Permutation Matrices

Last class, we introduced LU factorization of a matrix A with the assumption that no row swap is performed. This assumption unfortunately is not always satisfied – there is almost always some row swapping, either due to a zero diagonal entry, or a very small one that makes the elimination process unstable. We are blessed that the multiplying and adding/subtracting rows can be represented by matrix multiplications, namely, the $M^{(k)}$'s – the **Gaussian transformation matrix**. We turn now our attention to row swaps.

Example. Consider the square matrix

$$A = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix}.$$

We search through the first column and locate the row in which the largest (in magnitude) column entry resides and perform partial pivoting. Here, we identify that the third row and the first should swap,

$$\widehat{A} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 4 & 3 & 3 & 1 \\ 2 & 1 & 1 & 0 \\ 6 & 7 & 9 & 8 \end{bmatrix}$$

where now, dividing by 8 to find the multiplier becomes much more stable.

This procedure can be represented by the permutation operation on A, as in

$$P^{(1)}A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 4 & 3 & 3 & 1 \\ 2 & 1 & 1 & 0 \\ 6 & 7 & 9 & 8 \end{bmatrix}$$

Suppose now, we have done the row reduction for the first row and reached this following form,

$$M^{(1)}P^{(1)}A = A^{(2)} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{4} & -\frac{5}{4} & -\frac{5}{4} \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \end{bmatrix}$$

(HW exercise to find $M^{(1)}$). We are looking at the second row pivot with magnitude $\frac{1}{2}$. We prefer the last row here since $\frac{7}{4}$ is the largest magnitude from the 2nd row and on. The swap yields

$$P^{(2)}A^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{4} & -\frac{5}{4} & -\frac{5}{4} \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & -\frac{3}{4} & -\frac{5}{4} & -\frac{5}{4} \\ 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{3}{2} \end{bmatrix}.$$

We carry on, with an elimination step to find

$$M^{(2)}P^{(2)}A^{(2)} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{2}{7} & \frac{4}{7} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \end{bmatrix} = A^{(3)}.$$

Here, we prefer the fourth row, thus we perform

$$P^{(3)}A^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{2}{7} & \frac{4}{7} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \\ 0 & 0 & -\frac{2}{7} & \frac{4}{7} \end{bmatrix}.$$

One more step of elimination yields

$$M^{(3)}P^{(3)}A^{(3)} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \\ 0 & 0 & 0 & \frac{2}{3} \end{bmatrix}.$$

So, what did we achieve? Did we find the LU that makes A? Not quite. We found

$$PA = LU$$

where P is a permutation matrix,

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{3}{4} & 1 & 0 & 0 \\ \frac{1}{2} & -\frac{2}{7} & 1 & 0 \\ \frac{1}{4} & -\frac{3}{7} & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \\ 0 & 0 & 0 & \frac{2}{3} \end{bmatrix}.$$

Noting from the permutation matrices applied at each step of elimination,

$$P^{(1)} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad P^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad P^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

we observe that,

$$\begin{aligned} P &= P^{(3)}P^{(2)}P^{(1)} \\ &= P^{(3)} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= P^{(3)} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Remark. Did we just get lucky that the product of permutation matrices gives us what we wanted? This fact is not obvious, but the details may be too complicated for the course. A great explanation is provided in Trefethen and Bau (see lecture note appendix).

In practice, these permutation matrices don't actually show up – their effects are carried out by explicitly swapping the rows (or swapping the row indices). Algorithm 6.5 in your textbook outlines the procedures of finding L and U factors. **Though you are not required to code up this algorithm, you are responsible for knowing the reasons why step is set certain ways, or what each step accomplishes.**

Iterative Methods: an Introduction

In the previous section, we have studied direct methods of solving linear systems, in the sense that the error in our numerical solution arises purely from round-off errors. In this section, we study iterative methods, namely, approximating the true solution closer and closer, but only get close enough with a prescribed tolerance level.

Example. Suppose an iterative algorithm produces $x^{(k)} = \left(x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}\right)$ as an approximate solution to a linear system at the k^{th} iteration.

If we know the true solution is x, we want to see how close we are. We need a measure of the "difference" $\left(x^{(k)} - x\right)$.

If we don't know the true solution (almost always), we want to see how much we have improved from the previous step, that is, we want to look at $\left(x^{(k)} - x^{(k-1)}\right)$ and see how this "difference" is changing as k changes. Certainly, if this "difference" becomes smaller and smaller, and if we can also prove (in mathematical analysis) that this "difference" goes to 0 as $k \to \infty$, then we know the sequence is **convergent**.

Definition. A **vector norm** on $\mathbb{R}^n$ is a function, $\|\cdot\|$, from $\mathbb{R}^n$ to $\mathbb{R}$ with the following properties:

(1) (nonnegativity)$\|\boldsymbol{x}\| \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$,
(2) (positive definiteness/point-separating) $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$,
(3) (absolute homogeneity) $\|\alpha\boldsymbol{x}\| = |\alpha| \, \|\boldsymbol{x}\|$ for all $\alpha \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$,
(4) (triangle inequality/subadditivity)$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.