode itérative suivante x_0 vers la solution du système $Ax = b$.

E. V.4.2 Trouver la valeur de ω pour laquelle le taux de convergence est minimal, ce qui permet d'optimiser la vitesse de convergence.

D) Approfondissement

Ces exercices sont disponibles sous forme de Jupyter notebooks sur edunao.

Chapitre VIII : Corrections des exercices

Solution de Q. V.1.1 Let $A \in \mathcal{M}_q$. Then, there exist

- an upper triangular matrix T
- a unitary matrix U

such that $A = UTU^{-1}$.

Solution de Q. V.1.2 Appliquons la décomposition de Schur : soient $U \in U_q(\mathbb{C})$ et $T \in T_{q,sup}(\mathbb{C})$ telles que $A = UTU^*$. Alors $AA^* = UTU^*UT^*U^* = UTT^*U^*$ et $A^*A = UT^*TU^*$.

- Montrons la réciproque : supposons que $T = D$ avec D diagonale. On a $D^* = \overline{D}$. Alors $AA^* = UD\overline{D}U^*$ et $A^*A = U\overline{D}DU^*$. On en conclut que A est normale.
- Montrons le sens direct : supposons que A est normale. Examinons les éléments diagonaux de $TT^* = T^*T$. On a, pour tout $i \in \{1, \dots, q\}$,

$$(TT^*)_{ii} = \sum_{k=1}^{q} (T)_{ik}(T^*)_{ki} = \sum_{k=1}^{q} |T_{ik}|^2 = \sum_{k=i}^{q} |T_{ik}|^2$$
$$(T^*T)_{ii} = \sum_{k=1}^{q} (T^*)_{ik}(T)_{ki} = \sum_{k=1}^{q} |T_{ki}|^2 = \sum_{k=1}^{i} |T_{ki}|^2$$

Raisonnons par récurrence sur i pour montrer que les éléments extra-diagonaux de T sont nuls jusqu'à la ligne i :

- pour $i = 1$, on a :
$$\sum_{k=1}^{q} |T_{1k}|^2 = |T_{11}|^2.$$
D'où $\sum_{k=2}^{q} |T_{1k}|^2 = 0$ et $\forall k \geq 2$, $T_{1k} = 0$.
- Hypothèse de récurrence : supposons que, pour $i \geq 1$, pour tout $p \leq i$, $k \neq p$, $T_{pk} = 0$.
Montrons que la propriété est vraie au rang $i+1$, c'est-à-dire que, pour tout $k \leq i$, $T_{i+1,k} = 0$:
On a, sachant que, pour tout $k \neq i+1$, $T_{k,i+1} = 0$,
$$\sum_{k=i+1}^{q} |T_{i+1,k}|^2 = \sum_{k=1}^{i+1} |T_{k,i+1}|^2 = |T_{i+1,i+1}|^2$$
d'où $\sum_{k=i+2}^{q} |T_{i+1,k}|^2 = 0$. On en conclut que, pour tout $k \neq i+1$, $T_{i+1,k} = 0$.
- Conclusion : on a bien montré que la matrice T est diagonale. On note $T = D$.

Montrons que D contient les valeurs propres de A. Pour tout $i \in \{1, \dots, q\}$, notons v_i la i-ème colonne de la matrice U, c'est-à-dire $v_i = Ue_i$, avec e_i le i-ème vecteur canonique de $\mathbb{C}^q$. Rappelons que U unitaire implique $v_i^* v_j = \delta_{ij}$. Alors

$$Av_i = UDU^* v_i = UD \begin{pmatrix} v_1^* v_i \\ \vdots \\ v_i^* v_i \\ \vdots \\ v_q^* v_i \end{pmatrix} = UDe_i = D_{ii} Ue_i = D_{ii} v_i.$$

Les vecteurs v_i sont donc des vecteurs propres de A associés à D_{ii} valeur propre et les vecteurs v_i forment une base orthonormale de $\mathbb{C}^q$, car ce sont les colonnes de U. On a donc trouvé exactement q vecteurs propres formant une base de $\mathbb{C}^q$ et q valeurs propres associées : on en conclut que $\mathrm{Sp}(A) = \{D_{ii}, i \in \{1, \dots, q\}\}$.

Solution de Q. V.1.3 Si A est hermitienne, alors $A = A^*$ et donc A est normale. D'après la question V.1.2, il existe $D \in diag(\mathbb{C})$ et $U \in U_q(\mathbb{C})$ telles que $A = UDU^*$. De plus, comme $A = A^*$, $D = U^* AU = U^* UD^* U^* U = D^*$. On en conclut que les coefficients diagonaux de D sont leurs propres conjugués et sont donc réels !

Solution de Q. V.1.4 Si A est unitaire, alors $AA^* = I_q = A^* A$, donc A est normale. D'après la question V.1.2, A est diagonalisable dans une base de vecteurs propres orthogonaux car il existe $U \in U_q(\mathbb{C})$ et D diagonale telles que $A = UDU^*$. Comme $AA^* = I_q$, on a $DD^* = I_q$ et on a donc pour tout $i \in \{1, \dots, q\}$, $|D_{ii}|^2 = 1$, d'où $|D_{ii}| = 1$: le spectre de A est contenu dans le cercle-unité.

Solution de Q. V.2.1 On prend comme convention qu'un vecteur d'indices vides est nul : $x(1:0) = 0$.

Supposons que L est inversible. Soit x la solution du système $Lx = b$. Alors

$$\begin{aligned} L_{11} x_1 &= b_1 \\ &\vdots \\ \sum_{k=1}^{i} L_{ik} x_k &= b_i \\ &\vdots \\ \sum_{k=1}^{q} L_{qk} x_k &= b_q \end{aligned}$$

On peut donc calculer simplement x par récurrence sur i car

$$\forall i \in \{1, \dots, q\}, \quad L_{ii} x_i = b_i - \sum_{k=1}^{i-1} L_{ik} x_k.$$

L'algorithme s'écrit donc

Algorithme 1 : Résolution d'un système triangulaire inférieur

Donnée : L, b

Paramètre : $q = size(A, 1)$

1 **Pour** i=1 **à** q **faire**
2 **si** $L(i,i) \neq 0$ **alors**
3 $x(i) \leftarrow \dfrac{b(i) - L(i, 1:i-1) * x(1:i-1)}{L(i,i)}$
4 **sinon**
5 abandon : la matrice L n'est pas inversible
6 **finsi**
7 **finpour**

Résultat : x

Solution de Q. V.2.2 La seule différence de raisonnement avec la question précédente est le fait qu'on doive partir de la dernière composante de x !

Algorithme 2 : Résolution d'un système triangulaire supérieur

Donnée : U, b

Paramètre : $q = size(A, 1)$

1 **Pour** i=q **à** 1 **faire**
2 **si** $U(i,i) \neq 0$ **alors**
3 $x(i) \leftarrow \dfrac{b(i) - U(i, i+1:q) * x(i+1:q)}{U(i,i)}$
4 **sinon**
5 abandon : la matrice U n'est pas inversible !
6 **finsi**
7 **finpour**

Résultat : x

Solution de Q. V.2.3 On fait le calcul pour L triangulaire inférieure, le calcul étant le même pour une matrice triangulaire supérieure.
La ligne 3 nécessite

- $i-1$ multiplications et $i-1$ additions (en fait, $i-2$ additions et une soustraction)
- une division.

En sommant ces opérations sur i, on obtient : $q(q-1)/2$ additions, $q(q+1)/2$ multiplications (en fait, $q(q-1)/2$ multiplications et q divisions). La complexité est donc en $O(q^2)$ (quadratique).

Solution de Q. V.1 Soit $A = PLU$ avec P matrice de permutation. Le coût de la multiplication par la matrice P est linéaire, si on fait bien le produit avant de résoudre le système : $PLUx = b$ se résout en $Pz = b$ (linéaire car il suffit de changer les coefficients de place), puis $Ly = z$ et enfin $Ux = y$, ces deux opérations étant en $O(q^2)$.
Le calcul de la matrice A^2 nécessite q^2 fois $q-1$ additions et q multiplications, soit de l'ordre de $2q^3$ opérations. Notons N_{mult} ce nombre d'opérations. La complexité de la décomposition PLU est de l'ordre de q^3 également. Notons N_{PLU} le nombre d'opérations requis. On a montré dans l'exercice V.2 que la résolution des deux systèmes triangulaires a une complexité de l'ordre de q^2, que nous notons N_{triang}. En calculant A^2 puis en décomposant $A^2 = P_1L_1U_1$ puis en résolvant $P_1L_1U_1x = b$, on a donc besoin de faire $N_{mult} + N_{LU} + N_{triang} = O(q^3)$ opérations.

En faisant une décomposition de $A = P_2L_2U_2$ puis en résolvant $P_2L_2U_2L_2U_2x = b$, on n'a besoin que de faire $N_{LU} + 2N_{triang}$. Le calcul nécessite le même ordre de grandeur d'opérations, mais, la multiplication matricielle étant aussi chère que la décomposition PLU, la deuxième méthode est plus économique, en moyenne le temps de calcul est divisé par 2.

```
clear
temps=[]; q=50;

for n=1:500
A=rand(q); b=rand(q,1);
t1=tic; B=A*A; [L1,U1,P1]=lu(B); z1=P1*b; y1=L1\z1; x1=U1\y1; tmult=toc(t1);
t2=tic; [L2,U2,P2]=lu(A); zz2=P2*b; yy2=L2\zz2; xx2=U2\yy2; z2=P2*xx2; y2=L2\z2; x2=U2\y2;
tplu=toc(t2);
t3=tic; x=B\b; tml=toc(t3);
temps=[temps;[tmult tplu tml]];
end
```

Solution de Q. V.2.1 **Ceci est un calcul classique qui sera très utile par la suite !**

Calculons $v^T Av$, en posant $v_0 = 0$ et $v_{q+1} = 0$:

$$\begin{aligned} v^T Av &= \sum_{i=1}^{q} v_i (Av)_i = \sum_{i=1}^{q} v_i \left(\sum_{j=1}^{q} A_{ij} v_j \right) \\ &= v_1(A_{1,1}v_1 + A_{1,2}v_2) + \sum_{i=2}^{q-1} v_i(A_{i,i-1}v_{i-1} + A_{i,i}v_i + A_{i,i+1}v_{i+1}) + v_q(A_{q,q-1}v_{q-1} + A_{q,q}v_q) \\ &= \sum_{i=1}^{q} v_i(-v_{i-1} + (2+c_i)v_i - v_{i+1}) = \sum_{i=1}^{q} c_i v_i^2 + \sum_{i=1}^{q} v_i(v_i - v_{i-1}) + \sum_{i=1}^{q} v_i(v_i - v_{i+1}) \\ &= \sum_{i=1}^{q} c_i v_i^2 + \sum_{i=1}^{q} v_i(v_i - v_{i-1}) + \sum_{i=2}^{q+1} v_{i-1}(v_{i-1} - v_i) \\ &= \sum_{i=1}^{q} c_i v_i^2 + \sum_{i=2}^{q} (v_i - v_{i-1})^2 + v_1^2 + v_q^2. \end{aligned}$$

Solution de Q. V.2.2 La matrice A est clairement symétrique. On déduit du calcul précédent que A est symétrique définie positive (SDP) puisque, pour tout $v \in \mathbb{R}^q$, $v^T Av \geq 0$, et, $v^T Av = 0$ implique que $v_1 = v_q = 0$ et $\forall i \in \{2, \ldots, q\}$, $v_i = v_{i-1} = v_1$ donc $v = 0$. D'après le théorème du cours, A admet donc une décomposition de Cholesky.

Solution de Q. V.2.3 Il suffit de faire le calcul de $(BB^T)_{i,j}$ pour $i \geq j$, car BB^T est symétrique. On a immédiatement $(BB^T)_{1,1} = (B_{1,1})^2 = 2$.
Soient $i \in \{2, \ldots, q\}$ et $j \geq i$.
Alors Since BB^T is symmetric, we can simply compute $(BB^T)_{i,j}$ for $i \geq j$.
$(BB^T)_{1,1} = (B_{1,1})^2 = 2$ is obvious.
Let $i \in \{2, \ldots, q\}$ and $j \geq i$, then

$$(BB^T)_{i,j} = \sum_{k=1}^{q} B_{i,k}B_{j,k} = B_{i,j-1}B_{j,j-1} + B_{i,j}B_{j,j}.$$

- si $j - i \geq 2$, $(BB^T)_{i,j} = 0$,
- si $i \leq q-1$ et $j = i+1$, $(BB^T)_{i,i+1} = B_{i,i}B_{i+1,i} = -\sqrt{\frac{i+1}{i}}\sqrt{\frac{i}{i+1}} = -1$,
- si $i = j$, $(BB^T)_{i,i} = B_{i,i-1}^2 + B_{i,i}^2 = \frac{i-1}{i} + \frac{i+1}{i} = 2$.

Par unicité de la décomposition de Cholesky, la matrice B fournie est celle de la décomposition.

Solution de Q. V.3.1 **Ceci est un raisonnement très important !**

Montrons que la matrice A^TA est SDP. Elle est clairement symétrique. Soit $v \in \mathbb{R}^q$. Notons que $v^TA^TAv = (Av)^TAv = \|Av\|_2^2$. Donc $v^TA^TAv > 0$ si $Av \neq 0$. Or A est inversible donc $Av = 0$ est équivalent à $v = 0$. La matrice A^TA est donc bien SDP. D'après le théorème du cours, il existe donc $R \in T_{q,sup}$ à diagonale strictement positive telle que $A^TA = R^TR$.

Solution de Q. V.3.2 Posons $Q = AR^{-1}$. Montrons que Q est orthogonale, c'est-à-dire $Q^TQ = I_q$:

$$Q^TQ = (R^{-1})^TA^TAR^{-1} = (R^{-1})^TR^TRR^{-1} = I_q.$$

Solution de Q. V.3.3 Supposons qu'il existe deux couples de matrices (Q_1, R_1) et (Q_2, R_2) permettant une décomposition QR.
Alors $A = Q_1R_1 = Q_2R_2$ implique que $Q_1^TQ_2 = R_1R_2^{-1}$: il y a donc égalité entre une matrice orthogonale ($O_q(\mathbb{R})$ est stable par multiplication) et une matrice triangulaire supérieure à coefficients diagonaux strictement positifs. On en déduit que c'est la matrice identité : $Q_1 = Q_2$ et $R_1 = R_2$. Il y a donc unicité de la décomposition QR.

Solution de Q. V.4.1 On suppose bien entendu que $\omega \neq -1$.
La matrice d'itération de la méthode est

$$\mathcal{M}(\omega) = \frac{1}{1+\omega}P^{-1}(N + \omega P) = \frac{1}{1+\omega}(P^{-1}N + \omega I_q).$$

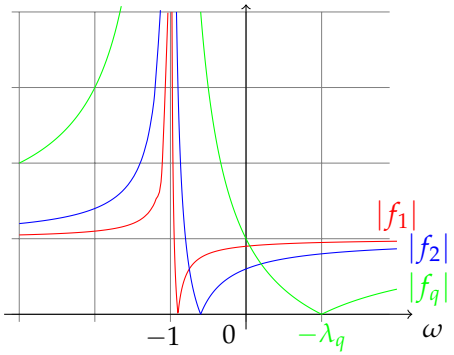
On a vu qu'une méthode itérative converge si et seulement si $\rho(\mathcal{M}(\omega)) < 1$.
Étudions le spectre de $\mathcal{M}(\omega)$:

$$\begin{aligned}\mu \in \mathrm{Sp}(\mathcal{M}(\omega)) &\iff \mu \in \mathrm{Sp}\left(\frac{1}{1+\omega}(P^{-1}N + \omega I_q)\right)\\ &\iff \mu \in \left\{\frac{\lambda_i + \omega}{1+\omega},\ i \in \{1,\dots,q\}\right\}\end{aligned}$$

A ω fixé, la méthode converge donc si et seulement si

$$\max_{i\in\{1,\dots,q\}} \left|\frac{\lambda_i + \omega}{1+\omega}\right| < 1.$$

Pour $i \in \{1,\dots,q\}$, notons $f_i : \omega \mapsto (\lambda_i + \omega)/(1+\omega)$. Il est clair que, pour tout $i \in \{1,\dots,q\}$, $|f_i|_{]-\infty,-1[} > 1$.



De plus,

$$\{\omega \in]-1,+\infty[/ |f_i(\omega)| < 1\} = \left]-\frac{1+\lambda_i}{2},+\infty\right[.$$

On en conclut donc que

$$\{\omega \in]-1,+\infty[:\ \rho(\mathcal{M}(\omega)) < 1\} = \bigcap_{1\leq i\leq q} \left]-\frac{1+\lambda_i}{2},+\infty\right[= \left]-\frac{1+\lambda_q}{2},+\infty\right[.$$

Solution de Q. V.4.2 Notons que

$$\rho(\mathcal{M}) : \omega \mapsto \begin{cases} |f_q(\omega)| & \text{si } \omega \in]-\infty,\omega_0[\backslash\{-1\} \\ |f_1(\omega)| & \text{si } \omega \in]\omega_0,+\infty[\end{cases}$$

avec ω_0 la valeur pour laquelle les fonctions f_1 et f_q sont égales, satisfaisant à

$$|\lambda_1+\omega_0| = \lambda_1+\omega_0 = |\lambda_q+\omega_0| = -\lambda_q-\omega_0$$

c'est-à-dire

$$\omega_0 = -\frac{\lambda_1+\lambda_q}{2}.$$

Le taux minimal de convergence vaut donc

$$\rho(\mathcal{M}(\omega_0)) = \frac{\lambda_1-\lambda_q}{2-(\lambda_1+\lambda_q)}.$$