1111: Linear Algebra I

Dr. Vladimir Dotsenko (Vlad)

Lecture 6

Gauss–Jordan elimination

Last time we discussed bringing matrices to reduced row echelon form. If a system of equations has a reduced row echelon matrix, it is very easy to describe all solutions.

If the last non-zero equation reads $0 = 1$, the system is clearly inconsistent.

If the pivot of last non-zero equation is a coefficient of some unknown, the system is consistent, and all solutions are easy to describe. For that, we shall separate unknowns into two groups, the *principal* (*pivotal*) unknowns, that is unknowns for which the coefficient in one of the equations is the pivot of that equation, and all the other ones, that we call *free* unknowns.

Once we assign arbitrary numeric values to free unknowns, each of the equations gives us the unique value of its pivotal unknown which makes the system consistent. Thus, we described the solution set in a parametric form using free unknowns as parameters.

Gauss–Jordan elimination: an example

Continuing with our example, the matrix

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 & -3 & 11 \\ 0 & 0 & 1 & 0 & 5 & -15 \\ 0 & 0 & 0 & 1 & -1 & 5 \end{pmatrix}$$

is in reduced row echelon form. The corresponding system of equations is

$$\begin{cases} x_1 + 2x_2 - 3x_5 = 11, \\ x_3 + 5x_5 = -15, \\ x_4 - x_5 = 5. \end{cases}$$

The pivotal unknowns are x_1, x_3, and x_4, and the free unknowns are x_2 and x_5. Assigning arbitrary parameters $x_2 := t_2$ and $x_5 := t_5$ to the free unknowns, we obtain the following description of the solution set:

$$x_1 = 11 - 2t_2 + 3t_5, x_2 = t_2, x_3 = -15 - 5t_5, x_4 = 5 + t_5, x_5 = t_5.$$

where t_2 and t_5 are arbitrary numbers.

Some remarks

In practice, we don't have to do first the row echelon form, then the reduced row echelon form: we can use the "almost pivotal" entry (the first non-zero entry of the row being processed) to cancel all other entries in its column, thus obtaining the reduced row echelon form right away.

The reduced row echelon form, unlike the row echelon form, is unique, that is does not depend on the type of row operations performed (there is freedom in which rows we swap etc.). We shall not prove it in this course.

For the intersection of two 2D planes in 3D, if the planes are not parallel, the reduced row echelon form will have one free variable, which can be taken as a parameter of the intersection line. More generally, one linear equation in n unknowns defines an $(n-1)$-dimensional plane, and we just proved that the intersection of several planes can be parametrised in a similar way (by free unknowns).

If we have one equation with one unknown, $ax = b$, then we can just write $x = b/a$. 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.

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:

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

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

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

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

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

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

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 $B_k = A_{k1}x_1 + \cdots + A_{kn}x_n$:

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

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:

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

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):

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