Sur la décomposition LU explicite

Jean-François Burnol, mars 2017

1 Introduction

Je considère une matrice A rectangulaire avec m lignes et n colonnes.

La méthode [1] du pivot de Gauss pour la réduction à une forme échelonnée est un algorithme fondamental dont la discussion est intimement liée à la notion de *rang* d'une matrice. Le *rang* de A est la dimension de l'espace engendré par les lignes et il est aussi la dimension de l'espace engendré par les colonnes. Le fait que les deux soient identiques n'est pas évident *a priori* et l'un des points les plus intéressants mis en évidence par la forme échelonnée.

En effet les opérations de lignes sont réversibles et donc à chaque étape l'espace vectoriel engendré par les lignes est invariant. À la fin de la réduction les lignes non nulles sont linéairement indépendantes à cause de la forme échelonnée. Donc la dimension de l'espace des lignes est le nombre de pivots (et on a obtenu une base, qui n'est cependant pas une base extraite des lignes initiales).

Au niveau des colonnes, un peu d'abstraction aide : les opérations de lignes correspondent à la multiplication sur la gauche par des matrices inversibles. Chaque colonne est transformée en une nouvelle colonne par une même application linéaire. Cela signifie qu'à chaque étape l'espace engendré par les colonnes peut changer, mais il est relié par une transformation linéaire inversible à l'espace des colonnes de l'étape précédente. À la fin il est clair que les colonnes contenant les pivots sont linéairement indépendantes et que toute autre colonne en est une combinaison linéaire. Donc les colonnes d'origine à la position des pivots forment une base (extraite) de l'espace initial des colonnes. Sa dimension est donc à nouveau égale au nombre des pivots.

Nous avons donc établi l'égalité de dimensions entre l'espace engendré par les lignes et celui engendré par les colonnes.

1. Il est abusif de parler de « la » méthode des pivots ; il y a plusieurs façons de gérer le cas où l'on doit échanger des lignes, et c'est même de nos jours un sujet de publication en algorithmique numérique, en particulier dans la perspective de la « parallélisation » des calculs.

Dans cette fiche je veux insister sur un aspect très concret et paradoxalement en déduire des formules explicites d'un intérêt plus théorique que pratique (encore que...).

2 Notations

Écrire des matrices ou des déterminants en LaTeX est assez pénible, et pour ne pas y passer des heures j'adopterai des notations plus concises. Surtout qu'il va me falloir écrire des matrices dont les éléments sont des déterminants, voire des quotients de déterminants.

En particulier si M est une matrice quelconque je noterai $M(i;k,l)$ la sous-matrice de taille $(i+1)\times(i+1)$ qui est obtenue en supprimant toutes les colonnes autres que les i premières et la l^e (on suppose $l>i$) ainsi que toutes les lignes autres que les i premières et la k^e (on suppose $k>i$). Faites un dessin pour bien visualiser.

Par exemple pour $i=0$, $M(0;k,l)$ est le coefficient de M en position (k,l), que je noterai $m_{k,l}$. Dans toute cette fiche les lignes sont numérotées à partir de 1 en partant du haut et les colonnes à partir de 1 en partant de la gauche.

Pour $i=1$, $M(1;k,l)$ est la matrice $\begin{pmatrix} m_{1,1} & m_{1,l} \\ m_{k,1} & m_{k,l} \end{pmatrix}$.

Toutes nos sous-matrices $M(i;k,l)$ sont des matrices carrées et je noterai $m(i;k,l)$ le déterminant de $M(i;k,l)$ (qui est donc un mineur de M). Je rappelle que $m(i;k,l)$ n'est défini que pour $k>i$ et $l>i$. Le cas $k=l=i+1$ nous intéressera particulièrement et je vais carrément adopter la notation $M(i+1)$ alors, et $m(i+1)$ pour son déterminant qui est donc le $(i+1)^e$ mineur principal fondamental. J'aurai besoin aussi de la notation $m(0)=1$.

Il y a exactement $\min(m,n)$ mineurs principaux fondamentaux.

3 Déroulement de l'algorithme dans le cas générique

Bon, donc partons d'une matrice A de taille $m\times n$ et appliquons la méthode de réduction par lignes. À chaque étape je vais supposer être dans la situation générique.

Si A n'a qu'une seule ligne, il n'y a rien à faire ! plus précisément, on désigne le premier coefficient non nul, et ce sera $a(1,1)$ génériquement, comme un « pivot ». Mais nous ne faisons rien de spécial avec ce pivot.

Supposons donc $m > 1$.

Première étape : le coefficient $a(1,1)$ est génériquement non nul et va servir de pivot. Pour chaque $i > 1$ on retire de la i^e ligne $\ell_{i,1} = a(i,1)/a(1,1)$ fois la première ligne. On note T_1 la matrice carrée de taille $m \times m$ qui ne diffère de la matrice identité que par sa première colonne ${}^t[1, -\ell_{2,1}, -\ell_{3,1}, \ldots, -\ell_{m,1}]$. Alors A a été transformée en $A_1 = T_1 A$.

Je suppose maintenant $m > 1$. Il y a trois possibilités :

1. Il n'y a dans A qu'une seule colonne. L'algorithme s'arrête là.
2. A a au moins deux colonnes mais seulement deux lignes. L'algorithme s'arrête là, enfin après avoir déclaré $a_1(2,2)$ comme un pivot (il sera non nul, génériquement, cf plus bas) mais on ne fait rien avec.
3. A a au moins deux colonnes et au moins trois lignes. On continue.

Deuxième étape : le coefficient $a_1(2,2)$ est génériquement non nul. Pourquoi ? eh bien la sous-matrice principale de A de taille 2×2 a été multipliée par une matrice 2×2 triangulaire inférieure avec des 1 sur la diagonale, donc le déterminant n'a pas changé, et ainsi $a_1(1,1)a_1(2,2) = a(2)$ (voir plus haut pour mes notations condensées). Génériquement $a(2)$ est non nul, d'où la conclusion.

Donc on pose, pour chaque $i > 2$, (il en existe par hypothèse !) $\ell_{i,2} = a_1(i,2)/a_1(2,2)$, et l'on note T_2 la matrice carrée de taille $m \times m$ qui ne diffère de la matrice identité que par sa deuxième colonne ${}^t[0, 1, -\ell_{3,2}, -\ell_{4,2}, \ldots, -\ell_{m,2}]$. Alors A_1 est transformée en $A_2 = T_2 A_1 = T_2 T_1 A$.

Je suppose maintenant $m > 2$. Il y a trois possibilités :

1. Il n'y a dans A que deux colonnes. L'algorithme s'arrête là.
2. $n \geqslant 3$ et $m = 3$, l'algorithme s'arrête là. Simplement on déclare $a_2(3,3)$ comme étant un pivot (il est non nul, sous notre hypothèse de généricité), mais on ne fait rien avec de spécial.
3. $n \geqslant 3$ et $m > 3$. On continue.

Troisième étape : le coefficient $a_2(3,3)$ est génériquement non nul. Pourquoi ? eh bien la sous-matrice principale de A de taille 3×3 a été multipliée par une matrice 3×3 triangulaire inférieure avec des 1 sur la diagonale, donc le déterminant n'a pas changé, et ainsi $a_2(1,1)a_2(2,2)a_2(3,3) = a(3)$ (voir plus haut pour mes notations condensées). Génériquement $a(3)$ est non nul, d'où la conclusion.

Donc on pose, pour chaque $i > 3$, (il en existe par hypothèse !) $\ell_{i,3} = a_2(i,3)/a_2(3,3)$, et l'on note T_3 la matrice carrée de taille $m \times m$ qui ne diffère de la matrice identité que par sa troisième colonne ${}^t[0,0,1,-\ell_{4,3},\dots,-\ell_{m,3}]$. Alors A_2 est transformée en $A_3 = T_3A_2 = T_3T_2T_1A$.

Je répète que je me situe ici dans la situation générique où je trouve toujours un pivot dans la colonne suivante, immédiatement à la première ligne suivante.

Quand est-ce que l'algorithme s'arrête dans cette situation générique ? soit parce qu'il n'y a plus de lignes à modifier, soit parce qu'il n'y a plus de colonne disponible pour y trouver un pivot. Comme les pivots sont le long de la diagonale, cela veut dire que l'on arrête avec $r = \min(m,n)$ pivots. Il y a eu exactement $r-1$ étapes, et on a transformé A en $U = A_{r-1} = T_{r-1}\cdots T_1A$. La matrice U n'a que r lignes non nulles. Elle est triangulaire supérieure. Sur sa diagonale principale les coefficients sont non nuls. La matrice $T = T_{r-1}\cdots T_1$ est une matrice carrée $m \times m$, triangulaire inférieure avec des 1 sur la diagonale.

Dans le cas où l'on avait plus de lignes que de colonnes ($m > n$, $r = n$), la matrice T n'a que ses r premières colonnes qui sont intéressantes, ensuite on a juste un bloc de taille $(m-n) \times (m-n)$ qui est une matrice identité. La matrice U dans ce cas à la forme d'une matrice rectangulaire $m \times r$ avec ses $m-r$ dernières lignes identiquement nulles et en haut un bloc $r \times r$ triangulaire supérieur inversible.

Ainsi

$$A = \underbrace{T_1^{-1}\dots T_{r-1}^{-1}}_{L} \cdot U$$

4 L, c'est mieux que T

Il y a une bonne raison pour laquelle on parle de décomposition $A = LU$ et pas de décomposition $TA = U$.

En effet revenons par exemple à T_2T_1. Multiplier à gauche par T_2 cela signifie modifier les lignes d'indices $i > 1$ en leur soustrayant des multiples de la première ligne. Donc T_2T_1 n'a PAS la même première colonne que T_1.

Mais regardons $T_1^{-1}T_2^{-1}$. D'abord c'est quoi T_1^{-1} ? On a dit que T_1 est la matrice carrée de taille $m \times m$ qui ne diffère de la matrice identité que par sa première colonne ${}^t[1, -\ell_{2,1}, -\ell_{3,1}, \ldots, -\ell_{m,1}]$. La multiplication à gauche par T_1 remplace chaque i^e ligne par elle-même moins $\ell_{i,1}$ fois la première ligne. L'opération inverse est de remplacer chaque i^e ligne par elle-même plus $\ell_{i,1}$ fois la première ligne. Donc T_1^{-1} est la matrice carrée de taille $m \times m$ qui ne diffère de la matrice identité que par sa première colonne ${}^t[1, +\ell_{2,1}, +\ell_{3,1}, \ldots, +\ell_{m,1}]$.

De même T_2^{-1} est la matrice carrée de taille $m\times m$ qui ne diffère de la matrice identité que par sa deuxième colonne ${}^t[0, 1, +\ell_{3,2}, +\ell_{4,2}, \ldots, +\ell_{m,2}]$.

Lorsque je multiplie T_2^{-1} à gauche par T_1^{-1}, je modifie les lignes avec $i > 1$ par un multiple de la première ligne : mais cette première ligne est juste $[1, 0, \ldots, 0]$! Donc je ne fais que modifier la première colonne de T_2^{-1} qui de celle de la matrice identité devient exactement la première colonne de T_1^{-1}.

Plus généralement en examinant la construction de

$$L = T_1^{-1} \ldots T_{r-1}^{-1}$$

de la droite vers la gauche, on voit qu'on ne fait que mettre à jour les colonnes de la droite vers la gauche. Bien sûr, nous ou l'ordinateur on va simplement remplir les colonnes de L de la gauche vers la droite au fur et à mesure de l'avancement de l'algorithme et de la construction de L et de l'évolution de A vers sa transformée échelonnée U.

Donc L ne nécessite AUCUN calcul supplémentaire. La matrice L est la matrice carrée triangulaire inférieure avec des 1 sur la diagonale dont les entrées $\ell_{i,j}$, $(i > j)$ sont les coefficients qui sont apparus au cours du déroulement de l'algorithme.

Maintenant on pourrait m'objecter (lorsque $n = m$ et aussi U est une matrice carrée inversible) que lorsque je veux résoudre

$$A\mathbf{x} = L \cdot U\mathbf{x} = \mathbf{y}$$

je vais passer par

$$\mathbf{x} = U^{-1} \cdot L^{-1}\mathbf{y}$$

et donc j'ai besoin de $T = L^{-1}$.

FAUX ! la matrice L est triangulaire inférieure, le système $L\mathbf{z} = \mathbf{y}$ se résout en cascade $z_1 = y_1$ etc... et connaître explicitement la matrice L^{-1} n'apporte RIEN de mieux au niveau de l'efficacité des calculs. On n'a PAS besoin de connaître explicitement cette matrice.

Et ça tombe bien, car la matrice que l'on a obtenue facilement c'est L pas $T = L^{-1}$.

5 Des formules explicites

Je veux une formule exacte pour $L_{i,j} = \ell_{i,j} = \ell(i,j)$, $i > j$.

Suffit de demander. On calcule $\ell(i,j)$ en examinant $A_{j-1} = T_{j-1} \dots T_1 \cdot A$.

Les $j-1$ premières colonnes de A_{j-1} ont leurs m lignes échelonnées. Notons $d_1, d_2, \dots, d_{j-1}$ les éléments diagonaux (ils sont, génériquement, non nuls). En-dessous il n'y a que des zéros.

La j^e ligne de A_{j-1} débute donc par $0, 0, .., 0$ jusqu'au j^e coefficient que je vais noter $\lambda_j = d_j$. Il est, génériquement, non-nul.

La i^e ligne débute par $0, 0, .., 0$ jusqu'au j^e coefficient que je vais noter λ_i.

On a $\ell(i,j) = \lambda_i/\lambda_j$. On est passé de A à A_{j-1} par des manipulations de lignes. Il est fondamental de comprendre que la i^e ligne n'a jusqu'à présent été modifiée qu'en lui retirant une combinaison linéaire des $j-1$ premières lignes !

D'où les deux équations puisque les déterminants restent invariants par ces transformations :

$$d_1 d_2 \dots d_{j-1} \lambda_j = a(j-1; j, j) = a(j)$$
$$d_1 d_2 \dots d_{j-1} \lambda_i = a(j-1; i, j)$$

et notre formule EXACTE pour :

$$i > j \implies L_{i,j} = \ell(i,j) = \frac{a(j-1; i, j)}{a(j)}$$

qui exprime DIRECTEMENT les coefficients de L comme des quotients de mineurs de la matrice d'origine A.

Revenons aux éléments diagonaux de la matrice U. Ils vérifient par ce qui précède :

$$\begin{aligned} d_1 &= a(1) \\ d_1 d_2 &= a(2) \\ d_1 d_2 d_3 &= a(3) \\ \ldots &= \ldots \\ d_1 \ldots d_r &= a(r) \end{aligned}$$

D'où les formules (je rappelle $a(0) = 1$ par convention)

$$u_{1,1} = d_1 = \frac{a(1)}{a(0)}, \quad u_{2,2} = d_2 = \frac{a(2)}{a(1)}, \quad \ldots \quad , u_{r,r} = \frac{a(r)}{a(r-1)}$$

Maintenant on veut les formules pour $u_{i,j}$, $j > i$ puisque tous les autres coefficients en dehors de la diagonale sont nuls.

Le coefficient $u_{i,j}$ est situé sur la i^e ligne. Cette ligne ne bouge plus après la $(i-1)^e$ étape. Lors des étapes précédentes on a modifié les lignes en leur soustrayant des combinaisons linéaires de celles au-dessus. Donc le mineur $u(i-1; i, j)$, $j \geqslant i$, est identique au mineur initial $a(i-1; i, j)$.

En particulier pour $j = i$ on retrouve notre argument que le mineur $u(i) = d_1 d_2 \ldots d_i$ vaut $a(i)$.

Et pour $j > i$ on peut faire la remarque que $u(i-1; i, j) = d_1 d_2 \ldots d_{i-1} u_{i,j}$ à cause de la forme triangulaire.

Nous obtenons l'équation :

$$\frac{u(i-1; i, j)}{u(i-1; i, i)} = \frac{d_1 d_2 \ldots d_{i-1} u_{i,j}}{d_1 d_2 \ldots d_i}$$

puis la formule

$$u_{i,j} = d_i \frac{a(i-1; i, j)}{a(i-1; i, i)} = \frac{a(i)}{a(i-1)} \frac{a(i-1; i, j)}{a(i)} = \frac{a(i-1; i, j)}{a(i-1)}$$

D'où ce beau théorème :

Théorème 1. *Soit* A *une matrice rectangulaire* $m \times n$ *et* $r = \min(m, n)$. *À condition que les* r *mineurs principaux de* A *soient non nuls, on a la factorisation*

$$A = L \cdot U \tag{1}$$

avec L *une matrice carrée* $m \times m$ *triangulaire inférieure, avec des* 1 *sur la diagonale, dont les coefficients sont*

$$j \leqslant r \quad \& \quad i < j \implies L_{i,j} = 0 \tag{2a}$$

$$j \leqslant r \quad \& \quad i = j \implies L_{i,i} = 1 \tag{2b}$$

$$j \leqslant r \quad \& \quad i > j \implies L_{i,j} = \ell_{i,j} = \frac{a(j-1; i, j)}{a(j)} \tag{2c}$$

$$j > r \implies L_{i,j} = \delta(i, j) \tag{2d}$$

et U *une matrice rectangulaire* $m \times n$, *triangulaire supérieure, dont les coefficients sont donnés par :*

$$i \leqslant r \quad \& \quad j < i \implies u_{i,j} = 0 \tag{3a}$$

$$i \leqslant r \quad \& \quad j \geqslant i \implies u_{i,j} = \frac{a(i-1; i, j)}{a(i-1)} \tag{3b}$$

$$i > r \implies u_{i,j} = 0 \tag{3c}$$

En particulier les coefficients diagonaux sont les quotients successifs des mineurs principaux de A. *(rappel :* $a(0) = 1$*)*

Il faut peut-être insister que L n'est pas uniquement déterminée lorsqu'il y a plus de lignes que de colonnes ($m > n$). Seul le bloc $n \times n$ en haut à gauche de L l'est. Mais les formules ci-dessus donnent la solution unique telle que le bloc $(m-n) \times (m-n)$ est une matrice identité.

6 Les formulations LDU

Il est assez naturel de mettre en facteur la matrice $r \times r$ des éléments diagonaux de U. On obtient le théorème suivant :

Théorème 2. *Soit* A *une matrice* $m \times n$ *dont tous les mineurs principaux* $a(1)$, ..., $a(r)$, $r = \min(m, n)$ *sont non nuls. Alors*

$$
\mathtt{A} = \begin{pmatrix}
1 & & & & & & \\
\cdots & 1 & & & & & \\
\vdots & \vdots & \ddots & & & & \\
\vdots & \vdots & \vdots & \ddots & & & \\
\cdots & \cdots & \frac{a(j-1;i,j)}{a(j)} & \cdots & 1 & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1
\end{pmatrix} \cdot \left(\begin{array}{c|c} \mathbf{D} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right) \cdot \begin{pmatrix}
1 & \cdots & \cdots & \cdots & \cdots & & \\
 & \ddots & \vdots & \vdots & & & \\
 & & 1 & \cdots & \frac{a(i-1;i,j)}{a(i)} & \cdots & \cdots \\
 & & & \ddots & \vdots & \vdots & \\
 & & & & 1 & \cdots & \cdots \\
 & & & & & \ddots & \\
 & & & & & & 1 \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & 0
\end{pmatrix}
$$

(on a supposé $m > r = n$ *et on note* $\mathbf{0}$ *diverses matrices nulles. Il y a une matrice identité* $(m-r)\times(m-r)$ *en bas à droite de* L*.) avec* $\mathbf{D}$ *la matrice diagonale*

$$
D\left(\frac{a(1)}{a(0)}, \frac{a(2)}{a(1)}, \ldots, \frac{a(r)}{a(r-1)}\right)
$$

formée avec les quotients des mineurs principaux successifs ($a(0) = 1$*) et les* $a(i-1;i,j)$ *et* $a(j-1;i,j)$ *étant des mineurs de* A *décrits précédemment et utilisés à l'intersection de la* i^e *ligne et de la* j^e *colonne des matrices ci-dessus.*

Aussi (avec, pour changer, $r = m < n$*) :*

$$
\mathtt{A} = \begin{pmatrix}
a(1) & & & & & & \\
\cdots & a(2) & & & & & \\
\vdots & \vdots & \ddots & & & & \\
\vdots & \vdots & \vdots & a(j) & & & \\
\vdots & \vdots & \vdots & \vdots & \ddots & & \\
\cdots & \cdots & \cdots & a(j-1;i,j) & \cdots & \ddots & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & a(r)
\end{pmatrix} \cdot \mathbf{E} \cdot
$$

$$
\cdot \begin{pmatrix}
a(1) & \cdots & \cdots & \cdots & \cdots & & & & \\
 & \ddots & \vdots & \vdots & & & & & \\
 & & a(i) & \cdots & a(i-1;i,j) & \cdots & \cdots & \cdots & \cdots \\
 & & & \ddots & \vdots & \vdots & \vdots & & \\
 & & & & \ddots & \cdots & \cdots & \cdots & \cdots \\
 & & & & & \ddots & \cdots & & \\
 & & & & & & a(r) & \cdots & \cdots
\end{pmatrix}
$$

où E *est la matrice diagonale*

$$D(\frac{1}{a(0)a(1)}, \frac{1}{a(1)a(2)}, \dots, \frac{1}{a(r-1)a(r)})$$

7 Le cas de la dimension 3

Prenons

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

Alors

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{d}{a} & 1 & 0 \\ \frac{g}{a} & \frac{ah-gb}{ae-db} & 1 \end{pmatrix} \quad U = \begin{pmatrix} a & b & c \\ 0 & \frac{ae-db}{a} & \frac{af-dc}{a} \\ 0 & 0 & \frac{\det A}{ae-db} \end{pmatrix}$$

et donc il vaudrait mieux que l'identité

$$\frac{g}{a}c + \frac{ah-gb}{ae-db} \cdot \frac{af-dc}{a} + \frac{\det A}{ae-db} = i$$

soit valable, et je vous laisse la vérifier...

La formulation avec une matrice diagonale peut s'écrire :

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ d & ae-db & 0 \\ g & ah-gb & \det A \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{a} & 0 & 0 \\ 0 & \frac{1}{a(ae-db)} & 0 \\ 0 & 0 & \frac{1}{(ae-db)\det A} \end{pmatrix} \cdot \begin{pmatrix} a & b & c \\ 0 & ae-db & af-dc \\ 0 & 0 & \det A \end{pmatrix}$$

8 Et si l'on doit permuter des lignes ?

La dicussion générale de l'élimination de Gauss-Jordan doit discuter des situations où l'on trouve moins de pivots que $\min(m, n)$, ou lorsque l'on doit permuter des lignes.

On en est à $A_k = T_k \dots T_1 A_1$ et pas de bol on ne trouve pas de pivot en position $(k+1, k+1)$. Il faut donc permuter les lignes, sans toucher aux k premières. Mais peut être que ce qui reste de la $(k+1)^e$ colonne est nul. Dans ce cas il faut aussi aller voir plus à droite et éventuellement aussi permuter les lignes.

Dans les deux cas on peut donc être amené à remplacer A_k par PA_k avec P une matrice qui permute certaines des lignes $k+1$ à m (le plus simple est de prendre pour P une transposition mais ce n'est pas obligatoire, et même faire plutôt une « rotation » des lignes est intéressant dans certains contextes).

On écrit $PA_k = T'_k \dots T'_1 PA$ avec $T'_k = PT_kP^{-1}$, ..., $T'_1 = PT_1P^{-1}$. À quoi ressemblent les PT_jP^{-1} ? Imaginons qu'on multiplie la matrice identité à gauche par PT_jP^{-1} :

1. les lignes de $k+1$ à m sont permutées par P^{-1}, disons que la nouvelle i^e ligne $\mathcal{L}'_i$ est $\mathcal{L}_{\sigma^{-1}(i)}$,
2. T_j remplace les lignes $\mathcal{L}'_i$, $i > j$ par $\mathcal{L}''_i = \mathcal{L}'_i - \alpha(i,j)\mathcal{L}'_j$, avec $-\alpha(i,j)$ les coefficients de la j^e colonne de T_j.
3. P remplace les lignes $\mathcal{L}''_i$ par $\mathcal{L}'''_i = \mathcal{L}''_{\sigma(i)}$.

Traitons les différents cas :

1. $i < j$. Ni P ni T_j ni P^{-1} ne modifient ces lignes.
2. $j \leqslant i \leqslant k$. Ces lignes sont modifiées comme T_j le fait, car laissées invariantes par P.
3. $i > k$. On a $\mathcal{L}'''_i = \mathcal{L}''_{\sigma(i)} = \mathcal{L}'_{\sigma(i)} - \alpha(\sigma(i),j)\mathcal{L}'_j = \mathcal{L}_{\sigma^{-1}(\sigma(i))} - \alpha(\sigma(i),j)\mathcal{L}_{\sigma^{-1}(j)} = \mathcal{L}_i - \alpha(\sigma(i),j)\mathcal{L}_j$, car $\sigma(j) = j$.

La conclusion est que la matrice T'_j est exactement comme T_j sauf en ligne i et colonne j au lieu d'avoir un coefficient $-\alpha(i,j)$, on a le coefficient $-\alpha(\sigma(i),j)$. Donc pour les inverses L'_j et L_j c'est pareil : au lieu d'avoir des coefficients $+\alpha(i,j)$, on a des coefficients $+\alpha(\sigma(i),j)$.

Cela signifie concrètement que si l'on doit permuter les lignes de A d'une certaine façon à une certaine étape, alors on permute *de la même manière* dans chaque colonne les coefficients sous-diagonaux (PAS LES LIGNES : cela détruirait l'aspect triangulaire de la matrice L !) déjà obtenus de la matrice L en cours de formation. Une fois cela fait on remplit une nouvelle colonne de L. Et on continue. Il est crucial pour cette recette que l'on ne s'autorise que des permutations de lignes dans les transformées de A qui ne touchent plus aux lignes avec les pivots déjà calculés.

L'algorithme aboutit donc à $P \cdot A = L \cdot U$ avec une certaine matrice de permutation P qui accumule toutes les permutations de lignes effectuées.

En général la matrice U est échelonnée, mais les pivots ne se trouvent pas forcément dans les r premières colonnes : dans le cas général il faut encore multiplier à droite par une matrice Q de permutation de colonnes pour aboutir à une formule $PAQ = LU'$ avec un U' ayant les pivots dans les r premières colonnes.