Matrix multiplication and matrix equations
Reading: Lay 1.4

September 4, 2013

Remember that in the last lecture we introduced vectors and linear combinations, and showed that the problem of determining whether a vector $\mathbf{b}$ could be written as a linear combination was the same as the problem of determining whether a particular linear system is consistent.

In this lecture, we introduce the notation of "matrix multiplication" as a nice way to handle problems like the ones involving linear combinations. This notation will have added benefits.

1 Multiplying a vector and a matrix

Recall that we say that a matrix is $m \times n$ if it has m rows and n columns.

Definition 1.1. Suppose we are given an $m \times n$ matrix A and a vector $\mathbf{x} \in \mathbb{R}^n$, and suppose

$$A = [\mathbf{a}_1\, \mathbf{a}_2\, \ldots \mathbf{a}_n], \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Then we define the product $A\mathbf{x}$ to be the linear combination

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n.$$

Note:

- The product $A\mathbf{x}$ is an element of $\mathbb{R}^m$;
- The product is only defined when $\mathbf{x}$ has the same number of entries as the number of columns of A.

This definition is easy to remember with practice. We will work an example, which hopefully will convince you that these calculations are not so complicated.

Example 1.2. If

$$A = \begin{bmatrix} 3 & 6 & 9 \\ 1 & 0 & 2 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 10 \end{bmatrix},$$

then

$$A\mathbf{x} = 1 \begin{bmatrix} 3 \\ 1 \\ -1 \\ 0 \end{bmatrix} - 2 \begin{bmatrix} 6 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 10 \begin{bmatrix} 9 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 81 \\ 21 \\ -1 \\ 10 \end{bmatrix}.$$

The next example will introduce the main point of this lecture.

Example 1.3. Suppose we have vectors

$$\mathbf{u} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}; \quad \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \quad \mathbf{w} = \begin{bmatrix} -2 \\ -2 \\ -2 \end{bmatrix}.$$

We want to determine whether we can write the vector

$$\mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

as a linear combination of $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$. But rather than actually solving this problem (you have seen examples of this already), we just want to express the problem in the fancy new notation we have just made. So the question is whether there exist coefficients x_1, x_2, x_3 such that

$$x_1\mathbf{u} + x_2\mathbf{v} + x_3\mathbf{w} = \mathbf{b}.$$

In the notation that we just introduced above, this is the same as asking whether there is a vector $\mathbf{x}$ in $\mathbb{R}^3$ such that

$$A\mathbf{x} = \mathbf{b},$$

where

$$A = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 0 & -2 \\ -1 & 0 & -2 \end{bmatrix}.$$

2 Matrix equations

The last example shows that problems like "what possible ways can we write a vector $\mathbf{b}$ as a linear combination of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$" can be written using matrix multiplication notation as questions of the form "what vectors $\mathbf{x}$ solve the matrix equation $[\mathbf{v}_1\,\mathbf{v}_2 \ldots \mathbf{v}_p]\mathbf{x} = \mathbf{b}$?" We will write this a little more formally in the following theorem.

Theorem 2.1 (compare Lay 1.4, Theorem 2)**.** *If A is an $m \times n$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and if $\mathbf{b} \in \mathbb{R}^m$, the matrix equation*

$$A\mathbf{x} = \mathbf{b}$$

(where the vector $\mathbf{x} \in \mathbb{R}^n$ is our variable) has the same solution set as the vector equation

$$x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n = \mathbf{b},$$

which has the same solution set as the linear system with augmented matrix

$$[\mathbf{a}_1\,\mathbf{a}_2\,\ldots\,\mathbf{a}_n\,\mathbf{b}]\,. \tag{1}$$

This theorem will be very useful. A lot of "natural" applied problems can be stated in terms of matrix equations like $A\mathbf{x} = \mathbf{b}$. The fact that we can use tools such as the RREF on these problems makes their solution much easier in many cases.

3 Applying Theorem 2.1

The definition of the product $A\mathbf{x}$ and the discussion we have had so far show that

- The matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

That is, if $A = [\mathbf{a}_1 \dots \mathbf{a}_n]$, the matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is in $\text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$. We have already seen how to determine whether some *fixed* vector $\mathbf{b}$ is in the span, using the augmented matrix (1). A harder question is to ask for a description of all possible $\mathbf{b}$ which can be written as a linear combination of $\mathbf{a}_1, \dots \mathbf{a}_n$–this is the same as asking to solve for the *entire* span of the set $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.

We are first going to ask a simpler version of this question: how can we tell when $\text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} = \mathbb{R}^m$?

Definition 3.1. We say a set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ **spans** $\mathbb{R}^m$ if every vector in $\mathbb{R}^m$ can be written as a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_p$. That is, we say $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ spans $\mathbb{R}^m$ if

$$\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\} = \mathbb{R}^m.$$

The following example is somewhat on the "easy" side because I will be running short on time in lecture. There is a more complicated one in the book (Lay 1.4, Example 3) which I encourage you to read.

Example 3.2. Let

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}.$$

Do the columns of A span $\mathbb{R}^2$? That is, is

$$\text{span}\left\{ \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix} \right\}$$

equal to $\mathbb{R}^2$?

A vector $\mathbf{b}$ is in this span if and only if it has the form

$$\mathbf{b} = c_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = (2c_1 + 3c_2) \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

That is, every vector in the span is a multiple of the single vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. From drawing pictures in $\mathbb{R}^2$, we can see that this is all vectors that lie along a particular line–not all of $\mathbb{R}^2$. So the answer is no; the columns do not span $\mathbb{R}^2$.

Theorem 3.3. *Let A be an $m \times n$ matrix. The following statements are logically equivalent. That is, A satisfies a single one of the following statements if and only if it satisfies all three, and if there is one statement which A does not satisfy, then it does not satisfy any of them.*

- *For each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution;*
- *Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A;*
- *The columns of A span $\mathbb{R}^m$;*
- *A has a pivot position in each row.*

You should think about why the first three statements are equivalent. These three are equivalent by the definitions of matrix multiplication and by the definition of the word "spans"–they are essentially different ways to say the same thing.

The fourth statement is stranger. Imagine if

$$A = [\mathbf{a}_1\, \mathbf{a}_2\, \dots \mathbf{a}_n]$$

and we want to see whether these columns span $\mathbb{R}^m$–that is, whether **every** $\mathbf{x}$ can be written as a linear combination of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$. You could determine this by writing down the RREF of the augmented matrices

$$B = [\mathbf{a}_1\, \mathbf{a}_2\, \dots \mathbf{a}_n, \mathbf{x}]$$

for every possible choice of $\mathbf{x} \in \mathbb{R}^m$, and making sure that these are the matrices of consistent linear systems. Of course this would be impossible, since there are infinitely many choice of $\mathbf{x}$.

However, if we are computing the RREF of

$$B = [\mathbf{a}_1\, \mathbf{a}_2\, \dots \mathbf{a}_n, \mathbf{x}]$$

for some $\mathbf{x}$, we know that it will be the augmented matrix of a consistent linear system if and only if the last column is not a pivot column. This will be true if every row of A has a pivot position (then we will run out of possible pivot positions before we get to the last column). The theorem says that if every row of A does not have a pivot position, then the opposite happens–there is some vector $\mathbf{x}$ such that we can force the above augmented matrix to have the last column as its pivot column, and so $\mathbf{x}$ does not lie in $\text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.

4 Row-vector rule for computing $A\mathbf{x}$

Lay has a short bit about "the easy way" to compute matrix products, which might be helpful to you. I'll talk about it a bit in class but not discuss it in these notes (Lay's description is not so bad).

5 Properties of Matrix-Vector products.

In this section, we briefly discuss two properties of the product $A\mathbf{x}$ which are sometimes useful for computing.

Theorem 5.1. *Let A be an $m \times n$ matrix, let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}_n$, and let c be a scalar. Then*

- $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$;
- $A(c\mathbf{u}) = c(A\mathbf{u})$.

The facts in this theorem are easy to prove. We prove the first one here. Let

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \quad \text{and} \quad A = [\mathbf{a}_1, \ldots, \mathbf{a}_n].$$

Then, by the definition of the product of a matrix and a vector, we have

$$A\mathbf{u} = u_1\mathbf{a}_1 + \ldots + u_n\mathbf{a}_n, \quad \text{and} \quad A\mathbf{v} = v_1\mathbf{a}_1 + \ldots + v_n\mathbf{a}_n. \tag{2}$$

On the other hand,

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{bmatrix}.$$

So, using the definition of the product $A(\mathbf{u} + \mathbf{v})$ and rearranging terms gives

$$\begin{aligned} A(\mathbf{u} + \mathbf{v}) &= (u_1 + v_1)\mathbf{a}_1 + \ldots + (u_n + v_n)\mathbf{a}_n \\ &= (u_1\mathbf{a}_1 + \ldots + u_n\mathbf{a}_n) + (v_1\mathbf{a}_1 + \ldots + v_n\mathbf{a}_n). \end{aligned} \tag{3}$$

Comparing (2) with (3) completes the proof.