1111: Linear Algebra I

Dr. Vladimir Dotsenko (Vlad)

Michaelmas Term 2015

Matrix product

One definition is immediately built upon what we just defined before. Let A be an $m \times n$-matrix, and B an $n \times k$-matrix. Their product $A \cdot B$, or AB, is defined as follows: it is the $m \times k$-matrix C whose columns are obtained by computing the products of A with columns of B:

$$A \cdot (\mathbf{b}_1 \mid \mathbf{b}_2 \mid \ldots \mid \mathbf{b}_k) = (A \cdot \mathbf{b}_1 \mid A \cdot \mathbf{b}_2 \mid \ldots \mid A \cdot \mathbf{b}_k)$$

Another definition states that the product of an $m \times n$-matrix A and an $n \times k$-matrix B is the $m \times k$-matrix C with entries

$$C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj}$$

(here i runs from 1 to m, and j runs from 1 to k). In other words, C_{ij} is the "dot product" of the i-th row of A and the j-th column of B.

Examples

Let us take $U = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $V = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $W = \begin{pmatrix} 2 & 3 & 1 \\ 5 & 2 & 0 \end{pmatrix}$.

Note that the products $U \cdot U$, $U \cdot V$, $V \cdot U$, $V \cdot V$, $U \cdot W$, and $V \cdot W$ are defined, while the products $W \cdot U$, $W \cdot V$, and $W \cdot W$ are not defined.

We have $U \cdot U = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $U \cdot V = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $V \cdot U = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$,
$V \cdot V = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $U \cdot W = \begin{pmatrix} 5 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $V \cdot W = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 3 & 1 \end{pmatrix}$.

In particular, even though both matrices $U \cdot V$ and $V \cdot U$ are defined, they are not equal.

Matrix product: third definition

However, these two definitions appear a bit *ad hoc*, without no good reason to them. The third definition, maybe a bit more indirect, in fact sheds light on why the matrix product is defined in exactly this way.

Let us view, for a given $m \times n$-matrix A, the product $A \cdot \mathbf{x}$ as a rule that takes a vector $\mathbf{x}$ with n coordinates, and computes out of it another vector with m coordinates, which is denoted by $A \cdot \mathbf{x}$. Then, given two matrices, an $m \times n$-matrix A and an $n \times k$-matrix B, from a given vector $\mathbf{x}$ with k coordinates, we can first use the matrix B to compute the vector $B \cdot \mathbf{x}$ with n coordinates, and then use the matrix A to compute the vector $A \cdot (B \cdot \mathbf{x})$ with m coordinates.

By definition, the product of the matrices A and B is the matrix C satisfying

$$C \cdot \mathbf{x} = A \cdot (B \cdot \mathbf{x}) \ .$$

Equivalence of the definitions

The first and the second definition are obviously equivalent: the entry in the i-th row and the j-th column of the matrix

$$(A \cdot \mathbf{b}_1 \mid A \cdot \mathbf{b}_2 \mid \ldots \mid A \cdot \mathbf{b}_k)$$

is manifestly equal to $A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj}$. (Note that $\begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix}$ is precisely $\mathbf{b}_j$, the j-th column of B).

Equivalence of the definitions

For the third definition, note that the property $C \cdot \mathbf{x} = A \cdot (B \cdot \mathbf{x})$ must hold for all $\mathbf{x}$, in particular for $\mathbf{x} = \mathbf{e}_j$, the standard unit vector which has the j-th coordinate equal to 1, and all other coordinates equal to zero.

Note that for each matrix M the vector $M \cdot \mathbf{e}_j$ (if defined) is equal to the j-th column of M. In particular, $A \cdot (B \cdot \mathbf{e}_j) = A \cdot \mathbf{b}_j$. Therefore, we must use as C the matrix $A \cdot B$ from the first definition (whose columns are the vectors $A \cdot \mathbf{b}_j$): only in this case $C \cdot \mathbf{e}_j = A \cdot \mathbf{b}_j = A \cdot (B \cdot \mathbf{e}_j)$ for all j. To show that $C \cdot \mathbf{x} = A \cdot (B \cdot \mathbf{x})$ for all vectors $\mathbf{x}$, we note that such a vector can be represented as $x_1\mathbf{e}_1 + \cdots + x_k\mathbf{e}_k$, and then we can use properties of products of matrices and vectors:

$$
\begin{aligned}
&A \cdot (B \cdot \mathbf{x}) = A \cdot (B \cdot (x_1\mathbf{e}_1 + \cdots + x_k\mathbf{e}_k)) = \\
&= A \cdot (x_1(B \cdot \mathbf{e}_1) + \cdots + x_k(B \cdot \mathbf{e}_k)) = x_1 A \cdot (B \cdot \mathbf{e}_1) + \cdots + x_k A \cdot (B \cdot \mathbf{e}_k) = \\
&= x_1 C \cdot \mathbf{e}_1 + \cdots + x_k C \cdot \mathbf{e}_k = C \cdot (x_1\mathbf{e}_1 + \cdots + x_k\mathbf{e}_k) = C \cdot \mathbf{x}.
\end{aligned}
$$

Properties of the matrix product

Let us show that the matrix product we defined satisfies the following properties (whenever all matrix operations below make sense):

$$A\cdot(B+C)=A\cdot B+A\cdot C,$$
$$(A+B)\cdot C=A\cdot C+B\cdot C,$$
$$(c\cdot A)\cdot B=c\cdot(A\cdot B)=A\cdot(c\cdot B),$$
$$(A\cdot B)\cdot C=A\cdot(B\cdot C)$$

All these proofs can proceed in the same way: pick a "test vector" $\mathbf{x}$, multiply both the right and the left by it, and test that they agree. (Since we can take $\mathbf{x}=\mathbf{e}_j$ to single out individual columns, this is sufficient to prove equality).

For example, the first equality follows from

$$(A\cdot(B+C))\cdot\mathbf{x}=A\cdot((B+C)\cdot\mathbf{x})=A\cdot(B\cdot\mathbf{x}+C\cdot\mathbf{x})=$$
$$A\cdot(B\cdot\mathbf{x})+A\cdot(C\cdot\mathbf{x})=(A\cdot B)\cdot\mathbf{x}+(A\cdot C)\cdot\mathbf{x}=(A\cdot B+A\cdot C)\cdot\mathbf{x}$$

The identity matrix

Let us also define, for each n, the *identity* matrix I_n, which is an $n \times n$-matrix whose diagonal elements are equal to 1, and all other elements are equal to zero.

For each $m \times n$-matrix A, we have $I_m \cdot A = A \cdot I_n = A$. This is true because for each vector $\mathbf{x}$ of height p, we have $I_p \cdot \mathbf{x} = \mathbf{x}$. (The matrix I_p does not change vectors; that is why it is called the identity matrix). Therefore,

$$(I_m \cdot A) \cdot \mathbf{x} = I_m \cdot (A \cdot \mathbf{x}) = A \cdot \mathbf{x},$$
$$(A \cdot I_n) \cdot \mathbf{x} = A \cdot (I_n \cdot \mathbf{x}) = A \cdot \mathbf{x} \ .$$