The Fundamental Theorem of Linear Algebra

Dale Kim
Insomniac Games, Inc.
dkim@insomniacgames.com
@Roflraging

January 27, 2017

1 Overview

For many, linear algebra is simply a set of rules for manipulating a matrices, but this view conceals the beauty and insight that can be gained from a deeper understanding. A key part of developing this understanding is to focus on the vector spaces involved instead of the mechanical operations we use to compute useful results. The fundamental theorem of linear algebra is (wait for it...) fundamental. It's so fundamental, they had to say it was fundamental! This theorem describes very important results regarding the relationships between the four matrix subspaces:

- Row space $\boldsymbol{C(A^T)}$
- Nullspace $\boldsymbol{N(A)}$
- Column space $\boldsymbol{C(A)}$
- Left nullspace $\boldsymbol{N(A^T)}$

In this article, I will go into describing the theorem, what it means for each subspace, and how this information could be used to build more intuition about linear systems you may encounter. It is assumed that the reader has a working knowledge of the basic matrix operations and familiarity with the terminology and notation in linear algebra (such as orthogonality, linear independence, combinations, etc).

A few notes: the organization of this article may be poor or hard to follow. Many diagrams were planned but ultimately had to be cut due to time constraints. I may update this to contain them at some point in the future. Additionally, I am not a mathematician by trade or training, so it may contain errors!

2 The Theorem

Given an $\boldsymbol{m}$ by $\boldsymbol{n}$ matrix $\boldsymbol{A}$, the fundamental theorem of linear algebra states four key results:

1. $\dim \boldsymbol{C(A)} = \dim \boldsymbol{C(A^T)} = \boldsymbol{rank(A)}$
2. $\dim \boldsymbol{N(A)} = \boldsymbol{n} - \boldsymbol{rank(A)}$
3. $\dim \boldsymbol{N(A^T)} = \boldsymbol{m} - \boldsymbol{rank(A)}$
4. $\boldsymbol{C(A^T)}^{\perp} = \boldsymbol{N(A)}$ and $\boldsymbol{C(A)}^{\perp} = \boldsymbol{N(A^T)}$

2.1 Column Space and Row Space

The first result states that the column space $\boldsymbol{C(A)}$ and row space $\boldsymbol{C(A^T)}$ have the same number of dimensions, which is the $\boldsymbol{rank}$ of $\boldsymbol{A}$. It may be surprising at first that these two spaces are the same size, but recall that the $\boldsymbol{rank}$ of a matrix is the number of pivots, which can be found through elimination. The process of elimination reveals which columns and rows of a matrix are independent and which ones are not, thus revealing the true size of the matrix. Consider the following 3 by 3 matrix $\boldsymbol{A}$ and its reduced row echelon form $\boldsymbol{R}$:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

$$\boldsymbol{R} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{2}$$

If you were tasked to describe $\boldsymbol{C(A)}$, you may have difficulty describing it just by looking at it, but $\boldsymbol{R}$ is much clearer and tells you a lot of information about $\boldsymbol{A}$. Elimination has produced two pivots in columns 1 and 3 of $\boldsymbol{R}$ and it is these columns (in $\boldsymbol{A}$) which produce the basis vectors for $\boldsymbol{C(A)}$.

Notice how the vectors are different, but with only two independent vectors in 3D, we cannot possibly fill the entirety of $\mathbb{R}^3$ and are restricted to a plane (2D). The row space, is similar, but the non-zero rows of $\boldsymbol{R}$ (not $\boldsymbol{A}$) are the basis vectors!

As with the column space, the row space is a plane. Going back to the theorem, it stated that the dimensions of $\boldsymbol{C(A)}$ and $\boldsymbol{C(A^T)}$ are the same and they equal $\boldsymbol{rank(A)}$, which our exploration thus far agrees with...

2.2 Nullspace and left nullspace

$$\boldsymbol{Az} = \boldsymbol{0}$$

Above is the definition of the nullspace $\boldsymbol{N(A)}$, which consists of all vectors $\boldsymbol{z}$ that annihilate the column space to the zero vector. Result 2 states that the number of dimensions in $\boldsymbol{N(A)}$ is equal to the number of columns $\boldsymbol{n}$ - $\boldsymbol{rank(A)}$. Going back to $\boldsymbol{R}$ in equation (2), we can easily find the vectors $\boldsymbol{z}$ that create the nullspace:

$$\boldsymbol{Rz} = \boldsymbol{0}$$

$$\begin{gathered} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ \begin{aligned} z_1 + 2z_2 &= 0 \\ z_3 &= 0 \end{aligned} \\ \boldsymbol{z} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \end{gathered} \tag{3}$$

Equation (3) shows a special solution for the nullspace. All vectors along this line satisfy the definition of the nullspace for the matrix $\boldsymbol{A}$. This a basis vector for $\boldsymbol{N(A)}$, but are there more? No, there cannot be any more because of the theorem! Taking $\boldsymbol{n}$ - $\boldsymbol{rank(A)}$ gives us $3 - 2 = 1$. $\boldsymbol{N(A)}$ is one dimensional (a line) in $\mathbb{R}^3$ and we found the special solution by solving $\boldsymbol{Rx} = \mathbf{0}$ with the lone free variable $\boldsymbol{x_2} = 1$.

The process for finding the left nullspace $\boldsymbol{N(A^T)}$ is identical to $\boldsymbol{N(A)}$, just transpose $\boldsymbol{A}$ before you begin!

2.3 Orthogonality

Here comes one of the really meaty parts of linear algebra: orthogonality. This concept is central to linear algebra and highlights an astonishing amount about the relationships between the four subspaces. We begin with a definition of orthogonality for two subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$:

$$\boldsymbol{v^T w} = 0 \text{ for all } \boldsymbol{v} \text{ in } \boldsymbol{V} \text{ and for all } \boldsymbol{w} \text{ in } \boldsymbol{W}$$

The last result claims that the row space $\boldsymbol{C(A^T)}$ is the ***orthogonal complement*** of the nullspace $\boldsymbol{N(A)}$. Every vector in $\boldsymbol{C(A^T)}$ is orthogonal to every vector in $\boldsymbol{N(A)}$! Going back to equations (2) and (3), we have the two basis vectors $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ for the row space and one basis vector $\boldsymbol{z_1}$ for the nullspace:

$$\boldsymbol{r_1} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \qquad \boldsymbol{r_2} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \boldsymbol{z_1} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

Is $\boldsymbol{C(A^T)} \perp \boldsymbol{N(A)}$ in our example? We can check this easily by constructing arbitrary vectors in the row space and nullspace and using the definition of orthogonal subspaces:

$$\begin{aligned} \boldsymbol{r} &= a\boldsymbol{r_1} + b\boldsymbol{r_2} & \boldsymbol{z} &= c\boldsymbol{z_1} \\ \boldsymbol{r} &= a\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + b\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} & \boldsymbol{z} &= c\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \\ \boldsymbol{r} &= \begin{bmatrix} a \\ 2a \\ b \end{bmatrix} & \boldsymbol{z} &= \begin{bmatrix} -2c \\ c \\ 0 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \boldsymbol{r^T z} &= \begin{bmatrix} a & 2a & b \end{bmatrix} \begin{bmatrix} -2c \\ c \\ 0 \end{bmatrix} \\ \boldsymbol{r^T z} &= -2ac + 2ac + 0 \end{aligned}$$

$$\boldsymbol{r^T z} = 0$$

3 Big Picture

By this point, you may be saying *"So what? I still don't see what's so special about these spaces and the fact that some of them are orthogonal!"* It's time for what Gilbert Strang calls, **"The big picture"**[1]:

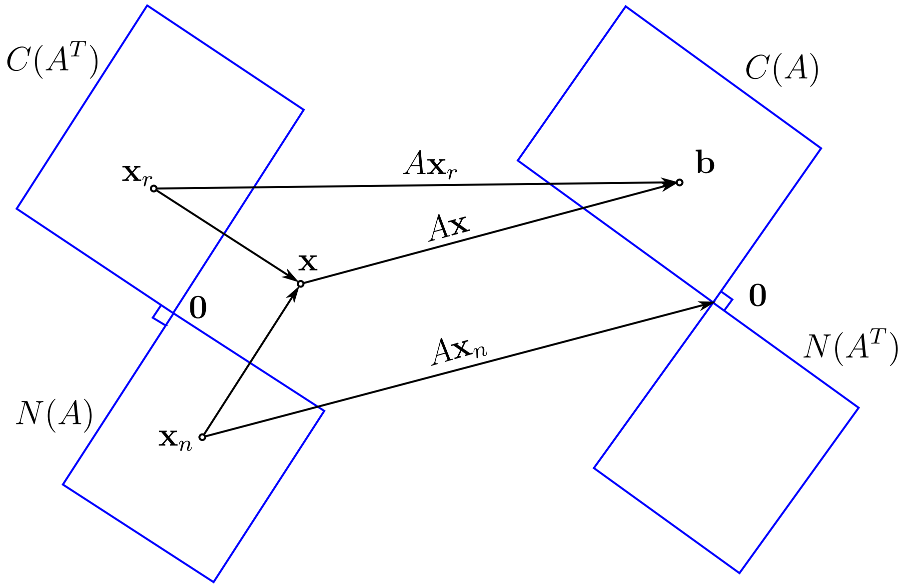

Figure 1: Four fundamental subspaces
[2]

3.1 A Closer Look at $Ax = b$

To fully appreciate one of the core problems of linear algebra, $\boldsymbol{Ax = b}$, it is helpful to keep in mind another perspective, namely the idea of functions. In Figure 1, there are arrows starting on the left in $\mathbb{R}^n$ and end in $\mathbb{R}^m$. Think of $\mathbb{R}^n$ as the ***domain***, where the input $\boldsymbol{x}$ must come from to participate in the matrix multiplication $\boldsymbol{Ax}$ and $\mathbb{R}^m$ as the ***codomain*** (take note of $\boldsymbol{Ax}$ in the figure, we will get to it later!). For now, focus on the innocuous looking line $\boldsymbol{Ax_r = b}$. It says something about the relationship between the row space and column space of $\boldsymbol{A}$. Can you guess what it is?

A natural question to ask when looking at this figure is if multiple vectors from the row space $\boldsymbol{x_r}$ can map to the same vector in $\boldsymbol{C(A)}$. No! Suppose:

$$\begin{aligned}\boldsymbol{Ax_r} &= \boldsymbol{Ax'_r}\\ \boldsymbol{Ax_r - Ax'_r} &= \boldsymbol{0}\\ \boldsymbol{A(x_r - x'_r)} &= \boldsymbol{0}\\ \boldsymbol{(x_r - x'_r)} &\in \boldsymbol{N(A)}\end{aligned}$$

We've concluded that $\boldsymbol{x_r - x'_r}$ is in the nullspace. $\boldsymbol{x_r}$ and $\boldsymbol{x'_r}$ were originally part of the row space, which is a subspace. Subspaces are closed under addition/subtraction. So $\boldsymbol{x_r - x'_r}$ is also in the row space. The only vector to be in both row space and nullspace is the zero vector!

$$\begin{aligned}\boldsymbol{x_r - x'_r} &= \boldsymbol{0}\\ \boldsymbol{x_r} &= \boldsymbol{x'_r}\end{aligned}$$

$\square$

Let's consider now the line $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{x} = \boldsymbol{x_r} + \boldsymbol{x_n}$. This may seem contradictory in light of the proof above, but it is not. The proof considered two vectors in the row space, but in this case we have only one vector in the row space. So then how is it possible that this can still map to $\boldsymbol{b} \in \boldsymbol{C(A)}$?

$$\begin{aligned} \boldsymbol{Ax} &= \boldsymbol{b} \\ \boldsymbol{A(x_r + x_n)} &= \boldsymbol{b} \\ \boldsymbol{Ax_r} + \boldsymbol{Ax_n} &= \boldsymbol{b} \\ \boldsymbol{Ax_r} + \boldsymbol{0} &= \boldsymbol{b} \\ \boldsymbol{Ax_r} &= \boldsymbol{b} \end{aligned}$$

□

$\boldsymbol{x_n}$ is in the nullspace and by the linearity of matrix multiplication, we can separate $\boldsymbol{Ax}$ into two separate multiplications. This result may not seem all that significant, but it explains why the possibility of having infinitely many solutions in a linear system exists! Consider a full rank $\boldsymbol{n}$ by $\boldsymbol{n}$ system, which has n pivots. By the fundamental theorem of linear algebra, the nullspace consists of only the zero vector and the row space spans all of $\mathbb{R}^n$. Using the previous two proofs, we know that for any $\boldsymbol{b} \in \boldsymbol{C(A)}$ it is uniquely mapped to by a vector $\boldsymbol{x_r} \in \boldsymbol{C(A^T)}$. Therefore, the only possible way for there to be multiple solutions is for the nullspace to be involved. But as we had just established, the nullspace has only one element in it, the zero vector:

$$\begin{aligned} \boldsymbol{Ax_r} + \boldsymbol{Ax_n} &= \boldsymbol{b} \\ \boldsymbol{Ax_r} + \boldsymbol{0A} &= \boldsymbol{b} \\ \boldsymbol{Ax_r} &= \boldsymbol{b} \end{aligned}$$

There simply aren't any free variables/vectors to create more solutions.

Now consider a rank deficient $\boldsymbol{m}$ by $\boldsymbol{n}$ system ($\boldsymbol{rank(A)} < \boldsymbol{min(m, n)}$). Using the same argument as before, any $\boldsymbol{b} \in \boldsymbol{C(A)}$ must be uniquely mapped by $\boldsymbol{x_r} \in \boldsymbol{C(A^T)}$. But this time, the nullspace has a dimension of 1 or greater, thus an infinite number of them. In $\boldsymbol{Ax_r} + \boldsymbol{Ax_n} = \boldsymbol{b}$, $\boldsymbol{x_r}$ is fixed but now we can choose an infinite number of nullspace vectors $\boldsymbol{x_n}$!

4 To Be Continued...

References

[1] Gilbert Strang. *Introduction to Linear Algebra, 4th Edition.* Welleslely - Cambridge Press, Wellesley, Massachusetts, 2009.

[2] Alexey Grigorev. http://www.itshared.org/2015/06/the-fundamental-theorem-of-linear.html

[3] http://mathworld.wolfram.com/FundamentalTheoremofLinearAlgebra.html

[4] https://en.wikipedia.org/wiki/Fundamental_theorem_of_linear_algebra