Développement : Critère de convergence itérative avec le rayon spectral

Analyse & Probabilités
Algèbre & Géométrie

Référence : [AK] Allaire G., Kaber S., *Algèbre linéaire numérique*, ellipses, 2002, p156

Pour les leçons :
- 153 : Valeurs propres, vecteurs propres. Calculs exacts ou approchés d'éléments propres. Applications.
- 156 : Endomorphismes trigonalisables. Endomorphismes nilpotents.
- 162 : Systèmes d'équations linéaires : opérations élémentaires, aspects algorithmiques et conséquences théoriques.
- 226 : Suites vectorielles et réelles définies par une relation de récurrence $u_{n+1} = f(u_n)$. Exemples. Applications à la résolution approchée d'équations.

Soit $n \in \mathbb{N}^*$. Si $A \in \mathcal{M}_n(\mathbb{C})$, on note $\rho(A) := \max_{\lambda \in \mathrm{Sp}(A)} |\lambda|$ *le rayon spectral de* A, et on note $T_n^+(\mathbb{C})$ l'ensemble des matrices carrées de taille n triangulaires supérieures à coefficients dans $\mathbb{C}$.

Lemme 1.

Soient $A \in \mathcal{M}_n(\mathbb{C})$ et $\varepsilon > 0$. Alors, il existe une norme subordonnée $|||\cdot|||$ matricielle telle que :

$$|||A||| < \rho(A) + \varepsilon$$

Preuve : Comme $A \in \mathcal{M}_n(\mathbb{C})$, A est trigonalisable dans $\mathcal{M}_n(\mathbb{C})$. Soient donc $T \in T_n^+(\mathbb{C})$ et $P \in \mathrm{GL}_n(\mathbb{C})$ telles que $T = P^{-1}AP$.
On considère la base canonique $(e_1, \dots, e_n)$ de $\mathbb{C}^n$. Soit $\delta > 0$. On pose, pour $i \in [\![1;n]\!]$, $e_i' = \delta^{i-1} e_i$. Pour $j \in [\![1;n]\!]$, on a :

$$\begin{aligned} Te_j' &= \delta^{j-1} T e_j \\ &= \delta^{j-1} \sum_{i=1}^{j} T_{i,j} e_i \\ &= \delta^{j-1} \sum_{i=1}^{j} T_{i,j} \frac{e_i'}{\delta^{i-1}} \\ &= \sum_{i=1}^{j} T_{i,j} \delta^{j-i} e_i'. \end{aligned}$$

Maintenant, soit $D = \mathrm{diag}(1, \delta, \dots, \delta^{n-1})$. On pose $T_\delta = D^{-1}TD = \begin{pmatrix} T_{1,1} & \delta T_{1,2} & \cdots & \delta^{n-1} T_{1,n} \\ & T_{2,2} & \ddots & \vdots \\ & & \ddots & \delta T_{n-1,n} \\ (0) & & & T_{n,n} \end{pmatrix}$.
Soit $\|\cdot\| : x \longmapsto \|(PD)^{-1}x\|_\infty$. On peut aisément vérifier que $\|\cdot\|$ est une norme sur $\mathbb{C}^n$. On lui associe sa norme subordonnée $|||\cdot|||$.
Pour $B \in \mathcal{M}_n(\mathbb{C})$, on a :

$$\begin{aligned} |||B||| &= \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Bx\|}{\|x\|} \\ &= \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|(PD^{-1})Bx\|_\infty}{\|(PD)^{-1}x\|_\infty}. \end{aligned}$$

En posant $y = (PD)^{-1}x$, $x = PDy$, on obtient :

$$|||B||| = \left|\left|\left|PD^{-1}BPD\right|\right|\right|_\infty.$$

Donc :

$$\begin{aligned} |||A||| &= \left|\left|\left|PD)^{-1}APD\right|\right|\right|_\infty \\ &= \left|\left|\left|D^{-1}TD\right|\right|\right|_\infty \\ &= |||T_\delta|||_\infty \\ &= \max_{i \in [\![1;n]\!]} \left(|T_{i,i}| + \sum_{j=i+1}^{n} \delta^{j-i} |T_{i,j}| \right). \end{aligned}$$

Remarque 2.

En effet, si $A = (A_{i,j})_{i,j\in[\![1;n]\!]}$, alors $|||A|||_\infty = \max_{i\in[\![1;n]\!]} \sum_{j=1}^n |A_{i,j}|$ (norme subordonnée à $\|\cdot\|_\infty$) : en fixant une ligne i, on fait la somme de tous les coefficients (en module) de la ligne i, et $|||A|||_\infty$ est le plus grand nombre d'entre eux.

Montrons-le. Cela revient à montrer que $\sup_{x\in\mathbb{C}^n, \|x\|_\infty=1} \|Ax\|_\infty = \max_{i\in[\![1;n]\!]} \sum_{j=1}^n |A_{i,j}|$.

$\star$ $\underline{\leqslant}$: Pour $i \in [\![1;n]\!]$ et $\|x\|_\infty = 1$, $|(Ax)_i| = \sum_{k=1}^n |A_{i,k}x_k| \leqslant \sum_{k=1}^n |A_{i,k}|$, puisque pour tout $k \in [\![1;n]\!]$, $|x_k| \leqslant \|x\|_\infty = 1$.
Par passage au sup sur i, on a l'inégalité $\leqslant$.
$\star$ $\underline{\geqslant}$: Il suffit de trouver un vecteur x réalisant l'égalité.

Soit $i_0 \in [\![1;n]\!]$ vérifiant $\sum_{j=1}^n |a_{i_0,j}| = \max_{i\in[\![1;n]\!]} \sum_{j=1}^n |a_{i,j}|$. Soit $x = (x_j)_{j\in[\![1;n]\!]} \in \mathbb{C}^n$ défini par :

$$\forall j \in [\![1;n]\!] \quad x_j = \begin{cases} \dfrac{A_{i_0,j}}{|A_{i_0,j}|} & \text{si } a_{i_0,j} \neq 0 \\ 0 & \text{sinon} \end{cases}.$$

Si $x = 0$, c'est que $\max_{i\in[\![1;n]\!]} \sum_{j=1}^n |a_{i,j}| = 0$, donc A est la matrice nulle. On a bien l'égalité souhaitée dans ce cas.
Sinon, $\|x\|_\infty = 1$, et :

$$\begin{aligned} \|Ax\|_\infty &= \max_{i\in[\![1;n]\!]} \sum_{j=1}^n |A_{i,j}x_j| \\ &= \max_{i\in[\![1;n]\!]} \sum_{j=1}^n |A_{i,j}|. \end{aligned}$$

Donc $\sup_{x\in\mathbb{C}^n, \|x\|_\infty=1} \|Ax\|_\infty \leqslant \max_{i\in[\![1;n]\!]} \sum_{j=1}^n |A_{i,j}|$, ce qui achève le calcul de $|||A|||_\infty$.

On choisit désormais $\delta > 0$ tel que $\sum_{j=i+1}^n \delta^{j-i}|T_{i,j}| < \varepsilon$. Alors, $|||A||| \leqslant \varepsilon + \max_{i\in[\![1;n]\!]} |t_{i,i}| = \varepsilon + \rho(A)$. □

Définition 3.

Soit $A \in \mathrm{GL}_n(\mathbb{R})$. On appelle *décomposition régulière de A* un couple (M, N) avec M facile à inverser. On définit la suite $(x_n)_{n\in\mathbb{N}}$ par $x_0 \in \mathbb{R}^n$ et :

$$\forall n \in \mathbb{N} \quad x_{n+1} = M^{-1}(Nx_n + b).$$

On dit que *la méthode itérative associée au couple (M, N) converge* lorsque la suite $(x_n)_{n\in\mathbb{N}}$ converge pour tout $x_0 \in \mathbb{R}^n$ (et dans ce cas, sa limite est la solution de $Ax = b$).

Théorème 4. ***Critère de convergence de la méthode itérative***

La méthode itérative converge si, et seulement si $\rho(M^{-1}N) < 1$.

<u>Preuve :</u> On pose $A = M - N \in \mathcal{M}_n(\mathbb{C})$ (on peut la voir comme matrice complexe). Soit x la solution exacte de $Ax = b$.
Pour $k \in \mathbb{N}$, on pose $e_k = x_k - x$. Alors :

$$\begin{aligned} e_{k+1} &= M^{-1}(Nx_k + B) - M^{-1}(Nx + B) \\ &= M^{-1}N(x_k - x) \\ &= M^{-1}Ne_k. \end{aligned}$$

Par récurrence immédiate :

$$\forall k \in \mathbb{N} \quad e_k = (M^{-1}N)^k e_0 = (M^{-1}N)^k (x_0 - x). \tag{1}$$

Distinguons deux cas :
$\star$ Si $\rho(M^{-1}N) < 1$, alors d'après le lemme, en prenant $\varepsilon = 1 - \rho(M^{-1}N) > 0$, on sait qu'il existe une norme subordonnée $|||\cdot|||$ (à une norme $\|\cdot\|$) telle que :

$$|||M^{-1}N||| < \rho(M^{-1}N) + \varepsilon = 1.$$

Pour $k \in \mathbb{N}$, il vient :

$$\begin{aligned} \|e_k\| &= \|(M^{-1}N)^k e_0\| \\ &\leqslant \underbrace{\left|\left|\left|M^{-1}N\right|\right|\right|^k}_{\underset{k\to+\infty}{\longrightarrow} 0} \|e_0\|, \end{aligned}$$

et $(\|e_k\|)_{k\in\mathbb{N}}$ converge vers 0.
$\star$ Si $\rho(M^{-1}N) \geqslant 1$, alors il existe $\lambda \in \mathrm{Sp}(M^{-1}N)$ tel que $|\lambda| \geqslant 1$.
Soit $u \in \mathbb{C}^n$ un vecteur propre associé à la valeur propre λ. On écrit $u = u_1 + iu_2$, avec $u_1, u_2 \in \mathbb{R}^n$. Alors :

$$(M^{-1}N)u = \lambda u.$$

Donc :

$$\forall k \in \mathbb{N} \quad (M^{-1}N)^k u = \lambda^k u. \tag{2}$$

Posons $x_0 = x + u_s$, avec $s \in \{1; 2\}$. D'après (1) :

$$\forall k \in \mathbb{N} \quad e_k = (M^{-1}N)^k u_s,$$

et d'après (2), il existe $s \in \{1; 2\}$ tel que $((M^{-1}N)^k u_s)_{k\in\mathbb{N}}$ ne converge pas, puisque $|\lambda| > 1$.
Pour cette condition initiale $x_0 = x + u_s$, la méthode itérative ne converge donc pas, ce qui achève la preuve. $\square$