EISTI - DÉPARTEMENT MATHÉMATIQUES

CORRECTION DE L'EXAMEN DE RATTRAPAGE D'ANALYSE NUMÉRIQUE

6 juin 2006 – **DURÉE 3h00**

Exercice 1 : Décomposition LU

1. Soit la matrice A définie par :

$$A = \begin{pmatrix} 3 & -1 & 0 \\ -2 & 3 & -1 \\ 0 & -2 & 3 \end{pmatrix}$$

Décomposer A sous la forme LU. (On rappelle que L est à diagonale unité)
SOLUTION : On pose

$$L = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ 0 & l_{32} & 1 \end{pmatrix} \qquad U = \begin{pmatrix} u_{11} & u_{12} & 0 \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

et on calcule les termes de A successivement de haut en bas et de gauche à droite :
$a_{11} = 3 = u_{11}$
$a_{12} = -1 = l_{11}u_{12} \Rightarrow u_{12} = -1$ car $l_{11} = 1$
$a_{21} = -2 = l_{21}u_{11} \Rightarrow l_{21} = \frac{-2}{u_{11}} = \frac{-2}{3}$
$a_{22} = 3 = l_{21}u_{12} + u_{22} \Rightarrow u_{22} = 3 - \frac{-2}{3}.(-1) = \frac{7}{3}$
$a_{23} = -1 = l_{22}u_{23} \Rightarrow u_{23} = -1$
$a_{32} = -2 = l_{32}u_{22} \Rightarrow l_{32} = -\frac{6}{7}$
$a_{33} = 3 = l_{32}u_{23} + u_{33} \Rightarrow u_{33} = \frac{15}{7}$
On trouve donc

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{-2}{3} & 1 & 0 \\ 0 & -\frac{6}{7} & 1 \end{pmatrix} \qquad U = \begin{pmatrix} 3 & -1 & 0 \\ 0 & \frac{7}{3} & -1 \\ 0 & 0 & \frac{15}{7} \end{pmatrix}$$

2. On se place dans le cadre général où A est une matrice carrée de taille n. On suppose que A est tridiagonale c'est à dire de la forme :

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 & \cdots & \cdots & 0 \\ a_{21} & a_{22} & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{ij} & \ddots & 0 \\ & & \ddots & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & \cdots & 0 & a_{n,n-1} & a_{nn} \end{pmatrix}$$

On **admet** que les matrices L et U ont le même profil que A c'est à dire qu'en plus des termes nuls par définition dans L et U, seuls les termes diagonaux, surdiagonaux et sous-diagonaux peuvent être non nuls.

(a) Donner l'expression de a_{ij} en fonction des termes généraux des matrices L et U, en utilisant la structure de la matrice A

SOLUTION : Par définition du produit on a : $a_{ij} = \sum_{k=1}^{n} l_{ik}u_{kj}$ mais L et U ayant même profil que A, on a $l_{ik} = 0$ si $|i-k| > 1$ et $u_{kj} = 0$ si $|j-k| > 1$ donc $a_{ij} = \sum_{k=\max(i-1,j-1)}^{\min(i+1,j+1)} l_{ik}u_{kj}$. Mais U est triangulaire supérieure donc $u_{kj} = 0$ si $k > j$, et L est triangulaire inférieure donc $l_{ik} = 0$ si $k > i$, donc $a_{ij} = \sum_{k=\max(i-1,j-1)}^{\min(i,j)} l_{ik}u_{kj}$.

(b) Afin d'optimiser le stockage en mémoire de la matrice A on utilise des vecteurs et non une matrice. Soient donc $\overrightarrow{a}$ et $\overrightarrow{b}$ deux vecteurs de $\mathbb{R}^n$ donnés et $\overrightarrow{c}$ et $\overrightarrow{d}$ deux vecteurs de $\mathbb{R}^{n-1}$ donnés. On considère alors la matrice A définie par :

$$A = \begin{pmatrix} a_1 & c_1 & 0 & \cdots & \cdots & 0 \\ d_1 & a_2 & c_2 & \ddots & & \vdots \\ 0 & d_2 & a_3 & c_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ & & \ddots & \ddots & \ddots & c_{n-1} \\ 0 & \cdots & \cdots & 0 & d_{n-1} & a_n \end{pmatrix}$$

et on veut résoudre le système linéaire $A\overrightarrow{x} = \overrightarrow{b}$ en utilisant la décomposition LU de A. La décomposition est effectuée en écrasant au fur et à mesure les valeurs de A stockées, afin de ne comsommer aucune autre place mémoire.

i. Compléter l'algorithme suivant :
Entrées/ Sorties :
Entrées : $\overrightarrow{a}$, $\overrightarrow{c}$, $\overrightarrow{d}$
Sorties : la décomposition est stockée dans les vecteurs donnés en entrée

SOLUTION : **Algorithme de la décomposition LU:**
$d(1) := \frac{d(1)}{a(1)}$
Pour $i = 2$ à $n-1$ faire
. $a(i) := a(i) - \mathbf{d(i-1) * c(i-1)}$
. $d(i) := \mathbf{d(i)/a(i)}$
FPour
$a(n) := \mathbf{a(n) - d(n-1) * c(n-1)}$
Remarque :Quelques calculs sont nécessaires pour arriver à remplir les cases.

ii. Après la décomposition LU de la matrice, on effectue la résolution. Expliquez en quoi la décomposition LU est utile pour la résolution du système linéaire $A\overrightarrow{x} = \overrightarrow{b}$.

SOLUTION : Lorsque la matrice est décomposée sous la forme LU, pour résoudre le système $A\overrightarrow{x} = \overrightarrow{b}$, on n'a plus qu'à résoudre deux systèmes triangulaires. Il suffit donc de faire une descente pour résoudre $L\overrightarrow{y} = \overrightarrow{b}$ et une remontée pour résoudre $U\overrightarrow{x} = \overrightarrow{y}$.

iii. Compléter les algorithmes suivants :
Entrées/ Sorties :
Entrées : $\overrightarrow{a}$, $\overrightarrow{c}$, $\overrightarrow{d}$, $\overrightarrow{b}$

Sorties : $\overrightarrow{b}$ dans lequel on stocke la solution $\overrightarrow{y}$ puis la solution $\overrightarrow{x}$
SOLUTION : **Résolution du système linéaire** $L\overrightarrow{y} = \overrightarrow{b}$
Pour $i = 1$ à $n-1$ faire
. $\quad b(i+1) := \mathbf{b(i+1) - d(i).*b(i)}$
FPour

Résolution du système linéaire $U\overrightarrow{x} = \overrightarrow{y}$
$b(n) = b(n)/a(n)$
Pour $i = n-1$ à 1 par pas de -1 faire
. $\quad b(i) := [b(i) - \mathbf{c(i)*b(i+1)}]/\mathbf{a(i)}$
FPour

Exercice 2 : Moindres carrés

On considère une résistance dont la valeur R varie en fonction de la température θ selon la loi

$$R = a\theta + b\theta^2$$

où a et b sont deux paramètres inconnus que l'on cherche à estimer à partir de m mesures (θ_k, R_k)

1. Proposer une méthode.(On exposera et on détaillera soigneusement la résolution de système permettant d'obtenir le meilleur résultat numérique en machine.)
   SOLUTION : On constitue une matrice M à l'aide des mesures de θ_k et θ_k^2 placées en ligne, les colonnes étant constituées des valeurs de θ_k et θ_k^2 respectivement. De cette façon le problème s'exprime $MK = R$ avec $K = \begin{pmatrix} a \\ b \end{pmatrix}$ vecteur des inconnues, et $R = (R_k)_{k=1,\dots m}$. On calcule donc les équations normales : $M^TMK = M^TR$. A l'aide de la SVD on décompose $M = U\sum V^T$ où U et V sont des matrices orthogonales et on en déduit $K = \left(M^TM\right)^{-1}M^TR = \left(V\sum^T U^TU\sum V^T\right)^{-1}V\sum^T U^TR = \left(V\sum^T\sum V^T\right)^{-1}V\sum^T U^TR = V\left(\sum^T\sum\right)^{-1}V^TV\sum^T U^TR = V\left(\sum^T\sum\right)^{-1}\sum^T U^TR$. On pose alors $\sum^{-1} = \left(\sum^T\sum\right)^{-1}\sum^T$ et on obtient $K = V\sum^{-1}U^TR = M^+R$ où M^+ désigne le pseudo inverse de Moore Penrose.

2. Application numérique : $m = 3$ et on a effectué les mesures suivantes :

θ	R
0	1
1	2
2	5

   Déterminer les valeurs des paramètres a et b.
   SOLUTION : On constitue M et R :

$$M = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 2 & 4 \end{pmatrix} \qquad R = \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}$$

   On résout le système des équations normales par inversion $M^TMK = M^TR \Rightarrow M^TMK = (M^TM)^{-1}M^TR$ avec

$$M^TM = \begin{pmatrix} 5 & 9 \\ 9 & 17 \end{pmatrix} \qquad \text{et } \left(M^TM\right)^{-1} = \frac{1}{4}\begin{pmatrix} 17 & -9 \\ -9 & 5 \end{pmatrix}$$

On en déduit

$$K = \frac{1}{4}\begin{pmatrix} 17 & -9 \\ -9 & 5 \end{pmatrix}\begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 1/2 \end{pmatrix}$$

3. On constate au cours des mesures, qu'une erreur entache la troisième mesure, c'est à dire que la valeur $\theta = 2$ est remplacée par $\theta = 2 + \varepsilon$ sans que la valeur de R soit modifiée. On suppose que ε est de la forme 10^{-i} pour $i \in \mathbb{N}$ et tel que $0 \leq i \leq 9$.
   On note
   $$A_\varepsilon = \begin{pmatrix} 1 + (2+\varepsilon)^2 & 1 + (2+\varepsilon)^3 \\ 1 + (2+\varepsilon)^3 & 1 + (2+\varepsilon)^4 \end{pmatrix}$$
   et
   $$\Delta A = A_\varepsilon - A_0$$
   On suppose que l'influence de cette erreur est négligeable au second membre au sens où le vecteur $\begin{pmatrix} 12 + 5\varepsilon \\ 22 + 10\varepsilon + 5\varepsilon^2 \end{pmatrix}$ peut être remplacé par $\begin{pmatrix} 12 \\ 22 \end{pmatrix}$.
   La résolution du problème est effectuée sur les mesures entachées d'erreur avec la même méthode que sur des mesures exactes.Différents indicateurs sont évalués et représentés sur le graphique ci-dessous (deltaA correspond à ΔA, et A correspond à A_0) .
   On veut savoir pour quelles valeurs de i on aura une erreur relative sur la solution obtenue inférieure à 10^{-5} (c'est à dire si x est la solution exacte et Δx l'erreur sur cette solution, $\frac{\|\Delta x\|}{\|x\|} \leq 10^{-5}$) .

   Question : Pouvez-vous, à partir du graphique et de résultats sur les majorations d'erreur, donner les valeurs de i assurant une erreur relative sur la solution obtenue inférieure à 10^{-5} ?
   Indication : $\ln(10^{-5}) \simeq -11.5$

SOLUTION : Si on note M_ε la matrice des données perturbées, la résolution se ramène à $M_\varepsilon K_\varepsilon = R$. En prémultipliant par M_ε^T on obtient $M_\varepsilon^T M_\varepsilon K_\varepsilon = M_\varepsilon^T R$. On s'aperçoit que $\begin{pmatrix} 12 + 5\varepsilon \\ 22 + 10\varepsilon + 5\varepsilon^2 \end{pmatrix} = M_\varepsilon^T R$ et que ce vecteur peut être assimilé à $M^T R$, donc on résout :$M_\varepsilon^T M_\varepsilon K_\varepsilon = M^T R$. On remarque que $A_\varepsilon = M_\varepsilon^T M_\varepsilon$ donc on est ramené à résoudre $A_\varepsilon K_\varepsilon = M^T R$. Le système initial sans perturbation est : $A_0 K = AK = M^T R$.On sait d'après les résultats de majoration d'erreur que

$$\frac{\|\Delta K\|}{\|K\|} \leq cond(A)\frac{\|\Delta A\|}{\|A\|}$$

où $\Delta K = K_\varepsilon - K$, et $\Delta A = A_\varepsilon - A$.
Une erreur relative sur la solution obtenue inférieure à 10^{-5} sera assurée si on a $cond(A)\frac{\|\Delta A\|}{\|A\|} \leq 10^{-5}$ soit $\ln\left(cond(A)\frac{\|\Delta A\|}{\|A\|}\right) \leq \ln(10^{-5}) \simeq -11.5$. Il suffit donc de lire sur le graphique pour quelle valeur de i la droite représentative de $ln\left(cond(A)\frac{\|\Delta A\|}{\|A\|}\right)$ intersecte $y = -11.5$. On trouve $i \geq 7.5$ environ.

Exercice 3 : Intégration numérique

Soit $f : [0,1] \to \mathbb{R}$ de classe $\mathcal{C}^1$. On considére la formule d'intégration approchée :

$$\mathcal{F} : \int_0^1 f(t)dt \simeq J(f) \text{ où } J(f) = \frac{1}{2}[f(0) + f(1)] + \frac{1}{12}[f'(0) - f'(1)] \qquad (1)$$

1. Montrons que $\mathcal{F}$ peut être obtenue en faisant en sorte que la formule de calcul approché
$$\int_0^1 f(t)dt \simeq af(0) + bf(1) + cf'(0) + df'(1)$$
soit d'ordre maximal.
On a 4 constantes à déterminer donc on doit imposer 4 conditions, c'est à dire l'exactitude de la formule pour les polynômes de degré 0 à 3.
Pour que la formule soit d'ordre 0 il faut qu'elle soit exacte pour les constantes soit :
$$\int_0^1 \alpha dt = a.\alpha + b.\alpha = \alpha \Leftrightarrow a + b = 1 \tag{2}$$
Pour que la formule soit d'ordre 1 il faut qu'elle soit exacte pour $f : t \mapsto t$ soit :
$$\int_0^1 tdt = \frac{1}{2} = a.0 + b + c + d \Leftrightarrow b + c + d = \frac{1}{2} \tag{3}$$
Pour que la formule soit d'ordre 2 il faut qu'elle soit exacte pour $f : t \mapsto t^2$ soit :
$$\int_0^1 t^2dt = \frac{1}{3} = a.0 + b + 2c.0 + 2d \Leftrightarrow b + 2d = \frac{1}{3} \tag{4}$$
Pour que la formule soit d'ordre 3 il faut qu'elle soit exacte pour $f : t \mapsto t^3$ soit :
$$\int_0^1 t^3dt = \frac{1}{4} = a.0 + b + 3c.0 + 3d \Leftrightarrow b + 3d = \frac{1}{4} \tag{5}$$
On constitue donc le système avec les équations (**??**) à (**??**) et sa résolution conduit à : $a = \frac{1}{2}, b = \frac{1}{2}, c = \frac{1}{12}, d = -\frac{1}{12}$. On obtient donc bien (**??**)

2. Montrons que $\mathcal{F}$ peut aussi être obtenue en appliquant la méthode de Simpson au calcul de $\int_0^1 \left(t - \frac{1}{2}\right) f'(t)dt$ (on pourra intégrer par parties).
On calcule $\int_0^1 \left(t - \frac{1}{2}\right) f'(t)dt$ par parties et on obtient :
$$\int_0^1 \left(t - \frac{1}{2}\right) f'(t)dt = \frac{1}{2}\left[f(0) + f(1)\right] - \int_0^1 f(t)dt$$
On en déduit que
$$\int_0^1 f(t)dt = \frac{1}{2}\left[f(0) + f(1)\right] - \int_0^1 \left(t - \frac{1}{2}\right) f'(t)dt \tag{6}$$
La formule de Simpson est
$$\int_0^1 g(t)dt \simeq \frac{1}{6}\left[g(0) + 4g(\frac{1}{2}) + g(1)\right]$$
On l'applique pour approcher $\int_0^1 \left(t - \frac{1}{2}\right) f'(t)dt$ dans (**??**), soit :
$$\begin{aligned}\int_0^1 \left(t - \frac{1}{2}\right) f'(t)dt &\simeq \frac{1}{6}\left[\left(0 - \frac{1}{2}\right) f'(0) + 4\left(\frac{1}{2} - \frac{1}{2}\right) f'(\frac{1}{2}) + \left(1 - \frac{1}{2}\right) f'(1)\right] \\ &= \frac{1}{6}\left[-\frac{1}{2}f'(0) + \frac{1}{2}f'(1)\right] = \frac{1}{12}\left[-f'(0) + f'(1)\right]\end{aligned}$$
d'où d'après (**??**) :
$$\int_0^1 f(t)dt \simeq \frac{1}{2}\left[f(0) + f(1)\right] + \frac{1}{12}\left[f'(0) - f'(1)\right]$$

3. La méthode est exacte pour les polynomes de degré 3 puisqu'on a montré au 1)a) qu'elle peut être construite de cette façon.

4. La méthode est exacte pour les polynomes de degré 4 si la quantité

$$\int_0^1 t^4 dt = \left[\frac{t^5}{5}\right]_0^1 = \frac{1}{5}$$

coïncide avec $\frac{1}{2}[f(0)+f(1)] + \frac{1}{12}[f'(0)+f'(1)]$ pour $f(t)=t^4$. On a dans ce cas :

$$\frac{1}{2}[f(0)+f(1)] + \frac{1}{12}[f'(0)-f'(1)] = \frac{1}{2}[0+1] + \frac{1}{12}[0-4] = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

Les deux quantités ne coïncident pas donc la méthode n'est pas exacte pour les polynomes de degré 4.

5. L'ordre de la méthode $\mathcal{F}$ étant le degré maximal pour lequel elle est exacte, on déduit de ce qui précède que la méthode est d'ordre 3.

6. On suppose que f est de classe $\mathcal{C}^4$. En intégrant par parties, vérifier que

$$\int_0^1 f(t)dt = J(f) + \frac{1}{24}\int_0^1 [t\,(t-1)]^2 f^{(4)}(t)dt$$

On évalue $I = \int_0^1 [t\,(t-1)]^2 f^{(4)}(t)dt$ en intégrant par parties. On obtient :

$$\begin{aligned}\int_0^1 [t\,(t-1)]^2 f^{(4)}(t)dt &= \left[2\,[t\,(t-1)]\,(2t-1)f^{(3)}(t)\right]_0^1 - \int_0^1 2t\,(t-1)\,(2t-1)f^{(3)}(t)dt \\ &= -\int_0^1 2t\,(t-1)\,(2t-1)f^{(3)}(t)dt\end{aligned}$$

On intègre encore par parties :

$$\begin{aligned} I &= -2\int_0^1 t\,(t-1)\,(2t-1)f^{(3)}(t)dt \\ &= -2\left[t\,(t-1)\,(2t-1)f^{(2)}(t)\right]_0^1 + 2\int_0^1 (6t^2-6t+1)f^{(2)}(t)dt \\ &= 2\int_0^1 (6t^2-6t+1)f^{(2)}(t)dt \end{aligned}$$

On intègre encore par parties :

$$\begin{aligned} I &= 2\int_0^1 (6t^2-6t+1)f^{(2)}(t)dt \\ &= 2\left\{\left[(6t^2-6t+1)f'(t)\right]_0^1 - \int_0^1 (12t-6)f'(t)dt\right\} \\ &= 2\left\{f'(1)-f'(0) - 6\int_0^1 (2t-1)f'(t)dt\right\} \\ &= 2\,(f'(1)-f'(0)) - 12\int_0^1 (2t-1)f'(t)dt \end{aligned}$$

On intègre encore par parties :

$$\begin{aligned} I &= 2\left(f'(1)-f'(0)\right) - 12\left\{[(2t-1)f(t)]_0^1 - 2\int_0^1 f(t)dt\right\} \\ &= 2\left(f'(1)-f'(0)\right) - 12\left(f(1)+f(0)\right) + 24\int_0^1 f(t)dt \end{aligned}$$

On en déduit que

$$\frac{1}{24}I = \underbrace{-\frac{1}{12}\left(f'(0)-f'(1)\right) - \frac{1}{2}\left(f(1)+f(0)\right)}_{-J(f)} + \int_0^1 f(t)dt$$

dont on déduit immédiatement le résultat demandé.

7. On applique la formule de la moyenne aux fonctions $t \mapsto [t(t-1)]^2$ et $t \mapsto f^{(4)}(t)$ qui est continue d'après l'hypothèse $f \in \mathcal{C}^4$. On vérifie facilement que $t \mapsto [t(t-1)]^2$ est toujours positive sur $[0,1]$. On en déduit qu'il existe $c \in [0,1]$ tel que

$$\int_0^1 f(t)dt = J(f) + \frac{1}{24}f^{(4)}(c)\int_0^1 [t(t-1)]^2\, dt.$$

On vérifie que $\int_0^1 [t(t-1)]^2\, dt = \frac{1}{30}$ d'où

$$\int_0^1 f(t)dt = J(f) + \frac{1}{24x30}f^{(4)}(c) = J(f) + \frac{1}{720}f^{(4)}(c)$$

Exercice 4 : Recherche de racines

On considère la fonction continue $f(x) = \ln(x+1) - 2$ sur l'intervalle $I = [4, 12]$.
On rappelle l'arsenal théorique suivant :

- *Théorème 1 :* Une méthode de point fixe de la forme $x_{n+1} = \phi(x_n)$ converge vers x solution de $x = \phi(x)$ si et seulement si l'application ϕ est strictement contractante, i.e. il existe une constante $L < 1$ telle que ϕ est lispchitzienne de rapport L.

- *Définition :*Une fonction f définie sur un intervalle I, est lipschitzienne de rapport L si :

$$\forall x, y \in I, \quad |f(x) - f(y)| \leq L\,|x-y|$$

  On peut montrer que f est lipschitzienne de rapport L si et seulement si $\forall x \in I, \quad f'(x) \leq L$

- *Théorème des accroissements finis :* Soit $f : [a,b] \to R$ une fonction à valeurs réelles (a et b réels tels que $a < b$). Si :

  – f est continue sur l'intervalle fermé $[a,b]$
  – f est dérivable sur l'intervalle ouvert $]a,b[$,

  alors il existe $c \in]a,b[, \qquad f'(c) = \frac{f(b)-f(a)}{b-a}$

- Itérations de la méthode de Newton : $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$

- Formule de Taylor-Lagrange : Soit f une fonction de classe C^{n+1} sur un intervalle $[a,b]$, et $x \in [a,b]$. Alors il existe $c \in [a,x]$ vérifiant :

$$f(x) = f(a) + \sum_{k=1}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) + \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(c)$$

1. En sachant que $5 < e^2 < 13$, montrer qu'il existe un zéro α pour la fonction f dans I et qu'il est unique.

   SOLUTION : On a : $f(4) = \ln 5 - 2$ donc $e^{f(4)} = \frac{5}{e^2} < 1$ d'où $f(4) < 0$.
   De plus : $f(12) = \ln 13 - 2$ donc $f(12) > 0$. Comme par ailleurs f est continue et croissante sur I on applique le théorème des valeurs intermédiaires et on en déduit que $\exists \alpha \in I / \quad f(\alpha) = 0$.

2. On utilise la méthode de bissection pour déterminer α.

   (a) On note $[a,b]$ l'intervalle de départ et (a_n) et (b_n) les suites utilisées pour déterminer α. Etablir une majoration de $b_n - a_n$ en fonction de a, b, n.
   SOLUTION : On a $\left\{ \begin{array}{l} a_0 = a \\ b_0 = ab \end{array} \right.$ puis $b_1 - a_1 = \frac{b_0 - a_0}{2}$, $b_2 - a_2 = \frac{b_1 - a_1}{2} = \frac{b_0 - a_0}{2^2}$ et on en déduit par récurrence que $b_n - a_n = \frac{b_0 - a_0}{2^n} = \frac{b-a}{2^n}$

   (b) Estimer le nombre minimal d'itérations nécessaires pour calculer le zéro α avec une tolérance $tol = 10^{-10}$. On donne $\ln 5 \simeq 1.6$
   SOLUTION : Pour calculer le zéro α avec une tolérance $tol = 10^{-10}$ il suffit que $b_n - a_n \leq 10^{-10}$ soit $\frac{b-a}{2^n} \leq 10^{-10} \Leftrightarrow b-a \leq 2^n.10^{-10} \Leftrightarrow \ln(b-a) \leq n \ln 2 - 10 \ln 10 \Leftrightarrow \ln(b-a) + 10 \ln 10 \leq n \ln 2 \Leftrightarrow n \leq \frac{\ln(b-a) + 10 \ln 10}{\ln 2} = \frac{\ln 8 + 10 \ln 10}{\ln 2} = 13 + 10 \frac{\ln 5}{\ln 2}$. Cette quantité est environ égale à 36.

3. On considère la méthode de point fixe $x_{n+1} = \phi(x_n)$ avec $\phi(x) = x - \ln(x+1) + 2$.

   (a) Montrer que cette méthode est convergente sur l'intervalle I
   SOLUTION : D'après le théorème 1 il suffit de montrer que ϕ est contractante, et ceci équivaut à montrer que $\phi'(x) < 1$. $\phi(x) = x - \ln(x+1) + 2$. $\Rightarrow \phi'(x) = 1 - \frac{1}{x+1}$. Sur $I = [4,12]$, $x \leq 12 \Rightarrow x+1 \leq 13 \Rightarrow \frac{1}{x+1} \geq \frac{1}{13} \Rightarrow \phi'(x) = 1 - \frac{1}{x+1} \leq 1 - \frac{1}{13} = \frac{12}{13} < 1$ donc la méthode est convergente. De plus on sait que la constante de Lipschitz vaut $\frac{12}{13}$.

   (b) Etablir une majoration de $|x_n - \alpha|$ en fonction de x_0, α, L (coefficient de Lipschitz de ϕ), n.
   SOLUTION : $|x_n - \alpha| = |\phi(x_{n-1}) - \phi(\alpha)| \leq L |x_{n-1} - \alpha|$ car $\alpha = \phi(\alpha)$, donc on en déduit par récurrence que $|x_n - \alpha| \leq L^n |x_0 - \alpha|$

   (c) En supposant que $|x_0 - \alpha| \leq 10^{-1}$, trouver le nombre minimal d'itérations de point fixe nécessaire pour avoir une erreur inférieure à 10^{-10}.
   SOLUTION : Pour que l'erreur soit inférieure à 10^{-10}, il suffit que $L^n |x_0 - \alpha| \leq 10^{-10} \Leftrightarrow n \ln L \leq -10 \ln 10 - \ln |x_0 - \alpha| \Leftrightarrow n \geq \frac{-10 \ln 10 - \ln |x_0 - \alpha|}{\ln L}$ car $\ln L < 0$.
   En supposant que $|x_0 - \alpha| \leq 10^{-1}$, on a $\ln |x_0 - \alpha| \leq -\ln 10 \Rightarrow -\ln |x_0 - \alpha| \geq \ln 10$ donc la condition sur n revient à $n \geq \frac{-10 \ln 10 + \ln 10}{\ln L} = \frac{-9 \ln 10}{\ln 12 - \ln 13} \simeq 16.3$. Il faut donc faire au moins 16 itérations.

4. On considère enfin la méthode de Newton

(a) Ecrire le terme général y_n de la suite définie par la méthode de Newton pour trouver le zéro α de f.
SOLUTION :

$$y_{n+1} = y_n - \frac{f(y_n)}{f'(y_n)} = 3y_n + 2 - (y_n + 1)\ln(y_n + 1)$$

(b) Trouver une constante $C < 1$ telle que l'inégalité suivante

$$|y_{n+1} - \alpha| \leq C\,|y_n - \alpha|^2$$

soit satisfaite.
SOLUTION : $|y_{n+1} - \alpha| = |3y_n + 2 - (y_n + 1)\ln(y_n + 1) - \alpha|$. Appliquons la formule de Taylor Lagrange à f sur $[\alpha, x]$. On en déduit l'existence de $c \in [\alpha, x]$:

$$f(\alpha) = f(x) + (\alpha - x)\,f'(x) + \frac{(\alpha - x)^2}{2}f''(c)$$

Si on pose $x = y_n$ et que l'on divise par $f'(x)$ alors $\exists c_n \in [\alpha, y_n]$

$$\begin{aligned} 0 &= \frac{f(y_n)}{f'(y_n)} + (\alpha - y_n) + \frac{(\alpha - y_n)^2}{2f'(y_n)}f''(c_n) \\ &\Rightarrow \left| y_n - \frac{f(y_n)}{f'(y_n)} - \alpha \right| = |y_{n+1} - \alpha| = \left| \frac{(\alpha - y_n)^2}{2f'(y_n)}f''(c_n) \right| \end{aligned}$$

or

$$f''(x) = -\frac{1}{(x+1)^2} \text{ et } f'(x) = \frac{1}{x+1}\text{sur } I$$

donc

$$\left| f''(x) \right| = \frac{1}{(x+1)^2} \leq 5^{-2} \text{ et } f'(x) = \frac{1}{x+1} \geq \frac{1}{13} \text{ sur } I$$

d'où

$$|y_{n+1} - \alpha| = \left| \frac{(\alpha - y_n)^2}{2f'(y_n)}f''(c_n) \right| \leq \frac{13}{2.5^2}(\alpha - y_n)^2$$

(c) En supposant que $|y_0 - \alpha| \leq 10^{-1}$, trouver une majoration permettant de déterminer le nombre minimal n d'itérations de la méthode de Newton nécessaire pour avoir une erreur inférieure à 10^{-10}. (On ne demande pas de mener les calculs pour trouver la valeur numérique)
SOLUTION : $|y_{n+1} - \alpha| \leq C\,|y_n - \alpha|^2$ donc on en déduit que

$$\begin{aligned} |y_{n+1} - \alpha| &\leq C^{1+2+...2^n}\,|y_0 - \alpha|^{2^{n+1}} \\ &\Leftrightarrow |y_n - \alpha| \leq C^{.2^n - 1}\,|y_0 - \alpha|^{2^n} \end{aligned}$$

Pour que $|y_n - \alpha| \leq 10^{-10}$ il suffit que $C^{.2^n-1}\,|y_0 - \alpha|^{2^n} \leq 10^{-10}$ soit

$$(2^n - 1)\ln C + 2^n \ln|y_0 - \alpha| \leq -10\ln 10$$

Après calculs il vient :

$$2^n \geq \frac{\ln C - 10\ln 10}{\ln\left[C\,|y_0 - \alpha|\right]}$$

d'où

$$n \geq \ln\left(\frac{\ln C - 10\ln 10}{\ln\left[C\left|y_0 - \alpha\right|\right]}\right) * \frac{1}{\ln 2}$$

Numériquement on obtient :

$$n \geq \ln\left(\frac{\ln C - 10\ln 10}{\ln C - \ln 10}\right) * \frac{1}{\ln 2} \simeq 2,1$$

On peut donc conclure que 3 itérations suffisent.