Chapitre 9

Propagation des erreurs d'arrondi dans la méthode de Gauss

Il s'agit d'étudier sur ordinateur en arithmétique flottante les étapes de la factorisation théorique $\mathcal{A} = \mathcal{L}\,\mathcal{U}$ avec $\mathcal{A} \in \mathcal{M}_n(\mathbb{R})$
On suppose qu'il n'y a pas d'erreur de donnée : $\mathcal{A}= A$
Soit LU le résultat informatique, on a une matrice erreur E telle que $LU - A = E$

1 Analyse de l'erreur à l'étape k :

L'algorithme de Gauss utilisé dans la décomposition LU de A s'écrit sous forme matricielle $A^{(k)} = L^{(k)}\ A^{(k+1)}\ \ 1 \leq k \leq n-1$ avec $A^{(1)} = A$ et $A^{(n)} = U$ (matrice triangulaire supérieure) et $L^{(k)} = I + l^{(k)} e_k^t$ avec e_k le k^{eme} vecteur de la base canonique et $l^{(k)} = (0, \cdots, 0, l^{(k)}_{k+1}, \cdots, l^{(k)}_n)^t$

$$L^{(k)} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 0 & \ddots & & & \vdots \\ \vdots & & \vdots & l^{(k)}_{k+1} & \ddots & & \vdots \\ \vdots & & \vdots & \vdots & 0 & \ddots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \\ 0 & \cdots & 0 & l^{(k)}_n & 0 & \cdots\ 0 & 1 \end{pmatrix}$$

matrice d'élimination à l'étape k (passage de la matrice $A^{(k)}$ à la matrice $A^{(k+1)}$), triangulaire inférieure à diagonale unité avec la k^{eme} colonne le vecteur $e_k + l^{(k)}$.

On a les formules algébriques récurrentes classiques de calcul des cœfficients des matrices :

$$L^{(k)}_{ik} = \frac{A^{(k)}_{ik}}{A^{(k)}_{kk}}\ ,\quad k+1 \leq i \leq n$$

$$A^{(k+1)}_{ij} = A^{(k)}_{ij} - L^{(k)}_{ik} A^{(k)}_{kj}\ ,\quad k+1 \leq i, j \leq n$$

Remarque :

$$A^{(k+1)}_{ij} = A^{(k)}_{ij}\ ,\quad 1 \leq i \leq k \quad j = 1, \cdots, n$$
$$A^{(k+1)}_{ij} = A^{(k)}_{ij} = 0\ ,\quad k+1 \leq i \leq n \quad 1 \leq j \leq k-1$$

On définit la matrice erreur à l'étape k par $E^{(k)} = L^{(k)}A^{(k+1)} - A^{(k)}$
On a $E_{ij}^{(k)} = 0 \ , \quad 1 \leq i \leq k \quad j = 1, \cdots, n$ ou $i = 1, \cdots, n \quad 1 \leq j \leq k-1$

Majoration des cœfficients de la matrice $E^{(k)}$

<u>Pour $j = k$</u>
Par élimination on a $A_{ik}^{(k+1)} = 0 \ , \quad k+1 \leq i \leq n$
On obtient sur machine $L_{ik}^{(k)} = \dfrac{A_{ik}^{(k)}}{A_{kk}^{(k)}}(1 + \delta_{ik}^{(k)}) \ , \quad k+1 \leq i \leq n$ où δ représente l'erreur relative d'arrondi sur le calcul de la division.
d'où $E_{ik}^{(k)} = L_{ik}^{(k)}A_{kk}^{(k+1)} - A_{ik}^{(k)} = \dfrac{A_{ik}^{(k)}}{A_{kk}^{(k)}}(1 + \delta_{ik}^{(k)})A_{kk}^{(k+1)} - A_{ik}^{(k)}$
On a $A_{kk}^{(k+1)} = A_{kk}^{(k)}$
d'où $E_{ik}^{(k)} = A_{ik}^{(k)}(1 + \delta_{ik}^{(k)}) - A_{ik}^{(k)} = A_{ik}^{(k)}\delta_{ik}^{(k)} \ , \quad k+1 \leq i \leq n$
On peut prendre $|\delta_{ik}^{(k)}| \leq \mathbf{u}$ (unité d'arrondi) par application des erreurs d'arrondi sur les opérations arithmétiques élémentaires, d'où
$|E_{ik}^{(k)}| \leq \mathbf{u}|A_{ik}^{(k)}| \ , \quad k+1 \leq i \leq n$

<u>Pour $j > k$</u>
On a $E_{ij}^{(k)} = L_{ik}^{(k)}A_{kj}^{(k+1)} + A_{ij}^{(k+1)} - A_{ij}^{(k)} \ , \quad k+1 \leq i \leq n$
On a $A_{kj}^{(k+1)} = A_{kj}^{(k)} \ , \quad j = 1, \cdots, n$
d'où $E_{ij}^{(k)} = A_{ij}^{(k+1)} - (A_{ij}^{(k)} - L_{ik}^{(k)}A_{kj}^{(k)})$
En calculant sur machine on obtient $E_{ij}^{(k)} = A_{ij}^{(k+1)} - (A_{ij}^{(k)} - L_{ik}^{(k)}A_{kj}^{(k)})(1 + \gamma_{ij}^{(k)})$
où $\gamma_{ij}^{(k)}$ est l'erreur relative d'arrondi sur le calcul de l'expression $A_{ij}^{(k)} - L_{ik}^{(k)}A_{kj}^{(k)}$ qui est égale théoriquement (sans erreur) à $A_{ij}^{(k+1)}$.
$E_{ij}^{(k)}$ représente l'erreur absolue du calcul de $A_{ij}^{(k+1)}$ à partir de $A_{ij}^{(k)}$.

On obtient sur machine $L_{ik}^{(k)}A_{kj}^{(k)}(1 + \beta_{ij}^{(k)})$ où $\beta_{ij}^{(k)}$ est l'erreur relative d'arrondi sur le calcul du produit.
Puis $\left(A_{ij}^{(k)} - L_{ik}^{(k)}A_{kj}^{(k)}(1 + \beta_{ij}^{(k)})\right)(1+\alpha_{ij}^{(k)})$ où $\alpha_{ij}^{(k)}$ est l'erreur relative d'arrondi sur le calcul de la différence.

Théoriquement $A_{ij}^{(k+1)} = A_{ij}^{(k)} - L_{ik}^{(k)}A_{kj}^{(k)}$.
D'où $E_{ij}^{(k)} = A_{ij}^{(k)} - L_{ik}^{(k)}A_{kj}^{(k)} - \left(A_{ij}^{(k)} - L_{ik}^{(k)}A_{kj}^{(k)}(1 + \beta_{ij}^{(k)})\right)(1 + \alpha_{ij}^{(k)})$
C'est-à-dire $E_{ij}^{(k)} = A_{ij}^{(k)} - L_{ik}^{(k)}A_{kj}^{(k)} - \left(A_{ij}^{(k)}(1 + \alpha_{ij}^{(k)}) - L_{ik}^{(k)}A_{kj}^{(k)}(1 + \alpha_{ij}^{(k)})(1 + \beta_{ij}^{(k)})\right) = -A_{ij}^{(k)}\alpha_{ij}^{(k)} + L_{ik}^{(k)}A_{kj}^{(k)}\alpha_{ij}^{(k)} + L_{ik}^{(k)}A_{kj}^{(k)}\beta_{ij}^{(k)}(1+\alpha_{ij}^{(k)}) = -A_{ij}^{(k+1)}\alpha_{ij}^{(k)} + L_{ik}^{(k)}A_{kj}^{(k)}\beta_{ij}^{(k)}(1+\alpha_{ij}^{(k)}) = -A_{ij}^{(k+1)}\alpha_{ij}^{(k)} + (A_{ij}^{(k)} - A_{ij}^{(k+1)})\beta_{ij}^{(k)}(1 + \alpha_{ij}^{(k)})$
On peut prendre $|\alpha_{ij}^{(k)}|$ et $|\beta_{ij}^{(k)}| \leq \mathbf{u}$ (unité d'arrondi) par application des erreurs d'arrondi sur les opérations arithmétiques élémentaires, d'où
$|E_{ij}^{(k)}| \leq \mathbf{u}(3 + 2\mathbf{u})max(|A_{ij}^{(k+1)}|, |A_{ij}^{(k)}|) \ , \quad k+1 \leq i \leq n \quad j > k$

Pour $j \geq k$ on obtient :
$|E_{ij}^{(k)}| \leq \mathbf{u}(3 + 2\mathbf{u})max(|A_{ij}^{(k+1)}|, |A_{ij}^{(k)}|) \ , \quad k+1 \leq i \leq n \quad j \geq k$

à l'ordre 1 :
$|E_{ij}^{(k)}| \leq 3\mathbf{u}K_{ij}^{(k+1)}$, $\quad k{+}1 \leq i \leq n \quad k \leq j \leq n \quad$ avec $K_{ij}^{(k+1)} = max(|A_{ij}^{(k+1)}|, |A_{ij}^{(k)}|)$
d'où $|E_{ij}^{(k)}| \leq 3\mathbf{u}K_{k+1}$, $\quad k{+}1 \leq i \leq n \quad k \leq j \leq n \quad$ avec $K_{k+1} = \underset{\substack{1\leq i,j\leq n \\ 1\leq l\leq k+1}}{max} |A_{ij}^{(l)}|$

<u>Remarque</u> :
Les cœfficients $E_{ij}^{(k)}$, $\quad 1 \leq i \leq k \quad j = 1, \cdots, n$ ou $i = 1, \cdots, n \quad 1 \leq j \leq k-1$ sont nuls.

2 Analyse de l'erreur pour la décomposition LU :

La matrice $E^{(k)}$ représente l'erreur de calcul à l'étape k. Il s'agit d'étudier la propagation de ces erreurs au cours des $n-1$ étapes.
On a $L^{(k)}A^{(k+1)} = A^{(k)} + E^{(k)}$ **(0)**
On considère les matrices $L^{(i)} = I + l^{(i)}e_i^t$ et $F^{(k)} \in \mathcal{M}_n(\mathbb{R})$ (avec ses k premières lignes nulles).
On a $L^{(i)}F^{(k)} = F^{(k)}$ pour $i \leq k$ **(1)**
En effet, $L^{(i)}F^{(k)} = F^{(k)} + l^{(i)}e_i^t F^{(k)}$ mais $e_i^t F^{(k)} = 0$ (puisque la i^{eme} ligne de $F^{(k)}$ nulle pour $i \leq k$).
En utilisant les relations **(0)** et **(1)**, on a :
$L^{(1)}L^{(2)}\cdots L^{(k-1)}L^{(k)}A^{(k+1)} = L^{(1)}L^{(2)}\cdots L^{(k-1)}A^{(k)} + L^{(1)}\cdots L^{(k-1)}E^{(k)}$
$= L^{(1)}\cdots L^{(k-1)}A^{(k)} + E^{(k)}$
D'où $L^{(1)}L^{(2)}\cdots L^{(n-2)}L^{(n-1)}A^{(n)} = L^{(1)}\cdots L^{(n-2)}A^{(n-1)} + E^{(n-1)} =$
$L^{(1)}\cdots L^{(n-3)}A^{(n-2)} + E^{(n-2)} + E^{(n-1)} = \cdots = L^{(1)}A^{(2)} + E^{(2)} + \cdots + E^{(n-2)} + E^{(n-1)} = A^{(1)} + E^{(1)} + E^{(2)} + \cdots + E^{(n-2)} + E^{(n-1)}$
Soit

$$\underbrace{L^{(1)}L^{(2)}\cdots L^{(n-1)}A^{(n)}}_{matrice\ LU\ calcule\ sur\ machine} = A^{(1)} + E^{(1)} + \cdots + E^{(n-1)} = A + E$$

On note $|A| = (|A_{ij}|)$ et $A \leq B$ pour $A_{ij} \leq B_{ij}\ \forall i, j$

Pour majorer E (au sens des coefficients), on majore tous les nombres K_{k+1} par K_n

$$\text{On a } F^{(k)} = \begin{array}{c} \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \vdots \\ \vdots & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots\ 0 & 1 & \cdots & 1 \end{pmatrix} \\ \text{k}^{eme} \text{ colonne} \end{array} \text{k+1}^{eme} \text{ ligne}$$

D'où pour la majoration $|E_{ij}^{(k)}| \leq 3\mathbf{u}K_{k+1}$, on a $|E^{(k)}| \leq 3\mathbf{u}K_n F^{(k)}$
donc $|E| = |\sum_{k=1}^{n-1} E^{(k)}| \leq \sum_{k=1}^{n-1} |E^{(k)}| \leq 3\mathbf{u}K_n \sum_{k=1}^{n-1} F^{(k)}$

Notons $F = \sum_{k=1}^{n-1} F^{(k)}$, on a :

$$F = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 1 & \cdots & \cdots & \cdots & \cdots & 1 \\ 1 & 2 & \cdots & \cdots & \cdots & 2 \\ 1 & 2 & 3 & \cdots & \cdots & 3 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 2 & 3 \cdots & (n-2) & (n-2) & (n-2) \\ 1 & 2 & 3 \cdots & (n-2) & (n-1) & (n-1) \end{pmatrix}$$

d'où $|E| \leq 3\mathbf{u}K_nF$
K_n représente le plus grand cœfficient en valeur absolue de toutes les matrices $A^{(k)}$, $\quad k = 1, \cdots, n$ de la méthode de Gauss et $\mathbf{u}$ l'unité d'arrondi de l'arithmétique flottante.

Par définition du pivot partiel on a $|L_{ik}^{(k)}| \leq 1$, donc par les formules algébriques on a $|A_{ij}^{(k+1)}| \leq |A_{ij}^{(k)}| + |A_{kj}^{(k)}|$ d'où $K_{k+1} \leq 2K_k$ soit $K_n \leq 2^{n-1}K_1$
Dans ce cas, on a $PA = LU + E$ où P est une matrice de permutation qui ne produit pas d'erreur.
La borne K_n est assez pessimiste. Des majorations plus fines ont été obtenues pour des classes de matrices (à diagonale dominante,...) par J.H. Wilkinson.
K_n se calcule au fur et à mesure des étapes et permet une analyse a posteriori des erreurs d'arrondi et de la croissance des cœfficients (intermédiaires) des matrices $A^{(k)}$.