4 MÉTHODES DIRECTES DE RÉSOLUTION DE SYSTÈMES LINÉAIRES

Introduction

Le problème que nous allons maintenant considérer est celui de la résolution numérique d'un système linéaire $Au = b$, dont la matrice A est inversible.

Le principe des *méthodes directes* étudiées dans ce chapitre revient à la détermination d'une matrice M inversible telle que *la matrice* MA *soit triangulaire supérieure* (théorème 4.2-1) ; c'est ce qui correspond à la procédure d'*élimination*. Il reste ensuite à résoudre le système linéaire

$$MAu = Mb,$$

par la méthode de *remontée*, décrite au paragraphe 4.1 (dans les calculs *effectifs*, on ne calcule pas explicitement la matrice M, mais seulement la matrice MA et le vecteur Mb).

Ce principe est à la base de la *méthode de Gauss* pour les systèmes linéaires à matrices "quelconques" et de la *méthode de Cholesky* pour les systèmes linéaires à matrices symétriques définies positives, que nous décrivons aux paragraphes 4.2 et 4.4, respectivement. Nous étudions au passage le calcul des *déterminants* des matrices carrées, le cas particulier (très important dans les applications) des *matrices tridiagonales* par points ou par blocs, ainsi que *la méthode de Gauss-Jordan*, que l'on peut considérer comme une variante de la méthode de Gauss, particulièrement bien adaptée au calcul de l'*inverse* d'une matrice donnée. Insistons à cet égard sur l'*inutilité du calcul de l'inverse d'une matrice pour la résolution d'un système linéaire*, comme nous l'indiquons dès le paragraphe 4.1.

L'interprétation matricielle de la méthode de Gauss est *la factorisation* LU *d'une matrice* (théorème 4.3-1). Ce résultat *très important*, notamment par ses applications variées en Analyse Numérique Matricielle, montre que, à des permutations éventuelles de lignes près, toute matrice inversible peut s'écrire comme le produit d'une *matrice triangulaire inférieure* L par une *matrice triangulaire supérieure* U. Cette factorisation se simplifie quelque peu dans le cas des *matrices symétriques définies positives* ; elle devient alors la *factorisation de Cholesky* (théorème 4.4-1), à la base de la méthode de Cholesky, déjà signalée.

Alors que les méthodes de Gauss et de Cholesky sont basées sur la factorisation de la matrice du système linéaire en un produit d'une matrice triangulaire inférieure par une matrice triangulaire supérieure, la *méthode de Householder* est associée à la factorisation d'une matrice (non nécessairement symétrique) en un produit d'une matrice *orthogonale* Q par une matrice triangulaire supérieure R. Nous décrivons cette méthode au paragraphe 4.5, où nous montrons qu'une telle *factorisation* QR peut être effectuée de façon ingénieuse à l'aide de matrices orthogonales "élémentaires" particulières, appelées *matrices de Householder.*

Signalons que la factorisation QR d'une matrice interviendra comme étape essentielle de calcul dans la "méthode QR" d'approximation des valeurs propres d'une matrice quelconque (paragraphe 6.3), et que les matrices de Householder seront à nouveau utilisées pour la réduction à la forme tridigonale d'une matrice symétrique (paragraphe 6.2).

4.1. Deux remarques concernant la résolution des systèmes linéaires

Contrairement à ce qu'une analyse sommaire pourrait laisser supposer, *la solution d'un système linéaire* $\mathrm{A}u = b$, A : *matrice inversible, ne s'obtient pas en calculant la matrice* A^{-1}, *puis en calculant le vecteur* $\mathrm{A}^{-1}b$ *(première remarque).* Le calcul de la matrice A^{-1} est en effet équivalent à la résolution des n systèmes linéaires (n : ordre de la matrice A) :

$$\mathrm{A}u_j = e_j, \quad 1 \leqslant j \leqslant n,$$

où e_j est le j-ème vecteur de la base de $\mathbf{K}^n$. Autrement dit, on remplacerait par une telle méthode la résolution d'*un* système linéaire (le problème donné) par la résolution de n systèmes linéaires, suivie de la multiplication de la matrice A^{-1} par le vecteur b !

Les méthodes que nous allons étudier sont basées sur la *deuxième remarque* évidente suivante : *si la matrice inversible* A *est triangulaire supérieure* (ou triangulaire inférieure), *la résolution numérique d'un système linéaire* $\mathrm{A}u = b$ *est immédiate* ; il s'écrit en effet

$$\begin{aligned} a_{11}u_1 + \ \dots \ + a_{1,n-1}u_{n-1} + a_{1n}u_n &= b_1, \\ &\ \vdots \\ a_{n-1,n-1}u_{n-1} + a_{n-1,n}u_n &= b_{n-1}, \\ a_{nn}u_n &= b_n, \end{aligned}$$

et puisque $a_{11}a_{22} \dots a_{nn} = \text{dét}\,(\mathrm{A}) \neq 0$, on résout le système en calculant u_n de la dernière équation, puis u_{n-1} de l'avant dernière, etc., ce qui donne

$$\begin{aligned} u_n &= a_{nn}^{-1} b_n, \\ u_{n-1} &= a_{n-1,n-1}^{-1}(b_{n-1} - a_{n-1,n}u_n), \\ &\ \vdots \\ u_1 &= a_{11}^{-1}(b_1 - a_{12}u_2 - \ \dots \ - a_{1,n-1}u_{n-1} - a_{1n}u_n). \end{aligned}$$

Chaque composante u_i apparaissant ainsi comme une fonction linéaire de $b_i, b_{i+1}, \dots, b_n$, ceci montre au passage que *l'inverse d'une matrice triangulaire est une matrice triangulaire du même type (supérieure ou inférieure).*

La méthode ci-dessus, appelée *méthode de remontée*, nécessite donc au total

$$\begin{cases} 1+2+ \ \dots \ +(n-1) = \dfrac{n(n-1)}{2} \text{ additions,} \\[2ex] 1+2+ \ \dots \ +(n-1) = \dfrac{n(n-1)}{2} \text{ multiplications,} \\[2ex] n \text{ divisions.} \end{cases}$$

La méthode de rémontée s'étend aux *matrices triangulaires par blocs*. Ainsi par exemple la résolution du système linéaire

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ & A_{22} & A_{23} \\ & & A_{33} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

se ramène-t-elle à la résolution des systèmes linéaires successifs :

$$\begin{aligned} A_{33}u_3 &= b_3, \\ A_{22}u_2 &= b_2 - A_{23}u_3, \\ A_{11}u_1 &= b_1 - A_{12}u_2 - A_{13}u_3. \end{aligned}$$

Naturellement, ceci suppose que l'on sache résoudre les systèmes linéaires dont les matrices sont les sous-matrices diagonales A_{ii}.

Dans l'étude des méthodes de résolution de systèmes linéaires (aussi bien directes qu'itératives), nous utiliserons fréquemment l'écriture matricielle. En particulier, nous considérerons des systèmes linéaires "intermédiaires", disons $Cu = d$, que nous "résoudrons" sous la forme $u = C^{-1}d$, ce qui semble être en contradiction avec la première remarque énoncée plus haut.

En fait, *ceci n'est qu'une commodité d'écriture*, et si on regarde les choses de plus près, on s'apercevra que la matrice C est triangulaire, par points ou par blocs, de sorte que l'on résout le système $Cu = d$ en utilisant la méthode de rémontée, *sans calculer explicitement la matrice* C^{-1}. C'est pourquoi l'on dit parfois, *avec abus évident de langage*, que de telles matrices C sont "faciles à inverser".

4.2. La méthode de Gauss

La *méthode de Gauss* est une méthode générale de résolution d'un système linéaire

$$Au = b, \quad A : \text{matrice inversible.}$$

Elle comporte trois étapes :

(i) une procédure dite d'*élimination* (successive des inconnues) équivaut à déterminer une matrice inversible M telle que la matrice MA soit *triangulaire supérieure* ;

(ii) on calcule simultanément le vecteur Mb ;

(iii) on résout le système linéaire

$$MAu = Mb,$$

à matrice triangulaire supérieure, par la méthode de *rémontée* décrite au paragraphe précédent.

REMARQUE. En pratique, *on ne calcule pas explicitement la matrice* M, mais seulement la matrice MA et le vecteur Mb. L'introduction de la matrice M n'est qu'une *commodité d'écriture.* ■

Décrivons la *première étape de l'élimination.* L'un au moins des éléments a_{i1}, $1 \leq i \leq n$, de la première colonne de la matrice $A = (a_{ij})$ est différent de zéro, sans quoi la matrice serait singulière. On *choisit* alors l'un des coefficients *non nuls* de la première colonne de A (nous examinerons ultérieurement comment on choisit *effectivement* cet élément non nul ; pour l'instant, le choix effectif n'a pas d'importance), qu'on appelle le premier *pivot* de l'élimination.

Ensuite, *on échange la ligne du pivot avec la première ligne*, ce qui, en écriture matricielle, revient à multiplier la matrice A à gauche par une matrice de permutation P particulière. On vérifie en effet que l'échange des i_0-ème et i_1-ème lignes d'une matrice équivaut à la multiplier à gauche par la *matrice de transposition* (en supposant $i_0 < i_1$ pour fixer les idées) :

$$
T(i_0, i_1) = \begin{array}{cc}
\begin{array}{cccccccccc} & & i_0 & & & & i_1 & & & \\ & & \downarrow & & & & \downarrow & & & \end{array} & \\
\left(\begin{array}{cccccccccc}
1 &&&&&&&&& \\
& 1 &&&&&&&& \\
&& 0 & \cdots & \cdots & \cdots & 1 & \cdots & \cdots & \\
&& \vdots & 1 &&& \vdots &&& \\
&& \vdots && 1 && \vdots &&& \\
&& \vdots &&& 1 & \vdots &&& \\
&& 1 & \cdots & \cdots & \cdots & 0 & \cdots & \cdots & \\
&&&&&&& 1 && \\
&&&&&&&& 1 & \\
&&&&&&&&& 1
\end{array}\right) & \begin{array}{l} \\ \\ \longleftarrow i_0 \\ \\ \\ \\ \longleftarrow i_1 \\ \\ \\ \\ \end{array}
\end{array} .
$$

On notera au passage que

$$\text{dét}\,(T(i_0, i_1)) = -1.$$

On posera donc

$$P = \begin{cases} I \text{ si } a_{11} \text{ est le pivot, et alors dét (P)} = 1\ ; \\ T(1, i) \text{ si } a_{i1},\ i \neq 1, \text{ est le pivot, et alors dét } (P) = -1, \end{cases}$$

et la matrice

$$PA = (\alpha_{ij})$$

ainsi obtenue est telle que

$$\alpha_{11} \neq 0,$$

par construction. *Par des combinaisons linéaires appropriées de la première ligne et des autres lignes de la matrice* PA, *on annule ensuite tous les éléments de la première colonne de la matrice* PA *situés sous la diagonale*, la première ligne restant inchangée.

En écriture matricielle, ceci revient à multiplier la matrice PA à gauche par la matrice

$$E = \begin{pmatrix} 1 &&&&& \\ -\pi^{-1}\alpha_{21} & 1 &&&& \\ -\pi^{-1}\alpha_{31} && 1 &&& \\ \cdot &&& \cdot && \\ \cdot &&&& \cdot & \\ \cdot &&&&& \cdot \\ -\pi^{-1}\alpha_{n1} &&&&& 1 \end{pmatrix}, \quad \pi = \alpha_{11},$$

de sorte que la matrice

$$B = EPA$$

est de la forme

$$B = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \ldots \alpha_{1n} \\ & b_{22} & b_{23} \ldots b_{2n} \\ & b_{32} & b_{33} \ldots b_{3n} \\ & \cdot & \cdot \qquad \cdot \\ & \cdot & \cdot \qquad \cdot \\ & \cdot & \cdot \qquad \cdot \\ & b_{n2} & b_{n3} \ldots b_{nn} \end{pmatrix}.$$

REMARQUE. On ne saurait trop insister sur la simplicité des opérations effectuées (échange de deux lignes, combinaisons linéaires de lignes), quelque peu “masquée” par l'écriture matricielle. Il faudrait être bien naïf pour croire que l'on calcule effectivement les matrices P, E et que l'on effectue ensuite les produits de matrices PA, EPA .... A cet égard, l'exemple numérique détaillé plus loin devrait être instructif. ■

Comme

$$\text{dét}\,(\mathrm{E}) = 1,$$

on a donc

$$\text{dét}\,(\mathrm{B}) = \text{dét}\,(\mathrm{E})\ \text{dét}\,(\mathrm{P})\ \text{dét}\,(\mathrm{A}) = \pm\ \text{dét}\,(\mathrm{A}),$$

selon que l'on a, ou non, effectué un échange de lignes, ce qui montre au passage que la matrice B est encore inversible. Par conséquent l'un au moins des éléments b_{i2}, $2 \leqslant i \leqslant n$, est différent de zéro, et tout naturellement, la seconde étape de l'élimination consiste à effectuer les mêmes opérations que précédemment, mais seulement sur la sous-matrice (b_{ij}), $2 \leqslant i, j \leqslant n$, en laissant inchangée la première ligne, et ainsi de suite ....

Notant désormais

$$\mathrm{A} = \mathrm{A}_1 = (a_{ij}^1), \quad \mathrm{P} = \mathrm{P}_1, \quad \mathrm{P}_1\mathrm{A}_1 = (\alpha_{ij}^1),$$

$$\mathrm{E} = \mathrm{E}_1, \quad \mathrm{B} = \mathrm{A}_2 = \mathrm{E}_1\mathrm{P}_1\mathrm{A}_1 = (a_{ij}^2),$$

on obtient de proche en proche comme *résultat de la* $(k-1)$*-ème étape de l'élimination*, $2 \leqslant k \leqslant n$, une matrice

$$\mathrm{A}_k = \mathrm{E}_{k-1}\mathrm{P}_{k-1} \ldots \mathrm{E}_2\mathrm{P}_2\mathrm{E}_1\mathrm{P}_1\mathrm{A}_1,$$

qui est de la forme

$$\mathrm{A}_k = \begin{pmatrix} a_{11}^k & a_{12}^k & \cdots & & \cdots & a_{1n}^k \\ & a_{22}^k & \cdots & & \cdots & a_{2n}^k \\ & & \ddots & & & \vdots \\ & & & a_{kk}^k & \cdots & a_{kn}^k \\ & & & \vdots & & \vdots \\ & & & a_{nk}^k & \cdots & a_{nn}^k \end{pmatrix} = \begin{pmatrix} \alpha_{11}^1 & \alpha_{12}^1 & \cdots & & \cdots & \alpha_{1n}^1 \\ & \alpha_{22}^2 & \cdots & & \cdots & \alpha_{2n}^2 \\ & & \ddots & & & \vdots \\ & & & a_{kk}^k & \cdots & a_{kn}^k \\ & & & \vdots & & \vdots \\ & & & a_{nk}^k & \cdots & a_{nn}^k \end{pmatrix},$$

la seconde expression rappelant que *les i premières lignes restent inchangées après la i-ème étape*.

Décrivons alors la *k-ème étape de l'élimination*. Puisque $\text{dét}\,(\mathrm{A}_k) = \pm\,\text{dét}\,(\mathrm{A})$, la matrice A_k est inversible, et donc l'un au moins des éléments a_{ik}^k, $k \leqslant i \leqslant n$, est différent de zero. On choisit l'un de ces éléments *non nuls* comme *pivot*, et l'on échange la ligne du pivot avec la k-ème ligne de la matrice A_k. En écriture matricielle, ceci revient à multiplier la matrice A_k à gauche par une matrice P_k qui est soit l'identité, soit une matrice de transposition, de sorte que, par construction, l'élément α_{kk}^k de la matrice

$$\mathrm{P}_k\mathrm{A}_k = (\alpha_{ij}^k)$$

est différent de zéro.

L'élimination proprement dite équivaut à la multiplication de la matrice P_kA_k à gauche par la matrice

$$E_k = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & -\pi_k^{-1}\alpha_{k+1,k}^k & 1 & & & \\ & & \vdots & & \ddots & & \\ & & \vdots & & & \ddots & \\ & & -\pi_k^{-1}\alpha_{nk}^k & & & & 1 \end{pmatrix}, \quad \pi_k = \alpha_{kk}^k,$$

opération qui ne modifie pas les k premières lignes de la matrice P_kA_k.

Après la $(n-1)$-ème étape, *la matrice*

$$A_n = E_{n-1}P_{n-1} \dots E_2P_2E_1P_1A$$

est donc triangulaire supérieure ; on a ainsi trouvé une matrice inversible

$$M = E_{n-1}P_{n-1} \dots E_2P_2E_1P_1$$

telle que la matrice MA soit triangulaire supérieure, ce qui était, rappelons-le, l'objectif "matriciel" de la procédure d'élimination. Observons que

$$\text{dét}\,(M) = \begin{cases} +\,1 \text{ si l'on a effectué un nombre pair d'échanges de lignes,} \\ -\,1 \text{ si l'on a effectué un nombre impair d'échanges de lignes,} \end{cases}$$

de sorte que, comme "sous-produit" de la procédure d'élimination, *on obtient un procédé particulièrement rapide du calcul du déterminant de la matrice* A, qui n'est autre, au signe près, que le *produit des pivots*, puisque

$$\text{dét}\,(A) = \frac{\text{dét}\,(A_n)}{\text{dét}\,(M)} = \pm\alpha_{11}^1\alpha_{22}^2 \dots \alpha_{nn}^n,$$

en convenant que $\alpha_{nn}^n = (A_n)_{nn}$ est le n-ème *pivot*.

L'aspect matriciel de la procédure d'élimination se trouve résumé dans le résultat suivant :

Théorème 4.2-1. ***Soit* A *une matrice carrée, inversible ou non. Il existe (au moins) une matrice inversible* M *telle que la matrice* MA *soit triangulaire supérieure.***

DÉMONSTRATION. Ce résultat est déjà démontré lorsque la matrice A est inversible. Or, la matrice A est singulière si et seulement si les éléments a_{ik}^k, $k \leq i \leq n$, de la matrice A_k sont nuls pour au moins une valeur de l'entier k. Mais dans ce cas, la matrice A_k est déjà de la forme A_{k+1}, et il suffit de considérer que $P_k = E_k = I$. ∎

Voici un *exemple numérique* d'application de la méthode de Gauss à la résolution d'un système linéaire. On a entouré de crochets les pivots, et on a indiqué en marge les combinaisons linéaires de lignes ; les notations sont celles introduites précédemment.

$$\left.\begin{array}{rl}(\mathrm{I}_1) & [5]\,u_1+2u_2+\ \ u_3 = \ 12\\ (\mathrm{II}_1) & 5\,u_1-6u_2+2u_3 = -1\\ (\mathrm{III}_1) & -4\,u_1+2u_2+\ \ u_3 = \ 3\end{array}\right\},\qquad \mathrm{A}=\mathrm{A}_1=\begin{pmatrix}5 & 2 & 1\\ 5 & -6 & 2\\ -4 & 2 & 1\end{pmatrix},$$

$$\left.\begin{array}{rl}(\mathrm{I}_1)=(\mathrm{I}_2) & 5\,u_1+2u_2+u_3 = 12\\ -(\mathrm{I}_1)+(\mathrm{II}_1)=(\mathrm{II}_2) & [-8]u_2+u_3 = -13\\ \frac{4}{5}(\mathrm{I}_1)+(\mathrm{III}_1)=(\mathrm{III}_2) & \frac{18}{5}u_2+\frac{9}{5}u_3 = \frac{63}{5}\end{array}\right\},\quad \mathrm{P}_1=\mathrm{I},\ \mathrm{E}_1=\begin{pmatrix}1 & & \\ -1 & 1 & \\ \frac{4}{5} & & 1\end{pmatrix},\quad \mathrm{A}_2=\begin{pmatrix}5 & 2 & 1\\ & -8 & 1\\ & \frac{18}{5} & \frac{9}{5}\end{pmatrix},$$

$$\left.\begin{array}{rl}(\mathrm{I}_2)=(\mathrm{I}_3) & 5\,u_1+2u_2+u_3 = 12\\ (\mathrm{II}_2)=(\mathrm{II}_3) & -8u_2+u_3 = -13\\ \frac{9}{20}(\mathrm{II}_2)+(\mathrm{III}_2)=(\mathrm{III}_3) & \left[\frac{9}{4}\right]u_3 = \frac{27}{4}\end{array}\right\},\quad \mathrm{P}_2=\mathrm{I},\ \mathrm{E}_2=\begin{pmatrix}1 & & \\ & 1 & \\ & \frac{9}{20} & 1\end{pmatrix},\quad \mathrm{A}_3=\begin{pmatrix}5 & 2 & 1\\ & -8 & 1\\ & & \frac{9}{4}\end{pmatrix}.$$

La matrice A_3 étant triangulaire, la résolution du dernier système s'effectue immédiatement par la méthode de remontée, ce qui donne

$$u_3 = 3, \quad u_2 = 2, \quad u_1 = 1.$$

Par ailleurs,

$$\text{dét}\,(A) = a_{11}^1 a_{22}^2 a_{33}^3 = 5\times(-8)\times\frac{9}{4} = -90.$$

REMARQUE. On peut évidemment *calculer* la matrice

$$M = E_2E_1 = \begin{pmatrix} 1 & & \\ -1 & 1 & \\ \frac{7}{20} & \frac{9}{20} & 1 \end{pmatrix},$$

ce qui permet d'écrire

$$MA = A_3, \quad \text{soit} \quad \begin{pmatrix} 1 & & \\ -1 & 1 & \\ \frac{7}{20} & \frac{9}{20} & 1 \end{pmatrix}\begin{pmatrix} 5 & 2 & 1 \\ 5 & -6 & 2 \\ -4 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 2 & 1 \\ & -8 & 1 \\ & & \frac{9}{4} \end{pmatrix},$$

mais en général, *l'expression de la matrice* M *n'est d'aucune utilité*. Comme on le verra, c'est au contraire la *matrice* M^{-1} qui présente de l'intérêt ; il se trouve d'ailleurs que cette matrice M^{-1} s'écrit *immédiatement* à partir de l'expression des matrices E_k (lorsqu'il n'y a pas eu d'échanges de lignes) alors que, pour le calcul de la matrice M, il faut effectuer le calcul explicite du produit des matrices E_k. ■

Donnons ensuite quelques précisions quant au *choix du pivot* à chaque étape de l'élimination. Naturellement, si au départ de la k-ème étape, l'élément a_{kk}^k de la matrice A_k est différent de zéro, rien ne s'oppose *théoriquement* à ce qu'il soit choisi comme pivot (alors $P_k = I$).

Mais, *à cause des erreurs d'arrondi, cette façon de procéder est, dans certains cas, complètement déconseillée.* A cet égard, l'exemple numérique suivant (tiré de [1], page 35) est à la fois spectaculaire et instructif. Supposons les calculs effectués en virgule flottante, avec une mantisse à trois chiffres, et dans le système décimal (pour fixer les idées et, surtout, pour faciliter les calculs...) ; autrement dit, *les données et les résultats de calculs intermédiaires sont arrondis aux trois premiers chiffres significatifs* (se reporter à la discussion du paragraphe 2.1). Considérons alors le système "exact"

$$\begin{aligned} (I_1)\,[10^{-4}]\, u_1 + u_2 &= 1, \\ (II_1) \qquad u_1 + u_2 &= 2, \end{aligned}$$

dont la solution "exacte" est

$$u_1 = 1{,}00010 \ldots \doteqdot 1, \quad u_2 = 0{,}99990 \ldots \doteqdot 1.$$

On peut prendre le nombre $a_{11} = 10^{-4}$ comme pivot puisqu'il n'est pas nul, ce qui conduit à la procédure suivante d'élimination :

$$\begin{aligned} (I_1) = (I_2) \quad & 10^{-4}u_1 + u_2 = 1, \\ -10^4(I_1) + (II_1) = (II_2) \quad & \qquad -9\,990u_2 = -9\,990, \end{aligned}$$

(1) FORSYTHE G. E., MOLER C. B. — *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, Englewood Cliffs, 1967.

puisque les nombres $(-10^4+1) = -9\,999$ et $(-10^4+2) = -9\,998$ sont arrondis chacun au même nombre $-9\,990$. La "solution" trouvée de cette façon :

$$u_2 = 1, \quad u_1 = 0,$$

est très éloignée de la véritable solution !

Par contre, si l'on commence par *échanger* les deux équations, c'est-à-dire si le pivot est l'élément $a_{21} = 1$, on est conduit aux calculs suivants :

$$\begin{array}{rr} (\mathrm{I}_1') & [1]\ u_1+u_2 = 2, \\ (\mathrm{II}_1') & 10^{-4}u_1+u_2 = 1, \\ (\mathrm{I}_1') = (\mathrm{I}_2') & u_1+u_2 = 2, \\ -10^{-4}(\mathrm{I}_1')+(\mathrm{II}_1') = (\mathrm{II}_2') & 0{,}999u_2 = 0{,}999, \end{array}$$

puisque les nombres $(-10^{-4}+1) = 0{,}9999$ et $(-2\times 10^{-4}+1) = 0{,}9998$ sont arrondis au même nombre 0,999. La "solution" correspondante :

$$u_2 = 1, \quad u_1 = 1,$$

est cette fois très satisfaisante !

Cet exemple montre que des erreurs d'arrondis à l'effet désastreux proviennent de la *division par des pivots "trop petits"*. C'est pourquoi, *pratiquement*, on utilise l'une des deux stratégies suivantes, au début de la k-ème étape, $1 \leqslant k \leqslant n-1$, de l'élimination :

—*Stratégie du pivot partiel.* Le pivot est l'un des éléments a_{ik}^k, $k \leqslant i \leqslant n$, vérifiant

$$|a_{ik}^k| = \max_{k\leqslant p\leqslant n} |a_{pk}^k| \ ;$$

—*Stratégie du pivot total.* Le pivot est l'un des éléments a_{ij}^k, $k \leqslant i,\ j \leqslant n$, vérifiant

$$|a_{ij}^k| = \max_{k\leqslant p,\ q\leqslant n} |a_{pq}^k|.$$

Si le pivot choisi par cette stratégie n'est pas dans la k-ème colonne, il faut donc également effectuer *un échange de colonnes* (en plus d'un échange de lignes). Cette stratégie est donc semblable, mais non identique, à la procédure d'élimination que nous avons décrite, puisqu'une telle opération équivaut à multiplier la matrice A_k également *à droite* par une matrice de transposition (à cet égard, une indication est donnée à l'exercice 4.3-4).

Renvoyant les lecteurs intéressés par ce genre de questions aux ouvrages spécialisés cités dans les Commentaires Bibliographiques, nous retiendrons simplement ceci : *une stratégie de pivot* (partiel ou total) *est indispensable si l'on veut éviter de trop grandes erreurs d'arrondi lors de l'application de la méthode de Gauss à des systèmes linéaires de matrices "quelconques". Par contre, dans certains cas particuliers, il est inutile d'avoir une stratégie de pivots* ; c'est le cas notamment des systèmes linéaires à *matrices symétriques définies positives* (voir notamment ([1]), page 220), que nous étudions au paragraphe 4.4.

Comptons enfin le nombre d'opérations élémentaires requises dans la méthode de Gauss.

(i) *Élimination.* Pour passer de la matrice A_k à la matrice A_{k+1}, $1 \leqslant k \leqslant n-1$, on effectue $(n-k)$ divisions, $(n-k+1)(n-k) = (n-k)^2+(n-k)$ additions et multiplications, soit au total

$$\begin{cases} (n-1)^2+(n-2)^2+\ \dots\ +1^2+(n-1)+(n-2)+\ \dots\ +1 = \dfrac{n^3-n}{3} & \text{additions,} \\ (n-1)^2+(n-2)^2+\ \dots\ +1^2+(n-1)+(n-2)+\ \dots\ +1 = \dfrac{n^3-n}{3} & \text{multiplications,} \\ (n-1)+(n-2)+\ \dots\ +1 = \dfrac{n(n-1)}{2} & \text{divisions.} \end{cases}$$

([1]) WILKINSON J. H. — *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

(ii) *Pour passer du vecteur* $E_{k-1}P_{k-1}\dots E_1P_1b$ *au vecteur* $E_kP_k\dots E_1P_1b$, on effectue $(n-k)$ additions et multiplications, soit au total

$$\begin{cases} (n-1)+(n-2)+\ \dots\ +1 = \dfrac{n(n-1)}{2} & \text{additions,} \\ (n-1)+(n-2)+\ \dots\ +1 = \dfrac{n(n-1)}{2} & \text{multiplications.} \end{cases}$$

(iii) La *remontée* nécessite (*cf.* paragraphe 4.1)

$$\begin{cases} \dfrac{n(n-1)}{2} & \text{additions,} \\ \dfrac{n(n-1)}{2} & \text{multiplications,} \\ n & \text{divisions.} \end{cases}$$

En fin de compte, *la méthode de Gauss nécessite donc de l'ordre de*

$$\begin{cases} \dfrac{n^3}{3} & \text{additions,} \\ \dfrac{n^3}{3} & \text{multiplications,} \\ \dfrac{n^2}{2} & \text{divisions.} \end{cases}$$

REMARQUE. Dans l'estimation du temps total de calcul (on rappelle à cet égard qu'on a donné au paragraphe 2.1 des ordres de grandeurs des durées comparées de chaque opération élémentaire), il faut également prendre en compte le temps dû à la recherche du pivot, *notamment lorsqu'on emploie la stratégie du pivot total.* ■

On a cru pendant un certain temps que la fonction $\dfrac{2n^3}{3}$ était une borne inférieure de la partie principale du nombre total d'opérations élémentaires nécessaires à la résolution d'un système linéaire quelconque par une *méthode directe* (quelle qu'elle soit). Bien que la question ne soit pas encore complètement résolue à l'heure actuelle, il semble néanmoins que l'exposant 3 n'est pas loin d'être optimal, et *c'est précisément cette observation qui justifie l'emploi de la méthode de Gauss dans tous les cas où la matrice du système est "quelconque"*. A cet égard, voir Pan (1984).

Il est très instructif de comparer le nombre d'opérations élémentaires de la méthode de Gauss avec le nombre des opérations élémentaires nécessaires à l'application des *formules de Cramer* :

$$u_i = \frac{\text{dét}\,(B_i)}{\text{dét}\,(A)}, \quad \text{où} \quad B_i = \begin{pmatrix} a_{11} & \dots & a_{1,i-1} & b_1 & a_{1,i+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2,i-1} & b_2 & a_{2,i+1} & \dots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n,i-1} & b_n & a_{n,i+1} & \dots & a_{nn} \end{pmatrix},$$

pour lesquelles on doit évaluer $(n+1)$ déterminants et effectuer n divisions. Or le calcul "brutal" d'un déterminant exigeant $\{(n!)-1\}$ additions et $(n-1)n!$ multiplications, l'utilisation des formules de Cramer recquiert donc de l'ordre de

$$\begin{cases} (n+1)! & \text{additions,} \\ (n+2)! & \text{multiplications,} \\ n & \text{divisions.} \end{cases}$$

Pour $n = 10$ par exemple, on obtient environ :

$$\begin{cases} 700 \text{ opérations pour la méthode de Gauss,} \\ 400\,000\,000 \text{ opérations pour la méthode de Cramer.} \\ \qquad\qquad\text{No comment} \ldots \end{cases}$$

On retiendra que la méthode de Gauss est la méthode la plus couramment utilisée pour résoudre les systèmes linéaires dont les matrices n'ont pas de propriétés particulières. En particulier cette méthode s'emploie pour les systèmes à *matrices pleines.*

Une méthode voisine dans son principe de la méthode de Gauss est la *méthode de Gauss-Jordan :* au lieu de chercher une matrice M telle que la matrice MA soit triangulaire supérieure (A : matrice inversible donnée), *on cherche une matrice* $\tilde{M}$ *telle que la matrice* $\tilde{M}A$ *soit diagonale.* Le résultat de la $(k-1)$-ème étape est alors une matrice de la forme (comparer avec la procédure d'élimination de la méthode de Gauss):

$$\tilde{A}_k = \tilde{E}_{k-1}\tilde{P}_{k-1} \ldots \tilde{E}_2\tilde{P}_2\tilde{E}_1\tilde{P}_1 A = \begin{pmatrix} a_{11}^k & & & a_{1k}^k & \ldots & a_{1n}^k \\ & a_{22}^k & & a_{2k}^k & \ldots & a_{2n}^k \\ & & \ddots & \vdots & & \vdots \\ & & & a_{kk}^k & \ldots & a_{kn}^k \\ & & & \vdots & & \vdots \\ & & & a_{nk}^k & \ldots & a_{nn}^k \end{pmatrix}.$$

Comme dans la méthode de Gauss, l'un des éléments a_{ik}^k, $k \leqslant i \leqslant n$, est différent de zéro (la matrice A étant inversible) et on le choisit comme *pivot*, ce qui revient à multiplier la matrice $\tilde{A}_k$ à gauche par une matrice de permutation $\tilde{P}_k$ telle que l'élément α_{kk}^k (c'est-à-dire le pivot choisi) de la matrice $(\alpha_{ij}^k) = \tilde{P}_k\tilde{A}_k$ soit différent de zéro. La matrice $\tilde{E}_k$ est alors de la forme

$$\tilde{E}_k = \begin{pmatrix} 1 & & & -\pi_k^{-1}\alpha_{1k}^k & & & \\ & \ddots & & \vdots & & & \\ & & 1 & -\pi_k^{-1}\alpha_{k-1,k}^k & & & \\ & & & 1 & & & \\ & & & -\pi_k^{-1}\alpha_{k+1,k}^k & 1 & & \\ & & & \vdots & & \ddots & \\ & & & -\pi_k^{-1}\alpha_{nk}^k & & & 1 \end{pmatrix}, \quad \pi_k = \alpha_{kk}^k,$$

de sorte qu'après la $(n-1)$-ème étape, la matrice

$$\tilde{A}_n = \tilde{M}A, \quad \text{avec} \quad \tilde{M} = \tilde{E}_{n-1}\tilde{P}_{n-1} \ldots \tilde{E}_2\tilde{P}_2\tilde{E}_1\tilde{P}_1,$$

est diagonale.

La méthode de Gauss-Jordan est utilisée pour le calcul de l'inverse d'une matrice donnée on résout *simultanément* les n systèmes linéaires

$$Au_j = e_j, \quad 1 \leqslant j \leqslant n,$$

en appliquant à chaque second membre e_j les transformations (échanges et combinaisons linéaires de lignes) représentées par les matrices $\tilde{P}_k$ et $\tilde{E}_k$. On obtient ainsi

$$\tilde{A}_n u_j = \tilde{M} e_j, \quad 1 \leqslant j \leqslant n,$$

et chacun de ces systèmes est de résolution immédiate puisque la matrice $\tilde{A}_n$ est diagonale.

4.3. La factorisation LU d'une matrice

Supposons qu'on ne s'intéresse plus à l'effet des erreurs d'arrondi dans la méthode de Gauss et que, par conséquent, on ne cherche pas à appliquer de stratégie particulière de pivot. Autrement dit, si l'élément a_{11}^1 de la matrice $A_1 = A$ est différent de zéro, on le prend comme pivot ; si l'élément a_{22}^2 de la matrice $A_2 = E_1 A_1$ (ici $P_1 = I$) est différent de zéro, on le prend comme pivot ; et ainsi de suite .... S'il est possible de choisir de proche en proche $P_1 = P_2 = \ldots = P_{n-1} = I$ (les pivots étant par conséquent les éléments a_{kk}^k des matrices A_k), la matrice

$$M = E_{n-1} \ldots E_2 E_1,$$

produit de matrices triangulaires inférieures, est elle-même triangulaire inférieure (vérification immédiate), ainsi que la matrice

$$L \overset{\text{déf}}{=} M^{-1}$$

(*cf.* paragraphe 4.1), et *la matrice* A *s'écrit comme le produit* A = LU *d'une matrice triangulaire inférieure* L *et d'une matrice triangulaire supérieure* U. En résumé,

$$P_1 = P_2 = \ldots = P_{n-1} = I \Rightarrow A = LU, \quad \text{avec} \quad \begin{cases} L = (E_{n-1} \ldots E_2 E_1)^{-1}, \\ U = (E_{n-1} \ldots E_2 E_1) A. \end{cases}$$

REMARQUE. Les notations L et U sont d'origine anglo-saxonne : "L" pour "lower" (triangular) et "U" pour "upper" (triangular). ■

On sait déjà comment effectuer le calcul de la matrice U, puisque

$$U = A_n = \begin{pmatrix} a_{11}^1 & a_{12}^1 & \ldots & a_{1n}^1 \\ & a_{22}^2 & \ldots & a_{2n}^2 \\ & & \ddots & \vdots \\ & & & a_{nn}^n \end{pmatrix},$$

où (rappel) :

$$(a_{ij}^k) = E_{k-1} \ldots E_2 E_1 A = A_k, \quad 1 \leqslant k \leqslant n-1.$$

Il est tout à fait remarquable que la matrice L *se calcule immédiatement à partir des matrices* E_k, $1 \leqslant k \leqslant n-1$. Celles-ci étant en effet de la forme

$$E_k = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & -l_{k+1,k} & 1 & & \\ & & \vdots & & \ddots & \\ & & -l_{nk} & & & 1 \end{pmatrix}, \quad \text{avec} \quad l_{ik} \overset{\text{déf}}{=} \frac{a_{ik}^k}{a_{kk}^k}, \quad k+1 \leqslant i \leqslant n,$$

il est facile de vérifier que (attention aux changements de signes sur les éléments l_{ik} !)

$$(\mathrm{E}_k)^{-1} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & l_{k+1,k} & 1 & & \\ & & \vdots & & \ddots & \\ & & l_{nk} & & & 1 \end{pmatrix},$$

$$\mathrm{L} = (\mathrm{E}_1)^{-1}(\mathrm{E}_2)^{-1} \ldots (\mathrm{E}_{n-1})^{-1} = \begin{pmatrix} 1 & & & & \\ l_{21} & 1 & & & \\ l_{31} & l_{32} & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & l_{n3} & \ldots & 1 \end{pmatrix}.$$

Ainsi, dans l'exemple numérique du paragraphe 4.2 (où il n'y a pas eu d'échanges de lignes), on a trouvé

$$\mathrm{E}_1 = \begin{pmatrix} 1 & & \\ -1 & 1 & \\ \frac{4}{5} & & 1 \end{pmatrix}, \quad \mathrm{E}_2 = \begin{pmatrix} 1 & & \\ & 1 & \\ & \frac{9}{20} & 1 \end{pmatrix}, \quad \mathrm{A}_3 = \begin{pmatrix} 5 & 2 & 1 \\ & -8 & 1 \\ & & \frac{9}{4} \end{pmatrix},$$

ce qui permet d'écrire

$$\mathrm{A} = (\mathrm{E}_2\mathrm{E}_1)^{-1}\mathrm{A}_3 = \begin{pmatrix} 1 & & \\ 1 & 1 & \\ -\frac{4}{5} & -\frac{9}{20} & 1 \end{pmatrix} \begin{pmatrix} 5 & 2 & 1 \\ & -8 & 1 \\ & & \frac{9}{4} \end{pmatrix} = \mathrm{LU}.$$

Attention ! Si la matrice L s'obtient immédiatement à partir des matrices E_k, il n'en va pas de même de la matrice $\mathrm{M} = \mathrm{L}^{-1} = \mathrm{E}_{n-1} \ldots \mathrm{E}_2\mathrm{E}_1$, pour laquelle il n'existe pas d'expression simple des éléments à partir de ceux des matrices E_k. Mais, comme on l'a déjà signalé, il n'y a pas de raison de calculer la matrice M ....

Il reste à donner des conditions *suffisantes* pour qu'il n'y ait pas lieu d'effectuer d'échanges de lignes dans la méthode de Gauss, c'est-à-dire pour qu'on puisse "factoriser" une matrice sous la forme du produit d'une matrice triangulaire inférieure L par une matrice triangulaire supérieure U. On va même établir qu'une telle *factorisation* LU est unique si l'on impose la valeur 1 aux éléments diagonaux de la matrice L (c'est précisément la valeur trouvée avec la construction précédente), ce qui permet de définir sans ambiguïté une telle factorisation (pour un complément, voir l'exercice 4.3-1).

Théorème 4.3-1 (*factorisation LU d'une matrice*). ***Soit*** $\mathrm{A} = (a_{ij})$ ***une matrice carrée d'ordre n telle que les n sous-matrices diagonales***

$$\Delta_k = \begin{pmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{pmatrix}, \quad 1 \leqslant k \leqslant n,$$

soient inversibles.

Alors il existe une matrice triangulaire inférieure $L = (l_{ij})$ ***avec*** $l_{ii} = 1$, $1 \leqslant i \leqslant n$, ***et une matrice triangulaire supérieure*** U ***telles que***

$$A = LU.$$

De plus, une telle factorisation est unique.

DÉMONSTRATION. Puisque $a_{11} \neq 0$; on peut choisir $P_1 = I$. Supposons qu'on ait pu choisir

$$P_1 = P_2 = \dots = P_{k-1} = I,$$

de sorte que l'égalité

$$(E_{k-1} \dots E_2E_1)A = A_k$$

s'écrit

$$\begin{pmatrix} \boxed{\begin{matrix} 1 & & & \\ \times & 1 & & \\ \vdots & \vdots & \ddots & \\ \times & \times & \dots & 1 \end{matrix}} & & \\ \vdots \quad \vdots & \ddots & \\ \times \quad \times \quad \dots & & 1 \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix} a_{11} & \dots & a_{1k} \\ \vdots & \Delta_k & \vdots \\ a_{k1} & \dots & a_{kk} \end{matrix}} & \dots & \times \\ \vdots & & \vdots \\ \times \quad \dots & & \times \end{pmatrix} = \begin{pmatrix} \boxed{\begin{matrix} a_{11}^1 & \dots & a_{1k}^1 \\ & \ddots & \vdots \\ & & a_{kk}^k \end{matrix}} & \dots & \times \\ \vdots & & \vdots \\ \times & \dots & \times \end{pmatrix}$$

Utilisant les règles de la multiplication par blocs des matrices en tenant compte de la forme particulière des matrices, on obtient

$$\text{dét}\,(\Delta_k) = a_{11}^1 \dots a_{kk}^k.$$

Comme $\Delta_k \neq 0$ par hypothèse, on déduit que l'élément a_{kk}^k est non nul. On peut donc le choisir comme pivot, ce qui équivaut au choix $P_k = I$.

L'*existence* d'une factorisation LU possédant les propriétés annoncées dans le théorème se trouve donc établie en posant

$$A = (E_{n-1} \dots E_2E_1)^{-1}(E_{n-1} \dots E_2E_1A) \overset{\text{déf}}{=} LU.$$

Démontrons son *unicité* : de l'existence de deux telles factorisations

$$A = L_1U_1 = L_2U_2,$$

on déduit

$$L_2^{-1}L_1 = \begin{pmatrix} 1 & & & \\ \times & 1 & & \\ \vdots & \vdots & \ddots & \\ \times & \times & \dots & 1 \end{pmatrix} = \begin{pmatrix} \times & \times & \dots & \times \\ & \times & \dots & \times \\ & & \ddots & \vdots \\ & & & \times \end{pmatrix} = U_2U_1^{-1}.$$

Or cette égalité de matrices n'est possible que si $L_2^{-1}L_1 = U_2U_1^{-1} = I$, soit $L_1 = L_2$ et $U_1 = U_2$. ■

Un cas important où les conditions d'application du théorème précédent se trouvent vérifiées est celui où la matrice A est *symétrique définie positive*, mais comme on va le voir, on profite de la symétrie pour modifier légèrement (en la simplifiant) la factorisation LU. C'est pourquoi nous traitons ce cas séparément, au paragraphe suivant.

Un *intérêt majeur* de l'existence d'une factorisation LU est le suivant : si l'on a à résoudre *plusieurs* systèmes linéaires correspondant à la *même* matrice A, il suffit de conserver l'expression des deux matrices L et U une fois celles-ci calculées "une première