HW # 06 Solutions: Supplement

1. What does a subspace represent?

   Math Answer: We know that lines, circles, triangles, and ellipses are fundamental objects in math and physics, with a sphere being a generalization of a circle and a pyramid being one generalization of a triangle. A subspace is one way to generalize the notion of a line that passes through the origin. A subspace can have many dimensions, whereas a line just has one.

   A second way to think of a subspace of $\mathbb{R}^n$ is that it is a subset of $\mathbb{R}^n$ that is also a vector space. It's perhaps easier now for us to think of $\mathbb{R}^1$ and $\mathbb{R}^2$ as vector spaces, but what about $\mathbb{R}^1$ and $\mathbb{R}^2$ as subsets of $\mathbb{R}^3$ or $\mathbb{R}^5$? What's the difference between $\mathbb{R}^2$ and

$$V := \{(x_1, \dots, x_5) \in \mathbb{R}^5 \mid x_3 = 0, x_4 = 0, x_5 = 0\} = \text{null}\left(\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}\right) = \text{range}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\right)?$$

   Nothing really? V in each of its representations is a two-dimensional subset of $\mathbb{R}^5$ that is closed under linear combinations, and hence we can do Linear Algebra in V!

   What's the difference between $\mathbb{R}^2$ and

$$V := \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_3 = x_1 + x_2\} = \text{null}(\begin{bmatrix} 1.0 & 1.0 & -1.0 \end{bmatrix}) = \text{range}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}\right)?$$

   All the term $x_3 = x_1 + x_2$ does is "tilt" the "x_1 x_2-plane" in $\mathbb{R}^3$. It still passes through the origin. It's like if you rotated the x-axis to the y-axis in $\mathbb{R}^2$, or stopped half way at a 45^o angle, it's still a line, nothing complicated. Similarly, rotating the standard plane in $\mathbb{R}^3$ so that vectors have a non-zero x_3-component is not complicated. It's still a plane passing through the origin.

   Bottom line, subsets where we can do Linear Algebra are called subspaces!

   Engineering Answer: As you advance in Engineering (or other Applied STEM field), you will find that some problems are linear while some are nonlinear. Linear problems can be treated with Linear Algebra, which as Math goes, is easier to understand and use than many other areas of Math, but just as importantly, we have much better algorithms to solve problems using Linear Algebra than we do in many other areas of Math. Hence, in Engineering and the Applied Sciences, if we have a **subset** that contains all the solutions to a problem that we are about and it is a **subspace**, we go **BINGO! Got lucky here!**

   Stack Exchange: For example, in order to speak about dimension you need a subspace. If you have a system of linear equations, the solution set is a translated subspace. Eigenspaces are subspaces. Spans are subspaces. Orthogonal complements are subspaces. If you have a linear transformation between vector spaces, the image of a subspace is a subspace, and the preimage of a subspace is also a subspace – in particular, the image of the entire domain is a subspace, and the preimage of $\{0\}$ is a subspace. This alone can teach you a lot about the structure of sets in a vector space where concrete calculations may be difficult to carry out or even imagine intuitively. https://math.stackexchange.com/questions/1735390/what-is-the-point-of-subspaces/1735418

2. Once you show that a set (subset) passes the 0-vector test, how to prove that it is a subspace?

   Answer:

   (a) Form a linear combination of two general vectors in the set and show that the resulting vector is once again in the set.

   (b) Equivalently, you can do the above as two separate steps

   - Show the set is closed under vector sums.

- Show the set is closed under scalar times vector multiplication.

(c) Show the set is the null space of a matrix. Then it is automatically a subspace.

(d) Show the set can be written as the span of some set of vectors, then it is automatically a subspace.

3. What is a null space of a matrix and why is it important?

Answer: If A is an $n \times m$ matrix, then null$(A) := \{x \in \mathbb{R}^m \mid Ax = 0\}$, the set of all vectors in $\mathbb{R}^m$ that A maps to the zero vector in $\mathbb{R}^n$. That seems pretty boring! Well, not really! Let's consider $Ax = b$ instead and suppose that x_0 is a solution, that is $Ax_0 = b$. We can then ask if there are other solutions, and if so, what are they?

Fact: Suppose that x_0 satisfies $Ax_0 = b$, so it is a solution to $Ax = b$; it can be any solution. Then the **set of all solutions** to $Ax = b$ is given by

$$x_0 + \bar{x}, \text{ where } \bar{x} \in \text{null}(A).$$

We get some nice things from this formula:

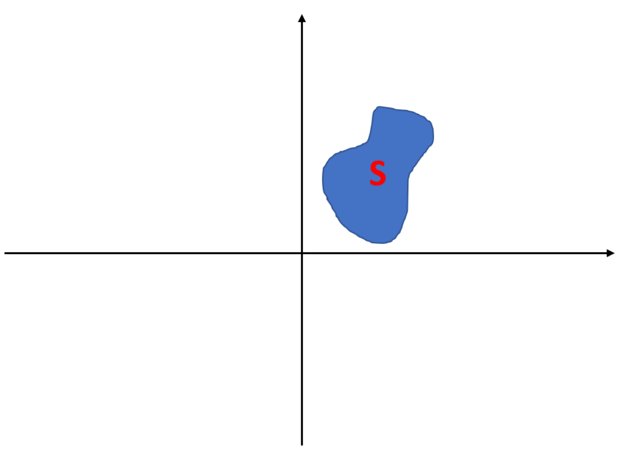
(a) x_0 is the unique solution to $Ax = b$ if, and only if, null$(A) = \{0_{m\times 1}\}$

(b) When null(A) contains more than just the zero vector, $\{0_{m\times 1}\}$, we can compute all other solutions once we find just one solution, which is kind of amazing.

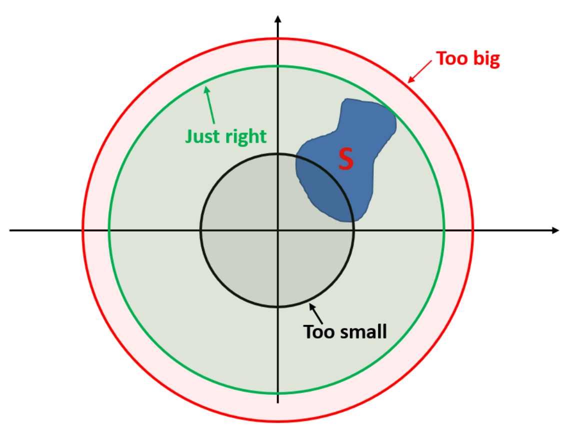
(c) Now, why would we care to know all solutions to an equation? Well, why do like that `google maps` knows many paths from your current location to your goal location? Because it can then optimize the length of your trip in terms of time, distance, or cost (tolls vs no tolls). That's cool! **A problem having more than one solution is awesome because then you get to choose the one you like best!** Exactly. And when the set of all solutions is easy to describe, such as $x_0 + \text{null}(A)$, then it is often very fast and easy to choose a "best" solution from the set of all solutions. We'll do this in Chapter 12.

4. Underdetermined equations, what's up with them?

Answer: See the above question about the null space of a matrix. It explains that under determined problems are ones that have more than one solution. Moreover, they allow engineers to find a best solution to a problem.



(a)



(b)

Figure 1: One attempt to build a picture in your mind about span$\{S\}$ making an analogy with the smallest circle about the origin that contains a given set, S. We will denote that operation by $\text{SmallDisk}\{S\}$. (a) is a subset of $\mathbb{R}^2$, while (b) shows three disks. One of the disks is too small to contain S, one contains S but is not the smallest disk that does this, and one is just right! The **just right disk** is what we define to be $\text{SmallDisk}\{S\}$.

5. Can you help me with the concept of span$\{S\}$ for $S \subset \mathbb{R}^n$?

Answer: Absolutely! To do something different than the textbook, the direction we want to go is "span$\{S\}$ is the smallest subspace of $\mathbb{R}^n$ that contains S." To make sense of this, we'll first do some geometry in $\mathbb{R}^2$. Figure 1 shows a set S. Instead of computing span$\{S\}$, we'll compute $\text{SmallDisk}\{S\}$. What? Hey, go with the flow for a moment. We want to compute the smallest disk (centered about the origin) that contains S. Recall that a disk is the circle plus everything inside the circle.

How could we "compute" or "construct" $\text{SmallDisk}\{S\}$? There are at least two ways:

(a) **Working from the inside and growing the set:** Looking at Figure 1, we can imagine starting with the **too small** circle, then take one a bit bigger, then a bit bigger yet, etc., until we just touch the outside boundary of S. In case you care how we might do that, at the k-th step, assume we have sampled points $\{p_1, p_2, \ldots, p_k\} \subset S$. We then compute which one has the largest Euclidean norm, and we build a disk

$$D_k := \{x \in \mathbb{R}^2 \mid ||x|| \leq \max\{||p_1||, ||p_2||, \ldots, ||p_k||\}\}.$$

Then we know that $\{p_1, p_2, \ldots, p_k\} \subset D_k$ and at least one of them is on the boundary. If it's also true that $S \subset D_k$, then we are done and we'll let you argue that $D_k = \text{SmallDisk}\{S\}$ is the smallest disk that contains S. If $S \not\subset D_k$, then there exists $p_{k+1} \in S$ such that $p_{k+1} \notin D_k$. We then define

$$D_{k+1} := \{x \in \mathbb{R}^2 \mid ||x|| \leq \max\{||p_1||, ||p_2||, \ldots, ||p_k||, ||p_{k+1}||\}\}$$

and repeat our reasoning. We either stop, when we have arrived at $\text{SmallDisk}\{S\}$, or continue growing the disk. In principle[1], we keep computing better and better approximations for $\text{SmallDisk}\{S\}$.

(b) **Working from the outside and shrinking the set:** Looking at Figure 1, we can imagine starting with the **too big** disk, then take one a bit smaller, then a bit smaller yet, etc., until we just touch the outside boundary of S. In case you care how we might do that, we assume that someone (often called an *oracle* in CS and Math) has given us **the collection of all disks** D_i**,** $i \in \mathcal{I}$ that

- are centered about the origin, and
- $S \subset D_i$, that is, each disk contains S.

Now, if D_{i_1} and D_{i_2} are two such disks, then their intersection

$$D_{i_{1,2}} := D_{i_1} \cap D_{i_2}$$

is also a disk centered about the origin and it contains S. Moreover, by construction, $D_{i_{1,2}} \subset D_{i_1}$ and $D_{i_{1,2}} \subset D_{i_2}$ and thus taking intersections "shrinks" the sets. It takes a bit more reasoning, but if you take the intersection of all the disks in the collection,D_i, $i \in \mathcal{I}$, you will arrive at the smallest disk that is centered at the origin and contains S. Hence

$$\text{SmallDisk}\{S\} = \cap_{i \in \mathcal{I}} D_i.$$

That's a long story, why did you tell it? Because if you can use your imagination to understand the "from the inside out" and the "from the outside in" processes on disks to compute $\text{SmallDisk}\{S\}$, then we hope that you can imagine the same processes being used for computing span$\{S\}$.

(a) **Working from the inside and growing the set:** At the k-th step, assume we have selected points $\{p_1, p_2, \ldots, p_k\} \subset S$. We then compute

$$V_k := \text{span}\{p_1, p_2, \ldots, p_k\} := \{\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_k p_k \mid \alpha_i \in \mathbb{R}\}.$$

If it's true that $S \subset V_k$, then we are done and we'll let you argue that $V_k = \text{span}\{S\}$ is the smallest subspace that contains S. If $S \not\subset V_k$, then there exists $p_{k+1} \in S$ such that $p_{k+1} \notin V_k$. We then define

$$V_{k+1} := \text{span}\{p_1, p_2, \ldots, p_k, p_{k+1}\}$$

and repeat our reasoning. At each step we either stop or we grow the subspace by one dimension. Because we can only increase the dimension at most n times, we converge in a finite number of steps.

[1]When you take Math 451, you'll learn that this particular algorithm does not have to converge when S has an infinite number of elements, a detail that we gloss over.

(b) **Working from the outside and shrinking the set:** We assume that someone (often called an *oracle* in CS and Math) has given us **the collection of all subsets** V_i**,** $i \in \mathcal{I}$ that contain S. Now, if V_{i_1} and V_{i_2} are two such subspaces, then their intersection

$$V_{i_{1,2}} := V_{i_1} \cap V_{i_2}$$

is also a subspace and it contains S. Moreover, by construction, $V_{i_{1,2}} \subset V_{i_1}$ and $V_{i_{1,2}} \subset D_{i_2}$ and thus taking intersections "shrinks" the sets. It takes a bit more reasoning, but if you take the intersection of all the subsets in the collection,V_i, $i \in \mathcal{I}$, you will arrive at the smallest subspace that contains S. Hence

$$\text{span}\{S\} = \cap_{i \in \mathcal{I}} V_i.$$

6. Why is orthogonality important?

   Answer 1: It's kind of the opposite of parallel. Figure 2 shows a range of vectors from being parallel and linearly dependent to vectors that are orthogonal and "what you might call maximally independent".

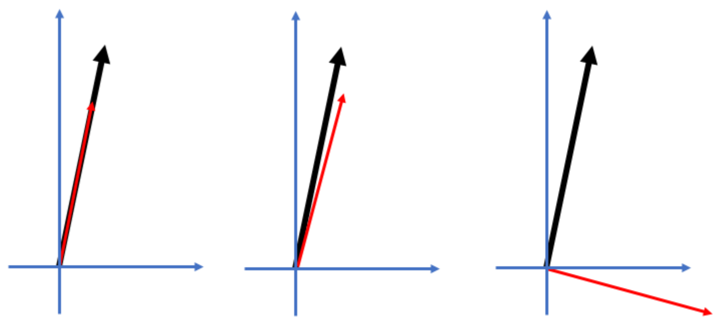


Figure 2: The left shows two parallel vectors, which are clearly linearly dependent. The middle shows two vectors that are linearly independent, but "just barely". The right shows two orthogonal vectors which are clearly linear independent!

   Answer 2: Two non-zero **orthogonal vectors** form a **right angle**. They also satisfy a version of the **Pythagorean Theorem**, if v and v are **orthogonal**, then

$$||u + v||^2 = ||u||^2 + ||v||^2.$$

7. What's the difference between orthogonal and orthonormal matrices?

   Answer: Not much! **Orthogonal matrices** must be square, say $n \times n$, and satisfy $Q^\top \cdot Q = Q \cdot Q^\top = I_n$, where as **orthonormal matrices** can be rectangular, say $n \times m$, and satisfy

   - $Q^\top \cdot Q = I_m$, if $n \geq m$, and
   - $Q \cdot Q^\top = I_n$, for $m \geq n$.

   The condition $Q^\top \cdot Q = I_m$ means the columns of Q are **orthonormal**, while the condition $Q \cdot Q^\top = I_n$ means the rows of Q are **orthonormal**.

8. Gram Schmidt (GS) is a real pain to do by hand. Why do we need it?

   Answer: Real engineers only do GS on a computer. We did a few examples by hand to (hopefully) clarify the calculations. When we program it up in Julia, we can easily include the normalization step, thereby producing **orthonormal vectors** instead of "just" **orthogonal** vectors. If we are running GS with normalization on the columns of a matrix A, then we produce an **orthonormal basis** for col span$\{A\}$.

9. QR Factorization is hard for me to grasp. Why is it related to back substitution?

   Answer: Suppose that want to solve $Ax = b$. One approach is **LU Factorization**, where we have $P \cdot A = L \cdot U$, and then we have **two triangular problems to solve**:

   - $Ly = Pb$ via **forward substitution** because L is **lower triangular**, and

- $Ux = y$, via **back substitution** because U is **upper triangular**.

If we use GS with normalization and write $A = Q \cdot R$, then we have one triangular set of equations to solve

- $Rx = Q^\top b$ via **back substitution** because R is **upper triangular**,
- where we have used the fact that

$$Ax = b \iff Q \cdot Rx = b \implies Q^\top \cdot Q \cdot Rx = Q^\top b \iff Rx = Q^\top b,$$

  because $Q^\top \cdot Q = I$ when the columns of A are linearly independent. When A has ore rows than columns, then we are computing a least squares solution.

10. I am overwhelmed by the number of new terms. Help?!?

    Answer: Dear Learner, You really need to make a cheat sheet. Sincerely, Broken Record.

11. I need to see the concepts applied in the context of real problems. What should I do?

    Answer: Prof. Stephen Boyd's book may help. It is free as an online PDF http://vmls-book.stanford.edu/. Otherwise, we go into depth in the three Projects, but of course, they do not use everything we have learned.