Analyse Numérique
Corrigé Série 8

14 novembre 2019

1. **Stabilité backward**. Considérer le problème suivant : résoudre un système linéaire inversible, c.-à-d., trouver $\boldsymbol{x} \in \mathbb{R}^s$ tel que $A\boldsymbol{x} = \boldsymbol{b}$, étant donnés $A \in \mathbb{R}^{s\times s}$ inversible et $\boldsymbol{b} \in \mathbb{R}^s$.

   (a) Écrire ce problème comme l'évaluation d'une fonction $f(\boldsymbol{z}) : \mathbb{R}^n \to \mathbb{R}^m$, c.-à-d., identifier les données $\boldsymbol{z}$, la fonction f et définir exactement $\mathbb{R}^n$ et $\mathbb{R}^m$.
   Indication : Ne confondez pas le $\boldsymbol{x}$ comme solution de $A\boldsymbol{x} = \boldsymbol{b}$ avec le $\boldsymbol{z}$ comme variable de $f(\boldsymbol{z})$.
   Sol.: *Les données sont A et $\boldsymbol{b}$. On définit $\boldsymbol{z} := \binom{\boldsymbol{a}}{\boldsymbol{b}} \in \mathbb{R}^{s^2+s}$, où $\boldsymbol{a} = (\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_s})^{\boldsymbol{T}}$, étant $\boldsymbol{a_i}$ la i-ème ligne de A.*
   La fonction est $f\colon \mathbb{R}^n \to \mathbb{R}^m$, définie par $\boldsymbol{z} \mapsto \boldsymbol{x} = A^{-1}\boldsymbol{b}$, avec $n = s^2 + s$ et $m = s$.

   (b) On note $\tilde{f}$ l'algorithme pour résoudre $A\boldsymbol{x} = \boldsymbol{b}$, ce qui donne une solution $\tilde{\boldsymbol{x}} \in \mathbb{R}^s$ t.q.
$$(A + \delta A)\tilde{\boldsymbol{x}} = \boldsymbol{b} + \delta\boldsymbol{b},$$
   où
$$\delta A \in \mathbb{R}^{s\times s} \quad \text{t.q.} \quad \|\delta A\|_\infty \le \varepsilon_{\text{mach}}\|A\|_\infty \quad \text{et} \quad \delta\boldsymbol{b} \in \mathbb{R}^s \quad \text{t.q.} \quad \|\delta\boldsymbol{b}\|_\infty \le \varepsilon_{\text{mach}}\|\boldsymbol{b}\|_\infty.$$
   Montrer que $\tilde{f}$ est backward stable au sens de la définition 3.11 pour la norme $\|\cdot\|_\infty$. Quelle est la valeur de la constante C ?
   Sol.: *Pour montrer que $\tilde{f}$ est backward stable au sens de la définition 3.11 il faut montrer que*
$$\forall\, \boldsymbol{z},\ \exists\, \tilde{\boldsymbol{z}} \ \ t.q. \ \ \tilde{f}(\boldsymbol{z}) = f(\tilde{\boldsymbol{z}}) \ \ et \ \ \frac{\|\tilde{\boldsymbol{z}} - \boldsymbol{z}\|}{\|\boldsymbol{z}\|} \le C\varepsilon_{\text{mach}} + o(\varepsilon_{\text{mach}}).$$
   On peut définir $\tilde{f} : \mathbb{R}^{s^2+s} \to \mathbb{R}^s$ de la même façon que f dans (a).
   - *Pour la première partie, $\tilde{f}(\boldsymbol{z}) = f(\tilde{\boldsymbol{z}})$, on a par définition*
$$\tilde{f}\left(\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix}\right) = f\left(\begin{pmatrix} \boldsymbol{a} + \delta\boldsymbol{a} \\ \boldsymbol{b} + \delta\boldsymbol{b} \end{pmatrix}\right),$$
   où $\delta\boldsymbol{a}$ est le vecteur des composantes de δA, $\delta\boldsymbol{a} \in \mathbb{R}^{s^2}$.
   - *Pour la deuxième partie, $\frac{\|\tilde{\boldsymbol{z}}-\boldsymbol{z}\|}{\|\boldsymbol{z}\|} \le C\varepsilon_{\text{mach}} + o(\varepsilon_{\text{mach}})$, il suffit de montrer que*
$$\frac{\left\|\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} - \begin{pmatrix} \boldsymbol{a} + \delta\boldsymbol{a} \\ \boldsymbol{b} + \delta\boldsymbol{b} \end{pmatrix}\right\|}{\left\|\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix}\right\|} = \frac{\left\|\begin{pmatrix} \delta\boldsymbol{a} \\ \delta\boldsymbol{b} \end{pmatrix}\right\|}{\left\|\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix}\right\|} \le C\varepsilon_{\text{mach}} + o(\varepsilon_{\text{mach}}).$$
   Or pour la norme $\|\cdot\|_\infty$ on a
$$\begin{aligned}\left\|\begin{pmatrix} \delta\boldsymbol{a} \\ \delta\boldsymbol{b} \end{pmatrix}\right\|_\infty &= \max\{\|\delta\boldsymbol{a}\|_\infty,\ \|\delta\boldsymbol{b}\|_\infty\} \le \max\{\varepsilon_{\text{mach}}\,\|A\|_\infty,\ \varepsilon_{\text{mach}}\,\|\boldsymbol{b}\|_\infty\} \\ &\le \max\{s\,\varepsilon_{\text{mach}}\,\|\boldsymbol{a}\|_\infty,\ \varepsilon_{\text{mach}}\,\|\boldsymbol{b}\|_\infty\} \\ &\le s\,\varepsilon_{\text{mach}} \max\{\|\boldsymbol{a}\|_\infty,\ \|\boldsymbol{b}\|_\infty\} = s\,\varepsilon_{\text{mach}} \left\|\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix}\right\|_\infty.\end{aligned}$$
   Donc $C = s$ et $\tilde{f}$ est backward stable.

(c) Montrer que $\tilde{f}$ est aussi backward stable pour la norme $\|\cdot\|_1$. Quelle est la constante C ?
Sol.: *On a obtenu*

$$\left\| \begin{pmatrix} \delta\boldsymbol{a} \\ \delta\boldsymbol{b} \end{pmatrix} \right\|_\infty \leq s\,\varepsilon_{\text{mach}} \left\| \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \right\|_\infty .$$

En utilisant l'équivalence des normes vectorielles sur $\mathbb{R}^n$ *on a*

$$\frac{1}{s^2+s} \left\| \begin{pmatrix} \delta\boldsymbol{a} \\ \delta\boldsymbol{b} \end{pmatrix} \right\|_1 \leq \left\| \begin{pmatrix} \delta\boldsymbol{a} \\ \delta\boldsymbol{b} \end{pmatrix} \right\|_\infty \leq s\,\varepsilon_{\text{mach}} \left\| \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \right\|_\infty \leq s\,\varepsilon_{\text{mach}} \left\| \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \right\|_1 .$$

Donc $\tilde{f}$ *est backward stable pour la norme* $\|\cdot\|_1$ *avec* $C = s(s^2+s)$.

(d) Qu'est-ce qu'on peut conclure par rapport à la stabilité de ce problème ?
Sol.: *Dans la définition 3.11 on demande que* C *ne soit pas trop grande. On voit que dans ce problème* C *dépend de* s*, donc elle peut devenir énorme pour des matrices de grande taille. Cependant, ce qui est importante est que pour* s *fixé,* C *reste la même* $\forall A \in \mathbb{R}^{s\times s}$, $b \in \mathbb{R}^s$.

2. **Erreurs des formules de quadrature**. Considérer les intégrales $\int_0^1 x^4\,\mathrm{d}x$ et $\int_0^1 x^5\,\mathrm{d}x$.

(a) Écrire les erreurs $E_s(f,0,1) := \int_0^1 f(x)\,\mathrm{d}x - \sum_{i=1}^s b_i f(c_i)$ pour approcher ces deux intégrales avec la règle du trapèze et de Simpson.
Sol.: *Pour approcher* $\int_0^1 x^4\,\mathrm{d}x = \frac{1}{5}$ *et* $\int_0^1 x^5\,\mathrm{d}x = \frac{1}{6}$ *avec la règle du trapèze on obtient respectivement les erreurs* $E_{\text{trap},1} = \frac{1}{5} - \frac{1}{2} = -\frac{3}{10}$ *et* $E_{\text{trap},2} = \frac{1}{6} - \frac{1}{2} = -\frac{1}{3}$. *Pour la règle de Simpson, on a* $E_{\text{Simp},1} = \frac{1}{5} - \frac{5}{24} = -\frac{1}{120}$ *et* $E_{\text{Simp},2} = \frac{1}{6} - \frac{9}{48} = -\frac{1}{48}$.

(b) Trouver la valeur de la constante α pour laquelle la règle du trapèze donne le résultat exact de $\int_0^1 (x^5 - \alpha x^4)\,\mathrm{d}x$.
Sol.: *L'erreur pour approcher* $\int_0^1 (x^5 - \alpha x^4)\,\mathrm{d}x = \frac{1}{6} - \frac{\alpha}{5}$ *avec la règle du trapèze est* $E_{\text{trap}} = \frac{1}{6} - \frac{\alpha}{5} - (\frac{1}{2} - \frac{\alpha}{2}) = \frac{3}{10}\alpha - \frac{1}{3}$. *Donc la valeur de la constante* α *pour laquelle la règle du trapèze donne le résultat exact est* $\alpha = \frac{10}{9}$.

(c) Montrer que, pour $\int_0^1 (x^5 - \alpha x^4)\,\mathrm{d}x$, la règle du trapèze donne un résultat plus précis de la règle de Simpson quand $\frac{15}{14} < \alpha < \frac{85}{74}$.
Sol.: *L'erreur pour approcher* $\int_0^1 (x^5 - \alpha x^4)\,\mathrm{d}x$ *avec Simpson est* $E_{\text{Simp}} = \frac{1}{6} - \frac{\alpha}{5} - (\frac{9}{48} - \frac{5}{24}\alpha) = \frac{1}{120}\alpha - \frac{1}{48}$. *La règle du trapèze donne un résultat plus précis de la règle de Simpson quand*

$$\left|\tfrac{3}{10}\alpha - \tfrac{1}{3}\right| < \left|\tfrac{1}{120}\alpha - \tfrac{1}{48}\right| .$$

À l'aide d'un graphique on remarque que cette inégalité est satisfaite quand α *se trouve entre les deux valeurs qui sont solutions de*

$$\tfrac{3}{10}\alpha - \tfrac{1}{3} = \tfrac{1}{120}\alpha - \tfrac{1}{48} \qquad \textit{et} \qquad \tfrac{3}{10}\alpha - \tfrac{1}{3} = -\left[\tfrac{1}{120}\alpha - \tfrac{1}{48}\right],$$

ce qui donne les solutions cherchées $\frac{15}{14}$ *et* $\frac{85}{74}$.

3. **($\star$) Formules de quadrature symétriques**. Soit (b_i, c_i) $(i = 1, \ldots, s)$ une formule de quadrature symétrique, c'est-à-dire avec $b_{s+1-i} = b_i$, $c_{s+1-i} = 1 - c_i$, $i = 1, \ldots, s$. Le but de cet exercice est de montrer que l'ordre de la formule de quadrature est pair. Autrement dit, si la méthode est d'ordre $2m-1$ avec $m \in \mathbb{N}^*$, alors elle est d'ordre $2m$.

(a) Montrer que tout polynôme $g(t)$ de degré $2m-1$ peut être écrit sous la forme

$$g(t) = C\,(t - 1/2)^{2m-1} + g_1(t),$$

où C est une constante et $g_1(t)$ est un polynôme de degré $\leq 2m-2$.
Sol.: *Soit* C *le coefficient qui multiplie* t^{2m-1} *dans le polynôme* $g(t)$, *c.-à-d.,* $g(t) = C\,t^{2m-1} + \ldots$. *Comme les polynômes* $g(t)$ *et* $C\,(t-1/2)^{2m-1}$ *ont le même coefficient qui multiplie* t^{2m-1}, *on en déduit que* $g_1(t) = g(t) - C\,(t-1/2)^{2m-1}$ *est un polynôme de degré* $\leq 2m-2$.

(b) Montrer que la formule de quadrature est exacte pour approcher l'intégrale $\int_0^1 (t-1/2)^{2m-1}\,\mathrm{d}t = 0$.

Sol.: *On a l'intégrale exacte* $\int_0^1 (t-1/2)^{2m-1}\,\mathrm{d}t = \frac{1}{2m}[(t-1/2)^{2m}]|_0^1 = \frac{1}{2m}[(1/2)^{2m} - (1/2)^{2m}] = 0$.
En utilisant la symétrie de la formule de quadrature, on a que pour tout i $b_i(c_i - 1/2)^{2m-1} + b_{s+1-i}(c_{s+1-i} - 1/2)^{2m-1} = 0$.

Ainsi, l'approximation numérique de $\int_0^1 (t-1/2)^{2m-1}\,\mathrm{d}t$ *est nulle, donc exacte.*

(c) Conclure.

Sol.: *Par linéarité de l'erreur de quadrature, l'erreur pour* $g(t)$ *est égale à l'erreur pour* $C\,(t - 1/2)^{2m-1}$ *(nulle d'après* (3b)*), plus l'erreur pour* $g_1(t)$ *(nulle par hypothèse sur l'ordre* $2m-1$ *de la formule de quadrature car* g_1 *est un polynôme de degré* $\leq 2m-2$*). La formule de quadrature est donc d'ordre* $2m$*.*