The Fundamental Theorem of Linear Algebra

Nicholas Hoell

Contents

These notes are going to present a striking result, fundamental to understanding how the concepts we've seen all semester relate to one another. But, equally important, we're going to see a remarkable picture which illustrates the theorem.

1 Prelude: Orthogonal Complements

Before we get to the main theorem, a bit on the terminology it uses. We'll begin with a natural definition.

Definition 1

Let $S \subseteq \mathbb{R}^n$ be a subspace. The set $S^\perp \subseteq \mathbb{R}^n$ is defined to be the collection of all vectors $\mathbf{v} \in \mathbb{R}^n$ orthogonal to S and called the **orthogonal complement of** S in $\mathbb{R}^n$. In other words,

$$S^\perp = \{\mathbf{u} \in \mathbb{R}^n \mid \mathbf{u} \cdot \mathbf{v} = 0, \mathbf{v} \in S\}$$

You should be able to prove that $S^\perp$ is a subspace of $\mathbb{R}^n$ (see exercises). The purpose of these notes is to establish a version of the **Fundamental Theorem of Linear Algebra**. The result can be thought of as a type of *representation* theorem, namely, it tells us something about *how vectors are* by describing the canonical subspaces of a matrix A in which they live. To understand this we consider the following representation theorem.

Theorem 1

Let $\mathbf{v} \in \mathbb{R}^n$ and let $S \subseteq \mathbb{R}^n$ be a subspace. Then there are vectors $\mathbf{s} \in S$, $\mathbf{s}_\perp \in S^\perp$ such that

$$\mathbf{v} = \mathbf{s} + \mathbf{s}_\perp$$

In other words, vectors can be expressed in terms of pieces living in orthogonal spaces.

The above theorem is sometimes expressed in the notation $\mathbb{R}^n = S \oplus S^\perp$ where the notation is called the "**direct sum**" of S and $S^\perp$.

Proof

The proof needs a small fact which you will see if you take MAT224. Here's the fact: not only does every subspace S of $\mathbb{R}^n$ have a basis, it actually has an *orthonormal* basis, namely a basis $\mathcal{B}$ consisting of vectors $\mathbf{s}_j$ where $\|\mathbf{s}_j\| = 1$ holds for all j and $\mathbf{s}_i \cdot \mathbf{s}_j = 0$ when $i \neq j$. The hard part of this statement actually is showing that subspaces have bases (which we've done). From there the idea is to use an algorithm to turn a given basis into one in which all the vectors are made orthogonal and of unit length. But it's a technical detail I don't want to describe here, I only want to use the result. So let's agree S has an orthonormal basis as described. Set $\mathbf{s} = (\mathbf{v} \cdot \mathbf{s}_1)\mathbf{s}_1 + \cdots + (\mathbf{v} \cdot \mathbf{s}_{\dim S})\mathbf{s}_{\dim S}$ which must clearly be an element of S. Then define $\mathbf{s}_\perp = \mathbf{v} - \mathbf{s}$. It's obvious that $\mathbf{v} = \mathbf{s} + \mathbf{s}_\perp$ and you can verify that $\mathbf{s}_\perp \in S^\perp$ giving the result. □

Where this is going: We'll soon see that for an $m \times n$ matrix A we have $(\ker A)^\perp = col(A^T)$. If this is true, then that means (using the notation above) that $\mathbb{R}^n = \ker(A) \oplus col(A^T)$. In this case a vector $\mathbf{x} \in \mathbb{R}^n$ can be written as $\mathbf{x} = \mathbf{p} + \mathbf{v}$ where $A\mathbf{v} = \mathbf{0}$ and $\mathbf{p} \in row(A)$ since $row(A) = col(A^T)$. In other words if $\mathbf{b} \in col(A)$ we can solve $A\mathbf{x} = \mathbf{b}$ and since $\mathbf{x} \in \mathbb{R}^n = \ker(A) \oplus col(A^T)$ we can write $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{p} + \mathbf{v}_h$$

In the above $\mathbf{p} \in row(A)$ since $row(A) = col(A^T)$. The space of possible $\mathbf{v}_h$'s is $\dim(ker(A))$-dimensional, so the nullity describes the lack of unique solvability for the linear system $A\mathbf{x} = \mathbf{b}$.

2 The Fundamental Theorem of Linear Algebra

We can now get on with proving the main theorem in this course, the capstone to our understanding what it means to solve systems of linear equations.

Proposition 1

Suppose A is an $m \times n$ matrix. Then

$$col(A^T)^\perp = \ker A$$

Proof

Suppose $\mathbf{v} \in \ker A$, then $A\mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} (row_1(A)) \cdot \mathbf{v} \\ (row_2(A)) \cdot \mathbf{v} \\ \vdots \\ (row_n(A)) \cdot \mathbf{v} \end{bmatrix}$ so $\mathbf{v}$ is orthogonal to $row_i(A)$ for $i = 1, ..., m$. Therefore $\mathbf{v} \in (row(A))^\perp = (col(A^T))^\perp$, i.e. $\ker A \subseteq (col(A^T))^\perp$. As well, if $\mathbf{v} \in (col(A^T))^\perp$ then, in particular, we must have that $\mathbf{v} \cdot (row_i(A)) = 0$ for $i = 1, ..., m$. But in this case we'd again have $\mathbf{v} \in \ker A$. Thus $(col(A^T))^\perp \subseteq \ker A$. Since the sets contain each other they must be equal and we're done.

□

If $\mathbf{v} \in (S^\perp)^\perp$ then clearly $\mathbf{v} \cdot \mathbf{s}_\perp = 0$ for all $\mathbf{s}_\perp \in S^\perp$. If $\{\mathbf{s}_1, ..., \mathbf{s}_{\dim S}\}$ is an orthonormal basis for S and $\{\mathbf{s}_1^\perp, ..., \mathbf{s}_{\dim S^\perp}^\perp\}$ is an orthonormal basis for $S^\perp$ we can write out

$$\begin{aligned} \mathbf{v} &= (\mathbf{v} \cdot \mathbf{s}_1)\mathbf{s}_1 + \cdots + (\mathbf{v} \cdot \mathbf{s}_{\dim S})\mathbf{s}_{\dim S} + (\mathbf{v} \cdot \mathbf{s}_1^\perp)\mathbf{s}_1^\perp + \cdots + (\mathbf{v} \cdot \mathbf{s}_{\dim S^\perp}^\perp)\mathbf{s}_{\dim S^\perp}^\perp \\ &= (\mathbf{v} \cdot \mathbf{s}_1)\mathbf{s}_1 + \cdots + (\mathbf{v} \cdots \mathbf{s}_{\dim S})\mathbf{s}_{\dim S} \in S \end{aligned}$$

So $(S^\perp)^\perp \subseteq S$. As well, clearly for $\mathbf{s} \in S$ we have $\mathbf{s} \in (S^\perp)^\perp$ since $\mathbf{s} \cdot \mathbf{s}_\perp = 0$ must be true for all $\mathbf{s}_\perp \in S^\perp$. Thus, $S \subseteq (S^\perp)^\perp$. From these facts it follows that

$$S = (S^\perp)^\perp$$

must hold for all subspaces $S \subseteq \mathbb{R}^n$. (Make sure you can verify this statement). Applying this general fact to the previous theorem we immediately get the following.

Theorem 2: Fundamental Theorem of Linear Algebra

Suppose A is an $m \times n$ matrix. Then

$$col(A^T) = (\ker A)^\perp$$

So that

$$\mathbb{R}^n = \ker(A) \oplus col(A^T)$$

gives an orthogonal decomposition of $\mathbb{R}^n$ into the null space and the row space of matrix A. Therefore, for $\mathbf{b} \in col(A)$ we have that $A\mathbf{x} = \mathbf{b}$ is solved by

$$\mathbf{x} = \mathbf{p} + \mathbf{v}_h$$

for $\mathbf{p} \in row(A)$ a particular solution, $A\mathbf{p} = \mathbf{b}$, and $\mathbf{v}_h \in \ker A$ a generic vector in $\ker A$.

2.1 The Diagram

The content of this theorem, the fundamental theorem of linear algebra, is encapsulated in the following figure. If a picture is worth a thousand words, this figure is worth at least several hours' thought.

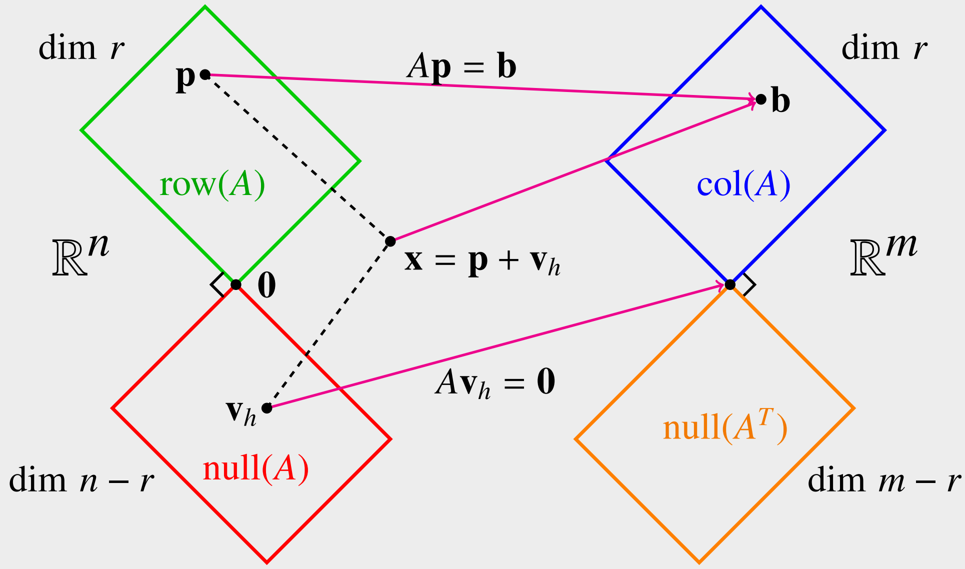

Figure 1: Solving $A\mathbf{x} = \mathbf{b}$ for an $m \times n$ matrix A with $rank(A) = r$.

Notice this picture has all the major players we've seen in this course: there are vectors, subspaces, dimensions, and they are all playing a role in the way we conceptualize solving a linear system of equations. Moreover, note the duality in the picture: the subspaces on the right-hand side, those living in $\mathbb{R}^m$, are the corresponding subspaces appearing on the left-hand side for the *transpose* of A rather than A (e.g. $col(A)$ appearing on the upper right is just $row(A^T)$ whereas we see $row(A)$ on the upper left-hand side). The subspaces meet at right angles because they're orthogonal complements of each other and the intersection of the spaces is the zero subspace of the respective ambient space ($\mathbb{R}^n$ or $\mathbb{R}^m$). Try to visualize how the picture changes as we look at limiting values of possible rank for A. Namely, how would this above change in the case where $rank(A) = n$ say? Or m? Or if A is square and invertible? If A is symmetric? Etc.

3 Exercises

You should be able to do the following.

1. Calculate the orthogonal complements of the improper subspaces of $\mathbb{R}^n$.
2. Let $\mathbf{v} \in \mathbb{R}^3$ be a nonzero vector. Describe $(\text{span}\{\mathbf{v}\})^\perp$.
3. Prove that $S \cap S^\perp = \{\mathbf{0}\}$ holds for any subspace $S \subseteq \mathbb{R}^n$.
4. Prove that $S^\perp$ is a subspace of $\mathbb{R}^n$ whenever S is.
5. Prove that for $S \subseteq \mathbb{R}^n$ a subspace we have $\dim S + \dim S^\perp = n$.
6. If S, W are subspaces of $\mathbb{R}^n$ show that $S \subseteq W \implies W^\perp \subseteq S^\perp$
7. Let $S = \{\mathbf{x} \in \mathbb{R}^4 \mid x_1 + x_2 + x_3 + x_4 = 0\}$. Prove that S is a subspace and find a basis for $S^\perp$.

8. Prove that every $m \times n$ matrix A defines a linear transformation from $row(A)$ **onto** $col(A)$, i.e.

$$T_A : row(A) \to col(A)$$

9. Consider A, an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Consider the following claim[1]: One and only one of the following systems is solvable

   (a) $A\mathbf{x} = \mathbf{b}$

   (b) $A^T\mathbf{y} = \mathbf{0}$ with $\mathbf{y}^T\mathbf{b} \neq \mathbf{0}$

   Prove that the above options *cannot BOTH hold*. In other words, if one of the above holds the other one mustn't.

[1] A variation of this strange-sounding claim is very important in numerous applications including, very importantly, in differential equations.