MT09-A2016 – 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)*

Soit A une matrice carrée de $\mathcal{M}_{nn}(\mathbb{R})$.

1. Montrer que si A admet une décomposition de Cholesky, alors A est symétrique définie positive.

   Réponse : cf. cours. □

2. La réciproque est-elle vraie? On ne demande pas de montrer le résultat.

   Réponse : Oui. □

3. Soit $\alpha \in \mathbb{R}$ et A la matrice définie par $A = \begin{bmatrix} 4 & 3 \\ 3 & \alpha \end{bmatrix}$. Faites le calcul de la décomposition de Cholesky de A. En déduire une condition nécessaire et suffisante sur α pour que A soit symétrique définie positive.

   Réponse : En écrivant $A = CC^T$ avec C triangulaire inférieure et $c_{ii} > 0$, on obtient

$$\left\{ \begin{array}{rcl} c_{11}^2 & = & 4 \\ c_{21}c_{11} & = & 3 \\ c_{21}^2 + c_{22}^2 & = & \alpha \end{array} \right. \iff \left\{ \begin{array}{rcl} c_{11} & = & 2 \\ c_{21} & = & \frac{3}{2} \\ c_{22} & = & \sqrt{\alpha - \frac{9}{4}}, \quad \textbf{ssi } \alpha > \frac{9}{4}. \end{array} \right.$$

   Donc A SDP $\iff$ la décomposition de Cholesky de A est faisable $\iff \alpha > \frac{9}{4}$. □

Exercice 2 *(barème approximatif : 1.5 points)*

On se place sur l'espace $\mathcal{M}_{nn}(\mathbb{R})$ des matrices carrées de taille $n > 1$. Soit $|||\cdot|||$ une norme matricielle subordonnée à la norme vectorielle $\|\cdot\|$.

1. Pour une matrice A, donner la définition de la norme subordonnée $|||A|||$ et du conditionnement $\chi(A)$.

   Réponse : cf. cours. □

2. Donner les propriétés de norme matricielle que vérifie $|||\cdot|||$.

   Réponse : cf. cours. □

3. Prouver l'inégalité triangulaire pour $|||\cdot|||$.

   Réponse : soit $x \in \mathbb{R}^n$, $x \neq 0$. Il vient

$$\begin{array}{rcl} \|(A+B)x\| & = & \|Ax + Bx\| \\ & \leq & \|Ax\| + \|Bx\| \quad \textbf{(inégalité triangulaire pour la norme vectorielle)} \\ & \leq & |||A|||\ \|x\| + |||B|||\ \|x\| \quad \textbf{(propriété de la norme matricielle subordonnée)}. \end{array}$$

   Donc, comme $x \neq 0$,

$$\frac{\|(A+B)x\|}{\|x\|} \leq |||A||| + |||B|||,$$

   le terme de gauche est un majorant indépendant de x, donc le $\sup$ qui est le plus petit des majorants reste plus petit que ce majorant. Il vient

$$|||A+B||| = \sup_{x\neq 0} \frac{\|(A+B)x\|}{\|x\|} \leq |||A||| + |||B|||.$$

□

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 $\mathcal{F}_2$ (en base 2). On expliquera ce que signifie les constantes t, L et U (notations du cours).

   Réponse : cf. cours.

   $$\mathcal{F}_2 = \left\{ \pm 0.d_1d_2\dots d_t \; 2^e \mid d_i \in \{0,1\} \; \forall i = 2,\dots,t, \quad d_1 = 1, \quad L \leq e \leq U \right\} \cup \{0\},$$

   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 $\mathcal{F}_2$ car 0 ne s'écrit pas comme un nombre flottant normal. □

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

   (a) Écrire tous les flottants compris dans $[1/2, 1]$.
   Réponse : sur $[1/2, 1[$, $e = 0$ et pour 1, $e = 1$. Les flottants valent :

   $$(0.100)_2 = \frac{1}{2},\ (0.101)_2 = \frac{5}{8},\ (0.110)_2 = \frac{3}{4},\ (0.111)_2 = \frac{7}{8},\ (1.00)_2 = (0.100)_2 \times 2^1 = 1.$$

   □

   (b) Calculer le flottant : $\tilde{x} = 1 \oplus \frac{5}{8}$.
   Réponse : cf. cours. On a :

   $$\tilde{x} = \mathrm{fl}(1.00 + 0.101) = \mathrm{fl}(1.101) = 1.10 \textbf{ ou } 1.11, \textbf{ selon les règles d'arrondi.}$$

   Donc $\tilde{x} = \frac{3}{2}$ ou $\frac{7}{4}$. Les deux valeurs sont possibles, mais une seule valeur est choisie par le système, selon la règle choisie pour les arrondis. Nous prenons par exemple $\tilde{x} = \frac{3}{2}$. □

   (c) Calculer l'erreur relative entre $\tilde{x}$ et $x = 1 + \frac{5}{8}$ On rappelle que $\varepsilon_{\mathrm{mach}} = 2^{-t}$: commenter le résultat.
   Réponse : on note e l'erreur relative définie par

   $$e = \frac{|\tilde{x} - x|}{|x|} = \frac{\frac{13}{8} - \frac{3}{2}}{\frac{13}{8}} = \frac{1}{13},$$

   qui est bien $\leq \varepsilon_{\mathbf{mach}} = 2^{-3} = \frac{1}{8}$, conformément au cours. Note : pour $\tilde{x} = \frac{7}{4}$, on obtiendrait

   $$e = \frac{|\frac{13}{8} - \frac{7}{4}|}{\frac{13}{8}} = \frac{1}{13} \leq \varepsilon_{\mathbf{mach}}.$$

   □

MT09-A2016- Examen médian
Durée : 1h30.
Polycopiés de cours et scilab autorisés - pas d'outils numériques

Questions de cours déjà traitées : environ *6 points.*

Exercice 1 : *(barème approximatif : 6 points)* **CHANGEZ DE COPIE**

Il est possible de traiter une question en admettant les résultats précédents.

1. Soit $C \in \mathcal{M}_{nn}(\mathbb{R})$ (avec $n \geq 1$). Soit λ une valeur propre de C.

   (a) Montrer que $C^T - \lambda I$ n'est pas inversible.
   Réponse : comme λ est une valeur propre de C,

   $$\begin{aligned} \det(C-\lambda I)=0 &\iff \det((C-\lambda I)^T)=0 \quad \textbf{car } \det(A^T)=\det(A) \\ &\iff \det(C^T-\lambda I)=0, \end{aligned}$$

   et donc $C^T - \lambda I$ n'est pas inversible (λ est valeur propre pour C^T). □

   (b) On rappelle qu'une matrice à diagonale strictement dominante est inversible. En utilisant ce qui précède, montrer que :

   $$\exists i \in \{1,2,\ldots,n\}, \text{ tel que : } |c_{ii}-\lambda| \leq \sum_{j=1}^{i-1} |c_{ji}| + \sum_{j=i+1}^{n} |c_{ji}|.$$

   Réponse : on rappelle que M à diagonale strictement dominante s'écrit :

   $$\forall i \in \{1,\ldots,n\}, \ |M_{ii}| > \sum_{j\neq i} |M_{ij}|. \tag{1}$$

   Comme $C^T - \lambda I$ n'est pas inversible, elle n'est pas à diagonale strictement dominante. On écrit le contraire logique de (1) pour $C^T - \lambda I$:

   $$\begin{aligned} \exists i \in \{1,\ldots,n\}, \ |(C^T-\lambda I)_{ii}| \leq \sum_{j\neq i} |(C^T-\lambda I)_{ij}| &\iff \exists i \in \{1,\ldots,n\}, \ |C_{ii}-\lambda| \leq \sum_{j\neq i} |(C^T)_{ij}| \\ &\iff \exists i \in \{1,\ldots,n\}, \ |C_{ii}-\lambda| \leq \sum_{j\neq i} |C_{ji}|. \end{aligned}$$

   □

2. Soit $A \in \mathcal{M}_{nn}(\mathbb{R})$. Dans toute la suite de l'exercice, on suppose que A^T est à diagonale strictement dominante.

   On décompose A conformément aux notations du cours en $A = D - E - F$.

   (a) Montrer que D est inversible.
   Réponse : D est une matrice diagonale, telle que $D_{ii} = A_{ii}$. Comme A^T est à diagonale strictement dominante, on a (attention à la transposition) :

   $$\forall i \in \{1,\ldots,n\}, \ |A_{ii}| > \sum_{j\neq i} |A_{ji}|. \tag{2}$$

   Ceci implique que $|D_{ii}| = |A_{ii}| > 0 \ \forall i$, donc la matrice diagonale D est inversible. □

   (b) Dans toute la suite de l'exercice, on définit $C = (E+F)D^{-1}$. Exprimer les coefficients de C à l'aide des coefficients de A.
   Réponse : D est inversible donc la matrice C est bien définie. Les règles du produit matriciel (cf. exercice TD 10, chap 0) impliquent pour la colonne C_j de C (on note e_j le $j$ème vecteur de la base canonique) :

   $$C_j = (E+F)(D^{-1})_j = (E+F)(\frac{1}{A_{ii}}e_j) = \frac{1}{A_{ii}}(E+F)_j = \frac{1}{A_{ii}}(E_j+F_j), \quad \forall j=1,\ldots,n.$$

(On a déjà remarqué que A_{ii} est non nul). Cela s'écrit encore (revoir les définitions de E et F) :

$$C_{ij} = \begin{cases} 0 & \textbf{si } i = j \\ -\dfrac{A_{ij}}{A_{ii}} & \textbf{si } i \neq j \end{cases}.$$

□

3. Utiliser les questions 1b) et 2b) pour montrer que si λ est valeur propre de C, alors $|\lambda| < 1$.

   Réponse : soit λ une valeur propre de C. Cela implique que $C^T - \lambda I$ n'est pas inversible et donc d'après la questions 1b),

$$\exists i \in \{1, \dots, n\},\ |C_{ii} - \lambda| \leq \sum_{j \neq i} |C_{ji}|.$$

   ce qui s'écrit en utilisant la question 2b) : il existe $i \in \{1, \dots, n\}$ tel que :

$$|-\lambda| \leq \sum_{j \neq i} \left| -\frac{A_{ji}}{A_{ii}} \right| \quad \Longrightarrow \quad |\lambda| \leq \frac{1}{|A_{ii}|} \sum_{j \neq i} |A_{ji}| < 1,$$

   d'après (2) (A^T est à diagonale strictement dominante). Cela implique que $|\lambda| < 1$ pour toute valeur propre de C et donc $\rho(C) < 1$. □

4. Soit M_1 et M_2 deux matrices appartenant à $\mathcal{M}_{nn}(\mathbb{R})$. On suppose que M_2 est inversible.

   (a) Montrer que $M_2^{-1}M_1$ et $M_1M_2^{-1}$ ont les mêmes valeurs propres.

   Réponse : soit (λ, y) un couple propre de $M_2^{-1}M_1$ ($y \neq 0$). On a $M_2^{-1}M_1y = \lambda y$. On pose $z = M_2y \iff y = M_2^{-1}z$. Il vient

$$\begin{aligned} M_2^{-1}M_1y = \lambda y \quad &\iff \quad M_2^{-1}M_1M_2^{-1}z = \lambda M_2^{-1}z \\ &\iff \quad M_1M_2^{-1}z = \lambda z, \end{aligned}$$

   en multipliant à gauche par M_2 inversible. Comme $y \neq 0$, z est lui aussi non-nul : c'est un vecteur propre pour $M_1M_2^{-1}$.
   Comme on a travaillé par équivalence, cela prouve que les valeurs propres de $M_2^{-1}M_1$ et de $M_1M_2^{-1}$ sont identiques.
   ***Autre preuve :* $M_2^{-1}M_1$ et $M_1M_2^{-1}$ sont semblables, car $M_2^{-1}M_1 = M_2^{-1}(M_1M_2^{-1})M_2$. Donc ces matrices ont le même polynôme caractéristique et donc les mêmes valeurs propres.**

   □

   (b) En déduire que $M_2^{-1}M_1$ et $M_1M_2^{-1}$ ont le même rayon spectral.

   Réponse : comme $\rho(A) = \max\limits_{\lambda \text{ v.p. de } A} |\lambda|$, le fait que les valeurs propres de $M_2^{-1}M_1$ et $M_1M_2^{-1}$ soient identiques implique en particulier que leur plus grande valeur propre en module est identique. Donc

$$\rho(M_2^{-1}M_1) = \rho(M_1M_2^{-1}).$$

   □

5. Utiliser les questions 3) et 4b) pour montrer que la méthode de Jacobi utilisée pour résoudre $Ax = b$ est convergente lorsque A^T est à diagonale strictement dominante.

   Réponse : d'après le cours, $J = D^{-1}(E + F)$ est la matrice de l'itération de Jacobi. La question 4b) implique que

$$\rho(D^{-1}(E + F)) = \rho((E + F)D^{-1}) \quad \iff \quad \rho(J) = \rho(C).$$

   D'après la question 3), puisque A^T est à diagonale strictement dominante, on sait que $\rho(C) < 1$. Donc finalement $\rho(J) < 1$, ce qui est une condition nécessaire et suffisante pour que la méthode de Jacobi converge. □

Exercice 2 : *(barème approximatif : 7 points)* CHANGEZ DE COPIE

Soit un entier $n \geq 1$. On rappelle que la factorisation LU d'une matrice de taille n nécessite de l'ordre de $\dfrac{n^3}{3}$ multiplications. Le coût de la résolution d'un système triangulaire est de l'ordre de $\dfrac{n^2}{2}$.

1. Soit M une matrice de $\mathcal{M}_{nn}$. Donner en le justifiant le coût en nombre de multiplications pour calculer M^{-1}.

Réponse : revoir le `TP1`. On pose $N = M^{-1}$. On a

$$MM^{-1} = I \iff MN = I \iff MN_j = I_j \ \forall j = 1, \ldots, n.$$

On doit résoudre n systèmes linéaires (1 par colonne de N) pour calculer M^{-1}. Il faut donc factoriser $M = LU$ (coût $\frac{n^3}{3}$), puis résoudre n systèmes linéaires triangulaires inférieurs ($LZ_j = I_j$, coût $n\frac{n^2}{2}$), et n systèmes linéaires triangulaires supérieurs ($UN_j = Z_j$, coût $n\frac{n^2}{2}$).

Le coût total est donc $\dfrac{4n^3}{3}$. □

2. Soit A une matrice inversible de $\mathcal{M}_{nn}$. On définit la matrice $K \in \mathcal{M}_{n^2n^2}$ par

$$K = \begin{bmatrix} A & -I & 0 & \cdots & \cdots & 0 \\ I & A & -I & 0 & \cdots & 0 \\ 0 & I & A & -I & 0 & \cdots \\ & & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & I & A & -I \\ 0 & \cdots & \cdots & 0 & I & A \end{bmatrix},$$

où $I \in \mathcal{M}_{nn}$ est la matrice identité et $0 \in \mathcal{M}_{nn}$ est la matrice nulle.

On veut résoudre $KX = B$, où $B \in \mathbb{R}^{n^2}$. Pour ce faire, on décompose X et B en blocs :

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix}, \quad \text{où } X_i \in \mathbb{R}^n, \text{ et } B_i \in \mathbb{R}^n, \forall i = 1, \ldots, n.$$

(a) On suppose que les matrices intervenant dans les calculs sont inversibles. On veut montrer que

$$KX = B \iff \left\{ \begin{array}{lcl} D_iX_i & = & D_{i-1}X_{i+1} + F_i, \quad i = 1, \ldots, n-1 \\ D_nX_n & = & F_n \end{array} \right., \text{ où } D_i \in \mathcal{M}_{nn}, F_i \in \mathbb{R}^n.$$

i. Calculer les matrices D_0, D_1 et le vecteur F_1.
Réponse : le système s'écrit par blocs (attention, on manipule des matrices et des vecteurs!) :

$$KX = B \iff \left\{ \begin{array}{rcl} AX_1 - X_2 & = & B_1 \\ X_{i-1} + AX_i - X_{i+1} & = & B_i, \quad i = 2, \ldots, n-1 \\ X_{n-1} + AX_n & = & B_n \end{array} \right. . \quad (3)$$

La première équation de (3) donne

$$D_1X_1 = D_0X_2 + F_1, \quad \textbf{en posant} \left\{ \begin{array}{lcl} D_1 & = & A \\ D_0 & = & I \\ F_1 & = & B_1 \end{array} \right. .$$

□

ii. Déterminer par récurrence D_i en fonction de A, D_{i-1} et D_{i-2}, pour $i = 2, \ldots, n$.
Déterminer également F_i en fonction de F_{i-1} et d'autres termes.
Réponse : la deuxième équation de (3) donne en multipliant à gauche par D_1 :

$$\begin{array}{rcl} X_1 + AX_2 - X_3 = B_2 & \Longrightarrow & D_1X_1 + D_1AX_2 = D_1X_3 + D_1B_2 \\ & \Longrightarrow & (D_0 + D_1A)X_2 = D_1X_3 + D_1B_2 - F_1 \\ & \Longrightarrow & D_2X_2 = D_1X_3 + F_2, \quad \textbf{en posant} \left\{ \begin{array}{lcl} D_2 & = & D_0 + D_1A \\ F_2 & = & D_1B_2 - F_1 \end{array} \right. . \end{array}$$

On montre par récurrence que $\forall i = \{2, \ldots, n\}$

$$D_iX_i = D_{i-1}X_{i+1} + F_i, \quad \textbf{avec} \left\{ \begin{array}{lcl} D_i & = & D_{i-2} + D_{i-1}A \\ F_i & = & D_{i-1}B_i - F_{i-1} \end{array} \right. .$$

C'est vrai pour $i = 2$, on passe de i à $i+1$. On écrit la ligne $i+1$:

$$\begin{array}{rcl} X_i + AX_{i+1} - X_{i+2} = B_{i+1} & \Longrightarrow & D_iX_i + D_iAX_{i+1} = D_iX_{i+2} + D_iB_{i+1} \\ & \Longrightarrow & (D_{i-1} + D_iA)X_{i+1} = D_iX_{i+2} + D_iB_{i+1} - F_i \\ & \Longrightarrow & D_{i+1}X_{i+1} = D_iX_{i+2} + F_{i+1}, \quad \textbf{en posant} \left\{ \begin{array}{lcl} D_{i+1} & = & D_{i-1} + D_iA \\ F_{i+1} & = & D_iB_{i+1} - F_i \end{array} \right. , \end{array}$$

ce qui achève la démonstration de la récurrence.
Il faut écrire l'équation pour $i = n$:

$$\begin{aligned} X_{n-1} + AX_n = B_n &\implies D_{n-1}X_{n-1} + D_{n-1}AX_n = D_{n-1}B_n \\ &\implies (D_{n-2} + D_{n-1}A)X_n = D_{n-1}B_n - F_{n-1} \\ &\implies D_nX_n = F_n, \quad \textbf{en posant} \left\{ \begin{array}{lcl} D_n & = & D_{n-2} + D_{n-1}A \\ F_n & = & D_{n-1}B_n - F_{n-1} \end{array} \right. , \end{aligned}$$

□

(b) On suppose que l'on dispose des fonctions du TP2 : `solsup, solinf, LU, inverse` (on ne demande pas de les réécrire ici).

Écrire une fonction scilab : `function [X] = resol(A, B)` qui, étant donnés la matrice `A` $\in \mathcal{M}_{nn}$ et le vecteur `B` $\in \mathbb{R}^{n^2}$ résout `KX = B` par la méthode décrite ci-dessus.

On sera attentif à la manipulation des vecteurs blocs et des matrices blocs.

Réponse : cet algorithme est une adaptation de celui de Richtmayer pour les matrices tridiagonales par blocs, qui exploite la forme particulière de K. Voici une implémentation possible. On a séparé la taille des blocs (n), du nombre de blocs ($nBloc$), ce qui permet de rendre l'algorithme plus clair.

```
===============================
function [XX] = richbloc(A, BB)
exec("LU.sci", 0);   exec("solinf.sci", 0);   exec("solsup.sci", 0);
n = size(A, 1);   N = length(BB);   nBloc = floor(N/n);

if nBloc ~= N/n then
      error('Not compatible sizes of matrix and vector');
end

DD = zeros(n, (nBloc+1)*n); // Stocke dans DD : blocs D_0 à D_nBloc
FF = zeros(N, 1); // Stocke dans FF : blocs F_1 à F_nBloc
Di = zeros(n, n);   Fi = zeros(n, 1); // tmp matrix and vector
X = zeros(N, 1);

DD(: , 1:n) = eye(n,n); // D_0 dans le bloc 1
DD(: , n+1:2*n) = A; // D_1 dans le bloc 2
FF(1:n) = BB(1:n); // F_1 dans le bloc 1

for ii = 2:nBloc
      indBlock_i = (ii)*n + 1 : (ii+1)*n
      indBlock_im1 = indBlock_i - n;
      indBlock_im2 = indBlock_i - 2*n;

      // décalage entre blocs des vecteurs et de la matrice,
      // car on commence à D_0 pour DD, alors qu'on commence à F_1 pour FF...
      indBlockVec_i = indBlock_im1;
      indBlockVec_im1 = indBlock_im2;

      // Calcul de D_i = D_{i-2} + D_{i-1} A :
      DD( : , indBlock_i ) = DD( : , indBlock_im2 ) + DD( : , indBlock_im1 ) * A;

      // Calcul de F_i = D_{i-1} B_i - F_{i-1} :
      FF( indBlockVec_i ) = DD( : , indBlock_im1 ) * BB( indBlockVec_i ) - FF( indBlockVec_im1 );
end

// Dernier bloc : résolution de D_n X_n = F_n
Di = DD( : , indBlock_i ); Fi = FF( indBlockVec_i );
[ L, U ] = LU( Di );   Z = solinf( L , Fi );   XX( indBlock_im1 ) = solsup( U, Z );

for ii = nBloc-1:-1:1
      indBlock_i = (ii)*n + 1 : (ii+1)*n;
      indBlock_im1 = indBlock_i - n;
      // décalage entre blocs des vecteurs et de la matrice
```

```
        indBlockVec_i = indBlock_im1;
        indBlockVec_ip1 = indBlock_i;

        // Résolution de D_i X_i = D_{i-1} X_{i+1} + F_i :
        Di = DD( : , indBlock_i );
        Fi = DD( : , indBlock_im1 ) * XX( indBlockVec_ip1 ) + FF( indBlockVec_i );
        [ L, U ] = LU( Di );  Z = solinf( L , Fi );   XX( indBlock_im1 ) = solsup( U, Z );
end
endfunction
=====================================
```

□

(c) Donner le coût en nombre de multiplications de cette fonction.

Réponse : Le coût d'un produit (matrice * matrice) de taille n est n^3. Le coût d'une factorisation $A = LU$ de taille n est $n^3/3$. On néglige les produits (matrice * vecteur) et les résolutions de systèmes triangulaires, car leur coût pour une taille n est de l'ordre de n^2 (n^2 pour mat*vec, $n^2/2$ pour une résolution de système triangulaire).

Dans la première boucle, on effectue $nBloc - 1$ produits (matrice * matrice) de taille n : le coût est donc $nBloc \times n^3$ (on néglige les termes d'ordre n^2).

Dans la deuxième partie du code, on effectue $nBloc$ factorisation LU de taille n, pour un coût de $nBloc \times n^3/3$.

Le coût total est donc $\dfrac{4nBloc \times n^3}{3}$ (qui vaut $4n^4/3$ si $nBloc = n$ comme dans l'exercice).

Par comparaison, une résolution brutale via LU de la matrice K de taille $nBloc \times n$ serait de $\dfrac{nBloc^3 \times n^3}{3}$ (soit ici $n^6/3$)! □

Exercice 3 : *(barème approximatif : 5 points)* CHANGEZ DE COPIE

On se donne la suite suivante, définie par récurrence :

$$\left\{ \begin{array}{rcll} x_{n+1} & = & -\dfrac{5}{2} - \dfrac{1}{x_n}, & n = 0, 1, \ldots, \\ x_0 & & \text{donné}. & \end{array} \right. \tag{4}$$

1. En posant $x_n = \dfrac{u_{n+1}}{u_n}$, transformer la récurrence (4) en une récurrence linéaire du type :

$$\left\{ \begin{array}{rcll} u_{n+2} & = & a u_{n+1} + b u_n, & n = 0, 1, \ldots, \\ u_0 & = & \alpha, & \\ u_1 & = & \beta. & \end{array} \right. \tag{5}$$

On donnera les valeurs de a et b.

Réponse : on remplace x_{n+1} et on trouve

$$\left\{ \begin{array}{rcll} u_{n+2} & = & -\dfrac{5}{2} u_{n+1} - u_n, & n = 0, 1, \ldots \end{array} \right.$$

□

2. (a) Calculer la solution exacte u_n en fonction de n, de α et β.

Réponse : l'équation caractéristique s'écrit

$$\lambda^2 + \frac{5}{2}\lambda + 1 = 0.$$

Les racines sont $-1/2$ et -2, donc la solution de la suite s'écrit

$$u_n = A\left(-\frac{1}{2}\right)^n + B\,(-2)^n.$$

Les conditions initiales u_0 et u_1 permettent de déterminer A et B :

$$\left\{ \begin{array}{rcl} \alpha & = & A + B \\ \beta & = & -A/2 - 2B \end{array} \right. \iff \left\{ \begin{array}{rcl} A & = & \dfrac{1}{3}(4\alpha + 2\beta) \\ B & = & \dfrac{1}{3}(-\alpha - 2\beta) \end{array} \right. ,$$

ce qui donne

$$u_n = \frac{1}{3}\left((4\alpha + 2\beta)\left(-\frac{1}{2}\right)^n - (\alpha + 2\beta)(-2)^n \right).$$

$\square$

(b) Quelle est la limite de u_n quand n tend vers l'infini? Discuter en fonction des valeurs de α et β.
Réponse : on sait que $(-1/2)^n$ tend vers 0 et que $|-2|^n$ tend vers l'infini.
Donc si $\alpha \neq -2\beta$, alors $|u_n|$ tend vers $+\infty$ (u_{2k} et u_{2k+1} tendent l'un vers $+\infty$, l'autre vers $-\infty$).
Si $\alpha = -2\beta$, alors u_n tend vers 0. $\square$

(c) Donner la limite de x_n quand n tend vers l'infini, si $\alpha = \frac{2}{3}$ et $\beta = -\frac{1}{3}$.
Réponse : on est dans le cas où $\alpha = -2\beta$, donc

$$u_n = \frac{2}{3}\left(-\frac{1}{2}\right)^n \quad \textbf{tend vers } 0.$$

Il vient

$$x_n = \frac{u_{n+1}}{u_n} = \frac{(-1/2)^{n+1}}{(-1/2)^n} = -\frac{1}{2} \quad \textbf{constant, donc tend vers } -\frac{1}{2}.$$

$\square$

3. On travaille à présent avec une arithmétique exacte, mais en tenant compte des erreurs d'arrondi qui sont faites sur la condition initiale : on suppose que

$$\tilde{u}_0 = \tilde{\alpha} = \frac{2}{3}(1+\delta_1) \quad \text{et} \quad \tilde{u}_1 = \tilde{\beta} = -\frac{1}{3}(1+\delta_2),$$

avec δ_1 et δ_2 petits.

(a) Calculer dans ce cas la solution pertubée $\tilde{u}_n$. Quelle est sa limite quand n tend vers l'infini?
Réponse : on reprend les calculs pour obtenir $\tilde{A}$ et $\tilde{B}$:

$$\left\{ \begin{array}{rcl} \tilde{\alpha} & = & \tilde{A} + \tilde{B} \\ \tilde{\beta} & = & -\tilde{A}/2 - 2\tilde{B} \end{array} \right. \quad \Longleftrightarrow \quad \left\{ \begin{array}{rcl} \tilde{A} & = & \frac{1}{3}(4\tilde{\alpha} + 2\tilde{\beta}) = \frac{2}{3}\left(1 + \frac{4\delta_1 - \delta_2}{3}\right) \\ \tilde{B} & = & \frac{1}{3}(-\tilde{\alpha} - 2\tilde{\beta}) = -\frac{2}{3}\left(\frac{\delta_1 - \delta_2}{3}\right) \end{array} \right. ,$$

ce qui donne

$$\tilde{u}_n = \frac{2}{3}\left(\left(1 + \frac{4\delta_1 - \delta_2}{3}\right)\left(-\frac{1}{2}\right)^n - \left(\frac{\delta_1 - \delta_2}{3}\right)(-2)^n \right).$$

Au départ la suite perturbée commence par tendre vers 0 car les δ_i sont petits, mais ensuite la suite perturbée $|\tilde{u}_n|$ tend comme 2^n vers $+\infty$, car $\delta_1 \neq \delta_2$ en général. $\square$

(b) Déterminer la limite de $\tilde{x}_n$ dans ce cas. Conclure.
Réponse : on obtient avec $\delta_1 \neq \delta_2$:

$$\tilde{x}_n = \frac{\tilde{u}_{n+1}}{\tilde{u}_n} = \frac{-\left(\frac{\delta_1 - \delta_2}{3}\right)(-2)^{n+1} + \left(1 + \frac{4\delta_1 - \delta_2}{3}\right)(-1/2)^{n+1}}{-\left(\frac{\delta_1 - \delta_2}{3}\right)(-2)^{n} + \left(1 + \frac{4\delta_1 - \delta_2}{3}\right)(-1/2)^{n}} \sim_{n\to\infty} -2$$

Donc la suite perturbée $\tilde{x}_n$ tend vers -2, alors que la suite x_n tend vers $-\frac{1}{2}$. Les erreurs d'arrondi sur la condition initiale suffisent pour provoquer ce changement de comportement surprenant *a priori*. $\square$