A THEOREM ON COLUMN SPACES

Supplement to Section 6.6 of Messer

Math 223

Let A be an $m \times n$ matrix. The *column space* of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A.

Theorem. *Let A be an $m \times n$ matrix, and let E be the reduced row echelon form of A. Then a basis for the column space of A consists of the columns of A corresponding to the leading 1's in E (the pivot columns). Furthermore, each column of E contains the coefficients needed to write the corresponding column of A as a linear combination of this basis.*

An example will clarify this. Suppose that we want to find a basis for the subspace of $\mathbb{R}^3$ spanned by the vectors $(1, 2, 3)$, $(3, -1, -5)$, $(1, -5, -11)$, $(1, -2, 1)$, and $(34, -8, -20)$. Furthermore, we would like the basis to consist of vectors in this collection. The Contraction Theorem (the theorem that says every spanning set contains a basis) tells us we can do this. In addition, we would like to know how the remaining vectors depend on the basis vectors. The theorem above gives us a computational way to do this.

Let the matrix A consist of these vectors as columns and compute its reduced row echelon form E. We get (as you should verify!)

$$A = \begin{pmatrix} 1 & 3 & 1 & 1 & 34 \\ 2 & -1 & -5 & -2 & -8 \\ 3 & -5 & -11 & 1 & -20 \end{pmatrix} \qquad \text{and} \qquad E = \begin{pmatrix} 1 & 0 & -2 & 0 & 5 \\ 0 & 1 & 1 & 0 & 8 \\ 0 & 0 & 0 & 1 & 5 \end{pmatrix}.$$

Note that E is not simply in echelon form, it is in *reduced* echelon form—the pivots are all 1's and the other entries in the pivot columns are 0's.

According to the theorem, a basis for the column space of A consists of the vectors $(1, 2, 3)$, $(3, -1, -5)$, and $(1, -2, 1)$, corresponding to the first, second, and fourth columns of E. (Since there are three vectors in the basis, it follows that the column space of A is actually all of $\mathbb{R}^3$.) The remaining columns of A are linear combinations of these, the coefficients being given by the corresponding columns of E. For example, the third columns of A and E tell us that

$$(1, -5, -11) = -2(1, 2, 3) + 1(3, -1, -5).$$

Similarly, the last columns tell us that

$$(34, -8, -20) = 5(1, 2, 3) + 8(3, -1, -5) + 5(1, -2, 1).$$

This is just another example showing that the row reduction process generates quite a variety of information!

The proof of this theorem is fairly simple, and is based on a careful examination of the process we use to solve linear equations. Consider the homogeneous linear system

$$\begin{gathered} a_{11}x_1 + \cdots + a_{1n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = 0. \end{gathered}$$

This is equivalent, of course, to the matrix equation $AX = 0$, where

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \qquad \text{and} \qquad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

As we have seen, if we write A in block form as $A = (\mathbf{a}_1 \dots \mathbf{a}_n)$, where $\mathbf{a}_1, \dots, \mathbf{a}_n$ are the columns of A, then

$$AX = (\mathbf{a}_1 \dots \mathbf{a}_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n.$$

Thus the matrix equation $AX = 0$ becomes

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}, \tag{1}$$

which is the equation from which we make conclusions about the independence or dependence of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$, including how some of them may be linear combinations of the others.

To solve the system $AX = 0$, we apply row operations to A to reduce it to echelon form. Suppose B is a matrix obtained from A by row operations,

$$B = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix}.$$

If we use B to write a system of equations,

$$\begin{gathered} b_{11}x_1 + \cdots + b_{1n}x_n = 0 \\ \vdots \\ b_{m1}x_1 + \cdots + b_{mn}x_n = 0, \end{gathered}$$

we know that this system is equivalent to the original system. This is because the row operations simply mimic algebraic operations we could use to solve the equations. The systems are equivalent because the operations are reversible. Consequently the systems have the same solutions. Writing B in block column form, $B = (\mathbf{b}_1 \ldots \mathbf{b}_n)$, this second system can be written as

$$x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n = \mathbf{0}. \tag{2}$$

The important point in comparing (1) and (2) is that any coefficients satisfying one of them also satisfy the other. If all of the coefficients have to be 0, then both collections of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ are linearly independent. If the coefficients don't all have to be 0, then both collections of vectors are linearly dependent.

The proof of the theorem involves pushing this just a bit more. Suppose that some column of B, say $\mathbf{b}_k$, is a linear combination of the other columns. It should be clear that this is the same as saying that coefficients $x_1, \ldots, x_n$ can be found that satisfy (2) such that $x_k = -1$:

$$x_1\mathbf{b}_1 + \cdots + x_{k-1}\mathbf{b}_{k-1} - \mathbf{b}_k + x_{k+1}\mathbf{b}_{k+1} + \cdots + x_n\mathbf{b}_n = \mathbf{0}.$$

By the equivalence of (1) and (2), we have

$$x_1\mathbf{a}_1 + \cdots + x_{k-1}\mathbf{a}_{k-1} - \mathbf{a}_k + x_{k+1}\mathbf{a}_{k+1} + \cdots + x_n\mathbf{a}_n = \mathbf{0}.$$

As a consequence, $\mathbf{a}_k$ is a linear combination of the other columns of A *using the same coefficients.*

Furthermore, suppose some collection of the columns of B are linearly dependent, that is, some non-trivial linear combination of them (meaning that not all of the coefficients are 0) equals $\mathbf{0}$. Then the *same* linear combination (*i.e.*, using the same coefficients) of the corresponding columns of A equals $\mathbf{0}$, and so the corresponding columns of A are also dependent. Conversely, if some collection of the columns of B are linearly independent, so are the corresponding columns of A.

Let's apply this letting B be E, the reduced echelon form of A. You may want to refer to the example on the first page, but our reasoning will be general. Note that the pivot columns of E—suppose there are ℓ of them—are the first ℓ vectors in the standard basis for $\mathbb{R}^m$. Thus they are linearly independent. It follows that the corresponding columns of A are independent. Denote these columns of A by $\tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_\ell$. (Note that $\tilde{\mathbf{a}}_j$ is not necessarily equal to $\mathbf{a}_j$. In the example, $\tilde{\mathbf{a}}_3$ is $\mathbf{a}_4$, that is, the third pivot column is the fourth column.) Furthermore, if $\mathbf{w}_k$ is a non-pivot column of E, it's clear how it is written as a linear combination of the pivot columns. It's the same as writing $\mathbf{w}_k$ in terms of the standard basis—the entries of $\mathbf{w}_k$ are the coefficients. It follows that

$$\mathbf{a}_k = w_{1k}\tilde{\mathbf{a}}_1 + \cdots + w_{\ell k}\tilde{\mathbf{a}}_\ell,$$

where $w_{1k}, \ldots, w_{\ell k}$ are the first ℓ coefficients of $\mathbf{w}_k$. (Note that the remaining coefficients of $\mathbf{w}_k$ are 0 because the rows of E that don't contain pivots are all zeros.) This is the second claim in the statement of the theorem. Since every column of A is a linear combination of $\tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_\ell$, it follows that these vectors form a basis for the column space of A.

This completes the proof of the theorem, but as an added bonus, it follows that the reduced echelon form E of A is unique! Here's why. The theorem gives us another way to construct the matrix E. Consider the columns of A one-by-one starting with $\mathbf{a}_1$. If the first few columns of A are all zeros, let the corresponding columns of a new matrix F be zeros. For the first non-zero column of A, let the corresponding column of F be the first standard basis vector of $\mathbb{R}^m$. For each subsequent column of A, if it is a linear combination of the previous columns, let the corresponding column of F consist of the coefficients used to form the linear combination (which are unique), otherwise let that column of F be the next standard basis vector. This process uniquely determines the matrix F, but it also describes the matrix E in the theorem. Therefore $E = F$ and so E is unique.

Robert L. Foote, March 2003