LU Factorization

After seeing Gaussian elimination with partial pivoting in its full algorithmic form, we wonder if there is even further simplifications of this process. More precisely, we ask if we can beat the $O\left(\frac{n^3}{3}\right)$ operation time to determine the solution vector $\boldsymbol{x}$ from $A\boldsymbol{x} = b$.

Suppose now we know

$$A = LU$$

where L is lower triangular and U is upper triangular. We can then solve $A\boldsymbol{x} = b$ in two steps:

(1) Let $\boldsymbol{y} = U\boldsymbol{x}$. We solve $L\boldsymbol{y} = \boldsymbol{b}$ for $\boldsymbol{y}$ which costs only $O\left(n^2\right)$ operations since L is lower triangular (via forward substitution).
(2) Since $\boldsymbol{y}$ is known, the upper triangular system $U\boldsymbol{x} = \boldsymbol{y}$ also requires just $O\left(n^2\right)$ operations to find $\boldsymbol{x}$.

Altogether, if we can write $A = LU$ as a factorized form, we only need $O\left(2n^2\right)$ operations instead of $O\left(n^3/3\right)$, a major improvement exemplified as follows:

n	$n^3/3$	$2n^2$	% reduction
10	3.3×10^2	2×10^2	40
100	3.3×10^5	2×10^4	94
1000	3.3×10^8	2×10^6	99.4

which clearly showcases the improvement when our system is large (and they can be even larger in practice).

However, there is no free lunch. Coming up with the specific L and U requires $O\left(n^3/3\right)$ operations. Nonetheless, once the factorization is determined, solving $A\boldsymbol{x} = b$ is extremely simple.

In fact, we have already come up with U. The very Gaussian elimination procedure gives us an upper triangular system, which we then perform back substitution.

$$A\boldsymbol{x} = b \implies U\boldsymbol{x} = \tilde{b}.$$

So what really happened in the $\implies$? Elimination did. In fact, row operations did. Can we represent row operations as a matrix multiplication on A?

Example. Notice that the identity matrix acting on A doesn't change anything, that is,

$$IA = A.$$

Example. We don't try until we realise we can. **Assume that we can perform Gaussian elimination without row swapping**. Consider the row operation using the first row.

$$\left(E_j - m_{j1}E_1\right) \to \left(E_j\right), \quad m_{j1} = \frac{a_{j1}}{a_{11}}.$$

What this really is involves a matrix of multipliers as follows:

$$\begin{aligned}
M^{(1)} &= \begin{bmatrix} 1 & 0 & . & . & . & . & 0 \\ -m_{21} & 1 & 0 & . & . & . & . \\ -m_{31} & 0 & . & . & & & . \\ . & . & . & . & . & & . \\ . & . & & . & . & . & . \\ . & . & & & 0 & . & 0 \\ -m_{n1} & 0 & . & . & . & . & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & . & . & . & . & 0 \\ -m_{21} & 0 & 0 & . & . & . & . \\ -m_{31} & 0 & . & . & & & . \\ . & . & . & . & . & & . \\ . & . & & . & . & . & . \\ . & . & & & 0 & . & 0 \\ -m_{n1} & 0 & . & . & . & . & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & . & . & . & . & 0 \\ 0 & 1 & 0 & . & . & . & . \\ . & . & . & . & & & . \\ . & . & . & . & . & & . \\ . & . & & . & . & . & . \\ . & . & & & . & . & 0 \\ 0 & 0 & . & . & . & 0 & 1 \end{bmatrix} \\
&= W^{(1)} + I_n
\end{aligned}$$

where I_n is the identity matrix with dimension n.

This operation is marked as

$$\begin{aligned} M^{(1)}A &= \left(W^{(1)} + I_n\right) A \\ &= W^{(1)}A + A \end{aligned}$$

Don't trust this? Carry it out.

$$\begin{aligned}
W^{(1)}A &= \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ -m_{21} & 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ -m_{31} & 0 & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & 0 & \cdot & 0 \\ -m_{n1} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & \cdot & \cdot & a_{2n} \\ a_{31} & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ -m_{21}a_{11} & -m_{21}a_{12} & \cdot & \cdot & \cdot & \cdot & -m_{21}a_{12} \\ -m_{31}a_{11} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -m_{n1}a_{11} & \cdot & \cdot & \cdot & \cdot & \cdot & -m_{n1}a_{1n} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ -a_{21} & -m_{21}a_{12} & \cdot & \cdot & \cdot & \cdot & -m_{21}a_{12} \\ -a_{31} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -a_{n1} & \cdot & \cdot & \cdot & \cdot & \cdot & -m_{n1}a_{1n} \end{bmatrix}
\end{aligned}$$

Now, $W^{(1)}A + A$ will clear up the first column

$$\begin{aligned}
W^{(1)}A + A &= \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ -a_{21} & -m_{21}a_{12} & \cdot & \cdot & \cdot & \cdot & -m_{21}a_{12} \\ -a_{31} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -a_{n1} & \cdot & \cdot & \cdot & \cdot & \cdot & -m_{n1}a_{1n} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & \cdot & \cdot & a_{2n} \\ a_{31} & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & \cdot & a_{1n} \\ 0 & a_{22}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{2n}^{(1)} \\ 0 & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & \cdot \\ 0 & a_{n2}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{nn}^{(1)} \end{bmatrix}
\end{aligned}$$

exactly as we wanted. We call $M^{(1)}$ the **first Gaussian transformation matrix**. Note that $M^{(1)}$ is **lower triangular**.

Remember, $M^{(1)}$ operates on **both** side of the equation $A\boldsymbol{x} = \boldsymbol{b}$, that is,

$$\boxed{A^{(2)}\boldsymbol{x}} = M^{(1)}A\boldsymbol{x} = M^{(1)}\boldsymbol{b} = \boxed{\boldsymbol{b}^{(2)}}.$$

We go on with the boxed term

$$A^{(2)}\boldsymbol{x} = \boldsymbol{b}^{(2)}$$

and multiply $M^{(2)}$ on the left of both sides to obtain

$$A^{(3)}\boldsymbol{x} = \boldsymbol{b}^{(3)}$$

where $A^{(3)} = M^{(2)}A^{(2)}$ and $\boldsymbol{b}^{(3)} = M^{(2)}\boldsymbol{b}^{(2)}$.

In general, with $A^{(k)}\boldsymbol{x} = \boldsymbol{b}^{(k)}$ formed, we multiply it then by the $\boldsymbol{k^{th}}$ **Gaussian transformation matrix**,

$$M^{(k)} = \begin{bmatrix} 1 & 0 & . & . & . & . & 0 \\ 0 & 1 & 0 & . & . & . & . \\ . & 0 & . & . & & & . \\ . & . & . & . & . & & . \\ . & . & & -m_{k+1,k} & . & . & . \\ . & . & & . & 0 & . & 0 \\ 0 & 0 & . & -m_{n,k} & . & 0 & 1 \end{bmatrix}$$

that is, an identity matrix with entries in the k^{th} column <u>strictly</u> below the diagonal modified by the multipliers.

From this general procedure, we realize that

$$A^{(k+1)}\boldsymbol{x} = M^{(k)}A^{(k)}\boldsymbol{x} = M^{(k)} \dots M^{(1)}A\boldsymbol{x} = M^{(k)}\boldsymbol{b}^{(k)} = \boldsymbol{b}^{(k+1)} = M^{(k)} \dots M^{(1)}\boldsymbol{b}$$

and we multiply a Gaussian transformation matrix on the left until we reach $A^{(n)}$, exactly the last step, that is,

$$A^{(n)} = M^{(n-1)}M^{(n-2)} \dots M^{(1)}A.$$

Now, if we can invert all these M matrices, we get

$$A = \left[M^{(n-1)}M^{(n-2)} \dots M^{(1)}\right]^{-1} A^{(n)}$$

where $A^{(n)}$ is upper triangular.

Curiously enough, the product of two lower triangular matrices is still lower triangular. More curiously enough, the inverse of a lower triangular matrix is still lower triangular! In fact, let's consider the inverse of the $\boldsymbol{k^{th}}$ **Gaussian transformation matrix**,

$$L^{(k)} = \left[M^{(k)}\right]^{-1} = \begin{bmatrix} 1 & 0 & . & . & . & . & 0 \\ 0 & 1 & 0 & . & . & . & . \\ . & 0 & . & . & & & . \\ . & . & . & . & . & & . \\ . & . & & m_{k+1,k} & . & . & . \\ . & . & & . & 0 & . & 0 \\ 0 & 0 & . & m_{n,k} & . & 0 & 1 \end{bmatrix}$$

because (the inverse is) to undo $M^{(k)}$, we simply hit the "reverse" button by adding the equations back instead of subtracting. With this, we see that

$$L := L^{(1)}L^{(2)} \cdots L^{(n-1)} = \begin{bmatrix} 1 & 0 & . & . & . & . & 0 \\ m_{21} & 1 & 0 & . & . & . & . \\ m_{31} & m_{32} & . & . & & & . \\ . & . & . & . & . & & . \\ . & . & & . & . & . & . \\ . & . & & 0 & & . & 0 \\ m_{n1} & m_{n2} & . & . & . & m_{n,n-1} & 1 \end{bmatrix}$$

with $L^{(k)} = \left[M^{(k)}\right]^{-1}$. Altogether, let

$$L = L^{(1)}L^{(2)} \cdots L^{(n-1)}$$

and

$$U = M^{(n-1)}M^{(n-2)} \dots M^{(1)}A,$$

we have

$$\begin{aligned} LU &= L^{(1)}L^{(2)} \cdots L^{(n-1)}M^{(n-1)}M^{(n-2)} \dots M^{(1)}A \\ &= \left[M^{(1)}\right]^{-1} \cdots \cancel{\left[M^{(n-1)}\right]^{-1} M^{(n-1)}} M^{(n-2)} \dots M^{(1)}A \\ &= A, \end{aligned}$$

that is, we have identified the LU-factorization of A.

Here, L is made up entirely of the multiplier values (and an identity matrix, accountable for 1's on the diagonal), so forming it is straightforward. Once L is found, we solve for $\boldsymbol{y}$ that satisfies

$$L\boldsymbol{y} = \boldsymbol{b}$$

where $\boldsymbol{y} := U\boldsymbol{x}$ as a definition. Since L is lower triangular, the system is solved by forward substitution (a process just with reversed index as in back substitution), which is $O\left(n^2\right)$.

To form U, we make use simultaneously of the entries in L because they are just multipliers that help us form each of the $M^{(k)}$'s. A detailed algorithm is given as Algorithm 6.4 in Section 6.5 of your textbook.