1 Gaussian elimination: LU-factorization

This note introduces the process of Gaussian[1] elimination, and translates it into matrix language, which gives rise to the so-called LU-factorization. Gaussian elimination transforms the original system of equations into an equivalent one, i.e., one which has the same set of solutions, by adding multiples of rows to each other. By eliminating variables from the equations, we end up with a "triangular" system, which can be solved simply through substitution. Let us follow the elimination process in a simple example of three equations in three unknowns.

$$
\begin{array}{lrcrcrcrl}
\hline
\epsilon_1 : & 2u & - & 2v & + & 3w & = & 1 & \\
\epsilon_2 : & 6u & - & 7v & + & 14w & = & 5 & \\
\epsilon_3 : & 4u & - & 8v & + & 30w & = & 14 & \\
\hline
\epsilon_1' = \epsilon_1 : & 2u & - & 2v & + & 3w & = & 1 & \\
\epsilon_2' = \epsilon_2 - 3\epsilon_1 : & & - & v & + & 5w & = & 2 & \\
\epsilon_3' = \epsilon_3 - 2\epsilon_1 : & & - & 4v & + & 24w & = & 12 & \\
\hline
\epsilon_1'' = \epsilon_1' : & 2u & - & 2v & + & 3w & = & 1 & \\
\epsilon_2'' = \epsilon_2' : & & - & v & + & 5w & = & 2 & \\
\epsilon_3'' = \epsilon_3' - 4\epsilon_2' : & & & & & 4w & = & 4 & \\
\hline
\text{Back-} & & & & & w & = & 1 & \\
\text{substitution:} & & - & v & + & 5 & = & 2 & \Rightarrow v = 3 \\
 & 2u & - & 6 & + & 3 & = & 1 & \Rightarrow u = 2 \\
\hline
\end{array}
\tag{1.1}
$$

We will now translate this process into matrix language.

A matrix is a rectangular array of numbers. Our example here works with 3 by 3 matrices, i.e., 3 rows and 3 columns. Actually, it is often best to think of a matrix as a collection of columns (or a collection of rows, depending on the context), rather than a rectangular array of numbers. It is also often useful, to think of a matrix B in terms of what it "does" when it acts on a vector x (or even some other matrix C). This is tantamount to interpreting the matrix B as an *operator* effecting the transformation $x \mapsto Bx$. Let us denote the standard basis vectors by $e_1, \ldots, e_k, \ldots$; for vectors with three

[1]Carl Friedrich Gauss, 1777-1855, one of the greatest mathematicians. The technique of "Gaussian" elimination has been known prior to Gauss.

components we have

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}; \tag{1.2}$$

note, that we always write our vectors as column vectors. The matrix which has the standard basis vectors as its columns is called the identity matrix, I,

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.3}$$

The identity matrix leaves vectors and matrices unchanged, i.e.,

$$Ix = x, \qquad IB = B,$$

for all vectors x and matrices B. On the other hand, the standard basis vectors can be used to pick out columns and rows from a matrix:

$$Be_k = k\text{-th column of } B; \qquad e_k^T B = k\text{-th row of } B;$$

the $\cdot^T$ symbol means transpose: it makes row vectors out of column vectors and vice versa. For a matrix, B^T is the matrix which has the rows of B as its columns. Finally,

$$e_j^T B e_k = b_{jk} = \text{ the element of } B \text{ in row } j \text{ and column } k.$$

We now return to our example. First, our given system of equations (1.1) is

$$\underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 6 & -7 & 14 \\ 4 & -8 & 30 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} u \\ v \\ w \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 1 \\ 5 \\ 14 \end{pmatrix}}_{b}. \tag{1.4}$$

The first step in the elimination process above is to subtract three times the first equation from the second. Therefore, in the coefficient matrix A, we replace the second row by the the difference between it and three times the first row. Of course, the factor 3 is chosen to obtain a 0 in the (2,1)-position of the matrix. One such elimination step corresponds to multiplying the matrix A by the **elementary** matrix E_{21} from the left:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{=:\, E_{21}} \underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 6 & -7 & 14 \\ 4 & -8 & 30 \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 0 & -1 & 5 \\ 4 & -8 & 30 \end{pmatrix}}_{E_{21}A}. \tag{1.5}$$

An elementary matrix is equal to the identity matrix I, except for one of the off-diagonal elements which is nonzero. Proceeding with our elimination steps we obtain

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}}_{=:\, E_{31}} \underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 0 & -1 & 5 \\ 4 & -8 & 30 \end{pmatrix}}_{E_{21}A} = \underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 0 & -1 & 5 \\ 0 & -4 & 24 \end{pmatrix}}_{E_{31}E_{21}A}, \tag{1.6}$$

and

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{pmatrix}}_{=:\, E_{32}} \underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 0 & -1 & 5 \\ 0 & -4 & 24 \end{pmatrix}}_{E_{31}E_{21}A} = \underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 0 & -1 & 5 \\ 0 & 0 & 4 \end{pmatrix}}_{E_{32}E_{31}E_{21}A \,=:\, U}. \tag{1.7}$$

Thus, our "final" matrix U is upper triangular, and corresponds to the coefficients of the final set of equations $(\epsilon_1'', \epsilon_2'', \epsilon_3'')$ in (1.1). In linear algebra courses this is often referred to as the *row echelon form* of the matrix A. The same matrices need to be applied to the vector b, to obtain our equivalent system of equations. Note that the three elementary matrices E_{21}, E_{31}, and E_{32} contain the information about how we proceeded in the process of Gaussian elimination; they are like entries in a bookkeeping journal. Therefore, it should be possible to reconstruct the original matrix A from U and the bookkeeping matrices E_{21}, E_{31}, and E_{32}. From (1.7) we have

$$\begin{aligned} E_{32}E_{31}E_{21}A &= U \\ E_{31}E_{21}A &= E_{32}^{-1}U \\ E_{21}A &= E_{31}^{-1}E_{32}^{-1}U \\ A &= \underbrace{E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}}_{=:\, L}U. \end{aligned} \tag{1.8}$$

This is the formula for reconstructing the matrix A. To find L, we need to find the inverses of elementary matrices. This is easy to do, if we remember the meaning of these elementary matrices. For example, E_{21} when applied to a matrix from the left, *subtracts* three times the first row from the second row. Clearly, the inverse operation (undoing what we have just done) is to *add* three times the first row to the second row.

$$E_{21} = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \Rightarrow \qquad E_{21}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ +3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \tag{1.9}$$

so inverses of elementary matrices are simply obtained by changing the sign of the nonzero off-diagonal element. Now, another interesting phenomenon occurs, when we perform the matrix multiplication to obtain L:

$$\begin{aligned} L &= E_{21}^{-1} E_{31}^{-1} E_{32}^{-1} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 4 & 1 \end{pmatrix}, \end{aligned} \tag{1.10}$$

so the matrix multiplication can be performed by just putting the nonzero off-diagonal elements of the elementary matrices into the appropriate positions in the matrix L. From an algorithmic point of view this means that the entries of L, which are just the multipliers in the Gaussian elimination process, can be easily stored during the process of Gaussian elimination. Please note, that the order in which those elementary matrices appear is crucial, as

$$L^{-1} = E_{32} E_{31} E_{21} = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 10 & -4 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -2 & -4 & 1 \end{pmatrix}. \tag{1.11}$$

Finally, putting everything together we obtain

$$\underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 6 & -7 & 14 \\ 4 & -8 & 30 \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 4 & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 2 & -2 & 3 \\ 0 & -1 & 5 \\ 0 & 0 & 4 \end{pmatrix}}_{U}, \tag{1.12}$$

i.e., Gaussian elimination gives us the LU-**factorization** (sometimes also called the LU-**decomposition**) of the matrix A, $A = LU$, where L is a lower triangular matrix with all diagonal entries equal to 1, and U is an upper triangular matrix.

Once we have this factorization, how can we make use of it to solve $Ax = b$? It is easy to see that

$$Ax = LUx = b$$

is equivalent to solving

$$Ux = L^{-1}b =: c.$$

In general, we do not want to compute the inverse of L. Instead, we can find $c = L^{-1}b$ through the process called *forward substitution*,

$$Lc = b;$$

since L is triangular this is as easy to do as the final step, namely *backward substitution*,

$$Ux = c.$$

As we shall see later, these two substitution processes are no more costly than computing $A^{-1}b$, given that A^{-1} is known, so in that sense knowing L and U is as good for solving systems as knowing the inverse. Once the factorization is obtained, it is very cheap to solve many systems with the same matrix A and different right hand sides.

Proposition 1.1 *If Gaussian elimination does not break down, i.e., if all pivots (*the leading diagonal elements at each step*) are nonzero, then the matrix A has a unique LU-factorization, where L is a lower triangular matrix with all diagonal entries equal to* 1*, and U is an upper triangular matrix.*

So what if Gaussian elimination "breaks down", i.e., if there is no LU factorization of our matrix A? For example, the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

does not have an LU factorization, even though the corresponding system of equations poses no difficulties. Indeed, if we simply switch the two rows (thus obtaining an equivalent system), the resulting matrix does have an LU factorization,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = LU,$$

where $L = U = I$, and I denotes the identity matrix of appropriate size. Usually, it is less obvious but no more difficult to fix the breakdown. For example, applying Gaussian elimination to

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 2 & -1 & -1 \end{pmatrix}$$

after two steps results into

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & -1 \end{pmatrix} A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & -2 \\ 0 & -3 & -3 \end{pmatrix};$$

to proceed we must switch rows 2 and 3, to move a nonzero element to the current diagonal position (2,2). How can switching rows be expressed in matrix language? Again, by matrix multiplication from the left[2], this time with a **permutation** matrix,

$$P = P_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Thus, we obtain

$$\begin{aligned} PA &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 2 & -1 & -1 \end{pmatrix} \\ = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 1 & 1 & -1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & -3 & -3 \\ 0 & 0 & -2 \end{pmatrix}, \end{aligned}$$

i.e., the matrix PA does have an LU-factorization. Of course, for larger matrices, more of those row exchanges may be necessary, which leads to a

[2]multiplying from the left will operate on the rows of a matrix, multiplying from the right will operate on the columns of a matrix

more complicated permutation of the rows in the original matrix A, but can easily be handled by good "bookkeeping".

Our strategy consists in switching rows to obtain a nonzero "pivot" (the current diagonal element). Naturally, this is not possible, if all the elements in the present column are zero. But in that case, we may just simply ignore the present column and move on to the next one. However, this scenario means that our matrix A is singular, i.e., $Ax = b$ has either no solution, or infinitely many. Although we will ignore this case in the following, we give an example.

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 3 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix} \rightarrow$$
$$\rightarrow \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = U,$$

with

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 2 & -2 & 1 \end{pmatrix}.$$

The resulting upper triangular matrix U has a whole row of zeros, clearly indicating singularity.

Summing up, the main result of this section is

Theorem 1.2 (LU-factorization) *For every n by n matrix A there exists a permutation matrix P, such that PA possesses an LU-factorization, i.e., $PA = LU$, where L is a lower triangular matrix with all diagonal entries equal to 1, and U is an upper triangular matrix.*