CC du 28 mars 2017, de 15h45 à 17h45.

Documents, calculatrices et ordinateurs ultraportables déconnectés du réseau autorisés.
Ce sujet comporte 2 pages. Barême donné à titre indicatif et non contractuel.
Les exercices 1 et 2 d'une part et l'exercice 3 d'autre part sont à rédiger sur des copies séparées

---

1. APPROXIMATION DE L'EXPONENTIELLE (7 PTS)

On interpole la fonction exponentielle sur $[-1,1]$ en $n+1$ points $x_0,...,x_n$ de $[-1,1]$.

(1) *Donner une majoration de l'erreur d'interpolation en x fonction de x,n et de $x_0,...,x_n$.*
La dérivée $n+1$-ième de l'exponentielle est elle-même, elle est majorée par e sur $[-1,1]$ donc :

$$|e^x - P_n(x)| \leq \frac{e}{(n+1)!}\prod_{j=0}^{n}(x-x_j)$$

(2) *On choisit $x_0,...,x_n$ en les racines du polynôme de Tchebyshev $T_{n+1}(\cos(x)) = \cos((n+1)x)$. Déterminer n pour assurer une erreur absolue inférieure à* `1e-10` *sur $[-1,1]$. Même question pour avoir une erreur relative inférieure à* `1e-10`.
Comme $T_{n+1}(x) = 2^n\prod_{j=0}^{n}(x-x_j)$, on a

$$|e^x - P_n(x)| \leq \frac{e}{(n+1)!2^n}$$

qui est plus petit que `1e-10` à partir de $n=10$. En erreur relative, il faut diviser par e^x, donc il suffit de majorer $e^2/(n+1)!/2^n$, $n=11$ convient.

(3) *Pour cette valeur de n, déterminer un majorant de l'erreur si on prenait des points $x_0,...,x_n$ équidistribués sur $[-1,1]$. Pour les études d'extremum de fonctions, vous pouvez donner le résultat calculé à la machine à condition d'indiquer les commandes utilisées sur la copie.*
Pour $n=10$, on a $x_k = -1+k/5, k=0..10$. On pose
`g:=product(x+1-k/5,k,0,10)`
puis on cherche les extrêmas de g
`l:=solve(g',x)`
la valeur de g en ces points
`L:=subst(g,x,l)`
et le maximum en valeur absolue (ils sont tous entre -1 et 1, et l'erreur est nulle en -1 et 1)
`max(abs(L))*e^2/11!` $\approx$ `1.6e-9`
Autre méthode
Si on note h le pas entre deux x_j successifs ($h=1/5$ pour $n=10$), on peut aussi majorer $\prod_{j=0}^{n}(x-x_j)$ par $h^{n+1}\frac{1}{2}n!$ car on est à distance au plus $h/2$ du plus proche, puis h du deuxième plus proche, puis $2h$ du suivant, etc. Donc l'erreur d'interpolation est majorée par $h^{n+1}\frac{1}{2(n+1)}$, ici $e^2/5^{11}/2/11 =$ `6.9e-9`, majoration un peu moins bonne que la précédente mais nécessitant moins de calculs à la machine.

(4) *Quelle valeur de n faut-il choisir pour obtenir la même précision en approchant l'exponentielle par son développement de Taylor à l'ordre n en $x=0$? Indication : il existe θ entre 0 et x tel que :*

$$f(x) = Taylor_n(x) + e^{\theta}\frac{x^{n+1}}{(n+1)!}$$

Le produit des $x-x_k$ est remplacé par x^{n+1}, il suffira donc que $e/(n+1)!$ soit plus petit que `1e-10`, ce qui se produit dès $n=13$.

(5) *Soit $x \in [1,2]$, on a*

$$e^x = \left(e^{\frac{x}{2}}\right)^2, \quad \frac{x}{2} \in [0,1]$$

on peut donc approcher e^x en prenant le carré de l'approximation de $e^{x/2}$. Que peut-on alors dire de l'erreur ?
L'erreur relative est multipliée par deux au premier ordre.

2. Approximation au sens des moindres carrés (3 pts)

On cherche un polynôme du second degré $P(t) = \alpha t^2 + \beta t + \gamma$ approchant le mieux possible au sens des moindres carrés l'exponentielle aux points d'abscisses $t_1 = -1$, $t_2 = -1/2$, $t_3 = 0$, $t_4 = 1/2$ et $t_5 = 1$.

(1) *On pose $x = (\gamma, \beta, \alpha)$ et b le vecteur de composantes $(e^{t_1}, ..., e^{t_5})$ le problème revient alors à minimiser $\|Ax - b\|_2$. Déterminer la matrice A en fonction des t_i.*
`A:=tran(tran(vandermonde([-1,-1/2,0,1/2,1]))[0..2])`

(2) *Résoudre le problème.*
En calcul exact, on peut faire :
`b:=exp([-1,-1/2,0,1/2,1]); linsolve(trn(A)*A,trn(A)*b)`
En calcul approché, on obtient un résultat plus précis avec la décomposition QR de A
`q,r:=qr(A); linsolve(r[0..2],trn(q)*b)`
On trouve `[0.994415410173,1.14859907711,0.547734459671]`.

(3) *Soit $f(t) = e^t - P(t)$, calculer f' et f'', en déduire le tableau de variations de f' puis donner une majoration de l'erreur $|f(t)|$ sur $[-1,1]$.*
On a $f' = e^t - (2\alpha t + \beta)$ et $f'' = e^t - 2\alpha$ donc $f'' < 0$ si et seulement si $t < \ln(2\alpha) = 0.09....$ Donc f' décroit de 0.31.. à -0.15.. sur $[-1, 0.09...]$ puis croit jusque 0.47.. et s'annule donc 2 fois en -0.49... et 0.58..., en ces points ou en -1 et 1 on atteint les valeurs extrêmes de $f(t)$, le maximum en valeur absolue est atteint en $t = 0.58..$ et vaut `5.9e-2` (arrondi par excès).

---

3. Transformée de Fourier rapide (10 pts) **(à rédiger sur une copie séparée)**

Pour tout entier N, on appelle *transformée de Fourier discrète* l'application de $\mathbb{C}^N$ dans $\mathbb{C}^N$ qui envoie le N-uplet $(x_0, \ldots, x_{N-1})$ sur le N-uplet $(X_0, \ldots, X_{N-1})$ défini par

$$\forall j \in \{0, \ldots, N-1\},\ X_j = \sum_{k=0}^{N-1} x_k e^{-\frac{2i\pi}{N} jk}.$$

De façon équivalente, on a

$$\begin{pmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{N-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & \omega & \omega^2 & \ldots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \ldots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \ldots & \omega^{(N-1)(N-1)} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix}$$

où $\omega = \omega_N = e^{-\frac{2i\pi}{N}}$. On notera $M_N = \left(\omega^{(j-1)(k-1)}\right)_{1 \le j,k \le N}$ la matrice ci-dessus.

(1) *Combien d'opérations (additions et multiplications de nombres complexes) faut-il réaliser pour calculer une transformée de Fourier discrète avec les formules ci-dessus ?*
Si on suppose que M_N a été calculée auparavant, il faut N multiplications et $N-1$ additions par composante de X, donc $O(N^2)$ opérations

(2) *Montrer que $M_N^* M_N = N I_N$ où M_N^* est la transposée de la conjuguée de M_N.*

$$(M_N)_{j,k} = \omega^{(j-1)(k-1)}, \quad (M_N^*)_{j,k} = \overline{\omega}^{(j-1)(k-1)} = \omega^{-(j-1)(k-1)}$$

donc

$$\begin{aligned}(M_N^* M_N)_{j,l} &= \sum_{k=0}^{N-1} (M_N^*)_{j,k}(M_N)_{k,l} \\ &= \sum_{k=0}^{N-1} \omega^{-(j-1)(k-1)} \omega^{(k-1)(l-1)} \\ &= \sum_{k=0}^{N-1} \omega^{(l-j)(k-1)} \\ &= \sum_{k=0}^{N-1} \left(\omega^{(l-j)}\right)^{k-1} \\ &= \frac{1-\omega^{(l-j)N}}{1-\omega^{(l-j)}}, j \neq l \\ &= 0 \text{ si } j \neq l\end{aligned}$$

Si $j = l$, on obtient N.

En déduire la norme et le conditionnement de M_N pour la norme subordonnée à la norme hermitienne de $\mathbb{C}^N$ (on rapelle que pour toute matrice $A \in \mathscr{M}_N(\mathbb{C})$ et pour tout $v \in \mathbb{C}^N$, $\|Av\| = \sqrt{\langle v|A^*Av\rangle}$).

Donc $\|M_N v\|^2 = N\|v\|^2$, la norme de M_N est $\sqrt{N}$ et le conditionnement est 1.

(3) *On suppose que le N-uplet $x = (x_0, \ldots, x_{N-1})$ est connue avec une imprécision Δx. On note ΔX l'imprécision sur sa transformée $X = (X_0, \ldots, X_{N-1})$. Montrer que (pour la norme hermitienne) $\|\Delta X\| = \sqrt{N}\|\Delta x\|$ et $\frac{\|\Delta X\|}{\|X\|} = \frac{\|\Delta x\|}{\|x\|}$.*

C'est une conséquence immédiate de $\|M_N v\|^2 = N\|v\|^2$ et de $\Delta X = M_N \Delta x$

(4) *Même si les données sont connues avec précision, l'arrondi en virgule flottante est une source d'erreur que l'on se propose d'estimer.*

(a) *Donner, en fonction de $(x_1, \ldots, x_N)$ et de la précision des nombres flottants, une majoration de l'erreur absolue causée par les erreurs d'arrondis dans le calcul de la transformée de Fourier discrète.*

Cette question est délicate ! Si ε est l'erreur d'arrondi, on a une erreur relative de ε introduite à chaque opération, donc une erreur absolue de ε multipliée par la norme. Pour la somme $x_0 + \omega x_1 + \omega^2 x_2 + \ldots + \omega^{N-1} x_{N-1}$, on a une erreur de $\varepsilon(|x_1| + |x_2| + \ldots + |x_{N-1}|)$ pour les produits par ω et pour les sommes :

$$\begin{aligned}&\varepsilon(|x_0 + \omega x_1| + |x_0 + \omega x_1 + \omega^2 x_2| + \ldots + |x_0 + \omega x_1 + \omega^2 x_2 + \ldots + \omega^{N-1} x_{N-1}|) \leq \\ &\qquad \leq \varepsilon((N-1)|x_0| + (N-1)|x_1| + (N-2)|x_2| + \ldots + |x_{N-1}|)\end{aligned}$$

qui peut donc théoriquement atteindre $O(N^2\varepsilon)$, mais cette estimation est statistiquement pessimiste, en effet il n'y a qu'une probabilité infime que toutes les erreurs d'arrondi se fassent dans le même sens d'une part, et que les valeurs absolues de somme soient égales aux sommes de valeurs absolues d'autre part.

On peut d'ailleurs améliorer l'erreur en $O(N\ln(N)\varepsilon)$ en implémentant la somme d'une liste d'éléments de taille paire comme étant la somme de la somme de deux listes de taille moitié (mais il faut alors stocker en mémoire la liste des éléments à sommer).

On peut aussi faire la somme partielle S_k normalement et faire en parallèle une somme correctrice des erreurs d'arrondis $x_k - ((S_k + x_k) - S_k)$ à ajouter à S_N à la fin de la sommation, on a une précision meilleure mais au prix d'un temps de calcul plus long.

Pour accélérer les calculs, particulièrement dans le cas $N = 2^n$, l'algorithme de Cooley-Tukey utilise une stratégie de type "diviser pour régner". On commence par calculer (récursivement) deux transformées de Fourier discrètes de taille moitié : le $N/2$-uplet $(A_0, \ldots, A_{N/2-1})$, transformée des termes d'indice pair $(x_0, x_2, \ldots, x_{N-2})$, et le $N/2$-uplet $(B_0, \ldots, B_{N/2-1})$, transformée des termes d'indice impair $(x_1, x_3, \ldots, x_{N-1})$.

(5) *Montrer que pout tout $j \in \{0, \ldots, N-1\}$,*

$$X_j = \sum_{k=0}^{N/2-1} x_{2k}\, e^{-\frac{4i\pi}{N} jk} + e^{-\frac{2i\pi}{N} j} \sum_{k=0}^{N/2-1} x_{2k+1}\, e^{-\frac{4i\pi}{N} jk}.$$

On remplace ω par sa valeur, on décompose la somme en indices pairs et impairs et on factorise un ω dans la somme d'indices impairs.

(6) *En déduire que pour tout* $j \in \{0,\dots,N/2-1\}$, $X_j = A_j + e^{-\frac{2i\pi}{N}j}B_j$ *et* $X_{j+N/2} = A_j - e^{-\frac{2i\pi}{N}j}B_j$ *(on fera attention que* A_j *et* B_j *ne sont a priori définis que pour* $0 \le j \le N/2-1$*).*
Pour faire apparaitre les A_j et B_j, lorsque $j \in [0,N/2-1]$, on utilise que ω^2 est une racine primitive 2^{n-1}-ième de 1. Pour les indices j supérieurs, remarquer que $e^{-\frac{2i\pi}{N}(j+N/2)} = -e^{-\frac{2i\pi}{N}j}$ et $e^{-\frac{4i\pi}{N}(j+N/2)k} = e^{-\frac{4i\pi}{N}jk}$.

(7) *En déduire une méthode de calcul de la transformée de Fourier discrète utilisant seulement* $O(n2^n) = O(N\log(N))$ *additions et multiplications de nombres complexes.*
Si on utilise le procédé, le temps de calcul $T(n)$ pour $N = 2^n$ est égal à celui des A_N et B_N et de leur combinaison soit 2 fois le temps de calcul pour $N = 2^{n-1}$ plus un $O(N)$

$$T(n) = 2T(n-1) + C2^n = 2(2T(n-2) + C2^{n-1}) + C2^n = 2^2T(n-2) + 2C2^n =$$
$$= 2^2(2T(n-3) + C2^{n-2}) + 2C2^n = 2^3T(n-3) + 3C2^n = ... = 2^nT(0) + nC2^n$$

ce qui est bien un $O(n2^n)$

(8) On s'intéresse à nouveau à l'impact des erreurs d'arrondi.

(a) *Pour tout j, on note* $\Delta A_j = \hat{A}_j - A_j$*, respectivement* $\Delta B_j = \hat{B}_j - B_j$ *l'erreur (absolue) sur le calcul de* A_j*, respectivement* B_j*. Montrer que l'erreur absolue pour le calcul de* X_j *est majorée en module par* $|\Delta A_j| + |\Delta B_j| + \varepsilon(|A_j| + |B_j|)$*, où* ε *est la précision des nombres flottants.*
En effet l'erreur d'arrondi est relative et $|A_j + B_j| \le |A_j| + |B_j|$

(b) *On suppose pour simplifier que tous les composants de* $(X_0,\dots,X_{N-1})$ *sont de module proche de* $\sqrt{N} = 2^{n/2}$*, et que ceux de* $(A_0,\dots,A_{N/2-1})$ *et* $(B_0,\dots,B_{N/2-1})$ *sont tous de module proche de* $\sqrt{N/2} = 2^{(n-1)/2}$*. On note* u_{n-1} *une borne sur l'erreur relative du calcul des* A_j *et des* B_j*. Montrer qu'une borne sur l'erreur relative du calcul des* X_j *est*

$$u_n = \sqrt{2}(u_{n-1} + \varepsilon).$$

On divise l'erreur absolue sur X_j par $|X_j|$ supposé équivalent à $2^{n/2}$ pour obtenir u_n, et on utilise que $|A_j|$ et $|B_j|$ sont équivalents à $2^{n/2}/\sqrt{2}$

(c) *En déduire que* u_n *est de l'ordre de* $\varepsilon\sqrt{2^n}$ *(indication : on pourra utiliser la commande* `rsolve`*). Comment ce résultat se compare-t-il à celui de la question 4 ?*
`rsolve(u(n)=sqrt(2)*(u(n-1)+eps),u(n),u(0)=eps)`
renvoie

$$u_n = 2^{\frac{n}{2}} \cdot \varepsilon(\sqrt{2}+3) + \varepsilon(-\left(\sqrt{2}\right) - 2)$$

C'est bien mieux ! Et comme précédemment, l'estimation est statistiquement pessimiste.

Question bonus : Une multiplication de deux nombres complexes stockés sous forme algébrique revient à faire 4 multiplications réelles, à moins que l'un des facteurs ne soit réel ou imaginaire pur. Pour simplifier les calculs, Rader et Brenner ont proposé de remplacer $(B_0,\dots,B_{N/2-1})$ par $(C_0,\dots,C_{N/2-1})$, transformée de Fourier discrète de $(x_1 - x_{N-1} - Q, x_3 - x_1 - Q, \dots, x_{N-1} - x_{N-3} - Q)$ avec $Q = \frac{2}{N}\sum_{k=0}^{N-1} x_{2k+1}$. On obtient alors les formules suivantes, que l'on ne demande pas de justifier :

$$X_0 = A_0 + C_0, \quad X_{N/2} = A_0 - C_0, \quad \text{et pour tout } j \neq 0, N/2, \ X_j = A_j + \frac{C_j}{2i\sin(\frac{2\pi}{N}j)}$$

(9) *Expliquer pourquoi cette méthode n'est pas satisfaisante en terme de stabilité numérique.*
Cette méthode est à priori instable car on divise C_j par $\sin(\frac{2\pi}{N}j)$ qui est tres petit pour j proche de 1 ou de $N/2-1$, l'erreur absolue sur C_j est alors fortement amplifiée.