Lecture 3e

More on Matrix Multiplication

(pages 124-127)

Now that we've had some practice multiplying matrices, let's take a look at some of its properties. The first thing I want to look at is an alternate definition for matrix multiplication. Let's recall our current definition:

Definition: Let B be an $m \times n$ matrix with rows $\vec{b}_1^T, \ldots, \vec{b}_m^T$, and let A be an $n \times p$ matrix with columns $\vec{a}_1, \ldots, \vec{a}_p$. Then we define BA to be the matrix whose ij-th entry is

$$(BA)_{ij} = \vec{b}_i \cdot \vec{a}_j$$

And now, let's take a closer look at the calculation $\vec{b}_i \cdot \vec{a}_j$:

$$\vec{b}_i \cdot \vec{a}_j = \begin{bmatrix} b_{i1} \\ b_{i2} \\ \vdots \\ b_{in} \end{bmatrix} \cdot \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix} =$$

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

We can rewrite this dot product using summation notation: $b_{i1}a_{1j} + b_{i2}a_{2j} + \cdots + b_{in}a_{nj} = \sum_{k=1}^{n} b_{ik}a_{kj}$. And this gives us a new definition for BA:

Definition: Let B be an $m \times n$ matrix and let A be an $n \times p$ matrix. Then the ij-th entry of BA is

$$(BA)_{ij} = \sum_{k=1}^{n} b_{ik}a_{kj} = \sum_{k=1}^{n} (B)_{ik}(A)_{kj}$$

This is, of course, exactly the same as the previous definition, but sometimes focusing on the actual calculation (instead of the bigger picture row-dot-column view) is helpful. For example, it is useful when proving the following theorem.

Theorem 3.1.3: If A,B, and C are matrices of the correct size so that the required products are defined, and $t \in \mathbb{R}$, then

(1) $A(B + C) = AB + AC$

(2) $t(AB) = (tA)B = A(tB)$

(3) $A(BC) = (AB)C$
(4) $(AB)^T = B^T A^T$

Proof of Theorem 3.1.3 (1) Suppose A,B, and C are matrices such that $A(B + C)$ is defined, and let $D = B + C$. Then $(D)_{ij} = d_{ij} = b_{ij} + c_{ij}$. So $(A(B+C))_{ij} = (AD)_{ij} = \sum a_{ik}d_{kj} = \sum a_{ik}(b_{kj} + c_{kj}) = \sum(a_{ik}b_{kj} + a_{ik}c_{kj}) = \sum a_{ik}b_{kj} + \sum a_{ik}c_{kj} = (AB)_{ij} + (AC)_{ij} = (AB + AC)_{ij}$. So, we see that $(A(B + C))_{ij} = (AB + AC)_{ij}$ for all i, j, and thus $A(B + C) = AB + AC$.

I won't go through the proofs of (2) and (3), as they use this same kind of symbol manipulation, but (4) is slightly different, so I'll show that proof too.

Proof of Theorem 3.1.3 (4): Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. Note that $(B^T)_{ik} = (B)_{ki}$ and $(A^T)_{kj} = (A)_{jk}$. This means that $(B^T A^T)_{ij} = \sum_{k=1}^{n}(B^T)_{ik}(A^T)_{kj} = \sum_{k=1}^{n}(B)_{ki}(A)_{jk} = \sum_{k=1}^{n}(A)_{jk}(B)_{ki} = (AB)_{ji}$. So, we've seen that $(B^T A^T)_{ij} = (AB)_{ji}$ for all i and j, which means that $B^T A^T = (AB)^T$.

Theorem 3.1.3 is a list of useful properties of matrix multiplication, but there are other useful properties that we would like to add. Unfortunately, we can't. That is, there are many properties we might like to have, but they aren't actually true. At the end of the previous lecture I pointed out the first of these–the fact that matrix multiplication is not commutative. That is, in general, $AB \neq BA$. In fact, we saw that BA may not even be defined even if AB is, and that even if both AB and BA are defined they might not even be the same size. Mind you, it does sometimes happen that $AB = BA$, so we can't even say that $AB \neq BA$ is always true.

The next thing we need to notice is that the cancelation law does not hold. That is, if $AB = AC$, this does not necessarily mean that $B = C$. This fact is best reinforced with an example.

Example: Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 4 & 8 \\ -1 & 3 \end{bmatrix}$, and $C = \begin{bmatrix} -5 & 5 \\ 8 & 6 \end{bmatrix}$. Then:

$$AB = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 4 & 8 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 11 \\ 6 & 22 \end{bmatrix}$$

and

$$AC = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} -5 & 5 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 11 \\ 6 & 22 \end{bmatrix} = AB$$

even though B does not equal C.

The fact that we don't have a cancelation law means that we can't define matrix division either. All in all, it is good to keep in mind that matrix multiplication is quite a bit different from the multiplication of real numbers. So you should pay extra special attention to the list from Theorem 3.1.3, because those are the only rules you can use.

There is something *similar* to the cancelation law that is true:

Theorem 3.1.4: If A and B are $m \times n$ matrices such that $A\vec{x} = B\vec{x}$ for every $\vec{x} \in \mathbb{R}^n$, then $A = B$.

Proof of Theorem 3.1.4: The key fact here is that $A\vec{x} = B\vec{x}$ FOR ALL $\vec{x}$, not just any one in particular. So, if we recall that $\vec{e}_j$ is the j-th standard basis vector (that is, it is a vector that has a 1 in the j-th component, and all other components are zeros), then we know that $A\vec{e}_j = B\vec{e}_j$ for all $1 \leq j \leq n$. But what is $A\vec{e}_j$? Well, first of all, note that it is an $n \times 1$ matrix, so we need only look at $(A\vec{e}_j)_{i1}$. Using our summation definition, we see that $(A\vec{e}_j)_{i1} = \sum_{k=1}^{n}(A)_{ik}(\vec{e}_j)_{k1}$. But since $(\vec{e}_j)_{k1}$ is zero for $k \neq j$ (and equals 1 when $k = j$), we have that $(A\vec{e}_j)_{i1} = (A)_{ij}$ for all i. This means that $A\vec{e}_j$ is the j-th column of A. Similarly, $B\vec{e}_j$ is the j-th column of B. But we have $A\vec{e}_j = B\vec{e}_j$, so this means that the j-th column of A is the same as the j-th column of B. And since this is true for every j, we see that $A = B$.

The fact that $A\vec{e}_j$ is the j-th column of A is actually quite useful. More specifically, the fact that if we line up all the products $A\vec{e}_1 \; A\vec{e}_2 \; \cdots \; A\vec{e}_n$, then we get another copy of A. And so we use this fact to define a very special matrix.

Definition: The $n \times n$ matrix $I_n = \text{diag}(1, 1, \ldots, 1)$ is called the identity matrix. That is, the identity matrix is a diagonal matrix, with all the diagonal entries equal to 1.

Examples: $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

Notation: Often we simply use I to denote an identity matrix, with the expectation that the size of I can be determined from context.

And the reason we care about the identity matrix is:

Theorem 3.1.5 If A is any $m \times n$ matrix, then $I_m A = A = A I_n$.

I've been giving the proof to a lot of things that the textbook leaves as exercises during this lecture, so I think this time I will go ahead and assign this proof as a practice problem.