# Algorithmique numérique

Contrôle de connaissances - lundi 13 janvier 2014 à 10h.

*Durée :* deux heures.
*Documents autorisés :* notes prises en cours.

Dans toute la suite $\mathbb{K}$ désigne un corps et $\mathbb{F}$ désigne l'ensemble des nombres flottants en base 2 et précision $p$ (sans contrainte sur la plage de l'exposant). Les notations $\mathrm{fl}(\cdot)$, $\mathrm{fl}_{\downarrow}(\cdot)$, $\mathrm{fl}_{\uparrow}(\cdot)$ précisent le fait que les opérations entre parenthèses sont effectuées respectivement en arrondi au plus proche, vers le bas, vers le haut. L'unité d'arrondi est désignée par $u$.

1. Soient $T \in \mathbb{K}^{n\times n}$ une matrice triangulaire supérieure inversible et $b \in \mathbb{K}^n$ un vecteur. On s'intéresse dans cette question à la résolution du système linéaire $Tx = b$. On notera $x_i$ le $i$ème élément de $x$, $b_i$ le $i$ème élément de $b$, et $t_{i,j}$ l'élément situé sur la $i$ème ligne et la $j$ème colonne de $T$.
   (a) Rappeler le coût $\mathrm{DP}(n)$ d'un produit scalaire en dimension $n$.
   (b) Exprimer $x_i$ en fonction de $x_{i+1}, \ldots, x_n$ et des éléments de $b$ et $T$.
   (c) En déduire le coût de résolution de $Tx = b$ en fonction de DP puis en fonction de $n$.
   (d) Savoir résoudre $Tx = b$ est-il utile pour résoudre des systèmes linéaires plus généraux, comme par exemple $Ax = b$ avec $A$ fortement régulière ? Pourquoi ?
2. Expliquez en quelques phrases ce que sont l'analyse directe et l'analyse inverse.
3. Soit $T \in \mathbb{K}^{n\times n}$ une matrice triangulaire supérieure. On s'intéresse dans cette question au calcul du déterminant $\det(T)$ en arithmétique flottante.
   (a) Justifier brièvement pourquoi $\det(T)$ est égal au produit $\prod_{i=1}^{n} t_{i,i}$.
   (b) On suppose que l'on évalue $\det(T)$ à l'aide de l'algorithme naïf suivant :

   $\widehat{r}_1 := t_{1,1}$;
   Pour $i$ de 2 à $n$ faire
   $\widehat{r}_i := \mathrm{fl}(\widehat{r}_{i-1} \times t_{i,i})$;
   Renvoyer $\widehat{r}_n$.

      i. En utilisant le premier modèle standard de l'arithmétique flottante, donner une majoration de l'erreur inverse (relative) commise sur la valeur $\widehat{r}_n$ retournée.
      ii. Déduire une majoration de l'erreur directe (relative) commise sur $\widehat{r}_n$. Quelle est la principale différence entre cette borne et celle vue en cours pour la sommation naïve de $n$ nombres flottants ?
4. Rappelez la définition mathématique d'une opération entre deux intervalles.
   Déduisez-en les formules pour la soustraction de deux intervalles $\mathbf{a} - \mathbf{b}$ où $\mathbf{a}$ représente l'intervalle $[\underline{a}, \overline{a}]$ et $\mathbf{b}$ représente l'intervalle $[\underline{b}, \overline{b}]$.
   De la même manière, rappelez la définition mathématique d'une fonction appliquée à un intervalle.
   Déduisez-en la formule pour la racine carrée d'un intervalle $\sqrt{\mathbf{a}}$ où là encore $\mathbf{a}$ représente l'intervalle $[\underline{a}, \overline{a}]$.
5. Soient $\mathbf{a} = [\underline{a}, \overline{a}]$ et $\mathbf{b} = [\underline{b}, \overline{b}]$ deux intervalles dont les bornes sont des nombres flottants. Le but de l'exercice est de trouver un algorithme (différent de celui vu en cours) pour calculer un intervalle $\mathbf{c} = [\underline{c}, \overline{c}]$ tel que $\mathbf{a} \times \mathbf{b} \subset \mathbf{c}$.

(a) Soient $m_{\mathbf{a}}, r_{\mathbf{a}} \in \mathbb{F}$ définis par $m_{\mathbf{a}} = \mathrm{fl}_{\downarrow}\left(\frac{\underline{a}+\overline{a}}{2}\right)$ et $r_{\mathbf{a}} = \mathrm{fl}_{\uparrow}(\overline{a} - m_{\mathbf{a}})$. Montrez que $r_{\mathbf{a}} \geqslant 0$, puis que $\forall a \in \mathbf{a}, |a - m_{\mathbf{a}}| \leqslant r_{\mathbf{a}}$.

(b) On définit similairement les deux flottants $m_{\mathbf{b}} = \mathrm{fl}_{\downarrow}\left(\frac{\underline{b}+\overline{b}}{2}\right)$ et $r_{\mathbf{b}} = \mathrm{fl}_{\uparrow}(\overline{b} - m_{\mathbf{b}})$. Soit $r_{\mathbf{c}} = \mathrm{fl}_{\uparrow}\big(|m_{\mathbf{a}}|r_{\mathbf{b}} + r_{\mathbf{a}}(|m_{\mathbf{b}}| + r_{\mathbf{b}})\big)$. Montrez que $\forall a \in \mathbf{a}$ et $\forall b \in \mathbf{b}$, $|ab - m_{\mathbf{a}}m_{\mathbf{b}}| \leqslant r_{\mathbf{c}}$.

(c) Montrez que $\forall a \in \mathbf{a}$ et $\forall b \in \mathbf{b}$, on a $\mathrm{fl}_{\downarrow}(m_{\mathbf{a}}m_{\mathbf{b}} - r_{\mathbf{c}}) \leqslant ab \leqslant \mathrm{fl}_{\uparrow}(m_{\mathbf{a}}m_{\mathbf{b}} + r_{\mathbf{c}})$.

(d) Déduire de la question précédente un algorithme flottant qui, étant donnés $\mathbf{a}$ et $\mathbf{b}$, calcule deux flottants $\underline{c}$ et $\overline{c}$ tels que $\mathbf{a} \times \mathbf{b} \subset \mathbf{c}$. Appelez `setround(-1)` ou `setround(+1)` pour passer respectivement dans le mode d'arrondi vers $-\infty$ ou vers $+\infty$. Supposez que le processeur est initialement dans le mode d'arrondi au plus proche.

6. On étudie le système d'équations

$$f(x) = 0 \text{ avec } f(x) = \begin{pmatrix} x_1^2 + 4x_2^2 - 4 \\ x_1 + 2x_2 - 2 \end{pmatrix}.$$

(a) Dessinez les courbes d'équations $x_1^2 + 4x_2^2 - 4 = 0$ et $x_1 + 2x_2 - 2 = 0$, puis vérifiez que les points d'intersection entre ces deux courbes sont $(0, 1)$ et $(2, 0)$.

(b) Écrivez la Jacobienne de $f$ en un point $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.

(c) Écrivez un pas de l'itération de Newton. Écrivez le système linéaire à résoudre à chaque pas.

(d) Écrivez un pseudo-code en Scilab implantant l'algorithme itératif de Newton.

(e) Appliquez un pas de Newton pour obtenir $x^1$ en partant de $x^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Voici les itérés suivants :

| numéro d'itération | $x_1$ | $x_2$ |
|---|---|---|
| 1 | | |
| 2 | 2.083333333333333 | -0.041666666666667 |
| 3 | 2.003205128205128 | -0.001602564102564 |
| 4 | 2.000005120013107 | -0.000002560006553 |
| 5 | 2.000000000013107 | -0.000000000006554 |

(f) Calculez la factorisation LU de la matrice $J = \begin{pmatrix} 2 & 8 \\ 1 & 2 \end{pmatrix}$.

Utilisez-la pour résoudre le système linéaire $J \cdot y = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

(g) Appliquez un pas de Newton pour obtenir $x^1$ en partant de $x^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Voici les itérés suivants :

| numéro d'itération | $x_1$ | $x_2$ |
|---|---|---|
| 1 | | |
| 2 | -0.083333333333333 | 1.041666666666667 |
| 3 | -0.003205128205128 | 1.001602564102564 |
| 4 | -0.000005120013107 | 1.000002560006554 |
| 5 | -0.000000000013107 | 1.000000000006554 |

(h) Pour conclure, que constatez-vous en observant ces deux suites d'itérés ?

* * *