1111: Linear Algebra I

Dr. Vladimir Dotsenko (Vlad)

Michaelmas Term 2015

FROM EQUATIONS TO MATRICES

For example, if we consider the system of equations

$$\begin{cases} x_1 + 2x_2 + x_3 + x_4 + x_5 = 1, \\ -3x_1 - 6x_2 - 2x_3 - x_5 = -3, \\ 2x_1 + 4x_2 + 2x_3 + x_4 + 3x_5 = -3, \end{cases}$$

then the corresponding matrix is

$$A = \begin{pmatrix} 1 & 2 & 1 & 1 & 1 & 1 \\ -3 & -6 & -2 & 0 & -1 & -3 \\ 2 & 4 & 2 & 1 & 3 & -3 \end{pmatrix}$$

(note the zero entry that indicates that x_4 is not present in the second equation).

Elementary row operations

We define *elementary row operations* on matrices to be the following moves that transform a matrix into another matrix with the same number of rows and columns:

- *Swapping rows*: literally swap the row i and the row j for some $i \neq j$, keep all other rows (except for these two) intact.
- *Re-scaling rows*: multiply all entries in the row i by a nonzero number c, keep all other rows (except for the row i) intact.
- *Combining rows*: for some $i \neq j$, add to the row i the row j multiplied by some number c, keep all other rows (except for the row i) intact.

Let us remark that elementary row operations are clearly reversible: if the matrix B is obtained from the matrix A by elementary row operations, then the matrix A can be recovered back. Indeed, each individial row operation is manifestly reversible.

From the equations viewpoint, elementary row operations are simplest transformations that do not change the set of solutions, so we may hope to use them to simplify the system enough to be easily solved.

Gauss–Jordan elimination

Gauss–Jordan elimination is a certain pre-processing of a matrix by means of elementary row operations.

- **Step 1:** Find the smallest k for which $A_{ik} \neq 0$ for at least one i, that is, the smallest k for which the kth column of the matrix A has a nonzero entry (this means x_k actually appears in at least one equation). Pick one such i, swap the first row of A with the ith one, and pass the new matrix A to Step 2.

- **Step 2:** If $m = 1$, terminate. Otherwise, take the smallest number k for which $A_{1k} \neq 0$. For each $j = 2, \ldots, m$, subtract from the jth row of A the first row multiplied by $\frac{A_{jk}}{A_{1k}}$. Divide the first row by A_{1k}. Finally, temporarily set aside the first row, and pass the matrix consisting of the last $m - 1$ rows as the new matrix A to Step 1.

Gauss–Jordan elimination

Once the procedure that we just described terminates, let us assemble together all the rows set aside along the way. The matrix A thus formed satisfies the following properties:

- For each row of A, either all entries of the row are equal to zero, or the first non-zero entry is equal to 1. (In this case we shall call that entry the *pivot* of that row).
- For each pivot of A, all entries in the same column below that pivot are equal to zero.

A matrix satisfying these two conditions is said to be in *row echelon form*.

Gauss–Jordan elimination

To complete pre-processing, for each row s of A that has nonzero entries, we do the following: for each $r < s$, subtract from the row r the row s multiplied by A_{rt}, where t is the position of the pivot in the row s.

As a result, the matrix A obtained after this is done satisfies the following properties:

- For each row of A, either all entries of the row are equal to zero, or the first non-zero entry is equal to 1. (In this case we shall call that entry the *pivot* of that row).
- For each pivot of A, all other entries in the same column are equal to zero.

A matrix satisfying these two conditions is said to be in *reduced row echelon form*. We proved an important theoretical result: every matrix can be transformed into a matrix in reduced row echelon form using elementary row operations.

Gauss–Jordan elimination: an example

Back to our example, we take the matrix we obtained and start transforming (writing down row operations to make it easy to check afterwards):

$$
\begin{pmatrix}
1 & 2 & 1 & 1 & 1 & 1 \\
-3 & -6 & -2 & 0 & -1 & -3 \\
2 & 4 & 2 & 1 & 3 & -3
\end{pmatrix}
\overset{(2)+3(1),(3)-2(1)}{\longmapsto}
$$

$$
\begin{pmatrix}
1 & 2 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 3 & 2 & 0 \\
0 & 0 & 0 & -1 & 1 & -5
\end{pmatrix}
\overset{(1)+(3),(2)+3(3),-1\times(3)}{\longmapsto}
$$

$$
\begin{pmatrix}
1 & 2 & 1 & 0 & 2 & -4 \\
0 & 0 & 1 & 0 & 5 & -15 \\
0 & 0 & 0 & 1 & -1 & 5
\end{pmatrix}
\overset{(1)-(2)}{\longmapsto}
$$

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

Gauss–Jordan elimination

Now we can use our results to solve systems of linear equations. We restore the unknowns, and look at the resulting system of equations. This system can be investigated as follows.

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.

Concluding 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 $x = b/a$, 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.

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

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

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

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

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.