Lesson 3

Reading: Larson, Section 2.1, Operations with Matrices; Section 2.2, Properties of Matrix Operations

Suggested exercises: Larson, Section 2.1: 1, 7, 11, 17, 21, 23, 35, 39, 41, 47, 49, 63; Section 2.2: 3, 13a, 17, 21, 23, 25, 27, 35, 39, 43, 67, 69

Submit: Lesson 3: Matrix arithmetic

Section 2.1: Operations with Matrices

This section introduces the basic matrix operations: addition, scalar multiplication, and multiplication. Addition of matrices and scalar multiplication are defined in what should seem to be a natural way, and should not present difficulties.

Matrix multiplication, on the other hand, may seem to be defined in a strange way at first. Why not just do what was done for addition and require the matrices to be the same size, and then multiply corresponding elements? The answer to this is that, of course, we *could* define such a multiplication of matrices, but it turns out not to be very useful. The original inspiration for our definition of matrix multiplication is a system of linear equations: we want to have a definition of matrix multiplication that makes it useful in representing and solving such a system. Therefore, we want a definition that will produce the right thing when presented with the left-hand side of a linear equation.

For example, suppose we have a linear system consisting of the single equation

$$x_1 + 2x_2 - x_3 = 2.$$

If we are going to represent this linear system with matrices, then we need a multiplication operation that will generate the $x_1 + 2x_2 - x_3$. If we have a matrix to contain the coefficients of this system, and a matrix to contain the variables, then we need something like this to happen:

$$\begin{bmatrix} 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \end{bmatrix}.$$

Notice that the operation consists of multiplying elements in corresponding locations in the matrices and adding the results. (If you have encountered the dot product of vectors before, this is what is being done here, if you think of the row matrix and the column matrix as vectors.) The matrices are arranged in the correct manner for the definition of matrix multiplication to work: the second matrix must have a number of rows equal to the number of columns of the first matrix so that every entry in the row matrix on the left has a corresponding entry in the column matrix on the right by which it is multiplied.

If we denote our 1×3 coefficient matrix by A, our 3×1 variable matrix by $\mathbf{x}$, and the right-hand side as the 1×1 matrix $\mathbf{b} = [2]$, then our system can be written in matrix form as $A\mathbf{x} = \mathbf{b}$.

This idea of multiplication of a row matrix and a column matrix extends to larger matrices in a consistent way: if we have an $m \times n$ matrix A and an $n \times p$ matrix B, then the product AB is defined to be an $m \times p$ matrix in which the entry in row i and column j of the result is the product of row i of A and column j of B.

Note how the sizes of A and B must match: the number of columns of A must be equal to the number of rows of B, in order for the product of a row of A and a column of B to be defined.

Thus, a linear system

$$
\begin{aligned}
a_{11}x + a_{12}y &= b_1 \\
a_{21}x + a_{22}y &= b_2
\end{aligned}
$$

has matrix form

$$
\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.
$$

There is another, equivalent, way of viewing matrix multiplication when the second matrix is a column matrix, which we can see by rearranging the matrix equation above. If we multiply the matrices on the left in the preceding equation, we get

$$
\begin{bmatrix} a_{11}x + a_{12}y \\ a_{21}x + a_{22}y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.
$$

Now separate the matrix on the left into a sum:

$$
\begin{bmatrix} a_{11}x \\ a_{21}x \end{bmatrix} + \begin{bmatrix} a_{12}y \\ a_{22}y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},
$$

and factor out the scalars x and y from these matrices:

$$x \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + y \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

We can see two things from this: first, that we can multiply a matrix by a column matrix by setting up a product of the columns of the first matrix by the corresponding entries of the second matrix, and second, that a linear system

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

has a solution if (and only if!) we can write $\mathbf{b}$ as a sum of scalar multiples of the columns of the matrix A. (We will encounter this idea later in the course as something called the *column space* of the matrix A.)

For example, the linear system

$$\begin{aligned} x + y + z &= 2 \\ 2x - y + z &= 5 \\ y + 2z &= 3 \end{aligned}$$

has solution $x = 1$, $y = -1$, $z = 2$. If we write the system in matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix},$$

then we can verify that our solution is correct either by computing the product on the left as

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} (1)(1) + (1)(-1) + (1)(2) \\ (2)(1) + (-1)(-1) + (1)(2) \\ (0)(1) + (1)(-1) + (2)(2) \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix},$$

or we can compute the product as

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = (1) \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + (-1) \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + (2) \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix},$$

which is equal to

$$\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}.$$

Section 2.2: Properties of Matrix Operations

In this section, the rules of algebra followed by the matrix operations are explored. Many of the rules are the same as the rules obeyed by the corresponding operation with real numbers, and therefore should feel familiar to you. For example, addition of matrices is commutative; in other words, as long as $A + B$ is defined, then the order does not matter: $A + B = B + A$.

There are some important exceptions, though, that you need to be familiar with. First, multiplication of matrices is *not* commutative. Not only might you get different matrices when you compute AB and BA, but it is even possible that one of the products is defined and the other is not! This occurs when the sizes work correctly for one product but not the other. For example, suppose A is a 2×3 matrix and B is a 3×4 matrix. Then AB is defined, and is a 2×4 matrix. The other product, BA, does not exist! The sizes are 3×4 and 2×3 in this order, and the number of columns of B is not equal to the number of rows of A, so the product BA is not defined.

Second, the cancellation property that you rely on when solving equations in real numbers does not hold for matrix equations. For example, if you are solving the real-number equation $2x = 2 \cdot 3$, then, without giving it much thought, you cancel the 2 to find $x = 3$. Unfortunately, this property simply does not hold with matrix multiplication. If you have a matrix equation $AB = AC$, then you cannot cancel the A to find $B = C$! There is an example showing this in the text, and one of the suggested exercises gives you a chance to see it for yourself.

The lack of the cancellation property leads to another difference with real-number algebra. If you are solving the real-number equation

$$(x - 1)(x - 2) = 0,$$

you use the property that a product of real numbers is zero only if at least one factor is zero, and conclude that either $x - 1 = 0$ or $x - 2 = 0$, leading to the two solutions $x = 1$ and $x = 2$. In other words, you have used this property:

If $ab = 0$ for real numbers a and b, then either $a = 0$ or $b = 0$.

This does not hold for matrices! It is possible to have matrices A and B whose product is the zero matrix of the appropriate size, but neither A nor B is a zero matrix. You will find an example in the text and in your suggested exercises.

The transpose of a matrix

The transpose operation takes a matrix A and produces a new matrix A^T by making row i of A be column i of A^T. This converts an $m \times n$ matrix into an $n \times m$ matrix. For example, for the 2×3 matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$

the transpose is the 3×2 matrix

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

You should note that the rule obeyed by transposes of products of matrices is as follows:

$$(AB)^T = B^T A^T.$$

Your instincts might tell you that it should be $A^T B^T$ on the right-hand side, but it is not; the order is reversed. If you do a size check on the matrices, you will see that it has to be this way. If A is $m \times n$ and B is $n \times p$, then the product AB is $m \times p$ and $(AB)^T$ is therefore $p \times m$. On the right-hand side of our equation, B^T is $p \times n$ and A^T is $n \times m$, so the product $B^T A^T$ is defined, and has size $p \times m$, which agrees with the size of $(AB)^T$. (This just means that the sizes behave correctly, which gives us a chance for equality: the fact that they are equal takes more work to show.)

On the other hand, $A^T B^T$ may not be defined at all: for this product to be defined, we have satisfy the additional requirement that $p = m$, so we really should not expect a property like $(AB)^T = A^T B^T$ to hold.