Lesson 4

Reading: Larson, Section 2.3, The Inverse of a Matrix; Section 2.4, Elementary Matrices

Suggested exercises: Larson, Section 2.3: 1, 7, 9, 13, 19, 33, 37, 41, 45, 47; Section 2.4: 1, 3, 7, 9, 13, 17, 21, 25, 31, 41, 45

Submit: Lesson 4: Matrix inverses and LU-factorization

Section 2.3: The Inverse of a Matrix

In this section, we encounter the matrix operation that corresponds to division of real numbers: computing the inverse of a square matrix. Division of real numbers is defined in terms of the *multiplicative inverse* of a nonzero real number: for $a \neq 0$, its multiplicative inverse is the unique nonzero real number, a^{-1}, satisfying $a \cdot a^{-1} = a^{-1} \cdot a = 1$. Once we have this, we define division of b by a in terms of multiplication: $b/a = b \cdot a^{-1}$. We do the same thing with matrices: for a square matrix A for which a multiplicative inverse exists, we define it to be the square matrix A^{-1} satisfying $A \cdot A^{-1} = A^{-1} \cdot A = I$, where I denotes the identity matrix of the same size as A. (Note that I plays the role in matrix multiplication that the real number 1 plays in real number multiplication.)

It is not as easy to say which square matrices will have multiplicative inverses as it is for real numbers. You can tell at a glance if a real number has a multiplicative inverse: if a real number a is not 0, then a^{-1} exists. For a square matrix A, there is a number that we can compute from A that will tell us if it has a multiplicative inverse; this number is called the *determinant* of A. If it is nonzero, then the matrix has a multiplicative inverse; if it is zero, then the matrix does not have a multiplicative inverse. The determinant will be the topic of chapter 3. (A small preview: a real number a can be considered as a 1×1 matrix $[a]$, and the determinant of a 1×1 matrix $[a]$ is the number a, so the condition for a real number a to have a multiplicative inverse is the same as the determinant condition for the 1×1 matrix $[a]$.)

There is a nice formula for the inverse of a 2×2 matrix that also gives a preview of the idea of the determinant:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

provided $ad - bc \neq 0$. The number $ad - bc$ is the determinant of the matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

so we see that the inverse is defined as long as the determinant is nonzero for a 2×2 matrix.

It is a sad fact that there are not simple formulas like this for inverses of larger square matrices; we will have to work for those. The process of finding the inverse is one that you are already familiar with, though. We take our square matrix A, adjoin an identity matrix of the same size on its right to form an augmented matrix, and row reduce this augmented matrix. If the portion that originally contained A can be reduced to the identity matrix, then the portion where we adjoined the identity matrix will contain A^{-1}. There is nothing magic going on here: we are just performing Gaussian elimination to solve for the inverse matrix. The row reduction solves, simultaneously, for all columns of A^{-1}.

If the portion that originally contained A cannot be reduced to the identity matrix, then A does not have an inverse.

We can represent this process symbolically as follows: $[A|I] \longrightarrow \left[I|A^{-1}\right]$, where the arrow represents the process of reduction to reduced row-echelon form.

When you compute the matrix this way, make sure that the elementary row operations that you perform are performed all the way across the augmented matrix. For example, to find the inverse of

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

we first form the augmented matrix

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right].$$

Interchange rows 2 and 3, making sure to perform the operation on the entire rows:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{array}\right].$$

Next, subtract row 3 from row 2 and then add row 3 to row 1, again making sure to do this with the entire rows:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{array}\right].$$

We have managed to reduce the 3×3 block on the left to the identity matrix, so this means that our matrix has an inverse, and the 3×3 block appearing on the right is it:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

You can, and probably should, check this by multiplying the matrix and its inverse, and observing that you get the 3×3 identity matrix.

Properties of the inverse

The inverse of a product of invertible matrices obeys a property similar to that of the transpose of a product: the inverse of a product of invertible matrices is the product of the inverses, in reverse order:

$$(AB)^{-1} = B^{-1}A^{-1}.$$

This is necessary so that the inverses line up correctly when you multiply AB and its inverse:

$$AB \cdot B^{-1}A^{-1} = A\left(BB^{-1}\right)A^{-1} = AIA^{-1} = AA^{-1} = I,$$

and similarly in the opposite order.

If a matrix A has an inverse, then so does its transpose A^T, and it has the nice property that $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Finally, if the coefficient matrix A of a linear system

$$A\mathbf{x} = \mathbf{b}$$

is invertible, then the linear system will necessarily have a unique solution, and it will be given by

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

What makes this exceptionally pleasant is that once you have gone to the effort of finding A^{-1}, then you can easily solve $A\mathbf{x} = \mathbf{b}$ for many different vectors $\mathbf{b}$ by simply performing the multiplication $A^{-1}\mathbf{b}$ for the various $\mathbf{b}$.

A pitfall you must avoid is assuming that a sum of invertible matrices has to be invertible, and assuming that the inverse of a sum of matrices is the sum of the inverses. No such law exists: it is very simple to find a case where this fails to be true. Probably the easiest way to see that this is not true is to consider the sum of an invertible matrix and its negative. Both of these matrices are invertible, but their sum is the zero matrix, which clearly has no multiplicative inverse.

You don't need to resort to such a special case, though: another example is provided by $\begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$ and $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. Check to see that these are both invertible, but their sum $\begin{bmatrix} 3 & 3 \\ 2 & 2 \end{bmatrix}$ is not. (For a 2×2 matrix, you can recognize a noninvertible matrix by seeing that one row is a scalar multiple of the other. The determinant will always be zero if this is the case, and hence the matrix will not be invertible.)

Section 2.4: Elementary Matrices

Each elementary row operation that we have used has a matrix associated with it, and multiplying a matrix A on the left by this matrix achieves the associated elementary row operation. The elementary matrix associated with a given elementary row operation is produced as follows: start with the identity matrix of the right size so that it can multiply A from the left, and perform the given elementary row operation on this matrix.

For example, suppose A is the 3×4 matrix

$$A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 4 & 1 & 0 \\ -1 & 1 & 2 & 2 \end{bmatrix}$$

and we want to add -2 times the first row to the second, in order to eliminate the 2 that leads the second row. Since A is 3×4, we choose a 3×3 identity matrix as the right size for multiplication from the left, and our elementary matrix E is the result of applying the desired row operation to this identity matrix:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Multiplying these together gives us what we want:

$$EA = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 4 & 1 & 0 \\ -1 & 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & -2 \\ -1 & 1 & 2 & 2 \end{bmatrix}.$$

Note that the product is, in fact, the matrix that results from performing the desired elementary row operation to A.

The elementary matrices are all invertible, and the inverse of an elementary matrix E is the elementary matrix that results from undoing the elementary row operation associated with E. For example,

$$E = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$

is an elementary matrix associated with the operation of adding 2 times the first row to the second. To undo this operation, you add -2 times the first row to the second, which means

$$E^{-1} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}.$$

You can (and should) check that these matrices are inverses.

A second example:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

is an elementary matrix, associated with multiplying row three by 4. This operation is undone by multiplying row three by $\frac{1}{4}$, so

$$E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}.$$

A final example:

$$E = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

is an elementary matrix, corresponding to the operation of interchanging rows one and three. To undo this operation, you interchange rows one and three again: in other words, you apply the same elementary row operation once more. Therefore,

$$E^{-1} = E = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

This always happens for row interchanges; each of these elementary matrices is equal to its inverse.

Note that the inverse of an elementary matrix is also an elementary matrix.

There are two key things that can be done with elementary matrices. First, if a square matrix is invertible, then a sequence of elementary row operations can be performed to row-reduce it to the identity matrix of the same size. Each elementary row operation corresponds to left-multiplication by an elementary matrix. Therefore, there is a sequence of elementary matrices $E_1, E_2, \ldots, E_m$ such that

$$E_m \cdots E_2 E_1 A = I.$$

We can unwind the product on the left by multiplying by the inverse of the outermost elementary matrix, and repeating until we have cleared all of the elementary matrices from the left-hand side. We do this to both sides of the equation, and we end up with

$$A = E_1^{-1} E_2^{-1} \cdots E_m^{-1}.$$

Each of the matrices on the right is an elementary matrix, and we draw the following remarkable conclusion: *any invertible matrix can be written as a product of elementary matrices.*

A simple illustration: the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

is invertible. We reduce to the identity using the following elementary row operations: (1) add -3 times the first row to the second, getting

$$\begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix},$$

(2) multiply the second row by $-1/2$, to get

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix},$$

and finally, (3) add -2 times the second row to the first to get

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This is equivalent to computing the following product:

$$\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

You should verify that the elementary matrices shown correspond to the elementary row operations used. Note the order: the matrices are added to the product on the left as they are performed, so the first operation performed corresponds to the rightmost elementary matrix, and the last operation corresponds to the leftmost.

Now, multiply on the left by the inverses of these matrices, peeling them off one by one, so that the inverse of the leftmost elementary matrix will appear farthest to the right on the right-hand side when we are done. Also note that the identity matrix disappears: it is equivalent to multiplying a real number by 1, so we don't need to show it, any more than we would need to write $2 \cdot 1$ instead of 2. We get:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}.$$

The next, and very important, use of elementary matrices is in computing an LU-factorization of a square matrix A. This idea is used heavily in applications in which very large matrices must be dealt with efficiently. If you use a computer system to solve a large system of linear equations, or compute the determinant of a large matrix, you can be pretty sure that an LU-factorization, or something very similar to it, is involved.

Pay close attention to the qualifier that begins the text's definition of the LU-factorization:

> *If* the $n \times n$ matrix A can be written as the product of a lower-triangular matrix L and an upper-triangular matrix U...

You need to be aware that not every square matrix A has an LU-factorization. It is necessary that A is reducible to row-echelon form without row interchanges. If a row interchange is required, then A cannot be guaranteed to have an LU-factorization.

For example, consider

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

(Do you agree that we cannot put this matrix into row-echelon form without switching the rows?) Suppose we could write this as a product of a lower-triangular matrix and an upper-triangular matrix. Then we would have

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \begin{bmatrix} d & e \\ 0 & f \end{bmatrix}.$$

If we multiply the matrices on the right, and equate components, we must have

$$\begin{aligned} ad &= 0 \\ ae &= 1 \\ bd &= 1 \\ be + cf &= 0. \end{aligned}$$

From the first equation, we see that either $a = 0$ or $d = 0$. But a can't be zero: look at the second equation. Therefore, it must be d that is zero. But it can't be zero, either: $d = 0$ won't satisfy the third equation. Therefore, there is no set of values for which we can have the matrix equation above, and consequently, this matrix does not have an LU-factorization.

Aside: it's beyond the scope of this course, but just in case you are curious, I'll let you know more of the story: while not every square matrix has an LU-factorization, it is true that, for every square matrix A, we have a factorization $PA = LU$, where P is a *permutation matrix*, that is, a matrix that interchanges two or more rows of A. The idea is that P shuffles the rows of A, so that PA can be reduced to upper-triangular form without row interchanges, and *that* matrix has an LU-factorization.

Second aside: the LU-factorization is due to the British mathematician Alan Turing, whose name you should know. In addition to other important mathematical work, he is responsible for two very important things: (1) he was a key figure in the Allied effort in breaking German codes in World War II, and (2) he developed the idea of the *Turing machine* in the 1930's, a model of computation from which modern computers are derived.

Now, back to LU-factorization: it is really nothing new for us. It is simply a matter of reducing the matrix A to row-echelon form, while keeping track of the elementary row operations used to get there. The upper-triangular matrix U is just row-echelon form, and the lower-triangular matrix comes from the elementary matrices used to get to row-echelon form. A very simple example: suppose

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Only one step is required to get to row-echelon form: add -3 times row one to row two:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}.$$

The elementary matrix

$$\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$$

performs this operation, so we have

$$\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}.$$

Now, just multiply both sides by the inverse of this elementary matrix:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}.$$

And that's it! That's our LU-factorization: L is the elementary matrix, and U is the row-echelon form.

If more than one elementary row operation is necessary, then the product of the inverses of the elementary matrices (in the correct order) will be L. For example, suppose

$$A = \begin{bmatrix} 1 & 1 & 2 \\ -2 & 1 & 3 \\ 1 & 0 & 4 \end{bmatrix}.$$

Then (1) adding 2 times row one to row two, and (2) adding -1 times row one to row three will give us (writing the elementary matrices achieving these row operations)

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ -2 & 1 & 3 \\ 1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 7 \\ 0 & -1 & 2 \end{bmatrix}.$$

Next, add $\frac{1}{3}$ times row two to row three. (Avoid the temptation to switch rows two and three to make the arithmetic easier! We can't use row interchanges in computing an LU-factorization.) This gives us

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ -2 & 1 & 3 \\ 1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 7 \\ 0 & 0 & \frac{13}{3} \end{bmatrix}.$$

Finally, remove the elementary matrices from the left-hand side by multiplying by their inverses:

$$\begin{bmatrix} 1 & 1 & 2 \\ -2 & 1 & 3 \\ 1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 7 \\ 0 & 0 & \frac{13}{3} \end{bmatrix},$$

and if we multiply the elementary matrices together, we get LU on the right-hand side:

$$\begin{bmatrix} 1 & 1 & 2 \\ -2 & 1 & 3 \\ 1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 7 \\ 0 & 0 & \frac{13}{3} \end{bmatrix}.$$