Why matrix multiplication?

It seems a bit strange the way we multiply two matrices. So let's see why it isn't.

1 Vectors and linear transformations

Let's assume column vectors.

$$x = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{1}$$

There's a special kind of transformation (function) on these vectors, linear transformations. They're special because it's easy to analyze linear systems, their kernels are vector spaces (physically speaking, we can superimpose solutions to $Ax = 0$), lots of things in physics behave linearly, and composition and addition of linear transformations are linear too.

How do we represent this abstract concept? Well, how do we represent vectors in the first place? By choosing a nice basis. So x really is

$$x = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{2}$$

where we have called our (orthonormal) basis $[1, 0, 0], [0, 1, 0], [0, 0, 1]$. Defining a representation for vectors in terms of coordinates (ie choosing an orthornormal basis) is almost arbitrary; we want to try and do as much as we can without it. Physically, we are born into a world where vectors seem to live in $\mathbb{R}^3$, but we aren't born with triplet of basis vectors, and it shouldn't matter which three we choose. Now observe what a linear transformation T does to x:

$$Tx = T \left(a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \tag{3}$$

$$= aT \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + bT \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + cT \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4}$$

which means to write down T we just need to write down $Tb_1, Tb_2...$ or its effect on the basis vectors. Let's just write them down next to each other.

$$T = \begin{bmatrix} | & | & | \\ T(e_1) & T(e_2) & T(e_3) \\ | & | & | \end{bmatrix} \tag{5}$$

Then Tx is a linear combination of the columns of T where the coefficients are the entries of x. This is of course the same as the normal rule for matrix multiplication in the special case where x is a column vector, but this way of interpreting Ax is, I think, more intuitive and meaningful.

In that case, let S be another matrix. What is ST? We don't really know, but we do know that $S(T(x))$ is a linear transformation of x. So let's say $ST = S \circ T$ because we would like to have $(ST)x = S(Tx)$. How do we represent ST? Clearly we must transform the columns of T.

$$ST = S \begin{bmatrix} | & | & | \\ T_1 & T_2 & T_3 \\ | & | & | \end{bmatrix} \tag{6}$$

$$= \begin{bmatrix} | & | & | \\ ST_1 & ST_2 & ST_3 \\ | & | & | \end{bmatrix} \tag{7}$$

and with a bit of fiddling we see that this is exactly the same as normal matrix multiplication. This also explains why matrix multiplication is associative; because function composition is associative.

2 Duality; inner and outer product

Instead of what we have been using so far, let

$$x = \begin{bmatrix} a & b & c \end{bmatrix} \tag{8}$$

and T be a 3×3 matrix. Tx is dimensionally wrong, but xT isn't. In our framework so far we would consider x to be a linear transformation, but let's call it a row vector instead. Then T is indeed a linear transformation.

$$\begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} - & T_1 & - \\ - & T_2 & - \\ - & T_3 & - \end{bmatrix} = xT \tag{9}$$

$$= a \begin{bmatrix} - & T_1 & - \end{bmatrix} \tag{10}$$

$$+ b \begin{bmatrix} - & T_2 & - \end{bmatrix} \tag{11}$$

$$+ c \begin{bmatrix} - & T_3 & - \end{bmatrix} \tag{12}$$

so xT is a linear combination of the rows of T where the coefficients are the entries of x.

Why do we choose to use column vectors instead of row vectors, since neither are "special"? Column vectors make application of linear transformations prefix rather than postfix operators. Also, with row vectors $S \circ T = TS$, a consequence of the postfix notation.

Any column of ST is a linear combination of columns of S, where the coefficients are entries from a column of T. Any row of ST is a linear combination of rows of T, where the coefficients are entries from a row of S.

$$\begin{bmatrix} - & - \\ - & - \end{bmatrix} \begin{bmatrix} | & | \\ | & | \end{bmatrix} \tag{13}$$

If you think about it long enough, then, the entries of ST must be the products of the rows of S and the columns of T. The diagram above neatly summarizes all this duality stuff. Mentally imagine the rows/columns transposing itself, grabbing rows/columns from the other matrix, and adding them up.

In fact, in more abstract treatments of vector spaces, one defines axiomatically the concept of a vector space over a field F, . Then the dual vector space F^* is the space of all continuous linear functions from $V \to F$. If we have the idea of a (positive-definite) inner product $\langle u, v \rangle = \langle v, u \rangle^*$ we see that $\langle u, \cdot \rangle$ is an element of F^*, and there is a theorem that any element of F^* is uniquely represented in this form. We see that inner products lead naturally to the idea of taking the conjugate transpose of a vector; likewise if we let A be a linear operator (in the finite case, represented by a matrix) and define $A^\dagger$ to be the operator such that $\langle A^\dagger u, v \rangle = \langle u, Av \rangle$ we are led to the idea of conjugate-transposing a matrix.

3 lecture 30

Projection matrix. Choose a basis to be on the projection line and off.

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{14}$$

In the standard basis

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{15}$$

What if
$D = \frac{d}{dx}$? Input: $c_0 + c_1 x^1 + c_2 x^2$
Output $= c_1 + 2c_2$

The reason we can take derivatives quickly is that the derivative is a linear transformation, otherwise it would take forever.

$$D = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{16}$$