Systems of Linear Equations

Math 108A: August 21, 2008
John Douglas Moore

Our goal in these notes is to explain a few facts regarding linear systems of equations not included in the first few chapters of the text [1], in the hopes of providing a better geometric picture of what the results in the text mean, and facilitate computations of examples. You are probably familiar with the dot product from calculus. We review the dot product in §1; this material is included in more general form in Chapter 6 of [1]. We advise you to skim over §1, and focus on §2 and §3, which will enable you to find bases for the subspaces that come up frequently when solving linear systems of equations. The material in §3 on elementary row operations should be a review of what you studied in Math 3C.

1 The dot product

You will recall that we discussed the dot product briefly before. If $\mathbf{x} = (x_1 \ldots, x_n)$ and $\mathbf{y} = (y_1 \ldots, y_n)$ are elements of $\mathbb{R}^n$, we define their *dot product* by

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \cdots + x_n y_n.$$

The dot product satisfies several key axioms:

1. it is symmetric: $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$;
2. it is bilinear: $(a\mathbf{x} + \mathbf{x}') \cdot \mathbf{y} = a(\mathbf{x} \cdot \mathbf{y}) + \mathbf{x}' \cdot \mathbf{y}$;
3. and it is positive-definite: $\mathbf{x} \cdot \mathbf{x} \geq 0$ and $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

The dot product is an example of an inner product on the vector space $V = \mathbb{R}^n$, as will be explained in Chapter 6 of [1].

Recall that the *length* of an element $\mathbf{x} \in \mathbb{R}^n$ is defined by

$$|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}.$$

Note that the length of an element $\mathbf{x} \in \mathbb{R}^n$ is always nonnegative.

Cauchy-Schwarz Theorem. *If* $\mathbf{x} \neq \mathbf{0}$ *and* $\mathbf{y} \neq \mathbf{0}$, *then*

$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}||\mathbf{y}|} \leq 1. \tag{1}$$

Sketch of proof (included to satisfy your curiosity—you may skip the proof if you want): If $\mathbf{v}$ is any element of $\mathbb{R}^n$, then $\mathbf{v} \cdot \mathbf{v} \geq 0$. Hence

$$(\mathbf{x}(\mathbf{y} \cdot \mathbf{y}) - \mathbf{y}(\mathbf{x} \cdot \mathbf{y})) \cdot (\mathbf{x}(\mathbf{y} \cdot \mathbf{y}) - \mathbf{y}(\mathbf{x} \cdot \mathbf{y})) \geq 0.$$

Expanding using the axioms for dot product yields

$$(\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y})^2 - 2(\mathbf{x} \cdot \mathbf{y})^2(\mathbf{y} \cdot \mathbf{y}) + (\mathbf{x} \cdot \mathbf{y})^2(\mathbf{y} \cdot \mathbf{y}) \geq 0$$

or

$$(\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y})^2 \geq (\mathbf{x} \cdot \mathbf{y})^2(\mathbf{y} \cdot \mathbf{y}).$$

Dividing by $\mathbf{y} \cdot \mathbf{y}$, we obtain

$$|\mathbf{x}|^2|\mathbf{y}|^2 \geq (\mathbf{x} \cdot \mathbf{y})^2 \quad \text{or} \quad \frac{(\mathbf{x} \cdot \mathbf{y})^2}{|\mathbf{x}|^2|\mathbf{y}|^2} \leq 1,$$

and (1) follows by taking the square root.

The key point of the Cauchy-Schwarz Inequality (1) is that it allows us to define angles between vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. It follows from properties of the cosine function that given a number $t \in [-1, 1]$, there is a unique angle θ such that

$$\theta \in [0, \pi] \qquad \text{and} \qquad \cos\theta = t.$$

Thus we can define the angle between two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ by requiring that

$$\theta \in [0, \pi] \qquad \text{and} \qquad \cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}||\mathbf{y}|}.$$

Then the dot product satisfies the formula

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos\theta.$$

In particular, we can say that two vectors vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are *perpendicular* or *orthogonal* if $\mathbf{x} \cdot \mathbf{y} = 0$. This provides much intuition for dealing with vectors in $\mathbb{R}^n$.

Thus if $\mathbf{a} = (a_1, \dots a_n)$ is a nonzero element of $\mathbb{R}^n$, the homogeneous linear equation

$$a_1x_1 + \cdots + a_nx_n = 0$$

describes the set of all vectors $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ that are perpendicular to $\mathbf{a}$. The set of solutions

$$W = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} = 0\}$$

to this homogeneous linear equation is a subspace of $\mathbf{R}^n$. We remind you that to see this we need to check three facts:

1. $\mathbf{a} \cdot \mathbf{0} = 0$, so $\mathbf{0} \in W$.

2. If $\mathbf{x} \in W$ and $\mathbf{y} \in W$, then $\mathbf{a} \cdot \mathbf{x} = 0$ and $\mathbf{a} \cdot \mathbf{y} = 0$, and it follows from the axioms for dot product that $\mathbf{a} \cdot (\mathbf{x} + \mathbf{y}) = 0$ so $\mathbf{x} + \mathbf{y} \in W$.

3. If $c \in \mathbb{R}$ and $\mathbf{x} \in W$, then $\mathbf{a} \cdot \mathbf{x} = 0$, and it follows from the axioms for dot product that $\mathbf{a} \cdot (c\mathbf{x}) = 0$, so $c\mathbf{x} \in W$.

One can also show that W is a subspace of $\mathbb{R}^n$ by showing that it is the null space of a linear map, as described in the following section. (Recall that the null space of a linear map is always a subspace by a theorem in [1].)

2 Linear systems and orthogonal complements

Linear algebra is the theory behind solving systems of linear equations, such as

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2, \\ \cdots\cdots & \cdot & \cdot \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m. \end{array} \tag{2}$$

Here the a_{ij}'s and b_i's are known elements of the field $\mathbb{F}$, and we are solving for the unknowns $x_1, \ldots, x_n$. This system of linear equations can be written in terms of the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

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

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ b_m \end{pmatrix}.$$

The matrix A defines a linear map

$$T_A : \mathbb{F}^n \to \mathbb{F}^m \quad \text{by} \quad T_A(\mathbf{x}) = A\mathbf{x}.$$

The range of T_A is just the space of vectors $\mathbf{b}$ for which the equation $A\mathbf{x} = \mathbf{b}$ has a solution, while the null space of T_A is the space of solutions to the associated *homogeneous* linear system

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0, \\ \cdots\cdots & \cdot & \cdot \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & 0. \end{array} \tag{3}$$

The fundamental Theorem 3.4 from the text [1] states that

$$n = \dim(\mathbb{F}^n) = \dim(\text{null}(T_A)) + \dim(\text{range}(T_A)). \tag{4}$$

In particular, it follows that

$$\dim(\text{null}(T_A)) = 0 \quad \Leftrightarrow \quad \dim(\text{range}(T_A)) = n. \tag{5}$$

Recall that in the special case where $n = m$, either (5) or Theorem 3.21 from [1] implies that T_A is surjective if and only if it is injective. Equivalently, the system (2) has a solution for every choice of vector $\mathbf{b} \in \mathbb{R}^n$ if and only if the associated homogeneous system (3) has only the zero solution. This is the content of problem 3-26 in [1].

In the special case where $\mathbb{F} = \mathbb{R}$, the notion of orthogonal complement gives a nice picture of the process of finding a basis for the space of solutions to a homogeneous linear system. Indeed, we can use the dot product to write the homogeneous linear system whose solutions form the null space of T_A in a particularly intuitive form. Let

$$\begin{array}{rcl} \mathbf{a}_1 & = & (a_{11}, a_{12}, \ldots, a_{1n}), \\ \mathbf{a}_2 & = & (a_{21}, a_{12}, \ldots, a_{2n}), \\ \cdot & \cdot & \cdots \\ \mathbf{a}_m & = & (a_{m1}, a_{m2}, \ldots, a_{mn}), \end{array}$$

and let $W = \text{span}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$. We can then say that

$$\text{null}(T_A) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}_1 \cdot \mathbf{x} = 0, \mathbf{a}_2 \cdot \mathbf{x} = 0, \cdots, \mathbf{a}_m \cdot \mathbf{x} = 0\}.$$

Note that the last space on the right-hand side is simply the collection of all vectors which are perpendicular to W. We call this space the *orthogonal complement* to W and denote it by $W^\perp$. We can then write

$$\text{null}(T_A) = W^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} = 0 \text{ for } \mathbf{a} \in W\ \}.$$

Orthogonal Complement Theorem. *If $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$ is a list of vectors in $\mathbb{R}^n$, $W = \text{span}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$ and $W^\perp$ is the space of solutions $\mathbf{x} \in \mathbb{R}^n$ to the homogeneous linear system of equations*

$$\mathbf{a}_1 \cdot \mathbf{x} = 0, \quad \mathbf{a}_2 \cdot \mathbf{x} = 0, \quad \cdots, \quad \mathbf{a}_m \cdot \mathbf{x} = 0,$$

then

1. *$\dim W + \dim W^\perp = n$, and*
2. *$\mathbb{R}^n = W \oplus W^\perp$.*

Proof: By Theorem 2.10 in [1], the list $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$ can be reduced to a basis $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k)$ for W, where $k = \dim W$. Let B be the $k \times n$ matrix whose rows are the elements of the list $(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k)$ and let $T_B : \mathbb{R}^n \to \mathbb{R}^k$ be the corresponding linear map. Then we can write $W^\perp = \text{null}(T_B)$, and it follows from (4) that

$$\begin{aligned} n = \dim(\mathbb{R}^n) = \dim(\text{null}(T_B)) + \dim(\text{range}(T_B)) &\leq \dim W^\perp + k \\ &= \dim W^\perp + \dim W. \end{aligned} \tag{6}$$

On the other hand, if $\mathbf{a} \in W \cap W^\perp$, then $\mathbf{a} \cdot \mathbf{a} = 0$ and hence $\mathbf{a} = \mathbf{0}$. Thus $W \cap W^\perp = \{\mathbf{0}\}$ and by Theorem 2.18,

$$\dim W + \dim W^\perp = \dim(W + W^\perp) \leq \dim \mathbb{R}^n = n. \tag{7}$$

Statement 1 follows from (6) and (7). Statement 1 together with Proposition 2.19 of [1] then implies that $\mathbb{R}^n = W \oplus W^\perp$, finishing the proof of the Theorem.

Corollary. *The dimension of the space W spanned by the rows of A equals the dimension of the space spanned by the columns of A.*

Proof: Comparison of (4) with the equation $n = \dim W^\perp + \dim W$ shows that $\dim W = \dim(\text{range}(T_A))$. But W is the space spanned by the rows of A, while $\text{range}(T_A)$ is the space spanned by the columns of A.

One sometimes expresses this corollary in the following form: Row rank equals column rank.

3 Elementary row operations

To find the *general solution* to the homogeneous linear system (3) means to find a basis for the space of solutions

$$\text{null}(T_A) = \{\mathbf{x} \in \mathbb{F}^n : A\mathbf{x} = \mathbf{0}\}.$$

Given such a basis $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, an arbitrary solution to the linear system can be written in a unique way as

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k.$$

We can think of $(c_1, \ldots, c_k)$ as coordinates in the solution space which correspond to the basis $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.

As described in the previous section, if $\mathbb{F} = \mathbb{R}$, $\text{null}(T_A)$ is just the orthogonal complement $W^\perp$ to the subspace W of $\mathbb{R}^n$ spanned by the rows of A. For general choice of field $\mathbb{F}$, we now describe a useful procedure, based upon the elementary row operations for finding a basis for the space W spanned by the rows of A and for the null space $\text{null}(T_A)$ at the same time.

Proposition. *Let V be a vector space over a field $\mathbb{F}$, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ a collection of elements of V. Then*

1. *If $\sigma : \{1, 2, \ldots, m\} \to \{1, 2, \ldots, m\}$ is one-to-one and onto. Then*

$$Span(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m) = Span(\mathbf{v}_{\sigma(1)}, \mathbf{v}_{\sigma(2)}, \ldots, \mathbf{v}_{\sigma(m)}).$$

2. *If $a \in F$ and $a \neq 0$, then*

$$Span(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m) = Span(a\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m).$$

3. *If* $a \in F$, *then*
$$Span(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m) = Span(\mathbf{v}_1 + a\mathbf{v}_2, \mathbf{v}_2, \ldots, \mathbf{v}_m).$$

With the techniques we have studied so far in the course, you should be able to provide a proof of this Proposition.

We can change an $m \times n$ matrix A to a new matrix A' by one of the *elementary row operations*:

1. Interchange two rows in A.
2. Multiply a row by a nonzero constant c.
3. Add a constant multiple of one row to another.

Each of these operations is reversible. Although they change the linear map T_A, the above Proposition, shows that the elementary row operations preserve the linear subspace W of $\mathbb{F}^n$ which is spanned by the rows of A. Moreover, since the elementary row operations transform the homogeneous linear system appearing in (3) into a homogeneous linear system with the same space of solutions, the elementary row operations preserve the null space of T_A, which is of course, just the orthogonal complement to W when $\mathbb{F} = \mathbb{R}$.

The idea is to perform elementary row operations to obtain a matrix B which is sufficiently simple that we can read off bases for W and the null space of T_A with ease.

A commonly used methodical way of doing this leads to replacing A by a matrix B in row-reduced echelon form. By definition, the matrix B is in row-reduced echelon form if it has the following properties:

1. The first nonzero entry in any row is a one.
2. If a column contains an initial one for some row all of the other entries in that column are zero.
3. If a row consists of all zeros, then it is below all of the other rows.
4. The initial one in a lower occurs to the right of the initial one in each previous row.

It can be proven that any $m \times n$ matrix can be put in row-reduced echelon form by elementary row operations. It is easy to see how to carry out the procedure. One starts by putting a one in the first row of the first nonzero column. We do this by interchanging rows if necessary to get a nonzero entry in the first row of the first nonzero column, and then divide the row by this entry. We then zero out all other elements in the first nonzero column. In the submatrix obtained by removing the first row, we then apply the same procedure obtaining an initial one in a second column and zeroing out all other entries in that column. Continuing with this procedure leads to a matrix in row-reduced echelon form.

Properly reformulated, the procedure we have described would give a proof that any matrix can be put in row-reduced echelon form by elementary row operations. We will not carry out all of the details, but you may be able to see how they would go.

It is usually easy to work this out in special cases, and from the resulting row-reduced echelon matrix we easily construct a basis for the space W spanned by the rows of A and a basis for the null space of T_A.

Indeed, one can show that the nonzero rows of the row-reduced echelon matrix form a basis for W. indeed, suppose there are k nonzero rows $\mathbf{b}_1, \ldots, \mathbf{b}_k$ in the row reduced echelon matrix B. Suppose there is a collection of elements $x_1, \ldots x_k$ from $\mathbb{F}$, such that

$$x_1\mathbf{b}_1 + \cdots + x_k\mathbf{b}_k = \mathbf{0}.$$

Each component of this equation corresponds to one of the columns of B and the equation corresponding to the j-th initial one is just the equation $x_j = 0$. Thus

$$x_1\mathbf{b}_1 + \cdots + x_k\mathbf{b}_k = \mathbf{0} \quad \Rightarrow \quad x_1 = \cdots = x_k = 0,$$

and the row vectors $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are linearly independent. Since they obviously span W, they form a basis for W.

If there are k nonzero rows and hence W has dimension k, one can solve for the k variables corresponding to the initial ones in those rows in terms of the $n - k$ variables corresponding to the other rows. The $n - k$ variables which do NOT correspond to initial ones are *free variables*, and can be thought of as coordinates for the space $W^\perp$. This gives a general solution to the original linear system, in which the $n - k$ free variables form the coordinates. One can check that the list of vectors multiplying the free variables are linearly independent and since they span null(T_A), they form a basis for null(T_A).

Finding a basis for the range of T_A is a quite different matter. In order to do this, you need to use the elementary **column** operations on the original matrix A, since the range of T_A is spanned by the columns of A. Note that the columns of A are **not** preserved by the elementary row operations.

References

[1] S. Axler, *Linear algebra done right*, Second edition, Springer, New York, 1997.