MT09-A2015 – Examen médian – Questions de cours
Durée : 30mn. Sans documents ni outils électroniques – Rédiger sur l'énoncé

NOM PRÉNOM :

ATTENTION, il y a QUATRE exercices indépendants pour cette partie questions de cours !

Exercice 1 *(barème approximatif : 1.5 points)*

1. Énoncer sans le démontrer le théoréme du point fixe. On précisera bien toutes les hypothèses et toutes les conclusions.

   Réponse : cf. cours, Proposition IV.2.2. Attention, TOUTES les hypothèses sont importantes : g doit être définie sur un intervalle fermé (borné) $[a,b]$ $(a<b)$, cet intervalle $[a,b]$ doit être stable par g $(g([a,b] \subset [a,b]))$, g doit être C^1 sur $[a,b]$ et g doit être contractante : il existe k, $0 \leq k < 1$ tel que pour tout x de $[a,b]$, $|g'(x)| \leq k$.

   Conclusion : existence et unicité du point fixe dans $[a,b]$ et convergence de la suite $(x_n)_{n\in\mathbf{N}}$ définie par $x_{n+1} = g(x_n)$ pour tout $x_0 \in [a,b]$ vers ce point fixe. □

2. Soit $f : \mathbb{R} \to \mathbb{R}$ et $g : \mathbb{R} \to \mathbb{R}$ deux fonctions. On suppose que $\sup_{x\in\mathbf{R}} |f(x)| < +\infty$ et $\sup_{x\in\mathbf{R}} |g(x)| < +\infty$. Montrer que

$$\sup_{x\in\mathbf{R}} |f(x)+g(x)| \leq \sup_{x\in\mathbf{R}} |f(x)| + \sup_{x\in\mathbf{R}} |g(x)|. \qquad (1)$$

   Réponse : on a d'après l'inégalité triangulaire :

$$\begin{aligned} \forall x \in \mathbb{R} \quad |f(x)+g(x)| &\leq |f(x)| + |g(x)| \\ &\leq \sup_{x\in\mathbf{R}} |f(x)| + \sup_{x\in\mathbf{R}} |g(x)| = M < +\infty, \end{aligned}$$

   où $M \in \mathbb{R}$ est un majorant indépendant de $x \in \mathbb{R}$. Donc le $\sup_{x\in\mathbf{R}} |f(x)+g(x)|$ existe et est fini, et comme c'est le plus petit des majorants, il reste inférieur à ce majorant M. On a montré (1).

   □

Exercice 2 *(barème approximatif : 1 point)*

Soit la matrice

$$A = \begin{bmatrix} -4 & 1 \\ 3 & -2 \\ 2 & 5 \end{bmatrix}.$$

1. Calculer $|||A|||_\infty$, $|||A|||_1$, $|||A|||_2$.

   Réponse : On a $|||A|||_\infty = \max_{i=1,2,3} \sum_{j=1}^{2} |A_{ij}| = \|\underline{A}_3\|_\infty = 7$, $|||A|||_1 = \max_{j=1,2} \sum_{i=1}^{3} |A_{ij}| = \|A_1\|_1 = 9$. Comme

$$A^T A = \begin{bmatrix} 29 & 0 \\ 0 & 30 \end{bmatrix}$$

   est une matrice diagonale, ses valeurs propres sont sur la diagonales et on a $|||A|||_2 = \sqrt{\rho(A^TA)} = \sqrt{30}$. □

Exercice 3 *(barème approximatif : 1.5 points)*

Soit une matrice $A \in \mathcal{M}_{mn}(\mathbb{R})$ (avec m et $n > 0$).

1. Montrer que $\mathrm{Ker}(A) = \mathrm{Ker}(A^TA)$.

   Réponse : On travaille par double inclusion.

   Soit $x \in \mathrm{Ker}(A)$, alors $Ax = 0$, donc $A^TAx = A^T0 = 0$ par linéarité, donc $\mathrm{Ker}(A) \subset \mathrm{Ker}(A^TA)$.

   Réciproquement, si $x \in \mathrm{Ker}(A^TA)$, alors $A^TAx = 0$, donc $x^TA^TAx = 0$. Or $x^TA^TAx = (Ax)^T(Ax) = \|Ax\|_2^2$. Donc $\|Ax\|_2^2 = 0$ implique que $Ax = 0$ d'après les propriétés de la norme, et donc $\mathrm{Ker}(A^TA) \subset \mathrm{Ker}(A)$.

   Finalement on a bien $\mathrm{Ker}(A) = \mathrm{Ker}(A^TA)$. □

2. Montrer que A^TA est symétrique semi-définie positive.

Réponse : A^TA est symétrique car $(A^TA)^T = A^T(A^T)^T = A^TA$. De plus, soit $x \in \mathbb{R}^n$, on a $x^TA^TAx = \|Ax\|_2^2 \geq 0$, donc A^TA est symétrique semi-définie positive. □

3. Montrer que les valeurs propres de A^TA sont positives ou nulles.

Réponse : soit $\lambda \in \mathbb{R}$ une valeur propre de A^TA (qui est symétrique donc ses valeurs propres sont réelles) et $y \neq 0$ un vecteur propre associé. Alors $A^TAy = \lambda y$ implique $\|Ay\|_2^2 = y^TA^TAy = \lambda y^Ty = \lambda\|y\|_2^2$, d'où on déduit (car $y \neq 0$) que $\lambda = \frac{\|Ay\|_2^2}{\|y\|_2^2} \geq 0$.

□

Exercice 4 *(barème approximatif : 1.5 points)*

1. Définir l'ensemble des flottants $\mathcal{F}_{10}$. On expliquera ce que signifie les constantes t, L et U (notations du cours).

Réponse : cf. cours.

$$\mathcal{F}_{10} = \left\{ \pm 0.d_1d_2\dots d_t\ 10^e \mid d_i \in \{0,1,2,\dots,9\}\ \forall i = 1,\dots,t, \quad d_1 \neq 0, \quad L \leq e \leq U \right\} \cup \{0\},$$

où t est le nombre de chiffres significatifs, L et U constituent les bornes inférieure et supérieure de l'exposant e. Par convention, l'exposant e est choisi de façon que le premier chiffre d_1 soit toujours non-nul. Le nombre 0 est explicitement inséré dans $\mathcal{F}_{10}$ car 0 ne s'écrit pas comme un nombre flottant normal. □

2. Dans le reste de cet exercice, on prend $t = 2$, $L = -1$, $U = 3$.

(a) Que vaut l'écart entre deux flottants successifs sur $[0.1,\ 1]$?

Réponse : sur $[0.1,\ 1]$, l'exposant vaut $e = 0 \in [-1,3]$. Les flottants de $\mathcal{F}_{10}$ sont : $0.10,\ 0.11,\ 0.12,\ 0.13,\ \dots,\ 0.98,\ 0.99,\ 1.0$. L'écart entre deux flottants successifs, noté δ_0, vaut : $\delta_0 = 0.01 = 10^{-t}$.

□

(b) Donner une majoration de l'écart relatif entre deux flottants successifs. Que vaut $\varepsilon_{\text{mach}}$?

Réponse : pour $e \in \{-1,0,1,2,3\}$, sur chaque intervalle $[0.1,\ 1] \times 10^e$, l'écart entre deux flottants successifs f_1 et f_2, noté $\delta_e = |f_2 - f_1|$, vaut : $\delta_e = 0.01 \times 10^e = 10^{-t+e}$. Comme ces deux flottants successifs sont supérieurs à 0.10×10^e, il vient :

$$\frac{|f_2 - f_1|}{|f_1|} = \frac{\delta_e}{|f_1|} \leq \frac{10^{-t+e}}{0.10 \times 10^e} = 10^{-t+1}$$

Dans le cours, on définit $\varepsilon_{\mathbf{mach}} = \frac{1}{2}10^{-t+1}$. C'est l'erreur relative maximale quand on approche un réel dans $[0.10 \times 10^L, 0.99 \times 10^U]$ par un flottant. □

3. (a) Calculer en addition flottante : $x = 100 \oplus 0.6$. Que vaut l'erreur relative?

Réponse : il faut aligner les nombres sur l'exposant le plus grand. Note: on prend ici la notation scientifique standard $d_1.d_2 \times 10^{e'}$ ($d_1 \neq 0$, $e' = e - 1 \in \{-2,-1,0,1,2\}$) que les humains sont en général plus habitués à manipuler.

$$\begin{aligned} x &= 100 \oplus 0.6 = 1.0 \times 10^2 + 6.0 \times 10^{-1} \\ &= 1.0 \times 10^2 + 0.006 \times 10^2 = 1.006 \times 10^2 \\ &= 1.0 \times 10^2 \end{aligned}$$

car on arrondit au plus proche flottant de $\mathcal{F}_{10}$. □

(b) Calculer en addition flottante : $y = (100 \oplus 0.6) \ominus 100$. Que vaut l'erreur relative? Commenter.

Réponse : il vient :

$$\begin{aligned} y &= (100 \oplus 0.6) \ominus 100 = x \ominus 100 \\ &= 1.0 \times 10^2 - 1.0 \times 10^2 \\ &= 0. \end{aligned}$$

Comme le calcul exact donne $\tilde{y} = 0.6$, l'erreur relative vaut $\frac{|y-\tilde{y}|}{|\tilde{y}|} = \frac{0.6-0}{0.6} = 100\%$. Il y a eu une erreur catastrophique lors du calcul de x (toute l'information sur 0.6 a été perdue lors de cette addition de 2 termes d'ordre de grandeur très différents), qui a été révélée lors du calcul de y (qui est complètement faux.) □

MT09-A2015- Examen médian
Durée : 1h30.
Polycopiés de cours et scilab autorisés - pas d'outils numériques

Questions de cours déjà traitées : environ *5 points.*

Exercice 1 : *(barème approximatif : 8 points)* **CHANGEZ DE COPIE**

La question 6 ne dépend que de la question 3. Les questions 4 et 5 sont largement indépendantes des autres.

Soit A une matrice inversible appartenant à $\mathcal{M}_{nn}(\mathbb{R})$ ($n \geq 1$). On désire effectuer la factorisation $A = UL$ de A (au lieu de LU), où

- U est une matrice triangulaire supérieure avec des 1 sur la diagonale,
- et L est une matrice triangulaire inférieure avec des termes diagonaux non-nuls.

On suppose que cette factorisation existe.

1. Montrer que
$$A_{ij} = \sum_{\alpha=r}^{n} U_{i\alpha} L_{\alpha j}, \quad \forall i, j = 1, \ldots, n,$$
où r est un entier dépendant de i et j à déterminer.

Réponse : $U_{ij} = 0$ si $i > j$ et $L_{ij} = 0$ si $j > i$. Donc si $i \leq \alpha$ alors $U_{i\alpha} \neq 0$ *a priori*, et si $j \leq \alpha$ alors $L_{\alpha j} \neq 0$ *a priori*. Donc pour $i, j = 1, \ldots, n$, il vient :

$$A_{ij} = \sum_{\alpha=1}^{n} U_{i\alpha} L_{\alpha j} = \sum_{\alpha=\max\{i,j\}}^{n} U_{i\alpha} L_{\alpha j}.$$

□

2. (a) Calculer $\underline{L}_n$ la dernière ligne de L.

Réponse : Pour $j = 1, \ldots, n$, $A_{nj} = \sum_{\alpha=\max\{n,j\}}^{n} U_{n\alpha} L_{\alpha j} = U_{nn} L_{nj} = L_{nj}$ car $U_{nn} = 1$. Donc finalement :

$$L_{nj} = A_{nj} \quad \forall j = 1, \ldots, n, \tag{2}$$

soit en vecteur ligne $\underline{L}_n = \underline{A}_n$. □

(b) Calculer U_n la dernière colonne de U. À quelle condition ce calcul est-il faisable?

Réponse : Pour $i = 2, \ldots, n$, $A_{in} = \sum_{\alpha=\max\{i,n\}}^{n} U_{i\alpha} L_{\alpha n} = U_{in} L_{nn}$, d'où on déduit :

$$U_{in} = \frac{A_{in}}{L_{nn}} \quad \forall i = 2, \ldots, n, \quad \textbf{si } L_{nn} \neq 0, \tag{3}$$

soit en vecteur colonne $U_n = \dfrac{1}{L_{nn}} A_n$. On remarque que l'on a bien $U_{nn} = \frac{A_{nn}}{L_{nn}} = 1$. Ce calcul est faisable si le "pivot" $L_{nn} \neq 0$. □

3. Soit $k \in \{1, \ldots, n-1\}$. On suppose qu'on connaît $\underline{L}_{k+1}, \ldots, \underline{L}_n$ et $U_{k+1}, \ldots, U_n$. En utilisant le produit matriciel, montrer, en le faisant, que l'on peut calculer $\underline{L}_k$ et U_k. (On pourra chercher à exprimer les termes de la ligne k de A puis de la colonne k de A).

À quelle condition ce calcul est-il faisable?

Réponse : on fait une récurrence sur k. Commençons par la ligne k de A pour déterminer $\underline{L}_k$. Pour $j = 1, \ldots, k$ (L est triangulaire inférieure),

$$\begin{aligned}
A_{kj} &= \sum_{\alpha=\max\{k,j\}}^{n} U_{k\alpha} L_{\alpha j} \quad \textbf{où } \max\{k,j\} = k \\
&= U_{kk} L_{kj} + \sum_{\alpha=k+1}^{n} U_{k\alpha} L_{\alpha j} \quad \textbf{où } U_{k\alpha} \in U_\alpha \textbf{ et } L_{\alpha j} \in \underline{L}_\alpha \textbf{ sont connues car } \alpha \geq k+1 \\
&= L_{kj} + \sum_{\alpha=k+1}^{n} U_{k\alpha} L_{\alpha j} \quad \textbf{car } U_{kk} = 1,
\end{aligned}$$

d'où on déduit

$$L_{kj} = A_{kj} - \sum_{\alpha=k+1}^{n} U_{k\alpha} L_{\alpha j} \quad \forall j = 1, \ldots, k. \tag{4}$$

Conseil : faites des dessins pour comprendre et expliquer!

Calculons la colonne k de A pour déterminer U_k. Pour $i = 1, \ldots, k-1$ (U est triangulaire supérieure et $U_{kk} = 1$ est connu),

$$\begin{aligned} A_{ik} &= \sum_{\alpha=\max\{i,k\}}^{n} U_{i\alpha} L_{\alpha k} \quad \textbf{où } \max\{i,k\} = k \\ &= U_{ik} L_{kk} + \sum_{\alpha=k+1}^{n} U_{i\alpha} L_{\alpha k} \quad \textbf{où } U_{i\alpha} \in U_\alpha \textbf{ et } L_{\alpha k} \in \underline{L}_\alpha \textbf{ sont connues car } \alpha \geq k+1 \end{aligned}$$

d'où on déduit

$$U_{ik} = \frac{1}{L_{kk}} \left(A_{ik} - \sum_{\alpha=k+1}^{n} U_{i\alpha} L_{\alpha k} \right) \quad \forall i = 1, \ldots, k-1, \quad \textbf{si } L_{kk} \neq 0. \tag{5}$$

On remarque que pour $i = k$, l'équation (5) est équivalente à $U_{kk} = 1$ d'après (4) (pour $j = k$).

Finalement, les équation (4) et (5) permettent de déterminer entièrement $\underline{L}_k$ et U_k.

La condition pour faire ce calcul est que le "pivot" $L_{kk} \neq 0$ pour $k = 2, \ldots, n$ ($k = 1$ n'est pas nécessaire ici).

Remarque : les pivots de la factorisation $A = UL$ se retrouvent sur la diagonale de L. Ils sont tous utilisés pour calculer U_k (cf. (5)) sauf pour $k = 1$ (première colonne), car $U_{11} = 1$ est déjà connu. □

4. Montrer que si la factorisation $A = UL$ existe, alors elle est unique.

   Réponse : Supposons qu'il existe deux factorisations $A = UL = \tilde{U}\tilde{L}$. Rappelons qu'une matrice triangulaire supérieure M avec des termes diagonaux M_{ii} non-nuls est inversible, et que son inverse M^{-1} est une matrice triangulaire supérieure vérifiant $(M^{-1})_{ii} = 1/M_{ii}$. De plus le produit de deux matrices triangulaires supérieures M et N est une matrice triangulaire supérieure dont les termes diagonaux valent $(MN)_{ii} = M_{ii} N_{ii}$.

   Idem pour les matrices triangulaires inférieures.

   Donc U et $\tilde{L}$ sont inversibles et il vient :

$$UL = \tilde{U}\tilde{L} \iff L\tilde{L}^{-1} = U^{-1}\tilde{U},$$

   où $L\tilde{L}^{-1}$ est triangulaire inférieure (comme produit de matrices triangulaires inférieures) et $U^{-1}\tilde{U}$ est triangulaire supérieure (comme produit de deux matrices triangulaires supérieures) et dont les termes diagonaux valent 1 (car les termes diagonaux de U^{-1} et de $\tilde{U}$ valent 1).

   Donc finalement $U^{-1}\tilde{U}$ est triangulaire supérieure et triangulaire inférieure, donc est diagonale, avec 1 sur la diagonale : c'est I. Il vient :

$$L\tilde{L}^{-1} = U^{-1}\tilde{U} = I \iff L = \tilde{L} \textbf{ et } U = \tilde{U},$$

   il y a donc unicité de la décomposition. □

5. Soit les matrices par blocs :

$$M = \begin{bmatrix} M_{11} & M_{12} \\ 0 & M_{22} \end{bmatrix}, \quad N = \begin{bmatrix} N_{11} & 0 \\ N_{21} & N_{22} \end{bmatrix}.$$

   On suppose que $M_{11} \in \mathcal{M}_{r_1,r_1}(\mathbb{R})$ et $M_{22} \in \mathcal{M}_{r_2,r_2}(\mathbb{R})$ et que $N_{11} \in \mathcal{M}_{s_1,s_1}(\mathbb{R})$ et $N_{22} \in \mathcal{M}_{s_2,s_2}(\mathbb{R})$.

   (a) Donner les conditions sur r_1, r_2, s_1 et s_2 pour que le produit par bloc ait un sens. Donner également les tailles de M_{12} et de N_{21}.

   Réponse : il faut que le produit matriciel de M_{ik} par N_{kj} soit possible, donc $r_1 = s_1$ et $r_2 = s_2$. Par ailleurs $M_{12} \in \mathcal{M}_{r_1,r_2}(\mathbb{R})$ et $N_{21} \in \mathcal{M}_{s_2,s_1}(\mathbb{R})$. □

   (b) Calculer le produit MN par blocs.

   Réponse :

$$MN = \begin{bmatrix} M_{11}N_{11} + M_{12}N_{21} & M_{12}N_{22} \\ M_{22}N_{21} & M_{22}N_{22} \end{bmatrix}. \tag{6}$$

□

(c) Pour $k = 0, \dots, n-1$, on appelle $[A]_{n-k}$ les sous matrices de A constituées des A_{ij} tels que $i = n-k, \dots, n$ et $j = n-k, \dots, n$. Quelle est la taille de $[A]_{n-k}$? Que valent $[A]_{n-k}$ pour $k = 0$ et $k = n-1$?

Réponse : $[A]_{n-k}$ est la partie inférieure droite de A :

$$[A]_{n-k} = \begin{bmatrix} A_{n-k,n-k} & A_{n-k,n-k+1} & \cdots & A_{n-k,n} \\ \vdots & \vdots & & \vdots \\ A_{n,n-k} & A_{n,n-k+1} & \cdots & A_{n,n} \end{bmatrix} \in \mathcal{M}_{k+1,k+1}(\mathbb{R}),$$

et donc pour $k = 0$ et $k = n-1$:

$$[A]_n = [A_{nn}] \in \mathcal{M}_{1,1}(\mathbb{R}), \quad [A]_1 = A \in \mathcal{M}_{n,n}(\mathbb{R}).$$

□

(d) À partir d'une décomposition par bloc des matrices U et L, écrire l'expression de $[A]_{n-k}$.

Réponse : pour $k = 0, \dots, n-1$, on décompose $A = UL$ en blocs de taille $r_1 = n-k-1$ et $r_2 = k+1$ de façon que

$$A = UL \iff \begin{bmatrix} \widetilde{A}_{11} & \widetilde{A}_{12} \\ \widetilde{A}_{21} & \widetilde{A}_{22} \end{bmatrix} = \begin{bmatrix} \widetilde{U}_{11} & \widetilde{U}_{12} \\ 0 & \widetilde{U}_{22} \end{bmatrix} \begin{bmatrix} \tilde{L}_{11} & 0 \\ \tilde{L}_{21} & \tilde{L}_{22} \end{bmatrix} \quad \textbf{où } \widetilde{A}_{ij}, \widetilde{U}_{ij}, \tilde{L}_{ij} \in \mathcal{M}_{r_1,r_2}$$

$$\iff \begin{bmatrix} \widetilde{A}_{11} & \widetilde{A}_{12} \\ \widetilde{A}_{21} & [A]_{n-k} \end{bmatrix} = \begin{bmatrix} \widetilde{U}_{11} & \widetilde{U}_{12} \\ 0 & [U]_{n-k} \end{bmatrix} \begin{bmatrix} \tilde{L}_{11} & 0 \\ \tilde{L}_{21} & [L]_{n-k} \end{bmatrix},$$

ce qui implique d'après le calcul par blocs (cf. (6)) que

$$[A]_{n-k} = [U]_{n-k}[L]_{n-k}.$$

□

(e) Déduire de la question précédente que si la factorisation $A = UL$ est faisable, alors $[A]_{n-k}$ est inversible, pour tout $k = 0, \dots, n-1$.

Réponse : si la factorisation $A = UL$ est faisable, alors pour tout $k = 0, \dots, n-1$, on peut effectuer le produit par bloc ci-dessus et donc $[A]_{n-k} = [U]_{n-k}[L]_{n-k}$. On obtient :

$$\det([A]_{n-k}) = \det([U]_{n-k}) \det([L]_{n-k}) = \det([L]_{n-k}) = \prod_{i=n-k}^{n} L_{ii}$$

car $[U]_{n-k}$ est triangulaire supérieure avec des 1 sur la diagonale et $[L]_{n-k}$ est triangulaire inférieure. Comme les L_{ii} doivent être non-nuls pour effectuer la factorisation $A = UL$, le déterminant de $[A]_{n-k}$ doit être non-nul et donc la matrice $[A]_{n-k}$ est inversible pour tout $k = 0, \dots, n-1$.

Remarque : pour $k = n-1$ (première colonne), on a vu que la condition $L_{11} \neq 0$ n'était pas strictement nécessaire (cf. remarque et (5)). Toutefois comme $\det(A) = \det([A]_1) = \det([U]_1)\det([L]_1) = \det(U)\det(L) = \prod_{i=1}^{n} L_{ii}$, le fait que A soit inversible (par hypothèse) implique que $L_{ii} \neq 0$ pour tout $i = 1, \dots, n$ et en particulier pour L_{11}. □

6. Écrire une fonction scilab : `function [U, L] = UL(A)` qui effectue la factorisation $A = UL$. (Un bonus sera accordé pour une version vectorielle du programme).

Réponse : l'algorithme est celui de Doolittle inversé (en partant du bas). Il est donné par les équations (4) et (5). Voici un exemple d'implémentation (utilisant le produit matrice vecteur) : □

```
function [U, L] = UL(A)
tol = 1e-12
n=size(A,1)
if ( size(A) ~= [n, n] )
  printf( "size(A)= %d, %d", size(A,1), size(A,2) );
  error('not a correct size')
end

L = zeros(n, n); U = eye(n, n); // initialisation
for k = n:-1:1 // boucle partant du bas
  // modification de la colonne Lk (utilise un bloc de L (bas a gauche))
```

```
  L(k, 1:k) = A(k, 1:k) - U(k, k+1:n) * L(k+1:n, 1:k);
  pivot = L(k,k);
  if ( abs( pivot ) < tol )
    if k > 1
      disp( pivot, [k,k]); error('Pivot nul');
    else // on ne retourne pas d'erreur si c'est le dernier pivot qui est nul
      warning("Dernier pivot nul, matrice de rang n-1"); return;
    end
  end
  // modification de la ligne Uk (rien n'est fait pour k=1)
  // (utilise un bloc de U (haut a droite))
  U(1:k-1, k) = ( A(1:k-1, k) - U(1:k-1, k+1:n) * L(k+1:n, k) )  / pivot;
end
endfunction
```

Il est possible de remplacer la modification de la colonne L_k par une boucle :

```
  for j = 1:k
    L(k, j) = A(k, j) - U(k, k+1:n) * L(k+1:n, j);
  end
```

ou même deux boucles :

```
  for j = 1:k
    s = 0
    for alpha = k+1:n
      s = s + U(k, alpha) * L(alpha, j);
    end
    L(k, j) = A(k, j) - s;
  end
```

Idem pour la modification de $\underline{U}_k$. Ces deux solutions sont acceptables, mais moins performantes sous scilab.

7. Écrire une fonction scilab : `function [x] = resol(A, b)` qui résout le système $Ax = b$ avec la factorisation UL. On suppose que l'on dispose des fonctions du TP2, qui ne seront donc pas réécrites ici.

   Réponse : en utilisant la factorisation, il vient :

$$Ax = b \Longleftrightarrow U(Lx) = b \Longleftrightarrow \begin{cases} Uy = b \\ Lx = y \end{cases}$$

□

```
function [x] = resol(A,b)
exec("UL.sci", -1); exec("solsup.sci", -1); exec("solinf.sci", -1);

[U, L] = UL(A);
y = solsup(U, b);
x = solinf(L, y);
endfunction
```

Exercice 2 : *(barème approximatif : 4 points)* **CHANGEZ DE COPIE**

Soit une matrice M appartenant à $\mathcal{M}_{nn}(\mathbb{R})$ ($n \geq 1$). Soit $|||\cdot|||$ une norme matricielle subordonnée à la norme vectorielle $\|\cdot\|$.

1. On suppose que $|||M||| < 1$.

   (a) Déterminer le noyau de $I+M$ et en déduire que $I+M$ est inversible.
   Réponse : soit $x \in \mathrm{Ker}(I+M)$, donc $(I+M)x = 0$ donc $-x = Mx$, ce qui implique que $\|-x\| = \|x\| = \|Mx\| \leq |||M|||\ \|x\|$ d'après les propriétés des normes subordonnées. Si $\|x\| \neq 0$, alors $|||M|||\ \|x\| < \|x\|$ ce qui est impossible, donc $\|x\| = 0 \Longleftrightarrow x = 0$.
   On conclut que $\mathrm{Ker}(I+M) = \{0\}$ et donc $(I+M)$ est injective donc bijective (car carrée). $\square$

   (b) On pose $N = (I+M)^{-1}$. En partant de la définition de l'inverse d'une matrice, montrer que

   $$|||N||| \leq \frac{1}{1 - |||M|||}.$$

   Réponse : la matrice N existe d'après la question précédente. On a $N(I+M) = I$, donc $N = I - NM$, ce qui implique que $|||N||| \leq |||I||| + |||NM|||$ d'après l'inégalité triangulaire. Pour une norme subordonnée $|||I||| = \sup_{x\neq 0} \frac{\|Ix\|}{\|x\|} = \sup_{x\neq 0} \frac{\|x\|}{\|x\|} = 1$ et $|||NM||| \leq |||N|||\ |||M|||$. Il vient donc $|||N||| \leq 1 + |||N|||\ |||M|||$, d'où $|||N|||(1 - |||M|||) \leq 1$. Comme $|||M||| < 1$, on peut diviser par $1 - |||M||| > 0$ et on obtient l'inégalité demandée. $\square$

2. Soit A dans $\mathcal{M}_{nn}(\mathbb{R})$ une matrice inversible et b dans $\mathbb{R}^n$. Nous appelons $x \in \mathbb{R}^n$ la solution du système $Ax = b$. Nous considérons le système perturbé $(A + \delta A)\hat{x} = b$, où $\delta A \in \mathcal{M}_{nn}(\mathbb{R})$ et $\hat{x} \in \mathbb{R}^n$. On pose $\delta x = \hat{x} - x$.

   (a) Quelle équation satisfait δx en fonction de A, δA et x?
   Réponse : en développant $(A+\delta A)\hat{x} = b$, et en utilisant le fait que $Ax = b$, on obtient $\delta A\, x + (A + \delta A)\delta x = 0 \Leftrightarrow (A+\delta A)\delta x = -\delta A\, x$, ce qui s'écrit encore avec A inversible

   $$(I + A^{-1}\delta A)\delta x = -A^{-1}\delta A\, x. \tag{7}$$

   $\square$

   (b) On suppose que $|||A^{-1}\delta A||| < 1$. Déduire des questions précédentes que

   $$\|\delta x\| \leq \frac{|||A^{-1}|||}{1 - |||A^{-1}\delta A|||}\ |||\delta A|||\ \|x\|.$$

   Réponse : comme $|||A^{-1}\delta A||| < 1$, d'après la question 1., la matrice $(I + A^{-1}\delta A)$ est inversible et son inverse vérifie :

   $$|||(I + A^{-1}\delta A)^{-1}||| \leq \frac{1}{1 - |||A^{-1}\delta A|||}.$$

   De l'équation (7), on tire donc $\delta x = -(I + A^{-1}\delta A)^{-1} A^{-1}\delta A\, x$ et donc

   $$\begin{aligned} \|\delta x\| &= \|(I + A^{-1}\delta A)^{-1} A^{-1}\delta A\, x\| \leq |||(I + A^{-1}\delta A)^{-1}|||\ |||A^{-1}|||\ |||\delta A|||\ \|x\| \\ &\leq \frac{1}{1 - |||A^{-1}\delta A|||}\ |||A^{-1}|||\ |||\delta A|||\ \|x\|. \end{aligned}$$

   $\square$

   (c) Écrire une majoration de l'erreur relative $\dfrac{\|\delta x\|}{\|x\|}$ en fonction du conditionnement de A et de l'erreur relative sur les données $\dfrac{|||\delta A|||}{|||A|||}$ uniquement.
   Réponse : pour $x \neq 0$, en utilisant le conditionnement de A noté $\chi(A) = |||A|||\ |||A^{-1}|||$, on obtient :

   $$\begin{aligned} \frac{\|\delta x\|}{\|x\|} &\leq \frac{1}{1 - |||A^{-1}\delta A|||}\ \chi(A)\ \frac{|||\delta A|||}{|||A|||} \\ &\leq \frac{1}{1 - |||A^{-1}|||\ |||\delta A|||}\ \chi(A)\ \frac{|||\delta A|||}{|||A|||} \qquad \textbf{si } |||A^{-1}|||\ |||\delta A||| < 1 \\ &\leq \frac{\chi(A)}{1 - \chi(A)\ \frac{|||\delta A|||}{|||A|||}}\ \frac{|||\delta A|||}{|||A|||}. \end{aligned}$$

   $\square$

Exercice 3 : *(barème approximatif : 4 points)* **CHANGEZ DE COPIE**

Soit T une matrice inversible appartenant à $\mathcal{M}_{2n,2n}(\mathbb{R})$ et H un vecteur colonne de $\mathbb{R}^{2n}$ ($n > 0$), on cherche V, vecteur colonne de $\mathbb{R}^{2n}$ vérifiant $TV = H$.

On décompose T, H et V en blocs de la façon suivante :

$$T = \begin{pmatrix} M & N \\ 0 & M \end{pmatrix}, \; V = \begin{pmatrix} X \\ Y \end{pmatrix}, \; H = \begin{pmatrix} E \\ F \end{pmatrix},$$

où $M, N \in \mathcal{M}_{nn}(\mathbb{R})$, 0 matrice nulle $\in \mathcal{M}_{nn}(\mathbb{R})$, E, F, X, Y vecteurs colonnes de $\mathbb{R}^n$.

On démontre, et nous l'admettrons, que

$$TV = H \Leftrightarrow \begin{cases} MX + NY = E \\ MY = F \end{cases} . \tag{8}$$

On suppose que l'on dispose des fonctions Scilab suivantes :

- `function [L, U] = LU(A)`
  Étant donnée la matrice A, cette fonction calcule sa factorisation LU : $A = LU$.
- `function [x] = resolLU(A, b)`
  Étant donnés la matrice A et le vecteur b, cette fonction calcule la factorisation LU de A puis résout $Ax = b$.
- `function [B] = inverse(A)`
  Étant donnée la matrice A, cette fonction calcule l'inverse $B = A^{-1}$ de A.
- `function [x] = solinf (L,b)`
  Étant donnés la matrice triangulaire inférieure L et le vecteur colonne b, cette fonction résout $Lx = b$.
- `function [x] = solsup (U,b)`
  Étant donnés la matrice triangulaire supérieure U et le vecteur colonne b, cette fonction résout $Ux = b$.

1. Écrire une fonction Scilab

   - `function [V] = sol(M, N, H)` qui étant donnés les matrices M et N et le vecteur H, calcule le vecteur V tel que $TV = H$, en utilisant la relation (8).

   Réponse : il ne faut surtout pas utiliser la fonction `inverse` ! Et il faut éviter de factoriser plusieurs fois la matrice M, car c'est l'opération la plus coûteuse : la fonction `resolLU` est également à proscrire !

   Voici une implémentation possible : □

```
function [V] = sol(M, N, H)
// resout : (M N; 0 M) V = H
exec("solsup.sci", -1); exec("solinf.sci", -1); exec("LU.sci", -1);
n = size(M,1);
[L, U] = LU(M); // factorisation
z = zeros(n, 1);
E = H(1:n); F = H(n+1:2*n); // extrait les 2 blocs de H
z = solinf(L, F);
y = solsup(U, z);
v = E - N*x; // calcul du second membre pour x
z = solinf(L, v);
x = solsup(U, z);
V = [x; y];
endfunction
```

2. Donner le nombre de multiplications nécessaires pour la résolution de ce système en utilisant (8).

   Quel serait le nombre de multiplications nécessaires pour résoudre directement $TV = H$? Expliquer.

   Réponse : d'après le cours, pour une matrice de taille n, le coût de `solinf` et `solsup` est de $n^2/2$ multiplications. Le coût de la factorisation LU est de $n^3/3$ multiplications. Le coût d'un produit matrice vecteur est de n^2 multiplications. Il y a 1 `LU`, 2 `solsup`, 2 `solinf` et un produit Nx, donc le coût de la fonction `sol` ci-dessus est de $n^3/3+3n^2 \approx n^3/3$ multiplications (les termes en n^2 sont négligeables devant $n^3/3$), donc essentiellement le coût de la factorisation de M.

   S'il avait fallu factoriser T directement qui est de taille $2n$, le coût aurait été de $(2n)^3/3 = 8n^3/3$ multiplications, soit 8 fois plus cher... □