Matrix Operations
Reading: Lay 2.1

September 23, 2013

You might remember from calculus that often we are interested in taking a function and changing it in some natural way to make a new function. This lecture is devoted to describing some natural ways to make new linear transformations from old.

As a side note, I approach this section's material a little differently from Lay, so note that the structure of these notes is not identical to that of Lay 2.1.

1 A Warm-Up

Let's forget about linear algebra for this warm-up section. Pretend you are back in calculus class and working with plain old functions of a real variable. Say we have functions $f, g : \mathbb{R} \longrightarrow \mathbb{R}$. For instance, let's think about $f(x) = x^2$ and $g(x) = e^x$.

One natural way to make new functions from old ones is to add them. For instance, we could define the new function h, which maps real numbers according to the rule

$$h(x) = f(x) + g(x) = x^2 + e^x.$$

We can also multiply a function by a constant to get a new function–for instance, the function H defined by

$$H(x) = 3f(x) = 3x^2.$$

1.1 Composition

The last big way we will cover here is composition of functions. Say we have the functions f and g from above. Remember that the function $f \circ g$ is defined by the following rule: when we want to evaluate $(f \circ g)(x)$, we do it by first plugging x into g, then we plug the result $g(x)$ into f. Sometimes this is written $f(g(x))$. So if $f(x) = x^2$ and $g(x) = \mathrm{e}^x$, then

$$f(g(x)) = \mathrm{e}^{2x}.$$

Note that in this case, $f \circ g$ is not the same as $g \circ f$. Indeed,

$$g(f(x)) = \mathrm{e}^{x^2}.$$

2 Matrix Sums, Scalar Multiples

Say we have two $m \times n$ matrices A and B (note they have the same dimensions). Multiplication by A and multiplication by B are two functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. We can define a new function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ with the following property: for every $\mathbf{x} \in \mathbb{R}^n$, we have

$$f(\mathbf{x}) = A\mathbf{x} + B\mathbf{x}. \tag{1}$$

This is analogous to the definition for functions given in the warm-up section.

The first question we should ask: is f a linear transformation? That is, is $f(\mathbf{x}) = C\mathbf{x}$ for some matrix C? It is easy to see that it is, and that the matrix C looks like we would expect. Let's number the columns of A and B:

$$A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}, \quad B = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_n \end{bmatrix}$$

Then

$$\begin{aligned} A\mathbf{x} + B\mathbf{x} &= x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n + x_1\mathbf{b}_1 + \ldots + x_n\mathbf{b}_n \\ &= x_1(\mathbf{a}_1 + \mathbf{b}_1) + \ldots + x_n(\mathbf{a}_n + \mathbf{b}_n). \end{aligned}$$

So we see that $f(\mathbf{x})$ is a linear transformation, and that for any $\mathbf{x} \in \mathbb{R}^n$, the value of $f(\mathbf{x})$ is given by the following matrix product:

$$\begin{bmatrix} \mathbf{a}_1 + \mathbf{b}_1 & \mathbf{a}_2 + \mathbf{b}_2 & \dots & \mathbf{a}_n + \mathbf{b}_n \end{bmatrix} \mathbf{x}$$

2.1 $A + B$

So using what we saw above, we can define the matrix sum $A + B$ to be the matrix whose columns are the sums of the corresponding columns of A and B. Such a matrix $A + B$ has the property that $(A + B)(\mathbf{x}) = A\mathbf{x} + B\mathbf{x}$.

Example 2.1. Say we have the 2×3 matrices

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

Then the matrix $A+B$ can be evaluated by adding the corresponding entries in A and B:

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

As we saw above, this $A + B$ has the property that $(A+B)\mathbf{x} = A\mathbf{x} + B\mathbf{x}$ for all $\mathbf{x}$. For instance,

$$(A+B)\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$$

and

$$A\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + B\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}.$$

2.2 cA

Say we are again given a matrix $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. We want to do something similar to the above but for constant multiples. That is, given some scalar c, we consider the function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, where

$$f(\mathbf{x}) = cA\mathbf{x}.$$

In this case, it is even easier to see that f is a linear transformation whose standard matrix is the matrix obtained from A by multiplying each entry of A by c. We denote this matrix by cA.

Example 2.2. Recall the matrix A from Example 2.1, and let us use the scalar $c = 3$. Then

$$3A = \begin{bmatrix} 3 & 9 & 9 \\ 6 & 0 & 0 \\ 3 & 3 & -3 \end{bmatrix}.$$

2.3 Addition and Scalar Multiplication Behave Nicely

We will first need a definition which is somewhat silly.

Definition 2.3. We call the matrix whose entries are all zeros the **zero matrix** and write it as 0. Note that this is the same symbol we use for the number zero. This is an unfortunate historical accident of mathematics. Please try not to confuse the zero matrix with the number zero.

It would be nice to have addition and scalar multiplication of matrices work together in the way you are used to. For instance, we would really be happy to have relationships like

$$A + A + A = 3A$$

be true.The good news is that such relationships *are* true. This is the content of the following theorem, which is not so difficult to prove using the definitions.

Theorem 2.4. *Let A, B, and C be $m \times n$ matrices (note: they are the same size!), and let r and s be scalars. Then the following relationships hold:*

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $A + 0 = A$
4. $r(A + B) = r A + r B$
5. $(r+s)A = r A + s A$
6. $r(sA) = (rs)A$.

3 Matrix Multiplication

Analogous to the composition $f \circ g$ from the warm-up, we want to consider what happens when we multiply a vector by a matrix B and then multiply the result by another matrix A. That is, we want to consider products of the form $A(B\mathbf{x})$, where A is $m \times n$ and B is $n \times p$ (note that these are not necessarily the same dimensions anymore!), and $\mathbf{x}$ is in $\mathbb{R}^p$.

So suppose

$$A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}, \quad B = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_p \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}.$$

Then

$$\begin{aligned} A(B\mathbf{x}) &= A(x_1\mathbf{b}_1 + \ldots + x_p\mathbf{b}_p) \\ &= x_1 A\mathbf{b}_1 + \ldots + x_p A\mathbf{b}_p \\ &= \begin{bmatrix} A\mathbf{b}_1 & \dots & A\mathbf{b}_p \end{bmatrix} \mathbf{x}. \end{aligned}$$

So multiplying $\mathbf{x}$ by B and then by A is the same as multiplying $\mathbf{x}$ by the matrix

$$\begin{bmatrix} A\mathbf{b}_1 & \dots & A\mathbf{b}_p \end{bmatrix}. \tag{2}$$

We will call the matrix in (2) by the name AB. What we have just proved is the following:

Theorem 3.1. *If A is an $m \times n$ matrix and B is an $n \times p$ matrix with columns $\mathbf{b}_1, \ldots, \mathbf{b}_p$, then for all $\mathbf{x} \in \mathbb{R}^p$, we have*

$$A(B(\mathbf{x})) = (AB)\mathbf{x},$$

where the $m \times p$ matrix AB is given by

$$AB = \begin{bmatrix} A\mathbf{b}_1 & \dots & A\mathbf{b}_p \end{bmatrix}.$$

Example 3.2. Suppose we have the matrices

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

Then

$$AB = \begin{bmatrix} A\begin{bmatrix}1\\2\\0\end{bmatrix} & A\begin{bmatrix}1\\-1\\3\end{bmatrix}\end{bmatrix} = \begin{bmatrix}5 & 7\\2 & 4\end{bmatrix}.$$

Please note what sizes A, B, and AB are in the definition of the product!

3.1 Row-Column Rule for AB

Say we again have matrices A and B, where A is an $m \times n$ matrix and B is an $n \times p$ matrix. Suppose that we write the entry in the ith row and jth column of A as a_{ij} and similarly for B and b_{ij}. Then probably the most efficient way to compute the entries of AB, which we write as $(AB)_{ij}$, is the following rule.

Theorem 3.3 (Row-column rule)**.** *The entry $(AB)_{ij}$, that is the entry of the matrix AB which is in the ith row and jth column, is given by the formula*

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj}.$$

It is good to do lots of practice computing matrix products.

Example 3.4. Say we have two matrices

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

What is $(AB)_{2\,3}$?

By the row-column rule, we can compute it using the 2nd row of A and the $3rd$ column of B:

$$(AB)_{2\,3} = 3(2) + 0(1) = 6.$$

Lay does a number of examples of computation with this method.

3.2 Properties of matrix multiplication

One thing to note is that if the product AB is defined, the product BA can still be undefined if the sizes of the matrices are not the appropriate size. Moreover, even if BA is defined, it is not necessarily true that $AB = BA$ (this is similar to the case of f and g from the warm-up section). Nevertheless, there are a number of properties of the matrix product which make life easier, which we list in the following theorem. First, a definition:

Definition 3.5. We define I_n to be the "$n \times n$ identity matrix": the matrix whose entries are all zero except for its diagonal entries (the entries $(I_n)_{11}$, $(I_n)_{22}$, etc), which are all one. For instance,

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The main property of the identity matrix is that if $\mathbf{x}$ is any vector in $\mathbb{R}^n$, then

$$I_n\mathbf{x} = \mathbf{x}.$$

Theorem 3.6. *Let A be an $m \times n$ matrix, let B and C have sizes so that the products listed below are defined, and let r be an arbitrary scalar. Then*

1. $A(BC) = (AB)C$
2. $A(B + C) = AB + AC$
3. $(B + C)A = BA + CA$
4. $r(AB) = (rA)B = A(rB)$
5. $I_mA = A = AI_n$.

All of the properties in Theorem 3.6 are again fairly simple to verify using the definition of matrix multiplication. If you want to see some proofs, let me know and I will add an appendix to these notes.

3.3 Caveats

Lay mentions some "warnings" about matrix multiplication which are good to note. One which you have already seen is that AB is not usually equal to BA. Another is that it is not always possible to cancel. That is, if $AB = AC$, then it is not necessarily true that $B = C$ (though sometimes it happens). Another fact is that if $AB = 0$, then it is not necessarily true that $A = 0$ or $B = 0$.

You should try to find matrices that illustrate the warnings above on your own. I will show you one such example.

Example 3.7. Let

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

Then

$$AB = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$