Cours–TD : Analyse matricielle

Dans tout ce qui suit, $n \in \mathbb{N}^*$, et $\mathbb{K}$ désigne $\mathbb{R}$ ou $\mathbb{C}$.

1 Normes matricielles

1.1 Normes sur $\mathbb{K}^n$

Rappel : si E est un e.v.n de dimension finie, toutes les normes sur E sont équivalentes, i.e. pour deux normes N, N' sur E, il existe $a, b > 0$ tel que pour tout $x \in E$

$$aN(x) \leq N'(x) \leq bN(x). \tag{1}$$

On dit que a est une constante *optimale* si $aN(x) \leq N'(x)$ pour tout $x \in E$, et s'il existe $x_0 \in E\backslash\{0\}$ pour lequel on a égalité. Idem pour b (avec un x_0 a priori différent).

Question 1.1. *On suppose que $E = \mathbb{K}^n$. Trouver les constantes a, b optimales de (1) lorsque*

1. $N(x) = ||x||_1 = \sum_{i=1}^n |x_i|$ *et* $N'(x) = ||x||_\infty = \max_{1\leq i\leq n} |x_i|$,
2. $N(x) = ||x||_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$ *et* $N'(x) = ||x||_\infty$.

Note : a et b ne doivent pas dépendre de x, mais peuvent dépendre de n !

Q1.1/ 1. Soit $x \in E$. Alors

$$||x||_\infty = \max_{1\leq i\leq n} |x_i| \leq \sum_{i=1}^n |x_i| = 1\,||x||_1,$$

et pour $x_0 = (1, 0, \dots, 0)$, on a égalité car $||x_0||_\infty = 1 = ||x_0||_1$. De plus,

$$||x||_1 = \sum_{i=1}^n |x_i| \leq \sum_{i=1}^n \max_{1\leq j\leq n} |x_j| = n \max_{1\leq j\leq n} |x_j| = n\,||x||_\infty,$$

et pour $x_0 = (1, \dots, 1)$, on a égalité car $\frac{1}{n}||x_0||_2 = \frac{n}{n} = 1 = ||x_0||_\infty$. Ainsi, pour tout $x \in E$, on a

$$\frac{1}{n}||x||_1 \leq ||x||_\infty \leq ||x||_1,$$

avec des constantes optimales.

Q1.1/ 2. Soit $x \in E$. Alors

$$||x||_\infty = \max_{1\leq i\leq n} |x_i| = \sqrt{\left(\max_{1\leq i\leq n} |x_i|\right)^2} \leq \sqrt{\sum_{i=1}^n |x_i|^2} = ||x||_2,$$

et pour $x_0 = (1, 0, \ldots, 0)$, on a égalité car $||x_0||_\infty = 1 = ||x_0||_2$. De plus,

$$||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \leq \sqrt{\sum_{i=1}^{n} \left(\max_{1\leq i\leq n} |x_i|\right)^2} \leq \sqrt{n} \max_{1\leq i\leq n} |x_i| = \sqrt{n}||x||_\infty,$$

et pour $x_0 = (1, \ldots, 1)$, on a égalité car $\frac{1}{\sqrt{n}}||x_0||_2 = 1 = ||x_0||_\infty$. Ainsi, pour tout $x \in E$, on a

$$\frac{1}{\sqrt{n}}||x||_2 \leq ||x||_\infty \leq ||x||_2,$$

avec des constantes optimales.

Question 1.2. *On suppose que $E = \mathbb{K}^n$. Montrer que pour tout $x \in E$,*

$$||x||_2 \leq ||x||_1 \leq \sqrt{n}||x||_2,$$

et que ces constantes sont optimales.

Indication : pour la deuxième inégalité, on pourra utiliser l'inégalité de Cauchy-Schwarz.

Q1.2/ Pour tout $x \in E$,

$$||x||_2^2 = \sum_{i=1}^{n} |x_i|^2 \leq \left(\sum_{i=1}^{n} |x_i|\right)^2 = ||x||_1^2,$$

d'où la première inégalité en passant à la racine carrée. Par ailleurs, on a égalité pour $x = (1, 0, \ldots, 0)$.

Pour la seconde inégalité, on remarque que

$$(x, y) = \sum_{i=1}^{n} x_i \overline{y_i}$$

est un produit scalaire sur E. Ainsi, par l'inégalité de Cauchy-Schwarz, on a pour tous $x, y \in E$

$$|(x, y)| \leq \sqrt{(x, x)}\sqrt{(y, y)} = ||x||_2||y||_2.$$

Soit $x \in E$. On définit $y \in E$ tel que pour tout $i \in [\![1, n]\!]$,

$$y_i = \begin{cases} \frac{x_i}{|x_i|} & \text{si } x_i \neq 0, \\ 0 & \text{si } x_i = 0. \end{cases}$$

de sorte que $x_i\overline{y_i} = |x_i|$. On en déduit que

$$|(x, y)| = \left|\sum_{i=1}^{n} x_i \overline{y_i}\right| = \left|\sum_{i=1}^{n} |x_i|\right| = ||x||_1 \leq ||x||_2||y||_2.$$

Or, comme $|y_i| \in \{0, 1\}$, on a

$$||y||_2 = \sqrt{\sum_{i=1}^{n} |y_i|^2} \leq \sqrt{\sum_{i=1}^{n} 1} = \sqrt{n},$$

d'où la seconde inégalité. Par ailleurs, on a égalité pour $x = (1, 1, \ldots, 1)$:

$$||x||_1 = \sum_{i=1}^{n} |1|^2 = n = \sqrt{n}\sqrt{n} = \sqrt{n}\sqrt{\sum_{i=1}^{n} |1|^2} = \sqrt{n}||x||_2.$$

Donc les constantes sont optimales.

1.2 Normes matricielles

Définition. Soit $||\cdot||$ une norme sur $M_n(\mathbb{K})$. On dit que c'est une norme *matricielle* si pour toutes matrices $A, B \in M_n(\mathbb{K})$

$$||AB|| \leq ||A||\,||B||.$$

Question 1.3. *Montrer que*

$$||A|| = \max_{1\leq i,j\leq n} |a_{i,j}|$$

définit une norme sur $M_n(\mathbb{K})$. Montrer que ce n'est PAS une norme matricielle.

Q1.3/ Pour toutes matrices $A, B \in M_n(\mathbb{K})$ et $\lambda \in \mathbb{K}$, on a de manière évidente $||\lambda A|| = |\lambda|\,||A||$ et

$$\begin{aligned}
||A+B|| &= \max_{1\leq i,j\leq n} |a_{i,j} + b_{i,j}| \\
&\leq \max_{1\leq i,j\leq n} (|a_{i,j}| + |b_{i,j}|) \qquad\qquad \text{(car } |\cdot| \text{ est une norme sur } \mathbb{K}\text{)} \\
&\leq \max_{1\leq i,j\leq n} |a_{i,j}| + \max_{1\leq i,j\leq n} |b_{i,j}| \\
&= ||A|| + ||B||.
\end{aligned}$$

Par ailleurs, $||A|| = 0$ implique $a_{i,j} = 0$ pour tous i, j, donc $A = 0$. Ainsi, $||\cdot||$ est bien une norme sur $M_n(\mathbb{K})$. En revanche, si $a_{i,j} = 1$ pour tous i, j, on a $A^2 = nA$ et donc

$$||A^2|| = ||nA|| = |n|\,||A|| > ||A||\,||A|| = 1 \times 1 = 1.$$

Ainsi, $||\cdot||$ n'est pas une norme matricielle.

Définition. Soit $||\cdot||_{\mathbb{K}}$ une norme sur $\mathbb{K}^n$. On peut définir une norme sur $M_n(\mathbb{K})$, dite norme subordonnée à $||\cdot||_{\mathbb{K}}$ par

$$||A|| = \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{||Ax||_{\mathbb{K}}}{||x||_{\mathbb{K}}} \qquad \left(= \sup_{x\in\mathbb{K}^n, ||x||_{\mathbb{K}}=1} ||Ax||_{\mathbb{K}} = \sup_{x\in\mathbb{K}^n, ||x||_{\mathbb{K}}\leq 1} ||Ax||_{\mathbb{K}} \right).$$

Question 1.4. *Soit $||\cdot||$ une norme subordonnée sur $M_n(\mathbb{K})$. Montrer que $||I_n|| = 1$ et que $||\cdot||$ est une norme matricielle.*

Q1.4/ On remarque que

$$||I_n|| = \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{||I_n x||_{\mathbb{K}}}{||x||_{\mathbb{K}}} = \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{||x||_{\mathbb{K}}}{||x||_{\mathbb{K}}} = \sup_{x\in\mathbb{K}^n\setminus\{0\}} 1 = 1.$$

Par ailleurs, pour toutes matrices $A, B \in M_n(\mathbb{K})$ et $x \in \mathbb{K}^n\setminus\{0\}$, on a

$$||ABx|| \leq ||A||\,||Bx|| \leq ||A||\,||B||\,||x||$$

d'où

$$||AB|| = \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{||ABx||_{\mathbb{K}}}{||x||_{\mathbb{K}}} \leq ||A||\,||B||.$$

Donc c'est une norme matricielle.

Dans la suite, on notera $||\cdot||_1$ (resp. $||\cdot||_2$, resp. $||\cdot||_\infty$) la norme sur $M_n(\mathbb{K})$ qui est subordonnée à la norme $||\cdot||_1$ (resp. $||\cdot||_2$, resp. $||\cdot||_\infty$) sur $\mathbb{K}^n$. Il y a donc un abus de notation pour lequel il faudra être vigilant.

Remarque. Comme $M_n(\mathbb{K})$ est de dimension finie, toutes ces normes subordonnées sont équivalentes entre elles.

Question 1.5. *Déduire de la question 1.2 que pour toute matrice $A \in M_n(\mathbb{K})$, on a*

$$\frac{1}{\sqrt{n}}||A||_2 \leq ||A||_1 \leq \sqrt{n}||A||_2,$$

et que ces constantes sont optimales (on pourra utiliser les expressions explicites des questions 1.1 et 1.2).

Q1.5/ Par la question 1.2, on a $||x||_2 \leq ||x||_1 \leq \sqrt{n}||x||_2$ pour tout $x \in \mathbb{K}^n$. Donc

$$||Ax||_1 \leq \sqrt{n}||Ax||_2,$$

et pour $x \neq 0$

$$\frac{1}{||x||_1} \leq \frac{1}{||x||_2},$$

donc en multipliant ces inégalités terme à terme et en passant au $\sup_{x\in\mathbb{K}^n\setminus\{0\}}$, on a

$$\sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{||Ax||_1}{||x||_1} \leq \sqrt{n} \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{||Ax||_2}{||x||_2},$$

c'est-à-dire $||A||_1 \leq \sqrt{n}||A||_2$. Pour montrer l'optimalité, posons

$$A = \begin{pmatrix} 1 & & \\ \vdots & & 0 \\ 1 & & \end{pmatrix} \Rightarrow A^*A = \begin{pmatrix} n & & \\ & & 0 \\ & & \end{pmatrix}.$$

On remarque que $||A||_1 = n$ et que $\rho(A^*A) = n$, si bien que $||A||_2 = \sqrt{n}$. D'où l'optimalité.

L'autre inégalité se déduit de même puisque

$$||Ax||_2 \leq ||Ax||_1,$$

$$\frac{1}{\sqrt{n}||x||_2} \leq \frac{1}{||x||_1}.$$

D'où $\frac{1}{\sqrt{n}}||A||_2 \leq ||A||_1$. Pour montrer l'optimalité, on remarque que

$$A = \begin{pmatrix} 1 & \cdots & 1 \\ & 0 & \end{pmatrix} \Rightarrow A^*A = \begin{pmatrix} & 1 & \end{pmatrix}.$$

D'une part, $||A||_1 = 1$, et d'autre part $\rho(A^*A) = n$. En effet, $(1, \ldots, 1)$ est un vecteur propre associé à n, tandis que 0 est valeur propre de multiplicité $n-1$ puisque $\mathrm{rg}(A^*A) = 1$. D'où l'optimalité.

On peut de même utiliser la question 1.1 pour déduire les constantes optimales pour les normes subordonnées associées.

Question 1.6. *Montrer que (ne faire qu'une seule norme, celle de votre choix)*

$$||A||_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{i,j}|,$$

$$||A||_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{i,j}|.$$

Q1.6/ On ne fera que $||A||_1$, l'autre norme étant similaire. Pour tout $x \in \mathbb{K}^n$, on a

$$\begin{aligned}
||Ax||_1 &= \sum_{i=1}^n |(Ax)_i| \\
&= \sum_{i=1}^n \left| \sum_{j=1}^n a_{i,j} x_j \right| \\
&\leq \sum_{j=1}^n \left(|x_j| \sum_{i=1}^n |a_{i,j}| \right) \\
&\leq \left(\max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}| \right) \times \sum_{j=1}^n |x_j| \\
&= \left(\max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}| \right) \times ||x||_1
\end{aligned}$$

Ainsi,

$$||A||_1 = \sup_{x \in \mathbb{K}^n, ||x||_1 = 1} ||Ax||_1 \leq \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}|.$$

Pour montrer l'égalité, il faut démontrer que pour toute matrice A, il existe un vecteur $x \in \mathbb{K}^n \backslash \{0\}$ tel que

$$||Ax||_1 = \left(\max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}| \right) \times ||x||_1.$$

Soit j_0 l'indice tel que

$$\max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}| = \sum_{i=1}^n |a_{i,j_0}|.$$

On pose

$$x_j = \begin{cases} \frac{\overline{a_{i,j}}}{|a_{i,j}|} & \text{si } j = j_0, \\ 0 & \text{sinon.} \end{cases}$$

Et dans ce cas, on a

$$\begin{aligned}
||Ax||_1 &= \sum_{i=1}^n \left| \sum_{j=1}^n a_{i,j} x_j \right| \\
&= \sum_{i=1}^n |a_{i,j_0}| \\
&= \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}|,
\end{aligned}$$

d'où le résultat car $||x||_1 = 1$.

Rappels : pour toute matrice $A \in M_n(\mathbb{K})$, on note $A^* =^t \overline{A}$ la matrice *adjointe* de A, i.e. la transposée de la matrice conjuguée de A. A est dite *auto-adjointe* si $A^* = A$. Si A est auto-adjointe, A est dite *positive* si $\langle Ax, x \rangle \in \mathbb{R}_+$ pour tout $x \in \mathbb{K}$. A est positive ssi toutes ses valeurs propres sont dans $\mathbb{R}_+$ (il suffit de diagonaliser avec le théorème spectral).

Rappels : on appelle *rayon spectral* de A le réel positif

$$\rho(A) = \sup_{\lambda \in \mathrm{Sp}A} |\lambda|,$$

où SpA représente l'ensemble des valeurs propres *complexes* de A.

Question 1.7. *Soit $A \in M_n(\mathbb{K})$. Vérifier que A^*A est autoadjointe et positive. En déduire que*

$$||A||_2 = \sqrt{\rho(A^*A)}.$$

Q1.7/ Comme $(AB)^* = B^*A^*$ et $A^{**} = A$, on a

$$(A^*A)^* = A^*A^{**} = A^*A,$$

donc A^*A est auto-adjointe. Comme $\langle Ax, y\rangle = \langle x, A^*y\rangle$ pour tous $x, y \in \mathbb{K}^n$, on a

$$\langle A^*Ax, x\rangle = \langle Ax, Ax\rangle = ||Ax||^2 \geq 0,$$

donc A^*A est positive.

Montrons la dernière propriété. Soit $x \in \mathbb{K}^n$. On a

$$\begin{aligned} ||Ax||_2^2 &= \langle Ax, Ax\rangle \\ &= \langle A^*Ax, x\rangle. \end{aligned}$$

Or, A^*A est positive par ce qui précède. Donc ses valeurs propres sont dans $\mathbb{R}_+$. Si on les compte avec multiplicité, on peut supposer que

$$0 \leq \lambda_1 \leq \cdots \leq \lambda_n.$$

Par le théorème spectral, il existe donc une base orthonormée $(e_1, \ldots, e_n)$ de $\mathbb{K}^n$ de vecteurs propres de A^*A, avec $(A^*A)e_i = \lambda_i e_i$. Ainsi, on peut décomposer x sur la base $(e_1, \ldots, e_n)$:

$$x = \sum_{i=1}^{n} \alpha_i e_i,$$

avec $\alpha_i \in \mathbb{K}$. Ainsi,

$$\begin{aligned} \langle A^*Ax, x\rangle &= \left\langle A^*A \sum_{i=1}^{n} \alpha_i e_i\,,\, \sum_{j=1}^{n} \alpha_j e_j \right\rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \overline{\alpha_j}\, \langle A^*Ae_i\,,\, e_j\rangle && \text{par linéarité} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \alpha_i \overline{\alpha_j} \langle e_i, e_j\rangle \\ &= \sum_{i=1}^{n} \lambda_i\, |\alpha_i|^2 && \text{car } (e_i) \text{ est orthonormée.} \end{aligned}$$

Or,

$$||x||_2^2 = \langle x, x\rangle = \sum_{i=1}^{n} |\alpha_i|^2.$$

Ainsi, pour tout $x \neq 0$, on a

$$\frac{||Ax||_2^2}{||x||_2^2} = \frac{\langle A^*Ax, x\rangle}{\langle x, x\rangle} = \frac{\sum_{i=1}^{n} \lambda_i\, |\alpha_i|^2}{\sum_{i=1}^{n} |\alpha_i|^2} \leq \lambda_n = \rho(A^*A).$$

Ainsi, en passant à la racine, on a bien

$$||A||_2 = \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{||Ax||_2}{||x||_2} \leq \sqrt{\rho(A^*A)}.$$

Par ailleurs, en prenant x tel que $\alpha_n = 1$ et $\alpha_i = 0$ pour $i < n$, on vérifie que

$$\frac{||Ax||_2^2}{||x||_2^2} = \lambda_n = \rho(A^*A),$$

d'où $||A||_2 = \sqrt{\rho(A^*A)}$.

Proposition (♦ admis)**.** *Soit* $||\cdot||$ *une norme matricielle sur* $M_n(\mathbb{K})$*. Alors elle vérifie*

$$\rho(A) \leq ||A||.$$

Réciproquement, pour toute matrice A *et* $\varepsilon > 0$*, il existe une norme* subordonnée $||\cdot||$ *(qui dépend de* A *et* ε*) telle que*

$$||A|| \leq \rho(A) + \varepsilon.$$

Problème 1.8 (Facultatif, pour vous entraîner)**.** Soit $||\cdot||$ une norme *matricielle* quelconque sur $M_n(\mathbb{K})$.

1. Montrer que $\rho(A) \leq ||A^k||^{1/k}$ pour tout $k \in \mathbb{N}^*$.
2. Soit $\varepsilon > 0$ et $A_\varepsilon = \frac{1}{\rho(A)+\varepsilon}A$. Vérifier que $\rho(A_\varepsilon) < 1$. En déduire qu'il existe $k_0 \in \mathbb{N}^*$ tel que pour tout $k \geq k_0$, on a
$$\rho(A) \leq ||A^k||^{1/k} \leq \rho(A) + \varepsilon.$$
3. En déduire la formule de Gelfand : $\rho(A) = \lim_{k\to+\infty} ||A^k||^{1/k}$.

Q1.8/ 1. Il suffit de montrer que $\rho(A)^k \leq ||A^k||$ pour tout $k \in \mathbb{N}^*$. Tout d'abord, on a $\rho(A)^k = \rho(A^k)$. En effet, en trigonalisant sur $\mathbb{C}$, on a

$$A \sim \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \Longrightarrow A^k \sim \begin{pmatrix} \lambda_1^k & & * \\ & \ddots & \\ 0 & & \lambda_n^k \end{pmatrix}$$

avec $\lambda_1, \dots, \lambda_n$ les valeurs propres de A comptées avec multiplicité. D'où $\rho(A)^k = \rho(A^k)$ car $\left(\sup_{1\leq i\leq n} |\lambda_i|\right)^k = \sup_{1\leq i\leq n} |\lambda_i|^k$. Ainsi, d'après la Proposition ♦, on a

$$\rho(A)^k = \rho(A^k) \leq ||A^k||,$$

ce qui prouve le résultat.

Q1.8/ 2. On a évidemment

$$\rho(A_\varepsilon) = \frac{1}{\rho(A)+\varepsilon}\rho(A) < 1.$$

Ainsi, par la Proposition ♦, il existe une norme subordonnée $||\cdot||_s$ telle que $\rho(A_\varepsilon) \leq ||A_\varepsilon||_s < 1$. Par ailleurs, par équivalence des normes, il existe $c > 0$ tel que pour tout $k \in \mathbb{N}$, on a

$$||A_\varepsilon^k|| \leq c||A_\varepsilon^k||_s$$

$$||A_\varepsilon^k||^{1/k} \leq c^{1/k} ||A_\varepsilon^k||_s^{1/k}$$

Or d'une part comme $|| \cdot ||_s$ est aussi une norme matricielle

$$||A_\varepsilon^k||_s^{1/k} \leq ||A_\varepsilon||_s < 1$$

et d'autre part comme $c^{1/k} \to 1$ quand $k \to +\infty$, il existe k_0 tel que pour tout $k \geq k_0$, on a $c^{1/k} \leq \frac{1}{||A_\varepsilon||_s}$. Donc on en déduit que pour $k \geq k_0$

$$||A_\varepsilon^k||^{1/k} \leq 1.$$

Ainsi, on a

$$\rho(A_\varepsilon) \leq ||A_\varepsilon^k||^{1/k} \leq 1,$$

la première inégalité venant de la question 1. En multipliant cette inégalité par $\rho(A) + \varepsilon$, et en utilisant le fait que

$$||A_\varepsilon^k||^{1/k} = \frac{1}{\rho(A) + \varepsilon} ||A^k||^{1/k},$$

on obtient bien

$$\rho(A) \leq ||A^k||^{1/k} \leq 1.$$

Q1.8/ 3. Evident d'après 2 et la définition d'une limite.

2 Suites et séries de matrices

2.1 Résultats de convergence

Rappels : une suite de matrices $(A_i)_{i\geq 1}$ de $M_n(\mathbb{C})$ converge vers A s'il existe une norme $||\cdot||$ telle que $\lim_{i\to+\infty} ||A_i - A|| = 0$. Une suite converge pour une norme ssi elle converge pour toute norme (car on est en dimension finie).

Question 2.1. *Soit $A \in M_n(\mathbb{C})$. Montrer que les quatres conditions suivantes sont équivalentes :*

1. $\lim_{k\to+\infty} A^k = 0$,
2. $\lim_{k\to+\infty}(A^k x) = 0$ *pour tout vecteur* $x \in \mathbb{C}^n$,
3. $\rho(A) < 1$,
4. *il existe au moins une norme subordonnée telle que* $||A|| < 1$.

Q2.1/ On montre que $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$. Commençons par $1 \Rightarrow 2$. Soit $||\cdot||$ une norme subordonnée. Alors

$$||A^k x|| \leq ||A^k||\,||x|| \xrightarrow[k\to 0]{} 0.$$

Montrons maintenant $2 \Rightarrow 3$. Soit $\lambda \in \mathrm{Sp}A$ tel que $\rho(A) = |\lambda|$. Soit $x \in \mathbb{C}$ le vecteur propre associé. Alors par hypothèse, et pour toute norme $||\cdot||$, on a

$$||A^k x|| = |\lambda|^k\,||x|| \xrightarrow[k\to 0]{} 0,$$

ce qui entraine nécessairement que $|\lambda| = \rho(A) < 1$. La propriété $3 \Rightarrow 4$ est une simple conséquence de la Proposition ♦.

Enfin, montrons $4 \Rightarrow 1$. Comme toute norme subordonnée est une norme matricielle, on a

$$||A^k|| \leq ||A||^k \xrightarrow[k\to 0]{} 0$$

car $||A|| < 1$. Donc $\lim_{k\to+\infty} A^k = 0$.

Rappel : $M_n(\mathbb{C})$ est complet (pour toute norme). En particulier, une suite (A_i) converge ssi elle est de Cauchy.

Question 2.2. *Prouver le résultat suivant. Soit $\sum_k a_k z^k$ une série entière de rayon de convergence $R > 0$ (éventuellement infini). Soit $A \in M_n(\mathbb{K})$ telle que $\rho(A) < R$. Alors la série $\sum_k a_k A^k$ converge dans $M_n(\mathbb{C})$.*

Indication : on pourra utiliser la Proposition ♦.

Q2.2/ Comme $\rho(A) < R$ et grâce à la Proposition ♦, il existe une norme subordonnée $||\cdot||$ telle que $||A|| < R$. Or, la série $\sum_k a_k A^k$ converge absolument, puisque

$$\begin{aligned}\sum_{k=0}^{N} \left\|a_k A^k\right\| &\leq \sum_{k=0}^{N} |a_k|\,||A^k|| \\ &\leq \sum_{k=0}^{N} |a_k|\,||A||^k.\end{aligned}$$

Or, cette dernière série converge puisque $\sum_k a_k z^k$ a pour rayon de convergence R et que $||A|| < R$. La convergence absolue entraine la convergence simple car $M_n(\mathbb{C})$ est complet.

Ainsi, si on note $f(z)$ la somme de la série entière ci-dessus, on peut définir

$$\begin{aligned} f\colon \{A \in M_n(\mathbb{C}), \rho(A) < R\} &\to M_n(\mathbb{C}) \\ A &\mapsto \sum_{k=0}^{\infty} a_k A^k. \end{aligned}$$

Exemple. Pour toute matrice $A \in M_n(\mathbb{C})$, on peut définir

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

On notera qu'on n'a pas besoin d'utiliser la Proposition ⧫ pour montrer que cette somme est bien définie, car le rayon de convergence de la série entière associée est $R = +\infty$.

Question 2.3. *Soit $A \in M_n(\mathbb{C})$ telle que $\rho(A) < 1$. Montrer que $I_n - A$ est inversible et que*

$$(I_n - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

Q2.3/ Comme la série $\sum_k z^k$ a pour rayon de convergence $R = 1$, et que $\rho(A) < 1$, on peut définir $\sum_{k=0}^{\infty} A^k$ par la question 2.2. De plus, pour tout $N \in \mathbb{N}$, on a

$$\begin{aligned} (I_n - A)\sum_{k=0}^{N} A^k &= \sum_{k=0}^{N} A^k - \sum_{k=0}^{N} A^{k+1} \\ &= A^0 - A^{N+1} \\ &= I_n - A^{N+1} \end{aligned}$$

et comme $\rho(A) < 1$, on a par la question 2.1 que $A^{N+1} \xrightarrow[N\to+\infty]{} 0$. Donc

$$\lim_{N\to+\infty} (I_n - A)\sum_{k=0}^{N} A^k = (I_n - A)\sum_{k=0}^{+\infty} A^k = I_n.$$

Ainsi, $I_n - A \in GL_n(\mathbb{C})$ avec $I_n - A = \sum_{k=0}^{+\infty} A^k$.

Remarque. On peut définir encore beaucoup d'autres fonctions sur les matrices :

$$\mathrm{ch}A = \sum_{k=0}^{\infty} \frac{A^{2k}}{(2k)!} = \frac{\exp A + \exp(-A)}{2}, \qquad \mathrm{sh}A = \dots$$

$$\cos A = \sum_{k=0}^{\infty} \frac{(-1)^k A^{2k}}{(2k)!} = \frac{\exp(iA) + \exp(-iA)}{2}, \qquad \sin A = \dots$$

$$(I_n + A)^{\alpha} = I_n + \sum_{k=1}^{\infty} \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!} A^k \qquad \text{avec } \alpha \in \mathbb{R}, \quad \rho(A) < 1.$$

2.2 Rappels d'exponentielle matricielle

L'exponentielle matricielle vérifie les propriétés suivantes.

$$AB = BA \Longrightarrow \exp(A+B) = \exp A \exp B, \qquad \text{et donc } (\exp A)^{-1} = \exp(-A),$$

$$\exp \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = \begin{pmatrix} e^{\lambda_1} & & * \\ & \ddots & \\ 0 & & e^{\lambda_n} \end{pmatrix}, \qquad \text{et donc } \det(\exp A) = \exp(\mathrm{tr} A).$$

Problème 2.4 (Facultatif, pour vous entraîner)**.**

1. Montrer que pour tout $A \in M_n(\mathbb{C})$ tel que $\rho(A) < 1$, on peut définir

$$\log(I_n + A) = \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} A^k.$$

On note l'ensemble des matrices nilpotentes et unipotentes de $M_n(\mathbb{C})$:

$$\begin{aligned} \mathrm{Nil} &= \{A \in M_n(\mathbb{C}) \mid A^n = 0\} = \{A \in M_n(\mathbb{C}) \mid \rho(A) = 0\}, \\ \mathrm{Uni} &= \{I_n + A \mid A \in \mathrm{Nil}\} \end{aligned}$$

Le but de ce problème est de montrer que exp réalise un homéomorphisme de Nil sur Uni, dont l'inverse est la fonction log.

2. Vérifier que exp(Nil) $\subset$ Uni, que log(Uni) $\subset$ Nil, et que ces deux applications sont continues.
3. Soit $X \in \mathrm{Nil}$. Montrer que pour tout $t \in \mathbb{R}$

$$\frac{d}{dt}(\log(\exp(tX)) = X.$$

   On pourra poser $Z(t) = \exp(tX) - I_n$ et remarquer que $Z(t) \in \mathrm{Nil}$.
4. En déduire que $\log(\exp(X)) = X$.
5. Montrer que exp est différentiable en la matrice nulle 0_n, de différentielle $D_{0_n} \exp(H) = H$.
6. En déduire qu'il existe un voisinage $V \subset \mathrm{Uni}$ de I_n tels que $\exp(\log Y) = Y$ pour tout $Y \in V$.
7. Soit $Y \in \mathrm{Uni}$. Montrer que $\exp(\log Y) = Y$. On pourra utiliser le polynôme (en t)

$$P(t) = \exp(\log(I_n + tX)) - (I_n + tX),$$

   avec $X \in \mathrm{Nil}$.
8. Conclure.

Q2.4/ 1. La série entière $\sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} z^k$ a pour rayon de convergence 1, donc on peut appliquer la question 2.2 et conclure.

Q2.4/ 2. Soit $N \in \mathrm{Nil}$. Alors $N^k = 0$ pour tout $k \geq n$, donc

$$\begin{aligned} \exp N &= I_n + \left(N + \cdots + \frac{1}{(n-1)!} N^{n-1} \right) \\ &= I_n + M \end{aligned}$$

avec $M \in \text{Nil}$. En effet, $M^n = N^n \left(I_n + \cdots + \frac{1}{(n-1)!}N^{n-2}\right)^n = 0$. Donc $\exp N \in \text{Uni}$. Par ailleurs, cette application est continue car polynômiale en les coefficients de N (vu que la somme ne comporte qu'un nombre fini de termes).

Soit maintenant $U \in \text{Uni}$. Alors comme $N = U - I_n \in \text{Nil}$, on a

$$\begin{aligned}\log U &= \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} N^k \\ &= \sum_{k=1}^{n-1} \frac{(-1)^{k+1}}{k} N^k \\ &= N \left(\sum_{k=0}^{n-2} \frac{(-1)^{k+1}}{k} N^k \right),\end{aligned}$$

et on montre de même que $(\log U)^n = 0$. Par ailleurs, puisque $N = U - I_n$, $\log U$ est également polynômial en les coefficients de U, donc l'application log est continue.

Q2.4/ 3. On remarque que

$$\begin{aligned}\log(\exp(tX)) &= \log\left(I_n + tX + \cdots + \frac{t^{n-1}}{(n-1)!}X^{n-1}\right) \\ &= \log\left(I_n + Z(t)\right) \\ &= \sum_{k=1}^{n-1} \frac{(-1)^{k+1}}{k} Z(t)^k\end{aligned}$$

Ainsi,

$$\begin{aligned}\frac{d}{dt}(\log(\exp(tX)) &= Z'(t) \sum_{k=1}^{n-1} (-1)^{k+1} Z(t)^{k-1} \\ &= Z'(t) \sum_{k=0}^{n-2} (-1)^k Z(t)^k \\ &= Z'(t)\left[(I_n + Z(t))^{-1} - (-1)^{n-1} Z(t)^{n-1}\right]\end{aligned}$$

en utilisant la question 2.3 car $\rho(Z(t)) = 0 < 1$ et $Z(t)^n = 0$. Or,

$$\begin{aligned}Z'(t) &= X + \frac{t}{2}X^2 \cdots + \frac{t^{n-2}}{(n-2)!}X^{n-1} \\ &= XA(t)\end{aligned}$$

avec $A(t) \in M_n(\mathbb{C})$, tandis que $Z(t)^{n-1} = X^{n-1}B(t)$ avec $B \in M_n(\mathbb{C})$. Ainsi, $Z'(t)Z(t)^{n-1} = 0$. On en déduit que

$$\begin{aligned}\frac{d}{dt}(\log(\exp(tX)) &= Z'(t)\left(I_n + Z(t)\right)^{-1} \\ &= X\exp(tX)\left(\exp(tX)\right)^{-1} \\ &= X.\end{aligned}$$

Q2.4/ 4. Notons $f(t) = \log(\exp(tX))$. Par la question 3, on a $f'(t) = X$. De plus,

$$f(0) = \log(\exp(0)) = \log(I_n) = 0.$$

Donc pour tout $t \in \mathbb{R}$

$$f(t) = f(0) + \int_0^t f'(s)ds = 0 + tX.$$

En particulier, $f(1) = \log(\exp(X)) = X$.

Q2.5/ 5. On part de la définition d'une différentielle en 0. Pour tout $H \in M_n(\mathbb{C})$,

$$\begin{aligned}\exp(0 + H) - \exp(0) &= \sum_{k=0}^{\infty} \frac{H^k}{k!} - I_n \\ &= \sum_{k=1}^{\infty} \frac{H^k}{k!} \\ &= H + H^2 \sum_{k=2}^{\infty} \frac{H^{k-2}}{k!}\end{aligned}$$

Or d'une part l'application $H \mapsto H$ est linéaire et trivialement continue, d'autre part, le deuxième terme est un $O(H^2)$ pour toute norme matricielle. Donc exp est différentiable en 0 et $D_{0_n} \exp(H) = H$.

Q2.5/ 6. La différentielle $D_{0_n} \exp$ est clairement bijective. De plus $\exp(0_n) = I_n$, donc par le théorème d'inversion locale, il existe un voisinage $W \subset \text{Nil}$ de 0_n et un voisinage $V \subset \text{Uni}$ de I_n, tels que exp est un difféomorphisme de W vers V. En particulier, si on note $\varphi \colon V \to W$ l'application réciproque, on a $\varphi \circ \exp|_W = \text{Id}_W$ et $\exp|_W \circ \varphi = \text{Id}_V$.

Or, par la question 4, on a déjà $\log|_V \circ \exp|_W = \text{Id}_W$, donc par unicité de l'application réciproque, on en déduit que $\varphi = \log|_V$, donc que $\exp \circ \log|_V = \text{Id}_V$.

Q2.5/ 7. Soit $Y \in \text{Uni}$ et $X = Y - I_n \in \text{Nil}$. On pose pour tout $t \in \mathbb{R}$

$$P(t) = \exp(\log(I_n + tX)) - (I_n + tX).$$

Par la question 6, on a $P(t) = 0$ si $I_n + tX \in V$. Or, V étant un voisinage de I_n, il existe $r > 0$ tel que $B(I_n, r) \subset V$, où $B(I_n, r)$ est la boule centrée en I_n, de rayon r, selon la norme $|| \cdot ||_2$ (par exemple). Ainsi, Y et X étant fixés, il existe t_0 tel que pour tout $|t| \leq t_0$ on a $||tX||_2 \leq t_0||X||_2 < r$. Donc $P(t) = 0$ pour tout $t \in [-t_0, t_0]$.

Or, $P(t)$ est un polynôme en t, comme on l'a vu à la question 2. Comme P a une infinité de racines, on en déduit que $P = 0$. Donc en particulier pour $t = 1$, on a

$$P(1) = 0 = \exp(\log(Y)) - Y.$$

Comme Y a été choisi quelconque, on a donc $(\exp \circ \log)(Y) = Y$ pour tout $Y \in \text{Uni}$.

Q2.5/ 8. Les questions précédentes montrent que exp et log sont des homéomorphismes réciproques l'un de l'autre, qui envoient respectivement Nil sur Uni et Uni sur Nil.

3 Analyse numérique matricielle

On s'intéresse à la résolution d'un système linéaire $Ax = b$ avec $A \in GL_n(\mathbb{K})$ et $b \in \mathbb{K}^n$.

3.1 Conditionnement

Il arrive que des petites perturbations de A et b conduisent à une solution x radicalement différente, comme le montre l'exemple suivant :

$$\begin{pmatrix} 8 & 6 & 4 & 1 \\ 1 & 4 & 5 & 1 \\ 8 & 4 & 1 & 1 \\ 1 & 4 & 3 & 6 \end{pmatrix} x = b = \begin{pmatrix} 19 \\ 11 \\ 14 \\ 14 \end{pmatrix} \Longrightarrow x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

$$\begin{pmatrix} 8 & 6 & 4 & 1 \\ 1 & 4 & 5 & 1 \\ 8 & 4 & 1 & 1 \\ 1 & 4 & 3 & 6 \end{pmatrix} x_\varepsilon = b_\varepsilon = \begin{pmatrix} 19,01 \\ 11,05 \\ 14,07 \\ 14,05 \end{pmatrix} \Longrightarrow x_\varepsilon = \begin{pmatrix} -2,34 \\ 9,745 \\ -4,85 \\ -1,34 \end{pmatrix}.$$

On remarque que l'erreur relative sur les données b et b_ε est grandement amplifiée quand on compare les solutions :

$$\frac{||x_\varepsilon - x||_\infty}{||x||_\infty} \simeq 2373 \frac{||b - b_\varepsilon||_\infty}{||b||_\infty} \quad !!$$

Pour étudier cette situation, pour $\varepsilon > 0$ fixé, on pose $A_\varepsilon = A + \varepsilon A'$ et $b_\varepsilon = b + \varepsilon b'$, de sorte à obtenir un système perturbé $A_\varepsilon x_\varepsilon = b_\varepsilon$, de solution x_ε. On souhaite avoir une majoration de l'erreur relative $\frac{||x_\varepsilon - x||}{||x||}$.

Question 3.1. *Montrer que $GL_n(\mathbb{K})$ est un ouvert de $M_n(\mathbb{K})$ (on pourra éventuellement utiliser la question 2.3). En déduire que $A_\varepsilon x_\varepsilon = b_\varepsilon$ admet une unique solution pour $\varepsilon > 0$ assez petit.*

Q3.1/ La méthode courte est d'utiliser le fait que l'application $A \mapsto \det A$ est continue (comme expression polynômiale en les coefficents de la matrice). Or, $GL_n(\mathbb{K}) = \det^{-1}(\mathbb{K}\backslash\{0\})$. Comme $\mathbb{K}\backslash\{0\}$ est ouvert, on en déduit que $GL_n(\mathbb{K})$ également (selon toute norme car elles sont équivalentes).

Ainsi, pour toute matrice A, il existe un voisinage V_A de A dans $GL_n(\mathbb{K})$, et donc il existe $\varepsilon_0 > 0$ tel que $A_\varepsilon = A + \varepsilon A' \in V_A$ pour tout $\varepsilon \leq \varepsilon_0$. Donc le système $A_\varepsilon x_\varepsilon = b_\varepsilon$ a pour unique solution $x_\varepsilon = A_\varepsilon^{-1} b_\varepsilon$.

Question 3.2. *On munit $M_n(\mathbb{C})$ d'une norme subordonnée $||\cdot||$. Montrer que quand $\varepsilon \to 0$,*

$$\left(I_n + \varepsilon A^{-1}A'\right)^{-1} = I_n - \varepsilon A^{-1}A' + O(\varepsilon^2),$$

$$x_\varepsilon = x + \varepsilon A^{-1}(b' - A'x) + O(\varepsilon^2).$$

En déduire que

$$\frac{||x_\varepsilon - x||}{||x||} \leq \varepsilon ||A^{-1}|| \, ||A|| \left(\frac{||b_\varepsilon - b||}{||b||} + \frac{||A_\varepsilon - A||}{||A||} \right) + O(\varepsilon^2). \tag{2}$$

Q3.2/ Pour $\varepsilon > 0$ assez petit, on a $\rho(\varepsilon A^{-1}A') < 1$. Ainsi, par la question 2.3, on a

$$\begin{aligned} \left(I_n + \varepsilon A^{-1}A'\right)^{-1} &= \sum_{k=0}^{\infty} \left(-\varepsilon A^{-1}A'\right)^k \\ &= I_n - \varepsilon A^{-1}A' + \varepsilon^2 \sum_{k=2}^{\infty} \varepsilon^{k-2} \left(-A^{-1}A'\right)^k \\ &= I_n - \varepsilon A^{-1}A' + O(\varepsilon^2). \end{aligned}$$

Ainsi, quand $\varepsilon \to 0$, on a bien

$$\begin{aligned}
x_\varepsilon &= A_\varepsilon^{-1} b_\varepsilon \\
&= (A + \varepsilon A')^{-1}(b + \varepsilon b') \\
&= \left(I_n + \varepsilon A^{-1}A'\right) A^{-1}(b + \varepsilon b') \\
&= \left(I_n - \varepsilon A^{-1}A'\right)(x + \varepsilon A^{-1}b') + O(\varepsilon^2) \\
&= x + \varepsilon(-A^{-1}A'x + A^{-1}b') + O(\varepsilon^2) \\
&= x + \varepsilon A^{-1}(-A'x + b') + O(\varepsilon^2).
\end{aligned}$$

Avec $||\cdot||$ une norme subordonnée, on en déduit que

$$\begin{aligned}
||x_\varepsilon - x|| &\leq \varepsilon ||A^{-1}||\,||A'||\,||x|| + \varepsilon ||A^{-1}||\,||b'|| + O(\varepsilon^2) \\
&\leq \varepsilon ||A^{-1}||\,||A|| \left(||x|| \frac{||A'||}{||A||} + \frac{||b'||}{||A||} \right) + O(\varepsilon^2)
\end{aligned}$$

Or, comme $b = Ax$, on a $||b|| \leq ||A||\,||x||$, ou encore,

$$\frac{1}{||A||} \leq \frac{||x||}{||b||}$$

ce qui permet de conclure que

$$\frac{||x_\varepsilon - x||}{||x||} \leq \varepsilon ||A^{-1}||\,||A||| \left(\frac{||A'||}{||A||} + \frac{||b'||}{||b||} \right) + O(\varepsilon^2),$$

ce qui est le résultat recherché.

On a majoré l'erreur relative entre x et x_ε en fonction de celle faite sur les données A et b. On dispose donc d'un « indicateur » de l'amplification de l'erreur : le *conditionnement* de A, noté

$$\text{cond}(A) = ||A||\,||A^{-1}|| \geq ||AA^{-1}|| = 1.$$

La situation la plus favorable est donc d'avoir un conditionnement très proche de 1. Par ailleurs, la valeur du conditionnement dépend de la norme subordonnée choisie. Ainsi, pour la matrice de l'exemple ci-dessus, on a

$$\text{cond}_\infty(A) = ||A||_\infty ||A^{-1}||_\infty \simeq 5367.$$

On peut montrer que la majoration de (2) est optimale (dans le sens où il existe b' et A' pour lesquels c'est une égalité). Cela étant, cette valeur reste pessimiste : dans la grande majorité des cas, on est très loin d'avoir une égalité dans (2).

Question 3.3. *Soit $A \in GL_n(\mathbb{K})$. Montrer que $Sp(A^{-1}) = \left\{\lambda^{-1} \mid \lambda \in SpA\right\}$ et que $SpA^*A \subset \mathbb{R}_+^*$. En déduire que*

$$cond_2(A) = \frac{\sigma_n(A)}{\sigma_1(A)},$$

où $\sigma_1(A)$ et $\sigma_n(A)$ sont respectivement la plus petite et la plus grande valeur propre de A^*A.

Q3.3/ Soit $\lambda \in \text{Sp}A$ et x un vecteur propre associé. Alors $\lambda \neq 0$ on aurait $A \notin GL_n(\mathbb{K})$. Alors

$$x = A^{-1}Ax = A^{-1}(\lambda x) = \lambda A^{-1}x,$$

donc $A^{-1}x = \lambda^{-1}x$, i.e. $\lambda^{-1} \in \text{Sp}(A^{-1})$. L'inclusion réciproque se montre de la même manière en utilisant que $(A^{-1})^{-1} = A$.

On a vu à la question 1.7 que A^*A est auto-adjointe et positive, donc $\mathrm{Sp}A^*A \subset \mathbb{R}_+$. Comme $A \in GL_n(\mathbb{K})$, il en va de même pour A^*, si bien que $A^*A \in GL_n(\mathbb{K})$. Donc $0 \notin \mathrm{Sp}(A^*A)$. On peut de la même façon montrer que AA^* vérifie les mêmes propriétés.

Par définition, il est clair que $||A||_2 = \rho(A^*A) = \sigma_n(A)$. Il suffit donc de montrer que $||A^{-1}||_2 = \frac{1}{\sigma_1(A)}$. Comme $(A^{-1})^* = (A^*)^{-1}$, on a

$$\begin{aligned}
||A^{-1}||_2 &= \rho\left[\left(A^{-1}\right)^* A^{-1}\right] \\
&= \rho\left[\left(A^*\right)^{-1} A^{-1}\right] \\
&= \rho\left[(AA^*)^{-1}\right] \\
&= \sup_{\lambda \in \mathrm{Sp}[(AA^*)^{-1}]} |\lambda| \\
&= \sup_{\lambda \in \mathrm{Sp}(AA^*)} \lambda^{-1} \qquad\qquad \text{car } \mathrm{Sp}(AA^*) \subset \mathbb{R}_+^* \\
&= \left(\inf_{\lambda \in \mathrm{Sp}(AA^*)} \lambda\right)^{-1} \qquad\qquad \text{idem}
\end{aligned}$$

par ce qui précède. Or, $\mathrm{Sp}(AA^*) = \mathrm{Sp}(A^*A)$ [1], ce qui permet d'affirmer que $||A^{-1}||_2 = \frac{1}{\sigma_1(A)}$.

Si A est une matrice normale (i.e. $AA^* = A^*A$), alors on peut montrer que $\mathrm{cond}_2(A) = \frac{\max|S|}{\min|S|}$, où $|S| = \{|\lambda| \mid \lambda \in \mathrm{Sp}A\}$. Toutefois, dans le cas général, il n'y a aucun lien entre les valeurs propres de A et A^*A. Par ailleurs, les relations d'équivalence entre les normes $||\cdot||_{1,2,\infty}$ permettent de déduire des relations d'équivalences entre $\mathrm{cond}_{1,2,\infty}$. Par exemple :

$$n^{-1}\mathrm{cond}_2(A) \leq \mathrm{cond}_1(A) \leq n\mathrm{cond}_2(A).$$

Il est donc toujours avantageux de minimiser $\mathrm{cond}_2(A)$ quelle que soit la norme considérée pour le conditionnement.

3.2 Méthodes itératives, généralités

Pour résoudre un système linéaire $Ax = b$, on a plusieurs méthodes possibles. Les méthodes directes cherchent à trouver la solution exacte, par exemple avec le pivot de Gauss. Bien que la solution obtenue soit exacte (modulo les erreurs / perturbations), l'inconvénient majeur est que pour une matrice carrée $A \in GL_n(\mathbb{K})$, la complexité est en $O(n^3)$. Ainsi, ces méthodes sont inapplicables pour les systèmes de très grande taille (par exemple $n \sim 10^6$). Les méthodes itératives se présentent alors comme une alternative plus adaptée.

Nous considérerons ici une approche très générale sur les méthodes de splitting, ou de décomposition.

Définition. Soit $A \in GL_n(\mathbb{K})$. Une paire de matrices (M, N) est une décomposition régulière de A si $M \in GL_n(\mathbb{K})$ et $A = M - N$. On définit alors une méthode itérative associée à (M, N) par

$$\begin{cases} x_0 \in \mathbb{K}^n \text{ donné}, \\ Mx_{k+1} = Nx_k + b. \end{cases}$$

Une méthode itérative est dite *convergente* si pour tout vecteur initial x_0, l'erreur entre x_k et la solution exacte x vérifie

$$e_k = x_k - x \xrightarrow[k \to +\infty]{} 0.$$

1. Pour toutes matrices $A, B \in M_n(\mathbb{K})$, toute valeur propre $\lambda \neq 0$ de AB est aussi valeur propre de BA. En effet si u est un vecteur propre de AB associé à λ, on a $BA(Bu) = B(ABu) = \lambda(Bu)$ avec $Bu \neq 0$ car $ABu = \lambda u \neq 0$.

On notera que $x_{k+1} = x_k$ ssi x_k est solution de $Ax = b$.

A chaque pas, on calcule donc $x_{k+1} = M^{-1}Nx_k + M^{-1}b$. En pratique, on prendra pour M une matrice facile à inverser, par exemple une matrice unitaire car alors $M^{-1} = \overline{M}^t$ (voir section suivante pour plus d'exemples). Ceci permet en général d'avoir une complexité en $O(n^2)$ par pas, ce qui représente un gain appréciable comparé aux méthodes directes (à condition de garder un nombre de pas petit devant n).

Question 3.4. *Trouver une relation entre e_k et e_{k-1}. En déduire que la méthode itérative associée à (M, N) est convergente ssi $\rho(M^{-1}N) < 1$.*

Q3.4/ Comme x est solution exacte de la méthode itérative, on a $x = M^{-1}Nx + M^{-1}b$. Ainsi, pour tout $k \geq 1$

$$\begin{aligned} e_k &= x_k - x \\ &= M^{-1}Nx_{k-1} + M^{-1}b - x \\ &= M^{-1}Ne_{k-1} + M^{-1}b + M^{-1}Nx - x \\ &= M^{-1}Ne_{k-1}. \end{aligned}$$

Ainsi, par récurrence immédiate $e_k = (M^{-1}N)^k e_0$. Par la question 2.1, on en déduit que $e_k \xrightarrow[k\to\infty]{} 0$ pour tout vecteur e_0 ssi $\rho(M^{-1}N) < 1$.

La méthode converge donc d'autant plus vite que $\rho(M^{-1}N)$ est petit devant 1. Du moins en théorie ! En effet, à chaque pas on résout un système linéaire $Mx_{k+1} = Nx_k + b$. Ainsi, d'après la question 3.2, toute erreur sur x_k est potentiellement amplifiée quand on calcule x_{k+1}. Que cette erreur vienne des données A et b, ou bien de la précision machine, il faut s'assurer qu'elle reste « petite » même après un grand nombre de pas.

Question 3.5. *Soit (M, N) une décomposition régulière de $A \in GL_n(\mathbb{K})$. On définit une méthode itérative « perturbée » par*

$$\begin{cases} x_0 \in \mathbb{K}^n \text{ donné}, \\ x_{k+1} = M^{-1}Nx_k + M^{-1}b + \varepsilon_k, \quad \forall k \geq 0. \end{cases}$$

On suppose que $||M^{-1}N|| < 1$, avec $||\cdot||$ une norme subordonnée, et qu'il existe $\varepsilon > 0$ tel que $||\varepsilon_k|| \leq \varepsilon$ pour tout k. Montrer qu'il existe une constante $K > 0$ qui dépend de $M^{-1}N$ mais pas de ε, telle que

$$\limsup_{k\to+\infty} ||x_k - x|| \leq K\varepsilon.$$

Q3.5/ On pose $e_k = x_k - x$. Similairement à la question 3.4, on a

$$e_{k+1} = M^{-1}Ne_k + \varepsilon_k.$$

Par récurrence, on a donc

$$e_{k+1} = (M^{-1}N)^{k+1}e_0 + \sum_{m=0}^{k} (M^{-1}N)^{k-m}\varepsilon_m.$$

On note $C = ||M^{-1}N|| < 1$. Alors

$$\begin{aligned} ||e_{k+1}|| &\leq ||M^{-1}N||^{k+1}||e_0|| + \sum_{m=0}^{k} ||M^{-1}N||^{k-m}||\varepsilon_m|| \\ &\leq C^{k+1}||e_0|| + \varepsilon \sum_{m=0}^{k} C^{k-m} \\ &\leq C^{k+1}||e_0|| + \varepsilon \sum_{m=0}^{k} C^{m} \\ &\leq C^{k+1}||e_0|| + \frac{\varepsilon}{1-C}, \end{aligned}$$

donc comme $C \in]0,1[$, on a

$$\limsup_{k\to+\infty} ||e_k|| \leq \frac{\varepsilon}{1-C},$$

d'où le résultat avec $K = (1-C)^{-1}$.

3.3 Méthodes de Jacobi et Gauss-Seidel

Exemple. Soit $A \in GL_n(\mathbb{K})$. On note $D = \mathrm{diag}(a_{11}, \ldots, a_{nn})$ la matrice diagonale formée à partir de la diagonale de A. La méthode de Jacobi consiste à choisir comme décomposition régulière

$$M = D, \qquad N = D - A.$$

On notera que M est très facile à inverser.

Exemple. On reprend les hypothèses et notations de l'exemple précédent. On note $-E$ et $-F$ les matrices triangulaires strictes qui correspondent respectivement à la partie triangulaire inférieure et supérieure de A, de sorte que

$$\begin{pmatrix} \ddots & & \text{-F} \\ & D & \\ \text{-E} & & \ddots \end{pmatrix}.$$

La méthode de Gauss-Seidel consiste à choisir comme décomposition régulière

$$M = D - E, \qquad N = F.$$

Comme M est triangulaire, le pivot de Gauss pour résoudre le système ne demande que $O(n^2)$ opérations.

Problème 3.6 (Facultatif, pour vous entrainer)**.** Soit $A \in M_n(\mathbb{K})$ une matrice auto-adjointe définie positive et (M, N) une décomposition régulière de A.

1. Montrer que $M^* + N$ est auto-adjointe.

On suppose maintenant que $M^* + N$ est définie positive.

2. Vérifier que $||x||_A = \sqrt{\langle Ax, x\rangle}$ est une norme sur $\mathbb{K}^n$. On note $||\cdot||$ la norme matricielle subordonnée à $||\cdot||_A$.
3. Montrer qu'il existe un vecteur $v \in \mathbb{K}^n$ tel que $||v||_A = 1$ et $||M^{-1}N|| = ||M^{-1}Nv||_A$.
4. En déduire que
$$||M^{-1}Nv||_A^2 = 1 - \langle (M^* + N)w, w\rangle,$$
où $w = M^{-1}Av$.
5. En déduire que $\rho(M^{-1}N) < 1$. On pourra utiliser la Proposition ⧫.

On considère le système linéaire $Ax = b$ avec $b \in \mathbb{K}^n$ (et A vérifie toujours les hypothèses ci-dessus).

6. Montrer que si $2D - A$ est définie positive, la méthode de Jacobi est convergente.
7. Montrer que la méthode de Gauss-Seidel est convergente.

Q3.6/ 1. On a

$$M - N = A = A^* = M^* - N^*$$

d'où $M^* + N = M + N^* = (M^* + N)^*$.

Q3.6/ 2. Il suffit de montrer que $(x, y) \mapsto \langle Ax, y\rangle$ définit un produit scalaire (hermitien si $\mathbb{K} = \mathbb{C}$). La sesquilinéarité est évidente, de même que la symétrie (hermitienne si $\mathbb{K} = \mathbb{C}$). Enfin,

$$\begin{aligned}(x, x) = 0 &\Leftrightarrow \langle Ax, x\rangle = 0\\ &\Leftrightarrow x = 0\end{aligned}$$

car A est définie positive.

Q3.6/ 3. Par définition, on a pour toute matrice P

$$||P|| = \sup_{x\in\mathbb{K}^n, ||x||_A=1} ||Px||_A.$$

Prouver l'existence de v revient à dire que le sup de cette définition est en fait un max. On note B la sphère unité de $\mathbb{K}^n$ pour la norme $||\cdot||_A$. B est compacte, car fermée et bornée dans $\mathbb{K}^n$ qui est de dimension finie. Or, l'application

$$\begin{aligned}\mathbb{K}^n &\to \mathbb{R}_+\\ x &\mapsto ||Px||_A\end{aligned}$$

est continue, donc atteint son max sur B. Donc il existe $v \in B$ tel que $||P|| = ||Pv||_A$. Il suffit alors d'appliquer le résultat à $P = M^{-1}N$.

Q3.6/ 4. Comme $N = M - A$, on a

$$\begin{aligned}||M^{-1}Nv||_A^2 &= \langle AM^{-1}Nv, M^{-1}Nv\rangle\\ &= \langle A(I_n - M^{-1}A)v, (I_n - M^{-1}A)v\rangle\\ &= \langle Av - Aw, v - w\rangle\\ &= \langle Av, v\rangle - \langle Av, w\rangle - \langle Aw, v\rangle + \langle Aw, w\rangle\\ &= 1 - \langle Av, w\rangle - \langle w, Av\rangle + \langle Aw, w\rangle \qquad \text{car } A^* = A\\ &= 1 - \langle Mw, w\rangle - \langle w, Mw\rangle + \langle Aw, w\rangle\\ &= 1 - \langle (M + M^* - A)w, w\rangle\\ &= 1 - \langle (M^* + N)w, w\rangle.\end{aligned}$$

Q3.6/ 5. Par la Proposition ♦, on a

$$\begin{aligned}\rho(M^{-1}N) &\leq ||M^{-1}N||\\ &= ||M^{-1}Nv||_A^2\\ &\leq 1 - \langle (M^* + N)w, w\rangle.\end{aligned}$$

Or, $M^* + N$ est supposée définie positive, et comme $v \neq 0$ (puisque $||v||_A = 1$), on a $w = M^{-1}Av \neq 0$ également (car $M^{-1}A \in GL_n(\mathbb{K})$). Donc $\langle (M^* + N)w, w\rangle > 0$, et finalement $\rho(M^{-1}N) < 1$.

Q3.6/ 6. Par hypothèse, on a

$$A = M - N = D - (D - A).$$

Par les questions précédentes, il suffit que la matrice suivante soit définie positive :

$$M^* + N = D^* + D - A$$

Il suffit donc de montrer que $D^* = D$ pour conclure. Si $\mathbb{K} = \mathbb{R}$, c'est évident. Si $\mathbb{K} = \mathbb{C}$, montrons que tous les coefficients de D sont réels. En effet, si on note $e_i \in \mathbb{C}^n$ le i-ème vecteur de la base canonique, alors comme A est supposée définie positive, on a

$$0 < \langle Ae_i, e_i \rangle = a_{i,i} = d_{i,i}$$

donc on a le résultat.

Q3.6/ 7. Par hypothèse, on a

$$A = M - N = (D - E) - F.$$

Par les questions précédentes, il suffit que la matrice suivante soit définie positive :

$$M^* + N = D^* - E^* + F = D^*.$$

En effet, $E^* = F$ car A est auto-adjointe. De plus, on a vu à la question précédent que $D^* = D$ est une matrice définie positive, d'où le résultat.

3.4 Gradient à pas constant

On suppose dans cette partie que $A \in S_n(\mathbb{R})$, i.e. A est une matrice symétrique réelle. On note $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ les valeurs propres de A, comptées avec multiplicité.

Définition. La méthode de gradient à pas constant est définie par la décomposition régulière

$$A = M - N = \frac{1}{\alpha} I_n - \left(\frac{1}{\alpha} I_n - A \right),$$

avec $\alpha \in \mathbb{R}^*$. En d'autres termes, elle se définit par

$$\begin{cases} x_0 \in \mathbb{K}^n \text{ donné}, \\ x_{k+1} = x_k + \alpha(b - Ax_k), \quad \forall k \geq 0. \end{cases}$$

Question 3.7. *Montrer les assertions suivantes :*

1. *Si $\lambda_1 \leq 0 \leq \lambda_n$, la méthode de gradient ne converge pas quelle que soit la valeur de α.*
2. *Si $0 < \lambda_1 \leq \cdots \leq \lambda_n$, la méthode de gradient converge ssi $0 < \alpha < 2/\lambda_n$.*

Q3.7/ En notant $e_k = x_k - x$, où x est la solution exacte, on a

$$\begin{aligned} e_{k+1} &= e_k + \alpha(b - Ax_k) \\ &= e_k + \alpha(b - Ae_k - Ax) \\ &= (I_n - \alpha A)e_k. \end{aligned}$$

Par un raisonnement similaire à celui de la question 3.4, la méthode converge ssi $\rho(I_n - \alpha A) < 1$. Or, comme I_n et A commutent, elles sont cotrigonalisables dans $\mathbb{C}$, avec leurs valeurs propres sur la diagonales. On en déduit que

$$\mathrm{Sp}(I_n - \alpha A) = \{1 - \alpha\lambda \mid \lambda \in \mathrm{Sp} A\}.$$

En particulier,

$$\min \mathrm{Sp}(I_n - \alpha A) = 1 - \alpha\lambda_n, \qquad \max \mathrm{Sp}(I_n - \alpha A) = 1 - \alpha\lambda_1.$$

La méthode converge donc ssi $|1 - \alpha\lambda_n| < 1$ et $|1 - \alpha\lambda_1| < 1$.

1. Supposons que $\lambda_1 \leq 0 \leq \lambda_n$. Alors si $\alpha \leq 0$, on a $1 - \alpha\lambda_n \geq 1$, et si $\alpha \geq 0$, on a $1 - \alpha\lambda_1 \geq 1$. Dans les deux cas, la méthode ne converge pas.

2. Supposons que $0 < \lambda_1 \leq \cdots \leq \lambda_n$. Alors

— Si $\alpha \leq 0$, on a $1 - \alpha\lambda_1 \geq 1$, donc la méthode ne converge pas.

— Si $\alpha \geq 2/\lambda_n$, on a $1 - \alpha\lambda_n \leq -1$, donc la méthode ne converge pas.

— Si $0 < \alpha < 2/\lambda_n$, on a $-1 < 1 - \alpha\lambda_n \leq 1 - \alpha\lambda_1 < 1$, donc la méthode converge.