

Matrix multiplication: why row times column?

Sean Fitzpatrick

2015-01-26

Multiplication

A few students have asked about the reasons for defining matrix multiplication the way we do. It might seem like it would be more natural to take to matrices of the same size and multiply the corresponding entries, the same as we do for addition. We'd even keep most of the same properties (and we'd gain one): suppose $A = [a_{ij}]$ and $B = [b_{ij}]$ are $m \times n$ matrices. We could define a product \widetilde{AB} (let's not call it AB since that's already been defined) by

$$\widetilde{AB} = [a_{ij}b_{ij}],$$

and then we'd have $\widetilde{AB} = \widetilde{BA}$ (and it's **not** true that $AB = BA$ for our definition), plus the other properties (distributive, associative, etc.).

So, it's not for want of nice algebraic properties that we make the definition that we do. One hint at a reason comes from our desire to re-write a system of linear equations in compact form: with the row-times-column rule, if

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \text{ and } B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

then the matrix equation $AX = B$ corresponds to the system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

OK, so this is nice enough, but it seems like a lot of trouble to go to for a bit of notational convenience. Is there another reason? There is, but it goes a bit beyond what we cover in Math 1410. The source of the rule comes from the theory of *linear transformations*. These are functions between vector spaces which have special properties from the point of view of linear algebra.

In this course we don't consider vector spaces in general. We stick to one set of examples: the vector spaces \mathbb{R}^n of $n \times 1$ column vectors, where n is a natural number. For example,

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x, y \in \mathbb{R} \right\}, \mathbb{R}^3 = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x, y, z \in \mathbb{R} \right\},$$

and so on. (We won't discuss these until later in the course.) A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (that is, the domain of f is \mathbb{R}^n and the range of f lies in \mathbb{R}^m) is called a **linear transformation** if it respects the addition and scalar multiplication in both spaces: $f(X + Y) = f(X) + f(Y)$ and $f(cX) = cf(X)$ for any vectors $X, Y \in \mathbb{R}^n$ and scalar $c \in \mathbb{R}$.

If $n = m = 1$, any linear transformation $f : \mathbb{R} \rightarrow \mathbb{R}$ is of the form $f(x) = ax$. (So the graph is a straight line through the origin.) Notice what happens if we have two such functions and we compose them: if $f(x) = ax$ and $g(x) = bx$, then $f(g(x)) = a(bx) = (ab)x$, so we just multiply the two constants.

Let's go up one dimension. A linear function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ looks like

$$f \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} a_{11}x + a_{12}y \\ a_{21}x + a_{22}y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

So far, so good: in one dimension, we have $f(x) = ax$. In two dimensions, we have $f(X) = AX$, which is the same pattern, except X is now a column vector, and A is now a 2×2 matrix. What if we have more than one such function? Let's say

$$g(X) = \begin{bmatrix} b_{11}x + b_{12}y \\ b_{21}x + b_{22}y \end{bmatrix} = BX,$$

where $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$. What is $f(g(X))$? If we write $g(X) = \begin{bmatrix} u \\ v \end{bmatrix}$, where $u = b_{11}x + b_{12}y$ and $v = b_{21}x + b_{22}y$, we get

$$\begin{aligned} f(g(X)) &= f \left(\begin{bmatrix} u \\ v \end{bmatrix} \right) \\ &= \begin{bmatrix} a_{11}u + a_{12}v \\ a_{21}u + a_{22}v \end{bmatrix} \\ &= \begin{bmatrix} a_{11}(b_{11}x + b_{12}y) + a_{12}(b_{21}x + b_{22}y) \\ a_{21}(b_{11}x + b_{12}y) + a_{22}(b_{21}x + b_{22}y) \end{bmatrix} \\ &= \begin{bmatrix} (a_{11}b_{11} + a_{12}b_{21})x + (a_{11}b_{12} + a_{12}b_{22})y \\ (a_{21}b_{11} + a_{22}b_{21})x + (a_{21}b_{12} + a_{22}b_{22})y \end{bmatrix} \\ &= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \end{aligned}$$

On the other hand, we might expect that we can just write

$$f(g(X)) = A(g(X)) = A(BX) = (AB)X.$$

For this to work, we would need to have

$$AB = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}.$$

But this is exactly the rule for multiplying two 2×2 matrices! From here, it's possible to show that the same is true in all other cases. (A general proof exists, but it's a bit complicated, so we'll omit it here. The interested reader can refer to any linear algebra textbook that covers linear transformations.)

So the moral of the story is that matrix multiplication is defined so that multiplying matrices corresponds to composing the corresponding linear transformations. But wait, there's more! This moral also explains the reasons for the sizes of the matrices being what they are. In general, we can show that any linear transformation

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

must be given by $f(X) = AX$, where A is an $m \times n$ matrix. (Recall that if A is $m \times n$, the equation $AX = B$ represents a system of m equations in n variables: the column vector X belongs to \mathbb{R}^n , and the column vector B belongs to \mathbb{R}^m .)

In order to form the composition $f(g(X))$ of two functions f and g , the range of g must be contained in the range of f . So if f is a linear transformation from \mathbb{R}^n to \mathbb{R}^m (given by an $m \times n$ matrix A), the domain of f is \mathbb{R}^n , so we need to have $g(X) \in \mathbb{R}^n$. That means we can take $g : \mathbb{R}^k \rightarrow \mathbb{R}^n$ for some k . Now, notice two things:

1. Since g is a function from \mathbb{R}^k to \mathbb{R}^n , it's given by an $n \times k$ matrix B . Since A is $m \times n$ and B is $n \times k$, the product AB makes sense.
2. The composite function $f(g(X))$ is a function from \mathbb{R}^k to \mathbb{R}^m (since X goes from \mathbb{R}^k to \mathbb{R}^n and then to \mathbb{R}^m), so it should be given by an $m \times k$ matrix, and that's exactly the size of matrix we get if A is $m \times n$ and B is $n \times k$.

Inverses

In calculus (or high school pre-calculus) we learn that if a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is *one-to-one*, then it's possible to define an *inverse function* $f^{-1} : \mathbb{R} \rightarrow \mathbb{R}$ by the property

$$y = f(x) \text{ if and only if } f^{-1}(y) = x$$

A little more generally, if the domain of f is $A \subseteq \mathbb{R}$ and the range is $B \subseteq \mathbb{R}$, then f is a function from A to B and f^{-1} is a function from B to A . The defining property above is equivalent to the following cancellation laws:

$$f^{-1}(f(x)) = x \text{ for all } x \in A \text{ and } f(f^{-1}(y)) = y \text{ for all } y \in B.$$

Now, as we noted above, any $m \times n$ matrix A defines a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ by

$$f(X) = AX \text{ for any } X \in \mathbb{R}^n.$$

We can define one-to-one functions in this context in the exact same way as for functions of one real variable. It's possible to show that f can only be one-to-one if $m \leq n$, and that the range of f is all of \mathbb{R}^n only if $m \geq n$. (The argument uses facts about dimension proved in Math 3410.) It follows that it's possible to define an inverse function

$$f^{-1} : \mathbb{R}^m \rightarrow \mathbb{R}^n$$

if and only if (a) f is one-to-one, and (b) $m = n$. So we can only talk about an inverse to the function $f(X) = AX$ if $m = n$, which means that A has to be an $n \times n$ *square* matrix. In fact, it's possible to prove the following:

Theorem 1. *Let A be an $n \times n$ matrix, and let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the function defined by*

$$f(X) = AX.$$

Then the function f is one-to-one if and only if the matrix A has an inverse, and $f^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is given by

$$f^{-1}(Y) = A^{-1}Y.$$

To see why this works, recall that the inverse of a matrix A is defined by the equations

$$AA^{-1} = I_n \quad \text{and} \quad A^{-1}A = I_n,$$

where I_n is the $n \times n$ identity matrix. Given $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, let

$$Y = f(X) = AX.$$

If we set $f^{-1}(Y) = A^{-1}Y$, then

$$f^{-1}(f(X)) = A^{-1}(f(X)) = A^{-1}(AX) = (A^{-1}A)X = I_n X = X,$$

and

$$f(f^{-1}(Y)) = A(f^{-1}(Y)) = A(A^{-1}Y) = (AA^{-1})Y = I_n Y = Y.$$

Thus, we get the same cancellation laws for the inverse function as in the one variable case.