VECTOR SPACES

PAUL SCHRIMPF
SEPTEMBER 23, 2013
UNIVERSITY OF BRITISH COLUMBIA
ECONOMICS 526

The last lecture introduced vector spaces. In this lecture we will explore vector spaces in more detail, which will eventually let us complete our characterization of the set of solutions to systems of linear equations.

Remember that our the vector space with which we are most interested is Euclidean space, $\mathbb{R}^n$. In fact, a good way to think about other vector spaces is that they are just variations of $\mathbb{R}^n$. The whole reason for defining and studying abstract vector spaces is to take our intuitive understanding of two and three dimensional Euclidean space and apply it to other contexts. If you find the discussion of abstract vector spaces and their variations to be confusing, you can ignore it and think of two or three dimensional Euclidean space instead. For the exams in this course, I am unlikely to ask about vector spaces other than $\mathbb{R}^n$. It is likely that you will not need to know anything about vector spaces other than $\mathbb{R}^n$ throughout the masters program. However, if you read enough articles in economics journals, you will come across abstract vector spaces, and hopefully what we have covered in this course will be helpful. Also, if you plan to continue and get a PhD it will be useful to know about abstract vector spaces.

1. NORMED VECTOR SPACES

One property of two and three dimensional Euclidean space is that vectors have lengths. Our definition of vector spaces from last lecture does not guarantee that we have a way of measuring length in all vector spaces, so let's define a special type of vector space where we can measure length.

Definition 1.1. A **normed vector space**, $(V, \mathbb{F}, +, \cdot, \|\cdot\|)$, is a vector space with a function, called the **norm**, from V to $\mathbb{F}$ and denoted by $\|v\|$ with the following properties:

(1) $\|v\| \geq 0$ and $\|v\| = 0$ iff $v = 0$,
(2) $\|\alpha v\| = |\alpha| \, \|v\|$ for all $\alpha \in \mathbb{F}$,
(3) The triangle inequality holds:
$$\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$$
for all $v_1, v_2 \in V$.

As in the previous lecture, when the field, addition, multiplication, and norm are clear from context, we will just write V instead of $(V, \mathbb{F}, +, \cdot, \|\cdot\|)$ to denote a normed vector space. Like length, a norm is always non-negative and only zero for the zero vector. Also, similar to length, if we multiply a vector by a scalar, the norm also gets multiplied by the scalar. The triangular inequality means that norm obeys the idea that the shortest distance between two points is a straight line. If you go directly from x to y you "travel" $\|x-y\|$. If you stop at point z in between, you travel $\|x-z\|+\|z-y\|$. The triangle inequality guarantees that

$$\|x-y\| \leq \|x-z\| + \|z-y\| .$$

1.1. **Examples.**

Example 1.1. $\mathbb{R}^3$ is a normed vector space with norm

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

This norm is exactly how we usually measure distance. For this reason, it is called the Euclidean norm.

More generally, for any n, $\mathbb{R}^n$, is a normed vector space with norm

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

The Euclidean norm is the most natural way of measuring distance in $\mathbb{R}^n$, but it is not the only one. A vector space can often be given more than one norm, as the following example shows.

Example 1.2. $\mathbb{R}^n$ with the norm

$$\|x\|_p = \left(\sum_{i=1}^{p} |x_i|^p\right)^{1/p}$$

for $p \in [1, \infty]$[1] is a normed vector space. This norm is called the p-norm.

For nearly all practical purposes, $\mathbb{R}^n$ with any p-norm is essentially the same as $\mathbb{R}^n$ with any other p-norm. $\mathbb{R}^n$ is the same collection of elements regardless of the choice of p-norm, and the choice of p-norm does not affect the topology of $\mathbb{R}^n$ or the definition of derivatives.[2] However, there are normed vector spaces where the choice of norm makes a difference.

Example 1.3. $\mathcal{L}^p(0,1)$ with p-norm

$$\|f\|_p = \left(\int_0^1 |f(x)|^p dx\right)^{1/p}$$

is a normed vector space. Moreover, $\mathcal{L}^p(0,1)$ is a different space for different p. For example, $\frac{1}{x^{1/p}} \notin \mathcal{L}^p(0,1)$, but $\frac{1}{x^{1/p}} \in \mathcal{L}^q(0,1)$ for $q < p$.

[1] Where $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$

[2] We will discuss topology next lecture, and derivatives soon afterward, so do not worry if you do not know what that means.

2. Inner product spaces

Another example of a normed vector space is any inner product space. Recall the definition of an inner product space from last lecture.

Definition 2.1. A real **inner product space** is a vector space over the field $\mathbb{R}$ with an additional operation called the inner product that is function from $V \times V$ to $\mathbb{R}$. We denote the inner product of $v_1, v_2 \in V$ by $\langle v_1, v_2 \rangle$. It has the following properties:

(1) Symmetry: $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$
(2) Linear: $\langle av_1 + bv_2, v_3 \rangle = a \langle v_1, v_3 \rangle + b \langle v_2, v_3 \rangle$ for $a, b \in \mathbb{R}$
(3) Positive definite: $\langle v, v \rangle \geq 0$ and equals 0 iff $v = 0$.

Any inner product space is also a normed vector space with norm

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Recall from the previous lecture that the inner product on $\mathbb{R}^n$ is the dot product,

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

The norm induced by the inner product is then

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2},$$

which is the usual Euclidean norm. Henceforth, whenever talking about inner product spaces we will use $\|x\|$ to denote the norm induced by the inner product (which is the same as the 2-norm or Euclidean norm).

Inner product spaces are special in another way. Remember that we are studying vector spaces and their variants to try to generalize our understand of three dimensional Euclidean space to other contexts. Vector spaces are places where we can add elements and multiply by scalars, just like in 3-d Euclidean space. In normed vector spaces, we can also measure distance. Another thing that we know how to do in Euclidean space is measure angles. Inner product spaces are vector spaces where we can also measure angles.

Suppose we have an inner product space then:

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle \\ &= \langle x, x \rangle + 2 \langle x, y \rangle + \langle y, y \rangle \end{aligned}$$

In $\mathbb{R}^n$ with the Euclidean norm when x and y are at right angles to one other, $\langle x, y \rangle = 0$, and we have the Pythagorean theorem:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

This motivates the following definition

Definition 2.2. Let $x, y \in V$, an inner product space. x and y are **orthogonal** iff $\langle x, y \rangle = 0$.

In Euclidean space, the inner product and angle are related by the following theorem,

Theorem 2.1. *Let $u, v \in \mathbb{R}^n$, then the angle between them is*

$$\theta = \cos^{-1} \frac{\langle u, v \rangle}{\|u\| \, \|v\|}.$$

Proof. We can prove this using the definition of the cosine and the Pythagorean theorem. See Blume and Simon theorem 10.3 for pictures. Imagine u and v in $\mathbb{R}^2$. Form a right-angle triangle by drawing a line orthogonal to the line from the origin to v and passing through u. Let tv, with $t \in \mathbb{R}$, be the point at the right-angle of the triangle. By definition, cosine is the ratio of the length of the adjacent side to the hypotenuse, which is

$$\cos\theta = \frac{\|tv\|}{\|u\|}.$$

Now we just need to relate $\|tv\|$ to $\langle u, v \rangle$. The opposite side of the triangle is length $\|u - tv\|$, so by the Pythagorean theorem,

$$\begin{aligned}
\|u\|^2 &= \|tv\|^2 + \|u - tv\|^2 \\
\|u\|^2 &= t^2 \|v\|^2 + \|u\|^2 - 2t \langle u, v \rangle + t^2 \|v\|^2 \\
2t \langle u, v \rangle &= 2t^2 \|v\|^2 \\
t &= \frac{\langle u, v \rangle}{\|v\|^2}
\end{aligned}$$

Plugging this result into the previous equation gives the conclusion. □

In inner product spaces other than $\mathbb{R}^n$, we could define angles to fit theorem 2.1. For example, this would allow us to talk about the angle between two functions in $\mathcal{L}^2(0,1)$. We will not use this definition of angle very much though. The really important thing to remember about inner product spaces and angles is that vectors can be orthogonal and the Pythagorean theorem holds.

2.1. **Useful inequalities.** When we start looking at limits next week, will we often need to prove that the norm of something is small. There are a number of inequalities that we will repeatedly use. The most common is the triangle inequality, which was part of our definition of norms. The triangle inequality has many implications, some of which are not obvious. These implications are often useful in proofs. The most common is what is known as the reverse triangle inequality.

Theorem 2.2 (Reverse triangle inequality)**.** *Let V be a normed vector space and $x, y \in V$. Then*

$$|\|x\| - \|y\|| \leq \|x - y\|.$$

Proof. By the usual triangle inequality,

$$\begin{aligned}
\|x\| + \|x - y\| &\geq \|y\| \\
\|x - y\| &\geq \|y\| - \|x\|
\end{aligned}$$

and

$$\|y\| + \|y - x\| \geq \|x\|$$
$$\|y - x\| \geq \|x\| - \|y\| .$$

Also, $\|y - x\| = \|(-1)(x - y)\| = |-1| \, \|x - y\| = \|x - y\|$ is greater than both $\|x\| - \|y\|$ and $\|y\| - \|x\|$ and

$$\|x - y\| \geq |\|x\| - \|y\|| .$$

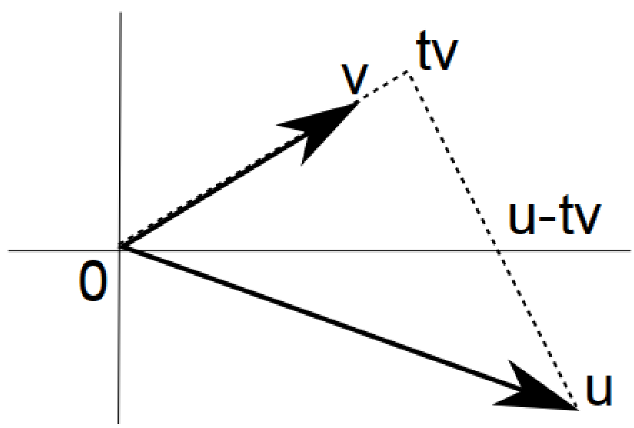
□

After the triangle inequality, arguably the most important inequality in mathematics is the Cauchy-Schwarz inequality.

Theorem 2.3 (Cauchy-Schwarz inequality)**.** *Let V be an inner product space and let $u, v \in V$. Then,*

$$|\langle u, v\rangle| \leq \|u\| \, \|v\| .$$

Proof. The idea of this proof can be illustrated in $\mathbb{R}^2$ by forming a right angle triangle with vertices at 0, u, and tv and sides of lengths $\|tv\|$, $\|u - tv\|$, and $\|v\|$, where t is chosen such that v and $u - tv$ are orthogonal, as shown in the following diagram.



We are choosing t so that $\langle v, u - tv\rangle = 0$, so solving for t,

$$\begin{aligned} \langle v, u - tv\rangle &= 0 \\ \langle v, v\rangle &= t \langle u, v\rangle \\ t &= \frac{\langle u, v\rangle}{\|v\|^2}. \end{aligned}$$

We can choose such a t in any inner product space, not just $\mathbb{R}^2$.

Now, let $z = u - tv$. By construction, $\langle z, v\rangle = 0$ and $u = tv + z$. Hence, by the Pythagorean theorem,

$$\begin{aligned} \|u\|^2 &= \|tv\|^2 + \|z\|^2 \\ &= \frac{\langle u, v\rangle^2}{\|v\|^2} + \|z\|^2 \end{aligned}$$

$\|z\|^2 \geq 0$, so

$$\|u\|^2 \geq \frac{\langle u, v \rangle^2}{\|v\|^2}$$
$$\|u\| \, \|v\| \geq | \langle u, v \rangle |.$$

□

Notice that in the proof, we also saw that $\|u\| \, \|v\| = | \langle u, v \rangle |$ if and only if $\|z\| = 0$. $\|z\|$ is zero whenever u and v are linearly dependent i.e. $u = \alpha v$ where $\alpha \in \mathbb{R}$.

3. Projections

The mapping from u to tv that we saw in the proofs of theorem 2.1 and 2.3 is so common that it has a name.

Definition 3.1. Let V be an inner product space and $x, y \in V$. The (orthogonal) **projection** of y onto x is

$$P_x y = \frac{\langle y, x \rangle}{\|x\|^2} x.$$

More generally, the projection of y onto a finite set $\{x_1, x_2, ..., x_k\}$ is

$$P_{\{x_j\}_{j=1}^k} y = \sum_{j=1}^{k} P_{x_j - P_{\{x_i\}_{i \neq j}} x_j} y.$$

Equivalently,

$$P_{\{x_j\}_{j=1}^k} y = X(X^T X)^{-1} X^T y$$

where X as an n by k matrix whose columns are x_k.

More generally still, if $X \subseteq V$ is a linear subspace, then the projection of y onto X is

$$P_X y = P_{\{b_j\}_{j=1}^k} y$$

where $b_1, ..., b_k$ are an orthogonal basis for X.

Finally, if $Y \subseteq V$ the projection of Y onto X is just the set consisting of the projection of each element of y onto X, i.e.

$$P_X Y = \{P_X y : y \in Y\}.$$

In $\mathbb{R}^2$ the projection of y onto x can be visualized by drawing a line that passes through y and is perpendicular to the line connecting x and the origin. In general, the projection of y onto a subspace X will be the point in x that is closest to y. This point will lie on a line that is orthogonal to X and passes through y. Projections are related to linear transformations and matrices as well.

Lemma 3.1. *Any projection is an idempotent linear transformation.*

Proof. First, we verify that projections have the two properties required for them to be linear transformations.

(1)

$$\begin{aligned} P_x(y_1+y_2) =& \frac{\langle x, y_1+y_2 \rangle}{\|x\|^2} x \\ =& \frac{\langle x, y_1 \rangle}{\|x\|^2} x + \frac{\langle x, y_2 \rangle}{\|x\|^2} x \\ =& P_x y_1 + P_x y_2 \end{aligned}$$

(2)

$$\begin{aligned} P_x(\alpha y) =& \frac{\langle x, \alpha y \rangle}{\|x\|^2} x \\ =& \alpha \frac{\langle x, y \rangle}{\|x\|^2} x \\ =& \alpha P_x y. \end{aligned}$$

Now, we show that projections are idempotent.

$$\begin{aligned} P_x(P_x y) =& P_x \left(\frac{\langle x, y \rangle}{\|x\|^2} x \right) \\ =& \frac{\left\langle x, \frac{\langle x, y \rangle}{\|x\|^2} x \right\rangle}{\|x\|^2} x \\ =& \frac{\langle x, y \rangle}{\|x\|^2} \frac{\langle x, x \rangle}{\|x\|^2} x \\ =& P_x y. \end{aligned}$$

□

It turns out that any symmetric idempotent linear transformation can be written in the form used to define (orthogonal) projections. Therefore, projections are sometimes defined as idempotent linear transformations instead. This definition is equivalent to ours. We will not prove this, but we will have the tools to prove it after this lecture so you might want to do it as an exercise.

4. Linear independence

Recall the definition of linear independence from last lecture.

Definition 4.1. A set of vectors $v_1, ..., v_k \in V$, is **linearly independent** if the only solution to

$$\sum_{j=1}^{k} c_j v_j = 0$$

is $c_1 = c_2 = ... = c_k = 0$.

4.1. **Checking linear independence.** Given a set of vectors, $\mathbf{v}_1, ..., \mathbf{v}_n \in \mathbb{R}^m$ (or any n-dimensional vector space), how do check whether they are linearly independent? Well, by definition, they are linearly independent if $c_1 = c_2 = ... = c_n = 0$ is the only solution to

$$\sum_{j=1}^{n} c_j \mathbf{v}_j = 0$$

If we write this condition as a system of linear equations we have

$$\begin{aligned} v_{11}c_1 + v_{12}c_2 + ... + v_{1n}c_n &= 0 \\ \vdots &= \vdots \\ v_{m1}c_1 + v_{m2}c_2 + ... + v_{mn}c_n &= 0 \end{aligned}$$

or in matrix form,

$$\begin{aligned} \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} =& 0 \\ V\mathbf{c} =& 0 \end{aligned}$$

We call any system with 0 on the right hand side a **homogeneous** system. Any homogeneous system always has $\mathbf{c} = 0$ as a solution. We know from lecture 2 that it will have other solutions if the rank of V is less than n. This proves the following lemma.

Lemma 4.1. *Vectors $\mathbf{v}_1, ..., \mathbf{v}_n \in \mathbb{R}^m$ are linearly independent if and only if*

$$\text{rank}(\mathbf{v}_1, ..., \mathbf{v}_n) = n.$$

Corollary 4.1. *Any set of $k > m$ vectors in $\mathbb{R}^m$ are linearly dependent.*

For sets of m vectors in $\mathbb{R}^m$ we can restate the lemma in terms of the determinant.

Corollary 4.2. *Vectors $\mathbf{v}_1, ..., \mathbf{v}_m \in \mathbb{R}^m$ are linearly independent if and only if*

$$\det(\mathbf{v}_1, ..., \mathbf{v}_n) \neq 0.$$

5. Row, column, and null space

5.1. **Row space.**

Definition 5.1. Let A be an m by n matrix. The **row space** of A, denoted $\text{Row}(A)$, is the space spanned by the row vectors of A.

The row space of A is a subspace of $\mathbb{R}^n$.

Lemma 5.1. *Performing Gaussian elimination does not change the row space of a matrix.*

Proof. Let $a_1, ..., a_m$ be the row vectors of A. Each step of Gaussian elimination transforms some a_j into $a_j + ga_k$ with $k \neq j$ or $g \neq -1$. It suffices to show that

$$\text{span}(a_1, ..., a_m) = \text{span}(a_1 + ga_k, ..., a_m).$$

If $x \in \text{span}(a_1, ..., a_m)$, then

$$\begin{aligned} x =& \sum_{i=}^{m} c_i a_i \\ =& c_1(a_1 + g a_k) + (\sum_{i=2}^{m} c_i a_i) - c_1 g a_k, \end{aligned}$$

so $x \in \text{span}(a_1 + g a_k, ..., a_m)$.

Conversely if $x \in \text{span}(a_1 + g a_k, ..., a_m)$, then

$$\begin{aligned} x =& c_1(a_1 + g a_k) + \sum_{i=2}^{m} c_i a_i \\ =& c_1 g a_k + \sum_{i=1}^{m} c_i a_i \end{aligned}$$

so $x \in \text{span}(a_1, ..., a_m)$. ☐

Corollary 5.1. *The dimension of the row space of a matrix is equal to its rank.*

Proof. Let $r_1, ..., r_m$ be the row vectors of the row echelon form of A. If the rank of A is $k \leq m$, then $r_{k+1} = 0, ..., r_m = 0$. Also, for $j <= k$, r_j has more leading zeros than r_{j-1}. A quick inductive argument then shows that $r_1, ..., r_k$ are linearly independent. For $k = 1$, linear independence follows from $r_k \neq 0$. For $k > 1$, assume $r_2, ..., r_k$ are linearly independent. We cannot write $r_1 = (r_{11}, r_{12}, ..., r_{1n})$ as a combination of $r_2, ..., r_k$ be they all begin with 0 and $r_{11} \neq 0$. Therefore $r_1, ..., r_k$ are linearly independent, and $\dim\text{Row}(A) = \text{rank}A$. ☐

5.2. **Column space.**

Definition 5.2. Let A be an m by n matrix. The **column space** of A, denoted $\text{Col}(A)$, is the space spanned by the column vectors of A.

We know that the column space of an n by n matrix is $\mathbb{R}^n$ if and only if the matrix is nonsingular. We will see that the column space plays an important role in describing the set of solutions to systems of linear equations.

Lemma 5.2. *Let A be an m by n matrix. Then $Ax = b$ has a solution iff $b \in \text{Col}(A)$.*

Proof. If x solves $Ax = b$, then b is in the column space of A by definition. Conversely if b is in the column space of A, then there must exist an x such that $Ax = b$. ☐

Let us examine the dimension of the column space of A.

Definition 5.3. A column of a matrix A is **basic** if the corresponding column of the row echelon form, A_r, contains a pivot.

Theorem 5.1. *The basic columns of A form a basis for $\text{Col}(A)$.*

Proof. Let A be $m \times n$ and denote its columns as $v_1, ..., v_n$. Let A_r be the row echelon form of A and denotes its columns as $w_1, ..., w_n$. Let $w_{i_1}, ..., w_{i_k}$ be the basic columns of A_r. Then w_{i_1} must end with more zeros than w_{i_2}, and w_{i_2} must end with more zeros than w_{i_3}, etc. By the same argument as in corollary 5.1, $w_{i_1}, ..., w_{i_k}$ are linearly independent. By definition of row echelon form, the final $m-k$ rows of A_r are all zero. Therefore $\dim\text{Col}(A_r) \leq k$, and $w_{i_1}, ..., w_{i_k}$ must span $\text{Col}(A_r)$ and be a basis.

Now we must show that $v_{i_1}, ..., v_{i_k}$ are a basis for $\text{Col}(A)$. Suppose

$$c_1 v_{i_1} + ... + c_k v_{i_k} = 0.$$

Then we could do Gaussian elimination to convert this system to

$$c_1 w_{i_1} + ... + c_k w_{i_k} = 0.$$

$w_{i_1}, ..., w_{i_k}$ are linearly independent so $c_1 = 0, ...c_k = 0$ and $v_{i_1}, ..., v_{i_k}$ are linearly independent too. Moreover, if we add any other v_j, $j \notin \{i_1, ..., i_k\}$, then by the same argument there must exist a non-zero c than solves

$$c_1 v_{i_1} + ... + c_k v_{i_k} + c_j v_j = 0.$$

Thus, $v_{i_1}, ..., v_{i_k}$ is a basis for $\text{Col}(A)$. □

Corollary 5.2. *The dimensions of the row and column spaces of any matrix are equal.*

Proof. The previous proof showed that $\dim\text{Col}(A) = \text{rank}A$, and earlier we showed that $\dim\text{Row}(A) = \text{rank}A$. □

This also proves that the ranks of A and A transpose are equal. We stated this fact earlier without proof.

Corollary 5.3. $\text{rank}A = \text{rank}A^T$.

When talking about linear transformations in general instead of just matrices, the column space is called the image or range of the transformation. All the results in this section still apply.

5.3. **Null space.**

Definition 5.4. Let A be m by n. The set of solutions to the homogeneous equation $Ax = 0$ is the **null space** (or kernel) of A, denoted by $\mathcal{N}(A)$ (or $\text{Null}(A)$).

Definition 5.5. Let $V \subseteq \mathbb{R}^n$ be a linear subspace, and let $c \in \mathbb{R}^n$ be a fixed vector. The set

$$\{x \in \mathbb{R}^n : x = v + c \text{ for some } v \in V\}$$

is called the set of **translates** of V by c, and is denoted $c + V$. Any set of translates of a linear subspace is called an **affine** space.

Like linear subspaces, affine spaces are points, lines, planes, and hyperplanes. Unlike linear subspaces, affine spaces do not necessarily contain 0.

Lemma 5.3. *Let $Ax = b$ be an m by n system of linear equations. Let x_0 be any particular solution. Then the set of solutions is $x_0 + \mathcal{N}(A)$.*

Proof. Let $w \in x_0 + \mathcal{N}(A)$. Then

$$\begin{aligned} Aw =& A(x_0) + A(\underbrace{w - x_0}_{\in \mathcal{N}(A)}) \\ =& b + 0. \end{aligned}$$

Let w be a solution to $Ax = b$. Then

$$A(w - x_0) = Aw - Ax_0 = 0$$

so $w - x_0 \in \mathcal{N}(A)$ and $w \in x_0 + \mathcal{N}(A)$. □

Theorem 5.2. *Let A be an m by n matrix. Then* $\dim\mathcal{N}(A) = n - \operatorname{rank}A$

Proof. Let $u_1, ..., u_k$ be a basis for $\mathcal{N}(A)$. If $k = n$, then $\dim\mathcal{N}(A) = n$ and $Ax = 0$ for all $x \in \mathbb{R}^n$. It follows that A must be all zeros so $\operatorname{rank}A = 0$ and the theorem is true.

If $k < n$, then we add $u_{k+1}, ..., u_n$ to $u_1, ..., u_k$ to form a basis for $\mathbb{R}^n$. We can do this because $u_1, ..., u_k$ do not span $\mathbb{R}^n$, so there must be $u_{k+1} \notin \operatorname{span}(u_1, ..., u_k)$. This implies that $u_1, ..., u_k, u_{k+1}$ are linearly independent. Repeating this argument we can add $u_{k+2}, ..., u_n$.

Now we will show that $Au_{k+1}, ..., Au_n$ are a basis for the column space. They are in the column space by definition. They are linearly independent because

$$c_{k+1}Au_{k+1} + ... + c_nAu_n = A(c_{k+1}u_{k+1} + ... + c_nu_n)$$

$A(c_{k+1}u_{k+1} + ... + c_nu_n)$ equals zero only if $(c_{k+1}u_{k+1} + ... + c_nu_n) \in \mathcal{N}(A)$, but by construction this is only possible for $c_{k+1} = ... = c_n = 0$.

Finally $Au_{k+1}, ..., Au_n$ span $\operatorname{Col}A$ since if $Ax = b$ we can write $x = c_1u_1 + ... + c_na_n$, and

$$\begin{aligned} b = Ax =& A\left(c_1u_1 + ... + c_na_n\right) \\ =& c_1A \underbrace{u_1}_{\in\mathcal{N}(A)} + ... + c_kA \underbrace{u_k}_{\in\mathcal{N}(A)} + c_{k+1}Au_{k+1} + ... + c_nAu_n \\ b =& 0 + c_{k+1}Au_{k+1} + ... + c_nAu_n \end{aligned}$$

so $b \in \operatorname{span}(Au_{k+1}, ..., Au_n)$. Thus,

$$\dim\operatorname{Col}(A) = n - k = n - \dim\mathcal{N}(A) = \operatorname{rank}A.$$

□

Collecting all the above results, we have finally completed our description of the set of all solutions to a system of linear equations.

Theorem 5.3 (Rouché-Capelli)**.** *A system of linear equations with n variables has a solution if and only if the rank of its coefficient matrix, A, is equal to the rank of its augmented matrix, $\hat{A}$. Equivalently, a solution exists if and only if $b \in \operatorname{Col}(A)$.*

If a solution exists and $\operatorname{rank}A$ *is equal to its number of columns, the solution is unique. If a solution exists and* $\operatorname{rank}A$ *is less than its number of columns, there are infinite solutions. In this case the set of solutions forms is an affine space, $x_0 + \mathcal{N}(A)$, where x_0 is any particular solution to $Ax = b$. This set of solutions is an affine subspace of dimension $n - \operatorname{rank}A$.*

5.4. **Relationship among row, column, and null space.** Let's examine how the row, column, and null spaces of a matrix are related. Suppose A is m by n, then $\operatorname{Col}(A) \subseteq \mathbb{R}^m$, $\operatorname{Row}(A) \subseteq \mathbb{R}^n$ and $\mathcal{N}(A) \subseteq \mathbb{R}^n$. Given that the row and null space are both subsets of $\mathbb{R}^n$, it is natural to wonder how they are related. Since the transpose just switches rows with columns, we know that $\text{Row}(A^T) = \text{Col}(A)$ and $\text{Col}(A^T) = \text{Row}(A)$. Suppose $x \in \mathcal{N}(A)$, then $Ax = 0$. The definition of the transpose requires that for all $w \in \mathbb{R}^m$,

$$\begin{aligned}\left\langle A^T w, x \right\rangle &= \langle w, Ax \rangle \\ \left\langle A^T w, x \right\rangle &= \langle w, 0 \rangle \\ \left\langle A^T w, x \right\rangle &= 0\end{aligned}$$

But, the set $\{A^T w : w \in \mathbb{R}^m\}$ is $\text{Col}(A^T) = \text{Row}(A)$. Thus, we can conclude that for any $x \in \mathcal{N}(A)$ and $y \in \text{Row}(A)$, $\langle y, x \rangle = 0$. In other words, the row and null spaces of a matrix are orthogonal. Similarly, $\mathcal{N}(A^T)$ and $\text{Row}(A^T) = \text{Col}(A)$ are orthogonal.