DIRO
IFT 2425

EXAMEN INTRA

Max Mignotte

DIRO, Département d'Informatique et de Recherche Opérationnelle, local 2377
Http : //www.iro.umontreal.ca/∼mignotte/ift2425/
E-mail : mignotte@iro.umontreal.ca

Date : *27/02/2007*

I	Mesure d'Incertitude et Amplification d'Erreur (20 pts).
II	Erreur en Arithmétique Flottante (20 pts).
III	Méthode du Point Fixe et de Newton (33 pts).
IV	Factorisation LU (23 pts)
V	Interpolation de Newton (15 pts)
Total	111 points.

Tous documents personnels, calculatrices et calculateurs autorisés

I. Mesure d'Incertitude et Amplification d'Erreur (20 pts)

Soit la fonction suivante,

$$f(x,y,z) = \alpha \, \frac{x^p \, y}{\sqrt{z}}$$

Dans laquelle $\alpha = 2$ et $p = 10$ connu exactement sans erreur, $x = 2.000$ a tous ses chiffres significatifs exacts (cse) et $y = 3.000$, $z = 4.000$ sont connu avec 0.2% et 0.3% d'erreur relative respectivement.

1. Donner l'approximation de $f(x,y,z)$ (i.e., $f^*(x,y,z)$).
   <**2 pts**>
2. Donner la borne supérieure de l'erreur absolue (i.e., incertitude totale de $f(x,y,z)$ ou $\Delta f(x,y,z)$) par la méthode de propagation d'erreur (utilisant l'approximation de Taylor de la fonction au premier ordre).
   <**8 pts**>
3. Souligner les chiffres significatifs "exacts" (cse) de ce résultat. Arrondir cette approximation aux nombres de cse adéquat.
   <**3 pts**>
4. Laquelle des variables (x,y,z) contribue le plus à l'erreur ?
   <**2 pts**>
5. Aurait-on pu utiliser la méthode de la fourchette pour trouver la borne supérieure de l'incertitude totale de $f(x,y,z)$? Si oui dite comment et faîtes le (et donner cette borne supérieure de l'incertitude totale de $f(x,y,z)$), si non justifier bien votre réponse.
   <**5 pts**>

Réponse

$\boxed{\textbf{1.}}$

On a pour valeur approchée,

$$f^*(x,y,z) = f(x^*,y^*,z^*) = 2 \times \frac{2^{10} \times 3.0}{\sqrt{4.0}} = 3072 \qquad < \textbf{2 pts} >$$

$\boxed{\textbf{2.}}$

Pour calculer l'incertitude (ou la borne supérieure de l'erreur absolue) de $f(x,y,z)$, on doit calculer la différentielle $\Delta f(x,y,z)$, i.e.,

$$\Delta f(x,y,z) = \left| \frac{\alpha \, p \, y}{\sqrt{z}} \, x^{p-1} \right| \Delta x \; + \; \left| \frac{\alpha \, x^p}{\sqrt{z}} \right| \Delta y \; + \; \left| -\frac{\alpha \, x^p \, y}{2 \, z^{3/2}} \right| \Delta z \qquad < \textbf{2 pts} >$$

Avec,

$$\begin{aligned} \Delta x = 0.5 \times 10^{-3} \qquad & <\textbf{1 pt}> \\ \Delta y = 6 \times 10^{-3} \qquad & <\textbf{1 pt}> \\ \Delta z = 12 \times 10^{-3} \qquad & <\textbf{1 pt}> \end{aligned}$$

Soit numériquement,

$$\begin{aligned} \Delta f(x=2.0, y=3.0, z=4.0) \quad &= \quad \left| \frac{2 \times 10 \times 3}{\sqrt{4}} \, 2^9 \right| \Delta x \; + \; \left| \frac{2 \times 2^{10}}{\sqrt{4}} \right| \Delta y \; + \; \left| -\frac{2 \times 2^{10} \times 3}{2 \times 4^{3/2}} \right| \Delta z \\ &= \quad 15360 \, \Delta x \; + \; 1024 \, \Delta y \; + \; 384 \, \Delta z \\ &= \quad (15360 \times 0.5 \times 10^{-3}) \; + \; (1024 \times 6 \times 10^{-3}) \; + \; (384 \times 12 \times 10^{-3}) \\ &= \quad 7.68 + 6.144 + 4.608 \\ &= \quad 18.432 \qquad < \textbf{3 pts} > \end{aligned}$$

$\boxed{\textbf{3.}}$

On obtient donc $f(.) \approx 3072 \pm 18.432$. Comme $18.432 < 0.5 \times 10^2$, le rang du dernier chiffre significatif est 2 et on peut écrire $f(.) \approx \underline{30}72 \pm 18.432$ <**1.5 pt**>. Si on arrondi finalement cette estimation aux nombres de cse adéquat, on obtient,

$$f(.) \approx 30 \times 10^2 \pm 50 \quad (\text{ou} \pm 18.5) \quad < \mathbf{1.5\ pt} >$$

$\boxed{\textbf{4.}}$

Le résultat de la question précédente nous montre que c'est la variable x qui contribue le plus à l'erreur. <**2 pts**>

$\boxed{\textbf{5.}}$

Oui on peut utiliser la méthode de la fourchette ; En utilisant les valeurs maximales pour x, y et la valeur minimale pour z, on peut déterminer la borne maximale de f et inversement sa borne minimale.

On a donc,

$$f^* - \Delta f = f(x_{\min}, y_{\min}, z_{\max}) \quad \text{et} \quad f^* + \Delta f = f(x_{\max}, y_{\max}, z_{\min})$$

Et donc,

$$\begin{aligned} 2\Delta f &= f(x_{\max}, y_{\max}, z_{\min}) - f(x_{\min}, y_{\min}, z_{\max}) \\ &= f(2 + 0.5 \times 10^{-3}, 3 + 6 \times 10^{-3}, 4 - 12 \times 10^{-3}) - f(2 - 0.5 \times 10^{-3}, 3 - 6 \times 10^{-3}, 4 + 12 \times 10^{-3}) \\ &\approx 3090.487236 - 3053.622973 \\ &\approx 36.864263 \end{aligned}$$

Donc $\Delta f \approx 18.4321315$ <**5 pts**>

Nota : On peut comparer cette valeur (exacte) avec celle approchée obtenue avec la méthode différentielle et constater qu'elles sont toute les deux très proches ! Les deux méthodes nous donnent le même nombre de cse pour le résultat final même si la méthode différentielle n'utilise qu'une approximation du premier ordre.

II. Erreur en Arithmétique Flottante (20 pts)

1. Expliquer pourquoi le calcul numérique de ces expressions,

   (a) $\frac{\sqrt{1+x^2}-1}{x^2}$ quand $x \to 0$

   (b) $(\ln x) - 1$ quand $x \to e$

   (c) $\frac{\sin(x)}{x-\sqrt{x^2-1}}$ quand $x \to ?$

   peuvent conduire à des problèmes d'erreurs numériques. Expliquer pourquoi (i.e., identifier et citer le problème numérique associé) et proposer une formule équivalente qui permettrait d'éviter ces problèmes numériques et qui permettrait d'augmenter la précision des calculs de ces trois expressions.
   <**9 pts**>

2. Trouver la limite de l'expression (a) quand $x \to 0$.
   <**2 pts**>.
   A partir de quel n votre calculatrice vous donne donc, pour $x = 10^{-n}$, un calcul erroné dans le calcul de l'expression (a) ($\frac{\sqrt{1+x^2}-1}{x^2}$) ?
   <**3 pts**>

3. Pour tout $x_0 > -1$, la séquence définie récursivement par

$$x_{n+1} = 2^{n+2}\left(\sqrt{1+2^{-(n+1)}\,x_n} - 1\right) \qquad (n \geq 0)$$

converge vers ≈ 0.810930 pour $x_0 = 1$.

(a) Identifiez le type de problème numérique auquel sera confronté cette formule itérative.
<**2 pts**>

(b) Arrangez cette formule de telle façon qu'elle évite donc le maximum de perte de précision en arithmétique flottante.
<**4 pts**>

Réponse

$\boxed{\textbf{1.}}$

Dans les trois cas, on aura un problème d'annulation de cse due à la soustraction de deux nombres approximés quasi égaux <**1 pt**>. Pour éviter ce problème, on peut écrire ces relations de telle façon que plus aucune soustraction de deux nombres incertains (i.e., imprécis) quasi égaux apparaissent.

On utilise le conjugué de l'expression $\sqrt{1+x^2}-1$ que l'on multiplie au numérateur et au dénominateur dans la première expression, on utilise les propriétés du logarithme ($\ln(a) - \ln(b) = \ln(a/b)$) et le fait que $\ln(\mathrm{e}) = 1$ dans la seconde expression et finalement en multipliant $x+\sqrt{1+x^2}$ au numérateur et dénominateur pour la dernière expression, on obtient les trois expressions équivalentes suivantes

$$\begin{aligned}
\frac{\sqrt{1+x^2}-1}{x^2} &= \frac{1}{\sqrt{1+x^2}+1} \qquad < \textbf{3 pts} > \\
\ln(x) - 1 &= \ln\left(\frac{x}{\mathrm{e}}\right) \qquad < \textbf{3 pts} > \\
\frac{\sin(x)}{x-\sqrt{x^2-1}} &= \left(x+\sqrt{x^2-1}\right)\,\sin(x) \quad (x \to \infty) \qquad < \textbf{3 pts} >
\end{aligned}$$

Nota : D'autres possibilités de formulation sont possibles

$\boxed{\textbf{2.}}$

On utilise donc le fait que $\frac{\sqrt{1+x^2}-1}{x^2} = \frac{1}{\sqrt{1+x^2}+1}$ pour trouver facilement que la limite, de cette expression quand $x \to 0$ est 0.5. <**2 pts**>

En partant de l'expression $\frac{\sqrt{1+x^2}-1}{x^2}<$, ma calculatrice me donne un calcul complètement erroné à partir de $x = 10^{-5}$. <**2 pts**>

$\boxed{\textbf{3.}}$

Pour de grand n, $2^{-(n+1)}$ décroîtra très vite vers 0. il en sera de même pour $2^{-(n+1)}x_n$. De ce fait, l'expression $\sqrt{1+2^{-(n+1)}\,x_n}-1$ fera apparaître la soustraction de deux nombres approximés quasi égaux avec une perte de précision qui en résultera. <**1 pt**> On a donc tout intérêt (numérique) à écrire cette relation différemment. Dans notre cas, c'est possible et

$$x_{n+1} = 2^{n+2}\left(\sqrt{1+2^{-(n+1)}\,x_n}-1\right) = 2^{n+2}\left(\frac{\left(2^{-(n+1)}\,x_n\right)}{\sqrt{1+2^{-(n+1)}\,x_n}+1}\right) = \frac{2\,x_n}{\sqrt{1+2^{-(n+1)}\,x_n}+1}$$

<**3 pts**>

Nota : Sur mon ordinateur, l'utilisation de la relation récursive brute pose problème dès la 50 ième itération tandis que la relation itérative ne faisant pas apparaître la soustraction de deux nombres approximés quasi égaux ne pose aucun problème.

III. Méthode du Point Fixe et de Newton (33 pts)

On se propose de trouver numériquement une valeur approchée de la racine de la fonction,

$$f(x) = 4\, e^{x/4} - 6 \tag{1}$$

1. **Méthode du point fixe**
   (a) Montrer qu'il existe une racine unique r pour cette Eq. (1) dans l'intervalle $J = [0, 2]$ et trouvez cette racine analytiquement.
   $<$**5 pts**$>$
   (b) Pour trouver numériquement cette racine, on se propose d'étudier les trois suites itératives ou méthodes de point fixes suivant ; $x_{n+1} = \Upsilon_i(x_n)$ pour $i = 1, 2, 3$,
      i. $\Upsilon_1(x) = x + 4e^{x/4} - 6$
      ii. $\Upsilon_2(x) = x - 4 + 6e^{-x/4}$
      iii. $\Upsilon_3(x) = x - \frac{e^{x/4}}{16} + \frac{3}{32}$

   En vous aidant aussi de la racine analytiquement trouvée à la question précédente, établissez laquelle ou lesquelles de ces suites itératives sont convergentes si on prend x_0 dans J et, dans le cas affirmatif, donner son ordre de convergence.
   $<$**14 pts**$>$
   (c) En utilisant la suite itérative $x_{n+1} = \Upsilon_3(x_n)$, déterminer analytiquement, grâce au théorème des accroissements finis (ou de la valeur moyenne), une majoration du type $|r_n - r| \leq K$, ou r_n désigne la valeur approchée, à la n-ième itération, de cette racine par la méthode itérative du point fixe. En déduire le nombre d'itérations n_* nécessaire pour obtenir, par cette méthode, une valeur approchée à 10^{-1} près de cette racine.
   $<$**8 pts**$>$
   Aurait-on pu, sans utiliser ce théorème des accroissement finis, prévoir que la vitesse de convergence de cette méthode du point fixe soit si lente ? Comment ? Justifier votre réponse.
   $<$**2 pts**$>$
2. **Méthode de Newton**
   (a) Soit $\Upsilon_4(x)$, la fonction intervenant dans la méthode itérative de Newton pour la résolution de la racine r de l'équation (1) dans J. Donner $\Upsilon_4(x)$ ainsi que la relation itérative $r_{n+1} = \Upsilon_4(r_n)$. Établissez si cette méthode itérative sera convergente si on prend x_0 dans J.
   $<$**4 pts**$>$

Réponse

1a.

L'étude des variations de la fonction f sur $J = [0, 2]$ montre que la fonction est continue et croissante sur cet intervalle (donc monotone) ($f'(x) = e^{x/4}$ et $f'(x) > 0\ \ \forall\, x \in [0, 2]$). $<$**1 pt**$>$

De plus, on a $f(0) = -2 < 0$ et $f(2) = 4e^{1/2} - 6 \approx 0.59488 > 0$ (donc $f(0)f(2) < 0$), il existe donc une racine r unique dans cet intervalle. $<$**2 pts**$>$

Analytiquement, on trouve facilement $r = 4 \ln \frac{3}{2}$. $<$**2 pts**$>$

1b.

On va tout d'abord calculer les dérivées $\Upsilon_1'(x)$, $\Upsilon_2'(x)$ et $\Upsilon_3'(x)$ pour vérifier quelles sont la valeur de leur dérivé au point correspondant à la racine (trouvée analytiquement i.e., en $x = r = 4 \ln \frac{3}{2}$). si $|\Upsilon_i'(x = 4 \ln \frac{3}{2})| > 1$ alors cela nous dira tout de suite que la méthode diverge (puisqu'on doit trouver un intervalle J dans lequel r et x_0 sont contenu et pour lequel $\forall x \in J, \quad |\Upsilon_i'(x)| < 1$). On trouve,

1. $\Upsilon_1'(x) = 1 + e^{x/4}$ <**1 pt**>
   et on a $\Upsilon_1'(r) = 1 + e^{\ln \frac{3}{2}} = 1 + \frac{3}{2} = \frac{5}{2} > 1$ donc la suite diverge.
   <**3 pts**>
2. $\Upsilon_2'(x) = 1 - \frac{3}{2}\, e^{-x/4}$ <**1 pt**>
   et on a $\Upsilon_2'(r) = 1 - \frac{3}{2}\, e^{-\ln \frac{3}{2}} = 1 - \frac{12}{12} = 0 < 1$. La suite pourrait converger ...vérifions :
   Sur $J = [0, 2]$, Υ_2' est croissante et $\forall x \in J, \Upsilon_2'(x) \in [-1/2, 1 - \frac{3}{2} e^{-1/2} \approx 0.09]$ et donc $\forall x \in J\ |\Upsilon_2'(x)| < 1$ donc la suite converge.
   <**3 pts**>
3. $\Upsilon_3'(x) = 1 - \frac{1}{64}\, e^{x/4}$ <**1 pt**>
   et on a $\Upsilon_3'(r) = 1 - \frac{1}{64} e^{\ln \frac{3}{2}} = 1 + \frac{3}{2} = \frac{125}{128} < 1$ donc la suite pourrait converger ...vérifions :
   Sur $J = [0, 2]$, Υ_3' est décroissante et $\forall x \in J, \Upsilon_3'(x) \in [1 - \frac{1}{64}, 1 - \frac{e^{1/2}}{64} \approx 0.9742]$ et donc $\forall x \in J$ $|\Upsilon_3'(x)| < 1$ donc la suite converge.
   <**3 pts**>

La méthode du point fixe associé à Υ_2 converge quadratiquement <**2 pts**> ($\Upsilon_2'(r) = 0$ et on vérifie que $\Upsilon_2''(r) \neq 0$), la méthode du point fixe associé à Υ_3 converge linéairement.

1c.

En utilisant le théorème de la valeur moyenne, on obtient, puisque $\Upsilon_3(r) = r$ et $r_{n+1} = \Upsilon_3(r_n)$ avec r_n la valeur approchée de la racine à la n-ième itération

$$\begin{aligned} (r_n - r) &= \frac{\Upsilon_3(r_{n-1}) - \Upsilon_3(r)}{(r_{n-1} - r)} \times (r_{n-1} - r) \\ &= \Upsilon_3'(\zeta) \times (r_{n-1} - r), \qquad \text{avec} \quad \zeta \in J. \qquad < \mathbf{2\ pts} > \end{aligned}$$

En utilisant l'inégalité $|\Upsilon_3'(x)| <\approx 0.9742$ <**1 pt**>, (i.e., la borne la plus pessimiste trouvée à la question 1b), on obtient les inégalités suivantes,

$$\begin{aligned} |r_n - r| \le (0.9742)\, |r_{n-1} - r| &\le (0.9742)^2\, |r_{n-2} - r| \\ &\le \ldots \\ &\le (0.9742)^n\, |r_0 - r| \\ &\le (0.9742)^n \times 2. \quad < \mathbf{3\ pts} > \end{aligned}$$

On obtiendra obligatoirement donc $|r_n - r| < 10^{-1}$, dès que,

$$\begin{aligned} (0.9742)^n \times 2 &< 10^{-1} \\ n \ln(0.9742) &< \ln(10^{-1}) - \ln 2. \end{aligned}$$

Soit pour tout $n \geqslant 148$ (ce qui montre que la convergence va être très lente ...) <**2 pts**>

On aurait put prévoir cette lenteur de convergence en voyant que $|\Upsilon_3'(x)| < 1$ mais tout juste, à cause de sa borne de droite très proche de 1 (donc la fonction n'est que tout juste contractante). <**2 pts**>

2a.

On retrouve $\Upsilon_2(x)$ et tout ce qu'on nous demande a déjà été démontré à la question 1.b. <**4 pts**>

IV. Factorisation LU (23 pts)

1. Décomposer la matrice A en produit LU (sans permutation) par la méthode assurant des uns ("1") sur la diagonale de la matrice U.
   <8 pts>

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 4 & -6 & 2 \\ -8 & 2 & 1 \end{pmatrix}.$$

2. Calculer le déterminant de A.
   <3 pts>
3. Soit a un paramètre réel. Utiliser la décomposition LU de la question précédente pour résoudre le système :
   <12 pts>

$$\begin{pmatrix} 4 & 2 & 1 \\ 4 & -6 & 2 \\ -8 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 12a+2 \\ 20a-6 \\ 2 \end{pmatrix}$$

Réponse

$\boxed{\textbf{1.}}$

On parle donc de la factorisation directe et on trouve,

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 4 & -6 & 2 \\ -8 & 2 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 4 & 0 & 0 \\ 4 & -8 & 0 \\ -8 & 6 & (15/4) \end{pmatrix}}_{L<\textbf{4 pts}>} \underbrace{\begin{pmatrix} 1 & (1/2) & (1/4) \\ 0 & 1 & (-1/8) \\ 0 & 0 & 1 \end{pmatrix}}_{U<\textbf{4 pts}>}.$$

$\boxed{\textbf{2.}}$

$$\begin{aligned} \det(\mathrm{A}) &= \det(\mathrm{L}) \times \det(\mathrm{U}) \\ &= 4 \times (-8 \times \frac{15}{4}) \\ &= -120 \quad < \textbf{3 pts} > \end{aligned}$$

$\boxed{\textbf{3.}}$

On doit d'abord résoudre le système $Ly = b$, i.e.,

$$\begin{pmatrix} 4 & 0 & 0 \\ 4 & -8 & 0 \\ -8 & 6 & (15/4) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 12a+2 \\ 20a-6 \\ 2 \end{pmatrix}$$

Par substitution avant, on trouve $y_1 = 3a + (1/2)$, $y_2 = -a + 1$, $y_3 = 8a$.

Puis on doit ensuite résoudre le système $Ux = y$, i.e.,

$$\begin{pmatrix} 1 & (1/2) & (1/4) \\ 0 & 1 & (-1/8) \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3a+1/2 \\ -a+1 \\ 8a \end{pmatrix}$$

qui, par substitution arrière permet de trouver $x_3 = 8a$, $x_2 = 1$ et enfin $x_1 = a$.

<12 pts>

V. Interpolation de Newton (15 pts)

On considère la fonction suivante,

$$f(x) = y = \frac{1}{4}\sin(2\pi x)$$

1. Appliquer la formule de Newton pour trouver un polynôme P de degré deux qui interpole la fonction f aux points $x_0 = 0$, $x_1 = \frac{1}{12}$, $x_2 = \frac{1}{4}$.
   <**8 pts**>
2. Montrer (analytiquement) que $\pi = \frac{1}{8I}$ avec $I = \int_0^{1/4} f(x)\,dx$.
   <**3 pts**>
3. En utilisant le polynôme de degré deux qui interpole la fonction f trouvé à la question 1., en déduire une approximation de la valeur de π (et essayer autant que possible de mettre cette approximation sous la forme d'une fraction). <**4 pts**>

Réponse

$\boxed{\textbf{1.}}$

On a donc les points suivants,

x_k	0	1/12	1/4
y_k	0	1/8	1/4

<**2 pts**>

Le tableau des différences divisées s'écrit,

x	y	Δy	$\Delta^2 y$
0	0		
		3/2	
1/12	1/8		-3
		3/4	
1/4	1/4		

<**4 pts**>

On obtient le polynôme suivant,

$$P_2(x) = \frac{3}{2}x - 3x(x - \frac{1}{12})$$

$$= -3x^2 + \frac{7}{4}x \quad < \textbf{2 pts} >$$

$\boxed{\textbf{2.}}$

$$I = \frac{1}{4}\int_0^{1/4} \sin(2\pi x)\,dx = -\frac{1}{4}\frac{1}{2\pi}\left[\cos(2\pi x)\right]_0^{\frac{1}{4}} = -\frac{1}{8\pi}(-1) = \frac{1}{8\pi}$$

On a donc bien $\quad \pi = \frac{1}{8I} \quad$ <**3 pts**>

$\boxed{\textbf{3.}}$

En utilisant l'approximation polynomiale de la fonction f trouvé à la question 1. , on peut donc donner une approximation rationnelle de la valeur de π en écrivant $\pi = \frac{1}{8I}$ avec

$$I \approx \int_0^{1/4} (-3x^2 + \frac{7}{4}x)\,dx = \left[-x^3 + \frac{7}{8}x^2\right]_0^{\frac{1}{4}} = \frac{5}{128}$$

Donc $\pi = \frac{1}{8I} \approx \frac{128}{8\times 5} = \frac{16}{5}$. <**4 pts**>

Nota : Ce genre de technique, entre autre, (mais avec des polynomes plus élevées) permettent donc de donner des approximation rationnelles de la valeur de π de plus en plus précises.