Analyse Numérique
Corrigé Série 6

31 octobre 2019

1. (**Normes matricielles**) Pour $A \in \mathbb{R}^{m\times n}$, la **norme d'opérateur** est donnée par :

$$\|A\|_p := \max_{\|\boldsymbol{x}\|_p=1} \|A\boldsymbol{x}\|_p = \max_{\boldsymbol{x}\neq 0} \frac{\|A\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p}.$$

(a) Montrer que

$$\|A\|_2 = \sqrt{\text{plus grande valeur propre de } A^\mathsf{T} A}. \tag{1}$$

On remarque que $A^\mathsf{T}A$ est symétrique et on rappelle le théorème suivant.

> **Théorème 1.** (Théorème spectral pour diagonaliser les matrices symétriques).
> Soit A une matrice symétrique réelle, alors il existe une matrice Q orthogonale et une matrice Λ diagonale dont tous les coefficients sont réels, telles que la matrice A est égale à $Q\Lambda Q^\mathsf{T}$.

Sol.: *On a*

$$\|A\boldsymbol{x}\|_2^2 = (A\boldsymbol{x})^\mathsf{T} A\boldsymbol{x} = \boldsymbol{x}^\mathsf{T}(A^\mathsf{T}A)\,\boldsymbol{x} = \boldsymbol{x}^\mathsf{T} Q\Lambda Q^\mathsf{T}\boldsymbol{x} = \boldsymbol{y}^\mathsf{T}\Lambda\boldsymbol{y} = \sum_{i=1}^n \lambda_i y_i^2,$$

où $\boldsymbol{y} := Q^\mathsf{T}\boldsymbol{x}$ et $A^\mathsf{T}A = Q\Lambda Q^\mathsf{T}$ est la diagonalisation de la matrice symétrique $A^\mathsf{T}A$, avec $\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$ $(\lambda_1 \geq \ldots \geq \lambda_n > 0)$. Puisque le changement de variable $\boldsymbol{y} = Q^\mathsf{T}\boldsymbol{x}$ vérifie $\|\boldsymbol{y}\|_2 = \|\boldsymbol{x}\|_2$, on en déduit que

$$\|A\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|A\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{y}\|_2=1} \sqrt{\sum_{i=1}^n \lambda_i y_i^2}.$$

Cette quantité est maximale pour $\boldsymbol{y} = \boldsymbol{e_1}$ (premier vecteur de la base canonique de $\mathbb{R}^n$) et vaut alors $\sqrt{\lambda_1}$. □

(b) Pour $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{n\times m}$, montrer que les valeurs propres non nulles de AB et de BA sont égales ; en déduire que $\|A\|_2 = \left\|A^\mathsf{T}\right\|_2$.

Sol.: *On considère $\lambda \neq 0$ valeur propre de AB et on appelle $\boldsymbol{v} \neq 0$ son vecteur propre. On a donc*

$$AB\boldsymbol{v} = \lambda\boldsymbol{v}, \tag{2}$$

en multipliant (2) par B à gauche on obtient

$$BA\,(B\boldsymbol{v}) = B\lambda\boldsymbol{v} = \lambda\,(B\boldsymbol{v})\,,$$

où $B\boldsymbol{v} \neq 0$, car sinon $AB\boldsymbol{v} = 0 \Rightarrow \lambda = 0$ en contradiction avec l'hypothèse : donc λ est aussi valeur propre de BA. En utilisant (1), et sachant que $A^\mathsf{T}A$ et AA^T ont les mêmes valeurs propres non nulles, alors

$$\|A\|_2 = \left\|A^\mathsf{T}\right\|_2.$$

□

(c) Dans le cas où A est symétrique, montrer $\|A\|_2 = |\lambda_{max}|$ où λ_{max} est la plus grande valeur propre de A en valeur absolue.

Sol.: *On applique (1) en remarquant que les valeurs propres de A^2 sont les carrés des valeurs propres de A.* □

(d) **(⋆)** Pour $A \in \mathbb{R}^{m\times n}$, montrer que $\|A\|_1 = \max_{1\le j\le n} \left\{ \sum_{i=1}^{m} |a_{ij}| \right\}$.

Sol.: *On remarque que, selon la définition de norme d'opérateur, on a* $\|A\|_1 := \max_{\|\boldsymbol{x}\|_1=1} \|A\boldsymbol{x}\|_1$. *Notons* A_j *la jème colonne de la matrice* A *et considérons* $\|A\boldsymbol{x}\|_1$ *:*

$$\|A\boldsymbol{x}\|_1 = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} a_{ij}x_j \right| \le \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}x_j| = \sum_{j=1}^{n} |x_j| \sum_{i=1}^{m} |a_{ij}| = \sum_{j=1}^{n} |x_j|\, \|A_j\|_1.$$

Soit K, $1 \le K \le n$, *l'indice pour lequel* $\|A_j\|_1$ *atteint sa valeur maximale, c.-à-d.,* $\|A_K\|_1 = \max_{1\le j\le n} \{\|A_j\|_1\}$. *Alors*

$$\|A\boldsymbol{x}\|_1 \le \|A_K\|_1 \sum_{j=1}^{n} |x_j| = \|A_K\|_1 \|\boldsymbol{x}\|_1 = \|A_K\|_1,$$

puisque $\|\boldsymbol{x}\|_1 = 1$. *Maintenant on veut montrer que cette borne est atteinte pour un certain* $\boldsymbol{x}$. *Par exemple, on choisi* $\boldsymbol{x} = \boldsymbol{e_K}$, *où* $\boldsymbol{e_K} = (0, 0, \ldots, 0, \underbrace{1}_{\text{position } K}, 0, \ldots, 0)^{\boldsymbol{T}}$. *Ce choix permet d'atteindre la borne car* $A\boldsymbol{e_K} = A_K$. □

2. (**Différence finie centrée et erreur d'arrondi**) Soit un intervalle $I = [x_0 - \varepsilon, x_0 + \varepsilon]$ et $f \in \mathcal{C}^3(I)$, et notons

$$\mathit{\Delta}_f(h) = \frac{f(x_0+h) - f(x_0-h)}{2h}.$$

(a) **(⋆)** Montrer que

$$|\mathit{\Delta}_f(h) - f'(x_0)| \le C\, h^2, \quad \text{pour} \quad |h| \le \varepsilon,$$

avec $C = \frac{1}{6} \max_{\eta\in I} |f^{(3)}(\eta)|$.

Sol.: *En faisant le développement de Taylor jusqu'à l'ordre 2 pour* $f(x_0 - h)$ *et* $f(x_0 + h)$ *on trouve*

$$\begin{aligned} f(x_0+h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \frac{h^3}{6}f^{(3)}(\eta^+), \qquad \eta^+ \in [x_0, x_0+h]\,, \\ f(x_0-h) &= f(x_0) - hf'(x_0) + \frac{h^2}{2}f''(x_0) - \frac{h^3}{6}f^{(3)}(\eta^-), \qquad \eta^- \in [x_0-h, x_0]\,. \end{aligned}$$

Ainsi,

$$f(x_0+h) - f(x_0-h) = 2hf'(x_0) + \frac{h^3}{6}\left(f^{(3)}(\eta^+) + f^{(3)}(\eta^-)\right),$$

d'où

$$\mathit{\Delta}_f(h) - f'(x_0) = \frac{h^2}{12}\left(f^{(3)}(\eta^+) + f^{(3)}(\eta^-)\right).$$

On conclut que

$$|\mathit{\Delta}_f(h) - f'(x_0)| \le \frac{h^2}{12}\left(|f^{(3)}(\eta^+)| + |f^{(3)}(\eta^-)|\right) \le \frac{1}{6}\max_{\eta\in I}|f^{(3)}(\eta)|\, h^2.$$

□

(b) **(⋆)** L'erreur ci-dessus est un résultat pour le calcul exact. En prenant compte des erreurs d'arrondi, vérifier que le h optimal (c.-à-d., le h pour lequel l'erreur atteint son minimum) est donné par :

$$h_{\text{opt}} \approx \sqrt[3]{\varepsilon_{\text{mach}}},$$

où $\varepsilon_{\text{mach}}$ est la précision de la machine.
Indication : Séparer l'erreur en deux termes Exacte+Arrondi et étudier le comportement du nouveau terme à droite en h.

Sol.: *Sur la machine, $\mathit{\Delta}_f(h)$ sera remplacé par $\mathrm{fl}(\mathit{\Delta}_f(h))$ Ainsi,*

$$\begin{aligned}
\mathrm{fl}(\mathit{\Delta}_f(h)) &= (\mathrm{fl}(f(x_0+h)) \ominus \mathrm{fl}(f(x_0-h))) \oslash (2\mathrm{fl}(h)) \\
&= \frac{(f(x_0+h)(1+\varepsilon_1) - f(x_0-h)(1+\varepsilon_2)))(1+\varepsilon_3)}{2h(1+\varepsilon_4)}(1+\varepsilon_5) \\
&= \frac{(f(x_0+h)(1+\varepsilon_1) - f(x_0-h)(1+\varepsilon_2)))(1+\varepsilon_3)}{2h}(1-\varepsilon_4+O(\varepsilon_4^2))(1+\varepsilon_5) \\
&= \frac{f(x_0+h)-f(x_0-h)}{2h} + \frac{\varepsilon_1 f(x_0+h) - \varepsilon_2 f(x_0-h)}{2h} + \frac{f(x_0+h)-f(x_0-h)}{2h}(\varepsilon_3+\varepsilon_5-\varepsilon_4) \\
&\quad + O(\varepsilon_{\mathrm{mach}}^2),
\end{aligned}$$

avec $|\varepsilon_i| \leq \varepsilon_{\mathrm{mach}}$ pour $i = 1, \ldots, 5$. Ainsi, l'erreur totale en arithmétique flottante devient

$$\begin{aligned}
|\mathrm{fl}(\mathit{\Delta}_f(h)) - f'(x_0)| &\leq |\mathit{\Delta}_f(h) - f'(x_0)| + \left|\frac{(\varepsilon_1+\varepsilon_3+\varepsilon_5-\varepsilon_4)f(x_0+h) - (\varepsilon_2+\varepsilon_3+\varepsilon_5-\varepsilon_4)f(x_0-h)}{2h}\right| \\
&\leq \underbrace{C_1\, h^2}_{\text{erreur exacte}} + \underbrace{C_2\, \varepsilon_{\mathrm{mach}}\, h^{-1}}_{\text{perturbation dû aux erreurs d'arrondi}} .
\end{aligned}$$

Donc l'erreur d'arrondi explose lorsque $h \to 0$. Trouver h_{opt} revient à maximiser le second membre en fonction de h, en négligeant les constantes on trouve que le terme à droite est minimal lorsque $h_{\mathrm{opt}}^2 \approx \varepsilon_{\mathrm{mach}} h_{\mathrm{opt}}^{-1}$, c.-à-d., $h_{\mathrm{opt}} \approx \sqrt[3]{\varepsilon_{\mathrm{mach}}}$. □

(c) Vérifier cette erreur en MATLAB, par exemple, pour la fonction $f(x) = \sin(x)$ au point $x = \pi/3$. Vous devriez obtenir une figure comme la Figure 1, qui montre que l'erreur décroît jusqu'à atteindre une valeur minimale en correspondance de $h_{\mathrm{opt}} \approx \sqrt[3]{\varepsilon_{\mathrm{mach}}} \approx 6.0555 \times 10^{-6}$ et puis augmente de nouveau pour des h plus petits que h_{opt}.

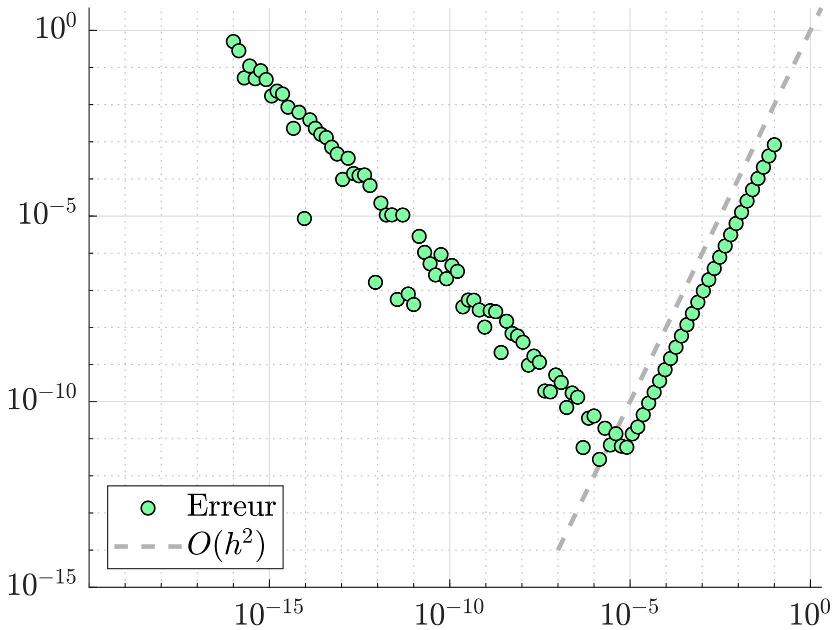

FIGURE 1 – Comportement logarithmique de l'erreur pour la fonction $f(x) = \sin(x)$ au point $x = \frac{\pi}{3}$ en fonction de h : l'erreur en arithmétique exacte (ligne) et l'erreur en arithmétique flottante (cercle).