DIRO
IFT 2425

EXAMEN FINAL

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 : *23/04/2015*

I	Moindres Carrés (31 pts)
II	Intégration Numérique (44 pts)
III	Interpolation & Dérivation Numérique (13 pts)
IV	Résolution d'Équation Différentielle (17 pts)
V	Système d'Équations Différentielles (13 pts)
Total	118 points

Tous documents personnels, calculatrices et calculateurs autorisés

I. Moindres Carrés (31 pts)

Soit un ensemble de N points expérimentaux $\{(x_i, y_i, z_i\}_{i=1}^{N}$ (points de données observés) dans le plan XYZ (espace de dimension 3) donnés dans le tableau 1. Le but de cet exercice est de chercher à estimer, au sens du critère des moindres carrés, les paramètres a, b et c pour que le plan d'équation $z = a + b\,x + c\,y$ s'ajuste le mieux possible aux points observés (appelé aussi *plan de meilleure approximation).*

Table 1 :

i	0	1	2	3
x_i	0	1	2	4
y_i	0	2	0	2
z_i	2	4	1	2

1. **Régression Linéaire**
   (a) Donner le résidu et ensuite trouver le système d'équations linéaires (sous forme matricielle) qui permettra d'estimer les paramètres a, b et c de ce plan de meilleure approximation au sens des moindres carrés.
   <**10 pts**>
   (b) Trouver les coefficients a, b c en résolvant ce système d'équations par la méthode de Gauss (sans pivotage) et en utilisant le Tableau 1 ($N = 4$).
   <**8 pts**>
2. **Système Surdéterminé**
   (a) On va maintenant tenter de résoudre ce problème d'ajustement de plan au sens des moindres carrés en identifiant la solution de notre problème (i.e., le vecteur solution $x = (a\ \ b\ \ c)^t$) comme étant la solution d'un système sur-dimensionné d'équations linéaires $\mathbf{A}\,x = \mathbf{b}$ dont les coefficients (de $\mathbf{A}$ et $\mathbf{b}$) sont entachés de bruit. Dans ce but, trouver la matrice $\mathbf{A}$ et $\mathbf{b}$ en fonction des $\{(x_i, y_i, z_i\}_{i=1}^{N}$.
   <**5 pts**>
   (b) Expliquer comment on peut retrouver les équations de la question 1(a) par cette méthode.
   <**8 pts**>

Réponse

1(a)

Le résidu que l'on cherchera à minimiser par cette technique d'estimation au sens des moindres carrés est :

$$R(a,b,c) = \sum_{i=1}^{N} \Big(z_i - (a + b\,x_i + c\,y_i) \Big)^2$$

<**3 pts**>

Trouver les coefficients a, b et c au sens des moindres carrés, c'est trouver a, b et c pour que $R(a,b,c)$ soit minimal, c'est-à-dire tel que $\frac{\partial R(a,b,c)}{\partial a} = 0$, $\frac{\partial R(a,b,c)}{\partial b} = 0$ et $\frac{\partial R(a,b,C)}{\partial c} = 0$. La dérivation par rapport à ces paramètres nous donne immédiatement :

$$\begin{aligned}
\frac{\partial R(a,b,c)}{\partial a} &= -2 \left(\sum_{i=1}^{N} [z_i - (a + b\,x_i + c\,y_i)] \right) = 0 \\
\frac{\partial R(a,b,c)}{\partial b} &= -2 \left(\sum_{i=1}^{N} [z_i - (a + b\,x_i + c\,y_i)]\,x_i \right) = 0 \\
\frac{\partial R(a,b,c)}{\partial c} &= -2 \left(\sum_{i=1}^{N} [z_i - (a + b\,x_i + c\,y_i)]\,y_i \right) = 0
\end{aligned}$$

qui conduise au système d'équations linéaires suivantes, pour les paramètres a, b et c :

$$\begin{cases} \sum_{i=1}^{N}(a + b\,x_i + c\,y_i) & = & \sum_{i=1}^{N} z_i \\ \sum_{i=1}^{N}(a\,x_i + b\,x_i^2 + c\,y_i\,x_i) & = & \sum_{i=1}^{N} z_i\,x_i \\ \sum_{i=1}^{N}(a\,y_i + b\,x_i\,y_i + c\,y_i\,x_i) & = & \sum_{i=1}^{N} z_i\,y_i \end{cases}$$

que l'on peut mettre sous forme matricielle :

$$\begin{pmatrix} N & \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i\,y_i \\ \sum_{i=1}^{N} y_i & \sum_{i=1}^{N} x_i\,y_i & \sum_{i=1}^{N} y_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} z_i \\ \sum_{i=1}^{N} z_i\,x_i \\ \sum_{i=1}^{N} z_i\,y_i \end{pmatrix}$$

<7 pts>

1(b)

Dans notre cas, on obtient le système d'équations linéaires suivantes :

$$\begin{pmatrix} 4 & 7 & 4 \\ 7 & 21 & 10 \\ 4 & 10 & 8 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 9 \\ 14 \\ 12 \end{pmatrix}$$

que l'on nous demande de résoudre par la méthode du pivot de Gauss sans pivotage ; c'est-à-dire en considérant d'abord les deux opérations élémentaires ; $R_2 = R_2 - \frac{7}{4}R_1$, $R_3 = R_3 - R_1$, et on obtient :

$$\begin{pmatrix} 4 & 7 & 4 \\ 0 & 35/4 & 3 \\ 0 & 3 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 9 \\ -7/4 \\ 3 \end{pmatrix}$$

On utilise une dernière opération élémentaire : $R_3 = R_3 - 4 \times \frac{3}{35}R_2$, i.e., $R_3 = R_3 - \frac{12}{35}R_2$, et on obtient :

$$\begin{pmatrix} 4 & 7 & 4 \\ 0 & 35/4 & 3 \\ 0 & 0 & 104/35 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 9 \\ -7/4 \\ 504/140 \end{pmatrix}$$

Ce qui nous permet d'obtenir par substitution arrière, $\boxed{c \approx 1.21}$, puis, $\boxed{b \approx -0.62}$, et enfin $\boxed{a \approx 2,12}$.

<8 pts>

2(a)

En disant que chacun des $N = 4$ points de données $\{(x_i, y_i, z_i)\}_{i=1}^{N}$ sont bien sur le plan d'équation $z = a + b\,x + c\,y$, à un bruit prêt entachant les données, on a facilement :

$$\underbrace{\begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} a \\ b \\ c \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}}_{b} \qquad \textit{ou} \qquad \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 0 \\ 1 & 4 & 2 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} a \\ b \\ c \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 2 \\ 4 \\ 1 \\ 2 \end{pmatrix}}_{b}$$

<5 pts>

On résout ce système sur-dimensionné par la méthode de la pseudo-inverse vue en cours et on obtient :

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \underbrace{(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t}_{\mathbf{B}} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$

la matrice $\mathbf{B}$ sera une matrice 3 qui sera la même que celle de la question 1(b) et on retrouvera donc les équations de cette question par la méthode la pseudo-inverse en multipliant par $\mathbf{A}^t$ de part et d'autre de cette équation :

$$\mathbf{A}^t\mathbf{A} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{A}^t \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$

<8 pts>

II. Intégration Numérique (44 pts)

Soit l'intégrale suivante :

$$I = \int_0^1 f(x)\,dx = \int_0^1 x\,e^{-x}\;dx \tag{1}$$

1. **Méthode de Romberg**

   (a) Donner une approximation du calcul de I (Eq. (1)) en appliquant la méthode d'extrapolation de Richarson combinée avec la quadrature composite du Trapèze (Méthode de Romberg) autant de fois que possible à partir des estimations obtenues avec $h = 1/2$, $h = 1/4$ et $h = 1/8$ et indiquer l'ordre du terme d'erreur obtenu par cette estimation.
   <14 pts>

   (b) Estimer le nombre d'intervalles nécessaires pour arriver, avec la méthode des quadratures composites du Trapèze, à la même précision que celle obtenue par la méthode précédente. Utiliser dans un premier temps l'ordre du terme d'erreur de ces deux méthodes et dans un deuxième temps la formule analytique de la borne supérieure d'erreur pour la méthode de la quadrature composite du Trapèze et une estimation réaliste de l'erreur pour la méthode de Romberg.
   <8 pts>

   (c) Soient $I_{\mathbf{1},1/8}$, $I_{\mathbf{1},1/4}$, $I_{\mathbf{1},1/2}$, étant respectivement les estimations obtenues par la méthode de la quadrature composite du Trapèze de la question 1.(a) pour estimer I avec $h = 1/2$, $h = 1/4$ et $h = 1/8$.
   Comment peut-on, en exploitant seulement la connaissance de ces trois estimations (i.e., $I_{\mathbf{1},1/8}$, $I_{\mathbf{1},1/4}$ et $I_{\mathbf{1},1/2}$) savoir (sans ambiguité) que la méthode numérique utilisée pour générer ces trois estimations est une méthode dont le terme d'erreur est $O(h^2)$?
   <5 pts>

   (d) Donner le coût calculatoire (en terme de nombre d'évaluations de fonction) que possède cette méthode de Romberg et l'ordre du terme d'erreur le plus bas possible (qu'il est possible théoriquement d'atteindre) correspondant en fonction de n, le nombre d'itérations ou d'étapes de la méthode de Rombeerg lorsque 1] on démarre avec une méthode de quadrature simple du Trapèze et 2] lorqu'on démarre avec une méthode de quadrature composite du Trapèze (comme c'est le cas dans cet exemple avec l'estimation de $I_{\mathbf{1},1/2}$).
   <5 pts>

   (e) Donner l'ordre du terme d'erreur le plus bas possible (qu'il est possible théoriquement d'atteindre) avec l'algorithme de Romberg lorsqu'on dispose de N données (on suposera que N est une puissance de 2). Donner cette ordre pour un tableau de 256 données discrétisées et expliquer clairement les raisons numériques qui expliquent pourquoi, pratiquement, on ne peut atteindre cette précision.
   <5 pts>

2. **Méthode de la Quadrature Gaussienne**
   Utiliser la méthode de la Quadrature Gaussienne (2 termes) pour estimer une valeur approchée de I.
   <7 pts>

Réponse

$\boxed{\textbf{1.(a)}}$

On va d'abord donner des estimation de $f(x)$ sur $x \in [0\,1]$ lorsque $h = 1/8$ (8 intervalles ou 9 points) :

x	0	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
$f(x)$	0	0.110312	0.194700	0.257733	0.303265	0.334538	0.354274	0.364754	0.367879

On obtient, par la méthode des trapèzes :

$$I_{\mathbf{1},1/8} = (h/2)\Big(f(0) + 2f(0.125) + 2f(0.25) + 2f(0.375) + 2f(0.50) + 2f(0.625) + 2f(0.75) + 2f(0.875) + f(1)\Big)$$

$$\begin{aligned}
I_{\mathbf{1},1/8} &\approx (0.125/2)\Big(0 + 2{\times}0.110312 + 2{\times}0.194700 + \ldots\ldots\ldots\ldots + 2{\times}0.364754 + 0.367879\Big) \\
I_{\mathbf{1},1/8} &\approx 0.262939 + O(h^2)
\end{aligned}$$

$$\begin{aligned}
I_{\mathbf{1},1/4} &= (h/2)\Big(f(0.00) + 2f(0.25) + 2f(0.50) + 2f(0.75) + f(1.00)\Big) \\
I_{\mathbf{1},1/4} &\approx (0.25/2)\Big(0 + 2{\times}0.194700 + 2{\times}0.303265 + 2{\times}0.354274 + 0.367879\Big) \\
I_{\mathbf{1},1/4} &\approx 0.259045 + O(h^2)
\end{aligned}$$

$$\begin{aligned}
I_{\mathbf{1},1/2} &= (h/2)\Big(f(0.00) + 2f(0.50) + f(1.00)\Big) \\
I_{\mathbf{1},1/2} &\approx (0.5/2)\Big(0 + 2{\times}0.303265 + 0.367879\Big) \\
I_{\mathbf{1},1/2} &\approx 0.243602 + O(h^2)
\end{aligned}$$

<8 pts>

En utilisant maintenant l'extrapolation de Richardson pour combiner deux à deux les formules numériques obtenues, on obtient puisque les estimations précédentes sont en $O(h^2)$:

$$\begin{aligned}
I_{\mathbf{2},1/8} &= I_{\mathbf{1},1/8} + \frac{1}{2^2-1}\big(I_{\mathbf{1},1/8} - I_{\mathbf{1},1/4}\big) \\
I_{\mathbf{2},1/8} &\approx 0.262939 + \frac{1}{3}(0.262939 - 0.259045) \\
I_{\mathbf{2},1/8} &\approx 0.264237 + O(h^4)
\end{aligned}$$

$$\begin{aligned}
I_{\mathbf{2},1/4} &= I_{\mathbf{1},1/4} + \frac{1}{2^2-1}\big(I_{\mathbf{1},1/4} - I_{\mathbf{1},1/2}\big) \\
I_{\mathbf{2},1/4} &\approx 0.259045 + \frac{1}{3}(0.259045 - 0.243602) \\
I_{\mathbf{2},1/4} &\approx 0.264192 + O(h^4)
\end{aligned}$$

<4 pts>

En utilisant encore l'extrapolation de Richardson pour combiner ces deux dernières formules numériques, on obtient finalement puisque les estimations précédentes sont en $O(h^4)$:

$$\begin{aligned}
I_{\mathbf{3},1/8} &= I_{\mathbf{2},1/8} + \frac{1}{2^4-1}\big(I_{\mathbf{2},1/8} - I_{\mathbf{2},1/8}\big) \\
I_{\mathbf{3},1/8} &\approx 0.264237 + \frac{1}{15}(0.264237 - 0.264192) \\
I_{\mathbf{3},1/8} &\approx 0.26424 + O(h^6)
\end{aligned}$$

<**2 pts**>

$\boxed{\textbf{1.(b)}}$

La méthode de la quadrature composite du Trapèze est en $O(h^2)$ mais la méthode de Romberg précédante, incluant deux itérations, est de l'ordre de $O(h^6)$ (avec $h = 1/8$ dans les deux cas). L'erreur de la méthode de Romberg est grossièrement de $\approx (1/8)^6 \approx 3.8 \times 10^{-6}$. (Grossièrement) il faudrait un $h \approx \sqrt{3.8 \times 10^{-6}} \approx 1.95 \times 10^{-3}$ pour arriver, avec la même précision que la méthode de la quadrature composite, soit ≈ 512 intervalles, i.e., 513 points.

Si on utilise la formule analytique de la borne supérieure d'erreur de la méthode de la quadrature composite du Trapèze, on arrivera à une meilleure estimation. Dans un premier temps, estimons l'erreur faîte sur I. On a $I_{\mathbf{3},1/8} \approx 0.264240$ et $I_{\mathbf{2},1/8} \approx 0.264237$. La différence entre ces deux estimations est de $\approx 3 \times 10^{-6}$; ce qui est cohérent avec la précision grossière trouvée précédemment et semble être une bonne estimation réaliste de l'erreur faîte par la méthode de Romberg.

Pour la méthode de la quadrature composite du Trapèze, cette formule d'erreur s'écrit analytiquement (avec $a = 0$ et $b = 1$, les bornes de l'intégrale) :

$$|E| = \frac{h^2}{12}(b-a)|f''(\xi)| = \frac{h^2}{12}|f''(\xi)| \quad \text{avec} \quad \xi \in [0\ 1].$$

Bornons cette estimation par 3×10^{-6} :

$$\begin{aligned}
\frac{h^2}{12} \max_{\xi \in [0\ 1]} |f''(\xi)| &< 3 \times 10^{-6} \\
h^2 &< \frac{12 \times 3 \times 10^{-6}}{\max_{\xi \in [0\ 1]} |f''(\xi)|} \\
h &< \sqrt{\frac{36 \times 10^{-6}}{\max_{\xi \in [0\ 1]} |f''(\xi)|}}
\end{aligned}$$

Reste à calculer $\max_{\xi \in [0\ 1]} |f''(\xi)|$:

On a $f(x) = x\,e^{-x}$, $f'(x) = e^{-x} - x\,e^{-x}$, $f''(x) = -2e^{-x} + x\,e^{-x}$ et $f'''(x) = 3e^{-x} - x\,e^{-x} > 2e^{-x} > 0$ (car $x \in [0\ 1]$), donc $f''(x)$ est croissante (et aussi négative sur $x \in [0\ 1]$) et donc bornée comme suit : $|f''(x)| < |f''(x = 0)| = 2$ et donc : $h < \sqrt{18 \times 10^{-6}} = 4.24 \times 10^{-3}$, i.e., $h \approx 235.8$, i.e. un nombre de points $n > 236$.

<**8 pts**>

$\boxed{\textbf{1.(c)}}$

Grossièrement $|I_{\mathbf{1},1/8} - I_{\mathbf{1},1/4}|$ représente une estimation de l'erreur faîte sur I. De même $|I_{\mathbf{1},1/4} - I_{\mathbf{1},1/2}|$ représente aussi une estimation de l'erreur faîte sur I mais pour un pas h deux fois plus petit.
Ainsi le rapport :

$$\frac{|I_{\mathbf{1},1/4} - I_{\mathbf{1},1/2}|}{|I_{\mathbf{1},1/8} - I_{\mathbf{1},1/4}|} = \frac{|0.259045 - 0.243602|}{|0.262939 - 0.259045|} \approx 3.97$$

$3.97 \approx 4$ represente une estimation du gain en précision lorsqu'on divise le pas h par deux. Ceci est typique d'une méthode numérique d'estimation avec un ordre d'erreur $O(h^2)$, ce qui est bien le cas ici.
<**5 pts**>

Nota -1- : On aurait pu raisonner de la façon suivante aussi : (qui conduirait aussi à $n = 2$). On sait que :

$$
\begin{aligned}
I &= I_{\mathbf{1},1/2} + O([h=1/2]^n) \\
I &= I_{\mathbf{1},1/4} + O([h=1/4]^n) \\
I &= I_{\mathbf{1},1/8} + O([h=1/8]^n)
\end{aligned}
$$

avec I, la vraie valeur de l'intégrale. La résolution de ce système de 3 équations à 3 inconnus conduirait aussi à $n \approx 2$.

Nota -2- : On peut se persuader que la méthode permet bien de trouver l'estimation de cet ordre de l'erreur en considérant un autre petit exercice :
Considérons cette fois $I_{\mathbf{1},1} \approx 0.1839395$ et donc pour la deuxième itération de l'extrapolation de l'extrapolation de Richardson : $I_{\mathbf{2},1/4} \approx 0.2634895$.
Ainsi $|0.264237 - 0.2634895|/|0.264192 - 0.264237| \approx 16.6 \approx 16$, qui est typique du gain en précision obtenu lorsqu'on divise le pas h par deux pour un modèle numérique d'une méthode avec un terme d'erreur en $O(h^4)$ comme c'est le cas pour la deuxième étape de l'extrapolation de Richardson sur la méthode des Trapèzes.

1.(d)

Dans notre exemple (question 1.(a) ou cas 2.), on a deux itérations de l'extrapolation de Richardson qui nécessitait 8 intervalles ou 9 points. Si on s'était arrêté à une itération de l'extrapolation de Richardson, on aurait eu besoin de seulement 4 intervalles ou 5 points ;

C'est-à-dire, une complexité calculatoire de $O(2^{n+1}+1)$ pour une précision avec un terme d'erreur de $O(h^{2n+2})$; et donc bien $O(h^6)$ dans notre cas.

Dans le cas 1. ; si on avait commencé par une quadrature simple du Trapèze, c'est-à-dire 1 intervalle (ou 2 points), on aurait eu une complexité calculatoire de $O(2^{n+1})$ pour une précision avec un terme d'erreur de $O(h^{2n+2})$.

<5 pts>

1.(e)

Si N est une puissance de 2, alors $1 + \log_2 N$ représente le nombre d'itérations possible maximal pour l'algorithme d'extrapolation de Richardson (ou Romberg) si on part à l'itération 0 avec une quadrature simple du Trapèze (et $\log_2 N$ si on part à l'itération 0 avec la plus petite quadrature composite du Trapèze) et donc la précision associée à cette méthode à un terme d'erreur de l'ordre de $O(h^{2(1+\log_2 N)+2})$.

Pour un tableau de 256 données discrétisées, on a une borne théorique de précision de $O(h^{2(1+\log_2 256)+2}) = O(h^{20})$!

On arrivera jamais à une telle précision à cause des erreurs numériques dans les évaluations de fonctions mais aussi plus précisement à cause de la relation utilisée dans la méthode de Romberg qui combine, à chaque itération, la méthode la plus précise (celle associée au pas h) avec celle qui est la moins précise (pas $2h$) :

$$
\begin{aligned}
&\text{Extrapolation de Richardson} = \\
&\text{méthode}_{\text{pas h}} + \frac{1}{2^{[\text{ordre de la méthode}]}-1}\Big(\text{méthode}_{\text{pas h}} - \text{méthode}_{\text{pas 2h}}\Big)
\end{aligned}
$$

et qui fait intervenir la soustraction de deux nombres approximées ($\text{méthode}_{\text{pas h}} - \text{méthode}_{\text{pas 2h}}$) d'autant plus quasi-égales que l'on avance dans les itérations et qui ainsi annule toute la précision (théorique) gagnée par une itération supplémentaire de l'algorithme de Romberg !

<5 pts>

2.

En faisant le changement de variable $x = \frac{1}{2}(t+1)$ ou $t = 2x - 1$, $(dx = (1/2)dt)$,

<1 pt>

nous obtenons :

$$
\begin{aligned}
I &= \int_0^1 x\, e^{-x}\, dx \\
&= \frac{1}{4}\int_{-1}^{1} \Big((t+1)\, e^{-1/2(t+1)}\Big)\, dt \qquad \textbf{< 2 pts >} \\
&= \frac{1}{4}\Big(([\sqrt{3}/3]+1)\, e^{-1/2([\sqrt{3}/3]+1)} + ([-\sqrt{3}/3]+1)\, e^{-1/2([-\sqrt{3}/3]+1)}\Big) \textbf{< 2 pts >} \\
&\approx 0.265 \qquad \textbf{< 2 pts >}
\end{aligned}
$$

III. Interpolation & Dérivation Numérique (13 **pts**)

Soit des valeurs discrétisées d'une fonction f espacée de $h = 0.25$:

x	0	0.25	0.50	0.75	1.0
$f(x)$	0	0.195	0.303	0.354	0.368

1. Trouver une estimation numérique de la dérivée de f en $x = 0.50$ avec une précision en $O(h^2)$ et le moins d'opérations possibles.
<4 pts>

2. On considère maintenant la méthode suivante : on utilise n points autours de $x = 0.5$ pour trouver le polynome de collocation d'ordre $n-1$ par la méthode de Lagrange que l'on dérive pour avoir une estimation de $f'(0.5)$, avec la même précision que la méthode précédente. Utiliser cette stratégie avec le bon n et les bon points de données pour trouver la même estimation que celle trouvée à la question précédente.
<5 pts>

3. En utilisant une interpolation par Lagrange du polynôme de collocation d'ordre 2, trouvez une stratégie pour obtenir une bonne approximation de $f'(0.375)$ en $O(h^2)$ et faîte le.
<4 pts>

Réponse

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

<4 pts> On utilise donc la formule centrée et on a : $f'(0.5) \approx \dfrac{0.354 - 0.195}{0.75 - 0.25} = 0.318$

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

La formule centrée précédente a été trouvé en fait en dérivant le polynôme de collocation d'ordre deux, donc en utilisant trois points centrés autour de l'estimation :

$$
\begin{aligned}
P_2(x) &= \frac{(x-0.50)(x-0.75)}{(0.25-0.50)(0.25-0.75)}\,0.195 + \frac{(x-0.25)(x-0.75)}{(0.50-0.25)(0.50-0.75)}\,0.303 + \frac{(x-0.25)(x-0.50)}{(0.75-0.25)(0.75-0.50)}\,0.354 \\
P_2(x) &= 1.56(x-0.50)(x-0.75) - 4.848(x-0.25)(x-0.75) + 2.832(x-0.25)(x-0.50) \\
P_2(x) &= -0.456\, x^2 + 0.774\, x + Cst
\end{aligned}
$$

<5 pts> puis $P_2'(x = 0.5) = -0.912 \times 0.5 + 0.774 = 0.318$

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

$$P_2'(x = 0.375) = -0.912 \times 0.375 + 0.774 = 0.432$$

<5 pts>

Nota : la réponse de la question précédente peut aussi se re-trouver en utilisant la formule centrée sur ce point d'abscisse $x = 0.375$ qui n'existe pas dans le tableau de donnés mais pour lequel on peut utiliser la formule centrée pour obtenir une (et en fait la même) estimation de $f'(x = 0.375)$. En effet ;

$$f'(x = 0.375) = \frac{0.303 - 0.195}{2 \times 0.125} = 0.432$$

IV. Résolution d'Équation Différentielle (17 pts)

1. Soit l'équation différentielle suivante :

$$y'(t) = f(t, y(t)) = t\,\sin(y(t)) \tag{2}$$

avec la condition initiale $y_0 = y(t = 0) = 2$.

(a) Résoudre cette équation différentielle avec la méthode d'Euler et avec un pas $h = 0.2$. Calculer seulement $y_1 = y(0.2)$ et $y_2 = y(0.4)$ (donner les résultats en arrondissant et en gardant au moins trois décimales).
<4 pts>

(b) Résoudre cette équation différentielle avec la méthode d'Euler modifiée et toujours avec un pas $h = 0.2$. Calculer seulement $y(0.2)$ et $y(0.4)$, (arrondisser et garder au moins quatre décimales).
<5 pts>

2. Donner les différences (avantages/inconvénients) fondamentales qui existent entre la méthode de Runge-Kutta et la méthode d'Adams-Moulton du même ordre.
<4 pts>

3. Quelles sont les raisons pour lesquelles les méthodes de Runge-Kutta d'ordre supérieur à 6 ne sont pas utilisées et que l'on se limite à une méthode de RK d'ordre 5 ou 6 généralement ?
<4 pts>

Réponse

1(a)

Par la méthode d'Euler, la relation itérative à considérer est donc, avec $h = 0.2$ et $y_0 = 2$

$$y_{n+1} = y_n + h\,\big(t_n \sin(y_n)\big)$$

<1 pt>

Les valeurs de t_n et y_n sont données dans le tableau ci-dessous

n	t_n	y_n
0	0.0	2.000
1	0.2	2.000
2	0.4	2.036

<3 pts>

1(b)

Par la méthode d'Euler modifiée, on a les relations itératives suivantes :

$$\tilde{y}_{n+1} = y_n + h\,\underbrace{\big(t_n \sin(y_n)\big)}_{f(t_n, y_n)} \quad \text{et} \quad y_{n+1} = y_n + \frac{h}{2}\big(f(t_n, y_n) + f(t_{n+1}, \tilde{y}_{n+1})\big) = y_n + \frac{h}{2}\big(t_n \sin(y_n) + t_{n+1} \sin(\tilde{y}_{n+1})\big)$$

<1 pt>

Les valeurs de t_n, y_n et $\tilde{y}_n$ sont reportées dans le tableau ci-dessous :

n	t_n	$\tilde{y}_n$	y_n
0	0.0	-	2.000
1	0.2	2.000	2.018
2	0.4	2.054	2.071

<4 pts>

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

La méthode de Runge-Kutta ne permet pas d'estimer précisement l'erreur (différence entre la solution analytique et la solution estimée numériquement) que fait la méthode en cours d'itérations ; par contre cette méthode permet, en cours d'itérations, de changer le pas et eventuellement de le diviser par deux si l'erreur entre la solution analytique et la soution estimée numériquement est trop grande.
C'est l'inverse pour la méthode d'Adams-Moulton ; on peut assez précisément estimer, en cours d'itérations, l'erreur de la méthode mais on ne peut corriger cette errer en cours de route en changeant le pas (divser par deux) en cours d'itérations. La seule façon est de recommencer la procédure itérative, dès le début, avec un pas moins grand.
<4 pts>

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

Les méthode de RK d'ordre 5 ou 6 sont celles où la précision est la plus intéressante pour un même coût calculatoire ou inversement, celle pour laquelle le coût calculatoire est le plus faible pour une même précision. N'oublions pas aussi que les méthodes de RK d'ordres elevées sont plus sensibles aux erreurs numériques car elle combine beaucoup plus d'opérations et de coefficients.
<4 pts>

V. Système d'Équations Différentielles (13 pts)

Soit l'équation différentielle suivante :

$$x'''(t) + t\,x(t) \quad = \quad 0$$

On considérera les conditions initiales $x(0) = 1$, $x'(0) = 0$ et $x''(0) = 1$. On vous demande de :

1. Ré-écrire ce système en un système d'équations différentielles du premier ordre (pour lequel on a précisé les conditions initiales).
   <4 pts>
2. Si on utilise la méthode d'Euler ordinaire pour résoudre ce système d'équations différentielles, préciser quelle serait le système de relations itératives qui serait obtenu.
   <4 pts>
3. À partir de $x(t = 0) = x_0 = 1.0$, trouver $x(t = 0.2)$ et $x(t = 0.4)$.
   <5 pts>

Réponse

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

En posant, $x'(t) = x_p(t)$ et $x'_p(t) = x_{pp}(t)$, on trouve immédiatement le système d'équations différentielles du premier ordre suivant :

$$\begin{cases} x'(t) &= x_p(t) \\ x'_p(t) &= x_{pp}(t) \\ x'_{pp}(t) &= -t\,x(t) \end{cases}$$

<3 pts>

Avec les conditions initiales $x(0) = 1$, $x_p(0) = 0$ et $x_{pp}(0) = 1$.

<1 pts>

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

On obtient, avec la méthode d'Euler ordinaire, le système de relations itératives suivant :

$$\left\{ \begin{array}{lcl} x(t_{n+1}) & = & x(t_n) + h\, x_p(t_n) \\ x_p(t_{n+1}) & = & x_p(t_n) + h\, x_{pp}(t_n) \\ x_{pp}(t_{n+1}) & = & x_{pp}(t_n) - h\, t_n\, x(t_n) \end{array} \right.$$

<4 pts>

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

En se souvenant que la fonction qui nous interesse (solution de l'équation différentielle) est $x(t)$ et en partant des conditions initiales $x(0) = 1$ et $x_p(0) = 0$ et $x_{pp}(0) = 1$, on obtient immédiatement pour $h = 0.2$:

$$\left\{ \begin{array}{lcl} x(0.2) & = & 1.0 + (0.2 \times 0.0) = 1.0 \\ x_p(0.2) & = & 0.0 + (0.2 \times 1.0) = 0.2 \\ x_{pp}(0.2) & = & 1.0 - 0.2 \times \big(0.0 \times 1\big) = 1 \end{array} \right.$$

<2.5 pts>

$$\left\{ \begin{array}{lcl} x(0.4) & = & 1.0 + (0.2 \times 0.2) = 1.04 \\ x_p(0.4) & = & 0.2 + (0.2 \times 1.0) = 0.4 \\ x_{pp}(0.4) & = & 1.0 - 0.2 \times \big(0.2 \times 1\big) = 0.96 \end{array} \right.$$

<2.5 pts>

Nota : $x_p(0.4)$ et $x_{pp}(0.4)$ n'était pas demandé.