MATH 304 - Linear Algebra

Last time we considered an $m \times n$ matrix $A = [a_{i,j}]$ and asked for which choices of $b_1, \dots, b_m$ the system

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + \cdots + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + \cdots + & a_{2,n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + \cdots + & a_{m,n}x_n & = & b_m \end{array}$$

is consistent. We will now look at this question from a different perspective.

First we introduce some concepts which will be very important for the rest of this course.

Definition. By $\mathbb{R}^n$ we denote the set of all n-tuples of numbers in $\mathbb{R}$:

$$\mathbb{R}^n = \{(u_1, u_2, \dots, u_n) : u_i \in \mathbb{R} \text{ for } i = 1, 2, \dots, n\}.$$

We say that the n-tuple $(u_1, u_2, \dots, u_n)$ is a solution to our system if $x_1 = u_1, x_2 = u_2, \dots, x_n = u_n$ is a solution.

We will often call elements of $\mathbb{R}^n$ **vectors**. Given any two vectors $u = (u_1, u_2, \dots, u_n)$ and $w = (w_1, w_2, \dots, w_n)$ we can add and subtract them as follows:

$$u + w = (u_1 + w_1, u_2 + w_2, \dots, u_n + w_n), \quad u - w = (u_1 - w_1, u_2 - w_2, \dots, u_n - w_n).$$

We can also multiply any vector $u = (u_1, u_2, \dots, u_n)$ by a number $c \in \mathbb{R}$ as follows

$$cu = (cu_1, cu_2, \dots, cu_n).$$

A lot of what will come in this course will be about these two operations on vectors.

We can take any vector $u = (u_1, u_2, \dots, u_n)$ in $\mathbb{R}^n$, insert $x_i = u_i$ into the left side of our equations, and compute $b_1, \dots, b_n$, which we consider as a vector $b = (b_1, \dots, b_m)$ in $\mathbb{R}^m$. Our original question can be now phrased as follows: which vectors b can be obtained form some choice of u? This looks a lot like a question about image of a function: we have a procedure which from every vector in $\mathbb{R}^n$ produces a vector in $\mathbb{R}^m$ and we ask about possible outcomes of such procedure.

In order to speak more efficiently about the procedure (a.k.a. function) just outlined, we introduce the following concept, which will be of fundamental importance to us.

Definition of dot product. Given two vectors $u = (u_1, u_2, \dots, u_n)$ and $w = (w_1, w_2, \dots, w_n)$ in $\mathbb{R}^n$ we define the dot product $u{\bullet}w$ of u and w to be the **number**

$$u{\bullet}w = u_1w_1 + u_2w_2 + \ldots + u_nw_n.$$

The dot product has the following nice properties. For any vectors u, v, w in $\mathbb{R}^n$ and any number $c \in \mathbb{R}$ we have

- $(u + v){\bullet}w = u{\bullet}w + v{\bullet}w$ and $(u - v){\bullet}w = u{\bullet}w - v{\bullet}w$
- $u{\bullet}w = w{\bullet}u$
- $(cu){\bullet}w = c(u{\bullet}w)$

To justify the first property set $u = (u_1, u_2, \dots, u_n)$, $v = (v_1, v_2, \dots, v_n)$, $w = (w_1, w_2, \dots, w_n)$ and follow the definitions:

$$(u+v){\bullet}w = (u_1+v_1)w_1+(u_2+v_2)w_2+\ldots+(u_n+v_n)w_n = u_1w_1+u_2w_2+\ldots+u_nw_n+v_1w_1+v_2w_2+\ldots+v_nw_n = u{\bullet}w+v{\bullet}w$$

The other properties are verified in a similar way.

Using the notion of the dot product, the equations in our system can be written as $r_1{\bullet}x = b_1$, $r_2{\bullet}x = b_2$,..., $r_m{\bullet}x = b_m$, where r_i denotes the i-th row of the matrix A (interpreted as an element of $\mathbb{R}^n$) and $x = (x_1, \dots, x_n)$.

We can now return to our idea of defining a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ in terms of the matrix A.

Definition. Given an $m \times n$ matrix $A = [a_{i,j}]$ we define a function $L_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ as follows

$$L_A(u) = (r_1 \bullet u, r_2 \bullet u, \ldots, r_m \bullet u),$$

where r_i denotes the i-th row of A.

Example. Consider the matrix

$$\begin{bmatrix} 3 & -1 & 2 & 3 & -1 \\ 1 & -1 & 2 & 3 & 5 \\ 2 & -3 & 6 & 9 & 4 \\ 7 & -2 & 4 & 6 & -3 \end{bmatrix}.$$

According to our definition, L_A is a function form $\mathbb{R}^5$ to $\mathbb{R}^4$ given by the formula

$$L_A(x_1, x_2, x_3, x_4, x_5) =$$

$$= (3x_1 - x_2 + 2x_3 + 3x_4 - x_5, x_1 - x_2 + 2x_3 + 3x_4 + 5x_5, 2x_1 - 3x_2 + 6x_3 + 9x_4 + 4x_5, 7x_1 - 2x_2 + 4x_3 + 6x_4 - 3x_5).$$

The system of equations we started this note with can be now written as $L_A(x) = b$ where $x = (x_1, \ldots, x_n)$ and $b = (b_1, \ldots, b_m)$. The system is consistent if and only if b is in the image (range) of the function L_A. Hence every system with coefficient matrix A is consistent if and only if the function L_A is onto. Every system with coefficient matrix A has at most one solution if and only if the function L_A is one-to-one (there is either none or exactly one x mapped by L_A onto any b). Finally, every system with coefficient matrix A has exactly one solution if and only if the function L_A is a bijection. Therefore we can restate the main result of the previous note as the following theorem.

Theorem. Let $A = [a_{i,j}]$ be an $m \times n$ matrix and let $L_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be the corresponding function. Then

- L_A is onto if and only if $\operatorname{rank}(A) = m$ (so $m \leq n$ in this case).
- L_A is one-to-one if and only if $\operatorname{rank}(A) = n$ (so $n \leq m$ in this case).
- L_A is a bijection if and only if $\operatorname{rank}(A) = m = n$.

In general, there is a homogeneous system of equations in m unknowns such that the image of L_A consists exactly of all solutions to this system.

Abusing the terminology, we will sometimes call a matrix one-to-one, onto, bijective if the function L_A has this property. It follows from the above theorem that if $n > m$ then A cannot be one-to-one, and if $n < m$ then A cannot be onto. We should think of n as a measure of size of $\mathbb{R}^n$. If the size of the domain of L_A is bigger that the size of codomain, then L_A cannot be injective; if the size of codomain is bigger that the size of the domain, then L_A cannot be surjective (note that the same is true when the domain and codomain are finite sets and size means the number of elements). Finally, as for finite sets of the same size, if $n = m$ then L_A is onto if and only if it is one-to-one.

Example. Find a system of homogeneous equations describing the image of the function L_A, where A is a 3×2 matrix

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & -2 \end{bmatrix}.$$

We follow the method outline in the previous note. We start with the matrix

$$\begin{bmatrix} 1 & -1 & b_1 \\ -1 & 1 & b_2 \\ 2 & -2 & b_3 \end{bmatrix}.$$

and perform elementary row operations $E_{2,1}(1)$, $E_{3,1}(-2)$ to get

$$\begin{bmatrix} 1 & -1 & b_1 \\ 0 & 0 & b_1 + b_2 \\ 0 & 0 & -2b_1 + b_3 \end{bmatrix}.$$

The coefficient part of this matrix (i.e. the matrix made of the first 2 columns) is now in a row-echelon form. The condition for $b = (b_1, b_2, b_3)$ to be in the image of L_A (i.e. for the corresponding system of equations to be consistent) is that no row can have its first non-zero entry in the last column. This means that we must have $b_1 + b_2 = 0$ and $-2b_1 + b_3 = 0$. Thus, the image of L_A consists of solutions $b = (b_1, b_2, b_3)$ to the system $b_1 + b_2 = 0$ and $-2b_1 + b_3 = 0$.