1111: Linear Algebra I

Dr. Vladimir Dotsenko (Vlad)

Michaelmas Term 2014

Two possibly useful remarks

I was made aware of a rather good (for some parts of the course, at least) linear algebra textbook, "Linear Algebra" by Jim Hefferon. The best thing is that it is available online for free:

`http://joshua.smcvt.edu/linearalgebra/`

It is also important that this book has many examples and exercises. I am fully aware that sometimes the amount of examples we do in class / in homeworks is not enough, so this book is a great source of exercises, and these exercises have answers online too. All in all, a wonderful practice kit.

As it happens every year, School of Maths is running "maths helprooms" for those of you who want to ask something about the things that are confusing in some of the modules. It takes place in the "old seminar room" of School of Maths (room 2.6) every day 1-2pm, and also 12-1pm on Tuesdays and Thursdays. Bring whatever concerns you have, and sophister students there will do their best to explain things in a down to earth way.

Previously on...

Last week, you learned about *reduced row echelon forms*, and got some experience of using *elementary row operations* to bring matrices to their reduced row echelon forms.

Roughly speaking, that teaches us that the solution set of a system of linear equations, when not empty, can be parametrised: you can pick some unknowns as arbitrary parameters, and then all others can be uniquely determined. Reduced row echelon forms teach us precisely which to choose, and how to determine others.

But... wait! If we have one equation with one unknown, $ax = b$, then we can just right $x = b/a$, solving everything in one go. Maybe we can do something similar for many unknowns? It turns out that there is a way to re-package our approach into something similar; this is our goal for this week.

Matrix arithmetic

Let us create an algebraic set-up for all that. Protagonists: *vectors* (*columns* of coordinates) and *matrices* (rectangular arrays of coordinates). Of course, a vector is a particular case of a matrix (with only one column).

We know that the two most basic operators on vectors are addition and re-scaling. The same works for matrices, component-wise. Of course, to add two matrices, they must have the same dimensions:

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \ldots & \ddots & \ldots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} + \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots & \ldots & \ddots & \ldots \\ B_{m1} & B_{m2} & \cdots & B_{mn} \end{pmatrix} =$$

$$= \begin{pmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \cdots & A_{1n}+B_{1n} \\ A_{21}+B_{21} & A_{22}+B_{22} & \cdots & A_{2n}+B_{2n} \\ \vdots & \ldots & \ddots & \ldots \\ A_{m1}+B_{m1} & A_{m2}+B_{m2} & \cdots & A_{mn}+B_{mn} \end{pmatrix}$$

Matrix arithmetic

Next, we define products of matrices and vectors. For that, we once again examine a system of m simultaneous linear equations with n unknowns

$$\begin{cases} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n = B_1, \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,n}x_n = B_2, \\ \qquad\qquad\qquad\qquad\qquad\qquad \cdots \\ A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n = B_m. \end{cases}$$

We introduce new notation for it, $A \cdot \mathbf{x} = \mathbf{b}$ (or even $A\mathbf{x} = \mathbf{b}$), where

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \ldots & \ddots & \ldots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{pmatrix}.$$

Note that this new notation is a bit different from the one last week, where A denoted the matrix including $\mathbf{b}$ as the last column.

Matrix arithmetic

In other words, for an $m \times n$-matrix A, and a column $\mathbf{x}$ of height n, we define the column $\mathbf{b} = A \cdot \mathbf{x}$ as the column of height m whose k-th entry is $B_k = A_{k1}x_1 + \cdots + A_{kn}x_n$:

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \dots & \ddots & \dots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} A_{11}x_1 + \cdots + A_{1n}x_n \\ A_{21}x_1 + \cdots + A_{2n}x_n \\ \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n \end{pmatrix}$$

A useful mnemonic rule is that the entries of $A \cdot \mathbf{x}$ are "dot products" of rows of A with the column $\mathbf{x}$.

Properties of $A \cdot \mathbf{b}$

The products we just defined satisfy the following properties:

$$A \cdot (\mathbf{x}_1 + \mathbf{x}_2) = A \cdot \mathbf{x}_1 + A \cdot \mathbf{x}_2,$$
$$(A_1 + A_2) \cdot \mathbf{x} = A_1 \cdot \mathbf{x} + A_2 \cdot \mathbf{x},$$
$$c \cdot (A \cdot \mathbf{x}) = (c \cdot A) \cdot \mathbf{x} = A \cdot (c \cdot \mathbf{x}).$$

Here A, A_1, and A_2 are $m \times n$-matrices, $\mathbf{x}$, $\mathbf{x}_1$, and $\mathbf{x}_2$ are columns of height n (vectors), and c is a scalar.

Now we have all the ingredients to define products of matrices in the most general context. There will be three equivalent definitions, each useful for some purposes.

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 both 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:

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