MATH 2210Q Applied Linear Algebra Notes

Arthur J. Parzygnat

These are my personal notes. This is *not* a substitute for Lay's book. I will frequently reference both recent versions of this book. The 4th edition will henceforth be referred to as [2] while the 5th edition will be [3]. In case comments apply to both versions, these two books will both be referred to as [Lay]. You will *not* be responsible for any **Remark**s in these notes. However, everything else, including what is in [Lay] (even if it's not here), is fair game for homework, quizzes, and exams. At the end of each lecture, I provide a list of recommended exercise problems that should be done after that lecture. Some of these exercises will appear on homework, quizzes, or exams! I also provide additional exercises throughout the notes which I believe are good to know. You should also browse other books and do other problems as well to get better at writing proofs and understanding the material.

Notes in light red are for the reader.

Notes in light green are reminders for me.

When a word or phrase is *underlined*, that typically means the definition of this word or phrase is being given.

Contents

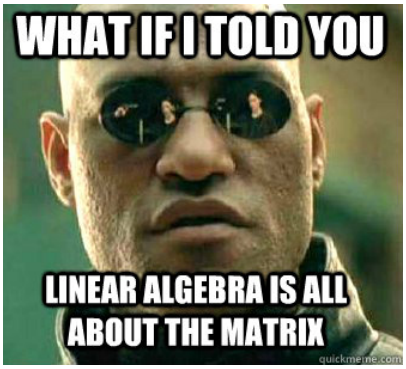
Sections with a * at the end are additional topics which can be covered if time permits.

Acknowledgments

I'd like to thank Christian Carmellini, Philip Parzygnat, Zachariah Pittman, Benjamin Russo, Xing Su, Yun Yang, and George Zoghbi for many helpful suggestions and comments.

1 Linear systems, row operations, and examples

Before saying what one studies in linear algebra, let us consider the following examples. These examples will illustrate the important concept of a mathematical object known as the matrix.



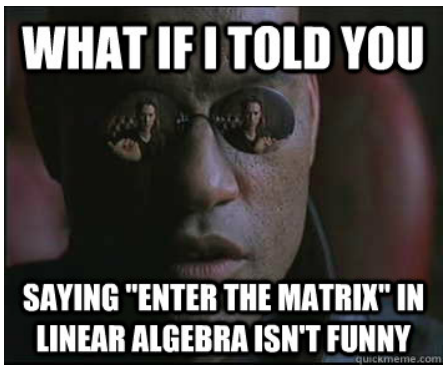


Figure 1: Memes "What if I told you Linear Algebra is all about the matrix" http://www.quickmeme.com/meme/3qwkiq and "What if I told you saying "enter the matrix" in linear algebra isn't funny" http://www.quickmeme.com/meme/36c7p8, respectively. Accessed on December 21, 2017.

Example 1.1. Queens, New York has several one-way streets throughout its many neighborhoods. Figure 2 shows an intersection in Middle Village, New York. We can represent the flow of traffic

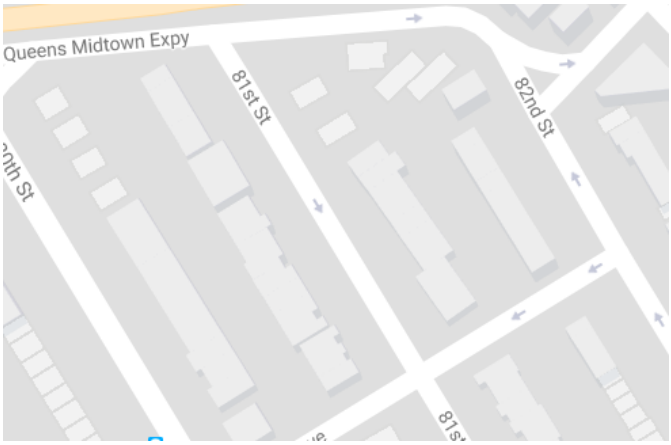
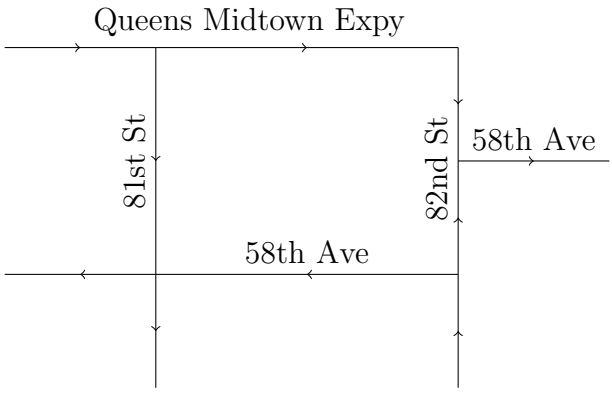


Figure 2: An intersection in Middle Village, New York in the borough of Queens. This image is obtained from Map data ©2017 Google https://www.google.com/maps/place/Middle+Village,+Queens,+NY/@40.7281297,-73.8802503,18.72z/data=!4m5!3m4!1s0x89c25e6887df03e7:0xef1a62f95c745138!8m2!3d40.717372!4d-73.87425

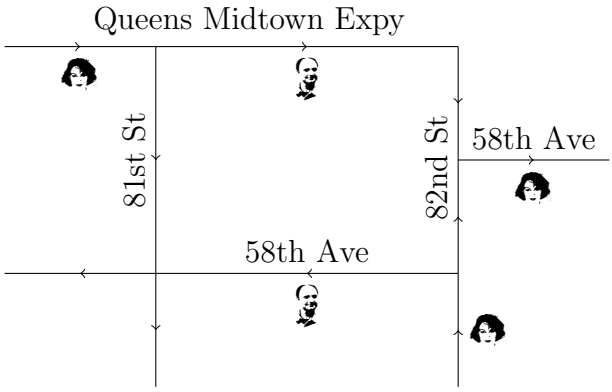
around 81st and 82nd streets diagrammatically as



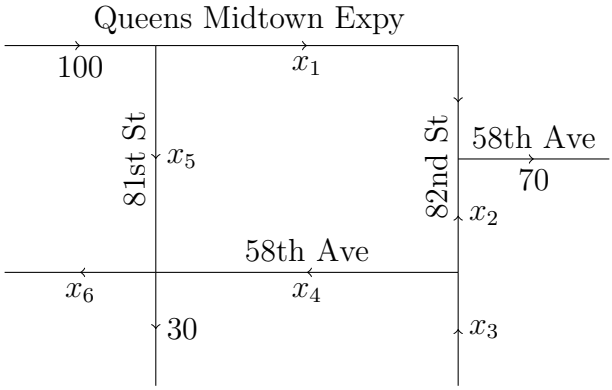
Imagine that we send out detectors (such as scouts) to record the average number of cars per hour along each street. What is the *smallest* number of scouts we will need to determine the traffic flow on *every* street? To answer this question, we first point out one important assumption that appears in several different contexts:

The net flow into an intersection equals the net flow out of an intersection.

From this, it actually follows that the net flow into the network itself is equal to the net flow out of the network. Each edge connecting any two intersections represents an unknown and each fact above provides an equation. Hence, this system has 9 unknowns and 4 equations. Therefore, one *expects* that the minimum number of scouts needed is $9 - 4 = 5$. However, this is certainly not a *proof* because some of these equations might be redundant! Furthermore, even if 5 is the minimum number of scouts needed, it does *not* mean that you can place these scouts anywhere and determine the entire traffic flow. For example, if you place the 5 scouts on the following streets



then you still don't know the traffic flow leaving 58th Ave and 81st Street on the bottom left. Let's see what happens explicitly by first sending out 3 scouts, which observe the following traffic flow per hour



The unknown traffic flows have been labelled by the variables $x_1, x_2, x_3, x_4, x_5, x_6$, which is where we did *not* send out any scouts. The equations for the "flow in" equals "flow out" are given by (they are written going clockwise starting at the top left)

$$\begin{aligned} 100 &= x_1 + x_5 \\ x_1 + x_2 &= 70 \\ x_3 &= x_2 + x_4 \\ x_4 + x_5 &= 30 + x_6 \end{aligned} \tag{1.2}$$

This system of linear equations can be rearranged in the following way

$$\begin{array}{cccccccc} x_1 & & & & +x_5 & & = & 100 \\ x_1 & +x_2 & & & & & = & 70 \\ & x_2 & -x_3 & +x_4 & & & = & 0 \\ & & & x_4 & +x_5 & -x_6 & = & 30 \end{array} \tag{1.3}$$

which makes it easier to see how to manipulate these expressions algebraically by adding or subtracting multiples of different rolls. When adding these rows, all we ever add are the coefficients and the variables are just there to remind us of our organization. We can therefore replace these equations with the *augmented matrix*

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 1 & 1 & 0 & 0 & 0 & 0 & 70 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right]. \tag{1.4}$$

Adding and subtracting rows here corresponds to the same operations for the equations. For example, subtract row 1 from row 2 to get

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 0 & 1 & 0 & 0 & -1 & 0 & -30 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \tag{1.5}$$

The result corresponds to the system of equations

$$\begin{array}{rrrrrrrr} x_1 & & & & +x_5 & & = & 100 \\ & x_2 & & & -x_5 & & = & -30 \\ & x_2 & -x_3 & +x_4 & & & = & 0 \\ & & & x_4 & +x_5 & -x_6 & = & 30 \end{array} \tag{1.6}$$

As we first learn about these operations, we will perform them one at a time and show what happens to them explicitly by the following notation

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 1 & 1 & 0 & 0 & 0 & 0 & 70 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \xrightarrow{R2 \mapsto R2-R1} \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 0 & 1 & 0 & 0 & -1 & 0 & -30 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \tag{1.7}$$

which is read as "row 2 becomes row 2 minus row 1." We implicitly understand that all the other rows remain *unchanged* unless explicitly written otherwise. Subtract row 2 from row 3 to get

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 0 & 1 & 0 & 0 & -1 & 0 & -30 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \xrightarrow{R3 \mapsto R3-R2} \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 0 & 1 & 0 & 0 & -1 & 0 & -30 \\ 0 & 0 & -1 & 1 & 1 & 0 & 30 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \tag{1.8}$$

Subtract row 4 from row 3 to get

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 0 & 1 & 0 & 0 & -1 & 0 & -30 \\ 0 & 0 & -1 & 1 & 1 & 0 & 30 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \xrightarrow{R3 \mapsto R3-R4} \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 0 & 1 & 0 & 0 & -1 & 0 & -30 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \tag{1.9}$$

Multiply row 3 by -1 to get rid of the negative coefficient for the x_3 variable (this step is not necessary and is mostly just for the A E S T H E T I C S)

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 0 & 1 & 0 & 0 & -1 & 0 & -30 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \xrightarrow{R3 \mapsto -R3} \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & 100 \\ 0 & 1 & 0 & 0 & -1 & 0 & -30 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 30 \end{array}\right] \tag{1.10}$$

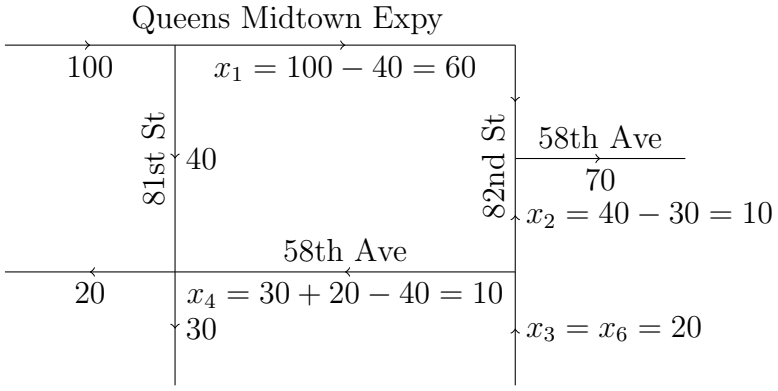
This tells us that the initial system of linear equations is equivalent to

$$\begin{array}{rrrrrrrr} x_1 & & & & +x_5 & & = & 100 \\ & x_2 & & & -x_5 & & = & -30 \\ & & x_3 & & & -x_6 & = & 0 \\ & & & x_4 & +x_5 & -x_6 & = & 30 \end{array} \tag{1.11}$$

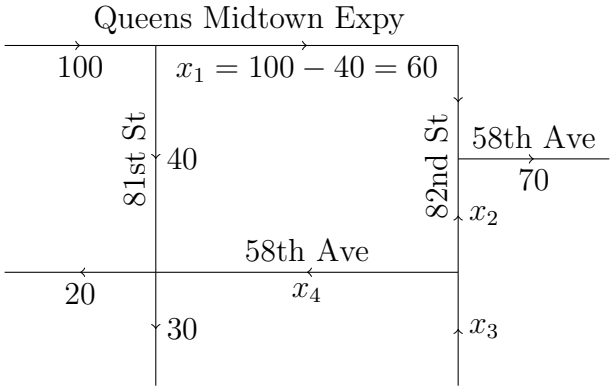
or

$$\begin{aligned} x_1 &= 100 - x_5 \\ x_2 &= x_5 - 30 \\ x_3 &= x_6 \\ x_4 &= 30 + x_6 - x_5 \end{aligned} \tag{1.12}$$

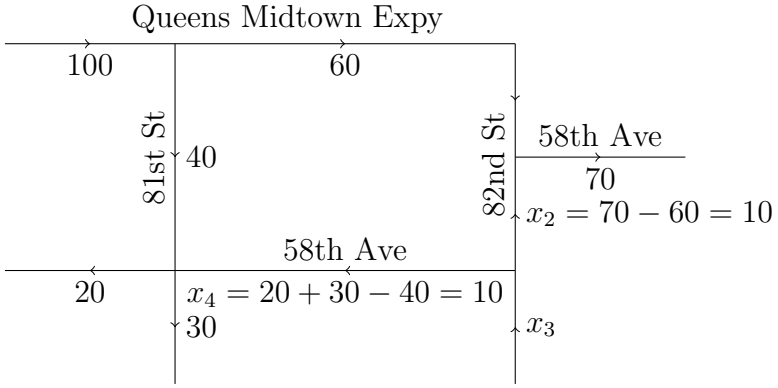
so that all of the traffic flows are expressed in terms of just x_5 and x_6. This choice is arbitrary and we could have expressed another four traffic flows in terms of the other two (again, not *any* four, but *some*). In this case, x_5 and x_6 are called *free variables*. In order to figure out the entire traffic flow through all streets, we can therefore send out two more scouts to observe x_5 and x_6. Let's suppose that the scouts discover that $x_5 = 40$ and $x_6 = 20$. Knowing these values, we can figure out the traffic flow through every street without sending out additional scouts by plugging these values into (1.12)



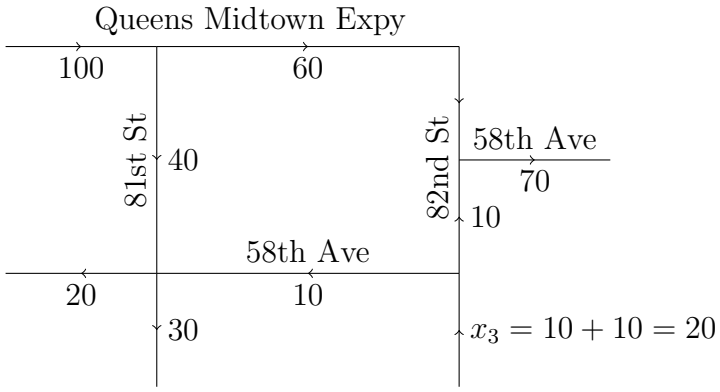
The method employed above determines the traffic flow at every street through the method of *row reduction*. We could have also figured out x_1, x_2, x_3, and x_4 by brute force by just implementing the intersection network flow conservation at every intersection without using any of the above methods. For instance, first we can figure out x_1.



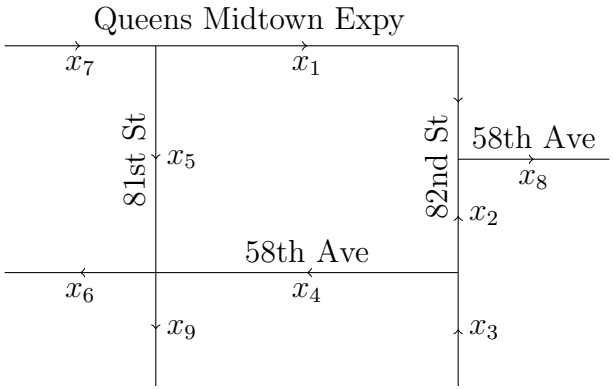
x_2 and x_4 can then each be solved immediately



and finally x_3.



Although the second method was much faster in this case, when dealing with large scale intersections over several blocks, this becomes much more difficult. The first method is more systematic and applies to all situations. For example, we can use the first method to *prove* that 5 is the minimum number of scouts needed. To do this, we place unknown variables at every street



This time, the equations for the "flow in" equals "flow out" are given by (they are written going clockwise starting at the top left)

$$\begin{aligned} x_7 &= x_1 + x_5 \\ x_1 + x_2 &= x_8 \\ x_3 &= x_2 + x_4 \\ x_4 + x_5 &= x_9 + x_6 \end{aligned} \tag{1.13}$$

This system of linear equations can be rearranged in the following way

$$\begin{array}{rrrrrrrrrcl} x_1 & & & & +x_5 & & -x_7 & & & = & 0 \\ x_1 & +x_2 & & & & & & -x_8 & & = & 0 \\ & x_2 & -x_3 & +x_4 & & & & & & = & 0 \\ & & & x_4 & +x_5 & -x_6 & & & -x_9 & = & 0 \end{array} \tag{1.14}$$

with corresponding augmented matrix

$$\left[\begin{array}{ccccccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 \end{array}\right] \tag{1.15}$$

Applying a similar row reduction procedure to before gives

$$\left[\begin{array}{ccccccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 \end{array}\right] \xrightarrow{R2 \mapsto R2 - R1} \left[\begin{array}{ccccccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 \end{array}\right]$$

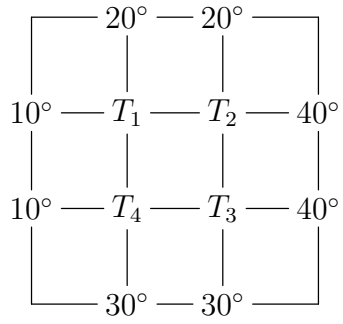
$$R3 \mapsto R3 - R2 - R1 \downarrow$$

$$\left[\begin{array}{ccccccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 \end{array}\right] \xleftarrow{R3 \mapsto -R3} \left[\begin{array}{ccccccccc|c} 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & -1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 \end{array}\right]$$

This tells us that the variables x_1, x_2, x_3, x_4, the dependent variables, are completely determined by the variables x_5, x_6, x_7, x_8, and x_9, the independent/free variables, which have no restriction among them (other than to guarantee that traffic flow is always nonnegative). Since there are five such variables, a minimum of five scouts is needed.

We will explain what choices one makes to reduce the linear system as we have, but first we will go through more examples.

Problem 1.16 (Exercise 1.1.33 in [Lay])**.** The temperature on the boundary of a cross section of a metal beam is fixed and known but is unknown at the intermediate points on the interior



(1.17)

Assume the temperature at these intermediate points equals the average of the temperature at the nearest neighboring points.[1] Calculate the temperatures T_1, T_2, T_3, and T_4.

[1] This is true to a good approximation and is in fact how approximation techniques can be used to solve problems like this though the mesh will usually be much finer, and the boundary might not look so nice. Furthermore, the solution we are obtaining is the steady state solution, which is what the temperatures will be after you wait long enough. For instance, if you dumped the beam into an ice bath, it would take time for the temperatures to be stable on the inside of the beam so that this method would work. We use the phrase "steady state" instead of "equilibrium" because something is forcing the temperatures to be different on the different edges of the beam.

Answer. The system of equations is given by

$$\begin{aligned} T_1 &= \frac{1}{4}(10 + 20 + T_2 + T_4) \\ T_2 &= \frac{1}{4}(T_1 + 20 + 40 + T_3) \\ T_3 &= \frac{1}{4}(T_4 + T_2 + 40 + 30) \\ T_4 &= \frac{1}{4}(10 + T_1 + T_3 + 30) \end{aligned} \tag{1.18}$$

Rewriting them so that the variables appear in order gives

$$\begin{aligned} 4T_1 - 1T_2 + 0T_3 - 1T_4 &= 30 \\ -1T_1 + 4T_2 - 1T_3 + 0T_4 &= 60 \\ 0T_1 - 1T_2 + 4T_3 - 1T_4 &= 70 \\ -1T_1 + 0T_2 - 1T_3 + 4T_4 &= 40. \end{aligned} \tag{1.19}$$

The coefficients in front of the unknown temperatures in (1.19) can be put together in an array[2]

$$\left[\begin{array}{cccc|c} 4 & -1 & 0 & -1 & 30 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & -1 & 4 & -1 & 70 \\ -1 & 0 & -1 & 4 & 40 \end{array}\right] \tag{1.20}$$

This *augmented matrix* will aid in implementing calculations to solve for the temperatures. From a course in algebra, you might guess that one way to solve for the temperatures is to solve for one and then plug in this value successively into the other ones. This becomes difficult when we have more than two variables. Some things we *can* do, which are more effective, are adding linear combinations of equations within the system (1.19). For instance, subtracting row 4 of (1.19) by row 2 gives

$$\begin{aligned} -1T_1 + 0T_2 - 1T_3 + 4T_4 &= 40 \\ -\big(-1T_1 + 4T_2 - 1T_3 + 0T_4 &= 60\big) \\ \hline 0T_1 - 4T_2 + 0T_3 + 4T_4 &= -20 \end{aligned} \tag{1.21}$$

for row 4. We are allowed to do this provided that a solution *exists* in the first place. We can also multiply this equation by $\frac{1}{4}$ without changing the values of the variables. This gives

$$0T_1 - 1T_2 + 0T_3 + 1T_4 = -5. \tag{1.22}$$

From this, we see that we are only manipulating the entries in the augmented matrix (1.20) and we don't have to constantly rewrite all the T variables. In other words, the augmented matrix

[2] [Lay] does not draw a vertical line to separate the two sides. I find this confusing. We will always draw this line to be clear.

becomes

$$\left[\begin{array}{cccc|c} 4 & -1 & 0 & -1 & 30 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & -1 & 4 & -1 & 70 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right] \tag{1.23}$$

after these two *row operations*. If we could get rid of T_2 from this last row, we could solve for T_4 (or vice versa). Similarly, we should try to solve for all the other temperatures by finding combinations of rows to eliminate as many entries from the left-hand-side of the augmented matrix. One possible sequence of row operations achieving this goal is

$$\left[\begin{array}{cccc|c} 4 & -1 & 0 & -1 & 30 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & -1 & 4 & -1 & 70 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right] \xrightarrow{R1 \mapsto R1 + 4R2} \left[\begin{array}{cccc|c} 0 & 15 & -4 & -1 & 270 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & -1 & 4 & -1 & 70 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 4 & -1 & 0 & -1 & 30 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & -1 & 4 & -1 & 70 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right] \xrightarrow{R1 \mapsto R1 + 4R2} \left[\begin{array}{cccc|c} 0 & 15 & -4 & -1 & 270 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & -1 & 4 & -1 & 70 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right] \downarrow R1 \mapsto R1 + 15R4$$

$$\left[\begin{array}{cccc|c} 0 & 0 & -4 & 14 & 195 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & 0 & 4 & -2 & 75 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right] \xleftarrow{R3 \mapsto R3 - R4} \left[\begin{array}{cccc|c} 0 & 0 & -4 & 14 & 195 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & -1 & 4 & -1 & 70 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right]$$

$$\downarrow R1 \mapsto R1 + R3$$

$$\left[\begin{array}{cccc|c} 0 & 0 & 0 & 12 & 270 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & 0 & 4 & -2 & 75 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right] \xrightarrow{R1 \mapsto \frac{1}{6}R1} \left[\begin{array}{cccc|c} 0 & 0 & 0 & 2 & 45 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & 0 & 4 & -2 & 75 \\ 0 & -1 & 0 & 1 & -5 \end{array}\right]$$

$$\downarrow R3 \mapsto R3 + R1 \quad R4 \mapsto R4 - \tfrac{1}{2}R1$$

$$\left[\begin{array}{cccc|c} 0 & 0 & 0 & 2 & 45 \\ -1 & 0 & -1 & 0 & -50 \\ 0 & 0 & 1 & 0 & 30 \\ 0 & -1 & 0 & 0 & -27.5 \end{array}\right] \xleftarrow[R3 \mapsto \frac{1}{4}R3]{R2 \mapsto R2 + 4R4} \left[\begin{array}{cccc|c} 0 & 0 & 0 & 2 & 45 \\ -1 & 4 & -1 & 0 & 60 \\ 0 & 0 & 4 & 0 & 120 \\ 0 & -1 & 0 & 0 & -27.5 \end{array}\right]$$

$$\downarrow R2 \mapsto R2 + R3$$

$$\left[\begin{array}{cccc|c} 0 & 0 & 0 & 2 & 45 \\ -1 & 0 & 0 & 0 & -20 \\ 0 & 0 & 1 & 0 & 30 \\ 0 & -1 & 0 & 0 & -27.5 \end{array}\right] \xrightarrow[R4 \mapsto -R3]{\substack{R1 \mapsto \frac{1}{2}R1 \\ R2 \mapsto -R2}} \left[\begin{array}{cccc|c} 0 & 0 & 0 & 1 & 22.5 \\ 1 & 0 & 0 & 0 & 20 \\ 0 & 0 & 1 & 0 & 30 \\ 0 & 1 & 0 & 0 & 27.5 \end{array}\right]$$

In other words, we have found a solution

$$\begin{aligned} T_1 &= 20 \\ T_2 &= 27.5 \\ T_3 &= 30 \\ T_4 &= 22.5 \end{aligned} \qquad \begin{array}{ccccc} & & 20^\circ & - & 20^\circ & & \\ & | & & | & & | & \\ 10^\circ & - & 20^\circ & - & 27.5^\circ & - & 40^\circ \\ & | & & | & & | & \\ 10^\circ & - & 22.5^\circ & - & 30^\circ & - & 40^\circ \\ & | & & | & & | & \\ & & 30^\circ & - & 30^\circ & & \end{array} \tag{1.24}$$

Because it helps to visualize this the same way, we can permute the rows and still have the same equations describing our problem

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 20 \\ 0 & 1 & 0 & 0 & 27.5 \\ 0 & 0 & 1 & 0 & 30 \\ 0 & 0 & 0 & 1 & 22.5 \end{array}\right] \tag{1.25}$$

This is another example of a *row operation*.

All of these examples have some features in common. In particular, they exhibit linear behavior of some sort. However, each system is quite different and one might think that to properly analyze these systems, one needs to work with each system separately. To a large extent, this is false. Instead, if one can abstract the crucial properties of linearity more precisely without the particular model one is looking at, then one can study these properties and make conclusions abstractly. Then by going back to the particular problem, one can apply these conclusions to say something about the particular system.

Linear algebra is the study of these abstract properties.

Not all problems in nature behave in such a linear fashion. Nevertheless, certain aspects of the system can be approximated linearly. This is where the techniques of linear algebra apply. Linear algebra is the study of systems of linear equations. Although not all physical situations are described by linear equations, the first order approximations of such systems are typically linear. Linear systems are much simpler to solve and give a decent approximation to the local behavior of a physical system.

Definition 1.26. A *linear system* (or a *system of linear equations*) in a finite number of variables is a collection of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \tag{1.27}$$

where the a_{ij} are real numbers (typically known constants), the b_i are real numbers (also typically known values), and the x_j are the variables (which we would often like to solve for). The *solution*

set of a linear system (1.27) is the collection of all $(x_1, x_2, \ldots, x_n)$ that satisfy (1.27). A linear system where the solution set is non-empty is said to be *consistent*. A linear system where the solution set is empty is said to be *inconsistent*.

It helps to start off immediately with some simple examples. We will slowly develop a more formal and rigorous approach to linear algebra as the semester progresses.

Example 1.28. Consider the linear system given by

$$\begin{aligned} -x - y + z &= -2 \\ -2x + y + z &= 1 \end{aligned} \tag{1.29}$$

These two equations are plotted in Figure 3.

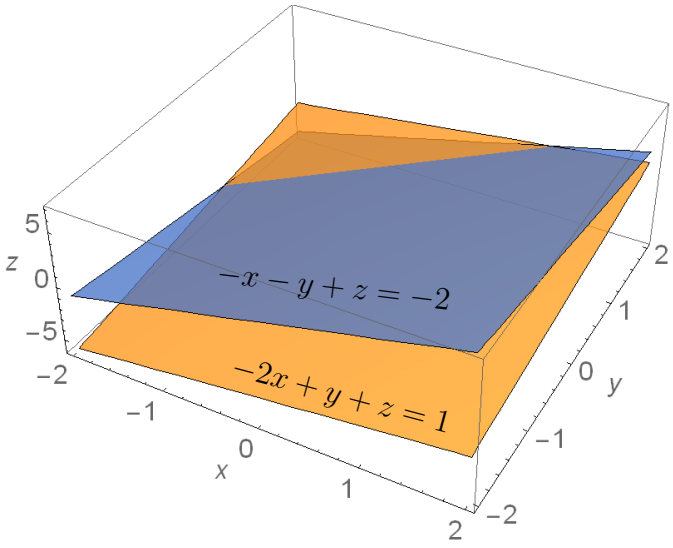


Figure 3: A plot of the equations $-x - y + z = -2$ and $-2x + y + z = 1$.

This picture shows that there are solutions, in fact a lines worth of solutions instead of a unique one (the intersection of the two planes is the set of solutions). How can we describe this line explicitly? Looking at (1.29), we can add the two equations to get[3]

$$-3x + 2z = -1 \iff z = \frac{1}{2}(3x - 1) \tag{1.30}$$

We can also subtract the second equation from the first to get

$$x - 2y = -3 \iff y = \frac{1}{2}(3 + x). \tag{1.31}$$

Hence, the set of points given by

$$\left(x, \frac{1}{2}(3 + x), \frac{1}{2}(3x - 1)\right) \tag{1.32}$$

[3]The $\iff$ symbol means "if and only if," which in this context means that the two equations are equivalent.

as x varies over real numbers, are all solutions of (1.29). In set-theoretic notation, the solution set would be written as

$$\left\{ \left(x, \frac{1}{2}(3+x), \frac{1}{2}(3x-1) \right) \in \mathbb{R}^3 \ : \ x \in \mathbb{R} \right\}. \tag{1.33}$$

We can plot this set, along with the two planes described by the two linear equations, in Figure 4.

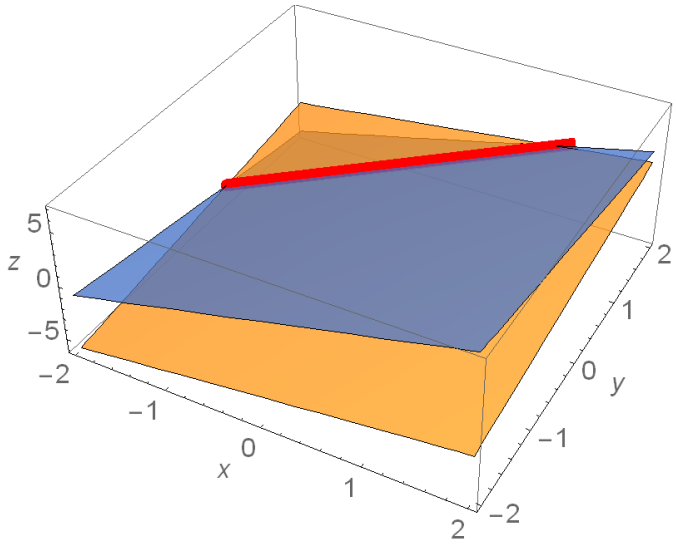


Figure 4: A plot of the equations $-x-y+z=-2$ and $-2x+y+z=1$ together with the intersection shown in red and given parametrically as $x \mapsto \left(x, \frac{1}{2}(3+x), \frac{1}{2}(3x-1)\right)$.

Hence, (1.29) is an example of a consistent system.

In this example, we saw that not only could we find solutions, but there were infinitely many solutions. Sometimes, a solution to a linear system need not exist at all!

Example 1.34. Let

$$\begin{aligned} 2x+3y &= 5 \\ 4x+6y &= -2 \end{aligned} \tag{1.35}$$

be two linear equations in the variables x and y. There is no solution to this system. If there were a solution, then dividing the second line by 2 would give $5=-1$, which is impossible.[4] This can also be seen by plotting these two equations in the plane as in Figure 5. These two lines do not intersect. Hence, (1.35) is an example of an inconsistent system.

To test yourself that you understand these definitions, try to answer the following true or false questions.

Problem 1.36. State whether the following claims are True or False. If the claim is true, be able to precisely deduce why the claim is true. If the claim is false, be able to provide an explicit counter-example.

[4]This is an example of a proof by contradiction.

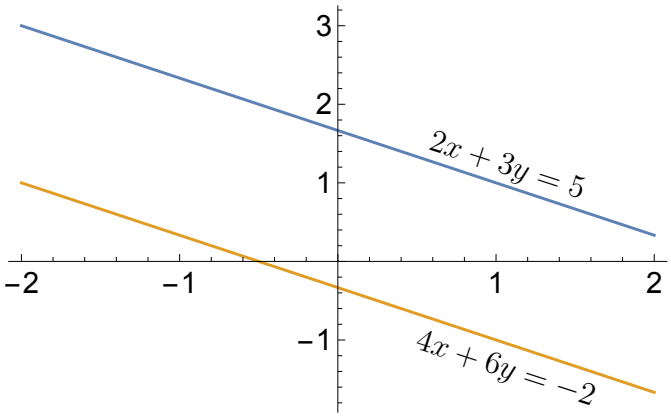


Figure 5: A plot of the equations $2x + 3y = 5$ and $4x + 6y = -2$.

(a) If a linear system has infinitely many solutions, then the linear system is inconsistent.

(b) If a linear system is consistent, then it has infinitely many solutions.

(c) Every linear system of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= 0 \\ a_{21}x_2 + a_{22}x_2 &= 0 \end{aligned} \tag{1.37}$$

is consistent for all real numbers a_{11}, a_{12}, a_{21}, and a_{22}.

Many of these questions have very simple answers. The difficulty is not just in knowing just which statement is true or false, but also being able to prove your claim. If there is ever a claim that seems insultingly simple, try to prove it. The reason for proving simpler claims is not so much to convince you of their validity but to get used to the way in which proofs are done in the simplest of examples. Therefore, we will present the solutions for now, but will eventually leave many such exercises throughout the notes.

Answer.

(a) False: A counterexample is in Example 1.28, which has infinitely many solutions.

(b) False: A counterexample will be presented shortly in Problem 1.16 below. The linear system described there is consistent and has only one solution.

(c) True: Setting $x_1 = 0$ and $x_2 = 0$ gives one solution regardless of what a_{11}, a_{12}, a_{21}, and a_{22} are.

In general, you should think about every definition that you are introduced to and be able to relate it to examples and general situations. Always compare definitions to understand the differences if some seem similar. We will now go through a more complicated and challenging linear system where it will be useful to introduce the concept of a matrix.

In this situation, we were lucky and a solution existed *and* was unique. In problem 1.16, there is only one element in the solution set. Occasionally, two arbitrary linear systems may have the same set of solutions.

Definition 1.38. Two linear systems of equations with the same variables that have the same set of solutions are said to be *equivalent*.

Hence, the two linear systems of equations given in (1.19) and (1.25) are equivalent. In going from one equation to another, several row operations were performed, none of which altered the set of solutions. In total, we have used three *row operations* to help us solve linear systems:

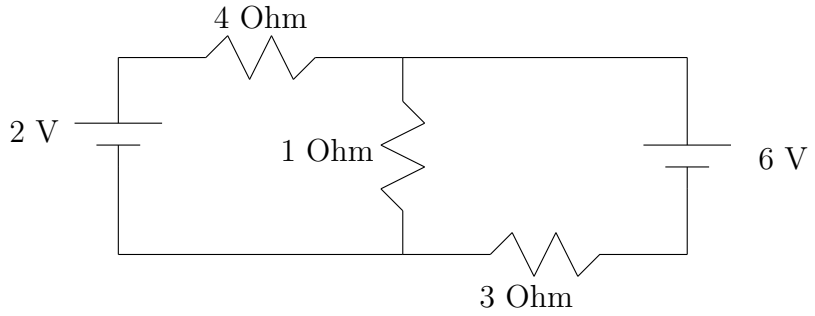
(a) scaling rows,

(b) adding rows, and

(c) permuting rows.

As we do more problems, we will get familiar with faster methods of solving systems of linear equations. We start with another problem from circuits with batteries and resistors.

Problem 1.39. Consider a circuit of the following form

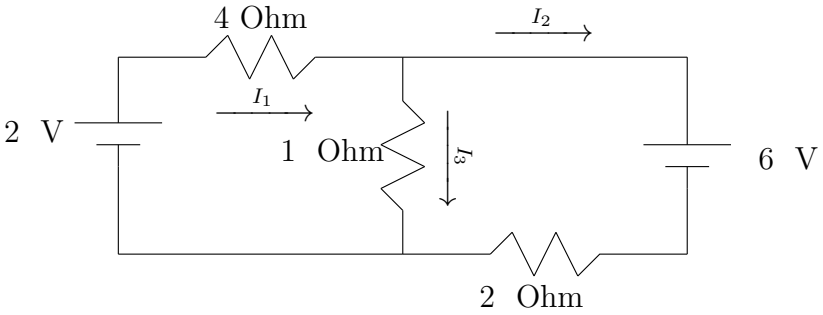


Here the jagged lines represent resistors and the two parallel lines, with one shorter than the other, represent batteries with the positive terminal on the longer side. The units of resistance are Ohms and the units for voltage are Volts. Find the current (in units of Amperes) across each resister along with the direction of current flow.

Answer. Before solving this, we recall a crucial result from physics, which is

Kirchhoff's rule: the voltage difference across any closed loop in a circuit with resistors and batteries is always zero.

Across a resistor, the voltage drop is the current times the resistance (this is called *Ohm's law*). Across a battery from the negative to positive terminal, there is a voltage increase given by the voltage of the battery. There is also the rule that says current is always conserved, meaning that at a junction, "current in" equals "current out", just as in Example 1.1. Knowing this, we label the currents in the wires by I_1, I_2, and I_3 as follows.



The directionality of these currents has been chosen arbitrarily. Conservation of current gives

$$I_1 = I_2 + I_3. \tag{1.40}$$

Kirchhoff's rule for the left loop in the circuit gives

$$2 - 4I_1 - 1I_3 = 0 \tag{1.41}$$

and for the right loop gives

$$-6 - 2I_2 + 1I_3 = 0. \tag{1.42}$$

These are three equations in three unknowns.

If you were lost up until this point, that's fine. You can start by assuming the following form for the linear system of equations.

Rearranging them gives

$$\begin{aligned} 1I_1 - 1I_2 - 1I_3 &= 0 \\ 0I_1 - 2I_2 + 1I_3 &= 6 \\ 4I_1 + 0I_2 + 1I_3 &= 2 \end{aligned} \tag{1.43}$$

and putting it in augmented matrix form gives

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & -2 & 1 & 6 \\ 4 & 0 & 1 & 2 \end{array}\right]. \tag{1.44}$$

Performing row operations to isolate as many unknowns as possible gives

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & -2 & 1 & 6 \\ 4 & 0 & 1 & 2 \end{array}\right] \xrightarrow{R3 \mapsto R3 - 4R1} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & -2 & 1 & 6 \\ 0 & 4 & 5 & 2 \end{array}\right] \xrightarrow{R3 \mapsto R3 + 2R2} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & -2 & 1 & 6 \\ 0 & 0 & 7 & 14 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & -2 & 1 & 6 \\ 0 & 0 & 7 & 14 \end{array}\right] \xrightarrow{R3 \mapsto \frac{1}{7}R3} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & -2 & 1 & 6 \\ 0 & 0 & 1 & 2 \end{array}\right] \xrightarrow{\substack{R3 \mapsto R3 + R1 \\ R2 \mapsto R2 - R3}} \left[\begin{array}{ccc|c} 1 & -1 & 0 & 2 \\ 0 & -2 & 0 & 4 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & -1 & 0 & 2 \\ 0 & -2 & 0 & 4 \\ 0 & 0 & 1 & 2 \end{array}\right] \xrightarrow{\substack{R1 \mapsto R1 - \frac{1}{2}R2 \\ R2 \mapsto -\frac{1}{2}R2}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

We have thus found our solution

$$\begin{aligned} I_1 &= 0 \text{ A} \\ I_2 &= -2 \text{ A} \\ I_3 &= 2 \text{ A} \end{aligned} \tag{1.45}$$

The negative sign means that the current is actually flowing in the opposite direction to what we assumed.

If you've noticed, there is usually a lot of freedom in the row reduction process, but the end goal is always similar. The idea is to separate the unknowns into two types. One type of unknown is a free/independent variable, and the other type is a dependent variable. Isolating which variables are dependent and independent is an important factor in understanding the set of solutions. Row reduction is the process by which one identifies the dependent and independent variables and writes the solution, dependent variables, in terms of the independent variables. If there are no independent variables, at most one solution exists.

Definition 1.46. Given a linear system of equations as in (1.27), which is written as an augmented matrix as

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right], \tag{1.47}$$

an *echelon form* of such an augmented matrix is an *equivalent* augmented matrix whose matrix components (to the left of the vertical line) satisfy the following conditions.

(a) All nonzero rows are above any rows containing only zeros.

(b) The first nonzero entry (from the left) of any row is always to the right of the first nonzero entry of the row directly above it.

(c) All entries in the column below the first nonzero entry of any row are zeros.

Once an augmented matrix is in echelon form, the first nonzero entry in a given row is called a *pivot*. The column containing a pivot is called a *pivot column* while a column that does not contain a pivot is called a *free variable column*. A matrix is in *reduced echelon form* if, in addition, the following hold.

(d) All pivots are 1.

(e) The pivots are the only nonzero entries in the corresponding pivot columns.

Conditions (b) and (e) together say that all entries above and below a pivot are all zero.

Exercise 1.48. State whether the following augmented matrices are in echelon form. If they are not, use row operations to find an equivalent matrix that is in echelon form. [Hint: identify the pivots first.]

(a) $\left[\begin{array}{cccc|c} 5 & 0 & 1 & -1 & 5 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

(b) $\left[\begin{array}{cccc|c} 5 & 0 & 1 & -1 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \end{array}\right]$

(c) $\left[\begin{array}{cccc|c} 5 & 0 & 1 & -1 & 5 \\ 5 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

(d) $\left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

(e) $\left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

Do the same thing for the previous examples but replace "echelon" with "reduced echelon."

For example, the pivot columns in the augmented matrix in part (a) are the first and third columns. The free variable columns are the second and fourth columns. This is because after row reduction, the solution set for this system is given by $x_1 = 1$, x_2 is free, $x_3 = x_4$, and x_4 is free.

It is a fact that the *reduced* row echelon form of a matrix is always unique provided that the linear system corresponding to it is consistent. This is why the word "the" is used to describe it. However, the word "an" was used for just echelon form because there could be many echelon forms associated to an augmented matrix. Furthermore, a linear system is consistent if and only if an echelon form of the augmented matrix does *not* contain any rows of the form

$$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & b \end{array}\right] \qquad \text{with } b \text{ nonzero.} \tag{1.49}$$

It is important that we look at an echelon form to determine if the system is consistent or not.

What is the point of having an augmented matrix in echelon form? Let us explain this via an example, such as (the pivot columns are colored red, in bold font, the pivots are underlined, and the free variable columns are colored blue, in regular font)

$$\left[\begin{array}{cccccc|c} \underline{\mathbf{3}} & 6 & \mathbf{3} & \mathbf{-6} & 3 & -9 & 15 \\ \mathbf{0} & 0 & \underline{\mathbf{2}} & \mathbf{4} & 0 & 0 & -8 \\ \mathbf{0} & 0 & \mathbf{0} & \underline{\mathbf{-1}} & 2 & 4 & 12 \end{array}\right] \tag{1.50}$$

If the variables representing this are given by $x_1, x_2, x_3, x_4, x_5, x_6$ in order, then the solutions are of the form

$$\begin{aligned} 3x_1 &= 15 - 6x_2 - 3x_3 + 6x_4 - 3x_5 + 9x_6 \\ 2x_3 &= -8 - 4x_4 \\ -x_4 &= 12 - 2x_5 - 4x_6 \end{aligned} \tag{1.51}$$

with x_2, x_5, x_6 free. However, this hasn't been reduced completely because, for example, x_4 appears in the equations for x_1 and x_3, but we know that x_4 can be expressed in terms of the free variables. In fact, $x_4 = -12 + 2x_5 + 4x_6$. Plugging this into the equations for x_1 and x_3 gives

$$\begin{aligned} 3x_1 &= 15 - 6x_2 - 3x_3 + 6(-12 + 2x_5 + 4x_6) - 3x_5 + 9x_6 = -57 - 6x_2 - 3x_3 + 9x_5 + 33x_6 \\ 2x_3 &= -8 - 4(-12 + 2x_5 + 4x_6) = 40 - 8x_5 - 16x_6 \\ x_4 &= -12 + 2x_5 + 4x_6 \end{aligned} \tag{1.52}$$

which gets rid of all occurrences of x_4 in the expressions for x_1 and x_3. Unfortunately, x_3 also appears in the expression for x_1 even though it is now apparent that x_3 is expressed solely in terms

of the free variables. Plugging this into the equation for x_1 gives

$$\begin{aligned} 3x_1 &= -57 - 6x_2 - 3(20 - 4x_5 - 8x_6) + 9x_5 + 33x_6 = -117 - 6x_2 + 21x_5 + 57x_6 \\ x_3 &= 20 - 4x_5 - 8x_6 \\ x_4 &= -12 + 2x_5 + 4x_6 \end{aligned} \tag{1.53}$$

Finally, dividing the equation involving x_1 by 3 gives

$$\begin{aligned} x_1 &= -39 - 2x_2 + 7x_5 + 19x_6 \\ x_3 &= 20 - 4x_5 - 8x_6 \\ x_4 &= -12 + 2x_5 + 4x_6 \end{aligned} \tag{1.54}$$

Thus, the point of having an augmented matrix in echelon form is to be able to write the set of solutions by successively plugging in expressions for the latter pivot variables into the former pivot variables if they ever appear. This procedure is precisely the last few steps that takes the echelon form augmented matrix to its *reduced* echelon form

$$\left[\begin{array}{cccccc|c} \underline{\mathbf{1}} & 2 & \mathbf{0} & \mathbf{0} & -7 & -19 & -39 \\ \mathbf{0} & 0 & \underline{\mathbf{1}} & \mathbf{0} & 4 & 8 & 20 \\ \mathbf{0} & 0 & \mathbf{0} & \underline{\mathbf{1}} & -2 & -4 & -12 \end{array}\right] \tag{1.55}$$

after which no further reduction/simplification can be made.

Recommended Exercises. Exercises 12, 16, 18, 20, and 28 in Section 1.1 and Exercises 7 and 14 (15 in [2]) in Section 1.6 in [3]. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems unless otherwise stated!

In this lecture, we finished Section 1.1 and worked through parts of Section 1.6 of [Lay]. We have also introduce a lot of terminology that you should get comfortable with. Understand each definition (in particular, if a definition is describing something, be sure to know *what* it is describing), have examples associated with each concept, and also have non-examples to know when definitions do not apply or when they fail.

Terminology checklist

linear system	
augmented matrix	
consistent	
inconsistent	
solution set	
equivalent (system)	
row operations	
flow in = flow out	
pivot	
pivot column	
free variable	
free variable column	
echelon form	
reduced echelon form	

2 Vectors and span

In our earlier examples of temperature on a rod and currents in a circuit, the arrays of numbers given by

$$\begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} \qquad \& \qquad \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} \tag{2.1}$$

are examples of *vectors* in $\mathbb{R}^4$ and $\mathbb{R}^3$, respectively. Here $\mathbb{R}$ is the set of real numbers and $\mathbb{R}^n$ is the set of n-tuples of real numbers where n is a positive integer being one of $1, 2, 3, 4, \ldots,$

$$\mathbb{R}^n := \big\{(x_1, \ldots, x_n) \ : \ x_i \in \mathbb{R} \quad \forall\, i = 1, \ldots, n\big\}. \tag{2.2}$$

Given two vectors in $\mathbb{R}^n$,

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \qquad \& \qquad \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \tag{2.3}$$

we can take their sum defined by

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} := \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix}. \tag{2.4}$$

We can also scale each vector by any number c in $\mathbb{R}$ by

$$c \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} := \begin{bmatrix} ca_1 \\ \vdots \\ ca_n \end{bmatrix}. \tag{2.5}$$

The above descriptions of vectors are algebraic and we've illustrated their algebraic structures (addition and scaling). Vectors can also be visualized when $n = 1, 2, 3$. Vectors are more than just points in space. For example, a billiard ball on an infinite pool table has a well-defined position.

•

You see it. It's right there. When it moves in one instant of time, the *difference* from the final position to the initial position provides us with a *length together with a direction.*

This is often called the *displacement*. Thus, to define vectors, a reference point must be specified. In the above example, the reference point is the initial position. A fixed observer (one that does not move in time) can also act as a reference point. This provides any other point with a length and a direction.

In many applications, the reference point will be called "zero" because often, the numerical values of the entries can be taken to be 0. One such example is the vector of temperatures, currents, traffic flows, etc.. Notice that the choice of 0 is really just a reference point and is not always universal in any sense. For example, for temperatures, we can choose 0 to be 0 in Celcius. In this case, 0 in Fahrenheit would actually be about -18 degrees Celcius, which is not 0. If we didn't specify which units we were using for temperature, 0 would be ambiguous, and we would not be able to define a temperature vector as in our previous example. Hence, to define vectors, we need to specify such a reference point. This is why we can define a vector as a tuple of numbers—this definition assumes we have already specified this reference point. We will often write vectors with an arrow over them as in $\vec{a}$ and $\vec{b}$ when n, the number of entries of said vector, is understood.

Definition 2.6. Let $S := \{\vec{v}_1, \ldots, \vec{v}_m\}$ be a set of m vectors in $\mathbb{R}^n$. The *span* of S is the set of all vectors of the form[5]

$$\sum_{i=1}^{m} a_i \vec{v}_i \equiv a_1 \vec{v}_1 + \cdots a_m \vec{v}_m, \tag{2.7}$$

where the a_i can be any real number. For a fixed set of a_i, the right-hand-side of (2.7) is called a *linear combination* of the vectors $\vec{v}_i$.

[5]Please do *not* confuse the notation $\vec{v}_i$ with the *components* of the vector $\vec{v}_i$. It can be confusing with these indices, but to be very clear, we could write the components of the vector $\vec{v}_i$ as

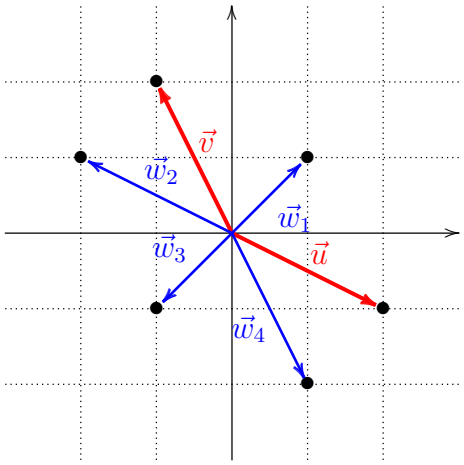
$$\begin{bmatrix} (v_i)_1 \\ \vdots \\ (v_i)_n \end{bmatrix}.$$

In set-theoretic notation, we would write the span in this definition as

$$\operatorname{span}(S) := \left\{ \sum_{i=1}^{m} a_i \vec{v}_i \in \mathbb{R}^n \ : \ a_1, \ldots, a_m \in \mathbb{R} \right\}. \tag{2.8}$$

The span of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ can be visualized quite nicely.

Problem 2.9. In the following figure, vectors $\vec{u}, \vec{v}, \vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4$ are depicted with a grid showing unit markings.



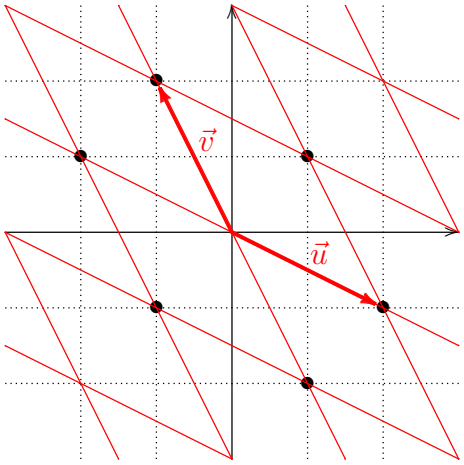
(2.10)

What linear combinations of $\vec{u}$ and $\vec{v}$ will produce the other bullets drawn in the graph?

Answer. To answer this question, it helps to draw the integral grid associated to the vectors $\vec{u}$ and $\vec{v}$. This is the set of linear combinations

$$a\vec{u} + b\vec{v} \tag{2.11}$$

such that $a, b \in \mathbb{Z}$, i.e. a and b are both integers. The *intersections* of the red lines in the following image depict these integral linear combinations.

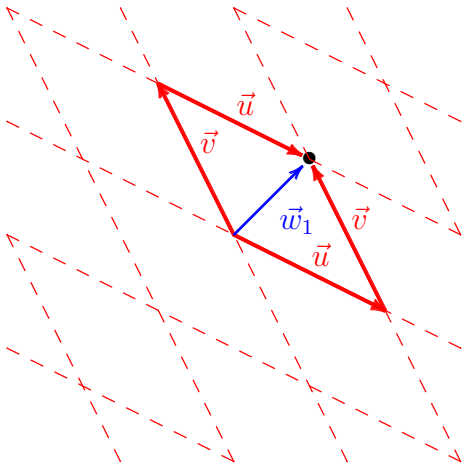


(2.12)

As you can see, the bullets lie exactly on these intersections. Hence, we should be able to find integers $a_i, b_i \in \mathbb{Z}$ such that

$$\vec{w}_i = a_i\vec{u} + b_i\vec{v} \tag{2.13}$$

for all $i = 1, 2, 3, 4$. For example, for $\vec{w}_1$, moving once along $\vec{u}$ and then $\vec{v}$ (or in the other order) gets us to $\vec{w}_1$

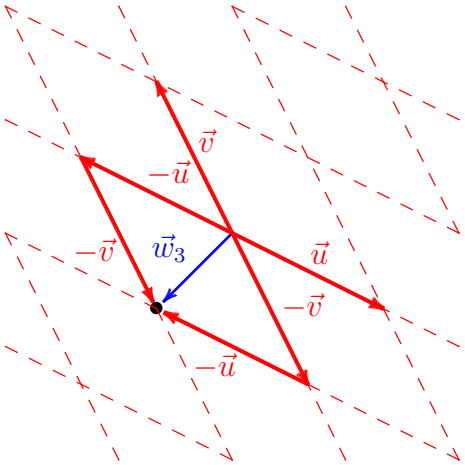


$$\vec{w}_1 = \vec{u} + \vec{v} \tag{2.14}$$

so that $a_1 = 1$ and $b_1 = 1$. $\vec{w}_2$ and $\vec{w}_4$ are relatively simple to see because they are just the vectors $\vec{u}$ and $\vec{v}$ flipped, i.e.

$$\vec{w}_2 = -\vec{u} \qquad \& \qquad \vec{w}_4 = -\vec{v}. \tag{2.15}$$

For $\vec{w}_3$, we illustrate the combinations



$$\vec{w}_3 = -\vec{u} - \vec{v} \tag{2.16}$$

The intersections of the red grid only depict integral linear combinations, but by scaling these vectors by any real number, the entire plane can be filled.

Problem 2.17. In the previous example, show that *every* vector

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{2.18}$$

can be written as a linear combination of $\vec{u}$ and $\vec{v}$. Thus $\{\vec{u}, \vec{v}\}$ spans $\mathbb{R}^2$.

Answer. To see this, note that

$$\vec{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad \& \qquad \vec{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}. \tag{2.19}$$

To prove the claim, we must find real numbers a_1 and a_2 such that

$$a_1\vec{u} + a_2\vec{v} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{2.20}$$

But the left-hand-side is given by

$$a_1\vec{u} + a_2\vec{v} = a_1 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + a_2 \begin{bmatrix} -1 \\ 2 \end{bmatrix} \overset{(2.5)}{=} \begin{bmatrix} 2a_1 \\ -a_1 \end{bmatrix} + \begin{bmatrix} -a_2 \\ 2a_2 \end{bmatrix} \overset{(2.4)}{=} \begin{bmatrix} 2a_1 - a_2 \\ -a_1 + 2a_2 \end{bmatrix}. \tag{2.21}$$

Therefore, we need to solve the linear system of equations given by

$$\begin{aligned} 2a_1 - a_2 &= b_1 \\ -a_1 + 2a_2 &= b_2, \end{aligned} \tag{2.22}$$

which should by now be a familiar procedure. Put it in augmented matrix form

$$\left[\begin{array}{cc|c} 2 & -1 & b_1 \\ -1 & 2 & b_2 \end{array}\right] \tag{2.23}$$

Permute the first and second rows

$$\left[\begin{array}{cc|c} -1 & 2 & b_2 \\ 2 & -1 & b_1 \end{array}\right] \tag{2.24}$$

Add two of row 1 to row 2 to get

$$\left[\begin{array}{cc|c} -1 & 2 & b_2 \\ 0 & 3 & b_1 + 2b_2 \end{array}\right] \tag{2.25}$$

This is now in echelon form. Multiply row 1 by -1 and divide row 2 by 3

$$\left[\begin{array}{cc|c} 1 & -2 & -b_2 \\ 0 & 1 & \frac{1}{3}(b_1 + 2b_2) \end{array}\right] \tag{2.26}$$

Add 2 of row 2 to row 1

$$\left[\begin{array}{cc|c} 1 & 0 & -b_2 + \frac{2}{3}(b_1 + 2b_2) \\ 0 & 1 & \frac{1}{3}(b_1 + 2b_2) \end{array}\right] \tag{2.27}$$

which is equal to

$$\left[\begin{array}{cc|c} 1 & 0 & \frac{1}{3}(2b_1 + b_2) \\ 0 & 1 & \frac{1}{3}(b_1 + 2b_2) \end{array}\right] \tag{2.28}$$

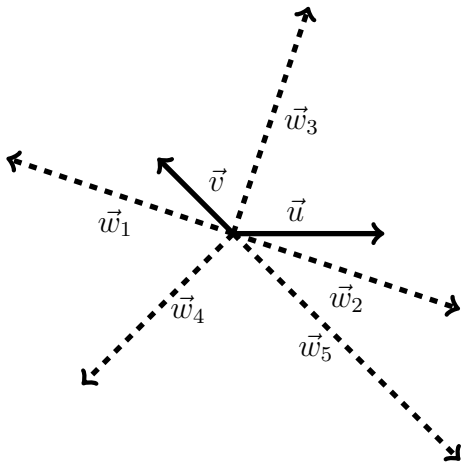
which says that

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \left(\frac{2b_1 + b_2}{3}\right)\vec{u} + \left(\frac{b_1 + 2b_2}{3}\right)\vec{v}. \tag{2.29}$$

For example,

$$\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \frac{4}{3}\vec{u} + \frac{2}{3}\vec{v}. \tag{2.30}$$

Exercise 2.31. Vectors $\vec{u}$ and $\vec{v}$ are drawn (as solid arrows) with the center point being the origin in $\mathbb{R}^2$. Express the vectors $\vec{w}_1, \vec{w}_2$, and $\vec{w}_3$ (drawn as dashed arrows) as linear combinations of $\vec{u}$ and $\vec{v}$.



$\vec{w}_1 = \_\_\_\_\_\vec{u} + \_\_\_\_\_\vec{v}$

$\vec{w}_2 = \_\_\_\_\_\vec{u} + \_\_\_\_\_\vec{v}$

$\vec{w}_3 = \_\_\_\_\_\vec{u} + \_\_\_\_\_\vec{v}$

$\vec{w}_4 = \_\_\_\_\_\vec{u} + \_\_\_\_\_\vec{v}$

$\vec{w}_5 = \_\_\_\_\_\vec{u} + \_\_\_\_\_\vec{v}$

You may have noticed that to express vectors as linear combinations of other vectors, we never needed to add any numbers together. All that was needed was that all of the vectors had the same reference point, i.e. origin. Once an origin is chosen, the span of the vectors given is the set of all linear combinations of the vectors, where linear combinations in this context means that we can scale the vectors and add them by putting them together "from head to toe." This gives us a geometric meaning of the concept of span for vectors that have the *same* origin. We cannot make sense of vector addition when the vectors do not have the same reference point. For example, if a train is traveling northwest towards Chicago at 50 mph and a completely different train is traveling northeast to Boston at 70 mph, then it doesn't make sense to add their velocities. However, if I decided to ride a skateboard through the halls of the train going towards Chicago at 10 mph towards the front of the train, I can put the velocities together in several possible ways. With respect to a reference point given by someone waiting outside at a station, I would appear to be moving at $50 + 10 = 60$ mph northwest. With respect to someone on the Chicago train, I'm just moving 10 mph northwest. From my perspective, if I didn't notice any friction or air resistance, I would think I'm not moving at all! Instead, the objects in the train would be moving at 10 mph behind me and the world outside the train would be moving at 60 mph behind me.

Recommended Exercises. Exercises 12, 16, 23, and 31 in Section 1.2 of [3]. Exercises 8, 25, 26, and 28 (you may use a calculator for exercise 28) in Section 1.3 of [3]. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems unless otherwise stated!

In this lecture, we finished Sections 1.2 and 1.3 of [Lay].

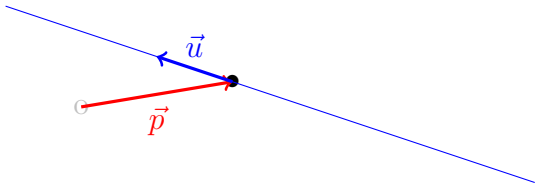
Terminology checklist

vector	
set notation $\{, :\}, \in, \forall, \mathbb{R}, \mathbb{Z}$ etc.	
linear combination	
span	

3 Solution sets of linear systems

We will skip Section 1.4 in [Lay] for now. We just discussed the span of vectors, but let's review it and discuss the relationship between the span and the solution set of a linear system.

Problem 3.1. In Example 1.28, we graphed two planes and their intersection, which was the set of solutions to the corresponding linear system. This intersection was a line. This line is spanned by a vector in the sense that if $\vec{p}$ is a vector on the line chosen as some reference point and $\vec{u}$ is a vector that points in a direction along the line whose origin is $\vec{p}$.



Give one example of a vector that spans the solution set of Example 1.28. In particular, give a reference point (origin) for this vector?

Answer. The set of solutions was given by

$$\left\{ \left(t, \frac{1}{2}(3+t), \frac{1}{2}(3t-1) \right) \in \mathbb{R}^3 \ : \ t \in \mathbb{R} \right\}. \tag{3.2}$$

We have used the variable t only because we will interpret it as time. Using the notation of vectors written vertically, this looks like

$$\left\{ \begin{bmatrix} t \\ \frac{1}{2}(3+t) \\ \frac{1}{2}(3t-1) \end{bmatrix} \in \mathbb{R}^3 \ : \ t \in \mathbb{R} \right\}. \tag{3.3}$$

We can split any vector in this set into a constant vector plus a vector multiplied by (the common factor) t as

$$\begin{bmatrix} t \\ \frac{1}{2}(3+t) \\ \frac{1}{2}(3t-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 3/2 \\ -1/2 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1/2 \\ 3/2 \end{bmatrix}. \tag{3.4}$$

As t varies over the set of real numbers, this traces out a straight line. This describes the solution to (1.29) in *parametric form*. This line coincides with the "span" of the vector

$$\begin{bmatrix} 1 \\ 1/2 \\ 3/2 \end{bmatrix} \tag{3.5}$$

whose origin is at

$$\begin{bmatrix} 0 \\ 3/2 \\ -1/2 \end{bmatrix}. \tag{3.6}$$

Again, the reason "span" is in quotes is because this is not exactly the span of the vector if we wrote its coordinates. It is the span of that vector when viewed as origin starting at $\vec{p}$ instead of $\vec{0}$. We will learn soon that this is technically the *affine span* of two vectors along the line. More precisely, the line in $\mathbb{R}^3$ described by the set of vectors of the form

$$\left\{ \begin{bmatrix} 0 \\ 3/2 \\ -1/2 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1/2 \\ 3/2 \end{bmatrix} \in \mathbb{R}^3 \ : \ t \in \mathbb{R} \right\} \tag{3.7}$$

can also be written as

$$\left\{ (1-t) \begin{bmatrix} 0 \\ 3/2 \\ -1/2 \end{bmatrix} + t \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \in \mathbb{R}^3 \ : \ t \in \mathbb{R} \right\}. \tag{3.8}$$

The latter expresses the line as the affine span of the vectors $\begin{bmatrix} 0 \\ 3/2 \\ -1/2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ because it describes the straight line through these two vectors. More about this will be discussed in a few sections.

Problem 3.9. Find two vectors with the same origin so that they span the solution set of the linear system

$$-x - y + z = -2 \tag{3.10}$$

with respect to that origin.

Answer. Since z can be solved in terms of x and y via $z = x + y - 2$, the set of solutions is given by

$$\left\{ \begin{bmatrix} x \\ y \\ x+y-2 \end{bmatrix} \in \mathbb{R}^3 \ : \ x, y \in \mathbb{R} \right\}. \tag{3.11}$$

Each such vector can be expressed as

$$\begin{bmatrix} x \\ y \\ x+y-2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix} + x \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}. \tag{3.12}$$

The origin can therefore be taken as

$$\begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}. \tag{3.13}$$

Since x and y can vary over the set of real numbers, two vectors that span this solution set are

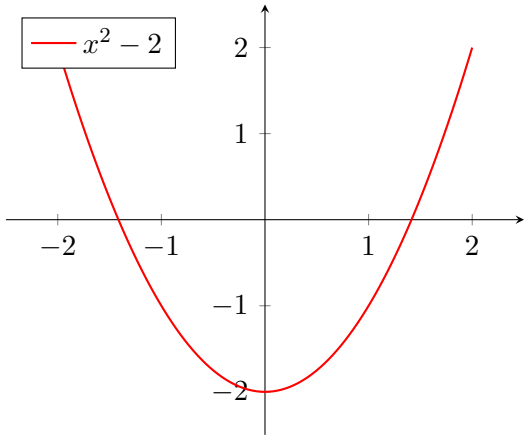
$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \& \qquad \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}. \tag{3.14}$$

Proposition 3.15. *Using the notation from Definition 1.26, if*

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad \text{and} \qquad \vec{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \tag{3.16}$$

are both solutions of the linear system (1.27), then every point on the straight line passing through both the vectors $\vec{x}$ and $\vec{y}$ is a solution to (1.27).

This result is surprising! In particular, it says that if we have two distinct solutions of a linear system, then we automatically have *infinitely many solutions*! To see that this fails for non-linear systems, consider the quadratic polynomial of the form $x^2 - 2$ (with x taking values in $\mathbb{R}$)



The two solutions are $y := \sqrt{2}$ and $z := -\sqrt{2}$ but there are no other solutions at all!

Proof. The straight line passing through $\vec{y}$ and $\vec{z}$ can be described parametrically as[6]

$$\mathbb{R} \ni t \mapsto (1-t)\vec{y} + t\vec{z} = \begin{bmatrix} (1-t)y_1 + tz_1 \\ \vdots \\ (1-t)y_n + tz_n \end{bmatrix}. \tag{3.17}$$

We have to show that each point on this line is a solution. It suffices to show this for the i-th equation in (1.27) for any $i \in \{1, \ldots, m\}$. Plugging in a point along the straight line, we get

$$\begin{aligned} a_{i1}\Big((1-t)y_1 + tz_1\Big) + \cdots + a_{in}\Big((1-t)y_n + tz_n\Big) &= (1-t)\Big(a_{i1}y_1 + \cdots + a_{in}y_n\Big) \\ &\quad + t\Big(a_{i1}z_1 + \cdots + a_{in}z_n\Big) \\ &= (1-t)b_i + tb_i \\ &= b_i \end{aligned} \tag{3.18}$$

using the distributive law (among other properties) for adding and multiplying real numbers. ∎

[6] You might have written down a different formula, but the line you get should be the same as the one we get here.

Problem 3.19 (Exercise 1.6.8 in [Lay])**.** Consider a chemical reaction that turns limestone $CaCO_3$ and acid H_3O into water H_2O, calcium Ca, and carbon dioxide CO_2. In a chemical reaction, all elements must be accounted for. Find the appropriate ratios of these compounds and elements needed for this reaction to occur without other waste products.

Answer. Introduce variables x_1, x_2, x_3, x_4 and x_5 for the coefficients of limestone, acid, water, calcium, and carbon dioxide, respectively. The elements appearing in these compounds and elements are H, O, C, and Ca. We can therefore write the compounds as a vector in these variables (in this order). For example, limestone, $CaCO_3$, is

$$\begin{bmatrix} 0 \\ 3 \\ 1 \\ 1 \end{bmatrix} \begin{matrix} \leftarrow \mathrm{H} \\ \leftarrow \mathrm{O} \\ \leftarrow \mathrm{C} \\ \leftarrow \mathrm{Ca} \end{matrix} \tag{3.20}$$

since it is composed of zero hydrogen atoms, three oxygen atoms, one carbon atom, and one calcium atom. Thus, the linear system we need to solve is given by

$$\begin{gathered} x_1\mathrm{CaCO_3} + x_2\mathrm{H_3O} = x_3\mathrm{H_2O} + x_4\mathrm{Ca} + x_5\mathrm{CO_2} \\ x_1 \begin{bmatrix} 0 \\ 3 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} = x_3 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} \end{gathered} \tag{3.21}$$

The associated augmented matrix together with the row reduction procedure is

$$\left[\begin{array}{ccccc|c} 0 & 3 & -2 & 0 & 0 & 0 \\ 3 & 1 & -1 & 0 & -2 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \end{array}\right] \xrightarrow[R2 \mapsto R2 - 3R4]{R3 \mapsto R3 - R4} \left[\begin{array}{ccccc|c} 0 & 3 & -2 & 0 & 0 & 0 \\ 0 & 1 & -1 & 3 & -2 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \end{array}\right]$$

$$\downarrow R1 \mapsto R1 - 3R2$$

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 3 & -2 & 0 \\ 0 & 0 & 1 & -9 & 6 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{array}\right] \xleftarrow{\text{permute}} \left[\begin{array}{ccccc|c} 0 & 0 & 1 & -9 & 6 & 0 \\ 0 & 1 & -1 & 3 & -2 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \end{array}\right]$$

$$\downarrow \begin{matrix} R1 \mapsto R1 + R4 \\ R2 \mapsto R2 + R3 \end{matrix}$$

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -6 & 4 & 0 \\ 0 & 0 & 1 & -9 & 6 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{array}\right] \xrightarrow[R2 \mapsto R2 + 6R4]{R3 \mapsto R3 + 9R4} \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & -3 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{array}\right]$$

Now the augmented matrix is in reduced echelon form. Notice that although solutions exist, they are not unique! We saw this happening in Example 1.28. Let us write the concentrations in terms of x_5, the concentration of calcium (this choice is somewhat arbitrary).

$$x_1 = x_5, \qquad x_2 = 2x_5, \qquad x_3 = 3x_5, \qquad x_4 = x_5, \qquad x_5 \text{ free.} \tag{3.22}$$

Thus, the resulting reaction is given by

$$x_5 CaCO_3 + 2x_5 H_3O \to 3x_5 H_2O + x_5 Ca + x_5 CO_2 \tag{3.23}$$

It is common to set the smallest quantity to 1 so that this becomes

$$CaCO_3 + 2H_3O \to 3H_2O + Ca + CO_2. \tag{3.24}$$

Nevertheless, we do not have to do this, and a proper way to express the solution is in parametric form in terms of the concentration of calcium (for instance) as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_5 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 1 \end{bmatrix}. \tag{3.25}$$

We did not have to choose calcium as the free variable. Any of the other elements would have been as good of a choice as any other, but in some instances, the resulting coefficients might be fractions.

The previous example leads us to the notion of *homogeneous linear systems*. For brevity, instead of writing the linear system (1.27) over and over again, we use the shorthand *notation* (for now, it is only notation!)

$$A\vec{x} = \vec{b}. \tag{3.26}$$

Definition 3.27. A linear system $A\vec{x} = \vec{b}$ is said to be *homogeneous* whenever $\vec{b} = \vec{0}$.

Note that a homogeneous linear system always has at least one solution (as we saw in Problem 1.36), namely $\vec{x} = \vec{0}$, which is called the *trivial solution*. For example, the trivial solution to Example 1.1 signifies that there are no cars going down any of the roads, a perfectly valid (yet perhaps suspicious) solution. We have also noticed in the example that there is a free variable in the solution. This is a generic phenomena:

Theorem 3.28. *The homogeneous equation $A\vec{x} = \vec{0}$ has a nontrivial solution if and only if* [7] *the corresponding system of linear equations has a free variable.*

[7] To prove a statement of the form "A if and only if B," one must show that A implies B and B implies A. In a proof, we often depict the former by ($\Rightarrow$) and the latter by ($\Leftarrow$).

Proof.
($\Rightarrow$) Let $\vec{x}$ be a non-zero vector (i.e. a non-trivial solution) that satisfies $A\vec{x} = \vec{0}$. Since $\vec{0}$ is also a solution, this system has two distinct solutions. By Proposition 3.15, all the points along the straight line

$$\mathbb{R} \ni t \mapsto (1-t)\vec{0} + t\vec{x} = t\vec{x} \tag{3.29}$$

are solutions as well. Since $\vec{x}$ is not zero, at least one of the components is non-zero. Suppose it is x_i for some $i \in \{1, \ldots, n\}$. Setting $t := \frac{1}{x_i}$ shows that the vector

$$\begin{bmatrix} x_1/x_i \\ \vdots \\ x_{i-1}/x_i \\ 1 \\ x_{i+1}/x_i \\ \vdots \\ x_n/x_i \end{bmatrix} \tag{3.30}$$

is also a solution. Hence, any constant multiple of this vector is also a solution. In particular, x_i can be taken to be a free variable for the linear system since

$$\begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} = x_i \begin{bmatrix} x_1/x_i \\ \vdots \\ 1 \\ \vdots \\ x_n/x_i \end{bmatrix}. \tag{3.31}$$

($\Leftarrow$) Suppose that x_i is a free variable. Setting $x_i = 1$ and all other free variables (if they exist) to zero gives a non-trivial solution to $A\vec{x} = \vec{0}$. ■

In (3.25), the solution of the homogeneous equation was written in the form

$$\vec{x} = \vec{p} + t\vec{v} \tag{3.32}$$

where in that example $\vec{p}$ was $\vec{0}$, t was x_5, and $\vec{v}$ was the vector

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 1 \end{bmatrix}. \tag{3.33}$$

This form of the solution of a linear equation is also in parametric form because its value depends on an additional unspecified parameter, which in this case is t. In other words, all solutions are valid as t varies over the real numbers. For a homogeneous equation, $\vec{p}$ is always $\vec{0}$. In fact, there could be more than one such parameter involved.

Theorem 3.34. *Suppose that the linear system described by $A\vec{x} = \vec{b}$ is consistent and let $\vec{x} = \vec{p}$ be a solution. Then the solution set of $A\vec{x} = \vec{b}$ is the set of all vectors of the form $\vec{p} + \vec{u}$ where $\vec{u}$ is any solution of the homogeneous equation $A\vec{u} = \vec{0}$.*

Proof. Let S be the solution set of $A\vec{x} = \vec{b}$, i.e.

$$S := \left\{ \vec{x} \in \mathbb{R}^n \ : \ A\vec{x} = \vec{b} \right\}. \tag{3.35}$$

The claim that we must prove is that for some *fixed* (also known as *particular*) solution $\vec{p} \in \mathbb{R}^n$ satisfying $A\vec{p} = \vec{b}$,

$$S = \left\{ \vec{p} + \vec{u} \in \mathbb{R}^n \ : \ A\vec{u} = \vec{0} \right\}. \tag{3.36}$$

Let's call the right-hand-side of (3.36) T. To prove that two sets are equal, $S = T$, we must show that each one is contained in the other. First let us show that T is contained in S, which is written mathematically as $T \subseteq S$. To prove this, let $\vec{x} := \vec{p} + \vec{u} \in T$ so that $\vec{p}$ satisfies $A\vec{p} = \vec{b}$ and $\vec{u}$ satisfies $A\vec{u} = \vec{0}$. By a similar calculation as in (3.18), we see that $A\vec{x} = \vec{0}$ (I'm leaving this calculation to you as an exercise). This shows that $\vec{x} \in S$ so that $T \subseteq S$ (because we showed that any *arbitrary* element in T is in S).

Now let $\vec{x} \in S$. This means that $A\vec{x} = \vec{b}$. Our goal is to find a $\vec{u}$ that satisfies the two conditions

(a) $A\vec{u} = \vec{0}$ and

(b) $\vec{x} = \vec{p} + \vec{u}$.

This would prove that $\vec{x} \in T$. Let's therefore define $\vec{u}$ to be $\vec{u} := \vec{x} - \vec{p}$. Then, by a similar calculation as in (3.18), we see that $A\vec{u} = \vec{0}$ (exercise!). Also, from this definition, it immediately follows that $\vec{p} + \vec{u} = \vec{p} + \left(\vec{x} - \vec{p}\right) = \vec{x}$. Hence, $S \subseteq T$.

Together, $T \subseteq S$ and $S \subseteq T$ prove that $S = T$. ∎

This theorem says that the solution set of a consistent linear system $A\vec{x} = \vec{b}$ can be expressed as

$$\vec{x} = \vec{p} + t_1\vec{u}_1 + \cdots + t_k\vec{u}_k, \tag{3.37}$$

where $\vec{p}$ is *one* solution of $A\vec{x} = \vec{b}$, k is a positive integer, $\{t_1, \ldots, t_k\}$ are the parameters (real numbers), and the set $\{\vec{u}_1, \ldots, \vec{u}_k\}$ spans the solution set of $A\vec{x} = \vec{0}$. A linear combination of solutions to $A\vec{x} = \vec{0}$ is a solution as well. Here's an application of the theorem.

Problem 3.38. Consider the linear system

$$\begin{aligned} 2x_1 + 4x_2 - 2x_5 &= 2 \\ -x_1 - 2x_2 + x_3 - x_4 &= -1 \\ x_4 - x_5 &= 1 \\ x_3 - x_4 - x_5 &= 0 \end{aligned} \tag{3.39}$$

Check that

$$\vec{p} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{3.40}$$

is a solution to this linear system. Then, find all the solutions of this system.

Answer. I'll leave the check that $\vec{p}$ is a solution to you. To find *all* the solutions, all we need to do now is solve the *homogeneous* system

$$\begin{aligned} 2x_1 + 4x_2 - 2x_5 &= 0 \\ -x_1 - 2x_2 + x_3 - x_4 &= 0 \\ x_4 - x_5 &= 0 \\ x_3 - x_4 - x_5 &= 0 \end{aligned} \tag{3.41}$$

This is a little easier than solving the original system because we have fewer numbers to keep track of (and therefore have a less likely probability of making a mistake!). After row reduction, the augmented matrix becomes (exercise!)

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{3.42}$$

which is in reduced echelon form. The set of solutions here are all of the form

$$\vec{u} = \begin{bmatrix} x_5 - 2x_2 \\ x_2 \\ 2x_5 \\ x_5 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 1 \end{bmatrix} \tag{3.43}$$

with x_2, x_5 free. Hence, the set of solutions of the linear system consists of vectors of the form

$$\underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{\vec{p}} + \underbrace{x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 1 \end{bmatrix}}_{\vec{u}} \tag{3.44}$$

where $x_2, x_5 \in \mathbb{R}$ are the free variables.

Exercise 3.45. State whether the following claims are True or False. If the claim is true, be able to precisely deduce why the claim is true. If the claim is false, be able to provide an explicit counter-example.

(a) If there is a nonzero (aka nontrivial) solution to a linear homogeneous system, then there are infinitely many solutions.

(b) If there are infinitely many solutions to a linear system, then the system is homogeneous.

(c) $\vec{x} = \vec{0}$ is always a solution to every linear system $A\vec{x} = \vec{b}$.

We have introduced several kinds of notation for linear systems so far. All of the following are equivalent ways of describing the same linear system:

(a) as a collection of *linear equations*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \tag{3.46}$$

(b) as an *augmented matrix* (with the understanding that the columns are read from left to right in the order $x_1, x_2, \ldots, x_n$)

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right] \tag{3.47}$$

(c) as a single *vector* equation

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \tag{3.48}$$

(d) as a *matrix* equation

$$A\vec{x} = \vec{b}. \tag{3.49}$$

The meaning of this last version will make more sense when we discuss linear transformations and matrices.

Recommended Exercises. Exercises 15, 27, and 40 in Section 1.5 of [Lay]. Exercises 7 and 15 (this one is similar to the circuit problem from last class) in Section 1.6 of [Lay]. You may (and are encouraged to) use any theorems we have done in class! Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

Today, we finished Sections 1.5 and 1.6 of [Lay] (we skipped Section 1.4). Whenever you see the equation $A\vec{x} = \vec{b}$, just read it as the associated linear system as in (1.27). We will not provide an algebraic interpretation of the expression "$A\vec{x}$" until a few lectures from now.

Terminology checklist

parametric form	
affine span	
homogeneous linear system	
trivial solution	
homogeneous solution	
particular solution	

4 Linear independence and dimension of solution sets

The solution sets of Problems 3.1 and 3.9 are visually different, and we would like to say that the span of one nonzero vector is a 1-dimensional line and the span of the two vectors in (3.14) is a 2-dimensional plane. But we need to be a bit more precise about what we mean by dimension. To get there, we first introduce the notion of linear independence. Heuristically, a solution set is n-dimensional if the *minimum* number of vectors needed to span it is n. This would answer our earlier question when we did Example 1.1 for traffic flow. The dimensionality of the solution set corresponds to the minimal number of people needed to count traffic to obtain the full traffic flow for a given set of streets and intersections. *Where* we place those people is related to a choice of linearly independent vectors that span the set of solutions.

Definition 4.1. A set of vectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$ in $\mathbb{R}^n$ is *linearly independent* if the solution set of the vector equation

$$x_1\vec{u}_1 + \cdots + x_k\vec{u}_k = \vec{0} \tag{4.2}$$

consists of only the trivial solution. Otherwise, the set is said to be *linearly dependent* in which case there exist some coefficients $x_1, \ldots, x_k$, not all of which are zero, such that (4.2) holds.

Example 4.3. The vectors

$$\begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} \qquad \& \qquad \begin{bmatrix} -3 \\ 6 \\ 0 \end{bmatrix} \tag{4.4}$$

are linearly dependent because

$$\begin{bmatrix} -3 \\ 6 \\ 0 \end{bmatrix} = -3 \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} \tag{4.5}$$

so that

$$3 \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + \begin{bmatrix} -3 \\ 6 \\ 0 \end{bmatrix} = 0. \tag{4.6}$$

Example 4.7. The vectors

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \& \qquad \begin{bmatrix} -1 \\ 1 \end{bmatrix} \tag{4.8}$$

are linearly independent for the following reason. Let x_1 and x_2 be two real numbers such that

$$x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{4.9}$$

This equation describes the linear system associated to the augmented matrix

$$\left[\begin{array}{cc|c} 1 & -1 & 0 \\ 1 & 1 & 0 \end{array}\right]. \tag{4.10}$$

Performing row operations,

$$\left[\begin{array}{cc|c} 1 & -1 & 0 \\ 1 & 1 & 0 \end{array}\right] \xmapsto{\text{R2}\to\text{R2}-\text{R1}} \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 2 & 0 \end{array}\right] \xmapsto{\text{R2}\to\frac{1}{2}\text{R2}} \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 2 & 0 \end{array}\right] \xmapsto{\text{R1}\to\text{R1}+\text{R2}} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right]. \tag{4.11}$$

The only solution to (4.9) is therefore $x_1 = 0$ and $x_2 = 0$. Thus, the two vectors in (4.8) are linearly independent.

Example 4.12. A set $\{\vec{u}_1, \vec{u}_2\}$ of two vectors in $\mathbb{R}^m$ is linearly dependent if and only if[8] one can be written as a scalar multiple of the other, i.e. there exists a real number c such that $\vec{u}_1 = c\vec{u}_2$ or[9] $c\vec{u}_1 = \vec{u}_2$.[10]

Proof. First[11] note that the associated vector equation is of the form

$$x_1\vec{u}_1 + x_2\vec{u}_2 = \vec{0}, \tag{4.13}$$

where[12] x_1 and x_2 are coefficients, or upon rearranging

$$x_1\vec{u}_1 = -x_2\vec{u}_2. \tag{4.14}$$

($\Rightarrow$) If the set is linearly dependent, then x_1 and x_2 cannot both be zero.[13] Without loss of generality, suppose that x_1 is nonzero.[14] Then dividing both sides of (4.14) by x_1 gives

$$\vec{u}_1 = -\frac{x_2}{x_1}\vec{u}_2. \tag{4.15}$$

Thus, setting $c := -\frac{x_2}{x_1}$ proves the first claim[15] (a similar argument can be made if x_2 is nonzero). ($\Leftarrow$) Conversely,[16] suppose that there exists a real number c such that[17] $\vec{u}_1 = c\vec{u}_2$. Then

$$\vec{u}_1 - c\vec{u}_2 = \vec{0} \tag{4.16}$$

[8] If A and B are statements, the phrase "A if and only if B" means two things. First, it means "A implies B." Second, it means "B implies A."

[9] In mathematics, the word "or" is never exclusive. If "A or B" are true, it always means that "at least one of A or B is true." It does *not* mean that if A is true, then B is false, or vice versa. If A happens to be true, we make no additional assumptions about B (and vice versa).

[10] In what follows, we will work through the proof very closely. We will try to guide you using footnotes so that you know what is part of the proof and what is based on intuition. Instead of first teaching you how to do proofs from scratch, we will go through several examples so that you see what they are like first. This is like learning a new language. Before learning the grammar, you want to first listen to people talking to get a feel for what the language sounds like. Then, when you learn the alphabet, you want to read a few passages before you start constructing sentences on your own. The point is not to know/memorize proofs. The point is to know how to read, understand, and construct proofs of your own.

[11] Before proving anything, we just recall what the vector equation is to remind us of what we'll need to refer to.

[12] If you introduce notation in a proof, please say what it is every time!

[13] What we have done so far is just state the definition of what it means for $\{\vec{u}_1, \vec{u}_2\}$ to be linearly dependent. Stating these definitions to remind ourselves of what we know is a large part of the battle in constructing a proof.

[14] We know from the definition that at least one of x_1 or x_2 is not zero but we do not know which one. It won't matter which one we pick in the end (some insight is required to notice this), so we may use the phrase "without loss of generality" to cover all other possible cases.

[15] Remember, we wanted to show that $\vec{u}_1$ is a scalar multiple of $\vec{u}_2$.

[16] We say "conversely" when we want to prove an assertion in the opposite direction to the previously proven assertion.

[17] Remember, this is literally the latter assumption in the claim.

showing that the set $\{\vec{u}_1, \vec{u}_2\}$ is linearly dependent since the coefficient in front of $\vec{u}_1$ is nonzero (it is 1).[18] ■

Example 4.17. Let

$$\hat{x} := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \hat{y} := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \& \qquad \hat{z} := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4.18}$$

be the three unit vectors in $\mathbb{R}^3$. A lot of different notation is used for this, sometimes $\hat{i}$, $\hat{j}$, and $\hat{k}$, and sometimes $\vec{e}_1, \vec{e}_2$, and $\vec{e}_3$, respectively. In addition, let $\vec{u}$ be any other vector in $\mathbb{R}^3$. Then the set $\{\hat{x}, \hat{y}, \hat{z}, \vec{u}\}$ is linearly dependent because $\vec{u}$ can be written as a linear combination of the three unit vectors. This is apparent if we write

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \tag{4.19}$$

since

$$\vec{u} = u_1\hat{x} + u_2\hat{y} + u_3\hat{z}. \tag{4.20}$$

Here's a slightly more complicated example.

Example 4.21. The vectors

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \qquad \& \qquad \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix} \tag{4.22}$$

are linearly dependent. This is a little bit more difficult to see so let us try to solve it from scratch. We must find x_1, x_2, and x_3 such that

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.23}$$

Putting the left-hand-side together into a single vector gives us an equality of vectors

$$\begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_2 - 2x_3 \\ x_1 + 3x_2 - 3x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.24}$$

We therefore have to solve the linear system whose augmented matrix is given by

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 1 & 3 & -3 & 0 \end{array}\right] \tag{4.25}$$

[18] Recall the definition of what it means to be linearly dependent and confirm that you agree with the conclusion.

which after some row operations is equivalent to

$$\left[\begin{array}{ccc|c} 1 & 0 & 3 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]. \tag{4.26}$$

This has non-zero solutions. Setting $x_3 = -1$ (we don't have to do this—we can leave x_3 as a free variable, but I just want to show that we can write the last vector in terms of the first two) shows that

$$\begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}. \tag{4.27}$$

The previous examples hint at a more general situation.

Theorem 4.28. *Let $S := \{\vec{u}_1, \ldots, \vec{u}_k\}$ be a set of vectors in $\mathbb{R}^n$. S is linearly dependent if and only if at least one vector from S can be written as a linear combination of the others.*

The proof of Theorem 4.28 will be similar to the previous example. Why should we expect this? Well, if $k = 3$, then we have $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ and we could imagine doing something very similar. Think about this! If you're not comfortable working with arbitrary k just yet, specialize to the case $k = 3$ and try to mimic the previous proof. Then try $k = 4$. Do you see the pattern? Once you're ready, try the following.[19]

Proof. The vector equation associated to S is

$$\sum_{j=1}^{k} x_j \vec{u}_j = \vec{0}, \tag{4.29}$$

where the x_j are coefficients.
($\Rightarrow$) If the set S is linearly dependent, then there exists[20] a nonzero x_i (for some i between 1 and k). Therefore,

$$\vec{u}_i = \sum_{\substack{j=1 \\ j \neq i}}^{k} \left(-\frac{x_j}{x_i}\right) \vec{u}_j, \tag{4.30}$$

where the sum is over all numbers j from 1 to k except i. Hence, the vector $\vec{u}_i$ can be written as a linear combination of the others.
($\Leftarrow$) Conversely, suppose that there exists a vector $\vec{u}_i$ from S that can be written as a linear combination of the others, i.e.

$$\vec{u}_i = \sum_{\substack{j=1 \\ j \neq i}}^{k} y_j \vec{u}_j, \tag{4.31}$$

[19] If this is your first time proving things outside of geometry in highschool, study how these proofs are written. Try to prove things on your own. Do not be discouraged if you are wrong. Keep trying. A good book on learning how to think about proofs is *How to Solve It* by G. Polya [4]. A course in discrete mathematics also helps. Practice, practice, practice!

[20] By definition of a linearly dependent set, at least one of the x_i's must be nonzero. This is phrased concisely by the statement "there exists a nonzero x_i...".

where the y_j are real numbers.[21] Rearranging gives

$$\vec{u}_i - \sum_{j \neq i}^{k} x_j \vec{u}_j = 0, \tag{4.32}$$

and we see that the coefficient in front of $\vec{u}_i$ is nonzero (it is 1). Hence S is linearly dependent. ■

Let's give a simple application of this theorem.

Example 4.33. On a computer, colors can be obtained from choosing three integers from the set of numbers $\{0, 1, 2, \ldots, 255\}$. These three integers represent the level of red, green, and blue. If we denote these three colors as constructing a column "vector",[22] we can write

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.34}$$

$$\text{R} \longleftrightarrow \begin{bmatrix} 255 \\ 0 \\ 0 \end{bmatrix} \qquad \text{G} \longleftrightarrow \begin{bmatrix} 0 \\ 255 \\ 0 \end{bmatrix} \qquad \text{B} \longleftrightarrow \begin{bmatrix} 0 \\ 0 \\ 255 \end{bmatrix}$$

The $\leftrightarrow$ should be read as "corresponds to." Because there are 256 numbers allowed for each of these three colors, the total number of vectors allowed is

$$(256)^3 = 16,777,216. \tag{4.35}$$

8-bit computer displays[23] work using these colors. Therefore, each pixel on your computer has this many possibilities. Multiply that by the number of pixels on your computer display. That's a lot of information. These colors are all obtained from linear combinations of the form

$$x_R \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_G \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_B \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4.36}$$

with $x_R, x_G, x_B \in \{0, 1, 2, 3, \ldots, 255\}$. For example,

$$\text{Y} \longleftrightarrow \begin{bmatrix} 255 \\ 255 \\ 0 \end{bmatrix} = 255 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 255 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \longleftrightarrow \text{R} + \text{G}$$

[21] We call our variables y to avoid potentially confusing them with the previous variables x.

[22] Technically, these are not vectors. They are just arrays. The reason these are not vectors is because we cannot scale these arrays by an arbitrary number because the maximum value of any entry is 255. Similarly, we cannot add combinations of colors arbitrarily because of the maximum and minimum values allowed. Nevertheless, this example describes the content of the previous theorem with hopefully something you can relate to.

[23] I need a reference for this.

$$\text{M} \longleftrightarrow \begin{bmatrix} 255 \\ 0 \\ 255 \end{bmatrix} = 255 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 255 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \longleftrightarrow \text{R} + \text{B}$$

$$\text{C} \longleftrightarrow \begin{bmatrix} 0 \\ 255 \\ 255 \end{bmatrix} = 255 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 255 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \longleftrightarrow \text{G} + \text{B}$$

The colors R ,G ,and B are linearly independent in the sense of the above definition, namely the only solution to

$$x_R \text{R} + x_G \text{G} + x_B \text{B} = \text{Black} \tag{4.37}$$

is

$$x_R = x_G = x_B = 0. \tag{4.38}$$

Using the previous theorem, another way of saying this is that none of the colors R, G, and B can be expressed in terms of the other two as linear combinations. What about the colors M, R, and B ? Are these linearly independent? Or can we express any of these colors in terms of the others? Well, we already know we can express M in terms of R and B so the three are not linearly independent—they are linearly dependent. However, the colors Y, M, and C are linearly independent—none of these colors can be expressed in terms of the others.

Problem 4.39. Is the set of vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} \tag{4.40}$$

linearly independent? Explain your answer.

Answer. To answer this question, we need to solve the system

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.41}$$

for the variables x_1, x_2, x_3, x_4. Putting this into an augmented matrix gives

$$\left[\begin{array}{cccc|c} 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 \end{array}\right] \xrightarrow{R3 \mapsto R3 - R1} \left[\begin{array}{cccc|c} 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & -2 & 1 & 0 & 0 \end{array}\right] \xrightarrow{R3 \mapsto R3 - 2R2} \left[\begin{array}{cccc|c} 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 3 & 2 & 0 \end{array}\right] \tag{4.42}$$

This augmented matrix is in echelon form, it is consistent, and it has 3 pivots and 1 free variable. Therefore, there exists more than one solution, and the vectors in the question are linearly dependent.

The following two theorems will give quick methods to figure out whether a given set of vectors is linearly dependent.

Theorem 4.43. *Let $S := \{\vec{u}_1, \ldots, \vec{u}_k\}$ be a set of vectors in $\mathbb{R}^n$ with $k > n$. Then S is linearly dependent.*

Proof. Recall, S is linearly dependent[24] if there exist numbers $x_1, \ldots, x_k$ not all zero such that

$$\sum_{i=1}^{k} x_i \vec{u}_i = \vec{0}. \tag{4.44}$$

This equation can be expressed as a linear system

$$\begin{aligned} \sum_{i=1}^{k} x_i (u_i)_1 &= 0 \\ &\vdots \\ \sum_{i=1}^{k} x_i (u_i)_n &= 0, \end{aligned} \tag{4.45}$$

where[25] $(u_i)_j$ is the j-th component of the vector $\vec{u}_i$. In this linear system, there are k unknowns given by the variables $x_1, \ldots, x_k$ and there are n equations. Because $k > n$, there are more unknowns than equations, and hence there is at least one free variable.[26] Let x_p be one of these free variables. Then the other x_i's might depend on x_p so we may write $x_i(x_p)$.[27] Then by setting $x_p = 1$, we find

$$1\vec{u}_p + \sum_{\substack{i=1 \\ i \neq p}}^{k} x_i(x_p = 1)\vec{u}_i = \vec{0} \tag{4.46}$$

showing that S is linearly dependent (again since the coefficient in front of $\vec{u}_p$ is nonzero). ■

Warning! Using an *example* of $S := \{\vec{u}_1, \ldots, \vec{u}_k\}$ and showing that it is linearly dependent is not a proof! We have to prove the claim for *all* potential cases. Nevertheless, an example *helps* to see why the claim might be true in the first place.

Another warning! The theorem does *not* say that if $k < n$, then the set is linearly independent! This is an important distinction, one that comes from logic. If A is a statement and B is a statement, then the claim "A implies B" does *not* imply that "not A implies not B." Also, "A implies B" does not imply "B implies A." However, "A implies B" *does* imply "not B implies not A." To see this in a concrete example, suppose the manager of some company is in charge of giving

[24] Again, it is always helpful to constantly remind yourself and the reader of definitions that are crucial to solving the problem at hand. It is also helpful to use them to introduce notation that has not been introduced in the statement of the claim (the theorem).

[25] We have introduced some notation, so we should define it.

[26] But wait, how do we know that a solution even *exists*? If a solution doesn't exist, then our conclusion must be false! Thankfully, by our earlier comments from the previous lecture, we know that every homogeneous linear system has at least one solution, namely the trivial solution. Hence, the solution set is not empty.

[27] This is read as "x_i is a function of x_p."

his workers a raise, particularly to those who do not make so much money. If a worker's salary is less than \$50,000 a year (A), the manager will give them a raise (B). Does this statement imply that if a worker's salary is greater than \$50,000 a year (not A), then that worker will *not* get a raise (not B)? No, it doesn't. We don't know what happens in this situation. Similarly, if a worker received a raise (B), does this mean that the worker must have made less than \$50,000 a year (A)? No, it doesn't mean that either. If a worker does *not* receive a raise (not B), then what do we know? We know that this guarantees that the worker in question could not have a salary that is less than \$50,000 a year because otherwise that worker would get a raise! Hence, the worker must make more than \$50,000 a year (not B).

Theorem 4.47. *Let $S := \{\vec{u}_1, \ldots, \vec{u}_k\}$ be a set of vectors in $\mathbb{R}^n$ with at least one of the $\vec{u}_i$ being zero. Then S is linearly dependent.*

Proof. Suppose $\vec{u}_i = \vec{0}$. Then choose[28] the coefficient of $\vec{u}_j$ to be

$$x_j := \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \tag{4.48}$$

Then

$$\sum_{j=1}^{k} x_j \vec{u}_j = 1\vec{u}_i = 1(\vec{0}) = \vec{0} \tag{4.49}$$

because any scalar multiple of the zero vector is the zero vector. Since not all of the coefficients are zero (one of them is 1), S is linearly dependent. ■

Another way to see this result is to imagine $\vec{0}$ was one of the vectors, such as

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}. \tag{4.50}$$

The augmented matrix associated to finding out if these vectors are linearly independent or not is

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]. \tag{4.51}$$

The second column is all zeros and will therefore always correspond to a free variable. Because the system is consistent, there are infinitely many solutions so that the above vectors are linearly dependent.

Definition 4.52. Let $A\vec{x} = \vec{b}$ be a consistent linear system with a particular solution $\vec{p}$. The *dimension* of the solution set of $A\vec{x} = \vec{b}$ is the number $k \in \mathbb{N}$ of linearly independent homogeneous solutions $\{\vec{u}_1, \ldots, \vec{u}_k\}$ needed so that the solution set consists of all vectors of the form $\vec{p} + t_1\vec{u}_1 + \cdots + t_k\vec{u}_k$ with $t_1, \ldots, t_k \in \mathbb{R}$.

[28] To show that the set is linearly dependent, we have to find a set of coefficients, not all of which are zero, so that their linear combination results in the zero vector. The coefficients that I've chosen here are not the only coefficients that will work. You may choose have chosen others. All we have to do is exhibit the *existence* of one such choice. We do not have to exhaust all posibilities.

Intuitively, the dimension of the solution set of $A\vec{x} = \vec{b}$ is the number of free variables you would find after reducing the augmented matrix $\left[A \mid \vec{b}\right]$ to echelon form *provided the system is consistent* (we will formally state this soon). If the linear system is inconsistent, then there are no solutions, so we cannot talk about the dimension. Be careful: this is different from saying that the number of free variables is zero. If a system $A\vec{x} = \vec{b}$ is consistent, there might only be one solution, in which case the number of free variables is zero. There is also something else that is sneaky about the "definition" of dimension. If you find the vectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$ but your neighbor finds another linearly independent set of vectors $\{\vec{v}_1, \ldots, \vec{v}_l\}$ with $l \in \mathbb{N}$, then does $l = k$? In order for the above definition to make sense, the answer to this question better be yes. In terms of the augmented matrix, if you *redefine* the echelon form of a matrix so that the pivots do not have to go in the order we have demanded, then will the number of free variables still be the same? That's not completely obvious to me, so we should prove it.

Theorem 4.53. *Let $\{\vec{u}_1, \ldots, \vec{u}_k\}$ and $\{\vec{v}_1, \ldots, \vec{v}_l\}$ be two sets of linearly independent vectors that span the solution set to the homogeneous linear system $A\vec{x} = \vec{0}$. Then $k = l$.*

The proof of this introduces more ideas from logic, namely *proof by contradiction.* This logic is as follows. To prove that "A implies B" we can "assume that A is true and B is false." If we can deduce some logical contradiction (such as "A is false" or "C is false" where C is something that must be true provided A is true), then the initial assumption, namely that "A is true and B is false", is false. Since the claim assumes "A is true," we must conclude that "B is true." The following proof might not be easy to follow if this is your first time proving something by contradiction. What also makes the following proof difficult is that it is broken up into many steps. Before we prove it, we will prove an important preliminary fact.

Lemma 4.54. *Let V be the solution set to the homogeneous linear system $A\vec{x} = \vec{0}$ of m equations in n variables with $m, n \in \mathbb{N}$. Let $S := \{\vec{u}_1, \ldots, \vec{u}_k\}$ be a linearly independent set of vectors such that $\operatorname{span}(S) = V$ and let $T := \{\vec{v}_1, \ldots, \vec{v}_l\}$ be a set that spans V, where $k, l \in \mathbb{N}$. Then $k \leq l$.*

Proof. Since T spans V, $\vec{u}_1$ can be written as a linear combination of the $\vec{v}$'s, i.e. there exist coefficients $c_{11}, \ldots, c_{1l} \in \mathbb{R}$ such that

$$\vec{u}_1 = c_{11}\vec{v}_1 + \cdots + c_{1l}\vec{v}_l. \tag{4.55}$$

Since S is linearly independent, Theorem 4.47 says that $\vec{u}_1$ is not zero. Therefore, one of the c's is not zero, i.e. there exists an $i_1 \in \{1, \ldots, l\}$ such that $c_{1i_1} \neq 0$. Therefore, we can divide by it to write $\vec{v}_{i_1}$ in terms of the other vectors, namely (the second line just compactifies the expression)

$$\begin{aligned} \vec{v}_{i_1} &= \frac{1}{c_{1i_1}}\vec{u}_1 - \frac{c_{11}}{c_{1i_1}}\vec{v}_1 - \cdots - \frac{c_{1i_1-1}}{c_{1i_1}}\vec{v}_{i_1-1} - \frac{c_{1i_1+1}}{c_{1i_1}}\vec{v}_{i_1+1} - \cdots - \frac{c_{1l}}{c_{1i_1}}\vec{v}_l \\ &= \frac{1}{c_{1i_1}}\vec{u}_1 - \sum_{\substack{j=1 \\ j\neq i_1}}^{l} \frac{c_{1j}}{c_{1i_1}}\vec{v}_j. \end{aligned} \tag{4.56}$$

We might sometimes write this as

$$\vec{v}_{i_1} = \frac{1}{c_{1i_1}}\vec{u}_1 - \frac{c_{11}}{c_{1i_1}}\vec{v}_1 - \cdots - \widehat{i_1^{\text{th}} \text{ term}} - \cdots - \frac{c_{1l}}{c_{1i_1}}\vec{v}_l, \tag{4.57}$$

where the wide hat over an expression indicates to exclude it. Now, define

$$T_1 := \{\vec{u}_1, \vec{v}_1, \ldots, \widehat{\vec{v}_{i_1}}, \ldots, \vec{v}_l\}, \tag{4.58}$$

which appends $\vec{u}_1$ to T and removes $\vec{v}_{i_1}$. Notice that T_1 still spans V. Hence, there exist coefficients $d_{21}, c_{21}, \ldots, \widehat{c_{2i_1}}, \ldots, c_{2l} \in \mathbb{R}$ (the hat still means that we exclude this term—I'm only writing it to keep track of the numbers) such that

$$\vec{u}_2 = d_{21}\vec{u}_1 + c_{21}\vec{v}_1 + \cdots + \widehat{c_{2i_1}\vec{v}_{i_1}} + \cdots + \ldots, c_{2l}\vec{v}_l. \tag{4.59}$$

Claim. There exists an $i_2 \in \{1, \ldots, \widehat{i_1}, \ldots, l\}$ such that $c_{2i_2} \neq 0$. *Proof of claim.* Suppose to the contradiction that *all* of the c coefficients were zero in (4.59). Since $\vec{u}_2$ is not zero, this would say $\vec{u}_2 \propto \vec{u}_1$,[29] which contradicts that S is linearly independent (see Example 4.12). *End of proof of claim.* Therefore,

$$\vec{v}_{i_2} = \frac{1}{c_{2i_2}}\vec{u}_2 - \frac{d_{21}}{c_{2i_2}}\vec{u}_1 - \sum_{\substack{j=1 \\ j\neq i_1, j\neq i_2}}^{l} \frac{c_{2j}}{c_{2i_2}}\vec{v}_j \tag{4.60}$$

Hence, we can also remove this vector from T_1, define

$$T_2 := \{\vec{u}_1, \vec{u}_2, \vec{v}_1, \ldots, \widehat{\vec{v}_{i_1}}, \ldots, \widehat{\vec{v}_{i_2}}, \ldots, \vec{v}_l\}, \tag{4.61}$$

and T_2 still spans V. We can continue to remove one $\vec{v}_{i_j}$ at a time from T_{j-1} and replace it with a $\vec{u}_j$ to construct T_j (instead of using Example 4.12, we need to use Theorem 4.28 to show that there exist nonzero coefficients c_{ji_j}—I encourage you to do the next step to see this explicitly). This process ends at T_k, where we have

$$T_k := \{\vec{u}_1, \ldots, \vec{u}_k, \vec{v}_{r_1}, \ldots, \vec{v}_{r_{l-k}}\}, \tag{4.62}$$

where the $\vec{v}_r$'s are the leftover $\vec{v}$'s from this procedure. Note that there are exactly $l - k$ of them. T_k still spans V and we have that S is now a subset of T, written as $S \subseteq T_k$. Therefore $k \leq l$. ■

Proof of Theorem 4.53. Lemma 4.54 shows that $k \leq l$ and $l \leq k$. Hence $l = k$. ■

Theorem 4.63. *Let $A\vec{x} = \vec{b}$ be a consistent linear system. The dimension of the solution set of $A\vec{x} = \vec{b}$ is the number of free variables.*

Proof. Sorry, that last proof wore me out and so I'm leaving this as an exercise. However, you should be able to use an idea similar to the proof of Theorem 3.28. ■

Example 4.64. One of the conclusions of Example 1.1 was that there are a minimum of five scouts needed to count traffic in a certain part of Queens, New York. This number corresponded to the number of free variables we found when row reducing a system describing the traffic flow through these streets. The number five here corresponds to the dimension of the solution set to the homogeneous system described by (1.14).

[29] The notation $A \propto B$ means "A is proportional to B".

There have been many definitions introduced so far and it is important to not confuse them. For example, let's distinguish span from linear independence. Let $\vec{v}_1, \ldots, \vec{v}_k$ be vectors in $\mathbb{R}^n$.

The following are equivalent.

(a) $\vec{b} \in \text{span}\{\vec{v}_1, \ldots, \vec{v}_k\}$

(b) $\left[\begin{array}{ccc|c} | & & | & | \\ \vec{v}_1 & \cdots & \vec{v}_k & \vec{b} \\ | & & | & | \end{array}\right]$ is consistent

(c) there exist real numbers $x_1, \ldots, x_k$ such that $x_1\vec{v}_1 + \cdots + x_k\vec{v}_k = \vec{b}$.

The following are equivalent.

(a) $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent

(b) $\left[\begin{array}{ccc|c} | & & | & | \\ \vec{v}_1 & \cdots & \vec{v}_k & \vec{0} \\ | & & | & | \end{array}\right]$ has no free variables

(c) the only solution to $x_1\vec{v}_1 + \cdots + x_k\vec{v}_k = \vec{0}$ is $x_1 = \cdots = x_k = 0$.

There is also a grammatical difference between the notion of span and linear independence. Span is used as a *noun* associated to a collection of vectors while linear independence is used as a descriptor of such a set, so being linearly independent is an *adjective* associated to a set.

Recommended Exercises. Exercises 6, 8 (use a theorem!), 36, and 38 in Section 1.7 of [Lay]. You may (and are encouraged to) use any theorems we have done in class! Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we finished Section 1.7 of [Lay] and explored some ideas from Section 2.9 of [Lay]. Whenever you see the equation $A\vec{x} = \vec{b}$, just read it as the associated linear system as in (1.27).

Terminology checklist

linearly independent	
linearly dependent	
dimension	

5 Subspaces, bases, and linear manifolds

We need some more experience with vectors in Euclidean space ($\mathbb{R}^n$) and certain subsets of it when working with the systems of equations that appear in linear algebra. We have already seen that the set of solutions to a system of linear equations is always "linear" in the sense that it is either a point, a line, a plane, or a higher-dimensional plane. We often distinguish solution sets depending on whether they contain the zero vector or not.

Definition 5.1. A *subspace* of $\mathbb{R}^n$ is a set H of vectors in $\mathbb{R}^n$ satisfying the following conditions.

(a) $\vec{0} \in H$.

(b) For every pair of vectors $\vec{u}$ and $\vec{v}$ in H, their sum $\vec{u} + \vec{v}$ is also in H.

(c) For every vector $\vec{v}$ and constant c, the scalar multiple $c\vec{v}$ is in H.

Example 5.2. $\mathbb{R}^n$ itself is a subspace of $\mathbb{R}^n$. Also, the set $\{\vec{0}\}$ consisting of just the zero vector in $\mathbb{R}^n$ is a subspace.

Are there other subspaces?

Exercise 5.3. Let H be the set of points in $\mathbb{R}^3$ described by the solution set of

$$3x - 2y + z = 0, \tag{5.4}$$

which is depicted in Figure 6.

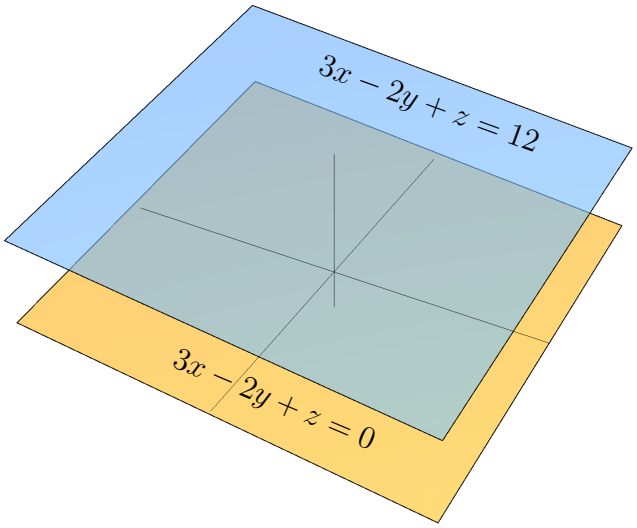


Figure 6: A plot of the planes described by $3x - 2y + z = 0$ (Exercise 5.3) and $3x - 2y + z = 12$ (Exercise 5.6)

Is $\vec{0}$ in H? Let

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \qquad \& \qquad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \tag{5.5}$$

be two vectors in H and let c be a real number. Is $\vec{u} + \vec{v}$ in H? Is $c\vec{v}$ in H?

Exercise 5.6. Is the set of solutions to

$$3x - 2y + z = 12 \tag{5.7}$$

a subspace of $\mathbb{R}^3$? See Figure 6 for a comparison of this solution set to the one from Exercise 5.3. If not, what goes wrong?

The last example was a subspace that has been shifted by a vector. To see this, notice that any point on the set of solutions to $3x - 2y + z = 12$ can be expressed as the set of vectors of the form (exercise!)

$$\begin{bmatrix} x \\ y \\ 12 - 3x + 2y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 12 \end{bmatrix} + x \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \tag{5.8}$$

for all $x, y \in \mathbb{R}$. This is almost the same as the set of solutions of $3x - 2y + z = 0$ except for the additional constant vector $(0, 0, 12)$. The set of solutions is of the form $\vec{p} + \vec{u}$ where

$$\vec{p} = \begin{bmatrix} 0 \\ 0 \\ 12 \end{bmatrix} \qquad \& \qquad \vec{u} = x \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \tag{5.9}$$

are particular and homogeneous solutions, respectively. In other words, if we denote the set of solutions to $3x - 2y + z = 0$ by H and the set of solutions to $3x - 2y + z = 12$ by S, then

$$S = \vec{p} + H, \tag{5.10}$$

where the latter notation means

$$\vec{p} + H := \{\vec{p} + \vec{u} \ : \ \vec{u} \in H\}. \tag{5.11}$$

Definition 5.12. A *linear manifold* (sometimes called an *affine subspace*[30]) in $\mathbb{R}^n$ is a subset S of $\mathbb{R}^n$ for which there exists a vector $\vec{p}$ such that the set

$$S - \vec{p} := \{\vec{v} - \vec{p} \ : \ \vec{v} \in S\} \tag{5.13}$$

is a subspace of $\mathbb{R}^n$.

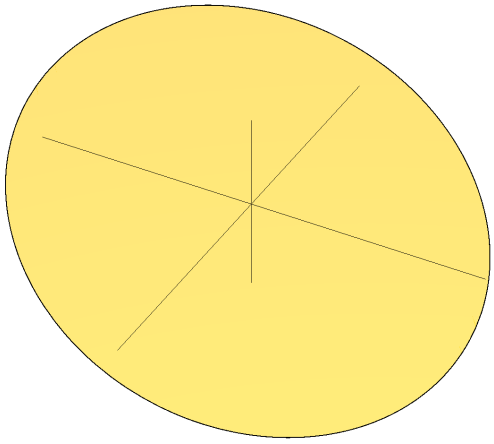
In other words, a linear manifold is a subspace that is shifted from the origin by some vector. Therefore, every subspace is a linear manifold (because all we have to do is shift by the zero vector) but not conversely, meaning that not every linear manifold is a subspace. In fact, we will soon see that a subspace is precisely a linear manifold that contains the zero vector. But before that, we will get a feeling for more of the geometric properties of subspaces and linear manifolds.

Exercise 5.14. Is the set of solutions to

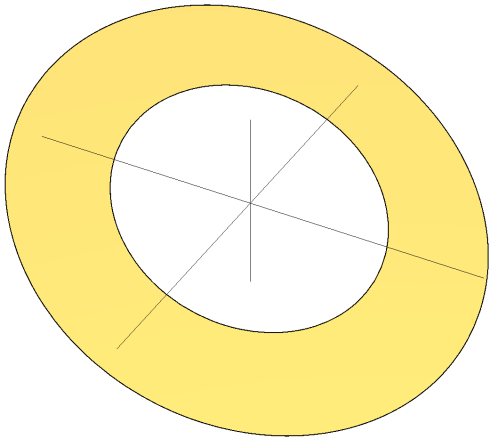
$$3x - 2y + z = 0 \tag{5.15}$$

with the constraint that

$$x^2 + y^2 \leq 1 \tag{5.16}$$



(a) A plot of the set of solutions to $3x - 2y + z = 0$ with the constraint $x^2 + y^2 \leq 1$.



(b) A plot of the set of solutions to $3x - 2y + z = 0$ with the constraint $\frac{1}{3} \leq x^2 + y^2 \leq 1$.

Figure 7: A plot of the set of solutions to $3x - 2y + z = 0$ with different constraints.

a subspace of $\mathbb{R}^3$? See Figure 7a. What goes wrong? Which of the three properties of the definition of subspace remain valid even in this example? What about the same linear system but with the constraint that

$$\frac{1}{3} \leq x^2 + y^2 \leq 1? \tag{5.17}$$

See Figure 7b. Are either of these linear manifolds?

The previous example leads to the following definition and hints at the following fact.

Theorem 5.18. *Let $A\vec{x} = \vec{b}$ be a consistent linear system of m equations in n unknowns. Then the set of solutions to this system is a linear manifold in $\mathbb{R}^n$. Furthermore, if $\vec{b} = 0$, then the set of solutions is a subspace.*

Proof. We first prove the second claim. Let H be the solution set of $A\vec{x} = \vec{0}$. Then $A\vec{0} = \vec{0}$ so that $\vec{0}$ is a solution. If $\vec{u}$ and $\vec{v}$ are solutions, then $A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v} = \vec{0} + \vec{0} = \vec{0}$ so that $\vec{u} + \vec{v}$ is a solution. If $\vec{u}$ is a solution, then $A(c\vec{u}) = cA\vec{u} = c\vec{0} = \vec{0}$ so that $c\vec{u}$ is a solution for all $c \in \mathbb{R}$.

To prove the first claim, let S be the solution set of $A\vec{x} = \vec{b}$ and let $\vec{p} \in S$. By Theorem 3.34, $S = H + \vec{p}$. Hence, $H = S - \vec{p}$ is a subspace by the previous paragraph. Therefore, S is a linear manifold. ■

Definition 5.19. A *basis* for a subspace H of $\mathbb{R}^n$ is a set of vectors that is both linearly independent and spans H. A *tangent basis* for a linear manifold S in $\mathbb{R}^n$ at a point $\vec{p} \in S$ is a basis for the subspace $H := S - \vec{p}$.

Why do we call it a tangent basis when talking about a linear manifold? There are two reasons. First of all, we use tangent basis because it should remind you of the tangent space to a surface in three-dimensional space. For example, the tangent space to the north pole on a sphere is a plane that is tangent to the sphere and has its origin at the north pole. A *tangent* basis for this plane consists of two tangent vectors whose origins begin at the north pole (see Figure 8). The

[30] I will avoid this terminology because we are already using the word "subspace" to mean something else.

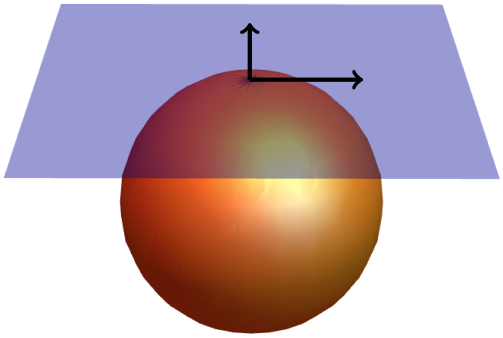


Figure 8: A sphere with tangent plane at the north pole together with a tangent basis.

second reason we call it a tangent basis in the context of linear manifolds is because we will later introduce the closely related concept of an affine basis. These will be vectors whose origin is the zero vector in Euclidean space (so the origin is *not* the north pole in the sphere example). One relationship between these two slightly different definitions is summarized in Figure 9.

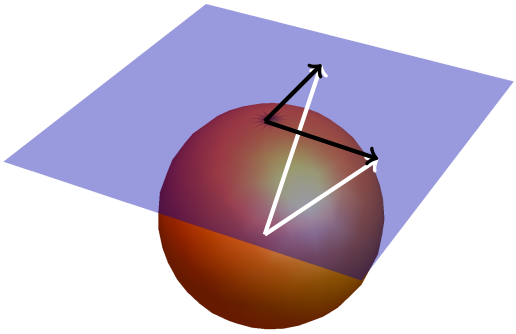


Figure 9: A sphere with tangent plane at the north pole together with a tangent basis (in black) and an affine basis (in white). The sphere is centered at the origin. The image has been tilted from Figure 8 so that it is easier to see the affine basis.

Exercise 5.20. Going back to our previous example of the plane in $\mathbb{R}^3$ specified by the linear system

$$3x - 2y + z = 0, \tag{5.21}$$

what is a basis for the vectors in this plane? Since the set of all vectors

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{5.22}$$

satisfying this linear system *define* this plane, we just need to find a basis for these solutions. We know that if we specify x and y as our free variables, then a general solution of this system is of the form

$$\begin{bmatrix} x \\ y \\ -3x + 2y \end{bmatrix} \tag{5.23}$$

with x and y free. How about testing the cases $x = 1$ with $y = 0$ and $x = 0$ with $y = 1$? This gives

$$\begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} \qquad \& \qquad \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \tag{5.24}$$

respectively. Any other vector in the solution set is a linear combination of these two vectors.

Definition 5.25. The number of elements in a basis for a subspace H of $\mathbb{R}^n$ is the *dimension* of H and is denoted by $\dim H$.

As before, the fact that this number is well-defined is not obvious. How can we be sure that any two choices of bases have the same number of vectors in them? However, the proof given for the dimension of a solution set in Theorem 4.53 is exactly the same as the proof of this claim. Given an arbitrary set of vectors, one can always reduce this set enough so that the left-over vectors form a basis for the span of the initial set of vectors.

Theorem 5.26. *Let $k \in \mathbb{N}$ and let $\{\vec{v}_1, \ldots, \vec{v}_k\}$ be a set of vectors in $\mathbb{R}^n$ and set $H := \operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_k\}$. Then either $H = \{\vec{0}\}$ or there exists a finite subset $\{\vec{v}_{i_1}, \ldots, \vec{v}_{i_l}\}$ for some $l \in \{1, \ldots, k\}$ of $\{\vec{v}_1, \ldots, \vec{v}_k\}$ that is linearly independent and $\operatorname{span}\{\vec{v}_{i_1}, \ldots, \vec{v}_{i_l}\} = H$, i.e. $\{\vec{v}_{i_1}, \ldots, \vec{v}_{i_l}\}$ is a basis for H.*

Proof. The only way that $H = \{\vec{0}\}$ is if every vector is the zero vector. Otherwise, we proceed by exhaustion. If $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent already, then we are done. Else, by Theorem 4.28, there exists a number $i_1 \in \{1, \ldots, k\}$ so that the vector $\vec{v}_{i_1} \in \{\vec{v}_1, \ldots, \vec{v}_k\}$ can be expressed as a linear combination of the others. Then

$$\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_{i_1-1}, \vec{v}_{i_1+1}, \ldots, \vec{v}_k\} = H. \tag{5.27}$$

If now $\{\vec{v}_1, \ldots, \vec{v}_{i_1-1}, \vec{v}_{i_1+1}, \ldots, \vec{v}_k\}$ is linearly independent, then we are done. Else, we repeat the procedure until we have obtained a linearly independent subset of vectors. This procedure must end because the number of vectors is finite and because at least one of the vectors is non-zero. ■

If it is not clear which vectors are linear combinations of the others, there is a systematic way of determining which vectors one can get rid of. Write

$$\left[\begin{array}{ccc|c} | & & | & | \\ \vec{v}_1 & \cdots & \vec{v}_k & \vec{0} \\ | & & | & | \end{array}\right] \tag{5.28}$$

as an augmented matrix and row reduce. The pivot columns of the *original* augmented matrix form a basis for the span of all the vectors. In other words, one can remove the vectors whose columns correspond to the free variables. The following example illustrates how this can be done.

Problem 5.29. Find a basis for the subspace spanned by the following set of vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\} \tag{5.30}$$

in $\mathbb{R}^5$.

Answer. Let H be the subspace spanned by the given vectors. Row reducing the associated augmented matrix gives

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & -1 & 0 \\ 0 & -1 & 1 & 1 & 0 \\ 1 & 0 & 2 & 1 & 0 \\ -1 & -2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 & 0 \end{array}\right] \mapsto \left[\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{5.31}$$

in reduce echelon form.[31] The first two columns are pivot columns. Hence,

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ -2 \\ -1 \end{bmatrix} \right\} \tag{5.32}$$

forms a basis for H.

A linear manifold is equivalently described by the geometric property that says it contains the straight line (going infinitely far in both directions) through any two points inside of it.

Theorem 5.33. *A nonempty subset $S \subseteq \mathbb{R}^n$ is a linear manifold if and only if for any two vectors $\vec{x}$ and $\vec{y}$ in S, the vector*

$$t\vec{x} + (1-t)\vec{y} \tag{5.34}$$

is in S for all $t \in \mathbb{R}$.

Proof.
($\Rightarrow$) Suppose that S is a linear manifold so that there exists a vector $\vec{p} \in S$ such that $H := S - \vec{p}$ is a subspace. Let $\vec{x}, \vec{y} \in S$ and let $t \in \mathbb{R}$. The goal is to show that $t\vec{x} + (1-t)\vec{y} \in S$. Because $H := S - \vec{p}$, it follows that $\vec{x} - \vec{p}, \vec{y} - \vec{p} \in H$. By definition of H being a subspace, $t(\vec{x} - \vec{p}) + (1-t)(\vec{s} - \vec{p}) \in H$ (this is combining two facts, the first of which is that any scalar multiples of $\vec{x} - \vec{p}$ and $\vec{y} - \vec{p}$ are in H and the second of which is that the sum of any two vectors in H is also in H). Therefore, adding back $\vec{p}$ gives us a vector in S, namely $t(\vec{x} - \vec{p}) + (1-t)(\vec{s} - \vec{p}) + \vec{p} \in S$ since $S = H + \vec{p}$. But this expression is equal to

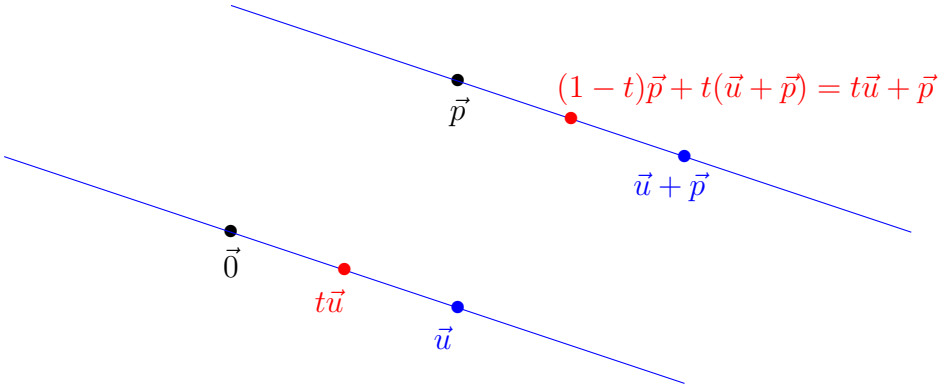
$$t(\vec{x} - \vec{p}) + (1-t)(\vec{s} - \vec{p}) + \vec{p} = t\vec{x} + (1-t)\vec{y}, \tag{5.35}$$

which shows that $t\vec{x} + (1-t)\vec{y} \in S$.
($\Leftarrow$) Suppose S is a set for which the straight line going through any two vectors in S is also in S. We must find a subspace H and a vector $\vec{p}$ such that $S = \vec{p} + H$. First, since S is nonempty, it contains at least one vector. Let $\vec{p}$ be any such vector in S and set $H := S - \vec{p}$. We show that H is a subspace.

[31] You only need to bring the augmented matrix to echelon form to find the pivot columns. I just wrote the reduced echelon form because the echelon form is not unique while the reduced echelon form is (provided you are following my ordering convention).

(a) $\vec{0} \in H$ because $\vec{0} = \vec{p} - \vec{p} \in S - \vec{p}$ since $\vec{p} \in S$.

(b) Let us actually check the scalar condition first since that is easier. Let $\vec{u} \in H$ and let $t \in \mathbb{R}$.

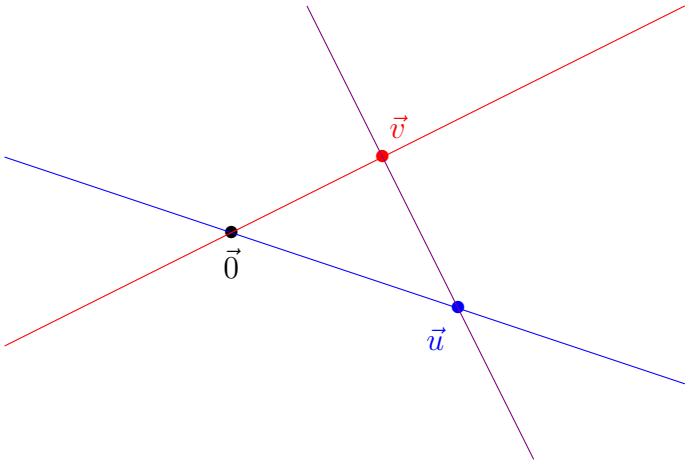


Then

$$t\vec{u} = \underbrace{(1-t)\vec{p} + t(\vec{u}+\vec{p})}_{\in S} - \vec{p} \tag{5.36}$$

showing that $t\vec{u} \in H = S - \vec{p}$.

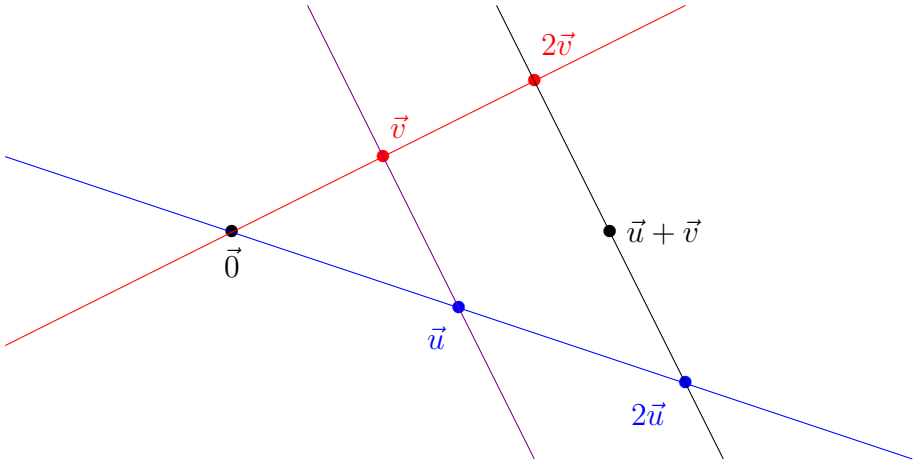
(c) Let $\vec{u}, \vec{v} \in H$. Since $\vec{0}$ is also in H, we can draw the straight lines through any pairs of these two vectors.



Drawing where $\vec{u} + \vec{v}$ shows that it lies along a line parallel to the line straight through $\vec{u}$ and $\vec{v}$ and is explicitly given by

$$\vec{u} + \vec{v} = \frac{1}{2}(2\vec{u}) + \frac{1}{2}(2\vec{v}). \tag{5.37}$$

A visualization of where this expression comes from is given by the following graphic.



Just as in (b), we can draw this in S by adding $\vec{p}$. Using our assumption gives $\vec{u}+\vec{v}+\vec{p} \in S$. This shows that $\vec{u}+\vec{v} \in H = S - \vec{p}$.

Hence, H is a subspace of $\mathbb{R}^n$ showing that S is a linear manifold. ■

The previous proof indicates how 3 distinct points determine a plane. Notice that the point $\vec{0}$ could have been anything and the geometric idea would still hold.

Definition 5.38. Let S be a set of vectors in $\mathbb{R}^n$. The *affine span* of S is the set of all linear combinations of vectors $\vec{v}$ in S of the form

$$\sum_{\vec{v}} a_{\vec{v}}\vec{v} \qquad \text{such that} \qquad \sum_{\vec{v}} a_{\vec{v}} = 1 \tag{5.39}$$

and all but finitely many $a_{\vec{v}}$ are zero. A linear combinations of this form is called an *affine linear combination*. The affine span of S is denoted by $\text{aff}(S)$. For example, if $S = \{\vec{v}_1, \ldots, \vec{v}_m\}$ is a finite set of m vectors, the affine span of S is all linear combinations of these vectors of the form

$$a_1\vec{v}_1 + \cdots + a_m\vec{v}_m \qquad \text{such that} \qquad a_1 + \cdots + a_m. \tag{5.40}$$

Example 5.41. In the proof of Theorem 5.33, $\vec{u}+\vec{v}$ is in the affine span of the vectors $\{\vec{0}, \vec{u}, \vec{v}\}$ while $\vec{u}+\vec{v}$ is not in the affine span of $\{\vec{u}, \vec{v}\}$. However, $\vec{u}+\vec{v}$ is in the affine span of $\{2\vec{u}, 2\vec{v}\}$.

The affine span, not the usual span of vectors, is used to add vectors in linear manifolds. The reason is because if we do not impose the additional condition that the sum of the coefficients is 1, we might "jump off" the linear manifold. We formalize this as follows.

Theorem 5.42. *A set S is a linear manifold if and only if for every subset $R \subseteq S$, the affine span of R is in S, i.e. $\text{aff}(R) \subseteq S$.*

Proof. Exercise. ■

The definition of linear dependence and independence of vectors can also be applied to linear manifolds and affine combinations.

Definition 5.43. A set of vectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$ in $\mathbb{R}^n$ is *affinely linearly dependent* if there exist real numbers $x_1, \ldots, x_k$, not all of which are zero, such that

$$x_1 + \cdots + x_k = 0 \qquad \text{and} \qquad x_1\vec{u}_1 + \cdots + x_k\vec{u}_k = \vec{0}. \tag{5.44}$$

Otherwise, the set of vectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$ is *affinely linearly independent*. A set of vectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$ in $\mathbb{R}^n$ is an *affine basis* for a linear manifold S in $\mathbb{R}^n$ iff $\{\vec{u}_1, \ldots, \vec{u}_k\}$ is affinely linearly independent and $\text{aff}\big(\{\vec{u}_1, \ldots, \vec{u}_k\}\big) = S$, the affine span of the vectors is equal to S.

Example 5.45. Going back to the tangent plane to the unit sphere at the north pole such as in Figure 9, let S be the $z = 1$ plane. The vectors

$$\vec{v}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \& \qquad \vec{v}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \tag{5.46}$$

(in black in Figure 9) form a tangent basis for S while the vectors

$$\vec{u}_1 := \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \vec{u}_2 := \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \& \qquad \vec{u}_3 := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{5.47}$$

(the first two are drawn in white in Figure 9) form an affine basis for S. The north pole specifies the origin of the tangent plane and is given by the vector

$$\vec{p} := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \tag{5.48}$$

which coincides with our choice of $\vec{u}_3$ in this case. The relationship between the tangent basis and the affine basis is given by

$$\vec{u}_1 = \vec{v}_1 + \vec{p}, \qquad \vec{u}_2 = \vec{v}_2 + \vec{p}, \qquad \vec{u}_3 = \vec{0} + \vec{p}. \tag{5.49}$$

What is the difference between affine independence and linear independence? In terms of augmented matrices, $\{\vec{u}_1, \ldots, \vec{u}_k\}$ in $\mathbb{R}^n$ is linearly independent iff the only solution to

$$\left[\begin{array}{ccc|c} | & & | & | \\ \vec{v}_1 & \cdots & \vec{v}_k & \vec{0} \\ | & & | & | \end{array}\right] \tag{5.50}$$

is the trivial solution (i.e. if there are no free variables). To the contrast, $\{\vec{u}_1, \ldots, \vec{u}_k\}$ in $\mathbb{R}^n$ is affinely independent iff the only solution to

$$\left[\begin{array}{ccc|c} 1 & \cdots & 1 & 0 \\ | & & | & | \\ \vec{v}_1 & \cdots & \vec{v}_k & \vec{0} \\ | & & | & | \end{array}\right] \tag{5.51}$$

is the trivial solution (i.e. if there are no free variables). This is because the first row is precisely the equation $x_1 + \cdots + x_k = 0$ while the rows below it collectively give $x_1\vec{v}_1 + \cdots + x_k\vec{v}_k = \vec{0}$.

Recommended Exercises. Exercises 3, 4, and 17 in Section 2.8 of [Lay] and Exercises 7 and 20 in Section 2.9 of [Lay]. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we went through parts of Sections 2.8, 2.9, 8.1, and 8.2 of [Lay].

Terminology checklist

subspace	
linear manifold (affine subspace)	
basis for a subspace	
tangent basis for a linear manifold at a point	
dimension	
affine span	
affine linear combination	
affinely linearly independent	
affine basis	

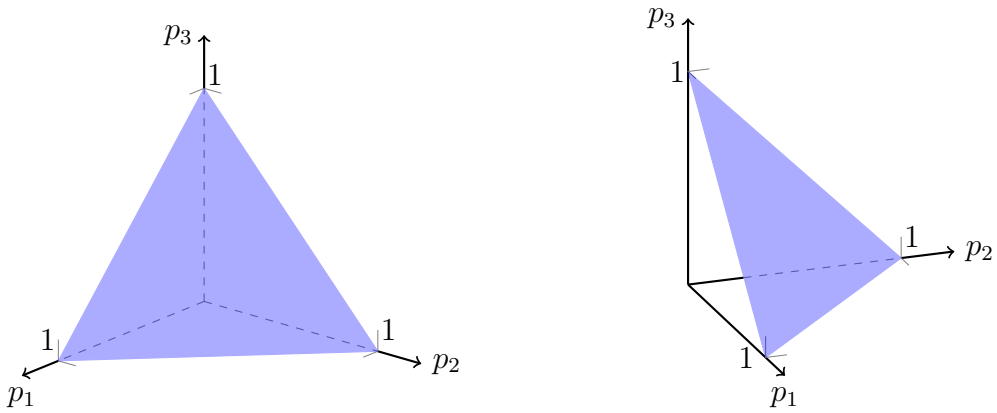
6 Convex spaces and linear programming

Linear manifolds with certain constraints are described by convex spaces. The quintessential example of a convex space that will occur in many contexts, especially probability theory, is that of a simplex.

Example 6.1. The set of all probability distributions on an n-element set can be described by a mathematical object known as the *standard $(n-1)$-simplex* and denoted by Δ^{n-1}. It is defined by

$$\Delta^{n-1} := \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n \ : \ \sum_{i=1}^{n} p_i = 1 \text{ and } p_i \geq 0 \ \forall \ i = 1, \ldots, n \right\}. \tag{6.2}$$

The interpretation of an $(n-1)$ simplex is as follows. For a set of events labeled by the numbers 1 through n, the probability of the event i taking place is p_i. For example, the 2-simplex looks like the following subset of $\mathbb{R}^3$ viewed from two different angles



The 1-simplex describes the probability space associated with flipping a weighted coin. Is the n-simplex a vector subspace of $\mathbb{R}^{n+1}$? Is it a linear manifold? Why or why not?

The previous example motivates the following definition.

Definition 6.3. A *convex space*[32] is a subset C of $\mathbb{R}^n$ such that if $\vec{u}, \vec{v}$ are any two vectors in C then

$$\lambda \vec{u} + (1-\lambda)\vec{v}, \tag{6.4}$$

with $\lambda \in [0,1]$, is also in C. The set of points between $\vec{u}$ and $\vec{v}$ is called the *interval* between $\vec{u}$ and $\vec{v}$.

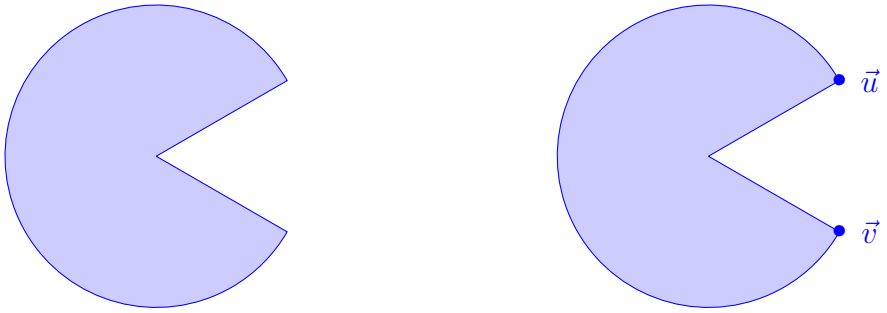
Example 6.5. Every linear manifold is a convex space. This is because for a linear manifold S, if the vectors $\vec{u}$ and $\vec{v}$ are in S, then all vectors of the form $\lambda\vec{u} + (1-\lambda)\vec{v}$ for all $\lambda \in \mathbb{R}$ are also in S. In particular, since all numbers between 0 and 1 are real numbers, $\lambda\vec{u} + (1-\lambda)\vec{v}$ are in S for all $\lambda \in [0,1]$ as well. However, not every convex space is a linear manifold. For example,

[32] I do not want to go into the technicalities of closed and open sets, but throughout, we will always assume that our convex spaces are also closed. Visually, it means that our convex spaces always include their boundaries, (faces, edges, vertices, etc.).

the n-simplex is a convex space but it is not a linear manifold. Intuitively, linear manifolds must extend infinitely far in all directions. Convex spaces can also extend in all directions, but they do not *have* to.

Exercise 6.6. Which of the examples in the previous section are convex spaces?

Example 6.7. PacMan



is not a convex space since the interval connecting $\vec{u}$ and $\vec{v}$ is not in the space.

Convex spaces are important in linear algebra because they often arise as the solution sets of systems of linear *inequalities* (instead of systems of equalities) of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\leq b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\leq b_m. \end{aligned} \tag{6.8}$$

You might ask why not also allow the reversed inequality on some of these equations? The reason for this is because we can multiply the whole inequality by -1, reverse the $\geq$ to a $\leq$, and then just rename the constant coefficients reproducing something of the form (6.8). How can we include equalities? This is done by replacing any equality, such as

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \tag{6.9}$$

by the two inequalities

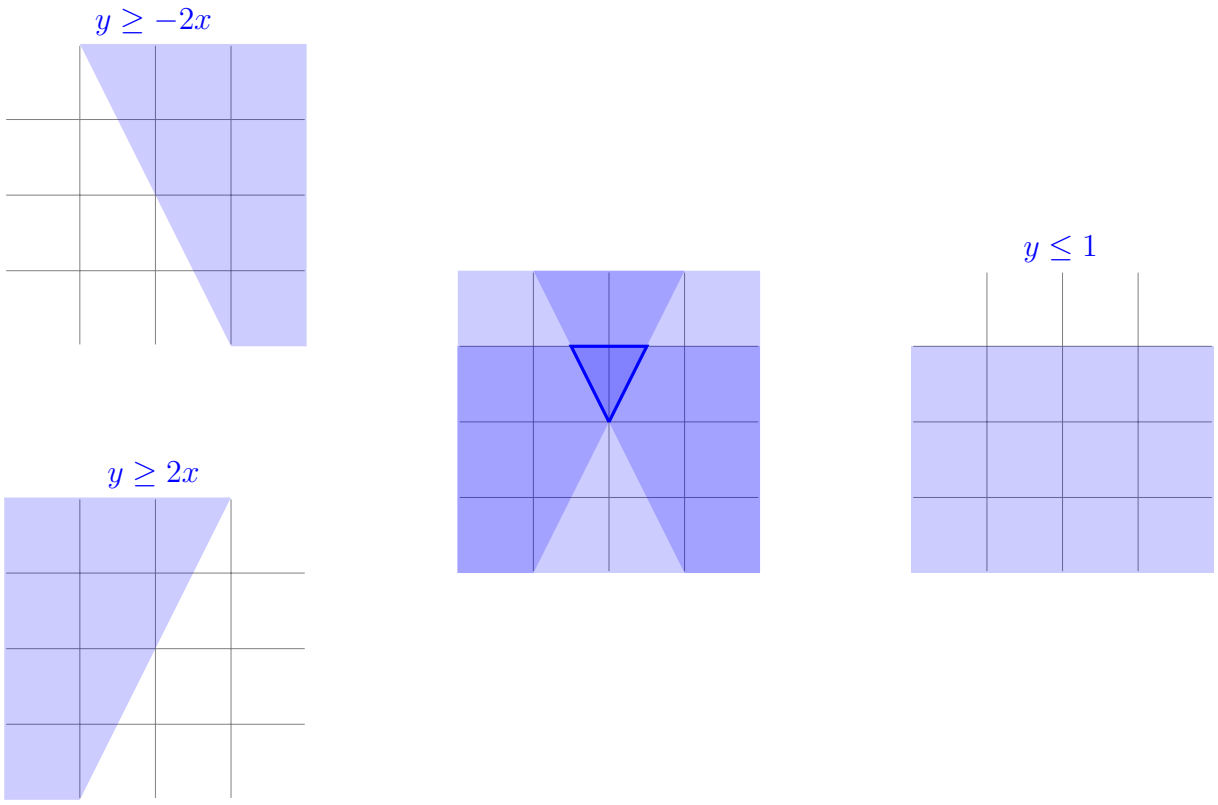
$$\begin{aligned} a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\leq b_2 \\ -a_{21}x_1 - a_{22}x_2 - \cdots - a_{2n}x_n &\leq -b_2 \end{aligned} \tag{6.10}$$

(the second inequality is equivalent to the first one with a $\geq$ instead after multiplying both sides by -1).

Example 6.11. Consider the following linear system of inequalities

$$\begin{aligned} y &\leq 1 \\ 2x - y &\leq 0 \\ -2x - y &\leq 0 \end{aligned} \tag{6.12}$$

These regions are depicted in the following figure (the central point is the origin)



The intersection describes an isosceles triangle with vertices given by

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}, \quad \& \quad \begin{bmatrix} -\frac{1}{2} \\ 1 \end{bmatrix}, \tag{6.13}$$

Solving systems of inequalities is difficult in general. Row operations do not work because multiplying by negative numbers reverses the sign of the inequality. Sometimes, one deals with a combination of linear systems of equalities and inequalities such as in Example 6.1. There, the linear system consists of only a single equation in n variables given by

$$p_1 + p_2 + \cdots + p_n = 1 \tag{6.14}$$

and a system of n inequalities

$$\begin{aligned} p_1 &\geq 0 \\ &\vdots \\ p_n &\geq 0. \end{aligned} \tag{6.15}$$

Theorem 6.16. *The set of solutions to any linear system of inequalities (6.8) is a convex space.*

If you parse through the definition of a convex space, this says that if $\vec{y}$ and $\vec{z}$ are two solutions to (6.8), then the *interval* connecting these two vectors, i.e. the set of points of the form $\lambda\vec{y} + (1-\lambda)\vec{z}$ with $\lambda \in [0, 1]$, is *also* in the solution set.

Proof. The proof is almost the same as it was for a linear system of equalities (Proposition 3.15) with one important difference. To see this let the system be described by $A\vec{x} \leq \vec{b}$ and let $\vec{y}$ and $\vec{z}$ be two solutions. Fix $\lambda \in [0,1]$. Then,

$$\begin{aligned} a_{11}\big(\lambda y_1 + (1-\lambda)z_1\big) + \cdots + a_{1n}\big(\lambda y_n + (1-\lambda)z_n\big) & \\ = \lambda(a_{11}y_1 + \cdots + a_{1n}y_n) + (1-\lambda)(a_{11}z_1 + \cdots + a_{1n}z_n) & \\ \leq \lambda b_1 + (1-\lambda)b_1 & \\ = b_1. & \end{aligned} \tag{6.17}$$

Note that it was crucial in the second last step that $\lambda \in [0,1]$ since if λ was not in this set, either λ or $1-\lambda$ would be negative, and then one of the inequalities would have to flip. ■

There are many convex spaces that do not come from linear inequalities.

Exercise 6.18. Which of the following are convex spaces?

(a) $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \ : \ x^2 + y^2 \leq 1 \right\}$

(b) $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \ : \ x^2 + y^2 \geq 1 \right\}$

(c) $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \ : \ x^2 - y^2 \leq 1 \right\}$

(d) $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \ : \ x^2 - y^2 \geq 1 \right\}$

(e) $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \ : \ x^2 - y^2 \geq 1, \ x \geq 0 \right\}$

(f) $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \ : \ x^2 - y^2 \geq 1, \ y \geq 0 \right\}$

(g) $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \ : \ x^2 - y^2 \geq -1, \ x \geq 0 \right\}$

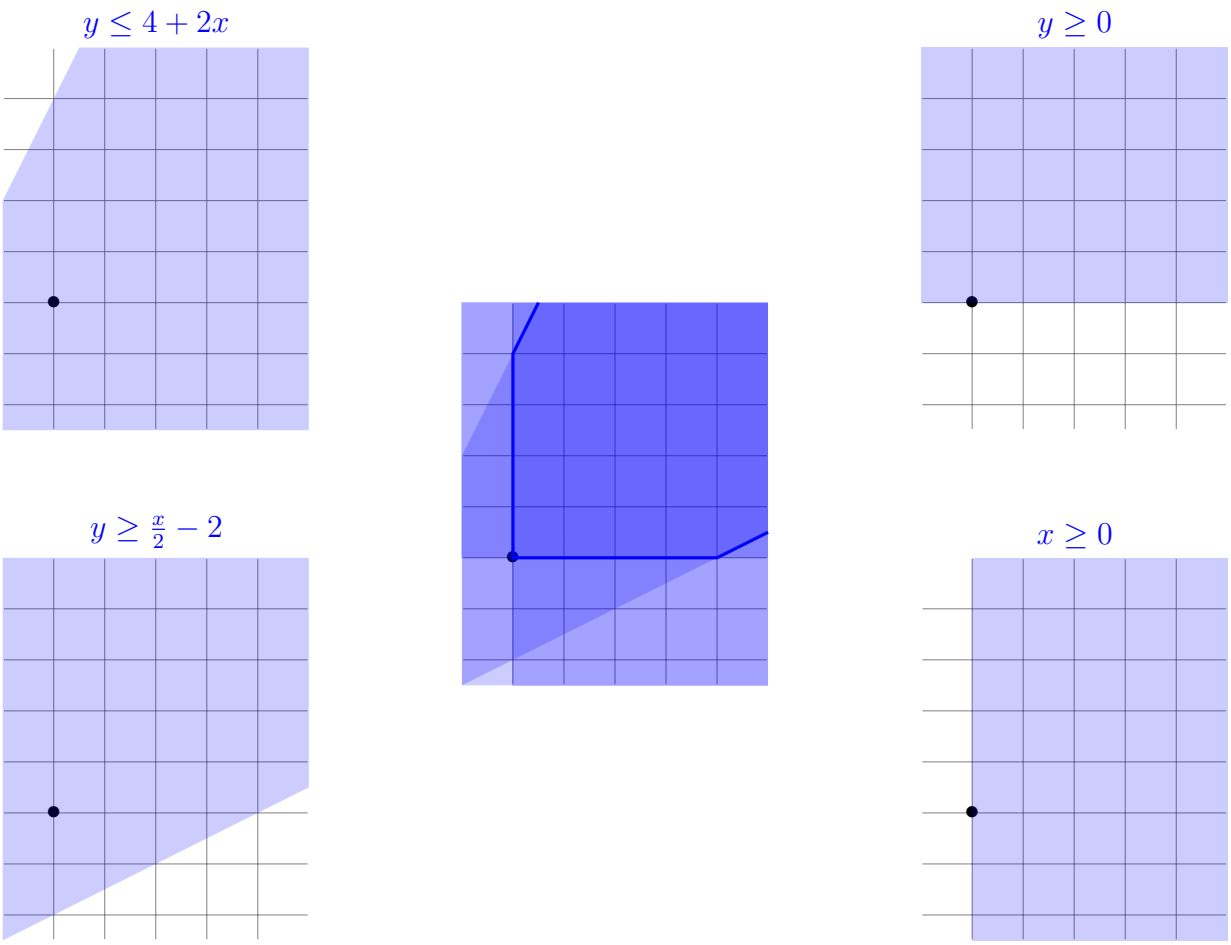
(h) $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \ : \ x^2 - y^2 \geq -1, \ y \geq 0 \right\}$

Problem 6.19 (Problem 9.2.9 in [2])**.** Find all solutions to the linear system of inequalities

$$\begin{aligned} -2x + y &\leq 4 \\ x - 2y &\leq -4 \\ -x &\leq 0 \\ -y &\leq 0. \end{aligned} \tag{6.20}$$

Draw the solution set as a convex subset of $\mathbb{R}^2$.

Answer. These regions are depicted in the following figure (the origin is represented by a •)



One often wants to optimize functions over linear constraints. As a result, one studies functions over convex spaces.

Definition 6.21. A *linear functional* on $\mathbb{R}^n$ is a function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying the condition

$$f(a\vec{u} + \vec{v}) = af(\vec{u}) + f(\vec{v}) \tag{6.22}$$

for all real numbers a and vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$.

For example, in Example 1 from Section 9.2 of [2], the function was given by

$$\mathbb{R}^2 \ni \begin{bmatrix} x \\ y \end{bmatrix} \mapsto 2x + 3y \in \mathbb{R}. \tag{6.23}$$

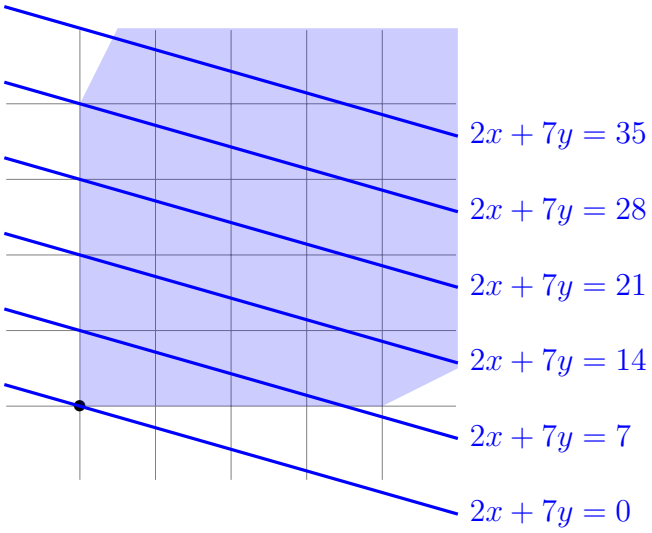
This described the profit obtained. In trying to maximize this function, we looked at the level sets of this function, which are described by the formula $2x + 3y = c$, with c a fixed constant.

Problem 6.24. Maximize the linear functional $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\mathbb{R}^2 \ni \begin{bmatrix} x \\ y \end{bmatrix} \mapsto 2x + 7y \in \mathbb{R} \tag{6.25}$$

over the convex set in Problem 6.19.

Answer. The level sets of this function are drawn for a few values below



Although the minimum of this function is attained, namely at $(x, y) = (0, 0)$, the maximum is not attained.

The fact that a maximum was not attained in the previous problem is due largely to the fact that the solution set of the linear system of inequalities is unbounded.

Definition 6.26. A subset $S \subseteq \mathbb{R}^n$ is *bounded* if there exists a positive number $R > 0$ such that $S \subseteq [-R, R] \times \cdots \times [-R, R]$, where $[-R, R]$ denotes the interval of diameter $2R$ centered at 0 and $[-R, R] \times \cdots \times [-R, R]$ denotes the n-dimensional cube centered at 0 and whose side lengths are all $2R$. If no such R exists, S is said to be *unbounded*.

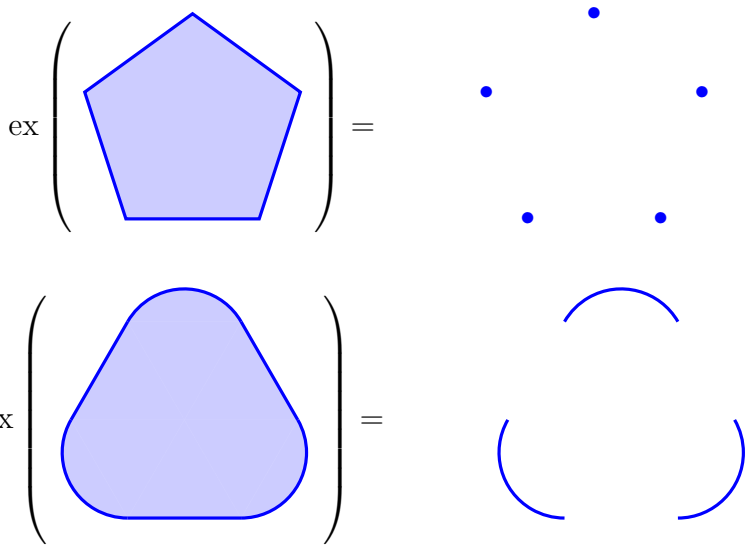
We have already learned that the set of solutions to a linear system of inequalities is a convex set. If this convex set is, in addition, bounded, then it is a polytope (in two dimensions, a polytope is just a polygon).

Theorem 6.27. *Let $A\vec{x} \leq \vec{b}$ be a consistent linear system of inequalities, let S denote the set of solutions, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear functional. If S is bounded, then f attains a maximum and a minimum on S.*

In fact, *much more* is true! The example we went over, namely Example 1 in Section 9.2 of [2], shows us that not only did a maximum occur on the polytope, but it occurred on a point of intersection given by the lines separating regions from the inequality. These points are incredibly significant because instead of checking infinitely many possible values to optimize a functional, one can merely check these extreme points.

Definition 6.28. An *extreme point* of a convex space C in $\mathbb{R}^n$ is a vector $\vec{u} \in C$ such that if $\vec{u} = \lambda\vec{v} + (1 - \lambda)\vec{w}$ for some $\vec{v}, \vec{w} \in C$ and $\lambda \in [0, 1]$, then $\vec{v} = \vec{u}$ and/or $\vec{w} = \vec{u}$. The set of extreme points of C is denoted by $\text{ex}(C)$.

Example 6.29. The following figures show two examples of extreme points of convex spaces illustrating the wide variety of possibilities.



Extreme points of a convex space are important because the entire convex space can be obtained via convex combinations of these points.

Definition 6.30. Let $\{\vec{u}_1, \ldots, \vec{u}_m\}$ be a set of vectors in $\mathbb{R}^n$. A *convex combination* of these vectors is a linear combination of these vectors of the form

$$p_1\vec{u}_1 + \cdots + p_m\vec{u}_m, \tag{6.31}$$

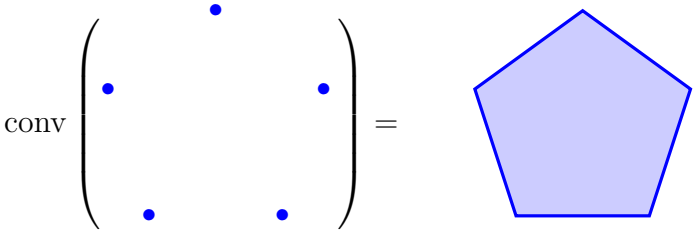
where $(p_1, \ldots, p_m) \in \Delta^{m-1}$, i.e.

$$p_i \geq 0 \quad \forall\, i \in \{1, \ldots, m\} \qquad \& \qquad \sum_{i=1}^{m} p_i = 1. \tag{6.32}$$

A convex combination of vectors is similar to an affine combination except that one can only obtain the vectors "between" the vectors given. We do not extend lines infinitely in both directions when taking a convex combination.

Definition 6.33. The *convex hull* of a set S of vectors in $\mathbb{R}^n$ is the set of all vectors in $\mathbb{R}^n$ that are (finite) convex combinations of the vectors in S. The convex hull of S is denoted by $\mathrm{conv}(S)$.

Example 6.34.



Theorem 6.35. *Let C be a bounded*[33] *convex set. Then*

$$C = \operatorname{conv}\big(\operatorname{ex}(C)\big). \tag{6.36}$$

Theorem 6.37. *Let $A\vec{x} \leq \vec{b}$ be a consistent linear system of inequalities, let S denote the set of solutions, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear functional. If S is bounded, then f attains a maximum and a minimum on* $\operatorname{ex}(S)$.

This theorem is related to something you may have learned in calculus. Given any continuously differentiable function $f : [a, c] \to \mathbb{R}$ on the domain $[a, c]$, (where $a, c \in \mathbb{R}$ and $a \leq c$) the maximum of f occurs at the critical points, i.e. where the derivative vanishes, or at the *extreme points*, which in this case are a and c. When f is a linear function, $f(x) = mx + b$ for some $m, b \in \mathbb{R}$, and its derivative is *always* nonzero provided that $m \neq 0$. Therefore, it has no critical points and its maximum occurs at a or c. If $m > 0$, the maximum occurs at c and if $m < 0$, the maximum occurs at a. Theorem 6.37 generalizes this result. If f is a linear function (now of any number of variables) on a (compact) convex space, the maximum of f must occur at the extreme points of the convex space.

Proof of Theorem 6.37. See Theorem 16 in Section 8.5 of [3]. ■

This theorem is supplemented by the important fact that the number of extreme points of a bounded convex set of solutions to a linear system of inequalities is *finite*. Let's use it to solve Example 1 from Section 9.2 in [2].

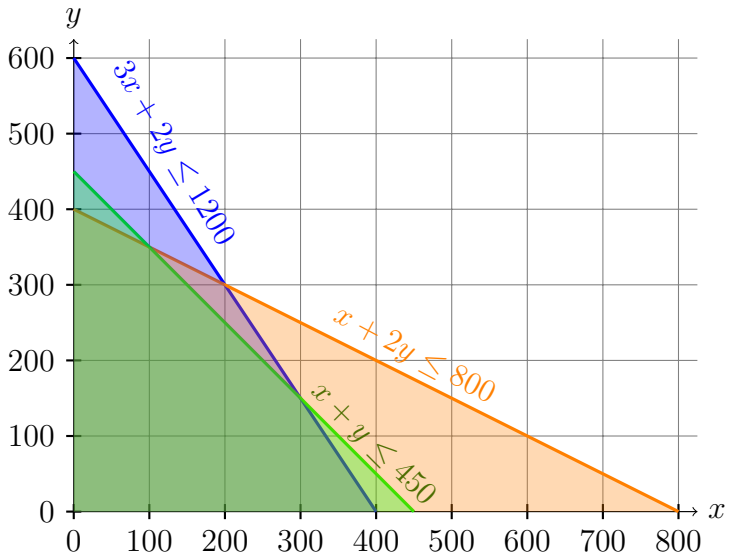
Problem 6.38. A company blends two types of seed mixtures, denoted by A and B. Each bag of A contains 3 pounds of fescue seed, 1 pound of rye seed, and 1 pound of bluegrass. Each bag of B contains 2 pounds of fescue, 2 pounds of rye, and 1 pound of bluegrass. The company has 1200 pounds of fescue, 800 pounds of rye, and 450 pounds of bluegrass. The company makes a profit of \$2 for each bag of A sold and \$3 for each bag of B sold. Assume that all bags are sold. How many bags of A and B should the company make so as to maximize its profit?

Answer. Let x and y denote the number of bags of A and B produced, respectively. Because the number of seeds of each time is positive and finite, there are inequalities enforced for the production of different quantities of bags A and B. These inequalities are given as follows

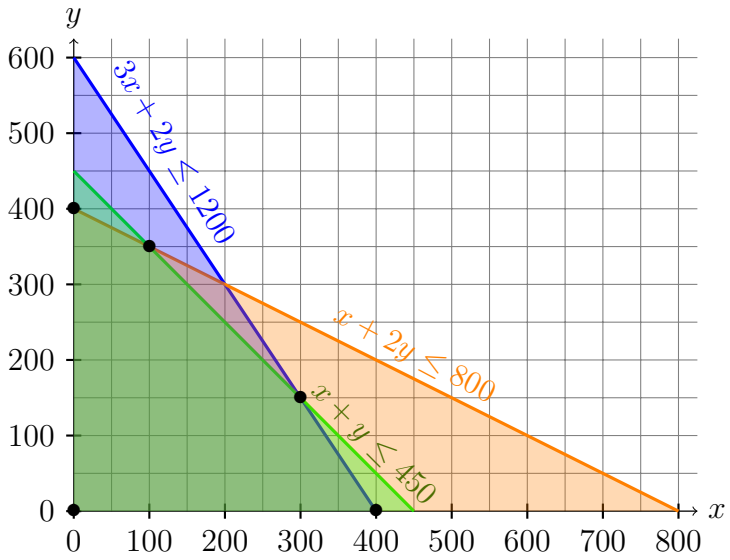
$$\begin{aligned} 0 \leq 3x + 2y &\leq 1200 \text{ for fescue} \\ 0 \leq x + 2y &\leq 800 \text{ for rye} \\ 0 \leq x + y &\leq 450 \text{ for bluegrass} \end{aligned} \tag{6.39}$$

and the corresponding regions are given by

[33] Technically, C must be compact for this to be true, since otherwise C might not have any extreme points. C is compact if and only if it is closed *and* bounded.



By looking at the common intersection of the above three regions, we can conclude that the extreme points are $(0, 0), (0, 400), (400, 0)$ and then the intersection of the lines described by $x + 2y = 800$ and $x + y = 450$ and the intersection of the lines described by $3x + 2y = 1200$ and $x + y = 450$. The solution of the first is given by $(100, 350)$ and the second is given by $(300, 150)$. These extreme points are bulleted in the graph below.



The profit function is given by

$$p(x, y) = 2x + 3y. \tag{6.40}$$

Hence, to maximize p subject to the above constraints, it suffices to compute p on the above extreme points

extreme points	profit
$(0, 0)$	0
$(0, 400)$	1200
$(100, 350)$	1250
$(300, 150)$	1050
$(400, 0)$	800

so that the maximum profit possible given the constraints is \$1250.

The following table summarizes the relationship between the three types of linear systems we have come across and their associated solution spaces.

Linear system	Solution set	allowed combinations for $\{\vec{v}_1, \ldots, \vec{v}_k\}$ a subset of the solution set
$A\vec{x} = \vec{0}$	subspace	all linear combinations $a_1\vec{v}_1 + \cdots + a_k\vec{v}_k$ with $a_1, \ldots, a_k \in \mathbb{R}$ called the span of $\{\vec{v}_1, \ldots, \vec{v}_k\}$ written $\operatorname{span}\big(\{\vec{v}_1, \ldots, \vec{v}_k\}\big)$
$A\vec{x} = \vec{b}$	linear manifold	all affine linear combinations $a_1\vec{v}_1 + \cdots + a_k\vec{v}_k$ with $a_1, \ldots, a_k \in \mathbb{R}$ satisfying $\sum_{i=1}^{k} a_i = 1$ called the affine span of $\{\vec{v}_1, \ldots, \vec{v}_k\}$ written $\operatorname{aff}\big(\{\vec{v}_1, \ldots, \vec{v}_k\}\big)$
$A\vec{x} \leq \vec{b}$	convex space	convex linear combinations $a_1\vec{v}_1 + \cdots + a_k\vec{v}_k$ with $a_1, \ldots, a_k \in \mathbb{R}$ satisfying $\sum_{i=1}^{k} a_i = 1$ and $a_1, \ldots, a_k \geq 0$ called the convex hull of $\{\vec{v}_1, \ldots, \vec{v}_k\}$ written $\operatorname{conv}\big(\{\vec{v}_1, \ldots, \vec{v}_k\}\big)$

Recommended Exercises. Exercises 5, 11, 12 in Section 8.3 and 1 in Section 8.5 of [3]. Exercises 1 and 15 in Section 9.2 of [2]. Be able to show all your work, step by step!

In this lecture, we went through parts of Sections 8.3, 8.5, and 9.2 of [Lay].

Terminology checklist

the standard simplex	
convex space	
linear system of inequalities	
linear functional	
bounded	
unbounded	
extreme point	
convex combination	
convex hull	

7 Linear transformations and their matrices

In the context of linear programming, we discussed the notion of a linear functional. These are special cases of linear transformations, which can be thought of as families of linear functionals. Linear transformations arise in many familiar situations.

Problem 7.1. Mark makes three types of sandwiches at the deli: BLT (bacon, lettuce, and tomato), HS (ham and swiss), and MS (meatball sub). It takes 3 minutes to make the BLT, 2 minutes to make the HS, and 1 minute to make the MS. The profit made is \$2 from the BLT, \$1 from the HS, and \$1 from the MS. Last Tuesday, Mark made 40 BLT's, 30 HS's, and 50 MS's. How much profit did Mark make for the deli? How much time did Mark spend making sandwiches?

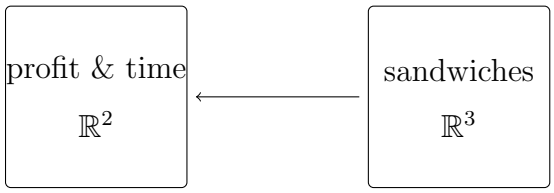
Answer. Let x, y, and z denote the number of BLT, HS, and MS, respectively, sold. The profit p, in dollars, made as a function of x, y, z is given by $p(x, y, z) = 2x + y + z$. The time t, in minutes, it took to make the sandwiches as a function of x, y, z is $t(x, y, z) = 3x + 2y + z$. We can put these two quantities together as a 2-component vector

$$\begin{bmatrix} p \\ t \end{bmatrix} = \begin{bmatrix} 2x + y + z \\ 3x + 2y + z \end{bmatrix} = x \begin{bmatrix} 2 \\ 3 \end{bmatrix} + y \begin{bmatrix} 1 \\ 2 \end{bmatrix} + z \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{7.2}$$

In our case, $x = 40, y = 30, z = 50$. Therefore,

$$\begin{bmatrix} p \\ t \end{bmatrix} = 40 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + 30 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 50 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 160 \\ 230 \end{bmatrix} \tag{7.3}$$

so that Mark will have made the deli a \$160 profit in 230 minutes, or roughly just under 4 hours. Notice that this procedure works for any x, y, z. In other words, the two linear equations describing the profit and time can be viewed as transforming the quantities of different sandwich types being made into their associated profit and time taken



because

$$x \begin{bmatrix} 2 \\ 3 \end{bmatrix} + y \begin{bmatrix} 1 \\ 2 \end{bmatrix} + z \begin{bmatrix} 1 \\ 1 \end{bmatrix} \longleftarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{7.4}$$

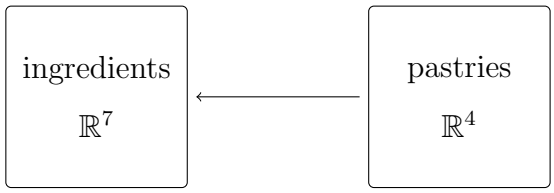
Example 7.5. A bakery produces pancakes, tres leches cake, strawberry shortcake, and egg tarts. Each of these items requires several ingredients, some of which are listed in the table below.

	Pancakes (makes 12)	Tres leches (makes 8 slices)	Strawberry shortcake (makes 8 slices)	Egg tarts (makes 16 tarts)
eggs	2	6	0	6
strawberries	$1+\frac{1}{2}$ lbs	0	$1+\frac{1}{2}$ lbs	0
heavy cream	1 cup	1 pint	3 cups	0
flour	3 cups	1 cup	4 cups	$3+\frac{3}{4}$ cups
butter	4 tbsp	0	$1+\frac{1}{4}$ cups	$1+\frac{1}{3}$ cups
sugar	3 tbsp	1 cup	$\frac{1}{2}$ cup	$\frac{2}{5}$ cup
milk	2 cups	$\frac{1}{3}$ cup	0	$\frac{1}{3}$ cup

How much flour is needed to make 4 batches of pancakes, 3 batches of tres leches cakes, 2 batches of strawberry shortcakes, and 4 batches of egg tarts? To answer this, we would simply multiply the number of batches for each item by the amount of flour needed for that item, and then add up all these quantities for the different recipes:

$$4\times 3+3\times 1+2\times 4+4\times\left(3+\frac{3}{4}\right)=38 \tag{7.6}$$

cups of flour. This procedure of decomposing a pastry into its ingredients in this way can be described by a transformation



On the right, we view the four different pastry items that we can think of as four linearly independent vectors. This is because no item, once made, can be obtained from the other items as a linear combination (we cannot take an egg tart, two strawberry shortcakes, and one tres leches cake, and turn them into a pancake). On the left, we view the seven different ingredients as seven linearly independent vectors. Of course, milk, butter, and heavy cream are closely related, but let's say that we do not want to put in the extra effort to turn one of them into another form.

Now let's try to provide a general abstract formula for which we can simply plug in numbers when needed. How much flour is needed to make p batches of pancakes, t batches of tres leches cakes, s batches of strawberry shortcakes, and e batches of egg tarts? Let F denote cups of flour so that F is a function of p, t, s, and e. The idea is similar to above and the flour function is

$$F(p,t,s,e)=3p+t+4s+\frac{15}{4}e. \tag{7.7}$$

Notice that the flour function is an example of a linear functional (see Definition 6.21). Similar equations can be written for the other ingredients. To describe the full transformation, we would have to write a list of such linear functionals for each ingredient. The functions for eggs, strawberries (by pounds), heavy cream (by cups), butter (by cups), sugar (by cups), and milk (by cups), respectively, would also all be linear functionals. When put together, they define a *linear transformation.*

Example 7.8. Figure 10 shows the difference between what an individual with protanopia sees (on the left) and what an individual without any colorblindness sees (on the right).[34] Since each

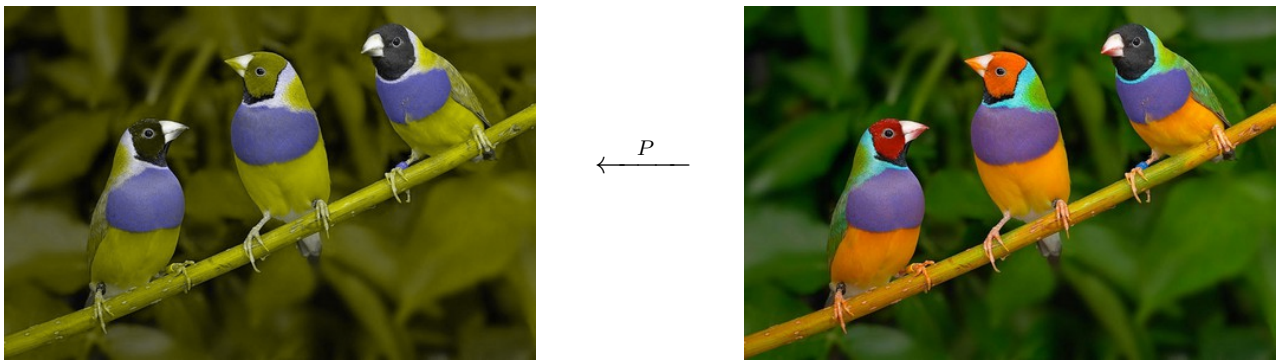
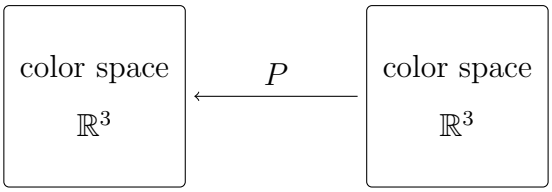
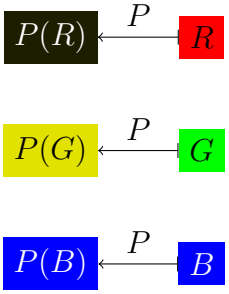


Figure 10: Protanopia colorblindness.

color can be identified numerically, by testing out several points, we can find out how each color changes. We can then hope to describe the relationship between normal eyesight and protanopia as a type of transformation



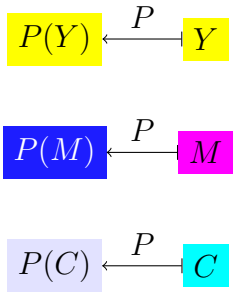
which we have called P. It is far from obvious, but it is true (to a reasonable degree),[35] that such a transformation is linear. In particular, this means that it can be determined by just knowing where three linearly independent colors go (recall Example 4.33). Under this transformation, the images of the pure RGB colors look like:



For the YMC colors, the transformation looks like:

[34]These figures were obtained from https://ssodelta.wordpress.com/tag/rgb-to-lms/ and the original photo is from https://animals.desktopnexus.com/wallpaper/480778/. If you have protanopia, you should see no difference between these images (note: I have done this experiment with somebody who has protanopia and although some of this is correct, not all of it seems to be correct... this requires further investigation).

[35]The filter might not be exactly linear. I need to do some research to figure out if there are corrections to linearity. Hence, take this with a grain of salt.



Example 7.9. An experiment[36] was done in 1926 to determine which color paint on walls helps a baby sleep more. In this study, it was not only found that different color paints are more conducive to healthier sleeping habits, but also that different genders were affected by colors differently. Table 1 shows the fractions of babies that had the healthiest sleeping habits to the corresponding color paints. A couple visited the doctor, who, upon analyzing their DNA, indicated that their odds of

	peach	lavender	sky blue	light green	light yellow
boy	0.15	0.2	0.2	0.2	0.25
girl	0.3	0.2	0.25	0.1	0.15

Table 1: Percentages for a study examining the healthiest sleeping habits for baby boys and girls depending on the color used for painting walls in a baby's room.

giving birth to a baby boy is actually 60%. The couple wants to finish the paint job in the baby's room well before the baby is born. What color should they paint their walls?

For this situation, we multiply all the respective probabilities for a boy by 0.6 and for a girl by 0.4 and then sum the results as in Table 2. The highest percentage occurs for sky blue. Hence,

	peach	lavender	sky blue	light green	light yellow
boy	0.09	0.12	0.12	0.12	0.15
girl	0.12	0.08	0.1	0.04	0.06
sum	0.21	0.2	0.22	0.16	0.21

Table 2: Percentages for 60% chance of giving birth to a boy

the couple should paint the room sky blue.

How would these results change if the doctor told them they are actually only 40% likely to give birth to a boy? For this situation, we multiply all the respective probabilities for a boy by 0.4 and for a girl by 0.6 and then sum the results as in Table 3. In this case, peach wins.

Definition 7.10. A *linear transformation* (sometimes called an *operator*) from $\mathbb{R}^n$ to $\mathbb{R}^m$ is an assignment, denoted by T, sending any vector $\vec{x}$ in $\mathbb{R}^n$ to a unique vector $T(\vec{x})$ in $\mathbb{R}^m$ satisfying

$$T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y}) \tag{7.11}$$

[36] This is an example of a poorly designed experiment, but let's just go with it...

	peach	lavender	sky blue	light green	light yellow
boy	0.06	0.08	0.08	0.08	0.1
girl	0.18	0.12	0.15	0.06	0.09
sum	0.24	0.2	0.23	0.14	0.19

Table 3: Percentages for 40% chance of giving birth to a boy

and

$$T(c\vec{x}) = cT(\vec{x}) \tag{7.12}$$

for all $\vec{x}, \vec{y}$ in $\mathbb{R}^n$ and all c in $\mathbb{R}$. Such a linear transformation can be written in any of the following ways[37]

$$T : \mathbb{R}^n \to \mathbb{R}^m, \qquad \mathbb{R}^n \xrightarrow{T} \mathbb{R}^m, \qquad \mathbb{R}^m \leftarrow \mathbb{R}^n : T, \qquad \text{or} \qquad \mathbb{R}^m \xleftarrow{T} \mathbb{R}^n. \tag{7.13}$$

Given a vector $\vec{x}$ in $\mathbb{R}^n$ and a linear operator $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$, the vector $T(\vec{x})$ in $\mathbb{R}^m$ is called the *image* of $\vec{x}$ under T. $\mathbb{R}^n$ is called the *domain* or *source* of T and $\mathbb{R}^m$ is called the *codomain* or *target*. The image of *all* vectors in $\mathbb{R}^n$ under T is called the *range* of T, i.e.

$$\text{image}(T) \equiv \text{range}(T) := \left\{ T(\vec{x}) \in \mathbb{R}^m \ : \ \vec{x} \in \mathbb{R}^n \right\}. \tag{7.14}$$

Do not confuse range with codomain/target. The range of a function is all the elements that are "hit" by the function, whereas the codomain/target just notifies the reader what kinds of values the function takes. For example, we say that the sin function is a real-valued function meaning that $\sin : \mathbb{R} \to \mathbb{R}$, but we know that sin cannot take on all values. sin can only take on values between -1 and 1 so we say its codomain/target is $[-1, 1]$, the interval between -1 and 1.

A linear transformation is completely determined by what it does to a basis.

Example 7.15. In Example 7.5, we only needed to know the values of p, t, s, and e to determine the amount of flour needed. The four pastry items pancakes, tres leches, strawberry shortcakes, and egg tarts, form a basis in the sense that no one of these items can be obtained from any combination of any other (linear independence) and all pastry items obtainable are precisely these (they span the possible products of the bakery). Let E, T, H, F, B, S, M denote the functions for eggs, strawberries (by pounds), heavy cream (by cups), flour (by cups), butter (by cups), sugar (by cups), and milk (by cups), respectively. Given any value of p, t, s, and e, which can be viewed as a vector in $\mathbb{R}^4$ as

$$\begin{bmatrix} p \\ t \\ s \\ e \end{bmatrix}, \tag{7.16}$$

[37] My preference are the second and last.

the resulting ingredients are given as a vector in $\mathbb{R}^7$ as

$$\begin{bmatrix} E \\ T \\ H \\ F \\ B \\ S \\ M \end{bmatrix} = p \begin{bmatrix} 2 \\ 3/2 \\ 1 \\ 3 \\ 1/4 \\ 3/16 \\ 2 \end{bmatrix} + t \begin{bmatrix} 6 \\ 0 \\ 2 \\ 1 \\ 0 \\ 1 \\ 1/3 \end{bmatrix} + s \begin{bmatrix} 0 \\ 3/2 \\ 3 \\ 4 \\ 5/4 \\ 1/2 \\ 0 \end{bmatrix} + e \begin{bmatrix} 6 \\ 0 \\ 0 \\ 15/4 \\ 4/3 \\ 2/5 \\ 1/3 \end{bmatrix} \tag{7.17}$$

We can then use the notation $A\vec{x} = \vec{b}$ to express this as

$$\begin{bmatrix} 2 & 6 & 0 & 6 \\ 3/2 & 0 & 3/2 & 0 \\ 1 & 2 & 3 & 0 \\ 3 & 1 & 4 & 15/4 \\ 1/4 & 0 & 5/4 & 4/3 \\ 3/16 & 1 & 1/2 & 2/5 \\ 2 & 1/3 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} p \\ t \\ s \\ e \end{bmatrix} = \begin{bmatrix} E \\ T \\ H \\ F \\ B \\ S \\ M \end{bmatrix}. \tag{7.18}$$

Example 7.19. Colorblindness can be modeled in terms of a linear transformation.[38] Referring back to Example 4.33 for notation, protanopia (a type of colorblindness) can be described in terms of a linear transformation. To obtain this linear transformation, we should write the corresponding RGB codes for what happens to R, G, and B when we view these colors as vectors in $\mathbb{R}^3$. These are given by

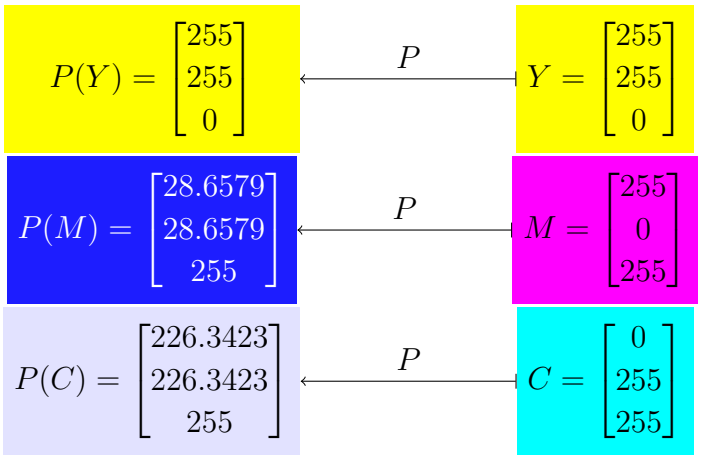
$$P(R) = \begin{bmatrix} 28.6579 \\ 28.6579 \\ 0 \end{bmatrix} \xleftarrow{P} R = \begin{bmatrix} 255 \\ 0 \\ 0 \end{bmatrix}$$

$$P(G) = \begin{bmatrix} 226.3423 \\ 226.3423 \\ 0 \end{bmatrix} \xleftarrow{P} G = \begin{bmatrix} 0 \\ 255 \\ 0 \end{bmatrix}$$

$$P(B) = \begin{bmatrix} 0 \\ 0 \\ 255 \end{bmatrix} \xleftarrow{P} B = \begin{bmatrix} 0 \\ 0 \\ 255 \end{bmatrix}$$

and for the YMC colors, the transformation looks like:

[38] The following is based on my understanding of the information on the website https://ssodelta.wordpress.com/tag/rgb-to-lms/.



From where R,G, and B go, the matrix

$$P := \begin{bmatrix} 0.112384 & 0.887617 & 0 \\ 0.112384 & 0.887617 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.20}$$

describes the filter for protanopia because

$$\begin{bmatrix} R_{\text{new}} \\ G_{\text{new}} \\ B_{\text{new}} \end{bmatrix} = \begin{bmatrix} 0.112384 & 0.887617 & 0 \\ 0.112384 & 0.887617 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = R \begin{bmatrix} 0.112384 \\ 0.112384 \\ 0 \end{bmatrix} + G \begin{bmatrix} 0.887617 \\ 0.887617 \\ 0 \end{bmatrix} + B \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{7.21}$$

For example, let's calculate $P($ $)$ where

$$\quad = \begin{bmatrix} 251 \\ 212 \\ 185 \end{bmatrix}. \tag{7.22}$$

By linearity,

$$\begin{aligned} P\left(\begin{bmatrix} 251 \\ 212 \\ 185 \end{bmatrix}\right) &= P\left(\begin{bmatrix} 251 \\ 0 \\ 0 \end{bmatrix}\right) + P\left(\begin{bmatrix} 0 \\ 212 \\ 0 \end{bmatrix}\right) + P\left(\begin{bmatrix} 0 \\ 0 \\ 185 \end{bmatrix}\right) \\ &= \begin{bmatrix} 28 \\ 28 \\ 0 \end{bmatrix} + \begin{bmatrix} 188 \\ 188 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 185 \end{bmatrix} \\ &= \begin{bmatrix} 216 \\ 216 \\ 185 \end{bmatrix}. \end{aligned} \tag{7.23}$$

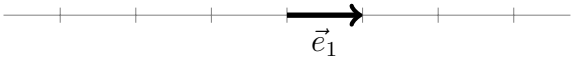
Using this notation, we can now make sense of linear systems and why they are expressed in the form $A\vec{x} = \vec{b}$.

Definition 7.24. Let $n \in \mathbb{N}$ be a positive integer and let $i \in \{1, \ldots, n\}$. The *i-th unit vector in* $\mathbb{R}^n$ is the vector

$$\vec{e}_i := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{-th entry}, \tag{7.25}$$

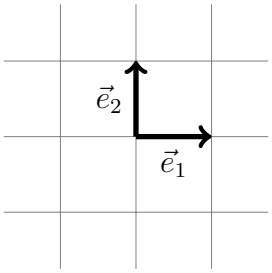
which is 0 in every entry except the i-th entry, where it is 1.

Example 7.26. In $\mathbb{R}$, there is only one unit vector, $\vec{e}_1$, and it can be drawn as



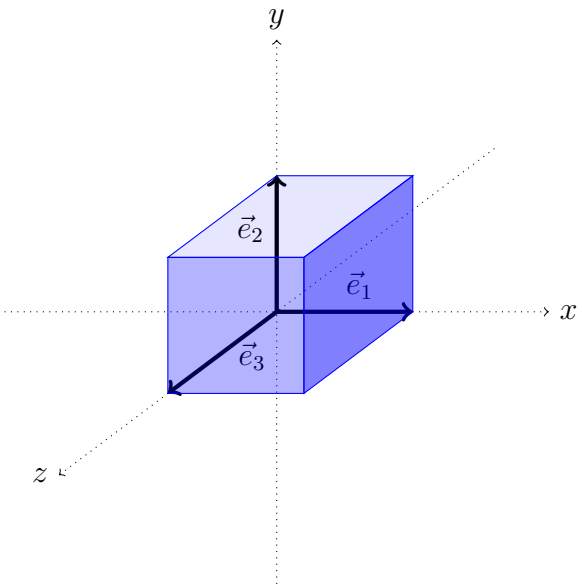
The distance between each pair of consecutive tick marks is 1.

Example 7.27. In $\mathbb{R}^2$, $\vec{e}_1$ and $\vec{e}_2$ can be drawn as



Example 7.28. In $\mathbb{R}^3$, all of the following notations are used to express certain vectors

$$\vec{e}_1 = \hat{x} = \hat{\imath} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \vec{e}_2 = \hat{y} = \hat{\jmath} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \& \qquad \vec{e}_3 = \hat{z} = \hat{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{7.29}$$



The blue box is merely drawn to help the viewer in visualizing the coordinates in three dimensions.

Definition 7.30. Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. The *$m \times n$ matrix associated to T* is the $m \times n$ array of numbers whose entries are given by

$$\begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix}. \tag{7.31}$$

More explicitly, if the vector $T(\vec{e}_i)$ is written as

$$T(\vec{e}_i) = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{bmatrix}, \tag{7.32}$$

then

$$\begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{7.33}$$

The relationship between matrices and augmented matrices from before is given as follows. An augmented matrix of the form (1.47) corresponding to a linear system (1.27) with variables $x_1, \ldots, x_n$, can be expressed as

$$A\vec{x} = \vec{b}, \tag{7.34}$$

where the notation $A\vec{x}$ stands for the *vector*[39]

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} := \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} \tag{7.35}$$

in $\mathbb{R}^m$. Notice that this vector can be decomposed, by factoring out the common factors, as

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}. \tag{7.36}$$

This is nothing more than the linearity of T expressed in matrix form. Equivalently, this equation can be written as

$$\begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 T(\vec{e}_1) + \cdots + x_n T(\vec{e}_n). \tag{7.37}$$

[39] The vector on the right-hand-side is a definition of the notation on the left-hand-side. Don't be confused by the fact that there are a lot of terms inside each component of the vector on the right-hand-side of (7.35)—it is *not* an $m \times n$ matrix!

In this way, we see the columns of the matrix A more clearly. Furthermore, an $m \times n$ matrix can be viewed as having an existence independent of a linear transformation, at least a-priori. Therefore, an $m \times n$ matrix acts on a vector in $\mathbb{R}^n$ to produce a vector in $\mathbb{R}^m$. This is a way of consolidating the augmented matrix and A is precisely the matrix corresponding to the linear system. One can express the matrix A as a row of column vectors

$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix} \tag{7.38}$$

where the i-th component of the j-th vector $\vec{a}_j$ is given by

$$(\vec{a}_j)_i = a_{ij}. \tag{7.39}$$

In this case, $\vec{b}$ is explicitly expressed as a linear combination of the vectors $\vec{a}_1, \ldots, \vec{a}_n$ via

$$\vec{b} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n. \tag{7.40}$$

Therefore, solving for the variables $x_1, \ldots, x_n$ for the linear system (1.27) is equivalent to finding coefficients $x_1, \ldots, x_n$ that satisfy (7.40). $A\vec{x} = \vec{b}$ is called a *matrix equation*.

Warning: we do not provide a definition for an $m \times n$ matrix acting on a vector in $\mathbb{R}^k$ with $k \neq n$.

Thus, there are three equivalent ways to express a linear system.

(a) m linear equations in n variables (1.27).

(b) An augmented matrix (1.47).

(c) A matrix equation $A\vec{x} = \vec{b}$ as in (7.35).

The above observations also lead to the following.

Theorem 7.41. *Let A be a fixed $m \times n$ matrix. The following statements are equivalent (which means that any one implies the other and vice versa).*

(a) For every vector $\vec{b}$ in $\mathbb{R}^m$, the solution set of the equation $A\vec{x} = \vec{b}$, meaning the set of all $\vec{x}$ satisfying this equation, is nonempty.

(b) Every vector $\vec{b}$ in $\mathbb{R}^m$ can be written as a linear combination of the columns of A, viewed as vectors in $\mathbb{R}^m$, i.e. the columns of A span $\mathbb{R}^m$.

(c) A has a pivot position in every row.

Proof. Let's just check part of the equivalence between (a) and (b) by showing that (b) implies (a). Suppose that a vector $\vec{b}$ can be written as a linear combination

$$\vec{b} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n, \tag{7.42}$$

where the $\{x_1, \ldots, x_n\}$ are some coefficients. Rewriting this using column vector notation gives

$$\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = x_1 \begin{bmatrix} (a_1)_1 \\ \vdots \\ (a_1)_m \end{bmatrix} + \cdots + x_n \begin{bmatrix} (a_n)_1 \\ \vdots \\ (a_n)_m \end{bmatrix} \tag{7.43}$$

We can set our notation and write

$$(a_j)_i \equiv a_{ij}. \tag{7.44}$$

Then, writing out this equation of vectors gives

$$\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} x_1 a_{11} + \cdots + x_n a_{1n} \\ \vdots \\ x_1 a_{m1} + \cdots + x_n a_{mn} \end{bmatrix} \tag{7.45}$$

by the rules about scaling and adding vectors from last lecture. The resulting equation is exactly the linear system corresponding to $A\vec{x} = \vec{b}$. Hence, the x's from the linear combination in (7.42) give a solution of the matrix equation $A\vec{x} = \vec{b}$. ■

Theorem 7.46. *Let A be an $m \times n$ matrix, let $\vec{x}$ and $\vec{y}$ be two vectors in $\mathbb{R}^n$, and let c be any real number. Then*

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} \qquad \& \qquad A(c\vec{x}) = cA\vec{x}. \tag{7.47}$$

In other words, every $m \times n$ matrix determines a linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$.

Exercise 7.48. Prove this! To do this, write out an arbitrary A matrix with entries as in (7.35) along with two vectors $\vec{x}$ and $\vec{y}$ and simply work out both sides of the equation using the rule in (7.35).

Recommended Exercises. Exercises 4, 13 (there is a typo in the 5th edition: you can ignore the symbol $\mathbb{R}^3$), 17, 25, in Section 1.4 of [Lay], Exercises 10, 12, 17, 18, 27 (26)—this is the line segment problem, 31, 33 in Section 1.8 of [Lay], and Exercise 3 in Section 1.10 of [Lay]. Be able to show all your work, step by step!

In this lecture, we went through parts of Sections 1.4, 1.8, 1.9, and 1.10 of [Lay].

Terminology checklist

linear transformation	
domain/source	
codomain/target	
image/range	
standard unit vectors $\vec{e}_i$	
equivalent (system)	
matrix associated to a linear transformation	
matrix equation	

8 Visualizing linear transformations

Every $m \times n$ matrix A acts on a vector $\vec{x}$ in $\mathbb{R}^n$ and produces a vector $\vec{b}$ in $\mathbb{R}^m$ as in

$$A\vec{x} = \vec{b}. \tag{8.1}$$

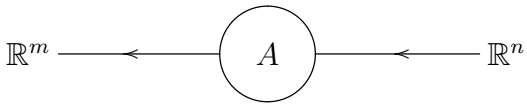
Furthermore, a matrix acting on vectors in $\mathbb{R}^n$ in this way satisfies the following two properties

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} \tag{8.2}$$

and

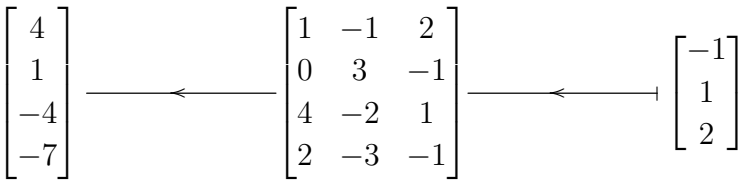
$$A(c\vec{x}) = cA\vec{x} \tag{8.3}$$

for any other vector $\vec{y}$ in $\mathbb{R}^n$ and any scalar c. Since $\vec{x}$ is arbitrary, we can think of A as an *operation* that acts on all of $\mathbb{R}^n$. Any time you input a vector in $\mathbb{R}^n$, you get out a vector in $\mathbb{R}^m$. We can depict this diagrammatically as



(8.4)

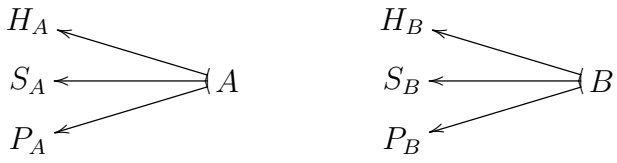
You will see right now (and several times throughout this course) why we write the arrows from right to left (your book does not, which I personally find confusing).[40] For example,[41]



(8.5)

is a 4×3 matrix (in the middle) acting on a vector in $\mathbb{R}^3$ (on the right) and producing a vector in $\mathbb{R}^4$ (on the left).

Example 8.6. In Exercise 1.3.28 in [Lay], two types of coal, denoted by A and B, respectively, produce a certain amount of heat (H), sulfur dioxide (S), and pollutants (P) based on the quantity of input for the two types of coal. Let H_A, S_A, and P_A denote these quantities for one ton of A and let H_B, S_B, and P_B denote these quantities for one ton of B. Visually, these can be described as a linear transformation



(8.7)

and the matrix associated to this transformation is

$$\begin{bmatrix} H_A & H_B \\ S_A & S_B \\ P_A & P_B \end{bmatrix} \tag{8.8}$$

[40] It doesn't matter how you draw it as long as you are consistent and you know what it means. It's not a 'rule' and only my preference.

[41] We use arrows with a vertical dash as in $\leftarrow\!\!|$ at the beginning when we act on specific vectors.

and it acts on vectors of the form

$$\begin{bmatrix} x \\ y \end{bmatrix} \tag{8.9}$$

where x is the number of tons of coal of type A and y is the number of tons of coal of type B. The rows of the matrix describe the type of output while the columns correspond to all outputs due to a given input (the type of coal used). Indeed, the net output given x tons of A and y tons of B is

$$\begin{bmatrix} H_A & H_B \\ S_A & S_B \\ P_A & P_B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} xH_A + yH_B \\ xS_A + yS_B \\ xP_A + yP_B \end{bmatrix} = x \begin{bmatrix} H_A \\ S_A \\ P_A \end{bmatrix} + y \begin{bmatrix} H_B \\ S_B \\ P_B \end{bmatrix} \tag{8.10}$$

as you probably already know from doing that exercise. The rows in the resulting vector correspond to the total heat, sulfur dioxide, and pollutant outputs, respectively. But thinking of the transformation (8.7) abstractly without matrices, it can be viewed as a linear transformation without reference to any given set of vectors. Abstractly, the transformation of the power plant produces 3 outputs (heat, sulfur dioxide, and pollutants) from 2 inputs (the two types of coal used).

From the above discussion, every $m \times n$ matrix is an example of a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. In the example above, namely (10.34), the image of

$$\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \tag{8.11}$$

under the linear operator given by the matrix

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & -1 \\ 4 & -2 & 1 \\ 2 & -3 & -1 \end{bmatrix} \tag{8.12}$$

is

$$\begin{bmatrix} 4 \\ 1 \\ -4 \\ -7 \end{bmatrix}. \tag{8.13}$$

Notice that the operator can act on any other vector in $\mathbb{R}^3$ as well, not just the particular choice we made. So for example, the image of

$$\begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix} \tag{8.14}$$

would be

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & -1 \\ 4 & -2 & 1 \\ 2 & -3 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} -5 \\ 10 \\ -7 \\ -8 \end{bmatrix}. \tag{8.15}$$

Maybe now you see why we wrote our arrows from right to left. It makes acting on the vectors with the matrix much more straightforward (as written on the page). If we didn't, we would have to flip the vector to the other side of the matrix every time to calculate the image. In this calculation, we showed

$$\begin{bmatrix} -5 \\ 10 \\ -7 \\ -8 \end{bmatrix} \longleftarrow \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & -1 \\ 4 & -2 & 1 \\ 2 & -3 & -1 \end{bmatrix} \longleftarrow\shortmid \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}. \tag{8.16}$$

Notice that the center matrix always stays the same no matter what vectors in $\mathbb{R}^3$ we put on the right. The matrix in the center is a rule that applies to *all* vectors in $\mathbb{R}^3$. When the matrix changes, the rule changes, and we have a different linear transformation.

Example 8.17. Consider the transformation that multiplies every vector by 2. Under this transformation, the vector

$$\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \tag{8.18}$$

gets sent to

$$\begin{bmatrix} 2 \\ 4 \\ 4 \end{bmatrix} \tag{8.19}$$

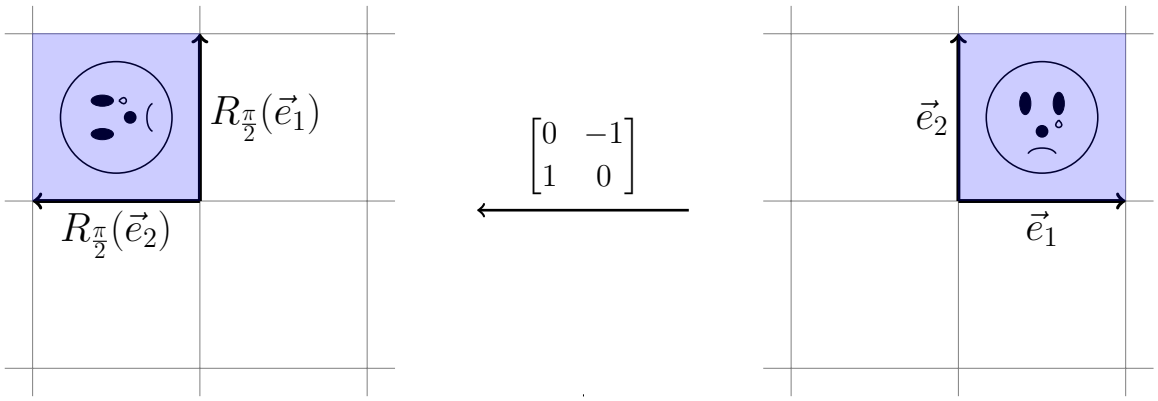
This transformation is linear and the matrix representing it is

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \tag{8.20}$$

Example 8.21. Let θ be some angle in $[0, 2\pi)$. Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates (counter-clockwise) all the vectors in the plane by θ degrees (for the pictures, let's say $\theta = \frac{\pi}{2}$). This transformation is linear and is represented by the matrix

$$R_\theta := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{8.22}$$

For $\theta = \frac{\pi}{2}$, this looks like



Visually, it is not difficult to believe that rotation by an angle θ is a linear transformation. However, to prove it is a bit non-trivial.

Problem 8.23. Prove that $\mathbb{R}^2 \xleftarrow{R_\theta} \mathbb{R}^2$, defined by rotating all vectors by θ, is a linear transformation.

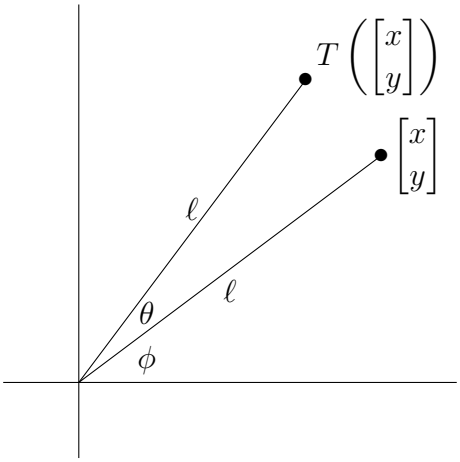
Answer. We will not prove this here as it is a homework problem, but we will set it up so that you know what is involved in proving such a claim. Any vector in $\mathbb{R}^2$ can be expressed in the following two ways

$$\begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{8.24}$$

where $x, y \in \mathbb{R}$ are the coordinates of the vector. First show that

$$R_\theta \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = x R_\theta \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) + y R_\theta \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \tag{8.25}$$

for all $x, y \in \mathbb{R}$. Therefore, you must calculate each side of this equality and prove that the results you obtain are the same. To do this, use trigonometry. Let's start this off by drawing a picture.



In this picture, ϕ is defined as the angle obtained from the vector $\begin{bmatrix} x \\ y \end{bmatrix}$ and ℓ is the length of this vector (use Pythagorean's theorem). The angle θ is the additional angle that we rotate this vector by. Once we have rotated by this angle, the total angle that we have is $\theta+\phi$. Hence, the coordinates of $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$ are given by $\begin{bmatrix} \ell\cos(\theta+\phi) \\ \ell\sin(\theta+\phi) \end{bmatrix}$. Use trigonometric identities to simplify this expression so that the answer is purely in terms of x, y, and θ. Then show that the resulting expression equals the right-hand-side of (8.25). The fact that this is a linear transformation will follow from this equality upon further work (see the HW assignment for hints).

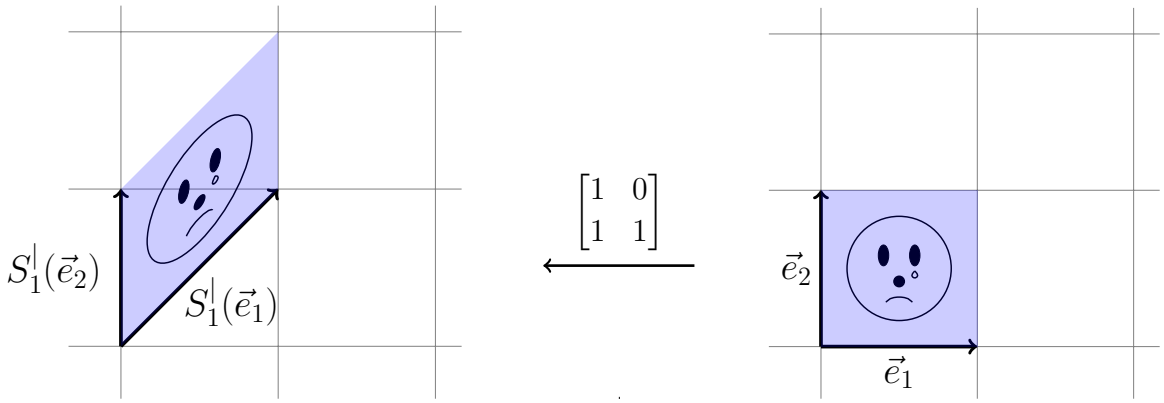
Example 8.26. A vertical shear in $\mathbb{R}^2$ is given by a matrix of the form

$$S_k^| := \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \tag{8.27}$$

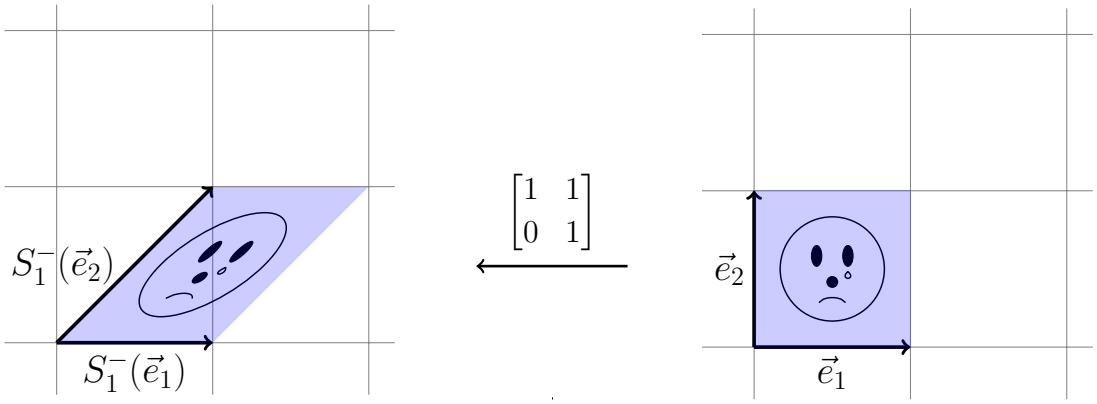
while a horizontal shear is given by a matrix of the form

$$S_k^- := \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \tag{8.28}$$

where k is a real number. When $k = 1$, the former is depicted by



while the latter is depicted by



Example 8.29. Many more examples are given in Section 1.9 of [Lay]. You should be comfortable with all of them!

Recommended Exercises. Exercises 6 (8) and 13 in Section 1.9 of [Lay]. Many of the Chapter 1 Supplementary Exercises are good as well! Be able to show all your work, step by step!

In this lecture, we finished Sections 1.4, and 1.9 of [Lay]. We still have a few concepts to cover from Section 1.8.

Terminology checklist

rotation	
vertical shear	
horizontal shear	
reflection across a line	

9 Subspaces associated to linear transformations

Definition 9.1. $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation with associated $m \times n$ matrix denoted by A. The *kernel* of T is the set of all vectors $\vec{x} \in \mathbb{R}^n$ such that $T(\vec{x}) = \vec{0}$

$$\ker(T) = \left\{ \vec{x} \in \mathbb{R}^n \ : \ T(\vec{x}) = \vec{0} \right\}. \tag{9.2}$$

Equivalently, the *null space* of A is the set of all solutions to the homogeneous equation

$$A\vec{x} = \vec{0}. \tag{9.3}$$

Problem 9.4. Jake bought stocks A and B in 2013 at a cost of C_A and C_B per stock, respectively. He spent a total of \$10,000. In 2017, he sold the stocks at a selling price of S_A and S_B per stock, respectively. Suppose that $S_B \neq C_B$ and $S_A \neq C_A$. In the end, he broke even, because he was a scrub and didn't diversify his assets. How many of each stock did Jake buy? What if he initially spent \$15,000 and still broke even?

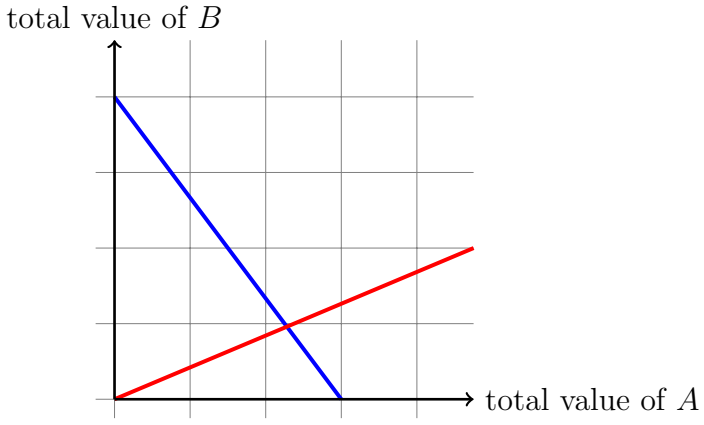
Answer. Let x and y denote the number of stocks (possibly not a whole number) of A and B that Jake had purchased in 2013. Because he spent \$10,000,

$$xC_A + yC_B = 10000. \tag{9.5}$$

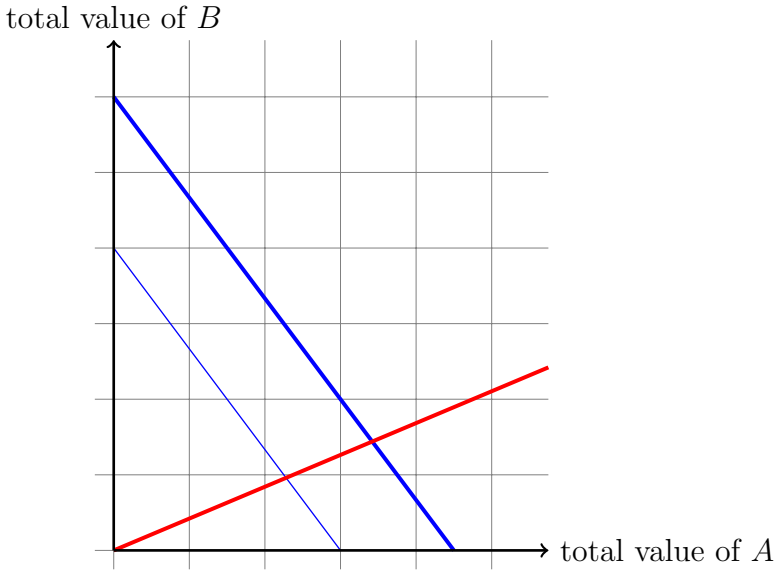
Because he broke even, his profit function $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \ni (x, y) \mapsto p(x, y)$ satisfies

$$x(S_A - C_A) + y(S_B - C_B) = 0. \tag{9.6}$$

The different possible combinations of purchasing stocks A and B and breaking even describes the kernel of the profit function. These two equations describe a linear manifold and a subspace of $\mathbb{R}^2$ sketched as follows



The intersection of these two lines indicates the quantity of stocks that were purchased. If Jake had spent \$15,000, only the first equation would change, and this would merely shift the blue line



Kernels can also be used to describe that information is lost in some sense. This will be discussed more precisely in Definition 9.43.

Example 9.7. Consider the example of protanopia colorblindness. The kernel of the protanopia filter P can be calculated by solving

$$\left[\begin{array}{ccc|c} 0.112384 & 0.887617 & 0 & 0 \\ 0.112384 & 0.887617 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 0.112384 & 0.887617 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]. \tag{9.8}$$

Therefore, the kernel is the set of vectors of the form

$$G \begin{bmatrix} -7.89807 \\ 1 \\ 0 \end{bmatrix}, \tag{9.9}$$

where G is a free variable. As you can tell, the only solution that physically makes sense is when $G = 0$ since colors cannot be chosen to be negative. Hence, although the kernel associated to the linear transformation P is spanned by the vector

$$\begin{bmatrix} -7.89807 \\ 1 \\ 0 \end{bmatrix} \tag{9.10}$$

no multiple of this vector intersects the set of allowed color values. So is there any information actually lost? Is it possible for a "reverse filter" to be applied to somebody with protanopia so that they *can* see in full color? I'll let you think about it.

Theorem 9.11. *The kernel (null space) of a linear transformation* $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ *is a subspace of* $\mathbb{R}^n$.

Proof. We must check the axioms of a subspace.

(a) The zero vector satisfies $T(\vec{0}) = \vec{0}$ because $T(\vec{0}) = T(0\vec{0}) = 0T(\vec{0}) = \vec{0}$ since 0 times any vector is the zero vector. Linearity of T was used in the second equality.

(b) Let $\vec{x} \in \ker(T)$ and let $c \in \mathbb{R}$. Then, $T(c\vec{x}) = cT(\vec{x}) = c\vec{0} = \vec{0}$. The first equality follows from linearity of T.

(c) Let $\vec{x}, \vec{y} \in \ker(T)$. Then $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}) = \vec{0} + \vec{0} = \vec{0}$. The first equality follows from linearity of T.

■

There is actually an important consequence in the above proof. We will illustrate why this is so in a short example.

Corollary 9.12. *Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. Then $T(\vec{0}) = \vec{0}$.*

This is important because it provides one quick method of showing that certain functions are not linear transformations. This is because this corollary says that it is *necessary* (i.e. it must be the case) that $\vec{0}$ is in the kernel for *every* linear transformation. In other words, if you show that $\vec{0}$ is not in the kernel of some function, then that means it *cannot* be linear.

Problem 9.13. Let $\mathbb{R}^3 \xleftarrow{T} \mathbb{R}^3$ be the function defined by

$$\mathbb{R}^3 \ni (x, y, z) \mapsto T(x, y, z) := (x + y - z, 2x - 3y + 2, 3x - 5z). \tag{9.14}$$

Show that T is not a linear transformation.

Answer. $T(0) = (0, 2, 0)$ so T is not a linear transformation.

Warning: showing that $T(\vec{0}) = \vec{0}$ does *not* mean that the function is linear.

Problem 9.15. Let $\mathbb{R}^3 \xleftarrow{T} \mathbb{R}^2$ be the function defined by

$$\mathbb{R}^3 \ni (x, y) \mapsto T(x, y) := \big((2 - y)x, x + 3y, 2x - y\big). \tag{9.16}$$

Show that T is not a linear transformation.

Notice that $T(0, 0) = (0, 0, 0)$ so this does not help in showing that T is not linear.

Answer. $2T(1, 1) = 2(1, 4, 1) = (2, 8, 2)$ while $T(2, 2) = (0, 8, 2)$. Since $2T(1, 1) \neq T(2, 2)$, T is not a linear transformation.

Note that all we had to show was one instance where linearity failed. Linearity is supposed to hold for *all* inputs so if we find just *one* case where it fails, the function cannot be linear. We now go to illustrating some examples of Theorem 9.11.

Example 9.17. Consider the linear system

$$3x - 2y + z = 0 \tag{9.18}$$

from an earlier section. The matrix corresponding to this linear system is just

$$A = \begin{bmatrix} 3 & -2 & 1 \end{bmatrix}, \tag{9.19}$$

a 1×3 matrix. Hence, it describes a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^1$. The nullspace of A exactly corresponds to the solutions of

$$\begin{bmatrix} 3 & -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}. \tag{9.20}$$

Definition 9.21. Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation with associated $m \times n$ matrix denoted by A. The *image* (also called *range*) of T is the set of all vectors in $\mathbb{R}^m$ of the form $T(\vec{x})$ with $\vec{x}$ in $\mathbb{R}^n$. Equivalently, the *column space* of A is the span of the columns of A.

The reason the image of transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ is the same as the column space is because the image of T is spanned by the vectors in the columns of the associated matrix

$$\begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix} \tag{9.22}$$

In other words, $\vec{b}$ is in the image of A if and only if there exist coefficients $x_1, \ldots, x_n$ such that

$$\vec{b} = x_1 T(\vec{e}_1) + \cdots + x_n T(\vec{e}_n). \tag{9.23}$$

Example 9.24. For the baker, the linear transformation from the batches of pastries to the ingredients required, the image of this transformation describes the quantity of ingredients needed to evenly make the batches of the pastries without any excess. For any point not in the image, the baker will either have an excess of certain ingredients or a lack of certain ingredients.

Theorem 9.25. *The image of a linear transformation* $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ *is a subspace of* $\mathbb{R}^m$.

To avoid confusion between "imaginary," the image of T will be denoted by $\text{ran}(T)$.

Proof. We must check the axioms of a subspace.

(a) Since every linear transformation takes $\vec{0}$ to $\vec{0}$, the zero vector satisfies $\vec{0} = T(\vec{0})$. Hence, $\vec{0} \in \text{ran}(T)$.

(b) Let $\vec{x} \in \text{ran}(T)$ and let $c \in \mathbb{R}$. By the first assumption, there exists a $\vec{z} \in \mathbb{R}^n$ such that $T(\vec{z}) = \vec{x}$. Then $c\vec{x} = cT(\vec{z}) = T(c\vec{z})$ which shows that $c\vec{x} \in \text{ran}(T)$.

(c) Let $\vec{x}, \vec{y} \in \text{ran}(T)$. Then there exist $\vec{z}, \vec{w}$ such that $T(\vec{z}) = \vec{x}$ and $T(\vec{w}) = \vec{y}$. Therefore, $\vec{x} + \vec{y} = T(\vec{z}) + T(\vec{w}) = T(\vec{z} + \vec{w})$ so that $\vec{x} + \vec{y} \in \text{ran}(T)$.

■

Definition 9.26. Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. The dimension of the image of T is called the *rank* of T and is denoted by $\operatorname{rank} T$,

$$\operatorname{rank}(T) = \dim(\operatorname{ran}(T)). \tag{9.27}$$

The rank of a linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ can be calculated by counting the number of pivot columns in the associated $m \times n$ matrix. In fact, the pivot columns (from the original matrix) form a basis for the image of T.

Proposition 9.28. *Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation then the pivot columns of the associated matrix*

$$\begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix} \tag{9.29}$$

form a basis for $\operatorname{ran}(T)$.

Proof. We have already established the the columns span $\operatorname{ran}(T)$. Let $i_1, \ldots, i_k$ denote the indices corresponding to the pivot columns. Then, removing the non-pivot columns from this matrix, we get

$$\left[\begin{array}{ccc|c} | & & | & 0 \\ T(\vec{e}_{i_1}) & \cdots & T(\vec{e}_{i_k}) & \vdots \\ | & & | & 0 \end{array}\right] \xrightarrow{\text{row operations}} \left[\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{array}\right] \tag{9.30}$$

after reducing to reduced row echelon form. This shows that the pivot columns of the original matrix are linearly independent. Hence, they form a basis for the image of T. ■

Example 9.31. Consider the linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$ described by the matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -3 & 2 \end{bmatrix}. \tag{9.32}$$

The images of the vectors $\vec{e}_1$ and $\vec{e}_2$ get sent to the columns of the matrix. They span the plane shown in Figure 11.

Problem 9.33. Let

$$A := \begin{bmatrix} 1 & -2 & 3 \\ -5 & 10 & -15 \end{bmatrix} \tag{9.34}$$

and set

$$\vec{b} := \begin{bmatrix} 2 \\ -10 \end{bmatrix}. \tag{9.35}$$

(a) Find a vector $\vec{x}$ such that $A\vec{x} = \vec{b}$.

(b) Is there more than one such $\vec{x}$ as in part (a)?

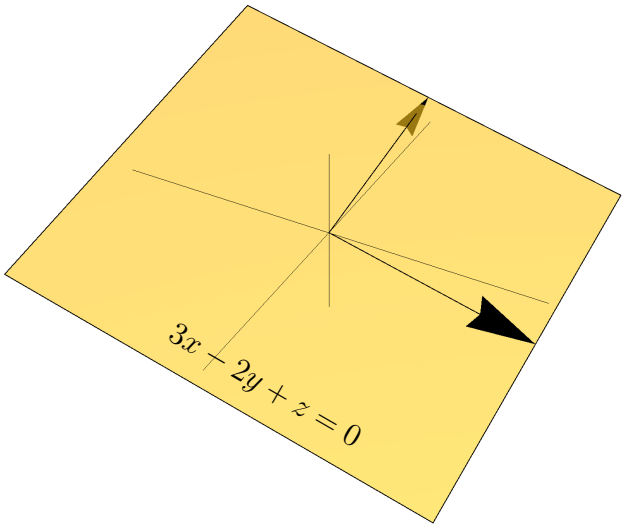


Figure 11: A plot of the plane described by $3x - 2y + z = 0$ along with two vectors spanning it.

(c) Is the vector

$$\vec{v} := \begin{bmatrix} 3 \\ 0 \end{bmatrix} \tag{9.36}$$

in the range of A viewed as a linear transformation?

Answer.

(a) To answer this, we must solve

$$\begin{bmatrix} 1 & -2 & 3 \\ -5 & 10 & -15 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -10 \end{bmatrix} \tag{9.37}$$

which we can do in the usual way we have learned

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 2 \\ -5 & 10 & -15 & -10 \end{array}\right] \xmapsto{\text{add 5 of row 1 to row 2}} \left[\begin{array}{ccc|c} 1 & -2 & 3 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{9.38}$$

There are two free variables here, say x_2 and x_3. Then x_1 is expressed in terms of them via

$$x_1 = 2 + 2x_2 - 3x_3. \tag{9.39}$$

Therefore, any vector of the form

$$\begin{bmatrix} 2 + 2x_2 - 3x_3 \\ x_2 \\ x_3 \end{bmatrix} \tag{9.40}$$

for *any* choice of x_2 and x_3 will have image $\vec{b}$.

(b) By the analysis from part (a), yes there is more than one such vector.

(c) To see if $\vec{v}$ is in the range of A, we must find a solution to

$$\begin{bmatrix} 1 & -2 & 3 \\ -5 & 10 & -15 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \tag{9.41}$$

but applying row operations as above

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 3 \\ -5 & 10 & -15 & 0 \end{array}\right] \xrightarrow{\text{add 5 of row 1 to row 2}} \left[\begin{array}{ccc|c} 1 & -2 & 3 & 3 \\ 0 & 0 & 0 & 15 \end{array}\right] \tag{9.42}$$

show that the system is inconsistent. This means that there are no solutions and therefore, $\vec{v}$ is not in the range of A.

Definition 9.43. A linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ is *onto* if every vector $\vec{b}$ in $\mathbb{R}^m$ is in the range of T and is *one-to-one* if for any vector $\vec{b}$ in the range of T, there is only a single vector $\vec{x}$ in $\mathbb{R}^n$ whose image is $\vec{b}$.

Theorem 9.44. *The following are equivalent for a linear transformation* $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$.

(a) T is one-to-one.

(b) The only solution to the linear system $T(\vec{x}) = \vec{0}$ is $\vec{x} = \vec{0}$.

(c) The columns of the matrix associated to T are linearly independent.

Proof. We will prove (a) $\implies$ (b) $\implies$ (c) $\implies$ (a).
((a) $\implies$ (b)) Suppose that T is one-to-one. Suppose there is an $\vec{x} \in \mathbb{R}^n$ such that $T(\vec{x}) = \vec{0}$. Since T is linear, $T(\vec{0}) = \vec{0}$. Since T is one-to-one, $\vec{x} = \vec{0}$.
((b) $\implies$ (c)) Suppose that the only solution to $T(\vec{x}) = \vec{0}$ is $\vec{x} = \vec{0}$. The goal is to show that $\{T(\vec{e}_1), \ldots, T(\vec{e}_1)\}$ is linearly independent, since these are precisely the columns of the matrix associated to T. The linear system

$$y_1 T(\vec{e}_1) + \cdots + y_n T(\vec{e}_n) = 0 \tag{9.45}$$

can be expressed as

$$T(y_1\vec{e}_1 + \cdots + y_n\vec{e}_n) = 0 \tag{9.46}$$

using linearity of T. By assumption, the only solution to this is

$$y_1\vec{e}_1 + \cdots + y_n\vec{e}_n = \vec{0}. \tag{9.47}$$

Since $\{\vec{e}_1, \ldots, \vec{e}_n\}$ is linearly independent, the only solution to this system is $y_1 = \cdots = y_n = 0$. Hence $\{T(\vec{e}_1), \ldots, T(\vec{e}_1)\}$ is linearly independent.
((c) $\implies$ (a)) Suppose that the columns of the matrix associated to T are linearly independent. Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ satisfy $T(\vec{x}) = T(\vec{y})$. The goal is to prove that $\vec{x} = \vec{y}$. By linearity of T, $T(\vec{x} - \vec{y}) = \vec{0}$. Since $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$ and similarly for $\vec{y}$, this reads

$$T\big((x_1 - y_1)\vec{e}_1 + \cdots + (x_n - y_n)\vec{e}_n\big) = \vec{0}. \tag{9.48}$$

By linearity of T, this can be expressed as

$$(x_1 - y_1)T(\vec{e}_1) + \cdots + (x_n - y_n)T(\vec{e}_n) = \vec{0}. \tag{9.49}$$

Since $\{T(\vec{e}_1), \ldots, T(\vec{e}_1)\}$ is linearly independent, the only solution to this system is

$$x_1 - y_1 = \cdots = x_n - y_n = 0. \tag{9.50}$$

In other words,

$$x_1 = y_1, \ldots, x_n = y_n \tag{9.51}$$

so that $\vec{x} = \vec{y}$. ■

Theorem 9.52. *The following are equivalent for a linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$.*

(a) T is onto.

(b) For every $\vec{b} \in \mathbb{R}^m$, the linear system $T(\vec{x}) = \vec{b}$ always has at least one solution.

(c) The columns of the associated $m \times n$ matrix A span $\mathbb{R}^m$, i.e.

$$\operatorname{span}\big(\{T(\vec{e}_1), \ldots, T(\vec{e}_n)\}\big) = \mathbb{R}^m. \tag{9.53}$$

Proof. This one is left as an exercise. Compare this to Theorem 7.41 (it's the same thing!). ■

Theorem 9.54. *Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. Then*

$$\operatorname{rank} T + \dim(\ker T) = n. \tag{9.55}$$

In other words,

$$\#\textit{ pivot columns } + \#\textit{ free variables } = \textit{ dimension of domain.} \tag{9.56}$$

Proof. Omitted. ■

The idea behind this "rank-nullity" theorem is that the information you start with (the domain) equals the information you obtained (the image) plus the information that was lost (the kernel). Notice that this has nothing to do with the codomain of the linear transformation.

As a summary, given a linear transformation

$$\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n \tag{9.57}$$

with domain (source) $\operatorname{domain}(T) = \mathbb{R}^n$ and codomain (target) $\operatorname{codomain}(T) = \mathbb{R}^m$, the image and kernel of T are given by

$$\operatorname{image}(T) := \big\{T(\vec{x}) \in \mathbb{R}^m \ : \ \vec{x} \in \mathbb{R}^n\big\} \tag{9.58}$$

and

$$\ker(T) := \big\{\vec{x} \in \mathbb{R}^n \ : \ T(\vec{x}) = \vec{0}\big\}. \tag{9.59}$$

Note that image$(T) \subseteq$ codomain(T) and $\ker(T) \subseteq$ domain(T). Since every linear transformation T has the matrix form

$$\begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix}, \tag{9.60}$$

the span of the columns is

$$\operatorname{span}\big\{T(\vec{e}_1), \ldots, T(\vec{e}_n)\big\} = \text{image}(T). \tag{9.61}$$

In fact, the pivot columns provide a *basis* for image(T). Hence,

$$\dim\big(\text{image}(T)\big) = \#\text{ of pivots of } T. \tag{9.62}$$

Solving

$$\left[\begin{array}{ccc|c} | & & | & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) & \vec{0} \\ | & & | & | \end{array}\right] \tag{9.63}$$

in parametric form will give you the basis for $\ker(T)$. Hence, the number of free variables is the dimension of this kernel

$$\dim\big(\ker(T)\big) = \#\text{ of free variables of (9.63)}. \tag{9.64}$$

Because of this, it immediately makes sense why the number of pivots plus the number of free variables is n.

Recommended Exercises. Exercises 10 and 12 in Section 1.8 of [Lay]. Exercises 22, 23, and 31 in Section 2.8 of [Lay]. Exercises 13, 15, and 20, in Section 2.9 of [Lay]. Be able to show all your work, step by step!

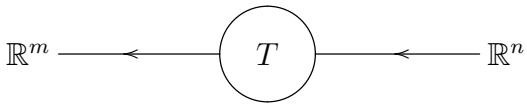
In this lecture, we finished Sections 1.8, 2.8, and parts of 2.9 of [Lay].

Terminology checklist

kernel/null space	
range/image/column space	
linear combination	
rank	
one-to-one/injective	
onto/surjective	

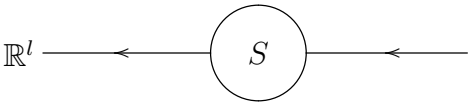
10 Iterating linear transformations—matrix multiplication

If you think of a linear transformation as a process (recall our examples: ingredients for a recipe is a decomposition, colorblindness can be described by a filter, etc.), you can perform processes in succession. For example, imagine you had two linear transformations



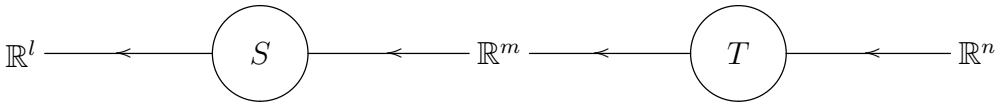
(10.1)

and



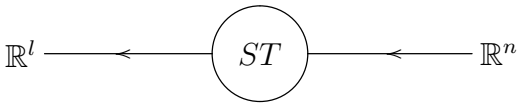
(10.2)

Then it should be reasonable to perform these operations in succession as



(10.3)

so that the result is some operation, denoted by ST, from $\mathbb{R}^n$ to $\mathbb{R}^l$



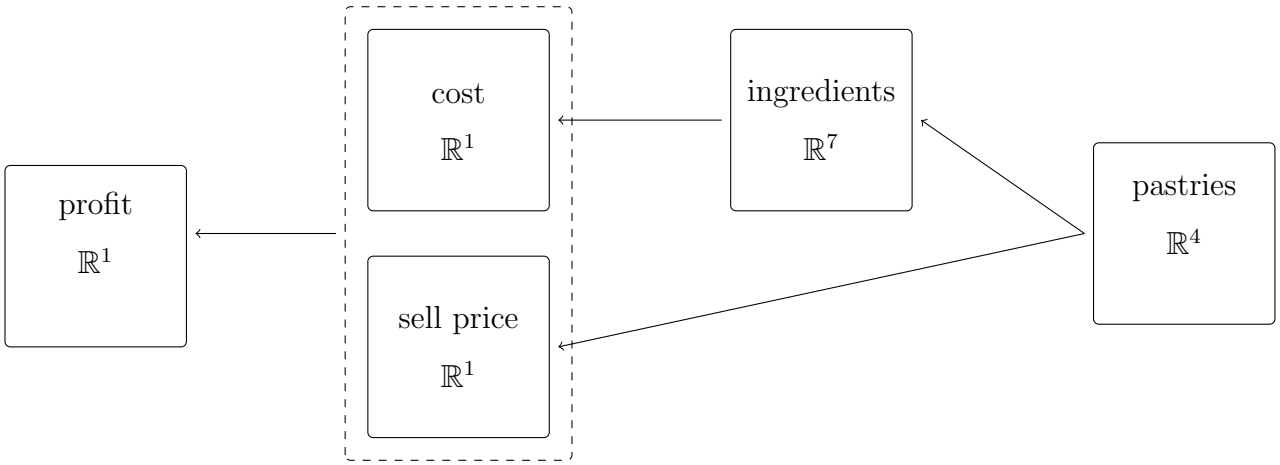
(10.4)

Problem 10.5. Most grocery stores give you discounts on items if you purchase more. In this case, the cost per quantity of a good is not typically linear. So imagine, for purposes of illustrating the example of iterating linear transformations, that a particular grocery store has a fixed price per quantity of goods regardless of how much you buy. Refer to Example 7.5 on baking pastries and their ingredients. For convenience, we have reproduced the table as well as the cost of ingredients per quantity. In addition, we have added the selling price at the bottom of the table per item (as opposed to per batch).

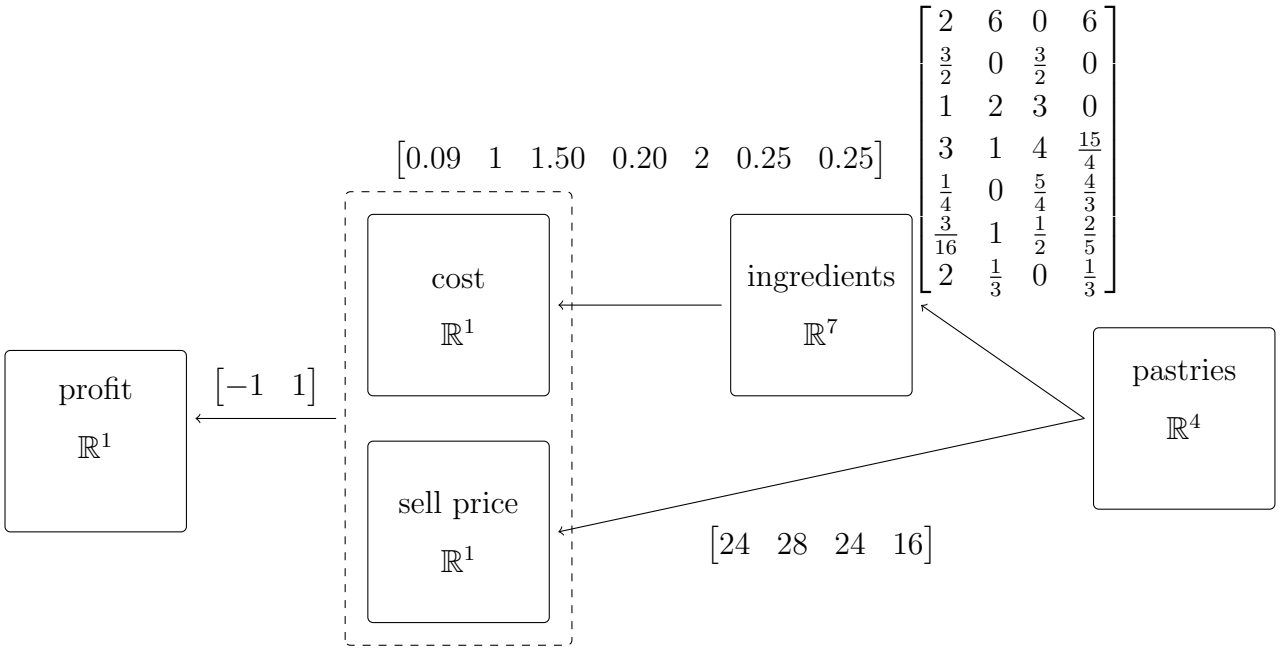
	Pancakes (makes 12)	Tres leches (makes 8 slices)	Strawberry shortcake (makes 8 slices)	Egg tarts (makes 16)	cost per ingredient
eggs	2	6	0	6	\$0.09 per egg
strawberries	$1+\frac{1}{2}$ lbs	0	$1+\frac{1}{2}$ lbs	0	\$1.00 per lb
heavy cream	1 cup	1 pint	3 cups	0	\$1.50 per cup
flour	3 cups	1 cup	4 cups	$3+\frac{3}{4}$ cups	\$0.20 per cup
butter	4 tbsp	0	$1+\frac{1}{4}$ cups	$1+\frac{1}{3}$ cups	\$2.00 per cup
sugar	3 tbsp	1 cup	$\frac{1}{2}$ cup	$\frac{2}{5}$ cup	\$0.25 per cup
milk	2 cups	$\frac{1}{3}$ cup	0	$\frac{1}{3}$ cup	\$0.25 per cup
selling price	\$2.00	\$3.50	\$3.00	\$1.00	

Note that there are 16 tablespoons in a cup and 2 cups in a pint. Ignoring the costs of maintaining a business, what is the profit of the bakery if they sell 4 batches of pancakes, 3 batches of tres leches cakes, 2 batches of strawberry shortcakes, and 4 batches of egg tarts?

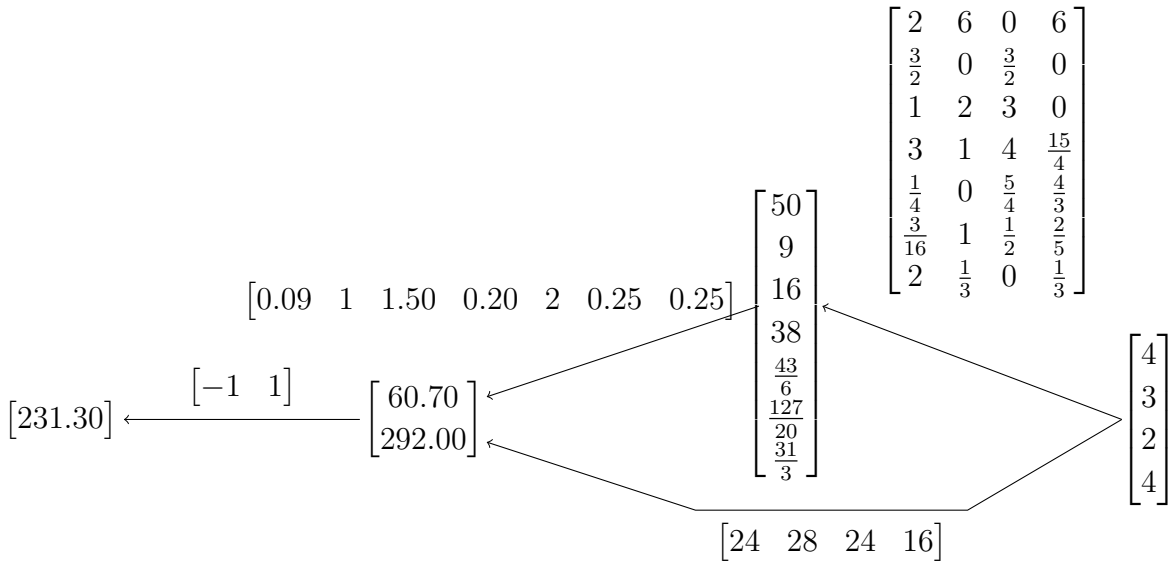
Answer. Before calculating, conceptually, we have the following list of linear transformations.



The cost and sell price can be calculated separately, but we have boxed them together because the profit is calculated as a difference of the two. The explicit matrices corresponding to these linear transformations are given by



Given the known quantities that are sold, we can calculate the images of the batches sold under these different linear transformations.

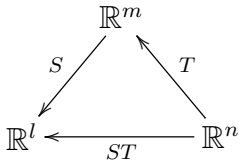


This leads to a profit of \$231.30.

Definition 10.6. The *composition* of $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ followed by $\mathbb{R}^l \xleftarrow{S} \mathbb{R}^m$ is the function $\mathbb{R}^l \xleftarrow{ST} \mathbb{R}^n$ defined by

$$\mathbb{R}^n \ni \vec{x} \mapsto S\big(T(\vec{x})\big). \tag{10.7}$$

In words, ST is the transformation that sends a vector $\vec{x}$ to $T(\vec{x})$ by applying T first and then applies S to the result, which is $T(\vec{x})$. Diagrammatically this looks like



(10.8)

Proposition 10.9. *The composition of a linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ followed by a linear transformation $\mathbb{R}^l \xleftarrow{S} \mathbb{R}^m$ is a linear transformation $\mathbb{R}^l \xleftarrow{ST} \mathbb{R}^n$.*

Proof. Let $c \in \mathbb{R}$ and $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then

$$\begin{aligned} S\big(T(\vec{x}+\vec{y})\big) &= S\big(T(\vec{x}) + T(\vec{y})\big) && \text{by linearity of } T \\ &= S\big(T(\vec{x})\big) + S\big(T(\vec{y})\big) && \text{by linearity of } S \end{aligned} \tag{10.10}$$

and

$$\begin{aligned} S\big(T(c\vec{x})\big) &= S\big(cT(\vec{x})\big) && \text{by linearity of } T \\ &= cS\big(T(\vec{x})\big) && \text{by linearity of } S. \end{aligned} \tag{10.11}$$

Hence, ST is a linear transformation. ■

Exercise 10.12. Show that $(ST)(\vec{0}) = \vec{0}$.

Because ST is a linear transformation, it must have a matrix associated to it. Let A be the matrix associated to S and let B be the matrix associated to T. Remember, this means

$$A = \begin{bmatrix} | & & | \\ S(\vec{e}_1) & \cdots & S(\vec{e}_m) \\ | & & | \end{bmatrix} \qquad \& \qquad B = \begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix}. \tag{10.13}$$

Notice the difference in the unit vector inputs! Let's try to figure out the matrix associated to ST. To do this, we need to figure out what the columns of this matrix are, and this should be given by

$$\begin{bmatrix} | & & | \\ (ST)(\vec{e}_1) & \cdots & (ST)(\vec{e}_n) \\ | & & | \end{bmatrix}. \tag{10.14}$$

Therefore, all we have to do is figure out what an arbitrary column in this matrix looks like. Therefore, pick some $i \in \{1, \ldots, n\}$. Our goal is to calculate this column

$$\begin{bmatrix} | \\ (ST)(\vec{e}_i) \\ | \end{bmatrix} = \begin{bmatrix} | \\ S\big(T(\vec{e}_i)\big) \\ | \end{bmatrix}. \tag{10.15}$$

By definition, $T(\vec{e}_i)$ is the i-th column of

$$B = \begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix}. \tag{10.16}$$

Let's therefore give some notation to the elements of this column:

$$\begin{bmatrix} | \\ T(\vec{e}_i) \\ | \end{bmatrix} =: \begin{bmatrix} b_{1i} \\ \vdots \\ b_{mi} \end{bmatrix}. \tag{10.17}$$

Notice that the indices make sense because B is an $m \times n$ matrix, so its columns must have m entries. We left the i index on the right because we want to keep track that it is the i-th column. Now we apply the linear transformation S to this vector, which we know how to compute

$$S\big(T(\vec{e}_i)\big) = \begin{bmatrix} | & & | \\ S(\vec{e}_1) & \cdots & S(\vec{e}_m) \\ | & & | \end{bmatrix} \begin{bmatrix} b_{1i} \\ \vdots \\ b_{mi} \end{bmatrix} = b_{1i}S(\vec{e}_1) + \cdots + b_{mi}S(\vec{e}_m). \tag{10.18}$$

Therefore, this particular linear combination of the columns of S is the i-th column of ST. Let's also put in some notation here. Writing the $l \times m$ matrix associated to S as

$$\begin{bmatrix} | & & | \\ S(\vec{e}_1) & \cdots & S(\vec{e}_m) \\ | & & | \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{l1} & \cdots & a_{lm} \end{bmatrix} \tag{10.19}$$

we can express the above linear combination as

$$\begin{aligned} S\big(T(\vec{e}_i)\big) &= b_{1i}S(\vec{e}_1) + \cdots + b_{mi}S(\vec{e}_m) \\ &= b_{1i}\begin{bmatrix} a_{11} \\ \vdots \\ a_{l1} \end{bmatrix} + \cdots + b_{mi}\begin{bmatrix} a_{1m} \\ \vdots \\ a_{lm} \end{bmatrix}. \\ &= \begin{bmatrix} b_{1i}a_{11} + \cdots + b_{mi}a_{1m} \\ \vdots \\ b_{1i}a_{l1} + \cdots + b_{mi}a_{lm} \end{bmatrix} \end{aligned} \tag{10.20}$$

Yes, this looks complicated. And remember, that this is only the i-th column of ST. If we now did this for all columns and entries, we would find that the matrix associated to ST is

$$AB = \begin{bmatrix} \sum_{k=1}^{m} a_{1k}b_{k1} & \sum_{k=1}^{m} a_{1k}b_{k2} & \cdots & \sum_{k=1}^{m} a_{mk}b_{kn} \\ \sum_{k=1}^{m} a_{2k}b_{k1} & \sum_{k=1}^{m} a_{2k}b_{k2} & \cdots & \sum_{k=1}^{m} a_{2k}b_{kn} \\ \vdots & \vdots & & \vdots \\ \sum_{k=1}^{m} a_{lk}b_{k1} & \sum_{k=1}^{m} a_{lk}b_{k2} & \cdots & \sum_{k=1}^{m} a_{lk}b_{kn} \end{bmatrix} \tag{10.21}$$

From this calculation, we see that the ij component (meaning the i-th row and j-th column entry) $(AB)_{ij}$ of the matrix AB is given by

$$(AB)_{ij} := \sum_{k=1}^{m} a_{ik}b_{kj}. \tag{10.22}$$

The resulting formula seems overwhelming, but there is a convenient way to remember it instead of this long derivation. The ij-th component of AB is given by multiplying the entries of the i-th row of A with the entries of the j-th column of B one by one in order and then adding them all together:

$$\begin{array}{c} \\ i\text{-th row} \rightarrow \end{array} \begin{bmatrix} \\ \boxed{a_{i1} \quad a_{i2} \quad \cdots \quad \cdots \quad a_{im}} \\ \\ \end{bmatrix} \begin{bmatrix} \\ \boxed{\begin{matrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{matrix}} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \displaystyle\sum_{k=1}^{m} a_{ik}b_{kj} \\ \\ \end{bmatrix} \begin{array}{c} j\text{-th column} \\ \downarrow \\ \\ \\ \\ \\ \end{array} \tag{10.23}$$

This operation makes sense because the number of entries in a row of A is m while the number of entries in a column of B is also m. Yet another way of thinking about the matrix product AB is if we write B as

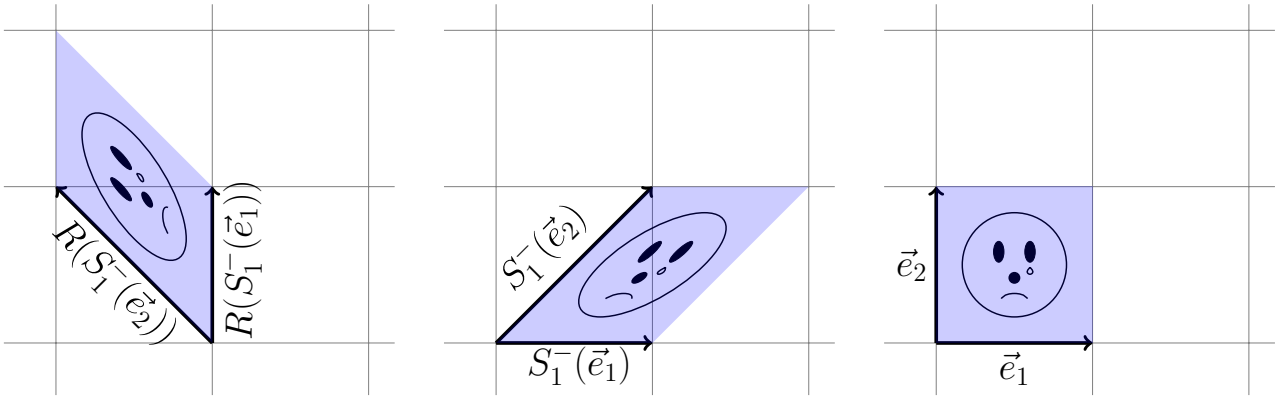
$$B = \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_n \\ | & & | \end{bmatrix} \tag{10.24}$$

Then AB is the matrix

$$AB = \begin{bmatrix} | & & | \\ A\vec{b}_1 & \cdots & A\vec{b}_m \\ | & & | \end{bmatrix}. \tag{10.25}$$

Example 10.26. Consider the following two linear transformations on $\mathbb{R}^2$ given by a shear S and then a rotation R by angle θ (in the figures, $k = 1$ and $\theta = \frac{\pi}{2}$).

$$\mathbb{R}^2 \longleftarrow \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \longleftarrow \mathbb{R}^2 \longleftarrow \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \longleftarrow \mathbb{R}^2 . \tag{10.27}$$



Let us compute the matrix associated to RS by calculating the first and second columns, i.e. $(RS)(\vec{e}_1)$ and $(RS)(\vec{e}_2)$. The first one is

$$R\big(S(\vec{e}_1)\big) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{10.28}$$

while the second is

$$R\big(S(\vec{e}_2)\big) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} k \\ 1 \end{bmatrix} = k \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} + \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} = \begin{bmatrix} k\cos\theta - \sin\theta \\ k\sin\theta + \cos\theta \end{bmatrix}. \tag{10.29}$$

Therefore, the resulting linear transformation is given by

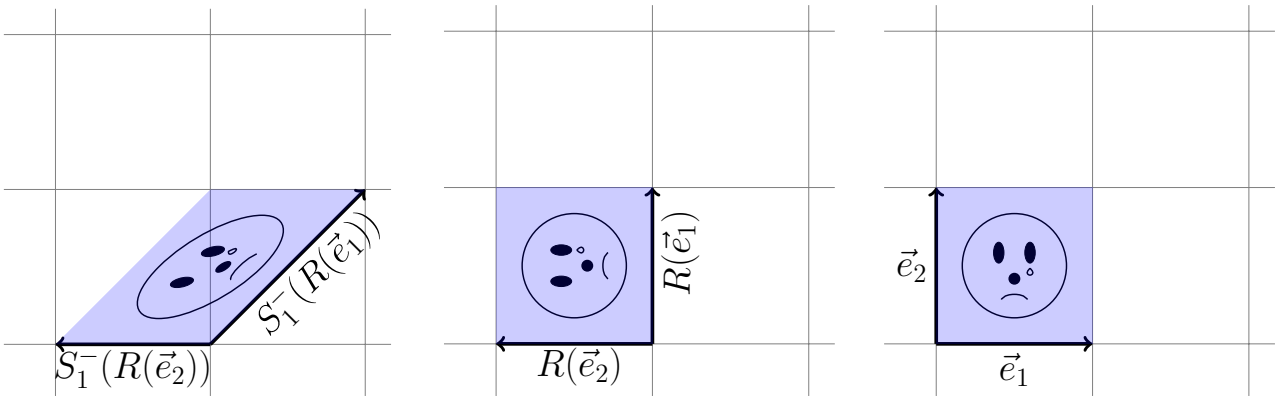
$$\mathbb{R}^2 \longleftarrow \begin{bmatrix} \cos\theta & k\cos\theta - \sin\theta \\ \sin\theta & k\sin\theta - \cos\theta \end{bmatrix} \longleftarrow \mathbb{R}^2 , \tag{10.30}$$

which with $k = 1$ and $\theta = \frac{\pi}{2}$ becomes

$$\mathbb{R}^2 \longleftarrow \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \longleftarrow \mathbb{R}^2 \tag{10.31}$$

If, however, we executed these operations in the opposite order

$$\mathbb{R}^2 \longleftarrow \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \longleftarrow \mathbb{R}^2 \longleftarrow \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \longleftarrow \mathbb{R}^2 \tag{10.32}$$



we would find the resulting linear transformation to be

$$\mathbb{R}^2 \longleftarrow \begin{bmatrix} k\sin\theta + \cos\theta & k\cos\theta - \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \longleftarrow \mathbb{R}^2\,, \tag{10.33}$$

which with $k = 1$ and $\theta = \frac{\pi}{2}$ becomes

$$\mathbb{R}^2 \longleftarrow \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \longleftarrow \mathbb{R}^2 \tag{10.34}$$

If A is an $m \times m$ matrix, then

$$A^k := \overbrace{A \cdots A}^{k \text{ times}} \tag{10.35}$$

Note that it does not make sense to raise an $m \times n$ matrix to some power other than 1. By definition,

$$A^0 := \mathbb{1}_m \tag{10.36}$$

is the identity $m \times m$ matrix.

Exercise 10.37. State whether the following claims are True or False. If the claim is true, be able to precisely deduce why the claim is true. If the claim is false, be able to provide an explicit counter-example.

(a) $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}^{15} = \begin{bmatrix} 1 & 15k \\ 0 & 1 \end{bmatrix}$ for all real numbers k.

(b) The matrix $\begin{bmatrix} -0.6 & 0.8 \\ -0.8 & -0.6 \end{bmatrix}$ represents a rotation.

Exercise 10.38. Compute the matrices from exercises 7-11 in Section 1.9 of [Lay] in the following two ways. First, calculate each of the individual matrices for the transformations and then matrix multiply (compose). Second, write the matrix associated to the over-all transformation. How are these two methods of calculating related?

Recommended Exercises. Exercises 9, 10, 11, and 12 in Section 2.1 of [Lay]. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we finished Section 2.1 of [Lay].

Terminology checklist

composition/matrix multiplication	
raising a square matrix to a power	

11 Hamming's error correcting code

We will review several concepts in the context of an example. This example and a lot of the wording comes directly from an exercise in [1]. Warning: our presentation differs slightly from the one often given in computer science courses.[42] We will try to include the computer science descriptions in remarks. You do not need to know all of the details about binary to solve any of the linear algebra problems, but if you already know binary, the remarks will help you understand the differences between our presentation and the one you might be accustomed to. Our purpose is to formulate Hamming's error correcting codes in such a way as to utilize linear algebra without having prior knowledge of computer science or binary.

Remark 11.1. In binary, one uses the numbers 0 and 1 only to represent arbitrary natural numbers. We are used to doing arithmetic in base 10. For binary, we use base 2. For example, the number 137 can be expressed as

$$137 = 1 \times 10^2 + 3 \times 10^1 + 7 \times 10^0. \tag{11.2}$$

Another way to decompose this number is in terms of sums of powers of 2 but our coefficients must be numbers less than 2, i.e. 0 or 1. So, 137, for example, is expressed as

$$\begin{aligned} 137 &= 128 + 8 + 1 \\ &= 2^7 + 2^3 + 2^0 \\ &= 1 \times 2^7 + 0 \times 2^6 + 0 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^1. \end{aligned} \tag{11.3}$$

Therefore, one can represent 137 as

$$137 \equiv 10001001 \tag{11.4}$$

in binary. Furthermore, one can do arithmetic analogous to how you learned how to do arithmetic in primary school. For example, to add the number 85 to 137 you would carry over terms

$$\begin{aligned} 137 + 85 &\equiv \left(1 \times 10^2 + 3 \times 10^1 + 7 \times 10^0\right) + \left(0 \times 10^2 + 8 \times 10^1 + 5 \times 10^0\right) \\ &= (1+0) \times 10^2 + (3+8) \times 10^1 + (7+5) \times 10^0 \\ &= 1 \times 10^2 + (10+1) \times 10^1 + (10+2) \times 10^0 \\ &= (1+1) \times 10^2 + (1+1) \times 10^1 + 2 \times 10^0 \\ &= 2 \times 10^2 + 2 \times 10^1 + 2 \times 10^0 \\ &\equiv 222. \end{aligned} \tag{11.5}$$

The shortcut method of doing this is what you might have learned in primary school

$$\begin{array}{r} 11 \\ 137 \\ +\ 85 \\ \hline 222 \end{array}$$

[42] We thank Christian Carmellini and Philip Parzygnat for helpful discussions on these points.

For binary, we proceed as follows. First, 85 is represented as

$$85 = 64 + 16 + 4 + 1 = 2^6 + 2^4 + 2^2 + 2^0 \equiv 1010101. \tag{11.6}$$

Adding these two numbers using the shortcut method gives

$$\begin{array}{r} \ 1 \\ 10001001 \\ +\ \ 1010101 \\ \hline 11011110 \end{array}$$

which equals

$$11011110 \equiv 128 + 64 + 16 + 8 + 4 + 2 = 222 \tag{11.7}$$

as expected. We can also make sense of multiplication analogous to how multiplication is defined for numbers using base 10. **We will *ignore* this arithmetic structure in what follows.** Instead, we will only keep track of the *parity* of all the arithmetic computations that we will perform. This will be explained in more detail in remarks as we progress.

Associated with every number is its parity. The parity will be indicated with either a 0 or a 1. 0 indicates that the number is even while 1 indicates that the number is odd. The way we will add these numbers is the same way we normally add numbers except with the rule that $1 + 1 = 0$. For example, $2017 = 1$ while $2018 = 0$. Hence we will *not* be using binary arithmetic to add our numbers. Multiplication of these numbers is also treated in the same was as with ordinary integers. This makes sense because, for example, an even number times an odd number is an even number while an odd number times an odd number is still an odd number. Just like the set of real numbers is so important that we give it notation, such as $\mathbb{R}$, the set $\{0, 1\}$ with this addition is so important that we also provide it with a symbol. Unfortunately, people will disagree on what letter to use. I prefer the notation $\mathbb{Z}_2$ to remind myself that I am working with integers where "2" is treated as 0. Furthermore, just as we can form n-component vectors of *real* numbers, (recall, this set is denoted by $\mathbb{R}^n$), we can form n-component vectors of 0's and 1's, and we denote this set by $\mathbb{Z}_2^n$. Most of the manipulations, definitions, and theorems that worked for vectors in $\mathbb{R}^n$ work for vectors in $\mathbb{Z}_2^n$.

Problem 11.8. Because there are infinitely many real numbers, there are infinitely many vectors in $\mathbb{R}^n$ for $n > 0$. How many vectors are there in $\mathbb{Z}_2^n$?

Answer. For each component of a vector in $\mathbb{Z}_2^n$, there are 2 possibilities: either a 0 or a 1. Therefore, for n components, this gives 2^n possible entries. In particular, the number of vectors in $\mathbb{Z}_2^n$ is finite.

Definition 11.9. An element of $\mathbb{Z}_2$ is typically called a *bit*, a vector in $\mathbb{Z}_2^8$ is typically called a *byte*, and a vector in $\mathbb{Z}_2^4$ is typically called a *nibble*.

Remark 11.10. One can think of each element of $\mathbb{Z}_2^n$ as encoding a specific number between 0 and 2^n. In this way, a binary representation of a number such as $137 \equiv 10001001$ can be encoded

in the vector

$$137 \equiv 10001001 \leftrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{Z}^7 \tag{11.11}$$

(yet another convention is to reverse the order). However, we will not be using the standard binary arithmetic in this representation. Instead, we will think of this vector as encoding some kind of information that need not be a natural number like 137. Instead, one can use these strings of 0's and 1's to encode a letter, word, sentence, phrase, etc. There are many ways to do this. For example, one can use ASCII to translate from strings to letters and symbols. Regardless, once a string of 0's and 1's is used to signify a message, it no longer makes sense to perform arithmetic with these strings in the usual way binary operations are performed. The vectors that will follow in this section should be interpreted in this way. Namely, they encode some kind of a message.

In 1950, Richard Hamming introduced a method of recovering transmitted information that was subject to certain kinds of errors during its transmission. A *Hamming matrix* with n rows is a matrix with $2^n - 1$ columns and whose columns consist of exactly all the non-zero vectors of $\mathbb{Z}_2^n$. For example, one such Hamming matrix with $n = 3$ rows (therefore $2^3 - 1 = 7$ columns) is given by

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}. \tag{11.12}$$

Problem 11.13. Express the kernel of H as the span of four vectors in $\mathbb{Z}_2^7$ of the form

$$\vec{v}_1 = \begin{bmatrix} * \\ * \\ * \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} * \\ * \\ * \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} * \\ * \\ * \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{v}_4 = \begin{bmatrix} * \\ * \\ * \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{11.14}$$

Answer. All we have to do is solve the augmented matrix problem (find the homogenous solution to)

$$\left[\begin{array}{ccccccc|c} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{array}\right] \tag{11.15}$$

and we see immediately (since this matrix is already in reduced echelon form) that the general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} -x_4 - x_6 - x_7 \\ -x_4 - x_5 - x_7 \\ -x_4 - x_5 - x_6 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = x_4 \begin{bmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{11.16}$$

where x_4, x_5, x_6, and x_7 are free variables. But, don't forget that $-1 = 1$ in $\mathbb{Z}_2$, so this actually becomes

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = x_4 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{11.17}$$

where x_4, x_5, x_6, and x_7 are free variables. By the way, this expression of the set of solutions is now in parametric form (only the free variables appear). From this, we can immediately read off the requested vectors:

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{v}_4 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{11.18}$$

Let's make sure this answer makes sense. H is a 3×7 matrix. The first three columns are pivot columns and the last four columns provide us with free variables. Therefore, we expect the kernel of H to be 4-dimensional. This agrees with the fact that we found the four vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$. Rank-Nullity tells us that these vectors form a basis for the kernel of H (but you can also check this explicitly by showing that these four vectors are linearly independent). Furthermore, because H is a 3×7 matrix, it describes a linear transformation $\mathbb{Z}_2^3 \xleftarrow{H} \mathbb{Z}_2^7$. Hence, the kernel should consist of vectors with 7 components. Again, this is consistent with the basis we found.

Using $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$, we can construct a new matrix

$$M := \begin{bmatrix} | & | & | & | \\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{11.19}$$

Problem 11.20. Show that $\text{image}(M) = \ker(H)$. In particular, what is the resulting matrix HM obtained by first performing M and then H?

Answer. The image of M is the span of the columns of the matrix associated with M. But these vectors also span the kernel of H by construction. Therefore,

$$\text{image}(M) = \text{span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\} = \ker(H). \tag{11.21}$$

This implies, in particular, that $\text{image}(M) \subseteq \ker(H)$. In other words, a vector $\vec{x}$ in $\text{image}(M)$ is also in $\ker(H)$, i.e. $H(M(\vec{x})) = \vec{0}$. This means that HM is the zero 3×4 matrix. Notice that we didn't even have to calculate HM explicitly.

Now, consider a "message", which will mathematically be described by a vector $\vec{u}$ in $\mathbb{Z}_2^4$, i.e. a nibble (so a message consisting of four bits), and suppose you would like to transmit this message to someone. While this message is traveling, the environment might perturb it slightly and might change some of the components of the vector $\vec{u}$.

We assume for simplicity that *at most one* entry can change during transmission.

Because our messages are encoded using only 0's and 1's, the way in which $\vec{u}$ might change is completely determined by which component gets altered.[43] Thus, there are only five possible outcomes for what happens to $\vec{u}$. These possible outcomes are

$$\vec{u} \qquad \vec{u} + \vec{e}_1 \qquad \vec{u} + \vec{e}_2 \qquad \vec{u} + \vec{e}_3 \qquad \vec{u} + \vec{e}_4 \tag{11.22}$$

since $\vec{u}$ is a four-component vector Notice that adding a vector $\vec{e}_i$ is the same as subtracting one since we are working with binary. For example, imagine we started with the vector

$$\vec{u} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}. \tag{11.23}$$

[43]For example, if our entries were allowed to be $0, 1$, or 2, and the number 2 gets altered, there are two possible numbers it could be: 0 or 1. Therefore, if the receiver sees 0 and somehow knows that the error occured in this entry, then the initial number could have been a 1 or a 2 and one needs additional information to figure this out. Using 0's and 1's only avoids this issue.

When this message is transmitted, if *at most one* error occurs, the five possible outcomes are

$$\vec{u} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \ \vec{u} + \vec{e}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \ \vec{u} + \vec{e}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \ \vec{u} + \vec{e}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ \vec{u} + \vec{e}_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \quad (11.24)$$

The information we are working with in our message is precious! We would like to not lose *any* information. How can we do this? Again, this only works assuming that at most one error occurs, but nevertheless it is an amazing idea. Hamming realized that you can *first* transform the vector $\vec{u} \in \mathbb{Z}_2^4$ into a *seven*-component vector $\vec{v} \in \mathbb{Z}_2^7$ by using the matrix M, i.e.

$$\vec{v} = M\vec{u} \quad (11.25)$$

and then transmit *that* vector. This is beneficial for the following several reasons. First of all, $\vec{u}$ is part of the vector $\vec{v}$. In fact, if we split up M into two parts (the top and bottom parts)

$$M^{\text{top}} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \qquad \& \qquad M^{\text{bot}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (11.26)$$

so we see that

$$M\vec{u} = \begin{bmatrix} M^{\text{top}}\vec{u} \\ \vec{u} \end{bmatrix} \quad (11.27)$$

so that all of the information of $\vec{u}$ is stored directly in the vector $\vec{v}$ since it is contained in the last four components. The three entries in $M^{\text{top}}\vec{u}$ are called *parity bits* (see Remark 11.41 for an explanation why). Secondly, there are now *eight* (instead of five) possible outcomes of $\vec{v}$ after transmission

$$\vec{v} \qquad \vec{v} + \vec{e}_1 \qquad \vec{v} + \vec{e}_2 \qquad \vec{v} + \vec{e}_3 \qquad \vec{v} + \vec{e}_4 \qquad \vec{v} + \vec{e}_5 \qquad \vec{v} + \vec{e}_6 \qquad \vec{v} + \vec{e}_7 \quad (11.28)$$

Problem 11.29. Using the matrix H, how can the receiver detect if there was an error during transmission? [Hint: consider the case that there was no error first and then analyze the situation if there was an error.]

Answer. If there is no error, then $\vec{v} = H\vec{u}$ is the message that the receiver sees. They write down this information. The receiver can then apply the transformation H to this message and will find that the resulting vector is $H\vec{v} = H(M\vec{u}) = \vec{0}$ by our calculations above. When they see the string of all 0's, they know that the message they received was the same as the original message and they keep only the last 4 rows of the vector $\vec{v}$ that they received since they know these 4 numbers give the vector $\vec{u}$.

On the other hand, if an error occurs during transmission, then the resulting vector that the receiver obtains will be of the form $\vec{v} + \vec{e}_i$ for some $i \in \{1, 2, \ldots, 7\}$. Remember, the receiver does not *know* that the vector is of this form (all they see is a string of 0's and 1's). Nevertheless, when they apply M they will obtain the vector

$$H(\vec{v} + \vec{e}_i) = H(M(\vec{u}) + \vec{e}_i) = H(M(\vec{u})) + H(\vec{e}_i) = \vec{0} + H(\vec{e}_i) = H\vec{e}_i, \quad (11.30)$$

where $H\vec{e}_i$ is the i-th column of the matrix H. Remember, this is because

$$H = \begin{bmatrix} | & | & | & | & | & | & | & | \\ H\vec{e}_1 & H\vec{e}_2 & H\vec{e}_2 & H\vec{e}_3 & H\vec{e}_4 & H\vec{e}_5 & H\vec{e}_6 & H\vec{e}_7 \\ | & | & | & | & | & | & | & | \end{bmatrix}. \tag{11.31}$$

By looking at the matrix H, notice that every single column in that matrix is different from every other column. Therefore, $H\vec{e}_i$ will inform the receiver what i is, and remember that i represents the component where the error occurred during transmission. First suppose the error occurred where $i \in \{1, 2, 3\}$. Because $\vec{u}$, the original message, was in the last four slots of the vector $M\vec{u}$, the receiver knows that the message they read before applying H is indeed the original message that was sent since the error only occurred in one of the first 3 entries and did not affect the last 4 entries (which is where $\vec{u}$ is). However, if $i \in \{4, 5, 6, 7\}$, then the receiver knows that an error occurred in the original message. Fortunately, because they can see $H\vec{e}_i$, they can identify which component the error occurred in. They can then fix this error (again because we are only using 0's and 1's, there is only one other number it could have been) and obtain the original message. Fascinating!

Example 11.32. In case you didn't quite get that, let's work with a concrete example. Suppose a sender has the initial message

$$\vec{u} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}. \tag{11.33}$$

After applying M, this vector becomes

$$\vec{v} = M\vec{u} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \tag{11.34}$$

since we must remember that $2 = 0$ in $\mathbb{Z}_2$. Notice how $\vec{u}$ is still preserved in the bottom four entries. So now let's say this message is transmitted and an error occurs in the second entry (of course, the receiver does not know this) so what the receiver sees first is the vector

$$\vec{v} + \vec{e}_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}. \tag{11.35}$$

The receiver writes this information down. Then applies the linear transformation M to the message and obtains

$$H(\vec{v}+\vec{e}_2) = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \tag{11.36}$$

Notice that the *math* tells us that $H(\vec{v}+\vec{e}_2) = H\vec{e}_2$ but the receiver does not *know beforehand* where the error occurred, so we should not express the above equation as equal to $H\vec{e}_2$ (even though it's true) because that would presume the receiver already knows the state—the receiver must apply the received vector *as they obtained it* because they are initially ignorant. So we see that we obtained the second column of H which tells us that an error occurred in the second entry of the transmitted message. Therefore, the last four entries were not altered and the receiver can safely conclude that the original message was indeed our starting vector $\vec{u}$.

Example 11.37. Using the same $\vec{u}$ as in (11.33), now consider the situation where the fifth entry of $\vec{v}$ gets altered during the transmission. Therefore, the receiver sees

$$\vec{v}+\vec{e}_5 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{11.38}$$

Applying H to this gives

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}. \tag{11.39}$$

Therefore, the receiver knows that the 5th component of the original message has an error because this resulting 3-component vector is the 5th column of H. By flipping this fifth entry and looking at the last four components, they get back the original message $\vec{u}$.

In summary, the receiver can perform a (linear) operation on the received message (namely, H) and figure out the *entire* original message *even if there was an error during transmission*! It

is very important to notice that neither H nor M were constructed in any way that depends on the original message! They apply to *all* transmissions of 4-bit messages with at most one error occurrence.

Remark 11.40. The previous experiment unfortunately fails if bits are replaced by their quantum analogues, known as *qubits*. The reason is that whenever the receiver looks at the message, they necessarily alter the state. This is what makes quantum cryptography challenging, but these aspects can also be used as strengths.

What do you do if you want to transmit longer messages? Although we will not answer this, the following remark partially addresses this question.

Remark 11.41. We will relate our discussion of Hamming error correcting codes to one that is often presented to the computer scientist in terms of what are called *parity bits*. Parity bits are the additional bits added to the initial message in the Hamming error correcting code and are used to identify the location of a possible error. To see how this works, consider an arbitrary message that the sender wishes to send

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}. \tag{11.42}$$

The matrix M changes this 4-bit message into a 7-bit message of the form

$$M\vec{u} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} u_1 + u_3 + u_4 \\ u_1 + u_2 + u_4 \\ u_1 + u_2 + u_3 \\ u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}. \tag{11.43}$$

The top three entries of $M\vec{u}$ are known as the parity bits. They add the bits from the entries of the original matrix in such a way so that if at most one error occurs, the receiver will perform the sums in the observed vector and compare it to what they see. If the entries in the parity bits do not obey the form on the right-hand-side of (11.43), then it means that the parity bit detects an error in one of the four entries it covers. For example, p_1 checks the entries p_1, u_1, u_3 and u_4. If $p_1 \neq u_1 + u_3 + u_4$, this means that p_1 detects an error, which means that an error definitely occurred in one of these four entries. If $p_1 = u_1 + u_3 + u_4$, this means that there is *no* error in *any* of these entries. The other parity bits detect errors in a similar fashion. One uses process of elimination to isolate precisely where the error occurred. For example, imagine that the receiver

sees the vector

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}. \tag{11.44}$$

The receiver is aware of the fact that the first three entries are the parity bits and that they should satisfy the sum formula of (11.43). Do they? If we take the bottom four entries (since this is where the initial message is contained) and apply the matrix M to it, instead we obtain

$$M \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \tag{11.45}$$

This tells us that the second parity bit detects an error while the first and third parity bits do not detect an error. The first parity bit (entry 1) checks itself and the 1st, 3rd, and 4th entries of the initial message (entries 1, 4, 6, and 7). The second parity bit (entry 2) checks itself and the 1st, 2nd, and 4th entries of the initial message (entries 2, 4, 5, and 7). The third parity bit (entry 3) checks itself and the 1st, 2nd, and 3rd entries of the initial message (entries 3, 4, 5, and 6). Because the first parity bit agrees with the message the receiver sees, this means that there is *no* error in the 1st, 4th, 6th, and 7th entries. Because the second parity bit *dis*agrees with the message the receiver sees, this means that there *is* an error in *one* of the 2nd, 4th, 5th, or 7th entries. By just looking at the first two parity bits, we know that the error occurs in the 2nd or 5th entries (it cannot be in the 4th or 7th because the first parity bit excludes this possibility), but we need the last parity bit to determine which one of these two possibilities it is. Because the third parity bit agrees with the message the receiver sees, this means that there is *no* error in the 3rd, 4th, 5th, and 6th entries. Hence, combining these results together, we conclude that the error occurred in the 2nd entry. This idea generalizes to arbitrary 4-bit messages and indicates how the parity bits work. The linear algebra method we have employed to identify the location of the error bypasses this elimination process.

Another subtle example is very helpful. Suppose that the receiver sees the vector

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}. \tag{11.46}$$

We check what the parity bits say (without using the linear algebra method)

$$\begin{aligned} p_1 \stackrel{?}{=} u_1 + u_3 + u_4 \quad &: \quad 1 \stackrel{?}{=} 0 + 1 + 1 \quad \text{✗} \\ p_2 \stackrel{?}{=} u_1 + u_2 + u_4 \quad &: \quad 0 \stackrel{?}{=} 0 + 0 + 1 \quad \text{✗} \\ p_3 \stackrel{?}{=} u_1 + u_2 + u_3 \quad &: \quad 1 \stackrel{?}{=} 0 + 0 + 1 \quad \text{✓} \end{aligned} \tag{11.47}$$

This means that parity bits p_1 and p_2 *both* detect an error. Since they themselves cannot *both* be where the error occurs (since this would mean two errors have occurred), it must be that one of the entries it checks in common has the error. This means that either u_1 or u_4 contains the error. Parity bit p_3, however, says that p_3, u_1, u_2, and u_3 are error-free. By process of elimination, this means that the error occurred in the entry u_4. We can also check this directly using the linear algebra method by applying the matrix H and we would find that H applied to the received vector indicates an error in the 7-th entry, which corresponds to an error in u_4.

There is one important difference between this method and the one computer scientists might be familiar with. This is the location of the parity bits and the location of the initial message. It is convenient to reorganize where the initial message and parity bits are. For an initial message of length four, as we have been discussing, the parity bits would actually be located in the 1st, 2nd, and 4th entries, and not the 1st, 2nd, and 3rd entries.

$$\begin{bmatrix} p_1 \\ p_2 \\ u_1 \\ p_3 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \tag{11.48}$$

One reason for doing this is so that one can change the length of the message easily by appending a longer string to the message. If we were to increase the length of our new message, we would have to include more parity bits, and where would we place them? Our convention for a message of length 4 placed all the parity bits on top, but if we want to add to our message, we have no choice but to place the next parity bit after our first string of 7 bits has left. For example, if you have 4 parity bits, you can actually send a message of length 11. The total vector containing the initial message and the parity bits might look something like (the column vector has been drawn as a row vector to fit it more easily on the page)

$$\begin{bmatrix} p_1 & p_2 & u_1 & p_3 & u_2 & u_3 & u_4 & p_4 & u_5 & u_6 & u_7 & u_8 & u_9 & u_{10} & u_{11} \end{bmatrix}. \tag{11.49}$$

The n-th parity bit is placed in the 2^{n-1}-st entry of a vector of length $2^n - 1$ so that the length of an initial message that can be sent with n parity bits is $2^n - 1 - n$. With our linear algebra method, we would have an awkward formula for where to place the parity bits.

What happens if you allow for more than just one error? This has also been addressed in the literature, but I'll let you think about it.

Recommended Exercises. See homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

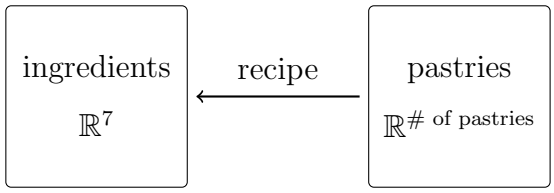
In this lecture, we reviewed many important concepts: kernel, image, matrix multiplication, bases, etc. all through an example that is an exercise in [1].

Terminology checklist

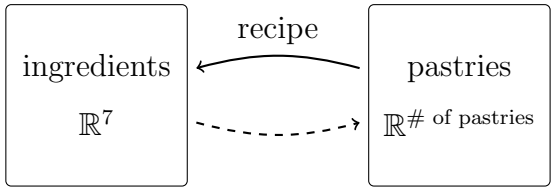
$\mathbb{Z}_2$	
bit, nibble, byte	
Hamming matrix	
Hamming's error correcting code	
parity bits	

12 Inverses of linear transformations

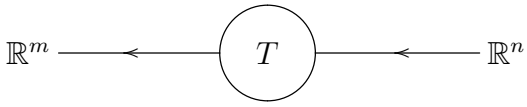
Refer back to Example 7.5 of the ingredients needed to make a set of pastries, but imagine one now considers *all* possible pastries (or at least a sufficiently large number of pastries, such as 20) one can make with those seven ingredients. As we discussed in that example, a *recipe* defines a linear transformation, which is, in particular, a function



Is there a way to go back?

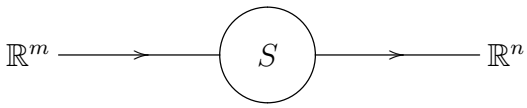


The way this question is phrased is a bit meaningless, because there are definitely many ways to go back. For example, you can just send every ingredient to the vector $\vec{0}$. A more meaningful question would be to ask if there is a way to go back that recovers the pastry you started with. Is this possible? Phrased differently, imagine being given a set of ingredients such as flour, milk, eggs, sugar, etc. What kinds of pastries can you make with your set of ingredients? Is there only one possibility? Of course not! Depending on the chef, one could make many different kinds of pastries with a given set of ingredients. Hence, there is no well-defined rule to go back, i.e. there is no function satisfying these requirements.[44] In the context of linear algebra, given a linear transformation



(12.1)

taking vectors with n components in and providing vectors with m components out, you might want to know if there is a way to go back to reverse the process. This would be a linear transformation going in the opposite direction (I've drawn it going backwards to our usual convention)

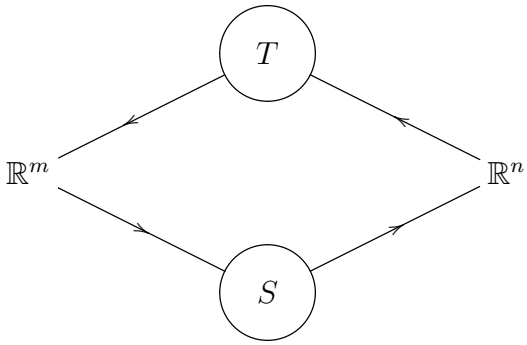


(12.2)

so that if we perform these two processes in succession, the result would be the transformation that does nothing, i.e. the *identity* transformation. In other words, going along any closed loop in

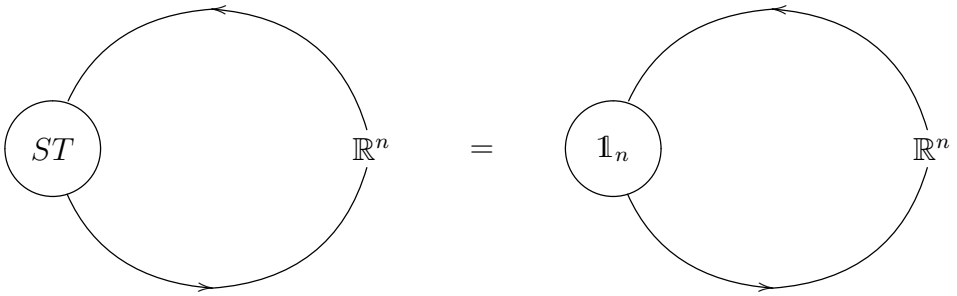
[44] If we had only used 4 pastries as in Example 7.5 and we used the recipes provided, then there actually is a way to go back because the columns of the matrix associated to the recipe are linearly independent. However, there are still *many* ways to go back and no unique choice.

the diagram



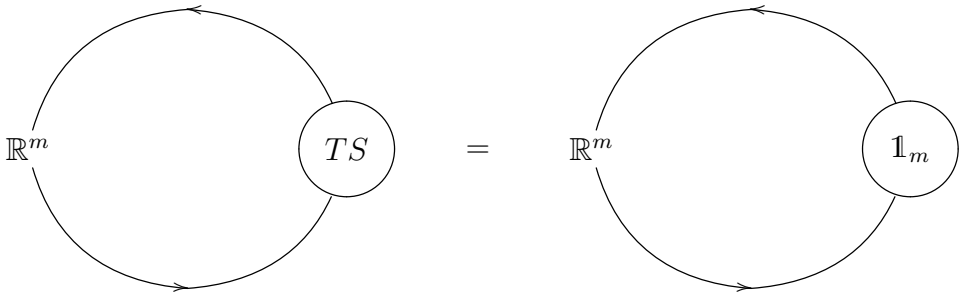
(12.3)

is the identity. Expressed another way, this means that



(12.4)

and



(12.5)

Here $\mathbb{1}_m$ is the identity transformation on $\mathbb{R}^m$ and similarly $\mathbb{1}_n$ on $\mathbb{R}^n$. Often, the inverse S of T is written as T^{-1} and the inverse T of S is written as S^{-1}. This is because inverses, if they exist, are unique.

Definition 12.6. A linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ is *invertible* (also known as *non-singular*) if there exists a linear transformation $\mathbb{R}^n \xleftarrow{S} \mathbb{R}^m$ such that

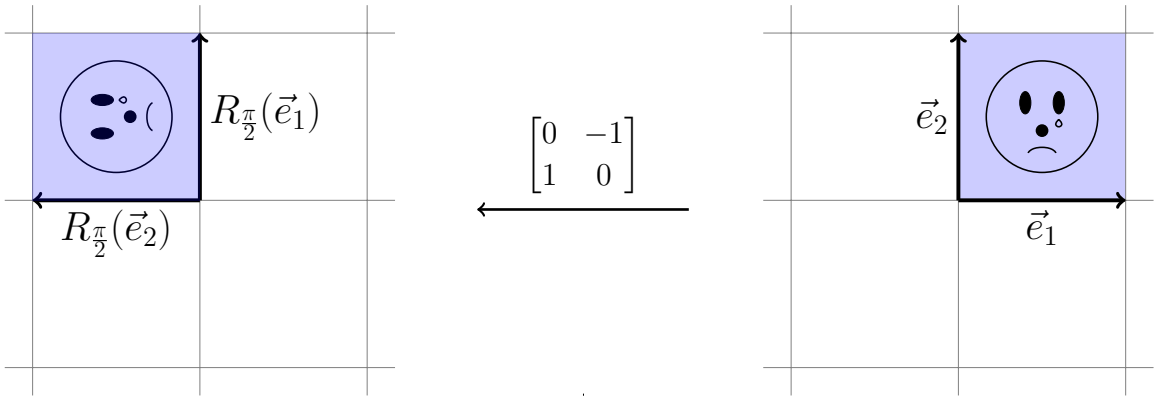
$$ST = \mathbb{1}_n \qquad \& \qquad TS = \mathbb{1}_m. \tag{12.7}$$

A linear transformation that is not invertible is called a *non-invertible* (also known as *singular*) linear transformation.

Example 12.8. Consider the matrix R_θ describing rotation in $\mathbb{R}^2$ counterclockwise about the origin by angle θ

$$\mathbb{R}^2 \longleftarrow \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \longleftarrow \mathbb{R}^2 \,. \tag{12.9}$$

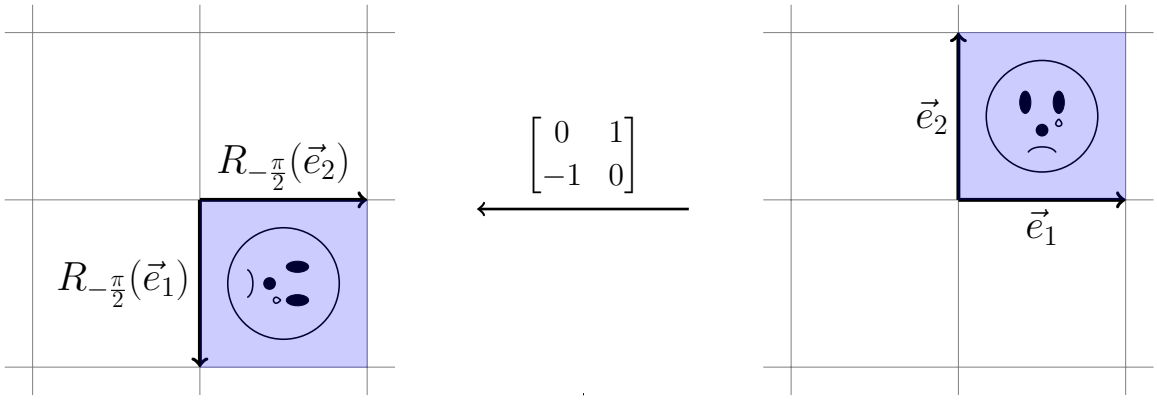
For $\theta = \frac{\pi}{2}$, this looks like



The inverse of such a transformation is very intuitive! We just want to rotate back by angle $-\theta$, i.e. clockwise by angle θ. This inverse *should* therefore be given by the matrix

$$R_{-\theta} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{12.10}$$

For $\theta = \frac{\pi}{2}$, this looks like



Is this really the inverse, though? We have to check the definition. Remember, this means we need to show

$$R_\theta R_{-\theta} = \mathbb{1}_2 \qquad \& \qquad R_{-\theta} R_\theta = \mathbb{1}_2. \tag{12.11}$$

It turns out that we only need to check any one of these conditions (this is one of the exercises in [Lay]), so let's check the second one.

$$\begin{aligned} R_{-\theta} R_\theta &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos^2\theta + \sin^2\theta & -\cos\theta\sin\theta + \sin\theta\cos\theta \\ -\sin\theta\cos\theta + \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{12.12}$$

There is something quite interesting about this last example, but to explain it, we provide the following definition.

Definition 12.13. The *transpose* of an $m \times n$ matrix A

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{12.14}$$

is the $n \times m$ matrix

$$A^T := \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}. \tag{12.15}$$

Another way of writing the transpose that makes it easier to remember is

$$\begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix}^T := \begin{bmatrix} — & \vec{a}_1 & — \\ & \vdots & \\ — & \vec{a}_n & — \end{bmatrix} \tag{12.16}$$

In other words, the columns become rows and vice versa. In the previous example of a rotation, we discovered that

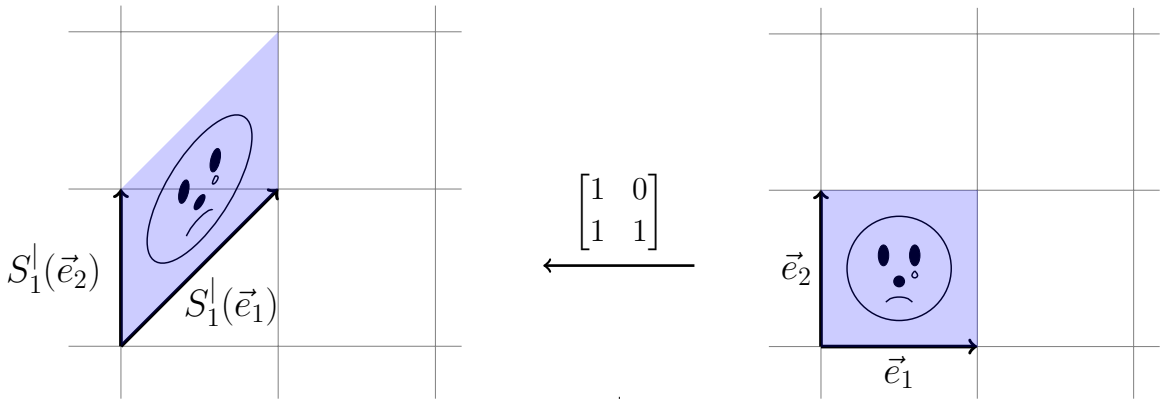
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^{-1} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^T. \tag{12.17}$$

These are special types of matrices, known as orthogonal matrices, and they will be discussed in more detail later in this course.

Example 12.18. Consider the matrix $S^{|}_k$ describing a vertical shear in $\mathbb{R}^2$ of length k

$$\mathbb{R}^2 \longleftarrow \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \longleftarrow \mathbb{R}^2 . \tag{12.19}$$

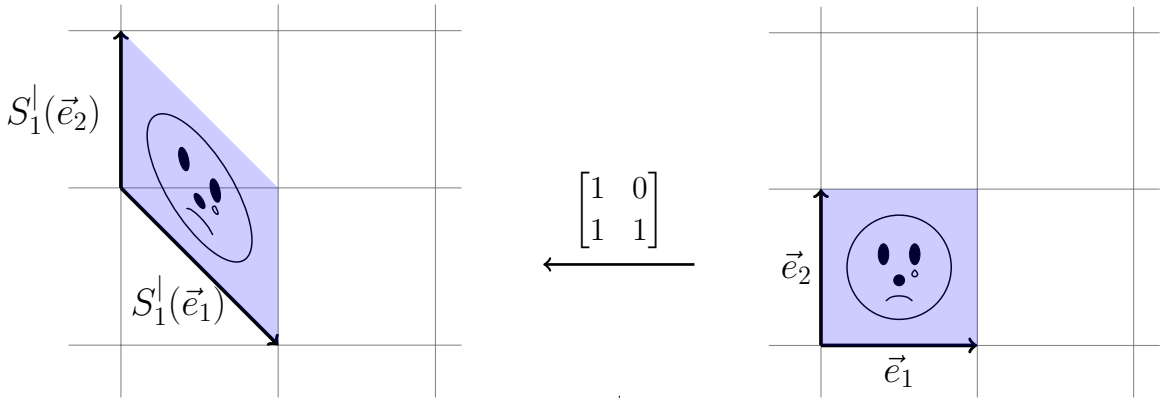
When $k = 1$, this transformation is depicted by



In this case as well, it seems intuitively clear that the inverse should be also vertical shear but where the shift is in the opposite vertical direction, namely, k should be replaced with $-k$. Thus, we propose that the inverse vertical shear, $S^{|}_{-k}$, is given by

$$S^{|}_{-k} = \begin{bmatrix} 1 & 0 \\ -k & 1 \end{bmatrix}. \tag{12.20}$$

When $k = 1$, this transformation is depicted by



We check that this works:

$$S_{-k}^| S_k^| = \begin{bmatrix} 1 & 0 \\ -k & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{12.21}$$

Theorem 12.22. *A 2×2 matrix*

$$A := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{12.23}$$

is invertible if and only if $ad - bc \neq 0$. When this happens,

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{12.24}$$

Proof. This is an if and only if statement so it's proof must be broken into two major steps.
($\Leftarrow$) Suppose that $ad - bc \neq 0$. Then the formula for A^{-1} shows that an inverse to A exists (matrix multiply to verify this). Hence, A is invertible.
($\Rightarrow$) Suppose that A is invertible. The goal is to show that $ad - bc \neq 0$. Since A is invertible, there exists a 2×2 matrix B such that $AB = \mathbb{1}_2 = BA$. Notice that this means solutions to the two systems

$$A\vec{x} = \vec{e}_1 \qquad \& \qquad A\vec{y} = \vec{e}_2 \tag{12.25}$$

are given by the respective columns of B since applying B to both sides of these two equations gives

$$\vec{x} = \mathbb{1}_2\vec{x} = (BA)(\vec{x}) = B(A\vec{x}) = B\vec{e}_1 \qquad \& \qquad \vec{y} = \mathbb{1}_2\vec{y} = (BA)(\vec{y}) = B(A\vec{y}) = B\vec{e}_2. \tag{12.26}$$

In other words, we have to be able to solve the system

$$\left[\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right] \tag{12.27}$$

for all $e, f \in \mathbb{R}$. In an earlier homework problem, we showed that if $a \neq 0$, then this row reduces to

$$\left[\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right] \mapsto \left[\begin{array}{cc|c} a & b & e \\ 0 & ad - bc & af - ec \end{array}\right]. \tag{12.28}$$

This is consistent for *all* $e, f \in \mathbb{R}$ provided that $ad - bc \neq 0$. But what if $a = 0$? Then it must be that $c \neq 0$ so that we can swap rows and row reduce in a similar way to make the same conclusion. Why can't both a and c be equal to 0? If this happened, then A would not have two pivot columns and it would not be possible to solve our two systems. Therefore, at least one of a or c is nonzero and the formula $ad - bc \neq 0$ must hold. ■

This actually concludes the proof, but you might wonder where the formula for A^{-1} comes from. Without loss of generality, suppose that $a \neq 0$ (we say without loss of generality because we can swap the rows to put c in the position of a and row reduction would give us an analogous result). Setting $e = 1$ and $f = 0$ gives

$$\left[\begin{array}{cc|c} a & b & 1 \\ c & d & 0 \end{array}\right] \mapsto \left[\begin{array}{cc|c} a & b & 1 \\ 0 & ad - bc & -c \end{array}\right], \tag{12.29}$$

which says

$$\left.\begin{aligned} ax_1 + bx_2 &= 1 \\ (ad - bc)x_2 &= 0 \end{aligned}\right\} \Rightarrow \vec{x} = \frac{1}{ad - bc}\begin{bmatrix} d \\ -c \end{bmatrix}. \tag{12.30}$$

Setting $e = 0$ and $f = 1$ gives

$$\left[\begin{array}{cc|c} a & b & 0 \\ c & d & 1 \end{array}\right] \mapsto \left[\begin{array}{cc|c} a & b & 0 \\ 0 & ad - bc & a \end{array}\right], \tag{12.31}$$

which says

$$\left.\begin{aligned} ay_1 + by_2 &= 0 \\ (ad - bc)y_2 &= a \end{aligned}\right\} \Rightarrow \vec{y} = \frac{1}{ad - bc}\begin{bmatrix} -b \\ a \end{bmatrix}. \tag{12.32}$$

Therefore,

$$B = \begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \tag{12.33}$$

which agrees with the formula for A^{-1} in the statement of the theorem.

Exercise 12.34. If $a = 0$, then $c \neq 0$. By going through a similar procedure, find B, the inverse of A, and show that it agrees with the formula we found.

Remark 12.35. You might have also tried to prove the second part of the theorem by writing the inverse of A as some matrix (the e and f here are not the same as in the above proof)

$$B = \begin{bmatrix} e & f \\ g & h \end{bmatrix} \tag{12.36}$$

and then matrix multiply with A to get the equation $AB = \mathbb{1}_2$ which reads

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{12.37}$$

This provides us with four equations in four unknowns (the knowns are a, b, c, d and the unknown variables are e, f, g, h)

$$\begin{aligned} ae + 0f + bg + 0h &= 1 \\ 0e + af + 0g + bh &= 0 \\ ce + 0f + dg + 0h &= 0 \\ 0e + cf + 0g + dh &= 1 \end{aligned} \quad , \tag{12.38}$$

which is a linear system described by the augmented matrix

$$\left[\begin{array}{cccc|c} a & 0 & b & 0 & 1 \\ 0 & a & 0 & b & 0 \\ c & 0 & d & 0 & 0 \\ 0 & c & 0 & d & 1 \end{array}\right] \tag{12.39}$$

By a similar argument to as before, at least one of a or c cannot be 0. Without loss of generality, assume that $a \neq 0$. Then we can row reduce this augmented matrix to

$$\left[\begin{array}{cccc|c} a & 0 & b & 0 & 1 \\ 0 & a & 0 & b & 0 \\ c & 0 & d & 0 & 0 \\ 0 & c & 0 & d & 1 \end{array}\right] \mapsto \left[\begin{array}{cccc|c} a & 0 & b & 0 & 1 \\ 0 & a & 0 & b & 0 \\ 0 & 0 & ad-bc & 0 & -c \\ 0 & 0 & 0 & ad-bc & a \end{array}\right] \tag{12.40}$$

Because a and c can't both be 0, in order for this system to be consistent, $ad - bc$ cannot be zero. This again concludes the proof since all that was needed to be shown was $ad - bc \neq 0$. But again, we can proceed and try to find the inverse by solving this augmented matrix completely. Proceeding with row reduction gives

$$\left[\begin{array}{cccc|c} a & 0 & b & 0 & 1 \\ 0 & a & 0 & b & 0 \\ 0 & 0 & ad-bc & 0 & -c \\ 0 & 0 & 0 & ad-bc & a \end{array}\right] \mapsto \left[\begin{array}{cccc|c} a & 0 & b & 0 & 1 \\ 0 & a & 0 & b & 0 \\ 0 & 0 & 1 & 0 & \frac{-c}{ad-bc} \\ 0 & 0 & 0 & 1 & \frac{a}{ad-bc} \end{array}\right] \mapsto \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & \frac{d}{ad-bc} \\ 0 & 1 & 0 & 0 & \frac{-b}{ad-bc} \\ 0 & 0 & 1 & 0 & \frac{-c}{ad-bc} \\ 0 & 0 & 0 & 1 & \frac{a}{ad-bc} \end{array}\right] \tag{12.41}$$

This gives us the matrix $B = A^{-1}$, and it agrees with our earlier result. Notice how much longer this construction was.

The quantity $ad - bc$ of a matrix as in this theorem is called the *determinant* of the matrix A and is denoted by $\det A$. In all of the examples, the matrices were square matrices, i.e. $m \times n$ matrices where $m = n$. It turns out that an $m \times n$ matrix cannot be invertible if $m \neq n$. Our examples from above are consistent with this theorem.

Example 12.42. In the 2×2 rotation matrix R_θ from our earlier examples, the determinant is given by

$$\det R_\theta = \cos\theta \cos\theta - \sin\theta(-\sin\theta) = \cos^2\theta + \sin^2\theta = 1. \tag{12.43}$$

Example 12.44. In the 2×2 vertical shear matrix $S_k^|$ from our earlier examples, the determinant is given by

$$\det S_k^| = 1 \cdot 1 - 0 \cdot k = 1. \tag{12.45}$$

You could imagine just by the form of the inverse of a 2×2 matrix that finding formulas for inverses of 3×3 or 4×4 matrices will be incredibly complicated. This is true. But we will still find that a certain number, also called the determinant, will completely determine whether an inverse *exists*. We will describe this in the next two sections. But before going there, let's look at some of the properties of the inverse of a linear transformation. We can still study properties even though we might not have an explicit formula for the inverse. Invertible matrices are quite useful for the following reason.

Theorem 12.46. *Let A be an invertible $m \times m$ matrix and let $\vec{b}$ be a vector in $\mathbb{R}^m$. Then the linear system*

$$A\vec{x} = \vec{b} \tag{12.47}$$

has a unique solution. Furthermore, this solution is given by

$$\vec{x} = A^{-1}\vec{b}. \tag{12.48}$$

Proof. The fact that $\vec{x} = A^{-1}\vec{b}$ is a solution follows from

$$A\big(A^{-1}\vec{b}\big) = (AA^{-1})\vec{b} = \mathbb{1}_2\vec{b} = \vec{b}. \tag{12.49}$$

To see that it is the only solution, suppose that $\vec{y}$ is another solution. Then by taking the difference of $A\vec{x} = \vec{b}$ and $A\vec{y} = \vec{b}$, we get

$$A(\vec{x} - \vec{y}) = \vec{0} \qquad \Rightarrow \qquad \underbrace{A^{-1}A}_{\mathbb{1}_2}(\vec{x} - \vec{y}) = A^{-1}\vec{0} \qquad \Rightarrow \qquad \vec{x} - \vec{y} = \vec{0} \tag{12.50}$$

so that $\vec{x} = \vec{y}$. ■

Exercise 12.51. Let

$$\vec{b} := \begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix} \tag{12.52}$$

and let $R_{\pi/6}$ be the matrix that rotates by $30°$ (in the counterclockwise direction). Find the vector $\vec{x}$ whose image is $\vec{b}$ under this rotation.

Steps:
(1) Write the matrix $R_{\pi/6}$ explicitly.
(2) Draw the vector $\vec{b}$.
(3) Guess a solution $\vec{x}$ by thinking about how $R_{\pi/6}$ acts.
(4) Use the theorem to calculate $\vec{x}$ to test your guess.
(5) Compare your results and then make sure it works.

Theorem 12.53. *If A is an invertible $m \times m$ matrix, then*

$$\big(A^{-1}\big)^{-1} = A. \tag{12.54}$$

If A and B are invertible $m \times m$ matrices, then AB is invertible and

$$(BA)^{-1} = A^{-1}B^{-1}. \tag{12.55}$$

This theorem is completely intuitive! To reverse two processes, you do each one in reverse as if you're rewinding a movie! The inverse of an $m \times m$ matrix A can be computed, if it exists, in the following way, reminiscent of how we solved linear systems. In fact, this idea is a generalization of the method we used to solve for the inverse of a 2×2 matrix. The idea is to row reduce the augmented matrix

$$\left[A \;\middle|\; \mathbb{1}_m\right] \tag{12.56}$$

to the form

$$\left[\mathbb{1}_m \;\middle|\; B\right] \tag{12.57}$$

where B is some new $m \times m$ matrix. If this can be done, $B = A^{-1}$.

Example 12.58. The inverse of the matrix

$$A := \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \tag{12.59}$$

can be calculated by some row reductions

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \mapsto \left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right] \mapsto \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right] \tag{12.60}$$

and then

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right] \mapsto \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right] \mapsto \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -1 & 1 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right] \tag{12.61}$$

So the supposed inverse is

$$A^{-1} = \begin{bmatrix} -1 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \tag{12.62}$$

To verify this, we should check that it works:

$$\begin{bmatrix} -1 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{12.63}$$

Exercise 12.64. A rotation by angle θ (about the origin) in $\mathbb{R}^3$ in the plane spanned by $\vec{e}_1$ and $\vec{e}_2$ is given by the matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12.65}$$

Show that the inverse of this matrix is

$$\begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12.66}$$

Theorem 12.67 (The Invertible Matrix Theorem)**.** *Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^m$ be a linear transformation with corresponding $m \times m$ matrix denoted by A. Then the following are equivalent (which means that if one condition holds, then all the other conditions hold).*

(a) T is invertible.

(b) The columns of A span $\mathbb{R}^m$, i.e. T is onto.

(c) The columns of A are linearly independent, i.e. T is one-to-one.

(d) For every $\vec{b} \in \mathbb{R}^m$, there exists a unique solution to $A\vec{x} = \vec{b}$.

(e) A^T is invertible.

Please see [Lay] for the full version of this theorem, which provides even more characterizations for a matrix to be invertible. Later, we will add other characterizing properties to this list as well.

Theorem 12.68. *Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. The following are equivalent.*

(a) T is one-to-one.

(b) There exists a linear transformation $\mathbb{R}^m \xrightarrow{S} \mathbb{R}^n$ such that $ST = \mathbb{1}_n$.

(c) The columns of the standard matrix associated to T are linearly independent.

(d) The only vector $\vec{x} \in \mathbb{R}^n$ satisfying $T\vec{x} = \vec{0}$ is $\vec{x} = \vec{0}$.

Theorem 12.69. *Let $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. The following are equivalent.*

(a) T is onto.

(b) There exists a linear transformation $\mathbb{R}^m \xrightarrow{S} \mathbb{R}^n$ such that $TS = \mathbb{1}_m$.

(c) The columns of the standard matrix associated to T span $\mathbb{R}^m$.

(d) For ever vector $\vec{b} \in \mathbb{R}^m$, there exists a vector $\vec{x} \in \mathbb{R}^n$ such that $T\vec{x} = \vec{b}$.

Exercise 12.70. State whether the following claims are True or False. If the claim is true, be able to precisely deduce why the claim is true. If the claim is false, be able to provide an explicit counter-example.

(a) $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}^{-15} = \begin{bmatrix} 1 & -15k \\ 0 & 1 \end{bmatrix}$ for all real numbers k.

(b) The inverse of the matrix $\begin{bmatrix} -0.6 & 0.8 \\ -0.8 & -0.6 \end{bmatrix}$ is the matrix $\begin{bmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{bmatrix}$.

(c) If A, B, C, and D are invertible 2×2 matrices, then $(ABCD)^{-1} = A^{-1}B^{-1}C^{-1}D^{-1}$.

Inverses can be used to compute the matrix of a linear transformation if it is known where a basis gets sent to under a linear transformation.

Problem 12.71. Let $\vec{u} = \begin{bmatrix} a \\ c \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} b \\ d \end{bmatrix}$ be a basis of $\mathbb{R}^2$ and let $\mathbb{R}^2 \xleftarrow{T} \mathbb{R}^2$ be a linear transformation. Suppose that

$$T(\vec{u}) = \begin{bmatrix} d \\ f \end{bmatrix} \qquad \& \qquad T(\vec{v}) = \begin{bmatrix} e \\ g \end{bmatrix}. \tag{12.72}$$

What is the matrix associated to T?

Answer. Let $A = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) \end{bmatrix}$ be the matrix associated to T. Then, A satisfies

$$A \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} d & e \\ f & g \end{bmatrix}. \tag{12.73}$$

Let's check this:

$$\begin{aligned} A \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} aT(\vec{e}_1) + cT(\vec{e}_2) & bT(\vec{e}_1) + dT(\vec{e}_2) \end{bmatrix} \\ &= \begin{bmatrix} T(a\vec{e}_1 + c\vec{e}_2) & T(b\vec{e}_1 + d\vec{e}_2) \end{bmatrix} \\ &= \begin{bmatrix} T(\vec{u}) & T(\vec{v}) \end{bmatrix}. \end{aligned} \tag{12.74}$$

Therefore, in order to solve for A, since $\{\vec{u}, \vec{v}\}$ is a linearly independent set, the invertible matrix theorem guarantees that the matrix $\begin{bmatrix} \vec{u} & \vec{v} \end{bmatrix}$ is invertible. Therefore,

$$\begin{aligned} A \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} d & e \\ f & g \end{bmatrix} &\iff A \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}}_{\mathbb{1}_2} = \begin{bmatrix} d & e \\ f & g \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \\ &\iff A = \begin{bmatrix} d & e \\ f & g \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}. \end{aligned} \tag{12.75}$$

Recommended Exercises. Exercises 7, 9 (except part (e)), 13, 14, 21, 22, 23, 24, 25, 26, 33, 34, and 35 in Section 2.2 of [Lay]. Exercises 11, 12, 13, 14, 21, 29, and 36 in Section 2.3 of [Lay]. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we finished Sections 2.2 and 2.3 of [Lay]. This concludes our study of Chapters 1 and 2 in [Lay]. In particular, we have skipped Sections 2.4, 2.5, 2.6, and 2.7.

Terminology checklist

invertible (non-singular) matrix	
non-invertible (singular) matrix	
transpose	
inverse of a 2×2 matrix	
determinant of a 2×2 matrix	
Calculating inverses using row reduction	
Invertible matrix theorem	

13 The signed volume scale of a linear transformation

We're going to do things a little differently from your book [Lay], so please pay close attention. Instead of starting with Section 3.1 on the formula for a determinant, we will explore some of the geometric properties of the determinant vaguely combining parts parts of Sections 3.2 and 3.3 (and also some stuff from Section 6.1). In the next lecture, we will talk about cofactor expansions (in fact, we will derive them). In the previous section, we defined the determinant of a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{13.1}$$

to be

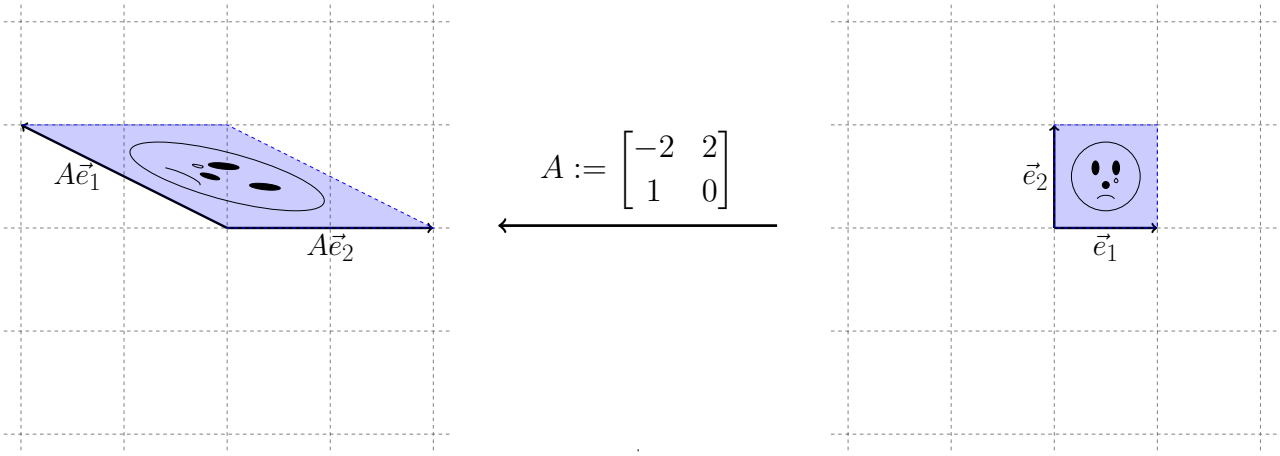
$$\det A := ad - bc. \tag{13.2}$$

We were partially motivated to give this quantity a special name because if $\det A \neq 0$, then the inverse of the matrix A is given by

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{13.3}$$

There is another perspective to determinants that allows a simple generalization to higher dimensions, i.e. for $m \times m$ matrices where m does not necessarily equal 2. To understand this generalization, we first explore some of the geometric properties of the determinant for 2×2 matrices.

Example 13.4. Consider the following linear transformation.



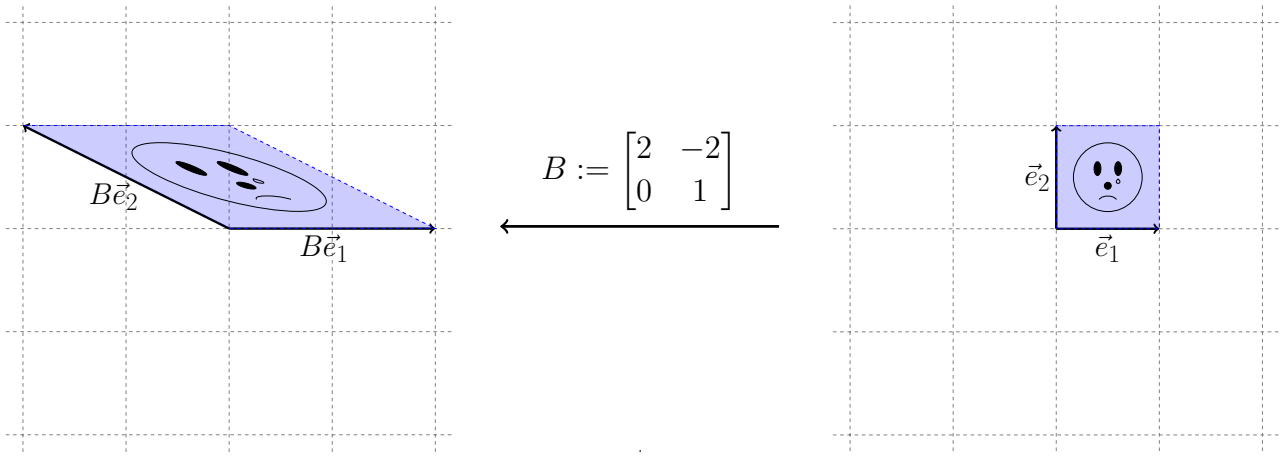
The square obtained from the vectors $\vec{e}_1$ and $\vec{e}_2$ gets transformed into a parallelogram obtained from the vectors $A\vec{e}_1$ and $A\vec{e}_2$. The area (a.k.a. 2-dimensional volume) of the square is initially 1. Under the transformation, the area becomes twice as big, so that gives a resulting area of 2. Also notice that the orientation of the face gets flipped once (the tear is initially on the left side of the face and after the transformation, it is on the right side). This is the same thing that happens to you when you look in the mirror. It turns out that

$$\det A = (\text{sign of orientation})(\text{volume of parallelogram}) = (-1)(2) = -2, \tag{13.5}$$

which we can check:

$$\det A = (-2)(0) - (1)(2) = -2. \tag{13.6}$$

Notice that if we swap the columns of A, then the transformation becomes



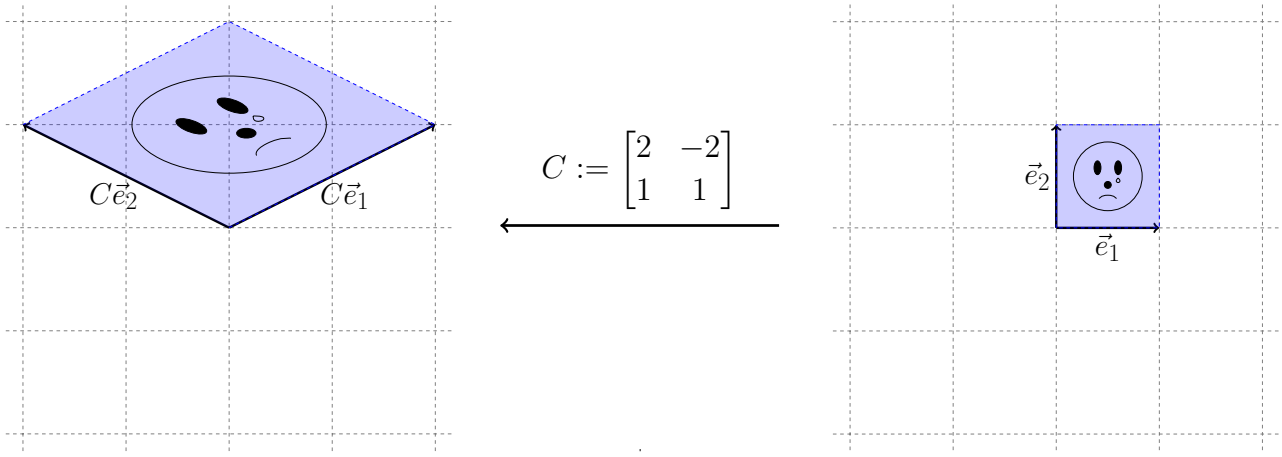
and the face is oriented the same way as in the original situation. The volume is scaled by 2 so we expect the determinant to be 2, and it is:

$$\det B = (2)(1) - (0)(-2) = 2. \tag{13.7}$$

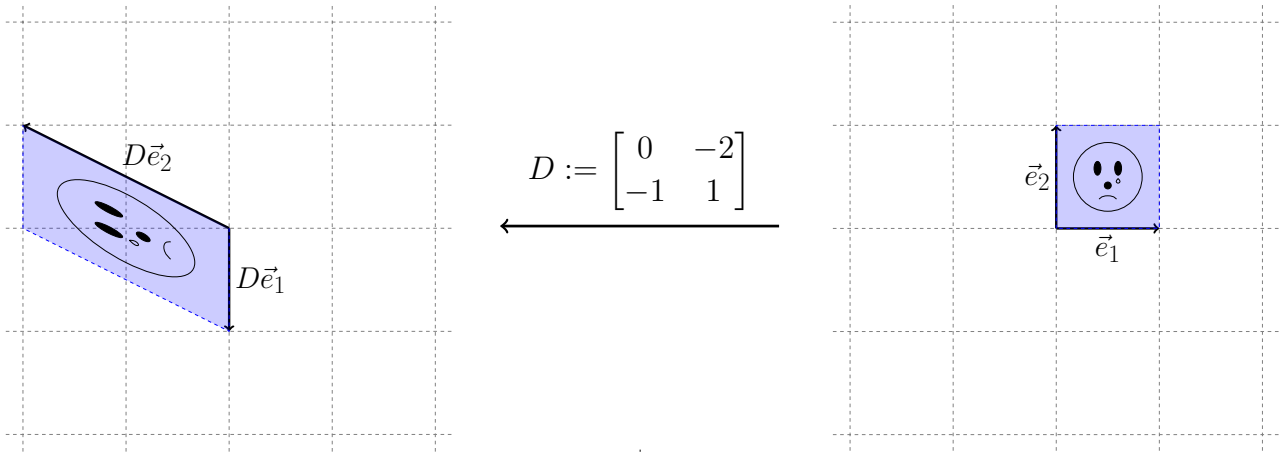
As another example, imagine writing the vector in the first column in the following way

$$\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix}. \tag{13.8}$$

Then how is the determinant of B related to the determinants of the transformations



and



A quick calculation shows that

$$\begin{aligned} \det B &= \det C + \det D \\ 2 &= 4 - 2. \end{aligned} \tag{13.9}$$

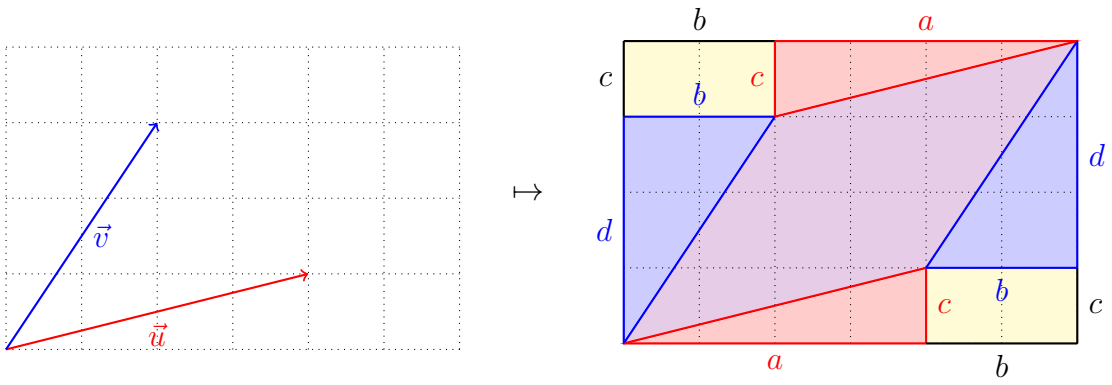
The previous example illustrates many of the basic properties of the determinant function. For a linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^m$, the determinant of the resulting matrix

$$\begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_m) \\ | & & | \end{bmatrix} \tag{13.10}$$

is the signed volume of the parallelepiped obtained from the column vectors in the above matrix. The sign of the determinant is determined by the resulting orientation: $+1$ if the orientation is right-handed and -1 if the orientation is left-handed. This definition has several important properties, many of which have been illustrated in the previous example. But we should gain more confidence that the determinant of a 2×2 matrix is really the area of the parallelogram whose two edges or obtained from the column vectors.

Theorem 13.11. *Let $a, b, c, d > 0$ and suppose that $a > b$ and $d > c$. Then the area of the parallelogram obtained from $\vec{u} := \begin{bmatrix} a \\ c \end{bmatrix}$ and $\vec{v} := \begin{bmatrix} b \\ d \end{bmatrix}$ is $ad - bc$.*

Proof. Drawing these two vectors and the resulting parallelogram



The area of the parallelogram is therefore

$$(a+b)(c+d) - ac - bd - 2bc = ad - bc, \tag{13.12}$$

which is the desired result. Notice that this agrees with the determinant of the matrix $\begin{bmatrix}\vec{u} & \vec{v}\end{bmatrix}$. Also notice that if the vectors $\vec{u}$ and $\vec{v}$ get swapped, the *orientation* of the parallelogram also changes. This accounts for the fact that the determinant could be negative. ■

The property of decomposing a column into two parts and calculating the determinant should also be proved, and its proof is actually more intuitive and provides sufficient justification for the result.

Theorem 13.13. *Let $a, b, c, d, e, f > 0$ and suppose that $a > c > e$ and $b > d > f$. Then*

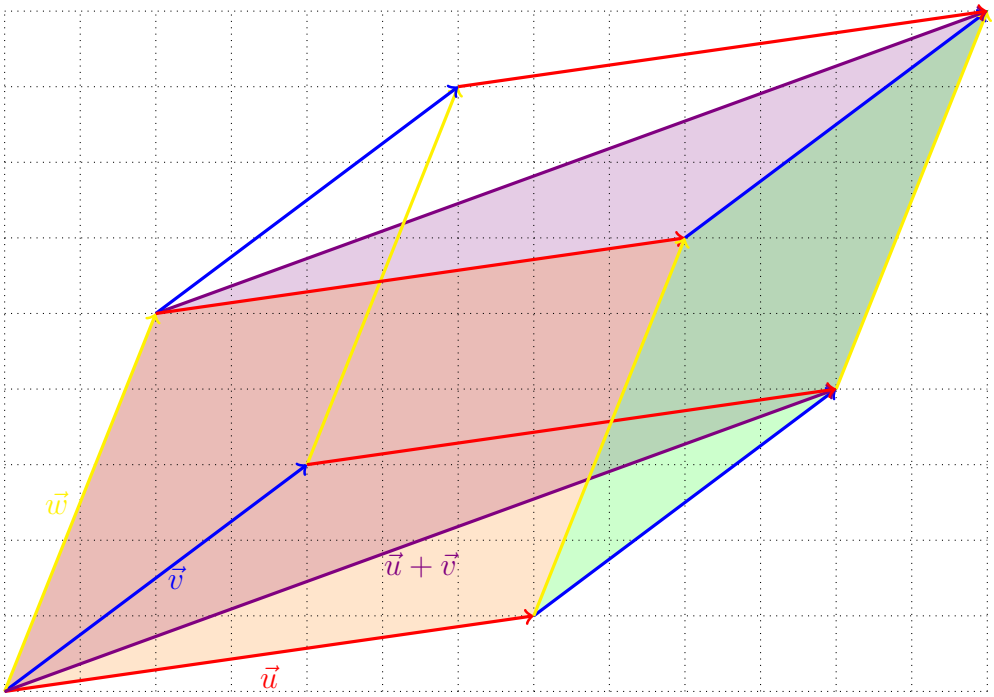
$$\det \begin{bmatrix} a+c & e \\ b+d & f \end{bmatrix} = \det \begin{bmatrix} a & e \\ b & f \end{bmatrix} + \det \begin{bmatrix} c & e \\ d & f \end{bmatrix} \tag{13.14}$$

The result is true regardless of the relationship between the numbers a, b, c, d, e, and f, but one has to keep track of signs.

Proof. Instead of proving this algebraically (which you should be able to do), let's prove it geometrically, which is far more intuitive. Set

$$\vec{u} := \begin{bmatrix} a \\ b \end{bmatrix}, \qquad \vec{v} := \begin{bmatrix} c \\ d \end{bmatrix}, \qquad \vec{w} := \begin{bmatrix} e \\ f \end{bmatrix}. \tag{13.15}$$

Then one obtains the following picture



The area of the orange shaded region, $af - be$, plus the area of the green shaded region, $cf - de$, is equal to the purple shaded region, $(a+c)f - (b+d)e$. This proves the theorem. ■

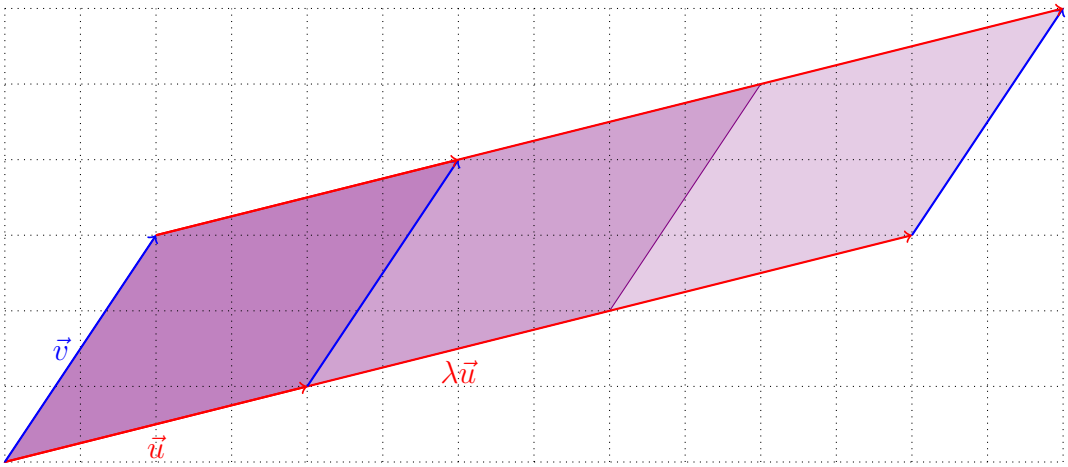
Another simple consequence of the area interpretation of the determinant is what happens when columns are scaled.

Theorem 13.16. *Let $a, b, c, d > 0$ and suppose that $a > b$ and $d > c$. Also, let $\lambda > 0$. Then*

$$\det \begin{bmatrix} \lambda a & b \\ \lambda c & d \end{bmatrix} = \lambda \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{13.17}$$

The result is true regardless of the relationship between the numbers a, b, c, d, and λ, but one has to keep track of signs.

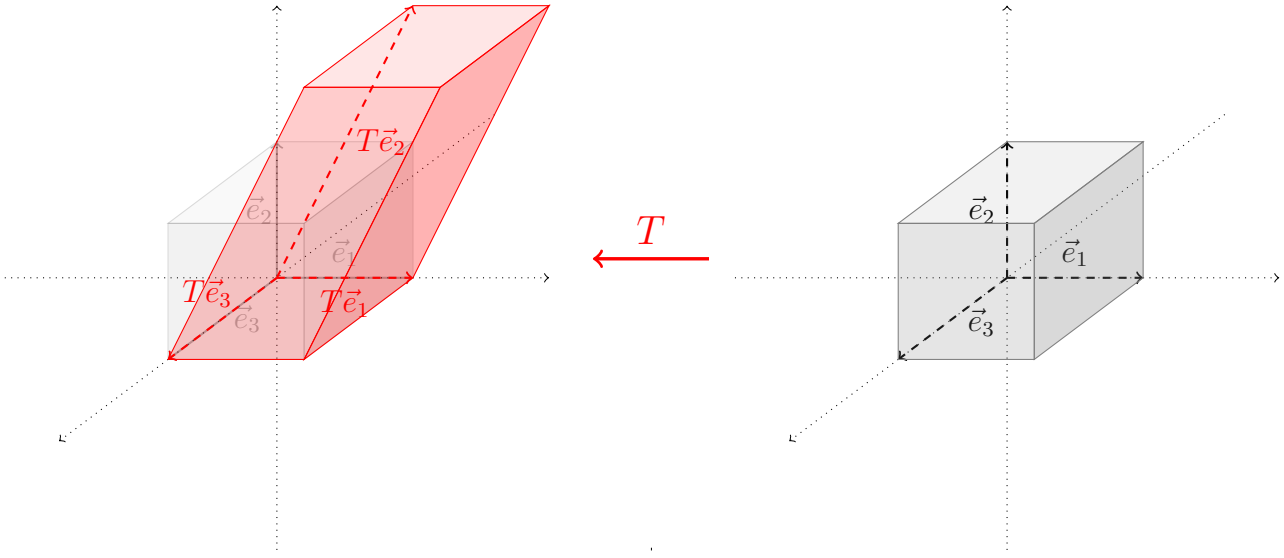
Proof. As before, set $\vec{u} := \begin{bmatrix} a \\ c \end{bmatrix}$ and $\vec{v} := \begin{bmatrix} b \\ d \end{bmatrix}$. For the purposes of the picture, suppose that $\lambda > 1$ (a completely analogous proof holds when $\lambda \leq 1$, and drawing the corresponding picture is left as an exercise). Using the generic picture for $\vec{u}$ and $\vec{v}$ as in the proof of Theorem 13.11 gives



which shows that the area increases by a factor of λ. This proves the claim. ∎

What happens in higher dimensions? Consider the following 3-dimensional example.

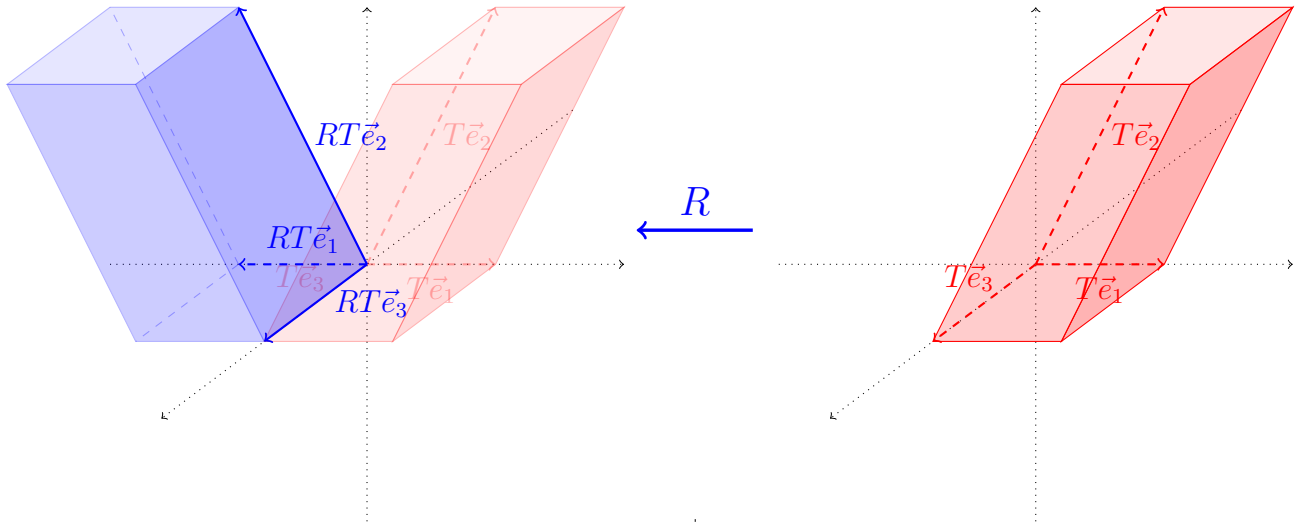
Example 13.18. Consider the transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$ that scales the second unit vector by a factor of 2 and shears everything by one unit along the first unit vector.



The matrix associated to T is

$$[T] = \begin{bmatrix} | & | & | \\ T(\vec{e}_1) & T(\vec{e}_2) & T(\vec{e}_3) \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{13.19}$$

Although we do not have a formula for the determinant yet, we can imagine that the determinant of this transformation is 2 since the volume doubles and the orientation stays the same. However, now consider reflecting through the $\vec{e}_2\vec{e}_3$-plane.



This reflect reverses the orientation and hence has determinant -1. Combined with the scale and shear from before, the transformation RT has determinant -2. As practice, it is useful to verify that the matrix associated to the transformation RT, which can be seen from the picture (by where the blue vectors are) to be

$$[RT] = \begin{bmatrix} | & | & | \\ RT(\vec{e}_1) & RT(\vec{e}_2) & RT(\vec{e}_3) \\ | & | & | \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{13.20}$$

is the matrix product of the transformations $[R]$ and $[T]$

$$[R][T] = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{13.21}$$

The first example also illustrates that the determinant for $m \times m$ matrices itself can be viewed as a function from m vectors in $\mathbb{R}^m$ to the real numbers $\mathbb{R}$. *These m vectors specify the parallelepiped in $\mathbb{R}^m$ and the function* det *gives the signed volume of this parallelepiped.* Before presenting the general *definition* of the determinant of $m \times m$ matrices in the most abstract version by highlighting its essential properties that we have discovered above, we will first describe another *formula* for the determinant of 3×3 matrices, which can, and will, be *derived* from the abstract definition. You might have learned in multivariable calculus that the volume of the parallelepiped P obtained from three vectors $\vec{v}_1, \vec{v}_2$, and $\vec{v}_3$ is given by

$$\left|(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3\right|, \tag{13.22}$$

where $\times$ is the cross product, $\cdot$ is the dot product, and $|\cdot|$ denotes the absolute value of a number. In fact, the orientation of the parallelepiped P is given by the sign, so it is better to write

$$(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3. \tag{13.23}$$

Recall, the dot product of two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ is a number and is given by

$$\vec{u} \cdot \vec{v} := \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = u_1 v_1 + u_2 v_2 + u_3 v_3. \tag{13.24}$$

The cross product of two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ is a vector in $\mathbb{R}^3$ and can be expressed in terms of determinants of 2×2 matrices. It is given by

$$\begin{aligned} \vec{u} \times \vec{v} &:= (u_2 v_3 - u_3 v_2)\vec{e}_1 - (u_1 v_3 - u_3 v_1)\vec{e}_2 + (u_1 v_2 - u_2 v_1)\vec{e}_3 \\ &= \det \begin{bmatrix} u_2 & v_2 \\ u_3 & v_3 \end{bmatrix} \vec{e}_1 - \det \begin{bmatrix} u_1 & v_1 \\ u_3 & v_3 \end{bmatrix} \vec{e}_2 + \det \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} \vec{e}_3. \end{aligned} \tag{13.25}$$

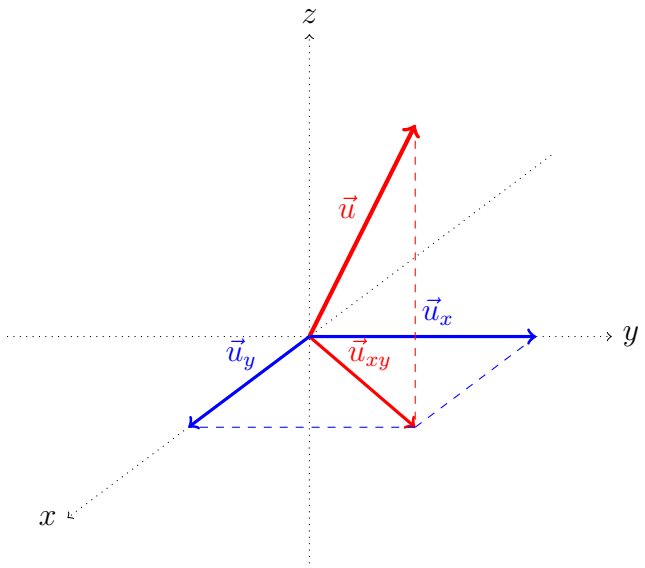
There is a lot of geometric meaning behind these formulas so we should explore them for the moment. Given two vectors $\vec{u}$ and $\vec{v}$, the dot product of $\vec{u}$ with $\vec{v}$ is given by (we'll prove this in a moment)

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos\theta, \tag{13.26}$$

where θ is the angle between $\vec{u}$ and $\vec{v}$ and $\|\cdot\|$ applied to a vector is the length of that vector. Note that the length $\|\vec{u}\|$ of a vector $\vec{u}$ in $\mathbb{R}^3$ is itself given in terms of the dot product by

$$\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}} \tag{13.27}$$

This last identity follows from the Pythagorean Theorem (used twice), as the following picture illustrates.

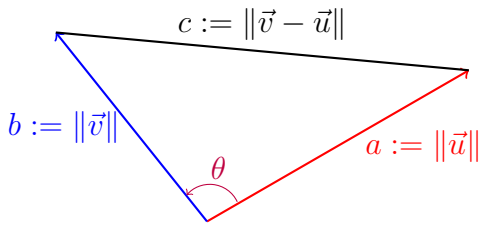


In this picture,

$$\vec{u}_{xy} := \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix}, \qquad \vec{u}_x := \begin{bmatrix} u_1 \\ 0 \\ 0 \end{bmatrix} \qquad \& \qquad \vec{u}_y := \begin{bmatrix} 0 \\ u_2 \\ 0 \end{bmatrix} \tag{13.28}$$

are the projections of the vector $\vec{u}$ onto the xy-plane, the x-axis, and the y-axis, respectively. Now let's prove (13.26).

Proof. (13.26) follows from the law of cosines



which says

$$c^2 = a^2 + b^2 - 2ab\cos\theta. \tag{13.29}$$

In terms of the dot product and the vectors $\vec{u}$ and $\vec{v}$, the law of cosines reads

$$(\vec{v} - \vec{u}) \cdot (\vec{v} - \vec{u}) = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2\|\vec{u}\|\|\vec{v}\| \cos\theta. \tag{13.30}$$

The left-hand-side of (13.30) reduces to

$$(\vec{v} - \vec{u}) \cdot (\vec{v} - \vec{u}) = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2\vec{u} \cdot \vec{v}. \tag{13.31}$$

Two of these terms cancel from equation (13.30) giving

$$-2\vec{u} \cdot \vec{v} = -2\|\vec{u}\|\|\vec{v}\| \cos\theta, \tag{13.32}$$

which reproduces formula (13.26) after dividing by -2. ■

The cross product of $\vec{u}$ and $\vec{v}$ satisfies the following condition[45]

$$\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\|\vec{v}\| \sin\theta, \tag{13.33}$$

where θ is the angle between $\vec{u}$ and $\vec{v}$. This provides the length of the vector $\vec{u} \times \vec{v}$. In fact, this length is the area of the parallelogram obtained from $\vec{u}$ and $\vec{v}$. Notice that it is zero if $\vec{u}$ and $\vec{v}$ are parallel (i.e. one is a scalar multiple of the other). The direction of $\vec{u} \times \vec{v}$ (up to a plus or minus sign) follows from the fact that it satisfies the conditions

$$(\vec{u} \times \vec{v}) \cdot \vec{u} = 0 \qquad \& \qquad (\vec{u} \times \vec{v}) \cdot \vec{v} = 0, \tag{13.34}$$

[45] I don't yet have a nice and easy-to-understand proof of this. I encourage you to think about one. The reason is because any geometric definition of the cross product I am aware of actually uses the right-hand-side (it is related the area of the parallelogram obtained by the vectors $\vec{u}$ and $\vec{v}$).

which can be verified via a quick computation. Setting φ_u be the angle between $\vec{u} \times \vec{v}$ and $\vec{u}$ and setting φ_v to be the angle between $\vec{u} \times \vec{v}$ and $\vec{v}$, these conditions say

$$0 = \|(\vec{u} \times \vec{v})\|\|\vec{u}\| \cos \varphi_u \qquad \& \qquad 0 = \|(\vec{u} \times \vec{v})\|\|\vec{v}\| \cos \varphi_v \tag{13.35}$$

i.e. $\vec{u} \times \vec{v}$ is orthogonal to both $\vec{u}$ and $\vec{v}$ (provided that the lengths of all of these vectors are not zero). So the only thing left to determine the vector $\vec{u} \times \vec{v}$ is the sign of the direction. This is determined by the right-hand-rule, which says that you chop your right hand pinky first onto $\vec{u}$, curl your fingers towards $\vec{v}$, and then your thumb points up towards $\vec{u} \times \vec{v}$.

Going back to our definition for the volume of a parallelepiped in $\mathbb{R}^3$, denote the components of the vectors $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$, as follows

$$\vec{v}_1 = \begin{bmatrix} v_{11} \\ v_{21} \\ v_{31} \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} v_{12} \\ v_{22} \\ v_{32} \end{bmatrix}, \qquad \& \qquad \vec{v}_3 = \begin{bmatrix} v_{13} \\ v_{23} \\ v_{33} \end{bmatrix}. \tag{13.36}$$

Then

$$\vec{v}_1 \times \vec{v}_2 := \begin{bmatrix} v_{21}v_{32} - v_{22}v_{21} \\ v_{12}v_{31} - v_{11}v_{32} \\ v_{11}v_{22} - v_{21}v_{12} \end{bmatrix} \tag{13.37}$$

while the dot product just multiplies corresponding entries and adds everything together

$$(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3 = (v_{21}v_{32} - v_{22}v_{31})v_{13} - (v_{12}v_{31} + v_{11}v_{32})v_{23} + (v_{11}v_{22} - v_{21}v_{12})v_{33} \tag{13.38}$$

In the parallelepiped example from above, we have

$$\left(\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right) \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = -2. \tag{13.39}$$

If we look closely at expression (13.38), we see that this is a linear combination of 2×2 determinants! Namely,

$$(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3 = v_{13} \det \left(\begin{bmatrix} v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix} \right) - v_{23} \det \left(\begin{bmatrix} v_{11} & v_{12} \\ v_{31} & v_{32} \end{bmatrix} \right) + v_{33} \det \left(\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \right). \tag{13.40}$$

We can visualize these determinants together with their appropriate factors in the following way

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} - \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} + \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}. \tag{13.41}$$

In fact, one can show that this is also equal to

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} - \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} + \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}, \tag{13.42}$$

i.e.

$$(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3 = v_{11} \det \begin{bmatrix} v_{22} & v_{23} \\ v_{32} & v_{33} \end{bmatrix} - v_{12} \det \begin{bmatrix} v_{21} & v_{23} \\ v_{31} & v_{33} \end{bmatrix} + v_{13} \det \begin{bmatrix} v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix}. \tag{13.43}$$

This equality of expressions can be thought of as a proof of the identity

$$(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3 = (\vec{v}_2 \times \vec{v}_3) \cdot \vec{v}_1. \tag{13.44}$$

In other words, under the permutation[46]

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \tag{13.45}$$

that sends 1 to 2, sends 2 to 3, and sends 3 to 1, the expression for the determinant does not change. If however, we swapped only two indices once, then we would get a negative sign (check this!):

$$(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3 = -(\vec{v}_2 \times \vec{v}_1) \cdot \vec{v}_3. \tag{13.46}$$

It helps to understand the properties a little more precisely in the 3×3 case.

Exercise 13.47. In this exercise, you will explore the properties of the dot product and then use these *properties* to *reconstruct* the formula for the dot product.

(a) Using the formula for the dot product of 3-component vectors, prove that

$$\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v} \tag{13.48}$$

for all vectors $\vec{v}, \vec{w} \in \mathbb{R}^3$.

(b) Using the formula for the dot product of 3-component vectors prove that

$$\vec{e}_i \cdot \vec{e}_j = \delta_{ij}, \tag{13.49}$$

where

$$\delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{13.50}$$

For example, $\vec{e}_1 \cdot \vec{e}_1 = 1$ while $\vec{e}_1 \cdot \vec{e}_2 = 0$.

(c) Using the formula for the dot product of 3-component vectors, prove that

$$(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w} \tag{13.51}$$

for all vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^3$.

(d) Now, using *only* the results from parts (a) and (c) of this exercise (and *not* the formula for the dot product in terms of the components of the vectors), prove that

$$\vec{v} \cdot (\vec{u} + \vec{w}) = \vec{v} \cdot \vec{u} + \vec{v} \cdot \vec{w} \tag{13.52}$$

for all vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^3$.

[46] The general definition of a permutation is given in the next lecture.

(e) Using *only* the *results* of the previous parts of this exercise (and *not* the formula for the dot product in terms of the components of the vectors) and the fact that all vectors $\vec{v}$ and $\vec{w}$ can be expressed as linear combinations of the unit vectors

$$\vec{v} = \sum_{i=1}^{3} v_i \vec{e}_i \qquad \& \qquad \vec{w} = \sum_{j=1}^{3} w_j \vec{e}_j, \tag{13.53}$$

prove that

$$\vec{v} \cdot \vec{w} = \sum_{i=1}^{3} v_i w_i \tag{13.54}$$

agreeing with the formula we had initially for the dot product.

This shows us that conditions (a), (b), and (c) are enough to *define* the dot product of vectors.

A similar fact holds for the cross product, as the following exercise shows.

Exercise 13.55. In this exercise, you will explore the properties of the cross product and then use these *properties* to *reconstruct* the formula for the cross product.

(a) Using the formula for the cross product of 3-component vectors, prove that

$$\vec{v} \times \vec{w} = -\vec{w} \times \vec{v} \tag{13.56}$$

for all vectors $\vec{v}, \vec{w} \in \mathbb{R}^3$.

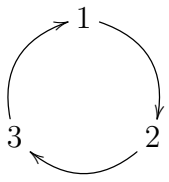
(b) Using the formula for the cross product of 3-component vectors prove that

$$\vec{e}_i \times \vec{e}_j = \sum_{k=1}^{3} \epsilon_{ijk} \vec{e}_k \tag{13.57}$$

where

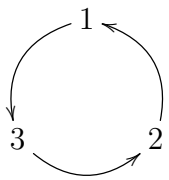
$$\epsilon_{ijk} := \begin{cases} 1 & \text{if } ijk \text{ is a cyclic permutation} \\ -1 & \text{if } ijk \text{ is an anti-cyclic permutation} \\ 0 & \text{if } i = j,\ j = k, \text{ or } i = k. \end{cases} \tag{13.58}$$

Note that a cyclic permutation of $1, 2, 3$ is one that is provided in clockwise order



(13.59)

beginning at any number. An anti-cyclic permutation of $1, 2, 3$ is one that is provided in counter-clockwise order



(13.60)

beginning at any number. So for example, 231 is cyclic while 321 is anti-cyclic. Hence, $\vec{e}_2 \times \vec{e}_3 = \vec{e}_1$ while $\vec{e}_3 \times \vec{e}_2 = -\vec{e}_1$.

(c) Using the formula for the cross product of 3-component vectors, prove that

$$(\vec{u}+\vec{v})\times\vec{w}=\vec{u}\times\vec{w}+\vec{v}\times\vec{w} \tag{13.61}$$

for all vectors $\vec{u},\vec{v},\vec{w}\in\mathbb{R}^3$.

(d) Now, using *only* the results from parts (a) and (c) of this exercise (and *not* the formula for the cross product in terms of the components of the vectors), prove that

$$\vec{v}\times(\vec{u}+\vec{w})=\vec{v}\times\vec{u}+\vec{v}\times\vec{w} \tag{13.62}$$

for all vectors $\vec{u},\vec{v},\vec{w}\in\mathbb{R}^3$.

(e) Using *only* the *results* of the previous parts of this exercise (and *not* the formula for the cross product in terms of the components of the vectors) and the fact that all vectors $\vec{v}$ and $\vec{w}$ can be expressed as linear combinations of the unit vectors

$$\vec{v}=\sum_{i=1}^{3}v_i\vec{e}_i \qquad \& \qquad \vec{w}=\sum_{j=1}^{3}w_j\vec{e}_j, \tag{13.63}$$

prove that

$$\vec{v}\times\vec{w}=\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}v_iw_j\epsilon_{ijk}\vec{e}_k, \tag{13.64}$$

and *then* express this result as a 3-component vector by showing that it equals

$$\vec{v}\times\vec{w}=\begin{bmatrix}v_2w_3-v_3w_2\\ v_3w_1-v_1w_3\\ v_1w_2-v_2w_1\end{bmatrix} \tag{13.65}$$

agreeing with the formula we had initially for the cross product.

This shows us that conditions (a), (b), and (c) are enough to *define* the cross product of vectors.

Therefore, since the determinant is expressed in terms of the cross product and the dot product, it, too, can be characterized by similar properties, as we will see later.

Exercise 13.66. Construct a geometric proof of the determinant formula for a 3×3 matrix analogous to the geometric construction of the determinant for 2×2 matrices. In other words, prove that the (signed) volume obtained from three linearly independent vectors $\vec{u},\vec{v},\vec{w}$ is

$$\det\begin{bmatrix}| & | & |\\ \vec{u} & \vec{v} & \vec{w}\\ | & | & |\end{bmatrix}. \tag{13.67}$$

As an initial step, first provide a geometric proof that

$$\det\begin{bmatrix}1 & 0 & 0\\ 0 & a & b\\ 0 & c & d\end{bmatrix}=\det\begin{bmatrix}a & b\\ c & d\end{bmatrix}. \tag{13.68}$$

Recommended Exercises. Please check HuskyCT for the homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we discussed parts of Sections 3.2, 3.3, and 6.1 of [Lay].

Terminology checklist

determinant as assigned volume	
length of a vector	
cross product of vectors in $\mathbb{R}^3$	
dot product of vectors	
permutation	

14 The determinant and the formula for the inverse of a matrix

All of the *properties* that we have discovered for the determinant of 2×2 and 3×3 matrices in terms of signed volumes can be turned into a *definition* for arbitrary $m \times m$ matrices. We will therefore think of the determinant of an $m \times m$ matrix as the signed volume of the parallelepiped obtained from the m columns of the matrix. In other words, the determinant should be a function sending m m-component vectors to some real number, which is to be interpreted as the signed volume of the parallelepiped obtained from these m vectors. It should satisfy all of the properties discussed in the previous section. Namely,

(a) the signed volume changes sign when any two vectors are swapped,

(b) the signed volume is linear in each column,

(c) the signed volume of the unit cube should be 1.

Definition 14.1. The *determinant* for $m \times m$ matrices is a function[47]

$$\det : \overbrace{\mathbb{R}^m \times \cdots \times \mathbb{R}^m}^{m \text{ times}} \to \mathbb{R}, \tag{14.2}$$

which we think of as assigning to m m-component vectors aligned as in a matrix

$$\begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} \mapsto \det(\vec{v}_1, \ldots, \vec{v}_n), \tag{14.3}$$

satisfying the following conditions.

(a) For every m-tuple of vectors $(\vec{v}_1, \ldots, \vec{v}_m)$ in $\mathbb{R}^m$,

$$\det\Big(\vec{v}_1, \ldots, \overset{\text{switch}}{\overbrace{\vec{v}_i}}, \ldots, \overset{\text{switch}}{\overbrace{\vec{v}_j}}, \ldots, \vec{v}_m\Big) = -\det\Big(\vec{v}_1, \ldots, \vec{v}_j, \ldots, \vec{v}_i, \ldots, \vec{v}_m\Big). \tag{14.4}$$

This is sometimes called the *skew-symmetry* of det .

(b) det is *multilinear*, i.e.

$$\det\Big(\vec{v}_1, \ldots, a\vec{v}_i + b\vec{u}_i, \ldots, \vec{v}_m\Big) = a\det\Big(\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_m\Big) + b\det\Big(\vec{v}_1, \ldots, \vec{u}_i, \ldots, \vec{v}_m\Big) \tag{14.5}$$

for all $i = 1, \ldots, m$ all scalars a, b and all vectors $\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_i, \vec{u}_i, \vec{v}_{i+1}, \ldots, \vec{v}_m$.

[47] I found this description of the determinant at http://math.stackexchange.com/questions/668/whats-an-intuitive-way-to-think-about-the-determinant

(c) The determinant of the unit vectors, listed in order, is 1:

$$\det\Big(\vec{e}_1,\ldots,\vec{e}_m\Big)=1. \tag{14.6}$$

For an $m\times m$ matrix A, the input for the determinant function consists of the columns of A written in order:

$$\det A:=\det\Big(A\vec{e}_1,\cdots,A\vec{e}_m\Big). \tag{14.7}$$

Corollary 14.8. *Let $(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_j,\ldots,\vec{v}_m)$ be a sequence of vectors in $\mathbb{R}^m$ where $\vec{v}_i=\vec{v}_j$ (and yet $i\neq j$). Then*

$$\det\Big(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_i,\ldots,\vec{v}_m\Big)=0. \tag{14.9}$$

Proof. By the skew-symmetry condition (a) in the definition of determinant,

$$\det\Big(\vec{v}_1,\ldots,\overset{\text{switch}}{\overbrace{\vec{v}_i,\ldots,\vec{v}_j}},\ldots,\vec{v}_m\Big)=-\det\Big(\vec{v}_1,\ldots,\vec{v}_j,\ldots,\vec{v}_i,\ldots,\vec{v}_m\Big) \tag{14.10}$$

but since $\vec{v}_i=\vec{v}_j$

$$-\det\Big(\vec{v}_1,\ldots,\vec{v}_j,\ldots,\vec{v}_i,\ldots,\vec{v}_m\Big)=-\det\Big(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big). \tag{14.11}$$

Putting these two equalities together gives

$$\det\Big(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big)=-\det\Big(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big). \tag{14.12}$$

A number can only equal the negative of itself if it is zero. Hence,

$$\det\Big(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big)=0 \tag{14.13}$$

whenever $\vec{v}_i=\vec{v}_j$ and $i\neq j$. ∎

Corollary 14.14. *Let $(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_j,\ldots,\vec{v}_m)$ be a list of m vectors in $\mathbb{R}^m$. Then*

$$\det\Big(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big)=\det\Big(\vec{v}_1,\ldots,\vec{v}_i+\vec{v}_j,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big) \tag{14.15}$$

for any j between 1 and m.

Proof. This follows immediately from multilinearity and the previous Corollary

$$\begin{aligned}\det\Big(\vec{v}_1,\ldots,\vec{v}_i+\vec{v}_j,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big)&=\det\Big(\vec{v}_1,\ldots,\vec{v}_i,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big)\\&\quad+\underbrace{\det\Big(\vec{v}_1,\ldots,\vec{v}_j,\ldots,\vec{v}_j,\ldots,\vec{v}_m\Big)}_{0}\end{aligned} \tag{14.16}$$

because $\vec{v}_j$ repeats itself in the argument of det. ∎

Let's check to make sure this definition reduces to the determinant of a 2×2 matrix. Let

$$A := \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{14.17}$$

Then the columns of A are

$$\left(\begin{bmatrix} a \\ c \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix} \right) = (a\vec{e}_1 + c\vec{e}_2, b\vec{e}_1 + d\vec{e}_2). \tag{14.18}$$

Therefore,

$$\begin{aligned} \det A &= \det(a\vec{e}_1 + c\vec{e}_2, b\vec{e}_1 + d\vec{e}_2) \\ &= a \det(\vec{e}_1, b\vec{e}_1 + d\vec{e}_2) + c \det(\vec{e}_2, b\vec{e}_1 + d\vec{e}_2) \\ &= ab \underbrace{\det(\vec{e}_1, \vec{e}_1)}_{0} + ad \underbrace{\det(\vec{e}_1, \vec{e}_2)}_{1} + cb \underbrace{\det(\vec{e}_2, \vec{e}_1)}_{-1} + cd \underbrace{\det(\vec{e}_2, \vec{e}_2)}_{0} \\ &= ad - bc. \end{aligned} \tag{14.19}$$

Wow! This abstract technique actually worked! And we didn't have to memorize the specific formulas for all the different cases of 2×2, 3×3, or more general matrices. All we have to remember are the three conditions in Definition 14.1. Granted, it takes *longer* to use this definition right now, and we will learn a particular pattern that will help us soon.

Example 14.20. Let's try an example in $\mathbb{R}^3$ this time finding out the more explicit formula. Let

$$A := \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}. \tag{14.21}$$

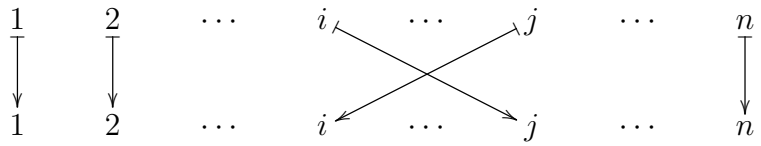
Then

$$\begin{aligned} \det A &= \det(A\vec{e}_1, A\vec{e}_2, A\vec{e}_3) \\ &= \det(\vec{e}_1 + 2\vec{e}_2, \vec{e}_1 - \vec{e}_3, \vec{e}_2 + \vec{e}_3) \\ &= \det(\vec{e}_1, \vec{e}_1 - \vec{e}_3, \vec{e}_2 + \vec{e}_3) + 2\det(\vec{e}_2, \vec{e}_1 - \vec{e}_3, \vec{e}_2 + \vec{e}_3) \\ &= \underbrace{\det(\vec{e}_1, \vec{e}_1, \vec{e}_2 + \vec{e}_3)}_{0} - \det(\vec{e}_1, \vec{e}_3, \vec{e}_2 + \vec{e}_3) \\ &\qquad + 2\det(\vec{e}_2, \vec{e}_1, \vec{e}_2 + \vec{e}_3) - 2\det(\vec{e}_2, \vec{e}_3, \vec{e}_2 + \vec{e}_3) \\ &= -\underbrace{\det(\vec{e}_1, \vec{e}_3, \vec{e}_2)}_{-1} - \underbrace{\det(\vec{e}_1, \vec{e}_3, \vec{e}_3)}_{0} + 2\underbrace{\det(\vec{e}_2, \vec{e}_1, \vec{e}_2)}_{0} \\ &\qquad + 2\underbrace{\det(\vec{e}_2, \vec{e}_1, \vec{e}_3)}_{-1} - 2\underbrace{\det(\vec{e}_2, \vec{e}_3, \vec{e}_2)}_{0} - 2\underbrace{\det(\vec{e}_2, \vec{e}_3, \vec{e}_3)}_{0} \\ &= -1. \end{aligned} \tag{14.22}$$

We can also calculate the determinant using our first formula in terms of the signed volume of the parallelepiped.

With this general idea, you can calculate determinants of very large matrices as well, though perhaps it may take a lot of time. We will try to do this by first recalling the definition of a permutation. We have already discussed the special case of permutations of two and three indices, but we need a more general definition for n indices.

Definition 14.23. A *permutation* of a list of numbers $(1, 2, \ldots, n)$ is a rearrangement of these same numbers in a different order $\big(\sigma(1), \sigma(2), \ldots, \sigma(n)\big)$. An *elementary permutation* is a permutation for which only two numbers switch places



(14.24)

Theorem 14.25. *Fix a positive integer n. Every permutation of $(1, 2, \ldots, n)$ is obtained from successive elementary permutations. Furthermore, the number of such elementary permutations is either always even or always odd for a given permutation.*

Definition 14.26. If a permutation can be expressed as an even number of elementary permutations, then its *sign* is $+1$. Otherwise, its sign is -1. The sign of a permutation σ is written as $\text{sign}(\sigma)$.

An arbitrary permutation is often written as

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix} \tag{14.27}$$

As an example, for a permutation of 3 numbers

$$\begin{pmatrix} 1 & 2 & 3 \\ i & j & k \end{pmatrix}, \tag{14.28}$$

the sign reproduces something we have already seen before

$$\text{sign} \begin{pmatrix} 1 & 2 & 3 \\ i & j & k \end{pmatrix} = \epsilon_{ijk}, \tag{14.29}$$

where ϵ_{ijk} was defined in (13.58). Now, let A be an arbitrary $m \times m$ matrix as in

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} \tag{14.30}$$

The j-th column of A is

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} = \sum_{i_j=1}^{m} a_{i_j j} \vec{e}_{i_j}. \tag{14.31}$$

Therefore,

$$\begin{aligned}\det A &= \det\left(\sum_{i_1=1}^{m} a_{i_1 1}\vec{e}_{i_1}, \ldots, \sum_{i_m=1}^{m} a_{i_m m}\vec{e}_{i_m}\right)\\ &= \sum_{i_1=1}^{m}\cdots\sum_{i_m=1}^{m} a_{i_1 1}\cdots a_{i_m m}\det\left(\vec{e}_{i_1}, \ldots, \vec{e}_{i_m}\right)\\ &= \sum_{\substack{i_1=1,\ldots,i_m=1\\ i_1\neq i_2\neq\cdots\neq i_m}}^{m} a_{i_1 1}\cdots a_{i_m m}\text{sign}(\sigma_{i_1\ldots i_m}),\end{aligned} \tag{14.32}$$

where $\sigma_{i_1\ldots i_m}$ is the permutation defined by

$$\begin{pmatrix}1 & 2 & \cdots & m\\ i_1 & i_2 & \cdots & i_m\end{pmatrix}. \tag{14.33}$$

In the second equality in (14.32), the multi-linearity of det was used. In the third equality, the skew-symmetry of det was used together with the fact that the determinant of the identity matrix is 1.

Example 14.34. Let's use this formula to compute the determinant of a 3×3 matrix and check that it agrees with our previous definition. Let the 3×3 matrix be of the form

$$A := \begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix} \tag{14.35}$$

Then,

$$\begin{aligned}\det A &= \sum_{\substack{i_1=1,i_2=1,i_3=1\\ i_1\neq i_2\neq i_3}}^{3} a_{i_1 1}a_{i_2 2}a_{i_3 3}\text{sign}(\sigma_{i_1 i_2 i_3})\\ &= a_{11}\sum_{\substack{i_2=2,i_3=2\\ i_2\neq i_3}}^{3} a_{i_2 2}a_{i_3 3}\text{sign}(\sigma_{1 i_2 i_3}) + a_{12}\sum_{\substack{i_1=2,i_3=2\\ i_1\neq i_3}}^{3} a_{i_1 1}a_{i_3 3}\text{sign}(\sigma_{i_1 1 i_3}) + a_{13}\sum_{\substack{i_1=2,i_2=2\\ i_1\neq i_2}}^{3} a_{i_1 1}a_{i_2 2}\text{sign}(\sigma_{i_1 i_2 1})\\ &= a_{11}\sum_{\substack{i_2=2,i_3=2\\ i_2\neq i_3}}^{3} a_{i_2 2}a_{i_3 3}\text{sign}(\sigma_{i_2 i_3}) - a_{12}\sum_{\substack{i_1=2,i_3=2\\ i_1\neq i_3}}^{3} a_{i_1 1}a_{i_3 3}\text{sign}(\sigma_{i_1 i_3}) + a_{13}\sum_{\substack{i_1=2,i_2=2\\ i_1\neq i_2}}^{3} a_{i_1 1}a_{i_2 2}\text{sign}(\sigma_{i_1 i_2})\\ &= a_{11}(a_{22}a_{33}-a_{32}a_{23}) - a_{12}(a_{21}a_{33}-a_{31}a_{23}) + a_{13}(a_{21}a_{32}-a_{31}a_{22})\\ &= a_{11}\det\begin{bmatrix}a_{22} & a_{23}\\ a_{32} & a_{33}\end{bmatrix} - a_{12}\det\begin{bmatrix}a_{21} & a_{23}\\ a_{31} & a_{33}\end{bmatrix} + a_{13}\det\begin{bmatrix}a_{21} & a_{22}\\ a_{31} & a_{32}\end{bmatrix}\end{aligned} \tag{14.36}$$

by some rearrangements of the terms. We therefore see that this agrees with our initial definition! We can also re-write this expression in terms of the 2×2 matrices by introducing the notation

$$\det A = a_{11}\det A_{11} - a_{12}\det A_{12} + a_{13}\det A_{13}. \tag{14.37}$$

A_{1j} is the resulting matrix obtained from A by deleting the 1-st row and j-th column

$$A_{11} := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad A_{12} := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \& \quad A_{13} := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{14.38}$$

Note that we could have chosen to do this rearrangement by separating out the second or third row instead of the first (we have already analyzed a similar situation earlier). Just to illustrate the slight difference, let's separate out the second row explicitly.

$$\begin{aligned} \det A &= \sum_{\substack{i_1=1,i_2=1,i_3=1 \\ i_1\neq i_2\neq i_3}}^{3} a_{i_1 1} a_{i_2 2} a_{i_3 3} \mathrm{sign}(\sigma_{i_1 i_2 i_3}) \\ &= a_{21} \sum_{\substack{i_2=1,i_3=1 \\ i_2\neq i_3\neq 2}}^{3} a_{i_2 2} a_{i_3 3} \mathrm{sign}(\sigma_{2 i_2 i_3}) + a_{22} \sum_{\substack{i_1=1,i_3=1 \\ i_1\neq i_3\neq 2}}^{3} a_{i_1 1} a_{i_3 3} \mathrm{sign}(\sigma_{i_1 2 i_3}) + a_{23} \sum_{\substack{i_1=1,i_2=1 \\ i_1\neq i_2\neq 2}}^{3} a_{i_1 1} a_{i_2 2} \mathrm{sign}(\sigma_{i_1 i_2 2}) \\ &= -a_{21} \sum_{\substack{i_2=1,i_3=1 \\ i_2\neq i_3\neq 2}}^{3} a_{i_2 2} a_{i_3 3} \mathrm{sign}(\sigma_{i_2 i_3}) + a_{22} \sum_{\substack{i_1=1,i_3=1 \\ i_1\neq i_3\neq 2}}^{3} a_{i_1 1} a_{i_3 3} \mathrm{sign}(\sigma_{i_1 i_3}) - a_{23} \sum_{\substack{i_1=1,i_2=1 \\ i_1\neq i_2\neq 2}}^{3} a_{i_1 1} a_{i_2 2} \mathrm{sign}(\sigma_{i_1 i_2}) \\ &= -a_{21}(a_{12}a_{33} - a_{32}a_{13}) + a_{22}(a_{11}a_{33} - a_{31}a_{13}) - a_{23}(a_{11}a_{32} - a_{31}a_{12}) \\ &= -a_{21} \det A_{21} + a_{22} \det A_{22} - a_{23} A_{23}, \end{aligned} \tag{14.39}$$

where

$$A_{21} := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad A_{22} := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \& \quad A_{23} := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{14.40}$$

The preceding example is quite general. A similar calculation gives an inductive formula for the determinant of an $m \times m$ matrix A as in (14.30) in terms of determinants of $(m-1) \times (m-1)$ matrices. A completely similar calculation, just more involved, gives

$$\det A = \sum_{j=1}^{m} (-1)^{j+1} a_{1j} \det A_{1j} \tag{14.41}$$

where A_{1j} is the $(m-1) \times (m-1)$ matrix obtained by deleting the first row of A and the j-th column of A as in

$$A_{1j} := \begin{bmatrix} a_{11} & \cdots & a_{1j-1} & a_{1j} & a_{1j+1} & \cdots & a_{1m} \\ a_{21} & \cdots & a_{2j-1} & a_{2j} & a_{2j+1} & \cdots & a_{2m} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj-1} & a_{mj} & a_{mj+1} & \cdots & a_{mm} \end{bmatrix}, \tag{14.42}$$

i.e.

$$A_{1j} = \begin{bmatrix} a_{21} & \cdots & a_{2j-1} & a_{2j+1} & \cdots & a_{2m} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj-1} & a_{mj+1} & \cdots & a_{mm} \end{bmatrix}. \tag{14.43}$$

More generally, we could have chosen the i-th row and the j-th column in calculating this determinant

$$\det A = \sum_{j=1}^{m} (-1)^{j+i} a_{ij} \det A_{ij}, \tag{14.44}$$

where

$$A_{ij} := \begin{bmatrix} a_{11} & \cdots & a_{1j-1} & a_{1j} & a_{1j+1} & \cdots & a_{1m} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i-11} & \cdots & a_{i-1j-1} & a_{i-1j} & a_{i-1j+1} & \cdots & a_{i-1m} \\ a_{i1} & \cdots & a_{ij-1} & a_{ij} & a_{ij+1} & \cdots & a_{im} \\ a_{i+11} & \cdots & a_{i+1j-1} & a_{i+1j} & a_{i+1j+1} & \cdots & a_{i+1m} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj-1} & a_{mj} & a_{mj+1} & \cdots & a_{mm} \end{bmatrix}, \tag{14.45}$$

i.e.

$$A_{ij} = \begin{bmatrix} a_{11} & \cdots & a_{1j-1} & a_{1j+1} & \cdots & a_{1m} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{i-11} & \cdots & a_{i-1j-1} & a_{i-1j+1} & \cdots & a_{i-1m} \\ a_{i+11} & \cdots & a_{i+1j-1} & a_{i+1j+1} & \cdots & a_{i+1m} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj-1} & a_{mj+1} & \cdots & a_{mm} \end{bmatrix}. \tag{14.46}$$

Definition 14.47. Let A be an $m \times m$ matrix as in the previous lecture. In the previous formula for the determinant,

$$C_{ij} := (-1)^{i+j} \det A_{ij} \tag{14.48}$$

is called the $\underline{(i,j)\text{-}cofactor}$ of A.

Theorem 14.49. *Let A be an $m \times m$ matrix. A is invertible if and only if* $\det A \neq 0$. *Furthermore, when this happens,*

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{m1} \\ C_{12} & C_{22} & \cdots & C_{m2} \\ \vdots & \vdots & & \vdots \\ C_{1m} & C_{2m} & \cdots & C_{mm} \end{bmatrix}, \tag{14.50}$$

where C_{ij} is the (i,j)-cofactor of A.

Recall the definition of the transpose of a matrix.

Definition 14.51. The $\underline{transpose}$ of an $m \times n$ matrix A

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{14.52}$$

is the $n \times m$ matrix

$$A^T := \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}. \tag{14.53}$$

We will provide several interpretations of the transpose later in this course.

Theorem 14.54. *Let A be an $m \times m$ matrix. Then*

$$\det A^T = \det A. \tag{14.55}$$

Proof. One can prove this directly from the formula for the determinant. ■

Exercise 14.56. Provide a geometric proof of

$$\det \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{14.57}$$

Definition 14.58. The matrix

$$\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{m1} \\ C_{12} & C_{22} & \cdots & C_{m2} \\ \vdots & \vdots & & \vdots \\ C_{1m} & C_{2m} & \cdots & C_{mm} \end{bmatrix} \equiv \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1m} \\ C_{21} & C_{22} & \cdots & C_{2m} \\ \vdots & \vdots & & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mm} \end{bmatrix}^T \tag{14.59}$$

in Theorem 14.49 is called the *adjugate* of A. It is often written as $\mathrm{adj}A$.

From the definition of the transpose of a matrix, the definition of the determinant, and Corollary 14.14, one can calculate the determinant using row operations. Swapping rows introduces a minus sign from the definition of det . Adding rows does not change anything by this theorem and Corollary 14.14. And in your homework, you are asked to find a formula for the determinant of an upper-triangular matrix. In other words, if a matrix A is row reduced to a matrix A' *without* any scaling of rows, then

$$\det A = (-1)^{\text{number of row swaps}} \det A'. \tag{14.60}$$

This is a faster method for computing the determinant.

Let's look back at Example 14.34 by using these results to calculate the inverse of a specific 3×3 matrix.

Example 14.61. Let's find the inverse of the matrix

$$A := \begin{bmatrix} 1 & 1 & 3 \\ 2 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{14.62}$$

The determinant of A is

$$\det A = 1 \cdot \det \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} = 1 - 6 = -5. \tag{14.63}$$

The adjugate matrix of A is given by

$$\begin{aligned}\mathrm{adj}A &:= \begin{bmatrix} \det\begin{bmatrix}-2 & 1\\ 1 & 0\end{bmatrix} & -\det\begin{bmatrix}2 & 1\\ 0 & 0\end{bmatrix} & \det\begin{bmatrix}2 & -2\\ 0 & 1\end{bmatrix} \\ -\det\begin{bmatrix}1 & 3\\ 1 & 0\end{bmatrix} & \det\begin{bmatrix}1 & 3\\ 0 & 0\end{bmatrix} & -\det\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix} \\ \det\begin{bmatrix}1 & 3\\ -2 & 1\end{bmatrix} & -\det\begin{bmatrix}1 & 3\\ 2 & 1\end{bmatrix} & \det\begin{bmatrix}1 & 1\\ 2 & -2\end{bmatrix}\end{bmatrix}^T \\ &= \begin{bmatrix}-1 & 0 & 2\\ 3 & 0 & -1\\ 7 & 5 & -4\end{bmatrix}^T \\ &= \begin{bmatrix}-1 & 3 & 7\\ 0 & 0 & 5\\ 2 & -1 & -4\end{bmatrix}\end{aligned}. \tag{14.64}$$

Therefore,

$$A^{-1} = \frac{1}{\det A}\mathrm{adj}A = \frac{1}{5}\begin{bmatrix}-1 & 3 & 7\\ 0 & 0 & 5\\ 2 & -1 & -4\end{bmatrix}. \tag{14.65}$$

Let's just make sure this works:

$$\begin{aligned}A^{-1}A &= \frac{1}{5}\begin{bmatrix}-1 & 3 & 7\\ 0 & 0 & 5\\ 2 & -1 & -4\end{bmatrix}\begin{bmatrix}1 & 1 & 3\\ 2 & -2 & 1\\ 0 & 1 & 0\end{bmatrix} \\ &= \frac{1}{5}\begin{bmatrix}-1+6+0 & -1-6+7 & -3+3+0\\ 0+0+0 & 0+0+5 & 0+0+0\\ 2-2+0 & 2+2-4 & 6-1+0\end{bmatrix} \\ &= \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}.\end{aligned} \tag{14.66}$$

Exercise 14.67. Let A be a square matrix such that $A^k = \mathbb{1}$ for some positive integer k and $A^j \neq \mathbb{1}$ for all $j \in \{2, \ldots, k-1\}$. Show that A is invertible and find its inverse.

Exercise 14.68. Let A be a 2×2 matrix such that $A^2 = 0$, where 0 is the 2×2 matrix of all zeros. Show that $\mathbb{1} - A$ is invertible and find its inverse.

Exercise 14.69. Let A be a square matrix such that $A^k = 0$ for some positive integer k. Show that $\mathbb{1} - A$ is invertible and find its inverse.

Recommended Exercises. See homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we finished Sections 3.1, 3.2, and 3.3 of [Lay]. Notice that we skipped Cramer's rule.

Terminology checklist

determinant	
skew-symmetric	
multilinear	
the sign of a permutation	
cofactor	
transpose	
adjugate	
inverse of a matrix	

15 Orthogonality

The dot product can be used to define the length and orthogonality of vectors in Euclidean space. Later, we will use this idea to diagonalize certain kinds of matrices. Along the way, we will discuss projections, which are special examples of linear transformations. To get a better understanding of the dot product, we have the following theorems, which include some useful inequelities for the dot product in Euclidean space. In what follows, the notation $\langle \vec{v}, \vec{u} \rangle$ will be used to denote the dot product, $\vec{v} \cdot \vec{u}$, which is also often referred to as the inner product of $\vec{v}$ and $\vec{u}$. We will use this notation to avoid confusing this with matrix multiplication.

Definition 15.1. The *Euclidean norm/length* on $\mathbb{R}^n$ is the function $\mathbb{R}^n \to \mathbb{R}$ defined by

$$\big\|(x_1, x_2, \ldots, x_n)\big\| := \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{15.2}$$

The *Euclidean inner product* on $\mathbb{R}^n$ is the function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\big\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \big\rangle := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n. \tag{15.3}$$

Often, the short-hand notation

$$\sum_{i=1}^{n} z_i := z_1 + z_2 + \cdots + z_n \tag{15.4}$$

will be used to denote the sum of n real numbers $z_i \in \mathbb{R}^n, i \in \{1, \ldots, n\}$.

Theorem 15.5. *$\mathbb{R}^n$ with these structures satisfies the following for all vectors $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ and for all numbers $c \in \mathbb{R}$,*

(a) $\langle \vec{x}, \vec{x} \rangle \geq 0$ *and* $\langle \vec{x}, \vec{x} \rangle = 0$ *if and only if* $\vec{x} = 0$,

(b) $\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}$,

(c) $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$,

(d) $\langle c\vec{x}, \vec{y} \rangle = c\langle \vec{x}, \vec{y} \rangle = \langle \vec{x}, c\vec{y} \rangle$,

(e) $\langle \vec{x} + \vec{z}, \vec{y} \rangle = \langle \vec{x}, \vec{y} \rangle + \langle \vec{z}, \vec{y} \rangle$ *and* $\langle \vec{x}, \vec{y} + \vec{z} \rangle = \langle \vec{x}, \vec{y} \rangle + \langle \vec{y}, \vec{z} \rangle$,

(f) $|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \|\vec{y}\|$ *and equality holds if and only if $\vec{x}$ and $\vec{y}$ are linearly dependent (Cauchy-Schwarz inequality),*

(g) $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ *(triangle inequality),*

(h) $\langle \vec{x}, \vec{y} \rangle = \frac{\|\vec{x}+\vec{y}\|^2 - \|\vec{x}-\vec{y}\|^2}{4}$ *(polarization identity).*

Proof.

(a) Let $(x_1, \ldots, x_n) := \vec{x}$ denote the components of $\vec{x}$. Then,

$$\langle \vec{x}, \vec{x} \rangle = \sum_{i=1}^{n} x_i^2 \geq 0 \tag{15.6}$$

since the square of any real number is always at least 0. Furthermore, the sum of squares is zero if and only if each term is zero, but the square root of zero is zero so $x_i = 0$ for all i if and only if $\langle \vec{x}, \vec{x} \rangle = 0$.

(b) This follows immediately from the definitions of $\| \cdot \|$ in (15.2) and $\langle \cdot \, , \, \cdot \rangle$ in (15.3).

(c) This follows from commutativity of multiplication of real numbers and the formula (15.3).

(d) This follows from commutativity and associativity of multiplication of real numbers and the formula (15.3).

(e) This follows from the distributive law for real numbers and the formula (15.3).

(f) Suppose that $\vec{x}$ and $\vec{y}$ are linearly independent. Therefore, $\vec{x} \neq \lambda \vec{y}$ for all all $\lambda \in \mathbb{R}$. Hence,

$$0 < \|\lambda \vec{y} - \vec{x}\|^2 = \sum_{i=1}^{n} (\lambda y_i - x_i)^2 = \lambda^2 \sum_{i=1}^{n} y_i^2 - 2\lambda \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} x_i^2. \tag{15.7}$$

In particular, this is a quadratic equation in the variable λ that has no real solutions. Hence,

$$\left(-2 \sum_{i=1}^{n} x_i y_i \right)^2 - 4 \left(\sum_{i=1}^{n} y_i^2 \right) \left(\sum_{j=1}^{n} x_i^2 \right) < 0. \tag{15.8}$$

Rewriting this and canceling out the common factor of 4 gives

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_i x_j y_j < \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2 y_j^2 \tag{15.9}$$

Applying the square root to both sides gives the desired result.

Now suppose $\vec{x} = \lambda \vec{y}$ for some $\lambda \in \mathbb{R}$. Then by parts (b) and (d), equality holds.

(g) Notice that by the Cauchy Schwarz inequality,

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 &= \sum_{i=1}^{n} (x_i + y_i)^2 \\ &= \sum_{i=1}^{n} x_i^2 + 2 \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 \\ &\leq \sum_{i=1}^{n} x_i^2 + 2 \left| \sum_{i=1}^{n} x_i y_i \right| + \sum_{i=1}^{n} y_i^2 \\ &\leq \|\vec{x}\|^2 + 2\|\vec{x}\|\|\vec{y}\| + \|\vec{y}\|^2 \\ &= \big(\|\vec{x}\| + \|\vec{y}\| \big)^2 \end{aligned} \tag{15.10}$$

Applying the square root and using parts (a) and (b) gives the desired result.

(h) This calculation is left to the reader.

■

Definition 15.11. Two vectors $\vec{u}$ and $\vec{v}$ are *orthogonal/perpendicular* whenever $\langle \vec{v}, \vec{u} \rangle = 0$.

The reason for this definition comes from the following. First, we can also use the inner product to define angles between vectors.

Definition 15.12. Let $\vec{v}$ and $\vec{u}$ be two nonzero vectors in $\mathbb{R}^n$. The *angle* θ from $\vec{v}$ to $\vec{u}$

$$\theta := \arccos\left(\frac{\langle \vec{v}, \vec{u} \rangle}{\|\vec{v}\|\|\vec{u}\|}\right). \tag{15.13}$$

The motivation for this definition comes from what happens in two and three dimensions, which we saw in (13.26). We can't prove this in arbitrary dimensions because we don't know what the law of cosines means in higher dimensions. Instead, we use this formula as a *definition*.[48] Now, in terms of the definition of angle, the definition of orthogonality says that the angle between $\vec{v}$ and $\vec{u}$ is an odd integer multiple of $\frac{\pi}{2}$, as we would expect from our common-day use of the word orthogonal/perpendicular.

Definition 15.14. A vector $\vec{u}$ in $\mathbb{R}^n$ of length 1 is called a *normalized/unit vector*. If $\vec{v}$ is any nonzero vector in $\mathbb{R}^n$, its *normalization* is given by

$$\frac{\vec{v}}{\|\vec{v}\|} = \frac{\vec{v}}{\sqrt{v_1^2 + \cdots + v_n^2}}, \tag{15.15}$$

where

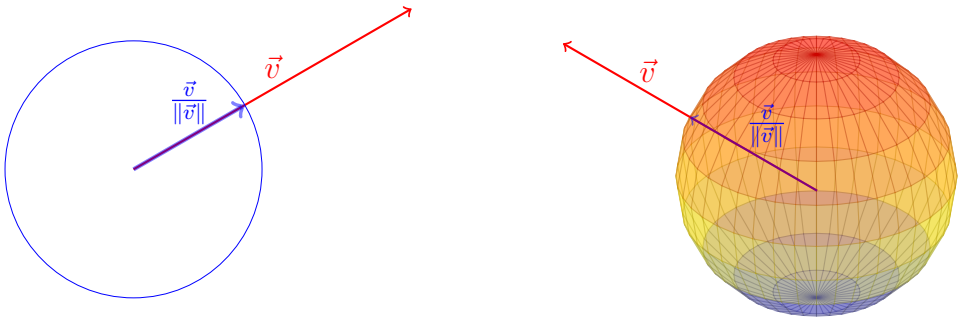
$$\vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \tag{15.16}$$

are the components of the vector $\vec{v}$. Occasionally, the notation $\hat{v}$ will be used for the normalized vector associated to $\vec{v}$.

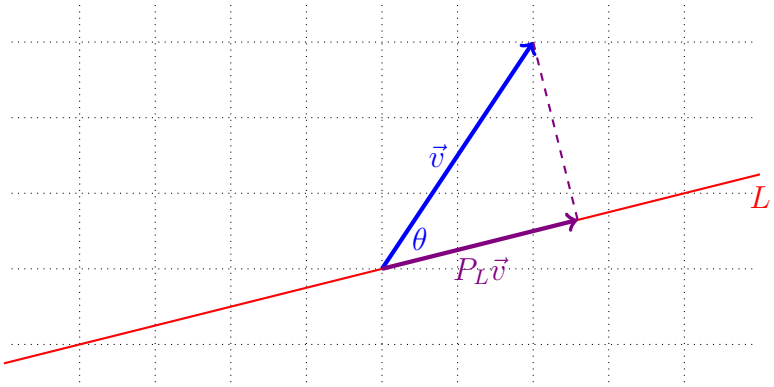
Normalized vectors always lie on the unit sphere in whatever Euclidean space the vector is in. Below are two drawings of the unit sphere along the normalized versions of vectors $\vec{v}$ in two dimensions and three dimensions[49]

[48] This is something that often happens in mathematics. One has different equivalent definitions for something in a few special cases. Some of these equivalent definitions can be generalized to include more cases. However, sometimes these generalizations can actually be different so that different choices will give different results. Which definition you choose depends on the context. I don't think we will have to worry about this in a course on linear algebra, but you may see this happen if you move on to more advanced mathematics courses.

[49] The sphere drawing was done using a modified version of code written by Christian Feuersänger and was found at http://tex.stackexchange.com/questions/124916/draw-sphere-pgfplots-with-axis-at-center.



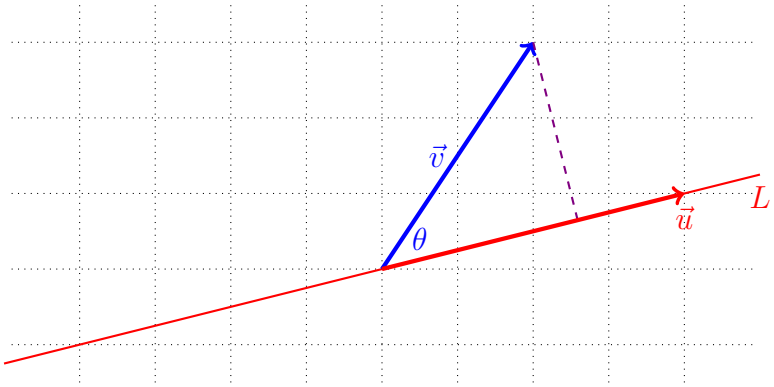
Given a vector $\vec{v}$ and $L \subseteq \mathbb{R}^2$ a one-dimensional subspace of $\mathbb{R}^2$, one can obtain a formula for the orthogonal projection $P_L\vec{v}$ of the vector $\vec{v}$ onto the line L as in the following figure



If θ is the angle from L to $\vec{v}$, then we can use trigonometry to figure out the length of this orthogonal projection. The result is

$$\|P_L\vec{v}\| = \|\vec{v}\| \cos\theta. \tag{15.17}$$

This is a scalar, and is called the scalar projection of $\vec{v}$ onto L. Let $\vec{u}$ be a vector pointing in this direction, such as depicted in the figure.



Therefore, the projection $P_L\vec{v}$ is a vector of length $\|\vec{v}\| \cos\theta$ pointing in the direction $\frac{\vec{u}}{\|\vec{u}\|}$. This specifies the projection as a vector and gives

$$\begin{aligned} P_L\vec{v} &= \big(\|\vec{v}\| \cos\theta\big) \frac{\vec{u}}{\|\vec{u}\|} \\ &= \big(\|\vec{v}\|\|\vec{u}\| \cos\theta\big) \frac{\vec{u}}{\|\vec{u}\|^2} \\ &= \langle \vec{v}, \vec{u} \rangle \frac{\vec{u}}{\|\vec{u}\|^2}, \end{aligned} \tag{15.18}$$

which gives the formula for the vector projection of $\vec{v}$ onto L. Although we have just derived this formula in two dimensions, we can also derive it in three dimensions, but then we can't attempt to do this in higher dimensions. Nevertheless, we use our definition of angle in arbitrary dimensions to motivate the definition of the scalar and vector projections of a vector onto a line L in arbitrary dimensions.

Definition 15.19. Let $\vec{v}$ and $\vec{u}$ be two vectors in $\mathbb{R}^n$ with $\vec{u}$ a nonzero vector and let $L := \text{span}\{\vec{u}\}$. The *vector projection* of $\vec{v}$ onto $\vec{u}$ (or onto L) is the vector $P_{\hat{u}}\vec{v}$ (sometimes written $P_L\vec{v}$) given by

$$P_{\hat{u}}\vec{v} := \frac{\langle \vec{v}, \vec{u} \rangle \vec{u}}{\|\vec{u}\|^2}. \tag{15.20}$$

In terms of the normalized vector $\hat{u}$, this is expressed as

$$P_{\hat{u}}\vec{v} := \langle \vec{v}, \hat{u} \rangle \hat{u}. \tag{15.21}$$

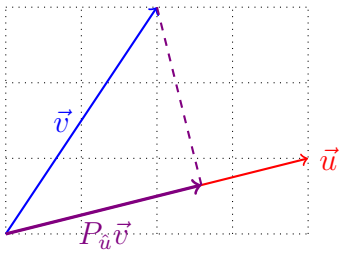
The *scalar projection* of $\vec{v}$ onto $\vec{u}$ (or L) is the Euclidean inner product of $\vec{v}$ with $\hat{u}$

$$S_{\hat{u}}\vec{v} := \langle \vec{v}, \hat{u} \rangle. \tag{15.22}$$

Notice that the scalar projection is a number while the vector projection is a vector. In fact,

$$P_{\hat{u}}\vec{v} = (S_{\hat{u}}\vec{v})\hat{u}. \tag{15.23}$$

The vector projection of $\vec{v}$ onto $\vec{u}$ is interpreted geometrically by the following picture



In this drawing,

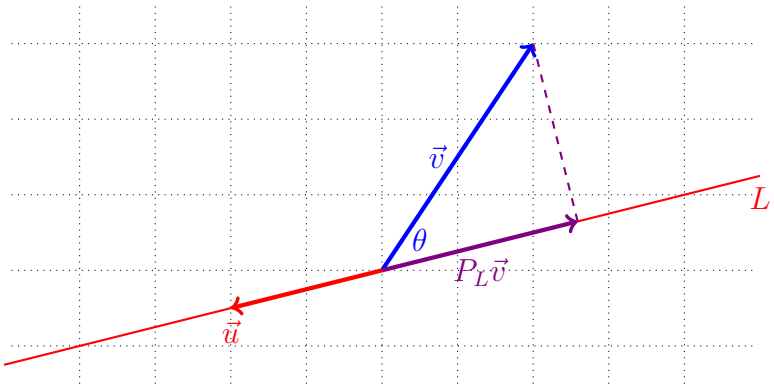
$$\vec{u} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \qquad \& \qquad \vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \tag{15.24}$$

so that

$$P_{\hat{u}}\vec{v} = \frac{\left\langle \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix} \right\rangle \begin{bmatrix} 4 \\ 1 \end{bmatrix}}{\left\| \begin{bmatrix} 4 \\ 1 \end{bmatrix} \right\|^2} = \frac{11}{17} \begin{bmatrix} 4 \\ 1 \end{bmatrix}. \tag{15.25}$$

The scalar projection is just the signed/oriented length of the projection. It is positive if $P_{\hat{u}}\vec{v}$ points in the same direction as $\vec{u}$ and it is negative if it points in the opposite direction.

Remark 15.26. It might seem annoying that there are two notations for the same concept. We use $P_L\vec{v}$ and $P_{\hat{u}}\vec{v}$ to both denote the projection of a vector onto a line. The only reason we also use the notation $P_{\hat{u}}\vec{v}$ because many books use it. However, you don't technically need all the data of a vector $\vec{u}$. A line is enough. There are many possible choices of $\vec{u}$ such that $L = \text{span}\{\vec{u}\}$ and the definition of the projection of a vector $\vec{v}$ onto this line does not depend on this choice. For example, imagine we picked $\vec{u}$ going in the other direction with a completely different magnitude such as



Here, the projection is actually pointing in the $-\frac{\vec{u}}{\|\vec{u}\|}$ direction. Hence, the projection is

$$P_L\vec{v} = -\big(\|\vec{v}\|\|\vec{u}\|\cos\theta\big)\frac{\vec{u}}{\|\vec{u}\|^2} = \big(\|\vec{v}\|\|\vec{u}\|\cos\left(\pi-\theta\right)\big)\frac{\vec{u}}{\|\vec{u}\|^2} = \langle\vec{v},\vec{u}\rangle\,\frac{\vec{u}}{\|\vec{u}\|^2} \tag{15.27}$$

because now $\pi - \theta$ is the angle between $\vec{u}$ and $\vec{v}$ so the formula in terms of the inner product is the same.

Theorem 15.28. *Let $\vec{u}$ be a nonzero vector in $\mathbb{R}^n$. Then the functions*

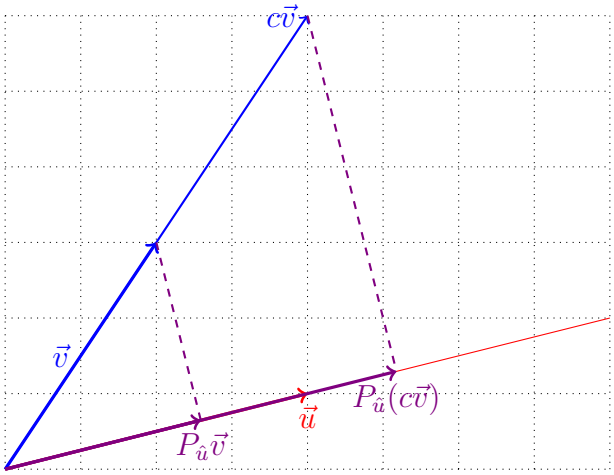
$$\begin{aligned} \mathbb{R}^n &\overset{S_{\hat{u}}}{\longmapsto} \mathbb{R} \\ \vec{v} &\mapsto S_{\hat{u}}\vec{v} \end{aligned} \tag{15.29}$$

and

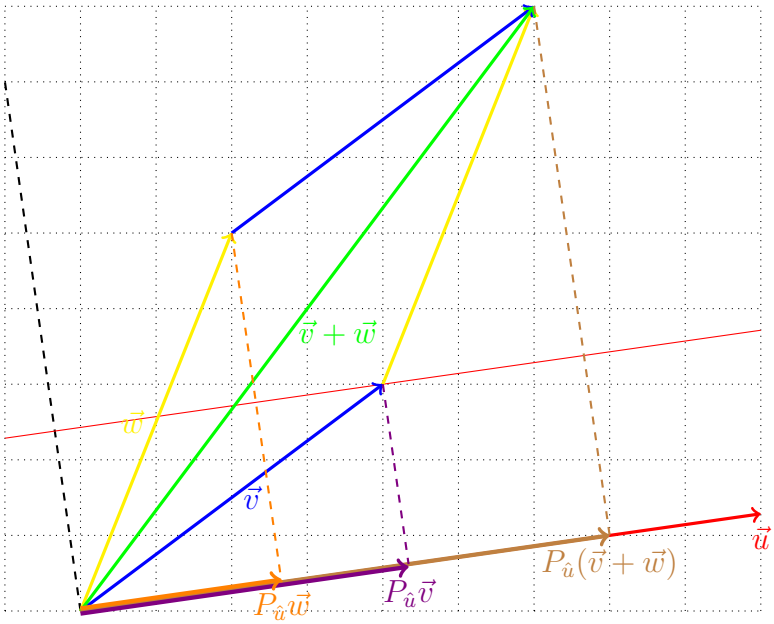
$$\begin{aligned} \mathbb{R}^n &\overset{P_{\hat{u}}}{\longmapsto} \mathbb{R}^n \\ \vec{v} &\mapsto P_{\hat{u}}\vec{v} \end{aligned} \tag{15.30}$$

are linear transformations. Furthermore, $P_{\hat{u}}$ is a linear transformation such that $P_{\hat{u}}^2 = P_{\hat{u}}$.

Proof. This can be proven algebraically, but there is a nice geometric proof. Similarity of triangles



proves that $P_{\hat{u}}(c\vec{v}) = cP_{\hat{u}}\vec{v}$ while geometry involving parallel lines



proves that $P_{\hat{u}}(\vec{v}+\vec{w}) = P_{\hat{u}}\vec{v}+P_{\hat{u}}\vec{w}$. The fact that $P_{\hat{u}}^2 = P_{\hat{u}}$ follows immediately from the geometric picture of the definition: once you have projected onto $\vec{u}$, you can't project anywhere else. The algebraic proof of this is less enlightening:

$$P_{\hat{u}}^2(\vec{v}) = P_{\hat{u}}\left(P_{\hat{u}}\vec{v}\right) = P_{\hat{u}}\left(\langle\vec{v},\hat{u}\rangle\hat{u}\right) = \langle\vec{v},\hat{u}\rangle P_{\hat{u}}\hat{u} = \langle\vec{v},\hat{u}\rangle \underbrace{\langle\hat{u},\hat{u}\rangle}_{1}\hat{u} = \langle\vec{v},\hat{u}\rangle\hat{u} = P_{\hat{u}}\vec{v} \tag{15.31}$$

for all vectors $\vec{v} \in \mathbb{R}^n$. Hence, $P_{\hat{u}}^2 = P_{\hat{u}}$. ■

In other words, the projection onto a vector is a linear transformation. In fact, the signed length of the projection is also a linear transformation. There are many other kinds of projections besides projections onto single vectors. One could imagine projecting onto a plane (such as when you shine a flashlight onto a figure and it makes a shadow on the wall as in Exercise 16.55). In higher dimensions, there are several ways to project onto different dimensional planes.

Definition 15.32. Let $\vec{v}$ be a nonzero vector in $\mathbb{R}^n$ and let $L := \text{span}\{\vec{v}\}$. The *orthogonal complement* of L is

$$L^\perp := \left\{\vec{u} \in \mathbb{R}^n \ : \ \langle \vec{v}, \vec{u}\rangle = 0\right\} \tag{15.33}$$

i.e. the set of vectors $\vec{u}$ in $\mathbb{R}^n$ that are orthogonal to $\vec{v}$.

Example 15.34. Consider the vector

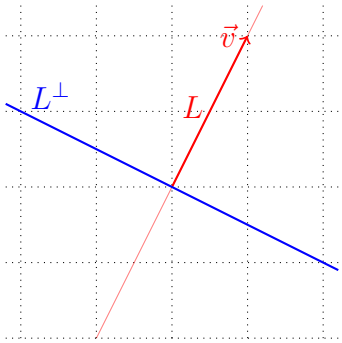
$$\vec{v} := \begin{bmatrix} 1 \\ 2 \end{bmatrix} \tag{15.35}$$

in $\mathbb{R}^2$ and let $L := \text{span}\{\vec{v}\}$. Then

$$L^\perp = \left\{aR_{\frac{\pi}{2}}(\vec{v}) \ : \ a \in \mathbb{R}\right\}, \tag{15.36}$$

where $R_{\frac{\pi}{2}}$ is the 2×2 matrix describing rotation by $\frac{\pi}{2}$, is the set of all scalar multiples of the vector $\vec{v}$ after it has been rotated by 90°. More explicitly,

$$L^\perp = \text{span}\left\{\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right\} \tag{15.37}$$



Example 15.38. Consider the vector

$$\vec{v} := \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \tag{15.39}$$

in $\mathbb{R}^3$ and let $L := \text{span}\{\vec{v}\}$. Then

$$\begin{aligned} L^\perp &= \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \ : \ \left\langle \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right\rangle = 0 \right\} \\ &= \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \ : \ 3x + 2y + z = 0 \right\} \end{aligned} \tag{15.40}$$

which describes the plane $z = -3x - 2y$. In other words,

$$\begin{aligned} L^\perp &= \left\{ x \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \in \mathbb{R}^3 \ : \ x, y \in \mathbb{R} \right\} \\ &= \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \right\}. \end{aligned} \tag{15.41}$$

Notice that the orthogonal complement of a line in $\mathbb{R}^3$ is a 2-dimensional plane. More generally, we can define the orthogonal complement of any subspace.

Definition 15.42. Let W be a k-dimensional subspace of $\mathbb{R}^n$. The *orthogonal complement* of W is

$$W^\perp := \left\{ \vec{v} \in \mathbb{R}^n \ : \ \langle \vec{v}, \vec{w} \rangle = 0 \text{ for all } \vec{w} \in W \right\} \tag{15.43}$$

i.e. the set of vectors $\vec{v}$ in $\mathbb{R}^n$ that are orthogonal to *all* vectors $\vec{w}$ in W.

Example 15.44. The orthogonal complement of the plane P described by the equation $z = -3x - 2y$ is precisely the one-dimensional subspace L from Example 15.34 because

$$\begin{aligned}
P^\perp &= \left\{ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \ : \ \left\langle \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, x \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \right\rangle = 0 \text{ for all } x, y \in \mathbb{R} \right\} \\
&= \left\{ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \ : \ v_1 x + v_2 y - (3x + 2y) v_3 = 0 \text{ for all } x, y \in \mathbb{R} \right\} \\
&= \left\{ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \ : \ x(v_1 - 3v_3) + y(v_2 - 2v_3) = 0 \text{ for all } x, y \in \mathbb{R} \right\} \\
&= \left\{ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \ : \ v_1 = 3v_3 \text{ and } v_2 = 2v_3 \right\} \\
&= \text{span} \left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\}.
\end{aligned} \tag{15.45}$$

If we have a basis for a subspace, we can use that basis to find the orthogonal complement.

Theorem 15.46. *Let $W \subseteq \mathbb{R}^n$ be a k-dimensional subspace and let $\{\vec{w}_1, \ldots, \vec{w}_k\}$ be a basis of W. Then $W^\perp$ is the solution set to the homogeneous system (of k equations in n unknowns)*

$$\begin{bmatrix} | & & | \\ \vec{w}_1 & \cdots & \vec{w}_k \\ | & & | \end{bmatrix}^T \begin{bmatrix} | \\ \vec{v} \\ | \end{bmatrix} = \vec{0}, \tag{15.47}$$

i.e. the set of $\vec{v} \in \mathbb{R}^n$ satisfying (15.47).

Proof. We first prove that

$$W^\perp = \left\{ \vec{v} \in \mathbb{R}^n \ : \ \langle \vec{v}, \vec{w}_1 \rangle = 0, \ldots, \langle \vec{v}, \vec{w}_k \rangle = 0 \right\}. \tag{15.48}$$

Let $\vec{v} \in W^\perp$. Then, by definition, $\langle \vec{v}, \vec{w} \rangle = 0$ for all $\vec{w} \in W$. In particular, $\langle \vec{v}, \vec{w}_i \rangle = 0$ for all $i \in \{1, \ldots, k\}$ since $\vec{w}_i \in W$ for all $i \in \{1, \ldots, k\}$. Hence

$$W^\perp \subseteq \left\{ \vec{v} \in \mathbb{R}^n \ : \ \langle \vec{v}, \vec{w}_1 \rangle = 0, \ldots, \langle \vec{v}, \vec{w}_k \rangle = 0 \right\}. \tag{15.49}$$

Now suppose $\vec{v}$ satisfies $\langle \vec{v}, \vec{w}_i \rangle = 0$ for all $i \in \{1, \ldots, k\}$. Let $\vec{w} \in W$. Since $\{\vec{w}_1, \ldots, \vec{w}_k\}$ is a basis of W, there exist coefficients $x_1, \ldots, x_k$ such that $\vec{w} = x_1\vec{w}_1 + \cdots + x_k\vec{w}_k$. Hence,

$$\langle \vec{v}, \vec{w} \rangle = \langle \vec{v}, x_1\vec{w}_1 + \cdots + x_k\vec{w}_k \rangle = x_1\langle \vec{v}, \vec{w}_1 \rangle + \cdots + x_k\langle \vec{v}, \vec{w}_k \rangle = 0 \tag{15.50}$$

proving that $\vec{v} \in W^\perp$. Hence,

$$W^\perp \supseteq \left\{ \vec{v} \in \mathbb{R}^n \ : \ \langle \vec{v}, \vec{w}_1 \rangle = 0, \ldots, \langle \vec{v}, \vec{w}_k \rangle = 0 \right\}. \tag{15.51}$$

This proves our first claim. Now, notice that the conditions $\langle \vec{v}, \vec{w}_i \rangle = 0$ for all $i \in \{1, \ldots, k\}$ is equivalent to

$$\begin{bmatrix} | & & | \\ \vec{w}_1 & \cdots & \vec{w}_k \\ | & & | \end{bmatrix}^T \begin{bmatrix} | \\ \vec{v} \\ | \end{bmatrix} = \vec{0} \tag{15.52}$$

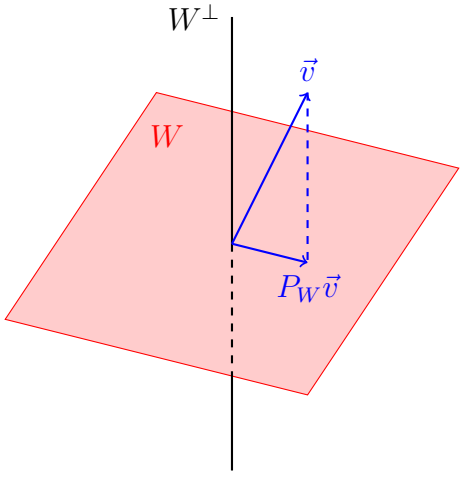
since

$$\begin{bmatrix} | & & | \\ \vec{w}_1 & \cdots & \vec{w}_k \\ | & & | \end{bmatrix}^T \begin{bmatrix} | \\ \vec{v} \\ | \end{bmatrix} = \begin{bmatrix} \langle \vec{w}_1, \vec{v} \rangle \\ \vdots \\ \langle \vec{w}_k, \vec{v} \rangle \end{bmatrix}. \tag{15.53}$$

■

The phenomenon of taking the orthogonal complement twice to get back what you started (this happened in Example 15.44) is true in general, but we will need an important result to prove it. A similar question we might ask, which is very intuitive, is the following. Given a subspace $W \subseteq \mathbb{R}^n$, is there a linear transformation that acts as a projection onto W? If so, how can one express this linear transformation as a matrix? Visually, the projection of $\vec{v}$ onto W should be a vector $P_W\vec{v}$ that satisfies the condition of being the closest vector to $\vec{v}$ inside W, i.e.

$$\|\vec{v} - P_W\vec{v}\| = \min_{\vec{w} \in W} \|\vec{v} - \vec{w}\|. \tag{15.54}$$



In the process of answering this question, we will prove that every subspace has an *orthonormal basis*.

Exercise 15.55. If $\mathbb{R}^m \xleftarrow{S} \mathbb{R}^n$ is a linear transformation, prove that

$$\langle \vec{w}, S\vec{v} \rangle = \langle S^T \vec{w}, \vec{v} \rangle \tag{15.56}$$

for all vectors $\vec{w} \in \mathbb{R}^m$ and $\vec{v} \in \mathbb{R}^n$. Furthermore, when $m = n$, show that $S = S^T$ if and only if $\langle \vec{w}, S\vec{v} \rangle = \langle S\vec{w}, \vec{v} \rangle$ for all vectors $\vec{w} \in \mathbb{R}^m$ and $\vec{v} \in \mathbb{R}^n$. This gives some geometric meaning to the transpose. [Hint: write out what the inner product is in terms of matrices and the transpose.]

Exercise 15.57. An *orthogonal $m \times m$ matrix* is an $m \times m$ matrix A such that $\langle A\vec{u}, A\vec{v} \rangle = \langle \vec{u}, \vec{v} \rangle$ for all vectors $\vec{u}, \vec{v} \in \mathbb{R}^m$. Show that $A^{-1} = A^T$ for such a matrix A. [Hint: first show that if for a fixed $\vec{w}$, the inner product $\langle \vec{u}, \vec{w} \rangle = 0$ for all vectors $\vec{u} \in \mathbb{R}^m$, then $\vec{w} = \vec{0}$.]

Recommended Exercises. Please check HuskyCT for the homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we covered Sections 6.1 and 6.2.

Terminology checklist

Euclidean norm/length	
Euclidean inner/dot product	
normalized/unit vector	
normalization	
vector projection	
scalar projection	
orthogonal/perpendicular	
orthogonal complement	

16 The Gram-Schmidt procedure

From the examples previously, we noticed some interesting properties of sets of orthogonal vectors when studying the orthogonal complement.

Theorem 16.1. *Let $\mathcal{S} := \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be an orthogonal set of nonzero vectors in $\mathbb{R}^n$. Then $\mathcal{S}$ is linearly independent. In particular, $\mathcal{S}$ is a basis for* $\text{span}(\mathcal{S})$.

Proof. Let $x_1, \ldots, x_k$ be coefficients such that

$$x_1\vec{v}_1 + \cdots + x_k\vec{v}_k = \vec{0}. \tag{16.2}$$

The goal is to show that $x_1 = \cdots = x_k = 0$. To see this, take the inner product of both sides of the above equation with the vector $\vec{v}_i$ for some $i \in \{1, \ldots, k\}$. This gives

$$x_1\langle\vec{v}_i, \vec{v}_1\rangle + \cdots + x_k\langle\vec{v}_i, \vec{v}_k\rangle = \langle\vec{v}_i, \vec{0}\rangle = 0. \tag{16.3}$$

Because $\mathcal{S}$ is orthogonal, the only nonzero term on the left is $x_i\langle\vec{v}_i, \vec{v}_i\rangle = x_i\|\vec{v}_i\|^2$. Since $\vec{v}_i \neq \vec{0}$, this means that $\|\vec{v}_i\|^2 > 0$. Therefore,

$$x_i\|\vec{v}_i\|^2 = 0 \qquad \Rightarrow \qquad x_i = 0. \tag{16.4}$$

Since $i \in \{1, \ldots, k\}$ was arbitrary, $x_1 = \cdots = x_k = 0$, which shows that $\mathcal{S}$ is linearly independent. ■

This theorem is useful because it can sometimes be used to quickly identify that a given set of vectors is linearly independent (however, it cannot be used to say that a set of vectors is linearly dependent!).

Problem 16.5. is the set of vectors

$$\left\{ \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 17 \\ -26 \end{bmatrix}, \begin{bmatrix} 8 \\ 2 \\ 1 \end{bmatrix} \right\} \tag{16.6}$$

linearly independent?

Answer. Denote these vectors by $\vec{v}_1, \vec{v}_2, \vec{v}_3$ in the order written above. Then $\langle\vec{v}_1, \vec{v}_2\rangle = 0$, $\langle\vec{v}_2, \vec{v}_3\rangle = 0$, and $\langle\vec{v}_3, \vec{v}_1\rangle = 0$. Hence, the set of vectors is orthogonal. Since none of the vectors is the zero vector, the set is linearly independent.

Given an orthogonal set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ of nonzero vectors in $\mathbb{R}^n$, the set

$$\left\{ \frac{\vec{v}_1}{\|\vec{v}_1\|}, \frac{\vec{v}_2}{\|\vec{v}_2\|}, \ldots, \frac{\vec{v}_k}{\|\vec{v}_k\|} \right\} \tag{16.7}$$

is an orthonormal set. An orthonormal set of vectors in $\mathbb{R}^n$ has many similar properties to the vectors $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_k\}$ in $\mathbb{R}^n$, where $k \leq n$. One such property is described in the following theorem.

Theorem 16.8. *Let $\mathcal{S} := \{\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_k\}$ be an orthonormal set in $\mathbb{R}^n$, and let $\vec{v}$ be a vector in* $\operatorname{span}(\mathcal{S})$. *Then*

$$\vec{v} = \langle \vec{v}, \hat{u}_1 \rangle \hat{u}_1 + \langle \vec{v}, \hat{u}_2 \rangle \hat{u}_2 + \cdots + \langle \vec{v}, \hat{u}_k \rangle \hat{u}_k. \tag{16.9}$$

Furthermore, since $\mathcal{S}$ is a basis for $\operatorname{span}(\mathcal{S})$, *these linear combinations are unique.*

The infinite-dimensional generalization of this theorem is what makes Fourier series work. We might discuss this later in the course.

Proof. Since $\mathcal{S}$ is a basis (because $\mathcal{S}$ are linearly independent and they span $\operatorname{span}(\mathcal{S})$ by definition), there exist unique coefficients $x_1, \ldots, x_k \in \mathbb{R}$ such that

$$\vec{v} = x_1 \hat{u}_1 + \cdots + x_k \hat{u}_k. \tag{16.10}$$

Fix some $i \in \{1, \ldots, k\}$. Applying the inner product with $\vec{u}_i$ to both sides gives

$$\langle \vec{u}_i, \vec{v} \rangle = x_i \underbrace{\langle \hat{u}_i, \hat{u}_i \rangle}_{1}, \tag{16.11}$$

which shows that $x_i = \langle \hat{u}_i, \vec{v} \rangle = \langle \vec{v}, \hat{u}_i \rangle$. Since $i \in \{1, \ldots, k\}$ was arbitrary, the above decomposition holds. ∎

Notice that since $\mathcal{S}$ is a basis for W, we know that there are *some* coefficients $x_1, \ldots, x_k$ such that

$$\vec{v} = x_1 \hat{u}_1 + \cdots + x_k \hat{u}_k. \tag{16.12}$$

Normally, to find these coefficients, we would have to row reduce the augmented matrix

$$\left[\begin{array}{ccc|c} | & & | & | \\ \hat{u}_1 & \cdots & \hat{u}_k & \vec{v} \\ | & & | & | \end{array}\right]. \tag{16.13}$$

This theorem tells us that we don't have to do this! All we have to do is compute the inner products $x_i = \langle \vec{v}_i, \hat{u}_i \rangle$ so we save ourselves the need to do any row reduction.

Example 16.14. As a simple case, let $\mathcal{S} = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_k\}$ be the first k standard unit vectors in $\mathbb{R}^n$ $(n \geq k)$. This theorem says that

$$\vec{v} = \langle \vec{v}, \vec{e}_1 \rangle \vec{e}_1 + \langle \vec{v}, \vec{e}_2 \rangle \vec{e}_2 + \cdots + \langle \vec{v}, \vec{e}_k \rangle \vec{e}_k \tag{16.15}$$

for every $\vec{v} \in \operatorname{span}(\mathcal{S})$. This is not a surprising result. Expressing both sides of these equations, this result just looks like

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_k \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{16.16}$$

You can see from the proof, specifically (16.11), how crucial it was that the vectors in $\mathcal{S}$ were not just a basis but an *orthonormal* basis. What happens if $\vec{v}$ is not in the subspace spanned by an orthonormal set? Given an orthonormal set of vectors, what is the linear transformation describing the projection onto the subspace spanned by these vectors? The following theorem answers these questions.

Theorem 16.17. *Let $\mathcal{S} := \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ be an orthonormal set of vectors in $\mathbb{R}^n$, let $W := \text{span}(\mathcal{S})$, and let $\vec{v}$ be a vector in $\mathbb{R}^n$. Then there exists a unique decomposition*

$$\vec{v} = \vec{w} + \vec{w}^{\perp} \tag{16.18}$$

with $\vec{w} \in W$ and $\vec{w}^{\perp} \in W^{\perp}$. In fact,

$$\vec{w} = \langle \vec{v}, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v}, \vec{u}_2 \rangle \vec{u}_2 + \cdots + \langle \vec{v}, \vec{u}_k \rangle \vec{u}_k \tag{16.19}$$

and

$$\vec{w}^{\perp} := \vec{v} - \big(\langle \vec{v}, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v}, \vec{u}_2 \rangle \vec{u}_2 + \cdots + \langle \vec{v}, \vec{u}_k \rangle \vec{u}_k \big). \tag{16.20}$$

Furthermore, the transformation $\mathbb{R}^n \xleftarrow{P_W} \mathbb{R}^n$ defined by sending $\vec{v}$ in $\mathbb{R}^n$ to

$$P_W(\vec{v}) := \langle \vec{v}, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v}, \vec{u}_2 \rangle \vec{u}_2 + \cdots + \langle \vec{v}, \vec{u}_k \rangle \vec{u}_k \tag{16.21}$$

is a linear transformation satisfying

$$P_W^2 = P_W \tag{16.22}$$

and

$$P_W^T = P_W. \tag{16.23}$$

Proof. The first thing we check is that $\vec{w}^{\perp}$ given by (16.20) is in $W^{\perp}$. For this, notice that

$$\begin{aligned} \langle \vec{w}^{\perp}, \vec{u}_i \rangle &= \Big\langle \vec{v} - \big(\langle \vec{v}, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v}, \vec{u}_2 \rangle \vec{u}_2 + \cdots + \langle \vec{v}, \vec{u}_k \rangle \vec{u}_k \big), \vec{u}_i \Big\rangle \\ &= \langle \vec{v}, \vec{u}_i \rangle - \sum_{j=1}^{k} \langle \vec{v}, \vec{u}_j \rangle \langle \vec{u}_j, \vec{u}_i \rangle \\ &= \langle \vec{v}, \vec{u}_i \rangle - \sum_{j=1}^{k} \langle \vec{v}, \vec{u}_j \rangle \delta_{ji} \\ &= \langle \vec{v}, \vec{u}_i \rangle - \langle \vec{v}, \vec{u}_i \rangle \\ &= 0. \end{aligned} \tag{16.24}$$

Since this is true for every $i = 1, 2, \ldots, k$, it follows that $\langle \vec{w}^{\perp}, \vec{u} \rangle = 0$ for every vector $\vec{u}$ in W, since every such vector is expressed as a linear combination of these elements in $\mathcal{S}$. Hence, $\vec{w}^{\perp}$ is in $W^{\perp}$.

The second thing we check is that $P_W^2 = P_W$ (you should check that P_W is in fact a linear transformation). For this, let $\vec{v}$ be a vector in V and calculating $P_W^2(\vec{v})$ gives

$$\begin{aligned} P_W^2(\vec{v}) &= P_W\left(\sum_{i=1}^{k}\langle\vec{v},\vec{u}_i\rangle\vec{u}_i\right) \\ &= \sum_{i=1}^{k}\left(\langle\vec{v},\vec{u}_i\rangle P_W(\vec{u}_i)\right) \\ &= \sum_{i=1}^{k}\left(\langle\vec{v},\vec{u}_i\rangle\sum_{j=1}^{k}\langle\vec{u}_i,\vec{u}_j\rangle\vec{u}_j\right) \\ &= \sum_{i=1}^{k}\left(\langle\vec{v},\vec{u}_i\rangle\sum_{j=1}^{k}\delta_{ij}\vec{u}_j\right) \\ &= \sum_{i=1}^{k}\left(\langle\vec{v},\vec{u}_i\rangle\vec{u}_i\right) \\ &= P_W(\vec{v}) \end{aligned} \tag{16.25}$$

as needed.

The final thing to check is $P_W^T = P_W$. By Exercise 15.55, it suffices to show that $\langle P_W\vec{v},\vec{u}\rangle = \langle\vec{v},P_W\vec{u}\rangle$ for all $\vec{u},\vec{v}\in\mathbb{R}^n$. This follows from

$$\begin{aligned} \langle P_W\vec{v},\vec{u}\rangle &= \left\langle\sum_{i=1}^{k}\langle\vec{v},\vec{u}_i\rangle\vec{u}_i,\vec{u}\right\rangle \\ &= \sum_{i=1}^{k}\langle\vec{v},\vec{u}_i\rangle\langle\vec{u}_i,\vec{u}\rangle \\ &= \sum_{i=1}^{k}\langle\vec{u},\vec{u}_i\rangle\langle\vec{v},\vec{u}_i\rangle \\ &= \left\langle\vec{v},\sum_{i=1}^{k}\langle\vec{u},\vec{u}_i\rangle\vec{u}_i\right\rangle \\ &= \langle\vec{v},P_W\vec{u}\rangle. \end{aligned} \tag{16.26}$$

■

The last conditions in Theorem 16.17 are what defines an orthogonal projection.

Definition 16.27. A linear transformation $\mathbb{R}^n \xleftarrow{S} \mathbb{R}^n$ is called a *projection* whenever $S^2 = S$. S is called an *orthogonal projection* whenever $S^2 = S$ and $S^T = S$.

A nice way to think about projections is given at http://math.stackexchange.com/questions/1303977/what-is-the-idea-behind-a-projection-operator-what-does-it-do. Are there projections that are not orthogonal projections? We will answer this shortly, though you should be able to find counter-examples at this point. Given an arbitrary subspace, we should be able to

define the orthogonal projection onto that subspace. However, we have only been able to construct such a projection when we have an orthonormal basis for such a subspace. Given an arbitrary subspace W, how do we know if a orthonormal set spanning W even *exists*? We know a *basis* always exists, but a basis is quite different from an orthonormal set (not every basis is an orthonormal set but every orthonormal set is a basis for the space it spans). The Gram-Schmidt process, a process we are about to describe, provides a *construction* of an orthonormal set for any subspace of a finite-dimensional inner product space. Hence, not only does it prove the existence of an orthonormal set for a subspace, it provides a step-by-step procedure for producing one!

Construction 16.28 (The Gram-Schmidt Procedure)**.** Let W be a nonzero subspace of $\mathbb{R}^n$. Because W is finite-dimensional, there exists a basis $\mathcal{B} := \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$, with $k \geq 1$, for W. Set

$$\vec{u}_1 := \frac{\vec{w}_1}{\|\vec{w}_1\|} \tag{16.29}$$

and let $W_1 := \text{span}(\{\vec{w}_1\}) \equiv \text{span}(\{\vec{u}_1\})$. Set

$$\vec{v}_2 := \vec{w}_2 - \langle \vec{w}_2, \vec{u}_1 \rangle \vec{u}_1, \tag{16.30}$$

i.e. the component of $\vec{w}_2$ perpendicular to W_1. Normalize it by setting

$$\vec{u}_2 := \frac{\vec{v}_2}{\|\vec{v}_2\|} \tag{16.31}$$

and set $W_2 := \text{span}\big(\{\vec{u}_1, \vec{u}_2\}\big)$. Proceed inductively. Namely, suppose that $\vec{u}_m$ and W_m have been constructed and $1 < m < n$. Then, set

$$\vec{v}_{m+1} := \vec{w}_{m+1} - \sum_{i=1}^{m} \langle \vec{w}_{m+1}, \vec{u}_i \rangle \vec{u}_i, \tag{16.32}$$

i.e. the component of $\vec{w}_{m+1}$ perpendicular to W_m. Normalize it by setting

$$\vec{u}_{m+1} := \frac{\vec{v}_{m+1}}{\|\vec{v}_{m+1}\|} \tag{16.33}$$

and set $W_{m+1} := \text{span}\big(\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_{m+1}\}\big)$. After this construction is done at step k (which must happen since n is finite), the set $\mathcal{S} := \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ of vectors has been constructed along with a sequence of subspaces $W_1, W_2, \ldots, W_k$ satisfying the following properties.

(a) The set $\mathcal{S}$ is orthonormal.

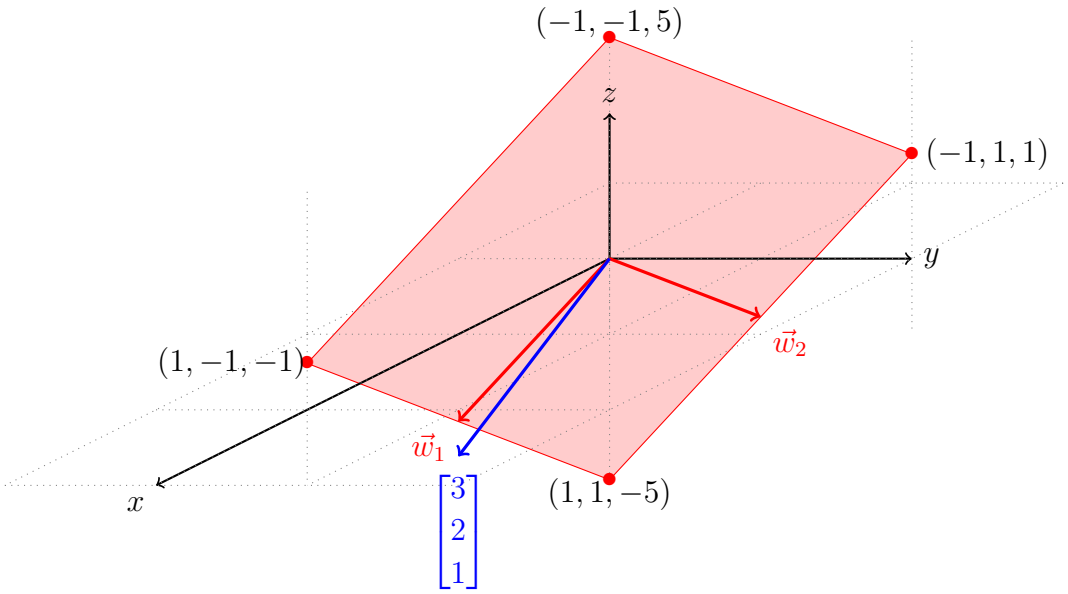
(b) $W_k = W$. In fact, $\text{span}(\mathcal{S}) = W$.

The first claim follows from the construction. The second claim follows from the fact that $\mathcal{S}$ is a basis of W_k containing k elements, but W_k is a subspace of W, which also has a basis containing k elements, which implies $W = W_k$.

Example 16.34. Going back to Example 15.38, a quick calculation shows that the basis

$$\mathcal{B} := \left\{ \vec{w}_1 := \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} \quad , \quad \vec{w}_2 := \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \right\} \tag{16.35}$$

for the plane described by the solutions to $3x + 2y + z = 0$ is not orthogonal.



It does not look like it, but the vector $(3, 2, 1)$ is orthogonal to the plane in the above figure. Applying the Gram-Schmidt procedure, the vector $\vec{u}_1$ is

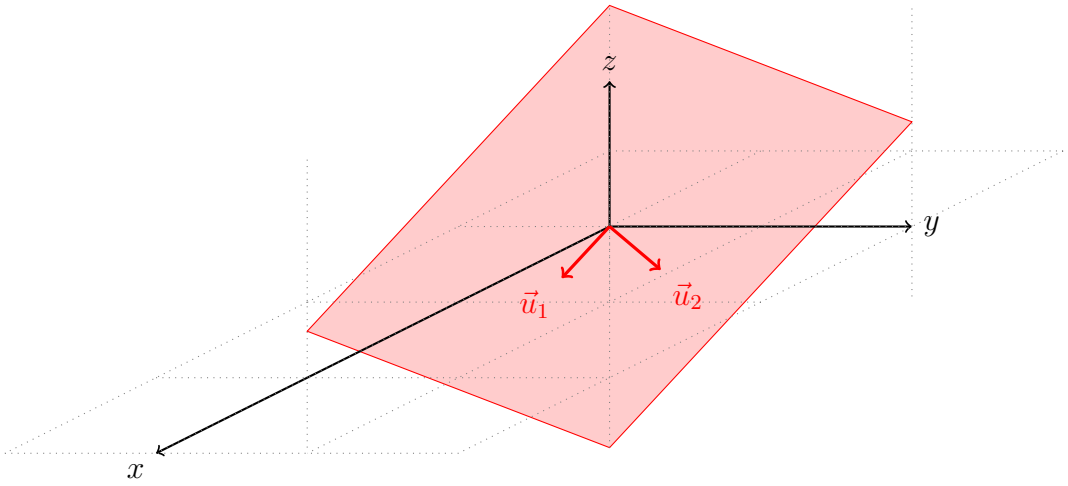
$$\vec{u}_1 := \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} \tag{16.36}$$

and $\vec{v}_2$ is

$$\begin{aligned} \vec{v}_2 &:= \vec{w}_2 - \langle \vec{w}_2, \vec{u}_1 \rangle \vec{u}_1 \\ &= \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} - \left\langle \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}, \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} \right\rangle \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} - \frac{3}{5} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} -3 \\ 5 \\ -1 \end{bmatrix}. \end{aligned} \tag{16.37}$$

Thus, the unit vector in this direction is given by

$$\vec{u}_2 := \frac{\vec{v}_2}{\|\vec{v}_2\|} = \frac{1}{\sqrt{35}} \begin{bmatrix} -3 \\ 5 \\ -1 \end{bmatrix}. \tag{16.38}$$

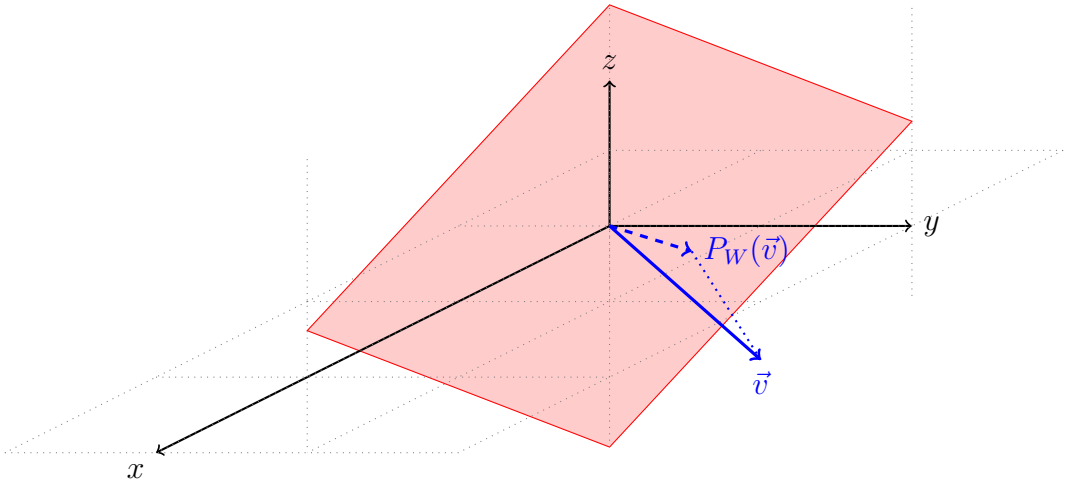


Now consider the vector

$$\vec{v} := \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}. \tag{16.39}$$

The projection of this vector onto the plane W spanned by $\mathcal{B}$ is given by

$$\begin{aligned} P_W(\vec{v}) &= \langle \vec{v}, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v}, \vec{u}_2 \rangle \vec{u}_2 \\ &= \left\langle \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} \right\rangle \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} + \left\langle \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \frac{1}{\sqrt{35}} \begin{bmatrix} -3 \\ 5 \\ -1 \end{bmatrix} \right\rangle \frac{1}{\sqrt{35}} \begin{bmatrix} -3 \\ 5 \\ -1 \end{bmatrix} \\ &= \frac{7}{10} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} + \frac{18}{35} \begin{bmatrix} -3 \\ 5 \\ -1 \end{bmatrix} \\ &= \frac{1}{35} \begin{bmatrix} -1 \\ 18 \\ 33 \end{bmatrix} \end{aligned} \tag{16.40}$$



We can calculate the matrix form of P_W in at least two ways. One way is to calculate $P_W\vec{e}_1, P_W\vec{e}_2$, and $P_W\vec{e}_3$. The other way is to use the relationship between inner products and the transpose. In

our case, since W is spanned by two orthonormal vectors $\vec{u}_1$ and $\vec{u}_2$, this is given by

$$\begin{aligned} P_W &= \vec{u}_1\vec{u}_1^T + \vec{u}_2\vec{u}_2^T \\ &= \frac{1}{10}\begin{bmatrix}1\\0\\-3\end{bmatrix}\begin{bmatrix}1 & 0 & -3\end{bmatrix} + \frac{1}{35}\begin{bmatrix}-3\\5\\-1\end{bmatrix}\begin{bmatrix}-3 & 5 & -1\end{bmatrix} \\ &= \frac{1}{10}\begin{bmatrix}1 & 0 & -3\\0 & 0 & 0\\-3 & 0 & 9\end{bmatrix} + \frac{1}{35}\begin{bmatrix}9 & -15 & 3\\-15 & 25 & -5\\3 & -5 & 1\end{bmatrix} \\ &= \frac{1}{14}\begin{bmatrix}5 & -6 & -3\\-6 & 10 & -2\\-3 & -2 & 13\end{bmatrix}. \end{aligned} \tag{16.41}$$

From this expression, you can see that P_W^T. It is a bit more cumbersome to calculate P_W^2, but it can be done and one sees that $P_W^2 = P_W$. However, we already know that we do not have to do this because we have constructed P_W in such a way so that it satisfies the conditions of an orthogonal projection.

Notice that you do not have to always normalize the vectors until the end so that you never need to work with squareroots. Let's do an example that illustrates this and also where you have to apply the Gram-Schmidt procedure twice.

Problem 16.42. Use the Gram-Schmidt procedure to find an orthonormal basis for the subspace

$$W := \operatorname{span}\left\{\vec{w}_1 := \begin{bmatrix}1\\0\\0\\1\end{bmatrix},\ \vec{w}_2 := \begin{bmatrix}1\\1\\1\\1\end{bmatrix},\ \vec{w}_3 := \begin{bmatrix}1\\1\\1\\0\end{bmatrix}\right\} \tag{16.43}$$

of $\mathbb{R}^4$.

Answer. First set $\vec{v}_1 := \vec{w}_1$ and note that $\|\vec{v}_1\|^2 = 2$. Set

$$\vec{v}_2 := \vec{w}_2 - \frac{\langle \vec{w}_2, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \frac{2}{2}\begin{bmatrix}1\\0\\0\\1\end{bmatrix} = \begin{bmatrix}0\\1\\1\\0\end{bmatrix} \tag{16.44}$$

so that $\|\vec{v}_2\|^2 = 2$. Then

$$\vec{w}_3 - \frac{\langle \vec{w}_3, \vec{v}_1\rangle}{\|\vec{v}_1\|^2}\vec{v}_1 - \frac{\langle \vec{w}_3, \vec{v}_2\rangle}{\|\vec{v}_2\|^2}\vec{v}_2 = \begin{bmatrix}1\\1\\1\\0\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\0\\0\\1\end{bmatrix} - \frac{2}{2}\begin{bmatrix}0\\1\\1\\0\end{bmatrix} = \begin{bmatrix}1/2\\0\\0\\-1/2\end{bmatrix} \tag{16.45}$$

so let's get rid of the fractions and set

$$\vec{v}_3 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}. \tag{16.46}$$

Hence, an orthogonal basis for W is

$$\left\{ \vec{v}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \ \vec{v}_2 := \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \ \vec{v}_3 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\} \tag{16.47}$$

and an orthonormal basis is obtained by normalizing these vectors

$$\left\{ \vec{u}_1 := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \ \vec{u}_2 := \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \ \vec{u}_3 := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\}. \tag{16.48}$$

Why do we need an orthonormal basis to construct an orthogonal projection onto a subspace? Let us look at what happens if we do not use an orthonormal basis but instead just use a basis of normalized vectors.

Example 16.49. Let

$$\vec{u}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \& \qquad \vec{u}_2 := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \tag{16.50}$$

and let W be the plane spanned by these two vectors (this is just the xy plane). We can define the linear transformation $\mathbb{R}^3 \xleftarrow{T_W} \mathbb{R}^3$ by

$$\mathbb{R}^3 \ni \vec{v} \mapsto T_W \vec{v} := \langle \vec{u}_1, \vec{v} \rangle \vec{u}_1 + \langle \vec{u}_2, \vec{v} \rangle \vec{u}_2 \tag{16.51}$$

exactly as we did for an orthonormal set of vectors. Again, we can calculate the matrix associated to T_W via finding its columns from $T_W \vec{e}_i$ or we can use the transpose method

$$T_W = \vec{u}_1 \vec{u}_1^T + \vec{u}_2 \vec{u}_2^T = \frac{1}{2} \begin{bmatrix} 3 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{16.52}$$

The range/image/column space of T_W is exactly W. However, even though $T_W^T = T_W$,

$$T_W^2 = \frac{1}{2} \begin{bmatrix} 5 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq T_W. \tag{16.53}$$

Therefore, it is not a projection. An intuitive way to see why T_W isn't a projection is to look at what happens to $\vec{e}_1$ and $\vec{e}_2$. Under a projection, these vectors must be fixed since they already lie in the plane W (this is the meaning of the equation $T_W^2 = T_W$ and is why it is such a crucial condition for the definition of a projection). Under these transformations

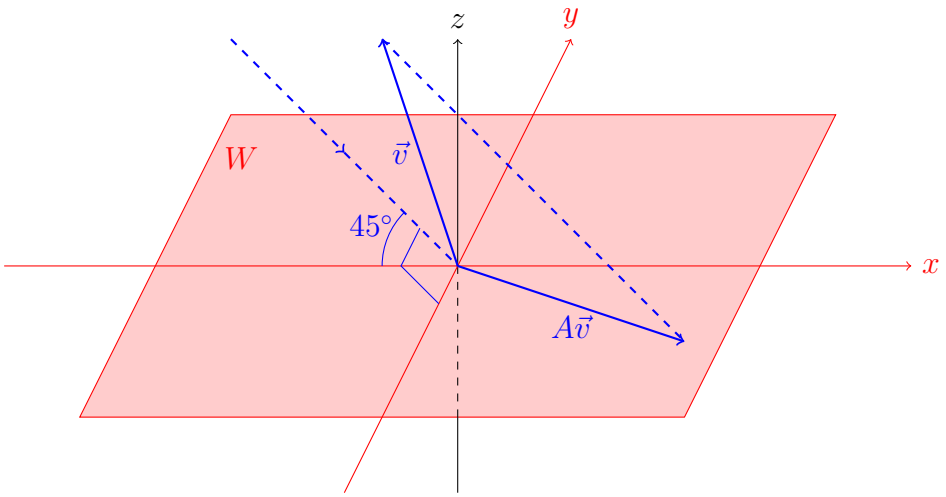
$$T_W\vec{e}_1 = \frac{3}{2}\vec{e}_1 + \frac{1}{2}\vec{e}_2 \qquad \& \qquad T_W\vec{e}_2 = \frac{1}{2}\vec{e}_1 + \frac{1}{2}\vec{e}_2 \tag{16.54}$$

neither of which are fixed.

An arbitrary projection, not necessarily orthogonal, onto a plane in $\mathbb{R}^3$ can be visualized as shining a flashflight onto that plane from some angle. If the angle makes a right angle with the plane, then the projection is an orthogonal projection. However, there are many other possible projections besides the orthogonal one. Consider, for example, shining the flashlight at a 45° angle with respect to the normal of a plane.

Exercise 16.55. Let $W = \text{span}\{\vec{e}_1, \vec{e}_2\}$ in $\mathbb{R}^3$. Express the matrix A associated to the linear transformation that is obtained from shining a flashlight at a 45° angle with respect to the $\vec{e}_3$ axis and pointing in the direction of $\vec{e}_1$, i.e. the flashlight points in the direction

$$\frac{1}{\sqrt{2}}(\vec{e}_1 - \vec{e}_3). \tag{16.56}$$



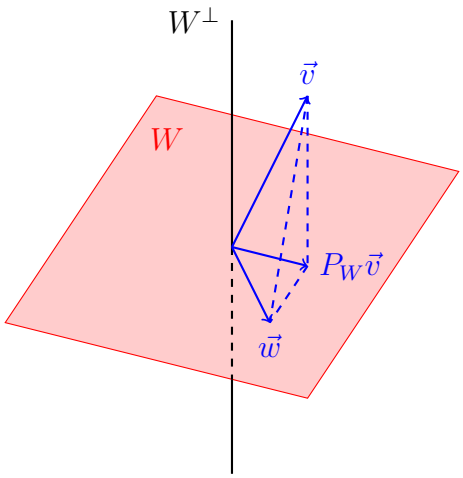
Theorem 16.57. *Let W be a subspace of $\mathbb{R}^n$. Then for every vector $\vec{v}$ in $\mathbb{R}^n$, there exist unique vectors $\vec{w}$ in W and $\vec{u}$ in $W^\perp$ satisfying $\vec{v} = \vec{w} + \vec{u}$. In fact, setting P_W to be the orthogonal projection onto W, this decomposition is written as*

$$\vec{v} = P_W\vec{v} + (\vec{v} - P_W\vec{v})\,. \tag{16.58}$$

Furthermore, $P_W\vec{v}$ minimizes the distance between the vector $\vec{v}$ and all vectors in W.

The difference between the first part of this Theorem and Theorem 16.17 is that Theorem 16.17 assumes different data. Theorem 16.17 assumes an orthonormal set is given. Theorem 16.57 on the other hand assumes that only a subspace is given. Having a basis is more information than having a subspace (since one is free to choose the basis).

Proof. Let $\mathcal{S}$ be an orthonormal set for W, the existence of which is guaranteed by the Gram-Schmidt procedure. Then apply Theorem 16.17 for uniqueness. To see that $P_W\vec{v}$ minimizes the distance from $\vec{v}$ to W, let $\vec{w}$ be another vector in W that is not equal to $P_W\vec{v}$. Then the vectors $\vec{v} - P_W\vec{v}$ and $P_W\vec{v} - \vec{w}$ are orthogonal because $\vec{v} - P_W\vec{v} \in W^\perp$ and $P_W\vec{v} - \vec{w} \in W$.



Hence, by Pythagorean's theorem,

$$\begin{aligned} \|\vec{v} - \vec{w}\| &= \|\vec{v} - P_W\vec{v} + P_W\vec{v} - \vec{w}\| \\ &= \|\vec{v} - P_W\vec{v}\| + \|P_W\vec{v} - \vec{w}\| \\ &> \|\vec{v} - P_W\vec{v}\|, \end{aligned} \tag{16.59}$$

which shows that $P_W\vec{v}$ minimizes the distance from $\vec{v}$ to W. ■

Theorem 16.60. *Let W be a k-dimensional subspace of $\mathbb{R}^n$. Then the orthogonal complement $W^\perp$ is an $(n-k)$-dimensional subspace of $\mathbb{R}^n$. Furthermore, the orthogonal complement of the orthogonal complement is the original subspace itself, i.e.*

$$(W^\perp)^\perp = W. \tag{16.61}$$

Proof. To prove this, it must be shown that $(W^\perp)^\perp \subseteq W$ and $(W^\perp)^\perp \supseteq W$.

Let us begin with the second claim first. Let $\vec{w} \in W$. The goal is to show that $\langle \vec{w}, \vec{u} \rangle = 0$ for all $\vec{u} \in W^\perp$. By definition of $W^\perp$, this means that $\langle \vec{u}, \vec{w}' \rangle = 0$ for all $\vec{w}' \in W$. Since $\vec{w} \in W$, this implies $\langle \vec{u}, \vec{w} \rangle = 0$ so that $\vec{w} \in (W^\perp)^\perp$.

Let $\vec{v} \in (W^\perp)^\perp$. The goal is to show that $\vec{v} \in W$. By Theorem 16.57, there exists unique vectors $\vec{w} \in W$ and $\vec{u} \in W^\perp$ such that $\vec{v} = \vec{w} + \vec{u}$. Taking the inner product of both sides with $\vec{u}$ gives

$$\underbrace{\langle \vec{u}, \vec{v} \rangle}_{=0 \text{ since } \vec{v} \in (W^\perp)^\perp} = \underbrace{\langle \vec{u}, \vec{w} \rangle}_{=0 \text{ since } \vec{u} \in W^\perp} + \langle \vec{u}, \vec{u} \rangle. \tag{16.62}$$

This implies $\vec{u} = \vec{0}$. Hence, $\vec{v} \in W$. ■

Recommended Exercises. Please check HuskyCT for the homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

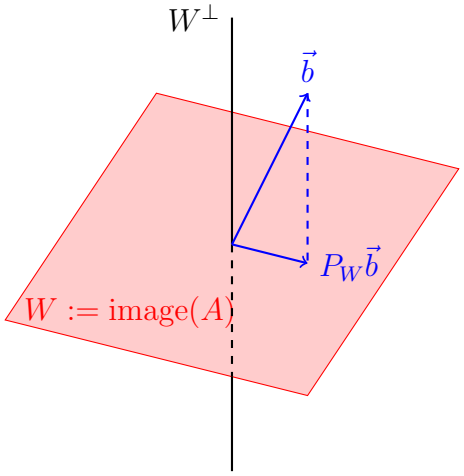
In this lecture, we covered Sections 6.3, 6.4, and parts of 6.7.

Terminology checklist

projection	
orthogonal projectiont	
normalized/unit vector	
Gram-Schmidt procedure	

17 Least squares approximation

If $\mathbb{R}^m \xleftarrow{A} \mathbb{R}^n$ is a linear transformation and $\vec{b} \in \mathbb{R}^m$ is a vector, a solution to $A\vec{x} = \vec{b}$ exists if and only if the system is consistent. In some applications, a system might be over-constrained, meaning that there are many more equations than unknowns. This corresponds to A having more rows than columns. In this situation, it is often the case that $A\vec{x} = \vec{b}$ will only be consistent for very special values of $\vec{b}$. However, given what we have learned, it should now be possible to find the best approximation to such a system



by projecting $\vec{b}$ onto the image of A. In this case, the best approximation solutions to $A\vec{x} = \vec{b}$ are the solutions to $A\vec{x} = P_W\vec{b}$. The latter is always consistent since W is precisely the image (column space) of A. The reason this is the best approximation to $A\vec{x} = \vec{b}$ follows from Theorem 16.57.

Definition 17.1. Let $\mathbb{R}^m \xleftarrow{A} \mathbb{R}^n$ be a linear transformation and let $\vec{b} \in \mathbb{R}^m$. A *least-squares approximation to $A\vec{x} = \vec{b}$* is a solution to $A\vec{x} = P_W\vec{b}$, where $W := \text{image}(A)$.

It takes a bit of work to calculate the projection onto a subspace. For example, we need an orthogonal basis of W to use the formulas we have discovered earlier. Fortunately, because the subspace is the image of a linear transformation, a drastic simplification can be made to the above definition.

Theorem 17.2. *Let $\mathbb{R}^m \xleftarrow{A} \mathbb{R}^n$ be a linear transformation and let $\vec{b} \in \mathbb{R}^m$. $\vec{x} \in \mathbb{R}^n$ is a least-squares approximation to $A\vec{x} = \vec{b}$ if and only if $\vec{x}$ is a solution to $A^TA\vec{x} = A^T\vec{b}$.*

Proof. For this proof, set $W := \text{image}(A)$.
($\Rightarrow$) Let $\vec{x} \in \mathbb{R}^n$ be a solution to $A\vec{x} = P_W\vec{b}$. Since $\vec{b} - P_W\vec{b} \in W^\perp$,

$$\langle A\vec{e}_k, \vec{b} - P_W\vec{b} \rangle = 0 \tag{17.3}$$

for all $k \in \{1, \ldots, n\}$. By the relationship between the inner product and the transpose of vectors, this equation says

$$(A\vec{e}_k)^T(\vec{b} - P_W\vec{b}) = 0 \tag{17.4}$$

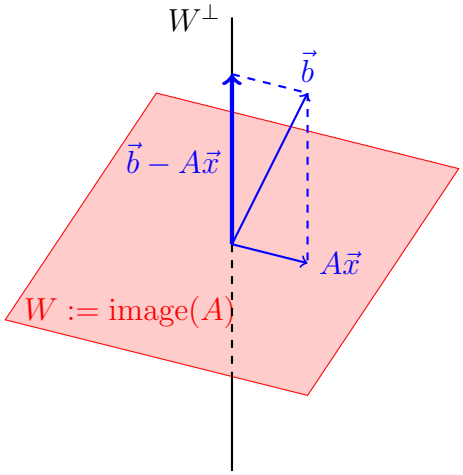
for all $k \in \{1, \ldots, n\}$. Hence,

$$A^T(\vec{b} - P_W\vec{b}) = \vec{0}. \tag{17.5}$$

By assumption, $P_W\vec{b} = A\vec{x}$ so that this says

$$A^T(\vec{b} - A\vec{x}) = 0 \qquad \Rightarrow \qquad A^TA\vec{x} = A^T\vec{b}. \tag{17.6}$$

($\Leftarrow$) Let $\vec{x} \in \mathbb{R}^n$ be a solution to $A^TA\vec{x} = A^T\vec{b}$. Then $A^T(\vec{b} - A\vec{x}) = \vec{0}$ implies that $\vec{b} - A\vec{x} \in W^\perp$.



By Theorem 16.57, $\vec{b}$ can be expressed uniquely as $\vec{b} = \vec{w} + \vec{u}$ with $\vec{w} \in W$ and $\vec{u} \in W^\perp$, but by the above calculations

$$\vec{b} = A\vec{x} + (\vec{b} - A\vec{x}) \tag{17.7}$$

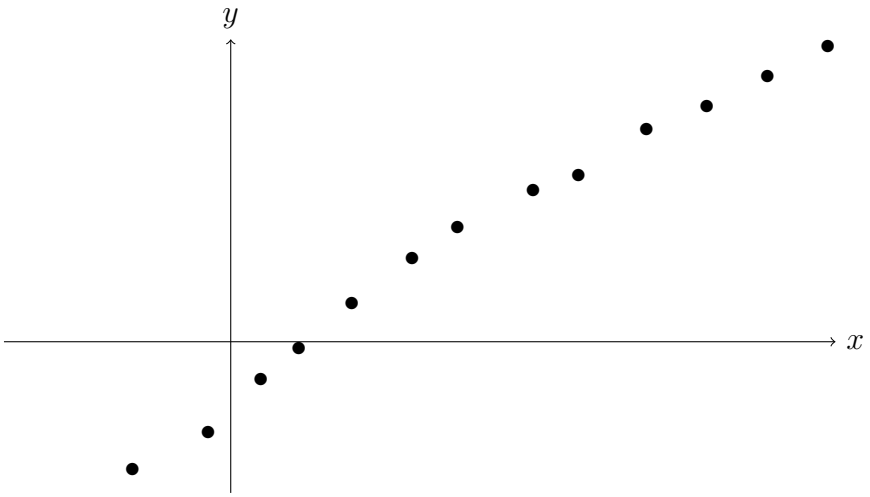
is such a decomposition. By uniqueness, this means that $A\vec{x} = P_W\vec{b}$. ■

Let us use this to work out, in full, the general linear regression problem for fitting data to a straight line.

Example 17.8. Consider a collection of d data points, where d is some positive integer (typically taken to be large), in $\mathbb{R}^2$. Let us denote these data points by

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}, \ldots, \begin{bmatrix} x_d \\ y_d \end{bmatrix}. \tag{17.9}$$

If the data points seem to lie close to a straight line, as in the following figure,



then it should be possible to try to find an equation describing this line of the form $y = mx + b$. In other words, we would like to find a slope m and a y-intercept b such that

$$\begin{aligned} y_1 &= mx_1 + b \\ y_2 &= mx_2 + b \\ &\vdots \\ y_d &= mx_d + b. \end{aligned} \tag{17.10}$$

It's typically unreasonable to expect to be able to solve a linear system of d equations in two unknowns (the unknowns here are m and b). In other words, it's usually not possible to find a straight line going through d many points in $\mathbb{R}^2$. Notice that (17.10) is equivalent to the matrix equation

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_d & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}. \tag{17.11}$$

This equation is of the form $A\vec{c} = \vec{y}$ where we would like to solve for $\vec{c}$.[50] As we stated above, we cannot solve this in general. However, Theorem 17.2 tells us that we can find a best approximation by solving $A^TA\vec{c} = A^T\vec{y}$ instead. Let us therefore solve this system. To do this, we'll calculate each side of this equation. The left-hand-side, ignoring the $\vec{c}$ for now, gives

$$A^TA = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_d & 1 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{d} x_j^2 & \sum_{j=1}^{d} x_j \\ \sum_{j=1}^{d} x_j & d \end{bmatrix}. \tag{17.12}$$

The right-hand-side gives

$$A^T\vec{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{d} x_j y_j \\ \sum_{j=1}^{d} y_j \end{bmatrix}. \tag{17.13}$$

The 2×2 matrix in (17.12) is not always invertible. However, the situations in which it is not invertible are rare. For the time being, let us therefore suppose that it is invertible. In this case, $A^TA\vec{c} = A^T\vec{y}$ can be solved by multiplying both sides of the equation by this inverse to yield $\vec{c} = (A^TA)^{-1}A^T\vec{y}$, which would tell us what the slope m and y-intercept b would be purely in terms of the data points. The inverse of A^TA is simple to calculate because it is a 2×2 matrix,

[50] We are using the variable name $\vec{c}$ now instead of $\vec{x}$ to avoid the potential confusion with the data points which are labelled using x_j.

and the result is

$$(A^TA)^{-1} = \frac{1}{d\sum_{j=1}^{d} x_j^2 - \sum_{i,j=1}^{d} x_i x_j} \begin{bmatrix} d & -\sum_{k=1}^{d} x_k \\ -\sum_{k=1}^{d} x_k & \sum_{k=1}^{d} x_k^2 \end{bmatrix}, \tag{17.14}$$

where I have relabeled some indices to avoid potential confusing. Multiplying this with our result for $A^T\vec{y}$ gives

$$\begin{aligned} (A^TA)^{-1}A^T\vec{y} &= \frac{1}{d\sum_{j=1}^{d} x_j^2 - \sum_{i,j=1}^{d} x_i x_j} \begin{bmatrix} d & -\sum_{k=1}^{d} x_k \\ -\sum_{k=1}^{d} x_k & \sum_{k=1}^{d} x_k^2 \end{bmatrix} \begin{bmatrix} \sum_{l=1}^{d} x_l y_l \\ \sum_{l=1}^{d} y_l \end{bmatrix} \\ &= \frac{1}{d\sum_{j=1}^{d} x_j^2 - \sum_{i,j=1}^{d} x_i x_j} \begin{bmatrix} d\sum_{l=1}^{d} x_l y_l - \sum_{k,l=1}^{d} x_k y_l \\ -\sum_{k,l=1}^{d} x_k x_l y_l + \sum_{k,l=1}^{d} x_k^2 y_l \end{bmatrix}. \end{aligned} \tag{17.15}$$

Pushing the overall factor in front inside the vector expression and dividing the numerator and denominator by d in the top entry, $\vec{c} = (A^TA)^{-1}A^T\vec{y}$ gives

$$m = \frac{\sum_{l=1}^{d} x_l y_l - \frac{1}{d}\sum_{k,l=1}^{d} x_k y_l}{\sum_{j=1}^{d} x_j^2 - \frac{1}{d}\sum_{i,j=1}^{d} x_i x_j} \qquad \& \qquad b = \frac{\sum_{k,l=1}^{d} x_k y_l (x_k - x_l)}{d\sum_{j=1}^{d} x_j^2 - \sum_{i,j=1}^{d} x_i x_j}. \tag{17.16}$$

In statistics, the numerator in the expression for m is typically called the *covariance* between the x and y data points while the expression in the denominator is called the *variance* of x

$$\mathrm{Var}[x] := \sum_{j=1}^{d} x_j^2 - \frac{1}{d}\sum_{i,j=1}^{d} x_i x_j \qquad \& \qquad \mathrm{Cov}[x,y] := \sum_{l=1}^{d} x_l y_l - \frac{1}{d}\sum_{k,l=1}^{d} x_k y_l. \tag{17.17}$$

This gives a complete solution to fitting data points in $\mathbb{R}^2$ to a straight line. However, what if A^TA is not invertible? This happens when the determinant is zero, which occurs when the variance in

the x data points vanishes. The variance can be rewritten as

$$\begin{aligned}
\mathrm{Var}[x] &= \sum_{j=1}^{d} x_j^2 - \frac{1}{d}\sum_{i,j=1}^{d} x_i x_j \\
&= \sum_{j=1}^{d} x_j^2 - \frac{2}{d}\sum_{i,j=1}^{d} x_j x_i + \frac{1}{d}\sum_{i,k=1}^{d} x_i x_k \\
&= \sum_{j=1}^{d} x_j^2 - \frac{2}{d}\sum_{i,j=1}^{d} x_j x_i + \frac{1}{d^2}\sum_{i,j,k=1}^{d} x_i x_k \\
&= \sum_{j=1}^{d}\left(x_j^2 - \frac{2x_j}{d}\sum_{i=1}^{d} x_i + \frac{1}{d^2}\sum_{i,k=1}^{d} x_i x_k\right) \\
&= \sum_{j=1}^{d}\left(x_j - \frac{1}{d}\sum_{i=1}^{d} x_i\right)^2
\end{aligned} \tag{17.18}$$

where we have used $d = \sum_{i=1}^{d} 1$ in the third line. This shows that the variance is always non-negative. Furthermore, since it is the sum of non-negative terms, the only way for $\mathrm{Var}[x]$ to be zero is if each of the terms in the final sum is zero, and this would mean that

$$x_j = \frac{1}{d}\sum_{i=1}^{d} x_i \tag{17.19}$$

for all $j \in \{1, \ldots, d\}$, i.e. if all of the x_j data points are equal. Therefore, if you are sampling data points and wish to make a graph out of these data points, it would be impossible to do so anyway if all of the x_j's were equal. Therefore, it is very reasonable to assume that the determinant of our $A^T A$ matrix is non-zero.

Problem 17.20. In her introductory physics lab, Joanna drops a ball from several different heights and counts how long it takes for the ball to reach the ground. She does several measurements for each height in increments of 50 centimeters from 0 to 2 meters. She then takes the average of her measurements for each height. Her results are listed in the following table without their uncertainties.

Height (meters)	0.5	1.0	1.5	2.0	2.5	3.0
Time (seconds)	0.38	0.49	0.55	0.66	0.74	0.76

She expects the relationship between the height, x, and time, t, to be quadratic and of the form $x(t) = \frac{1}{2}gt^2$, where g is a constant to be determined from these data. What is the best approximation to g?

Answer. Squaring the times gives the following table

Height (meters)	0.5	1.0	1.5	2.0	2.5	3.0
Time^2 (seconds^2)	0.14	0.24	0.30	0.44	0.55	0.58

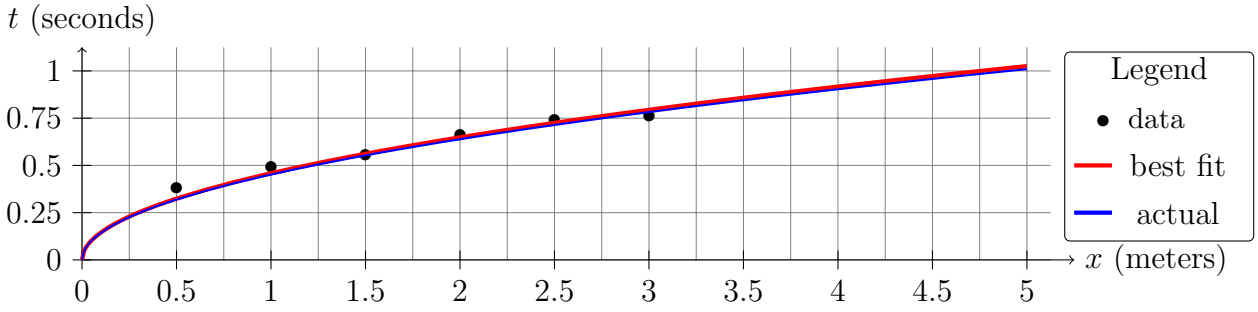
It therefore suffices to find the slope m so that the line $x = mt^2$ best fits these data. Let A be the 6×1 matrix for the input data, the square of time and let $\vec{b}$ be the vector for the output data, the height. In this case, $A\vec{x} = \vec{b}$ takes the form

$$\begin{bmatrix} 0.14 \\ 0.24 \\ 0.30 \\ 0.44 \\ 0.55 \\ 0.58 \end{bmatrix} \begin{bmatrix} m \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.0 \\ 1.5 \\ 2.0 \\ 2.5 \\ 3.0 \end{bmatrix} \tag{17.21}$$

and the goal is to solve for the best approximation to m. A^TA and $A^T\vec{b}$ are given by

$$0.9997m = 4.755 \qquad \Rightarrow \qquad m = 4.76 \qquad \Rightarrow \qquad g = 9.51. \tag{17.22}$$

The actual value of g is known to be 9.81 in Joanna's location. The data points, her best fit curve, and the actual curve are depicted in the following figure.



Problem 17.23. The Michaelis-Menten equation [7] is used as a model in biochemistry to describe the reaction rate, v, of an enzymatic reaction to the concentration, $[S]$, of the substrate. The formula is given by

$$v = \frac{v_{\max}[S]}{K_M + [S]}, \tag{17.24}$$

where $v_{\max}$ is a constant that describes the maximum possible achieved reaction rate and K_M is another constant that describes the substrate concentration when the reaction rate is half of $v_{\max}$. My friend did such an experiment and here is his data[51]

$[S]$ (millimolar)	0.900	0.675	0.450	0.225	0.090	0.045	0.0225
v (micromoles per min)	0.210	0.200	0.167	0.120	0.053	0.031	0.017

Find the coefficients $v_{\max}$ and K_M that best fit these data.

Answer. The Michaelis-Menten equation can be flipped so that it is of the form

$$\frac{1}{v} = \frac{K_M + [S]}{v_{\max}[S]} = \frac{1}{v_{\max}} + \frac{K_M}{v_{\max}} \frac{1}{[S]}. \tag{17.25}$$

In terms of the variables $\frac{1}{[S]}$ instead of $[S]$, this is a linear equation of the form $y = mx + b$ where m and b are the unknowns. Hence, we can apply the least-squares method. For this, we translate our data in terms of these reciprocated variables.

[51] I'd like to thank David Lei for sharing his data.

$1/[S]$	1.11	1.48	2.22	4.44	11.1	22.2	44.4
$1/v$	4.76	5.00	5.99	8.33	18.9	32.3	58.8

Therefore, set

$$A := \begin{bmatrix} 1 & 1.11 \\ 1 & 1.48 \\ 1 & 2.22 \\ 1 & 4.44 \\ 1 & 11.1 \\ 1 & 22.2 \\ 1 & 44.4 \end{bmatrix}, \qquad \vec{b} := \begin{bmatrix} 4.76 \\ 5.00 \\ 5.99 \\ 8.33 \\ 18.9 \\ 32.3 \\ 58.8 \end{bmatrix}, \qquad \& \qquad \vec{x} := \begin{bmatrix} 1/v_{\max} \\ K_M/v_{\max} \end{bmatrix}. \tag{17.26}$$

Solving $A^T A\vec{x} = A^T\vec{b}$ would give us the coefficients, which we can then use to find the best fit curve. This system then becomes

$$\begin{bmatrix} 7.00 & 87.0 \\ 87.0 & 2620 \end{bmatrix} \begin{bmatrix} 1/v_{\max} \\ K_M/v_{\max} \end{bmatrix} = \begin{bmatrix} 134 \\ 3600 \end{bmatrix}. \tag{17.27}$$

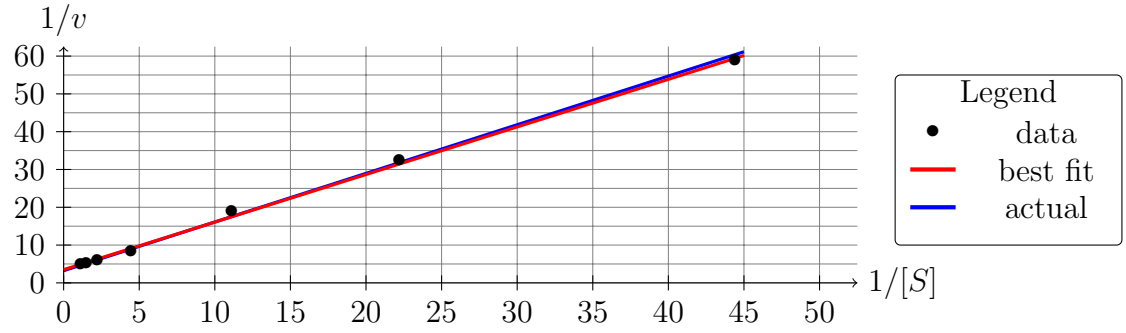
Row reduction then gives

$$\begin{bmatrix} 1/v_{\max} \\ K_M/v_{\max} \end{bmatrix} = \begin{bmatrix} 3.50 \\ 1.26 \end{bmatrix}. \tag{17.28}$$

Solving for $v_{\max}$ and K_M gives

$$v_{\max} = 0.286 \qquad \& \qquad K_M = 0.360. \tag{17.29}$$

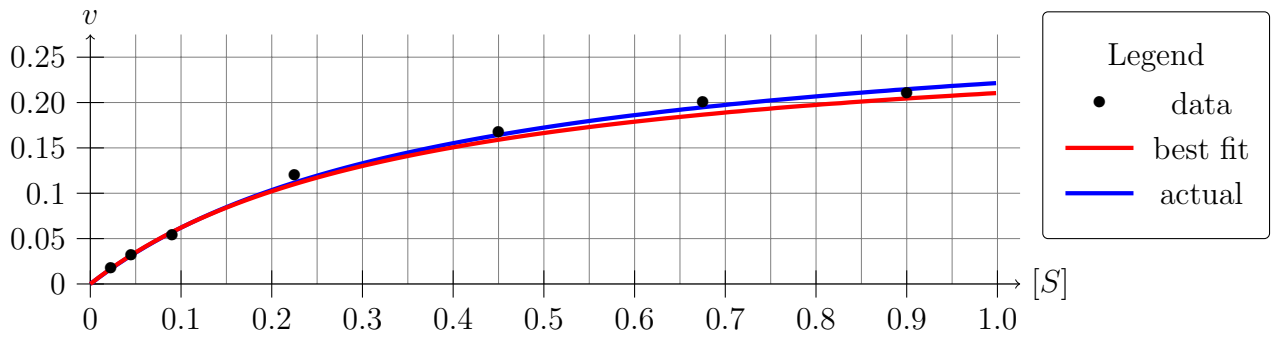
Plotting the inverse relation gives the following curve



Plotting the best fit curve

$$v\big([S]\big) = \frac{0.286[S]}{0.360 + [S]} \tag{17.30}$$

to the original Michaelis-Menten equation gives



The "actual" plot is actually the best fit curve obtained through a different method that is more standard for this specific kind of equation.[52] Nevertheless, the least-squares method provides a reasonable approximation.

In the previous examples, one could fit data to any curve provided that one is using linear combinations of functions that are linearly independent. This will make more sense when we discuss vector spaces and how certain spaces of functions are vector spaces, but for now we can provide a specific, yet somewhat imprecise, definition.

Definition 17.31. A set of functions $\{f_1, \ldots, f_k\}$ on some common domain in $\mathbb{R}$ is said to be *linearly indepedent* iff

$$a_1 f_1 + \cdots + a_k f_k = 0 \qquad \Rightarrow \qquad a_1 = \cdots = a_k = 0, \tag{17.32}$$

i.e.

$$a_1 f_1(x) + \cdots + a_k f_k(x) = 0 \text{ for all } x \text{ in the domain} \qquad \Rightarrow \qquad a_1 = \cdots = a_k = 0. \tag{17.33}$$

In this notation $a_1, \ldots, a_k$ are just some coefficients and the expression $a_1 f_1 + \cdots + a_k f_k$ is a *linear combination* of the functions in the set $\{f_1, \ldots, f_k\}$.

If you have data that must fit to some curve of the form

$$a_1 f_1 + \cdots + a_k f_k, \tag{17.34}$$

where $\{f_1, \ldots, f_k\}$ is some set of linearly independent functions, then your goal is to find the coefficients $\{a_1, \ldots, a_k\}$ so that the function $a_1 f_1 + \cdots + a_k f_k$ best fits your data. If your data inputs are $\{x_1, \ldots, x_d\}$ and your data outputs are $\{y_1, \ldots, y_d\}$, then your matrix A is given by

$$A := \begin{bmatrix} f_1(x_1) & \cdots & f_k(x_1) \\ \vdots & & \vdots \\ f_1(x_d) & \cdots & f_k(x_d) \end{bmatrix} \tag{17.35}$$

and the vector $\vec{b}$ is

$$\vec{b} := \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix}. \tag{17.36}$$

In the previous two examples, the functions are given as follows. For the ball being dropped from a height, there is actually only one function and it is given by $f(t) = t^2$. The coefficient is $\frac{g}{2}$. For the Michaelis-Menten equation, after taking the reciprocal, there are two functions. The first one is $f_1([S]) = 1$, which is just a constant, and the second one is $f_2([S]) = \frac{1}{[S]}$. Taking the reciprocal was important because this allowed us to express $\frac{1}{v}$ as a linear combination of these two functions, namely

$$\frac{1}{v} = \left(\frac{1}{v_{\max}}\right) f_1 + \left(\frac{K_M}{v_{\max}}\right) f_2. \tag{17.37}$$

[52] The reason the actual fit is used is because most of the data points are clustered near small values of $1/[S]$ as opposed to being distributed somewhat evenly. This means that the least-squares method we are using is not as accurate due to the lack of data for larger values of $1/[S]$. The original data is much more evenly distributed in terms of $[S]$.

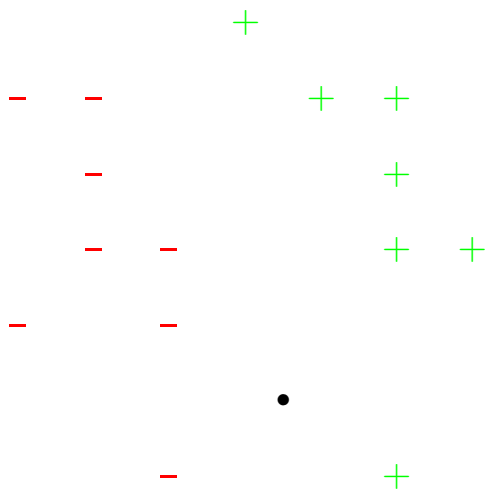
Recommended Exercises. Please check HuskyCT for the homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

Terminology checklist

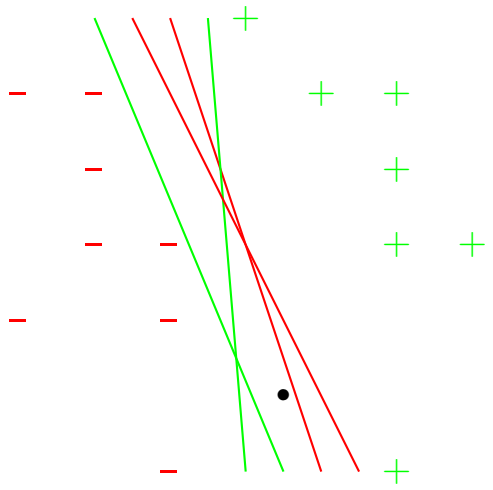
least-squares approximation/linear regression	
linear combination of functions	
linear independence of functions	

18 Decision making and support vector machines*

[Warning: this section has to be substantially edited for correctness. Update will come soon] We now move on to a different application of orthogonality in the context of machine learning and artificial intelligence.[53] The setup is that one has a large range of data $\mathcal{X} := \{\vec{x}_1, \ldots, \vec{x}_d\}$ described by vectors in $\mathbb{R}^n$ and these data separate into two types, $\mathcal{X}_+$ and $\mathcal{X}_-$. If a new data point $\vec{x}_{d+1}$ is provided, the machine must then decide to place this new data point in $\mathcal{X}_+$ or $\mathcal{X}_-$.



The new data point is drawn as a bullet •. To make this decision, the machine must draw a hyperplane, an $(n-1)$-dimensional linear manifold, that divides $\mathbb{R}^n$ into two parts in the most optimal way. Different hyperplanes will give different answers.



We would like to therefore establish a convention for a unique such hyperplane that is also the most optimal one to allow for the most accurate identification. How do we define the most optimal hyperplane? We will define a separating hyperplane and then define optimality, but first there are a few facts we should establish.

[53] I'd like to thank Benjamin Russo for helpful discussions on this topic.

Lemma 18.1. *Let $H \subseteq \mathbb{R}^n$ be a hyperplane. Then there exists a vector $\vec{w} \in \mathbb{R}^n \setminus \{\vec{0}\}$ and a real number $c \in \mathbb{R}$ such that H is the solution set of $\langle \vec{w}, \vec{x} \rangle - c = 0$.*

Proof. Let $\vec{h} \in H$. Then $H - \vec{h}$ is an $(n-1)$-dimensional subspace in $\mathbb{R}^n$. Hence, $(H - \vec{h})^\perp$ is a one-dimensional subspace spanned by some normalized vector $\hat{u}$. Because $\text{span}\{\hat{u}\}$ is perpendicular to H, there exists an $a \in \mathbb{R}$ such that $a\hat{u} \in H$. Then, H is the solution set of $\langle \hat{u}, \vec{x} \rangle - a = 0$. ■

Notice that the vectors $\vec{w}$ and numbers c need not be unique. Indeed, we can multiply the previous system by any non-zero real number λ to get $\langle \lambda\hat{u}, \vec{x} \rangle - \lambda a = 0$. Furthermore, notice that if H is the solution set of $\langle \vec{w}, \vec{x} \rangle - c = 0$ for some nonzero vector $\vec{w}$ and some number $c \in \mathbb{R}$, then the vector

$$\vec{x} = \frac{c}{\|\vec{w}\|}\hat{w} = \frac{c}{\|\vec{w}\|^2}\vec{w} \tag{18.2}$$

is in H. This is because

$$\left\langle \vec{w}, \frac{c}{\|\vec{w}\|}\hat{w} \right\rangle - c = c\langle \hat{w}, \hat{w} \rangle - c = c - c = 0. \tag{18.3}$$

This tells us that the orthogonal distance from the origin, the zero vector, to the hyperplane H is $\frac{c}{\|\vec{w}\|}$.

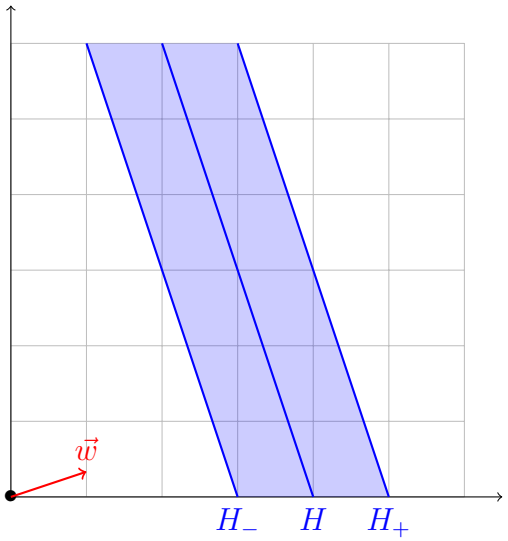
Definition 18.4. Let $\vec{w} \in \mathbb{R}^n \setminus \{\vec{0}\}$ and $c \in \mathbb{R}$ with associated plane H given by the solution set of $\langle \vec{w}, \vec{x} \rangle - c = 0$. The *marginal planes* H_+ and H_- associated to H are the solution sets to $\langle \vec{w}, \vec{x} \rangle - c = 1$ and $\langle \vec{w}, \vec{x} \rangle - c = -1$, respectively, i.e.

$$H_\pm := \left\{ \vec{x} \in \mathbb{R}^n \ : \ \langle \vec{w}, \vec{x} \rangle - c = \pm 1 \right\}. \tag{18.5}$$

For example, in $\mathbb{R}^2$, if

$$\vec{w} = \frac{1}{3}\begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad \& \qquad c = 4 \tag{18.6}$$

then these planes would look like the following



To check this, H is described by the linear system

$$\frac{1}{3}(3x+y)=4 \qquad \Rightarrow \qquad y=12-3x, \tag{18.7}$$

H_+ is described by

$$\frac{1}{3}(3x+y)=5 \qquad \Rightarrow \qquad y=15-3x, \tag{18.8}$$

and H_- is described by

$$\frac{1}{3}(3x+y)=3 \qquad \Rightarrow \qquad y=9-3x. \tag{18.9}$$

Notice that the vector

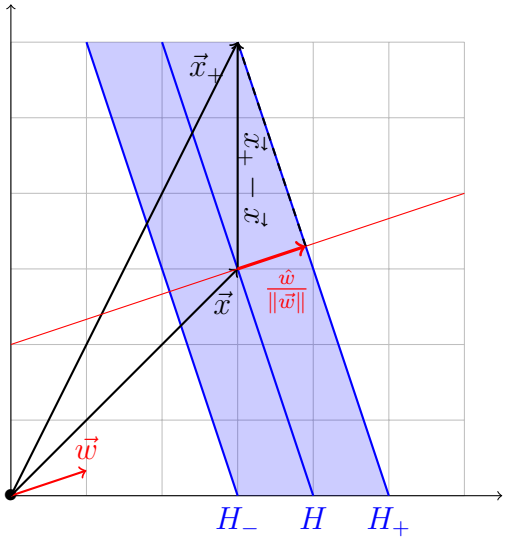
$$\left(\frac{c}{\|\vec{w}\|^2}\right)\vec{w}=\frac{6}{5}\begin{bmatrix}3\\1\end{bmatrix} \tag{18.10}$$

lies on the plane H.

Lemma 18.11. *Let $(\vec{w},c)\in(\mathbb{R}^n\setminus\{\vec{0}\})\times\mathbb{R}$ describe a hyperplane H. The perpendicular distance between H and H_+ is $\frac{1}{\|\vec{w}\|}$ and similarly for the distance between H and H_-.*

Proof. Let $\vec{x}_+\in H_+$ and $\vec{x}\in H$. The orthogonal distance between H and H_+ is given by

$$\left\langle \vec{x}_+-\vec{x},\frac{\vec{w}}{\|\vec{w}\|}\right\rangle=\frac{1}{\|\vec{w}\|}\big(\langle\vec{w},\vec{x}_+\rangle-\langle\vec{w},\vec{x}\rangle\big)=\frac{1}{\|\vec{w}\|}\big((1+c)-c\big)=\frac{1}{\|\vec{w}\|} \tag{18.12}$$



by the definition of H and H_+ in terms of $(\vec{w},c)$. A similar calculation holds for H_-. ∎

In these two cases, notice that the vectors

$$\vec{x}_\pm:=\left(\frac{c}{\|\vec{w}\|}\pm\frac{1}{\|\vec{w}\|}\right)\hat{w} \tag{18.13}$$

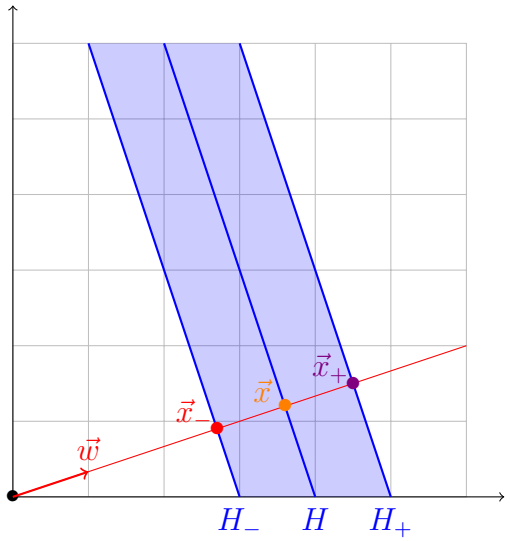
are vectors in $H_\pm$. This is because

$$\left\langle\vec{w},\left(\frac{c}{\|\vec{w}\|}\pm\frac{1}{\|\vec{w}\|}\right)\hat{w}\right\rangle-c=\frac{c}{\|\vec{w}\|}\langle\vec{w},\hat{w}\rangle\pm\frac{1}{\|\vec{w}\|}\langle\vec{w},\hat{w}\rangle-c=c\pm1-c=\pm1. \tag{18.14}$$

In the example we have been using, we have

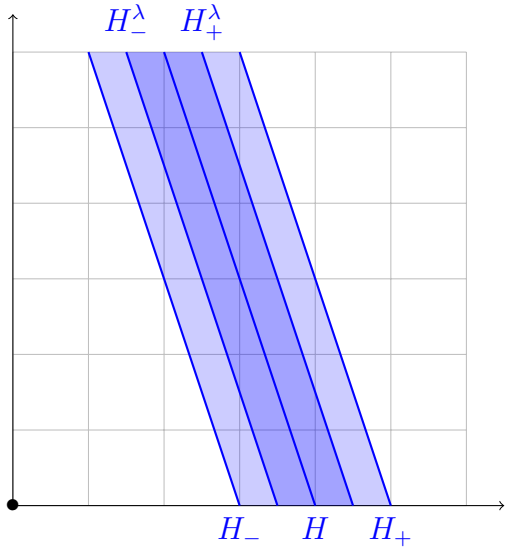
$$\vec{x}_- = \left(\frac{c-1}{\|\vec{w}\|^2}\right)\vec{w} = \frac{9}{10}\begin{bmatrix}3\\1\end{bmatrix}, \quad \vec{x} = \left(\frac{c}{\|\vec{w}\|^2}\right)\vec{w} = \frac{6}{5}\begin{bmatrix}3\\1\end{bmatrix}, \quad \& \quad \vec{x}_+ = \left(\frac{c+1}{\|\vec{w}\|^2}\right)\vec{w} = \frac{3}{2}\begin{bmatrix}3\\1\end{bmatrix} \tag{18.15}$$

as the three vectors that in the span of $\vec{w}$ and that pass through H_-, H, and H_+, respectively.



Definition 18.16. Let $(\vec{w}, c) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ describe a hyperplane H. The convex region between H_+ and H_- is called the *margin* of $(\vec{w}, c)$. The orthogonal distance between H_+ and H_-, which is given by $\frac{2}{\|\vec{w}\|}$, is called the *margin width* of $(\vec{w}, c)$.

Even though $(\vec{w}, c)$ can be scaled to $(\lambda\vec{w}, \lambda c)$ to give the same H, notice that the marginal planes are different. This is because the margin width has scaled by a factor of $\frac{1}{\lambda}$. For example, if we set $\lambda = 2$ in our example, the margin shrinks by $\frac{1}{2}$.



In this drawing, we've used the notation $H_\pm^\lambda$ to signify the resulting marginal planes for $(\lambda\vec{w}, \lambda c)$. If instead we only scale $\vec{w}$, but not c, to get $(\lambda\vec{w}, c)$, then we change the position of the hyperplane

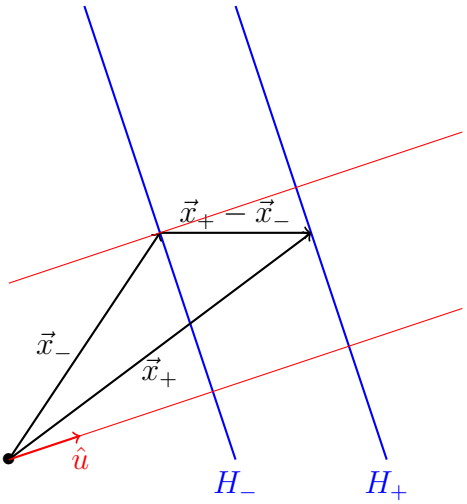
because the new equation that it is the solution set of is

$$\langle \lambda\vec{w}, \vec{x}\rangle - c = 0 \iff \langle \vec{w}, \vec{x}\rangle - \frac{c}{\lambda} = 0. \tag{18.17}$$

Therefore, the hyperplane $(\lambda\vec{w}, c)$ is equivalent to the hyperplane $\left(\vec{w}, \frac{c}{\lambda}\right)$. However, their margins, and hence their marginal planes, will be different. Therefore, think of the $\vec{w}$ in $(\vec{w}, c)$ as determining a direction as well as a margin width and think of c in $(\vec{w}, c)$ as determining the position of the central hyperplane. We make this relationship between $(\vec{w}, c)$ and such triples of hyperplanes formal in the following Lemma.

Lemma 18.18. *Two parallel hyperplanes H_- and H_+ in $\mathbb{R}^n$ determine a unique $(\vec{w}, c) \in \left(\mathbb{R}^n \setminus \{\vec{0}\}\right) \times \mathbb{R}$ whose marginal planes agree with H_- and H_+.*

Proof. Let $\vec{x}_+ \in H_+$ and pick $\hat{u} \in (H_+ - \vec{x}_+)^\perp$ such that if $\lambda\hat{u} \in H_-$ and $\mu\hat{u} \in H_+$, then $\lambda < \mu$ (i.e. choose a normal vector $\hat{u}$ perpendicular to H_+ that points from H_- to H_+). Also, let $\vec{x}_- \in H_-$ (any choice of vectors will work). The orthogonal separation between the planes H_+ and H_- is given by $\langle \vec{x}_+ - \vec{x}_-, \hat{u}\rangle$.



Therefore, set

$$\vec{w} := \left(\frac{2}{\langle \vec{x}_+ - \vec{x}_-, \hat{u}\rangle}\right)\hat{u}. \tag{18.19}$$

Now, pick any $\vec{x}_+ \in H_+$ and set

$$c := \langle \vec{w}, \vec{x}_+\rangle - 1. \tag{18.20}$$

Then $(\vec{w}, c)$ has H_+ and H_- as its marginal planes. ■

Exercise 18.21. Finish the proof by showing that $(\vec{w}, c)$ has H_+ and H_- as its marginal planes, i.e. show that H_+ is the solution set to $\langle \vec{w}, \vec{x}\rangle - c = 1$ and H_- is the solution set to $\langle \vec{w}, \vec{x}\rangle - c = -1$.

This result says that there is a 1-1 correspondence between the set of (ordered) pairs of parallel hyperplanes and the set $\left(\mathbb{R}^n \setminus \{\vec{0}\}\right) \times \mathbb{R}$.

Definition 18.22. Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ denote a non-empty set $\mathcal{X}$ of vectors in $\mathbb{R}^n$ that are separated into the two (disjoint) non-empty sets $\mathcal{X}_+$ and $\mathcal{X}_-$. Such a collection of sets is called a *training data set*. A hyperplane $H \subseteq \mathbb{R}^n$, described by $(\vec{w}, c) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$, *separates* $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ iff

$$\langle \vec{w}, \vec{x}_+ \rangle - c > 0 \qquad \& \qquad \langle \vec{w}, \vec{x}_- \rangle - c < 0 \tag{18.23}$$

for all $\vec{x}_+ \in \mathcal{X}_+$ and for all $\vec{x}_- \in \mathcal{X}_-$. In this case, H is said to be a *separating hyperplane* for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$. H *marginally separates* $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ iff

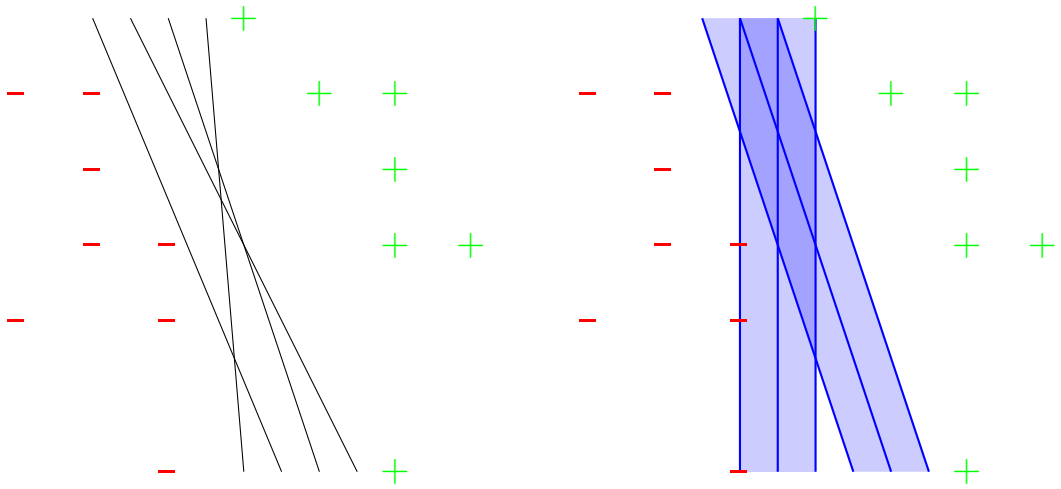
$$\langle \vec{w}, \vec{x}_+ \rangle - c \geq 1 \qquad \& \qquad \langle \vec{w}, \vec{x}_- \rangle - c \leq -1 \tag{18.24}$$

for all $\vec{x}_+ \in \mathcal{X}_+$ and for all $\vec{x}_- \in \mathcal{X}_-$. Let $\mathcal{S}_{\mathcal{X}} \subseteq (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ denote the set of hyperplanes that marginally separate $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$. Let $f : \mathcal{S}_{\mathcal{X}} \to \mathbb{R}$ be the function defined by

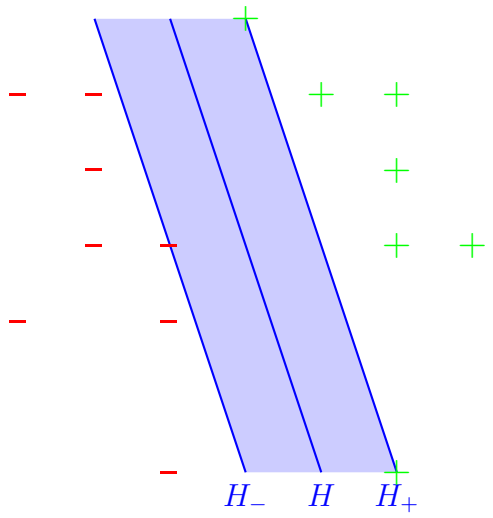
$$(\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R} \ni (\vec{w}, c) \mapsto f(\vec{w}, c) := \frac{2}{\|\vec{w}\|}, \tag{18.25}$$

i.e. the margin. A *support vector machine* (SVM) for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ is a maximum of f, i.e. an SVM is a pair $(\vec{w}, c) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ such that $\frac{1}{\|\vec{w}'\|} \leq \frac{1}{\|\vec{w}\|}$ for every other pair $(\vec{w}', c') \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$.

Some examples of separating hyperplanes and marginally separating hyperplanes are depicted in the following figures on the left and right, respectively.



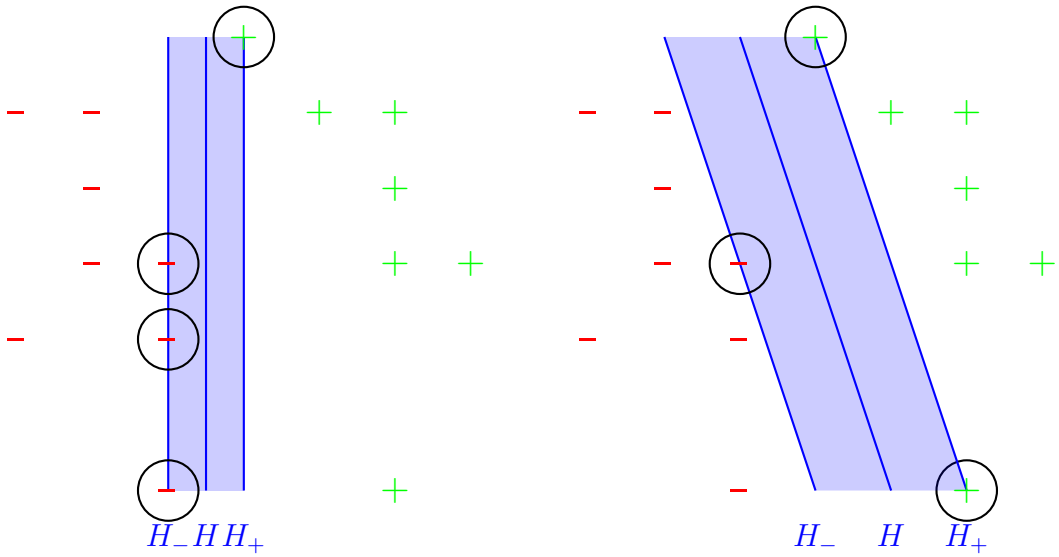
An SVM is a hyperplane that maximizes the margin, as in the following figure.



Because of this, it is useful to know when a given hyperplane that marginally separates a training data set can be enlarged. This will be useful because then instead of looking at the set of *all* marginally separating hyperplanes, we can focus our attention on those whose margins have been maximized. Afterwards, we will maximize the margin over *this* resulting set.

Definition 18.26. Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ be a training data set and let $(\vec{w}, c)$ be a marginally separating hyperplane for this set. The elements of $\mathcal{X} \cap H_\pm$ are called *support vectors* for $(\vec{w}, c)$. The set of support vectors is denoted by $H_{\mathcal{X}}^{\text{supp}}$. The notation $H_{\mathcal{X}_\pm}^{\text{supp}} := H_{\mathcal{X}}^{\text{supp}} \cap H_\pm$ will also be used to denote the set of positive and negative support vectors.

In the following figures, the support vectors have been circled for two different marginal hyperplanes.



Lemma 18.27. *Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ be a training data set and let $(\vec{w}, c)$ be a marginally separating hyperplane for this set. Then there exists a marginally separated hyperplane $(\vec{v}, d)$ such that*

$$\hat{v} = \hat{w} \qquad \& \qquad \frac{2}{\|\vec{v}\|} = \min_{\substack{\vec{x}_+ \in \mathcal{X}_+ \\ \vec{x}_- \in \mathcal{X}_-}} \langle \vec{x}_+ - \vec{x}_-, \hat{w} \rangle \,. \tag{18.28}$$

In other words, if the marginal planes do not contain any of the training data set, then the separating hyperplane can be translated and the margin width can be enlarged until the margin touches both positive and negative training data sets.

Proof. It will be convenient to define the function

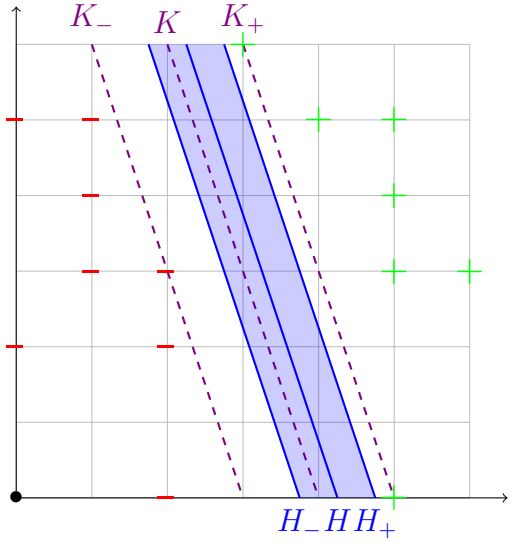
$$\mathcal{X} \ni \vec{x} \mapsto \theta(\vec{x}) := \begin{cases} +1 & \text{if } \vec{x} \in \mathcal{X}_+ \\ -1 & \text{if } \vec{x} \in \mathcal{X}_- \end{cases}. \tag{18.29}$$

By the discussions after Lemma 18.1 and Lemma 18.11, we have vectors in each H_-, H, and H_+ given by

$$\left(\frac{c-1}{\|\vec{w}\|}\right)\hat{w} \in H_-, \qquad \left(\frac{c}{\|\vec{w}\|}\right)\hat{w} \in H, \qquad \& \qquad \left(\frac{c+1}{\|\vec{w}\|}\right)\hat{w} \in H_+. \tag{18.30}$$

Set m_+ to be the remaining minimum orthogonal distance between H_+ and $\mathcal{X}_+$ and set m_- to be the remaining minimum orthogonal distance between H_- and $\mathcal{X}_-$, namely

$$m_\pm := \min_{\vec{x}_\pm \in \mathcal{X}_\pm} \theta(\vec{x}_\pm)\left(\langle \vec{x}_\pm, \hat{w}\rangle - \left(\frac{c \pm 1}{\|\vec{w}\|}\right)\right) \tag{18.31}$$



Therefore, the planes $K_\pm$ containing the vectors

$$\left(\frac{c \pm 1}{\|\vec{w}\|} \pm m_\pm\right)\hat{w} \tag{18.32}$$

that are perpendicular to $\hat{w}$ intersect $\mathcal{X}_\pm$ but do not contain points of $\mathcal{X}$ on the interior of their margin. By Lemma 18.18, there exists a $(\vec{v}, d) \in \left(\mathbb{R}^n \setminus \{\vec{0}\}\right) \times \mathbb{R}$ that describes these marginal separating hyperplanes, namely

$$\vec{v} := \left(\frac{2}{\frac{2}{\|\vec{w}\|} + m_+ + m_-}\right)\hat{w} \tag{18.33}$$

(since $\frac{2}{\|\vec{v}\|}$ is now the margin width between the new marginal hyperplanes) and

$$\begin{aligned} d &:= \left\langle \vec{v}, \left(\frac{c+1}{\|\vec{w}\|} + m_+ \right) \hat{w} \right\rangle - 1 \\ &= \frac{2\left(\frac{c+1}{\|\vec{w}\|} + m_+\right)}{\frac{2}{\|\vec{w}\|} + m_+ + m_-} - 1 \\ &= \frac{2c + \|\vec{w}\|(m_+ - m_-)}{2 + \|\vec{w}\|(m_+ + m_-)} \end{aligned} \tag{18.34}$$

(since this is the required number so that a vector on K_+ satisfies the positive marginal plane equation). ■

Exercise 18.35. Verify that $(\vec{v}, d)$ in the above proof defines marginally separating hyperplanes that are perpendicular to $\hat{w}$. Furthermore, explain why they cannot be enlarged any farther.

Theorem 18.36. *Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ be training data set for which there exists a separating hyperplane for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$. Then there exists a unique SVM for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$.*

Proof. By Lemma 18.27, it suffices to maximize the margin function f on the subset $\mathcal{S}_{\mathcal{X}}^{\text{supp}} \subseteq \mathcal{S}_{\mathcal{X}}$ consisting of marginally separating hyperplanes that have both positive and negative support vectors, namely on

$$\mathcal{S}_{\mathcal{X}}^{\text{supp}} := \big\{ (\vec{w}, c) \in \mathcal{S}_{\mathcal{X}} \ : \ H_\pm \cap \mathcal{X}_\pm \neq \varnothing \big\}. \tag{18.37}$$

The goal is therefore to maximize the margin function, which is a function of $(\vec{w}, c)$, subject to the constraint

$$\langle \vec{w}, \vec{x} \rangle - c \mp 1 = 0 \tag{18.38}$$

for all $\vec{x} \in \mathcal{S}_{\mathcal{X}}^{\text{supp}}$, or equivalently

$$\theta(\vec{x})\big(\langle \vec{w}, \vec{x} \rangle - c\big) - 1 = 0 \tag{18.39}$$

for all $\vec{x} \in \mathcal{S}_{\mathcal{X}}^{\text{supp}}$. Maximizing the margin function is equivalent to minimizing the function

$$(\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R} \ni (\vec{w}, c) \mapsto \frac{1}{2}\|\vec{w}\|^2 \tag{18.40}$$

subject to these same constraints. It is therefore equivalent to maximize the function g given by

$$(\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R} \ni (\vec{w}, c) \overset{g}{\mapsto} \frac{1}{2}\|\vec{w}\|^2 - \sum_{\vec{x} \in \mathcal{X}} \alpha_{\vec{x}} \Big(\theta(\vec{x})\big(\langle \vec{w}, \vec{x} \rangle - c\big) - 1 \Big). \tag{18.41}$$

Here, $\alpha_{\vec{x}} = 0$ for all $\vec{x} \in \mathcal{X} \setminus H_{\mathcal{X}}^{\text{supp}}$ and $\alpha_{\vec{x}}$ needs to be determined for all $\vec{x} \in H_{\mathcal{X}}^{\text{supp}}$. This condition guarantees that the function g equals f when restricted to $\mathcal{S}_{\mathcal{X}}^{\text{supp}}$ (but notice that it does not equal f on the larger domain $\mathcal{S}_{\mathcal{X}}$ of all marginally separating hyperplanes). The $\alpha_{\vec{x}}$ are called Lagrange multipliers. The extrema of g occur at points $(\vec{v}, d)$ for which the derivative of g vanishes with respect to these coordinates[54]

$$\frac{\partial g}{\partial \vec{w}}\Big|_{(\vec{w},c)} = 0 \qquad \& \qquad \frac{\partial g}{\partial c}\Big|_{(\vec{w},c)} = 0. \tag{18.42}$$

[54]Notice that it would not have made sense to take these derivatives if we had worked with the function f constrained to $\mathcal{S}_{\mathcal{X}}^{\text{supp}}$. This is because to define the derivative we need to take a limit of nearby points, but if $(\vec{w}, c) \in \mathcal{S}_{\mathcal{X}}^{\text{supp}}$, then it might not be true that $(\vec{w} + \vec{\epsilon}, c + \delta)$ is also in $\mathcal{S}_{\mathcal{X}}^{\text{supp}}$ for arbitrarily small vectors $\vec{\epsilon}$ and arbitrarily small numbers δ.

The first equation gives

$$\vec{w} = \sum_{\vec{x}\in\mathcal{X}} \alpha_{\vec{x}}\theta(\vec{x})\vec{x}, \tag{18.43}$$

which is the desired result, except that it has many unknown coefficients given by all of the Lagrange multipliers. The second equation gives

$$\sum_{\vec{x}\in\mathcal{X}} \alpha_{\vec{x}}\theta(\vec{x}) = 0, \tag{18.44}$$

which is a condition that the Lagrange multipliers have to satisfy. Plugging in these results back into the function g gives

$$\begin{aligned} g(\vec{w}, c) &= \frac{1}{2}\left\|\sum_{\vec{x}\in\mathcal{X}} \alpha_{\vec{x}}\theta(\vec{x})\vec{x}\right\|^2 - \sum_{\vec{x}\in\mathcal{X}} \alpha_{\vec{x}}\left(\theta(\vec{x})\left(\left\langle \sum_{\vec{y}\in\mathcal{X}} \alpha_{\vec{y}}\theta(\vec{y})\vec{y}, \vec{x}\right\rangle - c\right) - 1\right) \\ &= \frac{1}{2}\sum_{\vec{x},\vec{y}\in\mathcal{X}} \alpha_{\vec{x}}\alpha_{\vec{y}}\theta(\vec{x})\theta(\vec{y})\langle\vec{x},\vec{y}\rangle - \sum_{\vec{x},\vec{y}\in\mathcal{X}} \alpha_{\vec{x}}\alpha_{\vec{y}}\theta(\vec{x})\theta(\vec{y})\langle\vec{y},\vec{x}\rangle + c\sum_{\vec{x}\in\mathcal{X}} \alpha_{\vec{x}}\theta(\vec{x}) + \sum_{\vec{x}\in\mathcal{X}} \alpha_{\vec{x}} \\ &= \sum_{\vec{x}\in\mathcal{X}} \alpha_{\vec{x}} - \frac{1}{2}\sum_{\vec{x},\vec{y}\in\mathcal{X}} \alpha_{\vec{x}}\alpha_{\vec{y}}\theta(\vec{x})\theta(\vec{y})\langle\vec{x},\vec{y}\rangle \end{aligned} \tag{18.45}$$

Notice that although we have not yet solved the full problem, the maximizer only depends on the inner products between the vectors in the training data set. Setting (for the original function g)

$$\left.\frac{\partial g}{\partial \alpha_{\vec{x}}}\right|_{(\vec{w},c)} = 0 \tag{18.46}$$

for each $\vec{x} \in H_{\mathcal{X}}^{\text{supp}}$ will give additional conditions that the Lagrange multipliers have to satisfy. This equation then reads

$$\theta(\vec{x})\big(\langle\vec{w},\vec{x}\rangle - c\big) = 1 \tag{18.47}$$

for each $\vec{x} \in H_{\mathcal{X}}^{\text{supp}}$, and after plugging in the result for $\vec{w}$, this gives

$$\sum_{\vec{y}\in\mathcal{X}} \alpha_{\vec{y}}\theta(\vec{y})\langle\vec{y},\vec{x}\rangle - \theta(\vec{x}) = c \tag{18.48}$$

for each $\vec{x} \in H_{\mathcal{X}}^{\text{supp}}$. However, there is one subtle point, and that is that we do not know what $H_{\mathcal{X}}^{\text{supp}}$ is. Nevertheless, there is still an optimization procedure left over, and it is based on the different possible choices of $H_{\mathcal{X}}^{\text{supp}}$. For each choice of $H_{\mathcal{X}}^{\text{supp}}$, one has the linear system

$$\begin{aligned} \sum_{\vec{y}\in\mathcal{X}} \theta(\vec{y})\alpha_{\vec{y}} &= 0 \\ \sum_{\vec{y}\in\mathcal{X}} \theta(\vec{y})\langle\vec{y},\vec{x}\rangle\alpha_{\vec{y}} - c &= \theta(\vec{x}) \end{aligned} \tag{18.49}$$

in the variables $\{\alpha_{\vec{x}}\}_{\vec{x}} \in H_{\mathcal{X}}^{\text{supp}} \cup \{c\}$ obtained from equations (18.44) and (18.48). Notice that the second equation in this linear system is actually a set of $|H_{\mathcal{X}}^{\text{supp}}|$ equations. Therefore, this describes a linear system of $|H_{\mathcal{X}}^{\text{supp}}| + 1$ equations (+1 because of the first equation) in $|H_{\mathcal{X}}^{\text{supp}}| + 1$

variables (+1 because of the extra variable c). There are only a finite number of possible choices of $H_{\mathcal{X}}^{\text{supp}}$ and therefore only a finite number of linear systems one needs to solve. These systems are all consistent because we have assumed that the training data set can be separated. Hence, a solution to the SVM problem exists. ■

Some simple examples should help illustrate what could happen.

Problem 18.50. Find the SVM for the training data set given by

$$\mathcal{X}_- := \left\{ \vec{x}_- := \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\} \qquad \& \qquad \mathcal{X}_+ := \left\{ \vec{x}_+ := \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}. \tag{18.51}$$

Answer. In this case, there is only one positive and one negative vector. We expect the margin width to be 2 since this is the distance between the two points. Let us see that this works. The inner products are given by

$$\langle \vec{x}_-, \vec{x}_- \rangle = 1, \quad \langle \vec{x}_+, \vec{x}_+ \rangle = 1, \quad \langle \vec{x}_-, \vec{x}_+ \rangle = -1. \tag{18.52}$$

The associated linear system (18.49) is

$$\begin{aligned} \theta(\vec{x}_+)\alpha_{\vec{x}_+} + \theta(\vec{x}_-)\alpha_{\vec{x}_-} \qquad &= 0 \\ \theta(\vec{x}_+)\langle \vec{x}_+, \vec{x}_+ \rangle \alpha_{\vec{x}_+} + \theta(\vec{x}_-)\langle \vec{x}_-, \vec{x}_+ \rangle \alpha_{\vec{x}_-} - c &= \theta(\vec{x}_+) \\ \theta(\vec{x}_+)\langle \vec{x}_+, \vec{x}_- \rangle \alpha_{\vec{x}_+} + \theta(\vec{x}_-)\langle \vec{x}_-, \vec{x}_- \rangle \alpha_{\vec{x}_-} - c &= \theta(\vec{x}_-), \end{aligned} \tag{18.53}$$

which becomes

$$\begin{aligned} \alpha_{\vec{x}_+} - \alpha_{\vec{x}_-} \qquad &= 0 \\ \alpha_{\vec{x}_+} + \alpha_{\vec{x}_-} - c &= 1 \\ -\alpha_{\vec{x}_+} - \alpha_{\vec{x}_-} - c &= -1 \end{aligned} \tag{18.54}$$

after substitution. This linear system corresponds to the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & -1 & -1 \end{array}\right] \mapsto \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{array}\right] \tag{18.55}$$

so that the solution is

$$\alpha_{\vec{x}_+} = \frac{1}{2}, \quad \alpha_{\vec{x}_-} = \frac{1}{2}, \quad c = 0. \tag{18.56}$$

Plugging this into the equation for $\vec{w}$ (18.43) gives

$$\vec{w} = \alpha_{\vec{x}_+}\theta(\vec{x}_+)\vec{x}_+ + \alpha_{\vec{x}_-}\theta(\vec{x}_-)\vec{x}_- = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{18.57}$$

Therefore, the plane H is described as the set of vectors $\vec{x}$ such that $\langle \vec{w}, \vec{x} \rangle - c = 0$. Since $c = 0$, the set of solutions are all vectors $\vec{x}$ of the form

$$x \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{18.58}$$

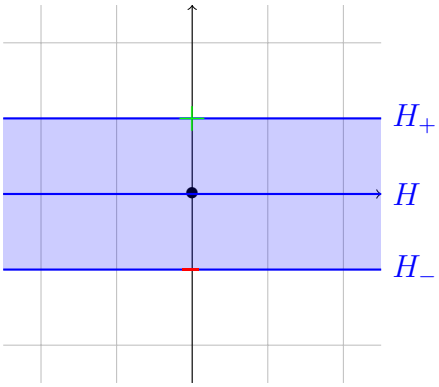
with $x \in \mathbb{R}$. The plane H_+ is the set of vectors $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ such that $\langle \vec{w}, \vec{x} \rangle - c = 1$. Since $c = 0$, this equation forces $y = 1$ but the x component is arbitrary, i.e. H_+ consists of all vectors of the form

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} + x \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{18.59}$$

with $x \in \mathbb{R}$. The plane H_- is the set of vectors $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ such that $\langle \vec{w}, \vec{x} \rangle - c = -1$. Since $c = 0$, this equation forces $y = -1$ but the x component is arbitrary, i.e. H_- consists of all vectors of the form

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix} + x \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{18.60}$$

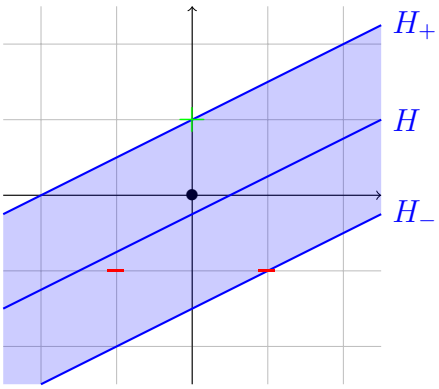
with $x \in \mathbb{R}$. Therefore, $\vec{w} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $c = 0$ indeed describes the following strip



Problem 18.61. Find the SVM for the training data set given by

$$\mathcal{X}_- := \left\{ \vec{x}_-^{\,1} := \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \vec{x}_-^{\,2} := \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\} \qquad \& \qquad \mathcal{X}_+ := \left\{ \vec{x}_+ := \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}. \tag{18.62}$$

Answer. If we solve for the SVM by including only one of the vectors from the negative training data set, then we expect to get a strip such as follows



and an analogous picture if we include only the other negative vector. Therefore, let us include all the points as support vectors. Their inner products are

$$\begin{aligned} \langle \vec{x}_-^1, \vec{x}_-^1 \rangle = 2, \quad \langle \vec{x}_-^1, \vec{x}_-^2 \rangle = 0, \quad \langle \vec{x}_-^1, \vec{x}_+ \rangle = -1, \\ \langle \vec{x}_-^2, \vec{x}_-^2 \rangle = 2, \quad \langle \vec{x}_-^2, \vec{x}_+ \rangle = -1, \quad \langle \vec{x}_+, \vec{x}_+ \rangle = 1. \end{aligned} \tag{18.63}$$

The associated linear system (18.49) is

$$\begin{aligned} \theta(\vec{x}_+)\alpha_{\vec{x}_+} \quad + \quad \theta(\vec{x}_-^1)\alpha_{\vec{x}_-^1} \quad + \quad \theta(\vec{x}_-^2)\alpha_{\vec{x}_-^2} \quad = 0 \\ \theta(\vec{x}_+)\langle \vec{x}_+, \vec{x}_+ \rangle \alpha_{\vec{x}_+} + \theta(\vec{x}_-^1)\langle \vec{x}_-^1, \vec{x}_+ \rangle \alpha_{\vec{x}_-^1} + \theta(\vec{x}_-^2)\langle \vec{x}_-^2, \vec{x}_+ \rangle \alpha_{\vec{x}_-^2} - c = \theta(\vec{x}_+) \\ \theta(\vec{x}_+)\langle \vec{x}_+, \vec{x}_-^1 \rangle \alpha_{\vec{x}_+} + \theta(\vec{x}_-^1)\langle \vec{x}_-^1, \vec{x}_-^1 \rangle \alpha_{\vec{x}_-^1} + \theta(\vec{x}_-^2)\langle \vec{x}_-^2, \vec{x}_-^1 \rangle \alpha_{\vec{x}_-^2} - c = \theta(\vec{x}_-^1) \\ \theta(\vec{x}_+)\langle \vec{x}_+, \vec{x}_-^2 \rangle \alpha_{\vec{x}_+} + \theta(\vec{x}_-^1)\langle \vec{x}_-^1, \vec{x}_-^2 \rangle \alpha_{\vec{x}_-^1} + \theta(\vec{x}_-^2)\langle \vec{x}_-^2, \vec{x}_-^2 \rangle \alpha_{\vec{x}_-^2} - c = \theta(\vec{x}_-^2) \end{aligned} \tag{18.64}$$

which becomes

$$\begin{aligned} \alpha_{\vec{x}_+} - \alpha_{\vec{x}_-^1} - \alpha_{\vec{x}_-^2} \quad = 0 \\ \alpha_{\vec{x}_+} + \alpha_{\vec{x}_-^1} + \alpha_{\vec{x}_-^2} - c = 1 \\ -\alpha_{\vec{x}_+} - 2\alpha_{\vec{x}_-^1} - 0\alpha_{\vec{x}_-^2} - c = -1 \\ -\alpha_{\vec{x}_+} - 0\alpha_{\vec{x}_-^1} - 2\alpha_{\vec{x}_-^2} - c = -1 \end{aligned} \tag{18.65}$$

after substitution. This linear system corresponds to the augmented matrix

$$\left[\begin{array}{cccc|c} 1 & -1 & -1 & 0 & 0 \\ 1 & 1 & 1 & -1 & 1 \\ -1 & -2 & 0 & -1 & -1 \\ -1 & 0 & -2 & -1 & -1 \end{array}\right] \mapsto \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 & 1/4 \\ 0 & 0 & 1 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] \tag{18.66}$$

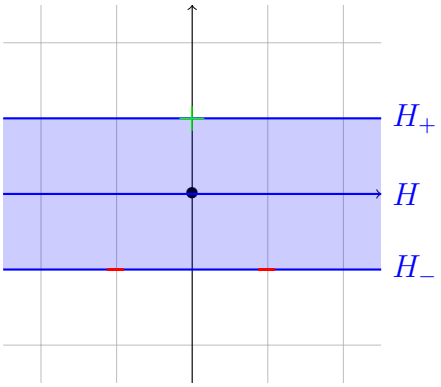
so that the solution is

$$\alpha_{\vec{x}_+} = \frac{1}{2}, \quad \alpha_{\vec{x}_-^1} = \frac{1}{4}, \quad \alpha_{\vec{x}_-^2} = \frac{1}{4}, \quad c = 0. \tag{18.67}$$

Plugging this into the equation for $\vec{w}$ (18.43) gives

$$\vec{w} = \alpha_{\vec{x}_+}\theta(\vec{x}_+)\vec{x}_+ + \alpha_{\vec{x}_-^1}\theta(\vec{x}_-^1)\vec{x}_-^1 + \alpha_{\vec{x}_-^2}\theta(\vec{x}_-^2)\vec{x}_-^2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{18.68}$$

which gives the following margin



The previous two examples assumed that all of the vectors given were actually support vectors. What if there are vectors in the training data set that are not support vectors? When this happens, we have to exclude them from the calculation. The difficulty with this is that it might not be clear apriori what the support vectors should be because we have not yet found the SVM. One then uses a method of exhaustion (trial and error, if you will). Because the training data set is finite, there are only a finite number of possibilities. However, as the training data set grows, the number of possibilities increases dramatically. One must then make educated guesses as to which combinations to try. The possibilities will usually be more transparent after drawing a visualization of the training data set.

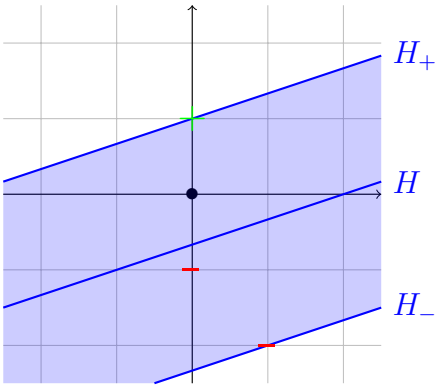
Problem 18.69. Find the SVM for the training data set given by

$$\mathcal{X}_- := \left\{ \vec{x}_-^1 := \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \vec{x}_-^2 := \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\} \qquad \& \qquad \mathcal{X}_+ := \left\{ \vec{x}_+ := \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}. \tag{18.70}$$

Answer. We will solve this problem by first showing what the solution is when $H_{\mathcal{X}}^{\text{supp}}$ is taken to be all of $\mathcal{X}$ and then what the solution is if $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_-^1, \vec{x}_+\}$ (there is still the other possibility of taking $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_-^2, \vec{x}_+\}$, but we will ignore this situation because an optimization for this would result in a non-separating solution). In either case, it is useful to have the inner products of these vectors handy:

$$\begin{aligned} &\langle \vec{x}_-^1, \vec{x}_-^1 \rangle = 1, \quad \langle \vec{x}_-^1, \vec{x}_-^2 \rangle = 2, \quad \langle \vec{x}_-^1, \vec{x}_+ \rangle = 0, \\ &\langle \vec{x}_-^2, \vec{x}_-^2 \rangle = 5, \quad \langle \vec{x}_-^2, \vec{x}_+ \rangle = -2, \quad \langle \vec{x}_+, \vec{x}_+ \rangle = 1. \end{aligned} \tag{18.71}$$

i. We can immediately throw out the case $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_-^2, \vec{x}_+\}$ because the resulting maximal margin would contain $\vec{x}_-^1$ as shown in the following figure



ii. If $H_{\mathcal{X}}^{\text{supp}} = \mathcal{X}$, we have the linear system

$$\begin{aligned} -\alpha_{\vec{x}_-^1} - \alpha_{\vec{x}_-^2} + \alpha_{\vec{x}_+} &= 0 \\ -\alpha_{\vec{x}_-^1} - 2\alpha_{\vec{x}_-^2} - \alpha_{\vec{x}_+} - c &= -1 \\ -2\alpha_{\vec{x}_-^1} - 5\alpha_{\vec{x}_-^2} - 2\alpha_{\vec{x}_+} - c &= -1 \\ \alpha_{\vec{x}_-^1} + 2\alpha_{\vec{x}_-^2} + \alpha_{\vec{x}_+} - c &= 1 \end{aligned} \tag{18.72}$$

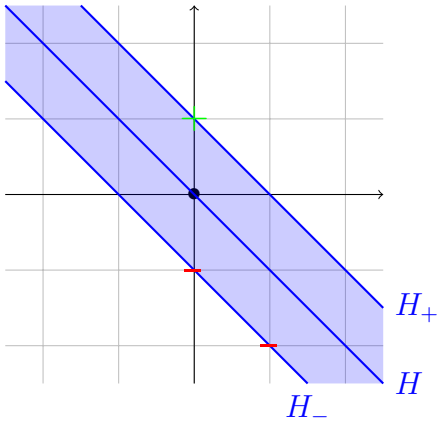
whose solution is

$$\alpha_{\vec{x}_-^1} = 2, \quad \alpha_{\vec{x}_-^2} = -1, \quad \alpha_{\vec{x}_+} = 1, \quad c = 0. \tag{18.73}$$

Therefore,

$$\vec{w} = \sum_{\vec{x}\in\mathcal{X}} \theta(\vec{x})\alpha_{\vec{x}}\vec{x} = -2\begin{bmatrix}0\\-1\end{bmatrix} + \begin{bmatrix}1\\-2\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix} \tag{18.74}$$

so that the margin width is $\sqrt{2}$. The resulting margin is depicted in the following figure.



iii. If $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_-^1, \vec{x}_+\}$, we have the linear system obtained from the first by removing all $\alpha_{\vec{x}_-^2}$ terms since $\alpha_{\vec{x}_-^2} = 0$, which is what the Lagrange multiplier must satisfy because $\vec{x}_-^2 \notin H_{\mathcal{X}}^{\text{supp}}$, i.e. $\vec{x}_-^2$ is not a support vector. We must also remove the equation obtained from $\frac{\partial g}{\partial \vec{x}_-^2} = 0$ since $\alpha_{\vec{x}_-^2} = 0$. The resulting linear system is

$$\begin{aligned} -\alpha_{\vec{x}_-^1} + \alpha_{\vec{x}_+} &= 0 \\ -\alpha_{\vec{x}_-^1} - \alpha_{\vec{x}_+} - c &= -1 \\ \alpha_{\vec{x}_-^1} + \alpha_{\vec{x}_+} - c &= 1 \end{aligned} \tag{18.75}$$

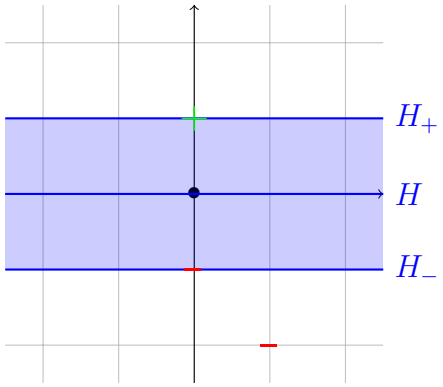
and its solution is

$$\alpha_{\vec{x}_-^1} = \frac{1}{2}, \quad \alpha_{\vec{x}_+} = \frac{1}{2}, \quad c = 0. \tag{18.76}$$

Therefore,

$$\vec{w} = \sum_{\vec{x}\in\mathcal{X}} \theta(\vec{x})\alpha_{\vec{x}}\vec{x} = -\frac{1}{2}\begin{bmatrix}0\\-1\end{bmatrix} + 0\begin{bmatrix}1\\-2\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix} \tag{18.77}$$

so that the margin width is 2. The resulting margin is depicted in the following figure.



From both of these solutions, we can read off the SVM by choosing the solution that has the largest margin, which is the second one.

Problem 18.78. Find the SVM for the training data set given by

$$\mathcal{X}_- := \left\{ \vec{x}_-^1 := \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \vec{x}_-^2 := \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\} \qquad \& \qquad \mathcal{X}_+ := \left\{ \vec{x}_+ := \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}. \tag{18.79}$$

Answer. The inner products are given by

$$\begin{aligned} &\langle \vec{x}_-^1, \vec{x}_-^1 \rangle = 5, \quad \langle \vec{x}_-^1, \vec{x}_-^2 \rangle = 1, \quad \langle \vec{x}_-^1, \vec{x}_+ \rangle = -2, \\ &\langle \vec{x}_-^2, \vec{x}_-^2 \rangle = 2, \quad \langle \vec{x}_-^2, \vec{x}_+ \rangle = -1, \quad \langle \vec{x}_+, \vec{x}_+ \rangle = 1. \end{aligned} \tag{18.80}$$

There are three cases to consider.

i. Assume $H_{\mathcal{X}}^{\text{supp}} = \left\{ \vec{x}_-^1, \vec{x}_+ \right\}$. The resulting linear system is

$$\begin{aligned} -\alpha_{\vec{x}_-^1} + \alpha_{\vec{x}_+} &= 0 \\ -5\alpha_{\vec{x}_-^1} - 2\alpha_{\vec{x}_+} - c &= -1 \\ 2\alpha_{\vec{x}_-^1} + \alpha_{\vec{x}_+} - c &= 1 \end{aligned} \tag{18.81}$$

and its solution is

$$\alpha_{\vec{x}_-^1} = \frac{1}{5}, \quad \alpha_{\vec{x}_+} = \frac{1}{5}, \quad c = -\frac{2}{5}. \tag{18.82}$$

Thus,

$$\vec{w} = -\frac{1}{5}\begin{bmatrix} -1 \\ -2 \end{bmatrix} + \frac{1}{5}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{18.83}$$

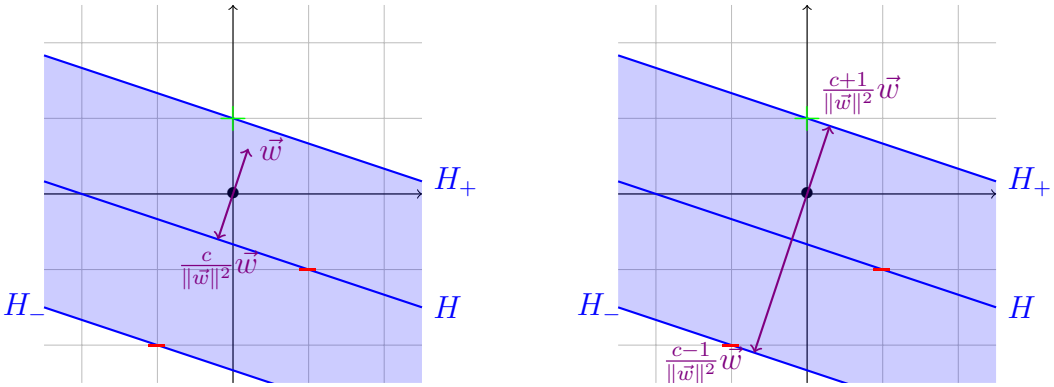
so that the margin width is

$$\frac{2}{\|\vec{w}\|} = \frac{2}{\left\| \frac{1}{5}\begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\|} = \frac{10}{\sqrt{10}} = \sqrt{10}. \tag{18.84}$$

The resulting margin along with $\vec{w}$ and

$$\frac{c}{\|\vec{w}\|^2}\vec{w} = \begin{pmatrix} -\frac{2}{5} \\ \frac{10}{25} \end{pmatrix} \left(\frac{1}{5}\begin{bmatrix} 1 \\ 3 \end{bmatrix} \right) = -\frac{1}{5}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = -\vec{w} \tag{18.85}$$

(which is a vector on the middle hyperplane H) are depicted in the following figure on the left



On the right, the two vectors

$$\frac{c+1}{\|\vec{w}\|^2}\vec{w} = \frac{3}{10}\begin{bmatrix}1\\3\end{bmatrix} \qquad \& \qquad \frac{c-1}{\|\vec{w}\|^2}\vec{w} = -\frac{7}{10}\begin{bmatrix}1\\3\end{bmatrix} \tag{18.86}$$

that are on the hyperplanes H_+ and H_- are drawn. The lines for these hyperplanes are obtained by solving the equations (the slope comes from the ratio of the y-component to the x-component of a vector orthogonal to $\vec{w}$)

$$\begin{aligned} y_+ &= -\frac{1}{3}x + b_+ \\ y &= -\frac{1}{3}x + b \\ y_- &= -\frac{1}{3}x + b_- \end{aligned} \tag{18.87}$$

by using the fact that these vectors are on these planes, i.e.

$$\begin{aligned} \left\langle \frac{c+1}{\|\vec{w}\|^2}\vec{w}, \vec{e}_2 \right\rangle &= -\frac{1}{3}\left\langle \frac{c+1}{\|\vec{w}\|^2}\vec{w}, \vec{e}_1 \right\rangle + b_+ \\ \left\langle \frac{c}{\|\vec{w}\|^2}\vec{w}, \vec{e}_2 \right\rangle &= -\frac{1}{3}\left\langle \frac{c}{\|\vec{w}\|^2}\vec{w}, \vec{e}_1 \right\rangle + b \\ \left\langle \frac{c-1}{\|\vec{w}\|^2}\vec{w}, \vec{e}_2 \right\rangle &= -\frac{1}{3}\left\langle \frac{c-1}{\|\vec{w}\|^2}\vec{w}, \vec{e}_1 \right\rangle + b_-, \end{aligned} \tag{18.88}$$

which reads

$$\begin{aligned} \frac{9}{10} &= -\frac{1}{3}\left(\frac{3}{10}\right) + b_+ \\ -\frac{3}{5} &= -\frac{1}{3}\left(-\frac{1}{5}\right) + b \\ -\frac{21}{10} &= -\frac{1}{3}\left(-\frac{7}{10}\right) + b_-, \end{aligned} \tag{18.89}$$

which gives the following equations for these lines

$$\begin{aligned} y_+ &= -\frac{1}{3}x + 1 \\ y &= -\frac{1}{3}x - \frac{2}{3} \\ y_- &= -\frac{1}{3}x - \frac{7}{3} \end{aligned} \tag{18.90}$$

This margin has a negative training data set in its interior so it cannot be an SVM because it is not described by a marginally separating hyperplane.

ii. Assume $H_{\mathcal{X}}^{\text{supp}} = \left\{\vec{x}_-^2, \vec{x}_+\right\}$. The resulting linear system is

$$\begin{aligned} -\alpha_{\vec{x}_-^2} + \alpha_{\vec{x}_+} &= 0 \\ -2\alpha_{\vec{x}_-^2} - \alpha_{\vec{x}_+} - c &= -1 \\ \alpha_{\vec{x}_-^2} + \alpha_{\vec{x}_+} - c &= 1 \end{aligned} \tag{18.91}$$

and its solution is

$$\alpha_{\vec{x}_-^2} = \frac{2}{5}, \quad \alpha_{\vec{x}_+} = \frac{2}{5}, \quad c = -\frac{1}{5}. \tag{18.92}$$

Thus,

$$\vec{w} = -\frac{2}{5}\begin{bmatrix}1\\-1\end{bmatrix} + \frac{2}{5}\begin{bmatrix}0\\1\end{bmatrix} = \frac{2}{5}\begin{bmatrix}-1\\2\end{bmatrix} \tag{18.93}$$

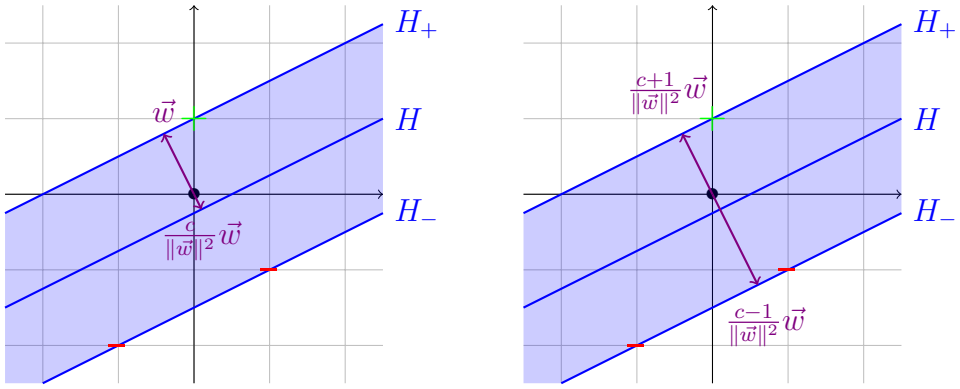
so that the margin width is $\sqrt{5}$. The other relevant quantities for obtaining the marginally separating hyperplane are

$$\frac{c-1}{\|\vec{w}\|^2}\vec{w} = \frac{3}{5}\begin{bmatrix}1\\-2\end{bmatrix}, \qquad \frac{c}{\|\vec{w}\|^2}\vec{w} = \frac{1}{10}\begin{bmatrix}1\\-2\end{bmatrix}, \qquad \& \qquad \frac{c+1}{\|\vec{w}\|^2}\vec{w} = \frac{2}{5}\begin{bmatrix}-1\\2\end{bmatrix} \tag{18.94}$$

Therefore, the lines describing the different hyperplanes are

$$\begin{aligned} y_+ &= \frac{1}{2}x + 1 \\ y &= \frac{1}{2}x - \frac{1}{4} \\ y_- &= \frac{1}{2}x - \frac{3}{2}. \end{aligned} \tag{18.95}$$

Hence, the resulting margin is given by



iii. Assume $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_-^1, \vec{x}_-^2, \vec{x}_+\}$. Since the previous case already contains $\vec{x}_-^1$ as a support vector, we already know the result will be the same. Hence, this is the SVM.

Exercise 18.96. Let $\mathcal{X}_+ = \{\vec{e}_1\}$, let $\mathcal{X}_- = \{\vec{e}_2\}$, and set $\mathcal{X} = \{\vec{e}_1, \vec{e}_2\}$.

(a) Sketch or describe $\mathcal{S}_{\mathcal{X}}$, the set of all marginally separating hyperplanes for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$. Note that $\mathcal{S}_{\mathcal{X}}$ must be a subset of $\left(\mathbb{R}^2 \setminus \{\vec{0}\}\right) \times \mathbb{R}$, which may be a bit challenging to draw.

(b) Sketch or describe $\mathcal{S}_{\mathcal{X}}^{\text{supp}}$, the set of all marginally separating hyperplanes for $\mathcal{X}$ for which their margin widths have been enlarged to include support vectors. Again, this should be a subset of $\left(\mathbb{R}^2 \setminus \{\vec{0}\}\right) \times \mathbb{R}$.

(c) Using the method employed in the preceding problems, find the SVM for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$.

(d) Draw the SVM in $\mathbb{R}^2$ together with $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$.

(e) What is the margin width of this SVM?

Recommended Exercises. Please check HuskyCT for the homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

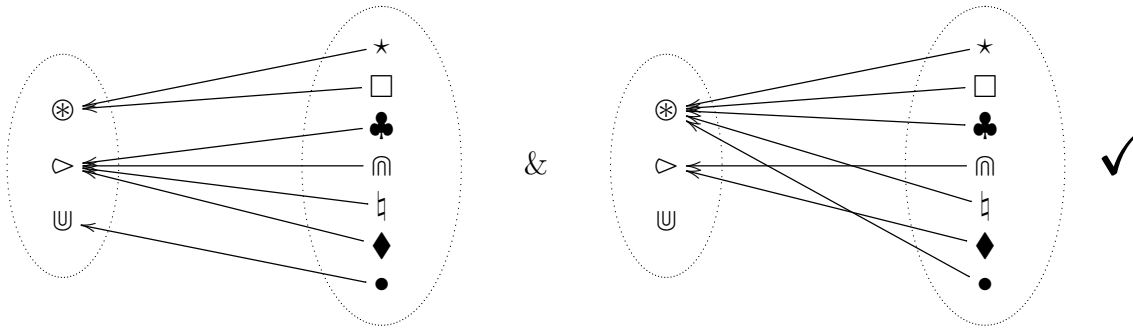
In this lecture, we covered Sections 6.5 and 6.6 in addition to several topics outside what is covered in [Lay].

19 Markov chains and complex networks*

Today we will cover some applications in the context of stochastic processes and Markov chains. To gain some motivation for this, we recall what a function is.

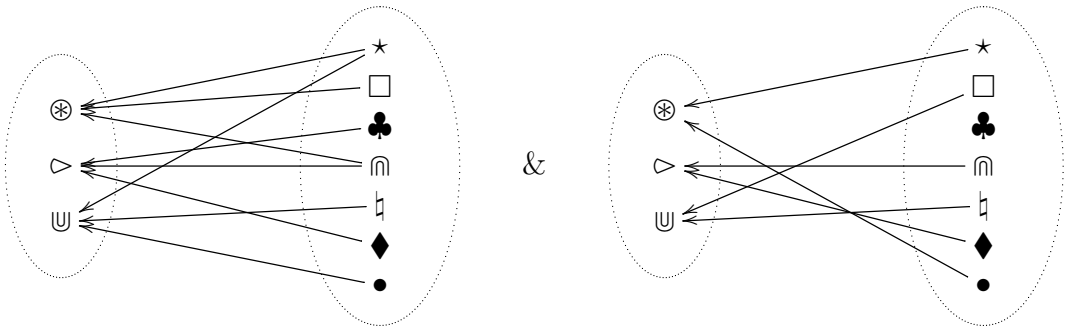
Definition 19.1. Let X and Y be two finite sets. A *function* f from X to Y written as $Y \xleftarrow{f} X$ is an assignment sending every x in X to a unique element, denoted by $f(x)$, in Y.

Example 19.2. The following illustrates two examples of a function



(19.3)

Example 19.4. The following two assignments are *not* functions.



(19.5)

The assignment on the left is not a function because, for instance, $\star$ gets assigned two entities, namely $\circledast$ and $\Cup$. The assignment on the right is not a function because, for instance, $\clubsuit$ is not assigned anything.

Today, we will think of the sets X, Y, and so on, as sets of *events* that could occur in a given situation. We will often denote the elements of X as a list $\{x_1, x_2, \ldots, x_n\}$. Thus, a function could be thought of as a deterministic process. What if instead of sending an element x in X to a *unique* element $f(x)$ in Y we instead *distributed* the element x *over* Y in some fashion? For this to be a reasonable definition, we would want the sum of the probabilities of the possible outcomes to be 1 so that *something* is always guaranteed to happen. But for this, we should talk about probability distributions.

Definition 19.6. A *probability distribution* on $X = \{x_1, x_2, \ldots, x_n\}$ is a function $\mathbb{R} \xleftarrow{p} X$ such that

$$p(x_i) \geq 0 \text{ for all } i \qquad \& \qquad \sum_{i=1}^{n} p(x_i) = 1. \tag{19.7}$$

Equivalently, such a probability distribution can be expressed as an n-component vector

$$\begin{bmatrix} p(x_1) \\ p(x_2) \\ \vdots \\ p(x_n) \end{bmatrix} \equiv \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} \tag{19.8}$$

again with the condition that each entry is at least 0 and the sum of all entries is equal to 1.

Exercise 19.9. Show that the set of all probability distributions on a finite set X is not a vector space. Is it a linear manifold? Is it a convex space?

Example 19.10. Let $X := \{H, T\}$, where H stands for "heads" and T stands for "tails." Let $\mathbb{R} \xleftarrow{p} X$ denote a "fair" coin toss, i.e.

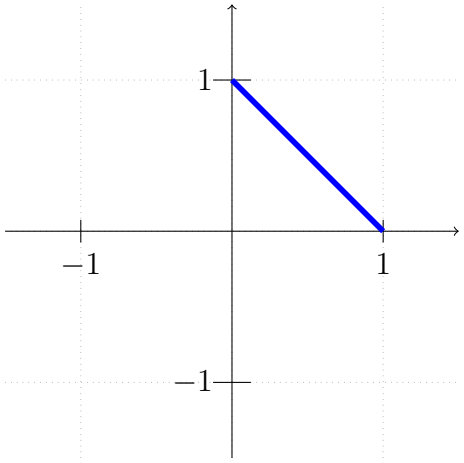
$$p(H) = \frac{1}{2} \qquad \& \qquad p(T) = \frac{1}{2}. \tag{19.11}$$

Then p is a probability distribution on X.

Example 19.12. Again, let $X := \{H, T\}$ be the set of events of a coin flip: either heads or tails. But this time, fix some weight r. r is some arbitrary number strictly between 0 and 1. Let $\mathbb{R} \xleftarrow{q_r} X$ be the probability distribution

$$q_r(H) = r \qquad \& \qquad q_r(T) = 1 - r. \tag{19.13}$$

Then q_r is a probability distribution on X. This is called an "unfair" coin toss if $r \neq \frac{1}{2}$. Thus, the set of all probability distributions looks like the following subset of $\mathbb{R}^2$.



Definition 19.14. Let X and Y be two finite sets. A *stochastic map/matrix* from X to Y is an assignment sending a probability distribution on X to a probability distribution on Y. Such a map is drawn as $T : X \rightsquigarrow Y$.

Let us parse out what this definition is saying. Write $X := \{x_1, x_2, \ldots, x_n\}$ and $Y := \{y_1, y_2, \ldots, y_m\}$. As we've already discussed, any probability distribution p on X can be expressed as a vector (19.8) and similarly on Y. Thus $T(p)$ is a probability distribution on Y, i.e. is some

vector (this time with m components). Is this starting to look familiar? T is an operation taking an n-component vector to an m-component vector. It almost sounds as if T is described by some matrix. Furthermore, we can look at the special probability distribution δ_{x_i} defined by

$$\delta_{x_i}(x_j) := \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{19.15}$$

(you may recognize this as the Kronecker-delta function). In other words, δ_{x_i} describes the probability distribution that says the event x_i will occur with 100% probability and no other event will occur. As a vector, this looks like

$$\delta_{x_i} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{-th entry} \tag{19.16}$$

Therefore, we might expect that the probability distribution $T(p)$ on Y is determined by the probability distributions δ_{x_i} since p itself can be written as a linear combination of these! Indeed, we have

$$p = \sum_{i=1}^{n} p(x_i)\delta_{x_i}, \tag{19.17}$$

or in vector form

$$\begin{bmatrix} p(x_1) \\ p(x_2) \\ \vdots \\ p(x_n) \end{bmatrix} = p(x_1) \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + p(x_2) \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + p(x_n) \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{19.18}$$

Furthermore, whatever T is, it has to send the Kronecker-delta probability distribution to *some* distribution on Y which is represented by an m-component vector

$$T(\delta_{x_i}) =: \begin{bmatrix} T_{1i} \\ T_{2i} \\ \vdots \\ T_{mi} \end{bmatrix}. \tag{19.19}$$

The meaning of this vector is as follows. Imagine that the event x_i takes place with 100% probability. Then the stochastic map says that after x_i occurs, there is a T_{1i} probability that the event y_1 will occur, a T_{2i} probability that the event y_2 will occur,..., and a T_{mi} probability that the event y_m will occur. This exactly describes the i-th column of a matrix. In other words, the stochastic

process is described by a matrix given by

$$T = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{m1} & T_{2m} & \cdots & T_{mn} \end{bmatrix} \tag{19.20}$$

where the i-th column represents physically the situation described in the past few sentences. Now let's go back to our initial probability distribution p on X. In this case, the event x_i takes places with probability $p(x_i)$ instead of 100%. Given this information, what is the probability of the event y_j taking place *after* the stochastic process? This would be obtained by taking the j-th entry of the resulting m-component vector from the matrix operation

$$\begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{m1} & T_{2m} & \cdots & T_{mn} \end{bmatrix} \begin{bmatrix} p(x_1) \\ p(x_2) \\ \vdots \\ p(x_n) \end{bmatrix} \tag{19.21}$$

In other words,

$$y_j = \sum_{i=1}^{n} T_{ji} p(x_i) \tag{19.22}$$

is the probability of the event y_j taking place given that the stochastic process T takes place and the initial probability distribution on X was given by p.

Example 19.23. Imagine a machine that flips a coin and is programmed to always obtain heads when given heads and always obtains tails when it is given tails. Unfortunately, machines are never perfect and there are always subtle changes in the environment that actually make the probability distribution slightly different. Oddly enough, the distribution for heads and tails was slightly different after performing the tests over and over again. Given heads, the machine is 88% percent likely to flip the coin and land heads again (leaving 12% for tails). Given tails, the machine is only 86% likely to flip the coin and land tails again (leaving 14% for heads). The matrix associated to this stochastic process is

$$T = \begin{bmatrix} 0.88 & 0.14 \\ 0.12 & 0.86 \end{bmatrix}. \tag{19.24}$$

Imagine I give the machine the coin heads up at first. After how many flips will the probability of seeing heads be less than 65%? After one flip, the probability of seeing heads is

$$\begin{bmatrix} 0.88 & 0.14 \\ 0.12 & 0.86 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.88 \\ 0.12 \end{bmatrix} \tag{19.25}$$

as we could have guessed. After another turn, it becomes (after rounding and suppressing the higher order terms)

$$\begin{bmatrix} 0.88 & 0.14 \\ 0.12 & 0.86 \end{bmatrix} \begin{bmatrix} 0.88 \\ 0.12 \end{bmatrix} = \begin{bmatrix} 0.79 \\ 0.21 \end{bmatrix} \tag{19.26}$$

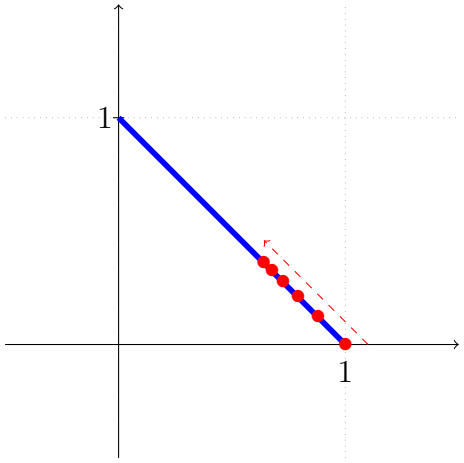
and so on

$$\begin{bmatrix}0.68\\0.32\end{bmatrix} \xleftarrow{T} \begin{bmatrix}0.73\\0.27\end{bmatrix} \xleftarrow{T} \begin{bmatrix}0.79\\0.21\end{bmatrix} \tag{19.27}$$

until after 5 turns we finally get

$$\begin{bmatrix}0.88 & 0.14\\0.12 & 0.86\end{bmatrix}^5 \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0.64\\0.36\end{bmatrix}. \tag{19.28}$$

If we draw these points on the space of probability distributions, they look as follows



which by the way makes it look like they are converging. We will get back to this soon.

Definition 19.29. Given a set X, a stochastic process $X \rightsquigarrow X : T$ from X to itself, and a probability distribution p on X, a *Markov chain* is the sequence of probability vectors

$$\big(p, T(p), T^2(p), T^3(p), \dots\big). \tag{19.30}$$

Example 19.31. In the previous example, what happens if we keep iterating the stochastic map? Does the resulting distribution eventually converge to some probability distribution on X? And if it does converge to some probability distribution q, does that probability remain "steady"? In other words, can we find a vector q such that $Tq = q$? Could there be *more* than one such "steady" probability distribution? Let's first try to find such a vector before answering all of these questions. We want to solve the equation

$$\begin{bmatrix}0.88 & 0.14\\0.12 & 0.86\end{bmatrix} \begin{bmatrix}q\\1-q\end{bmatrix} = \begin{bmatrix}q\\1-q\end{bmatrix} \tag{19.32}$$

Working out the left-hand-side gives the two equations

$$\begin{aligned} 0.88q + 0.14(1-q) &= q \\ 0.12q + 0.86(1-q) &= 1-q \end{aligned} \tag{19.33}$$

This is a bit scary: two equations and one unknown! But maybe we can still solve it... The first equation gives the solution

$$q = \frac{0.14}{0.26} \approx 0.54. \tag{19.34}$$

Fortunately, the second equation gives the same exact solution! What this is saying is that if I was 54% sure that I gave the machine a coin with heads up, then the probability of the outcome would be 54% heads *every single time*!

Definition 19.35. Let X be a set and T a stochastic process on X. A *steady state probability distribution for X and T* is a probability distribution p on X such that $T(p) = p$.

A more clever way to solve for steady state probability distributions is to rewrite the equation $T(p) = p$ as $(T - \mathbb{1})(p) = 0$, where $\mathbb{1}$ is the stochastic process that *does nothing* (in other words, it leaves every single probability distribution alone). Since $T - \mathbb{1}$ can be represented as a matrix and p as a vector, this amounts to solving a homogeneous system, which you are quite familiar with by now.

Go through Example 2 in Section 10.2 in [2].

Problem 19.36. If $S : X \rightsquigarrow Y$ is a stochastic map, what is the meaning of S^T, the transpose of the stochastic map?

Answer. If we write out the elements of X and Y as $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$, then S has the matrix form

$$S = \begin{bmatrix} | & & | \\ S\vec{e}_1 & \cdots & S\vec{e}_n \\ | & & | \end{bmatrix}. \tag{19.37}$$

It's helpful to write out the components of S explicitly

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix}. \tag{19.38}$$

Note that $S\vec{e}_k$, the k-th *column* of S, is the probability distribution associated to the stochastic map with a definitive starting value of x_k. In other words, it describes all possible outputs given the input x_k with their corresponding probabilities. The k-th *row* of S describes all ways of achieving the output y_k from all possible inputs with their corresponding probabilities. Notice that the sum of the entries in each row of S do not have to add up to 1. For example, if S gave the same output no matter what input was given, then it would look like a matrix of all 0's except for one row consisting of all 1's. So the transpose of S is in general not a stochastic matrix. Nevertheless, we still have an interpretation of the rows of S, which are the columns of S^T. Therefore, S^T assigns to each y_k the possible elements in X that could have lead to y_k as being the output of S together with the corresponding probability that that specific element in X lead to y_k. Stochastic matrices S for which S^T is also a stochastic matrix are called *doubly stochastic matrices*.

We now come to answering the many questions we had raised earlier.

Definition 19.39. Let X be a finite set. A T stochastic map on X is said to be *regular* if there exists a positive integer k such that the matrix associated to T^k has entries $(T^k)_{ij}$ satisfying $0 < (T^k)_{ij} < 1$ for all i and j.

Theorem 19.40. *Let X be a finite set and T a regular stochastic map on X. Then, there exists a unique probability distribution q on X such that $T(q) = q$. Furthermore, for any other probability distribution p, the sequence*

$$\big(p, T(p), T^2(p), T^3(p), T^4(p), \dots\big) \tag{19.41}$$

converges to q. In fact, when the probability distribution q is written as a vector $\vec{q}$ and T as a stochastic matrix

$$\lim_{n\to\infty} T^n = \begin{bmatrix} | & & | \\ \vec{q} & \cdots & \vec{q} \\ | & & | \end{bmatrix}. \tag{19.42}$$

This limit is meant to be interpreted pointwise, i.e. the limit of each of the entries.

We already saw this in the example above. We found a unique solution to the coin toss scenario and we also observed how our initial configuration tended towards the steady state solution. If T is not regular, the sequence might not converge to a steady state solution. Markov chains appear in several other contexts. For example, Google prioritizes search results based on stochastic matrices. The internet can be viewed as a directed graph where webpages are represented as vertices and a directed edge from one vertex to another means that the source webpage has a hyperlink to the target webpage.

Definition 19.43. A *directed graph* consists of a set $\mathcal{V}$, a set $\mathcal{E}$, and two functions $s, t : \mathcal{E} \to \mathcal{V}$ such that

(a) $s(e) \neq t(e)$ for all $e \in \mathcal{E}$,

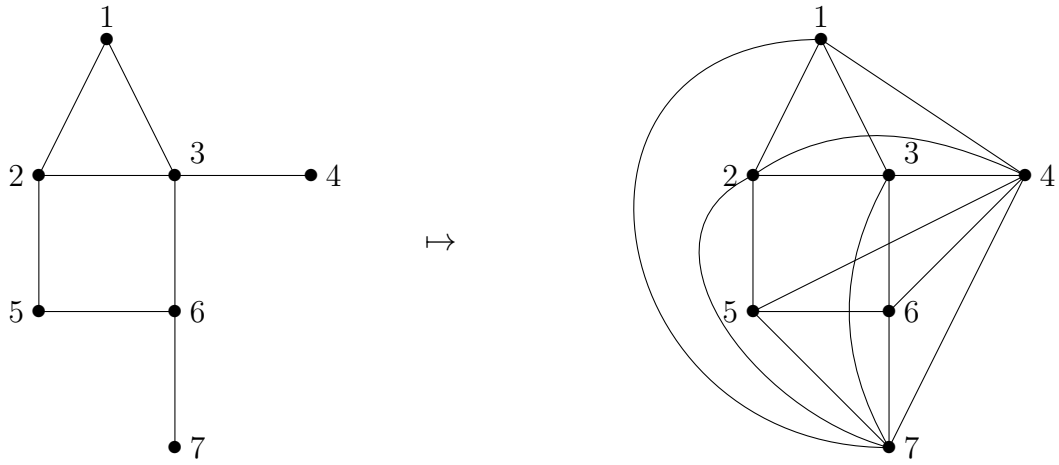
(b) if $s(e) = s(e')$ and $t(e) = t(e')$, then $e = e'$,

(c) for each $v \in \mathcal{V}$, there exists an $e \in \mathcal{E}$ such that either $s(e) = v$ or $t(e) = v$.

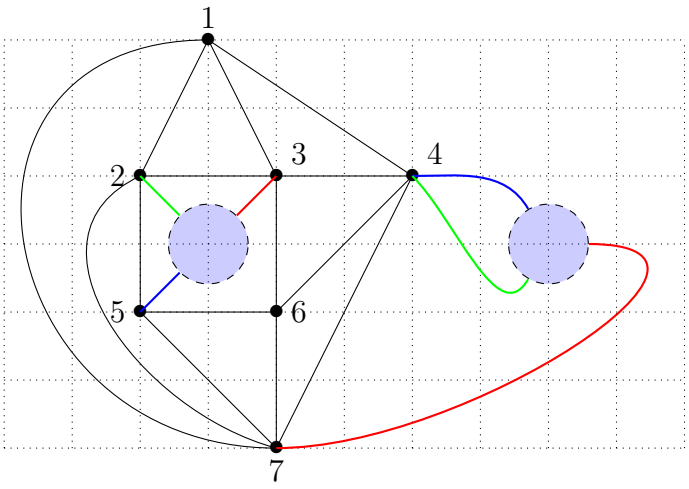
This definition is interpreted in the following way. The elements of $\mathcal{V}$ are called *vertices* (also called nodes) and the elements of $\mathcal{E}$ are called *directed edges*. The functions s and t are interpreted as the source and target of each directed edge, respectively. The first condition guarantees that there are no loops. In terms of the internet example, this means that there is no webpage that hyperlinks to itself (of course, some webpages do this, but we will not consider such cases). The second condition guarantees that there is at most one directed edge from one vertex to another. In terms of the internet example, this means that a webpage has at most one hyperlink to another webpage. The third condition guarantees that there are no isolated vertices. In terms of the internet example, this means that each webpage is connected to some other webpage either by having a hyperlink to another webpage or by being the hyperlink of another webpage. Note that we do allow directed edges to go in both direction between two vertices. This means that we allow the situation that a webpage A hyperlinks to B and B hyperlinks back to A.

Go through PageRank and the Google Matrix on page 19 in Section 10.2 of [2]. The main idea is that a surfer clicks a hyperlink with a uniform distribution. With this information, a stochastic matrix can be obtained. This stochastic matrix is not regular and two adjustments need to be made. The first adjustment has a wonderful geometric interpretation which will be explained below. The second adjustment is a convex combination with a uniform distribution allowing for the possibility of a surfer selecting a website at random regardless of whether or not a hyperlink exists on that webpage. These two modifications construct a regular stochastic matrix so that Theorem 19.40 holds.

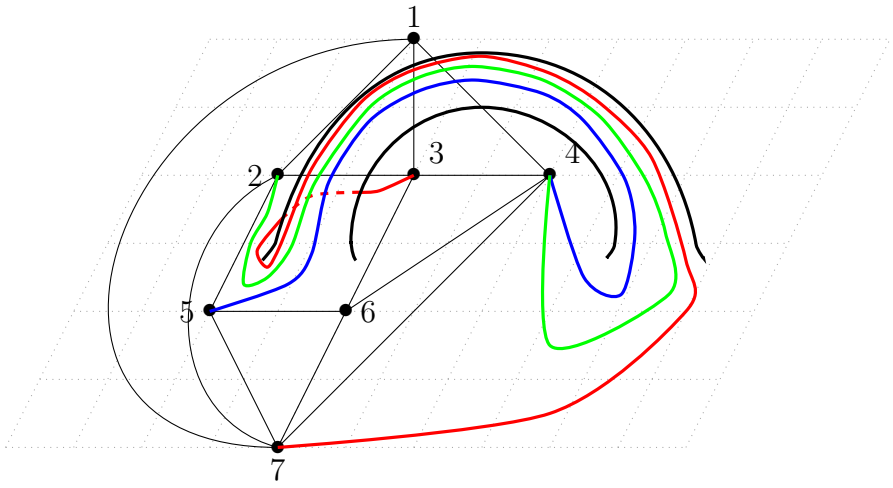
Note that adjustment 1 in Lay's book may change the topology of the graph in the sense that one cannot draw it on the plane without intersections. Naively drawing the adjusted graph results in



As you can see, there are three edges that cannot be draw without intersecting any other edge. If we could somehow cut out two holes in the plane and somehow glue the outer circles together, we could "tunnel" from the outside to the inside of the graph and connected the edges so that they do not intersect.



We'll go to three dimensions to see why the three different edges do not intersect each other.



If we view our original planar graph on a two-dimensional sphere (which we can always do since the graph is compact), cutting out two such holes and gluing the boundary circles together means that the adjusted graph actually lives on a torus (see Figure 12). Notice how none of the edges

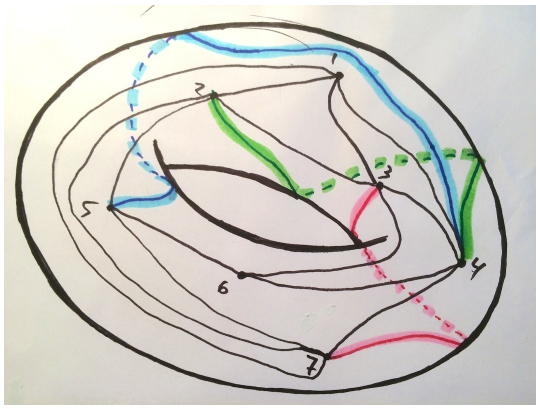


Figure 12: Embedding the graph on a torus

intersect each other now.

Such graphs are part of the more general area of study known as complex networks. Since the internet is a vastly larger network than the examples we illustrated above, computing the steady state vectors, and therefore obtaining the ranking of website importance, is a challenging task. Imagine trying to do this for such a large network (see Figure 13).[55]

There are methods to reduce such big data problems to more manageable ones, but this necessarily involves some approximations. Such methods are discussed in [3]. Figuring out the topology is a great help in obtaining certain features of the network. Unfortunately, the kind of topology we have discussed above is rarely touched on in a first course in topology, unless it is towards the end of the course. Such material is more often deferred to a course on *algebraic* topology or graph theory. If you'd like to get a good taste of topology for beginners, I recommend the book *The Shape of Space* by Weeks [6].

[55]This figure was obtained from Grandjean, Martin, "Introduction à la visualisation de données, l'analyse de réseau en histoire", Geschichte und Informatik 18/19, pp. 109128, 2015.

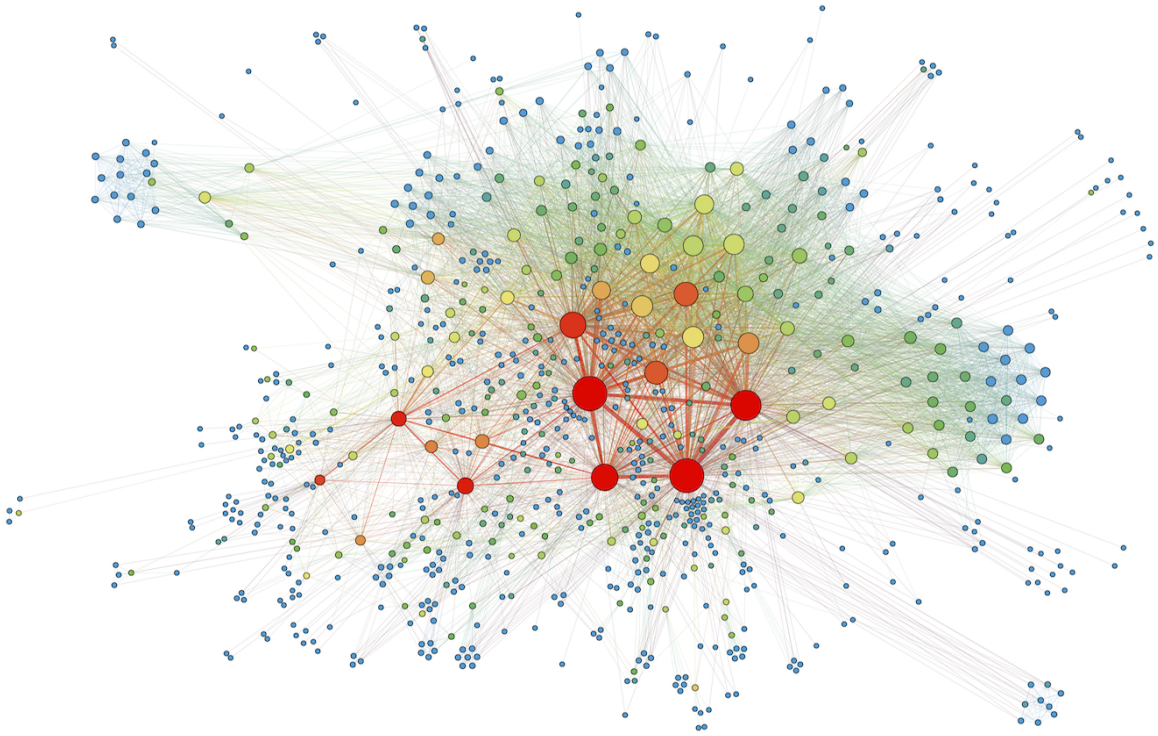


Figure 13: A complex network, similar to the one described by webpages and hyperlinks. The larger, warmer color, nodes depict higher importance.

Recommended Exercises. Exercises 4 and 18 in Section 4.9 of [Lay]. Exercises 8, 15, 23, and 24 in Section 10.1 of [2]. Exercises 3, 13, 27, 28, 34, and 35 in Section 10.2 of [2]. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we covered parts of Sections 4.9, 10.1, 10.2, and my own personal notes.

20 Eigenvalues and eigenvectors

The steady state vectors from the lecture on Markov chains are special cases of what are called *eigenvectors* with *eigenvalue* 1.

Definition 20.1. Let $\mathbb{R}^n \xleftarrow{T} \mathbb{R}^n$ be a linear transformation. An *eigenvector for T* is a non-zero vector $\vec{v} \in \mathbb{R}^n$ for which $T(\vec{v}) \propto \vec{v}$ (read $T(\vec{v})$ is proportional to $\vec{v}$), i.e. $T(\vec{v}) = \lambda\vec{v}$ for some scalar $\lambda \in \mathbb{R}$. The proportionality constant λ in the expression $T(\vec{v}) = \lambda\vec{v}$ is called the *eigenvalue* of the eigenvector $\vec{v}$.

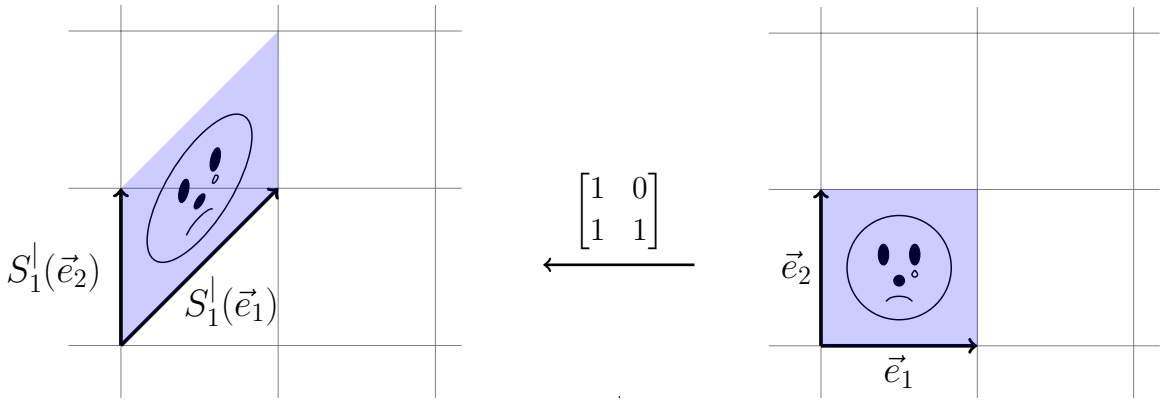
Equivalently, $\vec{v}$ is an eigenvector for T iff

$$\operatorname{span}\{T\vec{v}\} \subseteq \operatorname{span}\{\vec{v}\}. \tag{20.2}$$

Please note that although eigenvectors are assumed to be non-zero, eigen*values* can certainly be zero. For example, take the matrix $\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$ which has eigenvalues 1 and 0 with corresponding eigenvectors $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$, respectively. This is why the condition is *not* $\operatorname{span}\{T\vec{v}\} = \operatorname{span}\{\vec{v}\}$.

Besides the example studied in the previous lecture associated with Markov chains and stochastic processes, we have several other examples.

Example 20.3. Consider the vertical shear transformation in $\mathbb{R}^2$ given by



Visually, it is clear that the vector $\vec{e}_2$ is an eigenvector of eigenvalue 1. Let us check to see if this is true and if this is the only eigenvector for the vertical shear. The system we wish to solve is

$$\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \lambda\begin{bmatrix}x\\y\end{bmatrix} \tag{20.4}$$

for all possible values of x and y as well as λ. Following a similar procedure to what we did last class, we subtract

$$\lambda\begin{bmatrix}x\\y\end{bmatrix} = \lambda\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} \tag{20.5}$$

from both sides

$$\left(\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix} - \lambda\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}, \tag{20.6}$$

which becomes the homogeneous system

$$\begin{bmatrix} 1-\lambda & 0 \\ 1 & 1-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{20.7}$$

Notice that we are trying to find a *nontrivial* solution to this system. Rather than trying to manipulate these equations algebraically and solving these systems in a case by case basis, let us analyze this system from the linear algebra perspective. Finding an eigenvector for the vertical shear matrix amounts to finding a nontrivial solution to the system described by equation (20.7), which means that the kernel of the matrix $\begin{bmatrix} 1-\lambda & 0 \\ 1 & 1-\lambda \end{bmatrix}$ must be nonzero, which, by the Invertible Matrix Theorem, means the determinant of this matrix must be zero, i.e.

$$\det \begin{bmatrix} 1-\lambda & 0 \\ 1 & 1-\lambda \end{bmatrix} = 0. \tag{20.8}$$

Solving this, we arrive at the polynomial equation

$$(1-\lambda)^2 = 0. \tag{20.9}$$

The only root of this polynomial is $\lambda = 1$ (in fact, $\lambda = 1$ appears twice, which means it has *multiplicity* 2—more on this soon). Knowing this information, we can then solve the system (20.7) much more easily since the equation reduces to

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \tag{20.10}$$

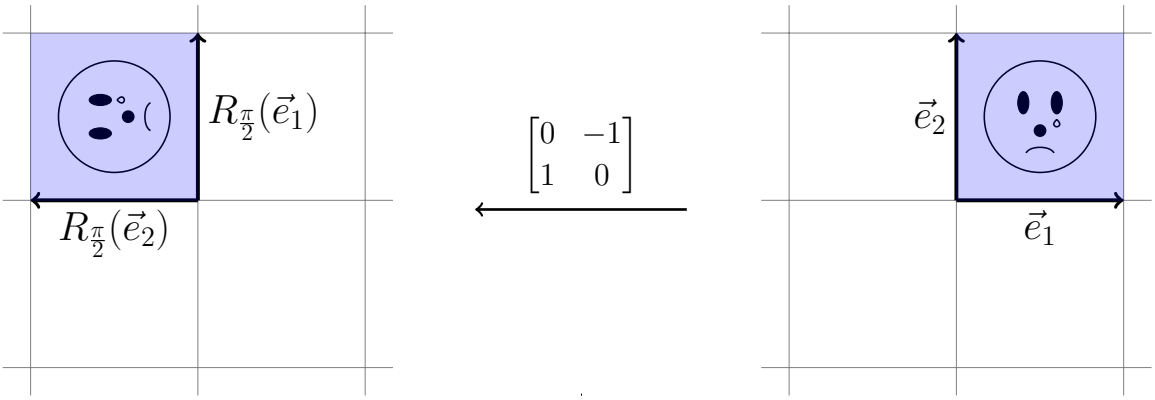
and the set of solutions to this system is

$$\left\{ t \begin{bmatrix} 0 \\ 1 \end{bmatrix} \ : \ t \in \mathbb{R} \right\}, \tag{20.11}$$

where t is a free variable. Therefore, all of the vectors of the form (20.11) are eigenvectors with eigenvalue 1.

Example 20.12. Let $\mathbb{R}^n \xleftarrow{P} \mathbb{R}^n$ be a projection, i.e. $P^2 = P$. If λ is an eigenvalue of P, then $\lambda \in \{0, 1\}$. To see this, suppose that $\vec{v}$ is an eigenvector of P. Then $P\vec{v} = \lambda\vec{v}$. Applying P once more, we obtain $P^2\vec{v} = \lambda P\vec{v} = \lambda^2\vec{v}$. But $P^2 = P$ also implies $P^2\vec{v} = P\vec{v} = \lambda\vec{v}$. Putting these two equations together gives $\lambda^2\vec{v} = \lambda\vec{v}$, which we can reorganize as $\lambda(1-\lambda)\vec{v} = \vec{0}$. Since $\vec{v}$ is an eigenvector, it is non-zero. This implies $\lambda(1-\lambda) = 0$. This proves that $\lambda \in \{0, 1\}$. This says that the eigenvalues of any projection can only be 0 or 1.

Example 20.13. Consider the rotation by angle $\frac{\pi}{2}$



Following a similar procedure to the previous example, namely solving

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix} \tag{20.14}$$

for all possible values of x and y as well as λ, we obtain

$$\begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{20.15}$$

Again, we want to find eigenvectors and eigenvalues for this system, and this means the matrix $\begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix}$ must be non-invertible so that

$$\det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = 0, \tag{20.16}$$

but the determinant is given by

$$\lambda^2 + 1 = 0. \tag{20.17}$$

This polynomial has no real root. The only roots are $\lambda = \pm\sqrt{-1}$. Therefore, there are no eigenvectors with real eigenvalues for the rotation matrix. This is plausible because if you rotate in the plane, nothing except the zero vector is fixed, agreeing with our intuition.

The previous example suggests that we can always find eigenvalues for a real matrix, but we would have to allow them to be complex. The more precise statement of this fact is provided in Theorem 20.70.

Example 20.18. As we saw in Example 20.13, the characteristic polynomial associated to the rotation by angle $\frac{\pi}{2}$ matrix given by

$$R_{\frac{\pi}{2}} := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{20.19}$$

is

$$\lambda^2 + 1 = 0. \tag{20.20}$$

Using just real numbers, no such solution exists. However, using complex numbers, we know exactly what λ should be. The possible choices are

$$\lambda_1 = \sqrt{-1} \qquad \& \qquad \lambda_2 = -\sqrt{-1} \tag{20.21}$$

since both satisfy $\lambda^2 = -1$. What are the corresponding eigenvectors? As usual, we solve a homogeneous problem for each eigenvalue. For λ_1, we have

$$\left[\begin{array}{cc|c} -\sqrt{-1} & -1 & 0 \\ 1 & -\sqrt{-1} & 0 \end{array}\right] \to \left[\begin{array}{cc|c} -1 & \sqrt{-1} & 0 \\ 1 & -\sqrt{-1} & 0 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & -\sqrt{-1} & 0 \\ 0 & 0 & 0 \end{array}\right] \tag{20.22}$$

which has solutions of the form

$$t \begin{bmatrix} -\sqrt{-1} \\ 1 \end{bmatrix} \tag{20.23}$$

with t a free variable. Hence one such eigenvector for λ_1 is

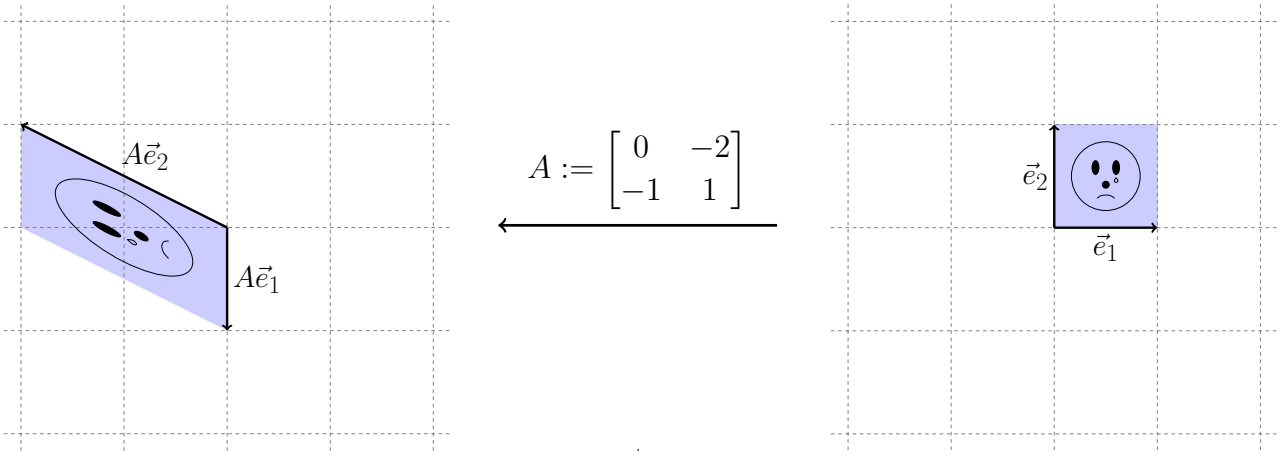
$$\vec{v}_1 = \begin{bmatrix} 1 \\ \sqrt{-1} \end{bmatrix} \tag{20.24}$$

(I multiplied throughout by $\sqrt{-1}$ so that the first entry is a 1). Similarly, for λ_2 one finds that one such eigenvector is

$$\vec{v}_2 = \begin{bmatrix} 1 \\ -\sqrt{-1} \end{bmatrix}. \tag{20.25}$$

Hence, the rotation matrix *does* have eigenvalues and eigenvectors—we just can't *see* them! Complex numbers are briefly reviewed at the end of this lesson. When we discuss ordinary differential equations, we will prove a physical interpretation for complex eigenvalues (briefly, they describe oscillations).

Example 20.26. Consider the transformation given by



At a first glance, it does not look like there are any eigenvectors for this transformation, but this is misleading. The possible eigenvalues are obtained by solving the quadratic polynomial

$$\det \begin{bmatrix} -\lambda & -2 \\ -1 & 1-\lambda \end{bmatrix} = -\lambda(1-\lambda) - 2 = 0 \iff \lambda^2 - \lambda - 2 = 0. \tag{20.27}$$

The roots of this quadratic polynomial are given in terms of the quadratic formula

$$\lambda = \frac{-(-1) \pm \sqrt{(-1)^2 - 4(1)(-2)}}{2(1)} = \frac{1}{2} \pm \frac{3}{2}, \tag{20.28}$$

which has two solutions. The eigenvalues are therefore

$$\lambda_1 = -1 \qquad \& \qquad \lambda_2 = 2. \tag{20.29}$$

Associated to the first eigenvalue, we have the linear system

$$\begin{bmatrix} -\lambda_1 & -2 \\ -1 & 1-\lambda_1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{20.30}$$

The solutions of this system are all scalar multiples of the vector

$$\vec{v}_1 := \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \tag{20.31}$$

Therefore, $\vec{v}_1$ is an eigenvector of $\begin{bmatrix} 0 & -2 \\ -1 & 1 \end{bmatrix}$ with eigenvalue -1 because

$$\begin{bmatrix} 0 & -2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = -1 \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \tag{20.32}$$

Similarly, associated to the second eigenvalue, we have the linear sytem

$$\begin{bmatrix} -\lambda_2 & -2 \\ -1 & 1-\lambda_2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} -2 & -2 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{20.33}$$

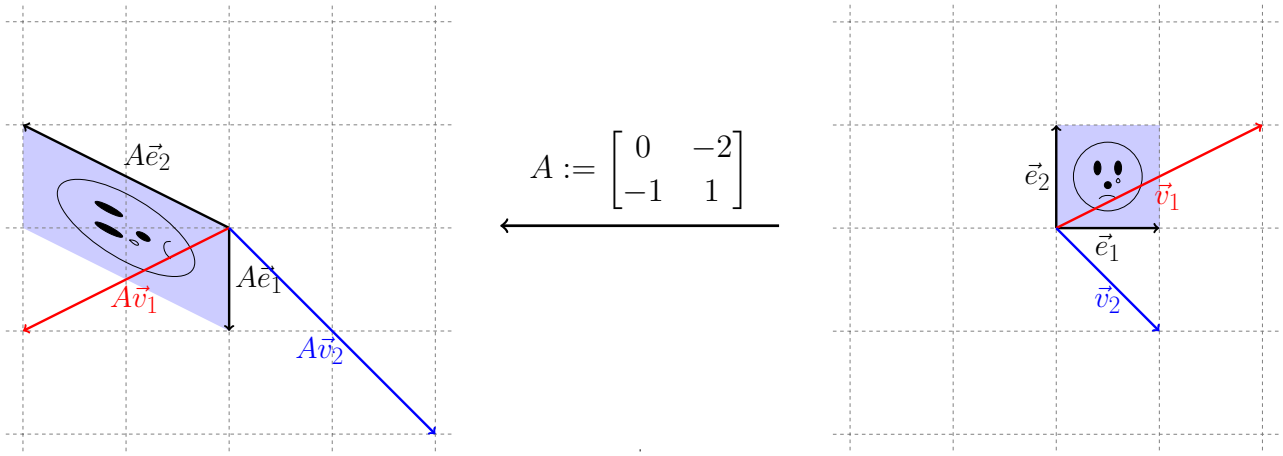
whose solutions are all scalar multiples of the vector

$$\vec{v}_2 := \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{20.34}$$

Again, this means that $\vec{v}_2$ is an eigenvector of $\begin{bmatrix} 0 & -2 \\ -1 & 1 \end{bmatrix}$ with eigenvalue 2 because

$$\begin{bmatrix} 0 & -2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{20.35}$$

Visually, under the transformation, these eigenvectors stay along the same line where they started.



I highly recommend checking out 3Blue1Brown's video https://www.youtube.com/watch?v=PFDu9oVAE-g on eigenvectors and eigenvalues for geometric animations describing what eigenvectors are. The previous examples indicate the following algebraic fact relating eigenvalues to determinants.

Theorem 20.36. *Let A be an $n \times n$ matrix. The eigenvalues for the linear transformation associated to A are given by the solutions to the polynomial equation*

$$\det(A - \lambda \mathbb{1}_n) = 0 \tag{20.37}$$

in the variable λ.

Proof. Let $\vec{v}$ be an eigenvector of A with eigenvalue λ. Then $A\vec{v} = \lambda\vec{v}$. Therefore, $(A - \lambda\mathbb{1})\vec{v} = \vec{0}$, i.e. $\vec{v} \in \ker(A - \lambda\mathbb{1})$. Since $\vec{v}$ is a non-zero vector, $A - \lambda\mathbb{1}$ is not invertible. Hence, $\det(A - \lambda\mathbb{1}) = 0$.

Conversely, suppose that λ satisfies $\det(A - \lambda\mathbb{1}) = 0$. Then $A - \lambda\mathbb{1}$ is not invertible. Since $A - \lambda\mathbb{1}$ is an $n \times n$ matrix, $A - \lambda\mathbb{1}$ is not injective, i.e. its nullspace is at least 1-dimensional, i.e. $(A - \lambda\mathbb{1})\vec{v} = \vec{0}$ has a nontrivial solution, i.e. there exists a nonzero vector $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$. ∎

Definition 20.38. For an $n \times n$ matrix A, the resulting polynomial $\det(A - \lambda\mathbb{1}_n)$ is called the *characteristic polynomial* of A. If an eigenvalue appears k times as a root, then k is called the *multiplicity* of that eigenvalue. The span of the eigenvectors of A for a particular eigenvalue λ is called the *eigenspace* associated to A and λ.

One consequence of the above theorem is the following.

Corollary 20.39. *λ is an eigenvalue of an $n \times n$ matrix A if and only if λ is an eigenvalue of A^T, the transpose of A.*

Proof. Since the determinant of a matrix is equal to the determinant of its transpose,

$$\begin{array}{ccc} \det(A^T - \lambda\mathbb{1}_n^T) & & \det\left((A - \lambda\mathbb{1}_n)^T\right) \\ \| & & \| \\ \det(A^T - \lambda\mathbb{1}_n) & & \det(A - \lambda\mathbb{1}_n) \end{array} \tag{20.40}$$

which, by Theorem 20.36 shows that the eigenvalues of A^T and A are the same. ∎

Another immediate consequence of the Theorem 20.36 is the following.

Theorem 20.41. *Let A be an upper-triangular $n \times n$ matrix of the form*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{nn} \end{bmatrix} \tag{20.42}$$

Then the eigenvalues of A are given by the roots of the polynomial

$$(a_{11}-\lambda)(a_{22}-\lambda)(a_{33}-\lambda)\cdots(a_{nn}-\lambda)=0 \tag{20.43}$$

in the variable λ.

Proof. Since

$$A-\lambda\mathbb{1}=\begin{bmatrix} a_{11}-\lambda & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22}-\lambda & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33}-\lambda & \cdots & a_{3n} \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{nn}-\lambda \end{bmatrix} \tag{20.44}$$

is an upper triangular matrix, its determinant is the product of the diagonal elements, i.e.

$$\det(A-\lambda\mathbb{1}_n)=(a_{11}-\lambda)(a_{22}-\lambda)(a_{33}-\lambda)\cdots(a_{nn}-\lambda). \tag{20.45}$$

■

Unfortunately, this theorem only tells us what the eigenvalues of an upper triangular matrix are. It is not so easy to write down a set of eigenvectors (if they even exist). It is now important to understand why we sometimes found complex eigenvalues even though we started out with matrices that only had real numbers in their entries. This is because an arbitrary polynomial of the form

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n=0 \tag{20.46}$$

with real coefficients has roots that are in general complex! So you might wonder if an arbitrary polynomial

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n=0 \tag{20.47}$$

with *complex* coefficients, are their roots that are not necessarily complex? It turns out the answer is no. Furthermore, how many roots are there? Hopefully there are n roots provided that $a_n\neq 0$. The answer is yes, and the following theorem makes both of these statements precise.

Theorem 20.48 (Fundamental Theorem of Algebra)**.** *Every polynomial of the form*

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n=0, \tag{20.49}$$

with all $a_0,a_1,\ldots,a_n\in\mathbb{C}$ and $a_n\neq 0$, has exactly n complex roots (possibly with multiplicity). In other words, there exist complex numbers $r_1,r_2,\ldots,r_n\in\mathbb{C}$ such that

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n=a_n(x-r_1)(x-r_2)\cdots(x-r_n). \tag{20.50}$$

Proof. There are many interesting proofs of this theorem, though we will not concern ourselves here with a proof. ■

For this reason, we may often work with complex numbers even if our original matrices were real-valued. This is because some eigenvalues can be complex! When this happens, we may find eigenvectors whose eigenvalues are complex, but these eigenvectors will also be complex. If we started with a real linear transformation $\mathbb{R}^n \xleftarrow{T} \mathbb{R}^n$ and we find a complex eigenvalue with an eigenvector with complex coefficients, then that vector does not live in $\mathbb{R}^n$, because $\mathbb{R}^n$ consists of vectors with only real entries. Instead, just like we enlarge our set of real numbers $\mathbb{R}$ to include complex numbers $\mathbb{C}$, we can enlarge $\mathbb{R}^n$ to include complex linear combinations of vectors. We denote the set of n-component vectors with complex entries by $\mathbb{C}^n$.

Another interesting fact that we did not point out explicitly from the examples but is also visible is the following.

Theorem 20.51. *Let $\{\vec{v}_1, \vec{v}_2\}$ be two eigenvectors of a linear transformation $\mathbb{R}^n \xleftarrow{T} \mathbb{R}^n$ with associated eigenvalues λ_1, λ_2 that are distinct. Then $\{\vec{v}_1, \vec{v}_2\}$ is a linearly independent set of vectors.*

Proof. We must show that the only solution to

$$x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0} \tag{20.52}$$

is $x_1 = x_2 = 0$. Without loss of generality, suppose that $\lambda_1 \neq 0$ (we know that at least one of λ_1 and λ_2 must be nonzero since they are distinct). Then, by applying T to

$$x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0} \tag{20.53}$$

we obtain

$$x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 = \vec{0}. \tag{20.54}$$

Since $\lambda_1 \neq 0$, we can divide by it to obtain

$$x_1\vec{v}_1 + x_2\frac{\lambda_2}{\lambda_1}\vec{v}_2 = \vec{0} \tag{20.55}$$

Subtracting this from the first equation, we get

$$\left(1 - \frac{\lambda_2}{\lambda_1}\right) x_2\vec{v}_2 = \vec{0}. \tag{20.56}$$

By assumption, $\lambda_1 \neq \lambda_2$ so that the term in parentheses is nonzero. Furthermore, $\vec{v}_2$ is nonzero because by definition of being an eigenvector, it must be nonzero. Hence $x_2 = 0$. Therefore, we are left with

$$x_1\vec{v}_1 = \vec{0} \tag{20.57}$$

and again, since $\vec{v}_1$ is nonzero by definition of being an eigenvector, this implies $x_1 = 0$. Hence, $\{\vec{v}_1, \vec{v}_2\}$ is a linearly independent set of vectors. ■

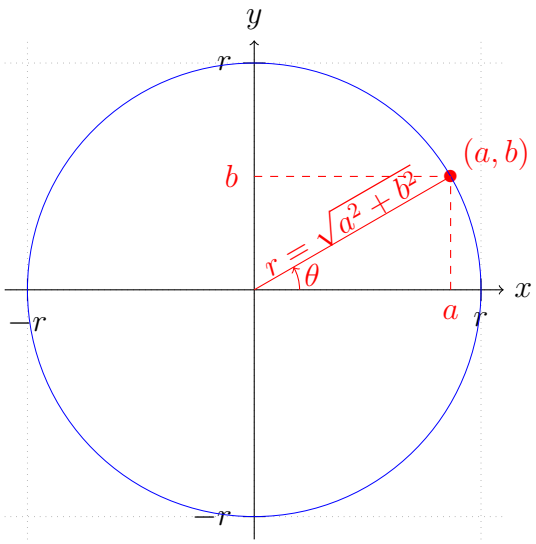
Theorem 20.58. *Suppose that $\{\vec{v}_1, \ldots, \vec{v}_k\}$ are eigenvectors of a linear transformation $\mathbb{R}^n \xleftarrow{T} \mathbb{R}^n$ with associated eigenvalues $\{\lambda_1, \ldots, \lambda_k\}$ all of which are distinct. Then the set $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent.*

Proof. Use the previous theorem and induction. ■

We now briefly discuss complex numbers, complex eigenvalues, and complex eigenvectors. A geometric visualization of complex numbers is obtained from the fact that every complex number has a polar decomposition.

Theorem 20.59. *Every complex number* $(a, b) \equiv a+b\sqrt{-1}$ *can be written in the form* $r(\cos\theta, \sin\theta) \equiv r\cos(\theta) + r\sin(\theta)\sqrt{-1}$ *for some unique non-negative number* r *and for some angle* θ *in* $[0, 2\pi)$. *Furthermore, if* a *and* b *are not both zero, then the angle* θ *is unique. If we define* $e^{\sqrt{-1}\theta} := \cos(\theta) + \sqrt{-1}\sin(\theta)$, *this says* $a + b\sqrt{-1} = re^{\sqrt{-1}\theta}$.

Proof. It helps to draw (a, b) in the plane.



Set

$$r := \sqrt{a^2 + b^2} \qquad \& \qquad \theta := \begin{cases} \arctan\left(\frac{b}{a}\right) & \text{for } a > 0, b \geq 0 \\ \frac{\pi}{2} & \text{for } a = 0, b > 0 \\ \pi + \arctan\left(\frac{b}{a}\right) & \text{for } a < 0 \\ \frac{3\pi}{2} & \text{for } a = 0, b < 0 \\ 2\pi + \arctan\left(\frac{b}{a}\right) & \text{for } a > 0, b < 0 \end{cases} \tag{20.60}$$

where θ is defined as long as both a and b are not zero. If $a = b = 0$, θ can be anything. ■

Exercise 20.61. Show that the eigenvalues of a matrix of the form

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{20.62}$$

are of the form

$$a \pm b\sqrt{-1}. \tag{20.63}$$

Complex eigenvalues and eigenvectors can be interpreted in this way as types of rotations in a particular plane. Hence, they always come in pairs.

Exercise 20.64. Write the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ as a rotation followed by a scaling. [Hint: compute the determinant to find the scaling factor and then use trigonometric identities to find the angle of rotation.]

Definition 20.65. Let $(a, b) \equiv a + b\sqrt{-1}$ be a complex number (a and b are real numbers). The *complex conjugate* of $(a, b) \equiv a + b\sqrt{-1}$ is

$$\overline{(a, b)} = (a, -b) \tag{20.66}$$

or in terms of the notation $a + b\sqrt{-1}$ this is written as

$$\overline{a + b\sqrt{-1}} = a - b\sqrt{-1}. \tag{20.67}$$

Definition 20.68. Let $\mathbb{C}^n$ be the set of n-component vectors but whose entries are all complex numbers. Addition, scalar multiplication (this time using complex numbers!), and the zero vector are analogous to how they are defined for $\mathbb{R}^n$. Similarly, $m \times n$ matrices can be taken to have complex numbers as their entries and are (complex) linear transformations $\mathbb{C}^m \leftarrow \mathbb{C}^n$.

Theorem 20.69. *Let A be a* real *$n \times n$ matrix. If λ is an eigenvalue of A then $\overline{\lambda}$ is also an eigenvalue.*

Proof. Let λ_0 be a complex eigenvalue of A that is not real. In particular, $\det(A - \lambda_0 \mathbb{1}_n) = 0$. The only complex number z for which $\lambda_0 z$ is real is $z = \overline{\lambda_0}$. Since $\det(A - \lambda \mathbb{1}_n)$ has $(\lambda_0 - \lambda)$ as a factor, i.e. there exists a degree $(n-1)$ polynomial p in the variable λ with $\det(A - \lambda \mathbb{1}_n) = (\lambda_0 - \lambda)p(\lambda)$, $p(\lambda)$ must have a factor of the form $(\overline{\lambda_0} - \lambda)$ in order for $\det(A - \lambda \mathbb{1}_n)$ to be real polynomial in λ. ■

Theorem 20.70. *Let A be an $n \times n$ matrix with complex entries. Then A has n (complex) eigenvalues including mutiplicity.*

Proof. The characteristic polynomial is given by $\det(A - \lambda \mathbb{1}_n)$. The coefficient in front of λ^n is always ± 1. Therefore, by the Fundamental Theorem of Algebra, this polynomial has n complex roots. ■

Recommended Exercises. Please check HuskyCT for the homework. Please show your work! Do *not* use calculators or computer programs to solve any problems! In this lecture, we covered parts of Sections 5.1, 5.2, 5.5, and Appendix B in [Lay].

Terminology checklist

eigenvector	
eigenvalue	
root of polynomial	
characteristic polynomial	
complex numbers	
complex eigenvalues	
complex conjugate	
multiplicity of eigenvalue	
eigenspace	

21 Diagonalizable matrices

As we saw in the previous lecture, it is easy to compute the eigenvalues of an upper-triangular matrix. It is equally simple to compute the eigenvalues of a diagonal matrix.

Theorem 21.1. *Let*

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \tag{21.2}$$

be a diagonal matrix. Then $\vec{e}_i$ is an eigenvector of D with eigenvalue λ_i for all $i \in \{1, \ldots, n\}$.

Proof. Exercise. ■

This Theorem tells us more than Theorem 20.41 since we know not only the eigenvalues but also the eigenvectors in the case of a diagonal matrix. However, not all matrices are diagonal. The good news is that many matrices are close to being diagonal in a sense we will describe soon.

Theorem 21.3. *Let A be an $n \times n$ matrix. Let P be an invertible matrix and define*

$$B := PAP^{-1}. \tag{21.4}$$

Then $\det(A) = \det(B)$.

Proof. This follows immediately from the product rule for determinants, which says

$$\det(B) = \det(PAP^{-1}) = \det(P)\det(A)\det(P^{-1}) = \det(P)\det(A)\det(P)^{-1} = \det(A). \tag{21.5}$$

■

Definition 21.6. An $n \times n$ matrix A is *similar* to an $n \times n$ matrix B iff there exists an invertible $n \times n$ matrix P such that $B = PAP^{-1}$.

Definition 21.7. Let A be an $n \times n$ matrix. A is *diagonalizable* iff it is similar to a diagonal matrix D, i.e. iff there exists an invertible $n \times n$ matrix P such that $A = PDP^{-1}$.

Note that solving for D in the above equation gives $D = P^{-1}AP$.

Theorem 21.8. *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. Furthermore, when A is diagonalizable and its eigenvectors are given by $\{\vec{v}_1, \ldots, \vec{v}_n\}$, setting*

$$P := \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}, \tag{21.9}$$

A can be expressed as

$$A = P \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} P^{-1}, \tag{21.10}$$

where λ_i is the eigenvalue corresponding to the eigenvector $\vec{v}_i$, namely

$$A\vec{v}_i = \lambda_i \vec{v}_i \tag{21.11}$$

for all $i \in \{1, \ldots, n\}$.

Proof.
($\Rightarrow$) Suppose $A = PDP^{-1}$ with D as in (21.10). Set $\vec{v}_i := P\vec{e}_i$ for each $i \in \{1, \ldots, n\}$. Then

$$AP\vec{e}_i = \underbrace{PP^{-1}}_{\mathbb{1}} AP\vec{e}_i = PD\vec{e}_i = P\lambda_i \vec{e}_i = \lambda_i P\vec{e}_i = \lambda_i \vec{v}_i \tag{21.12}$$

showing that $\vec{v}_i$ is an eigenvector of A with eigenvalue λ_i. Because P is invertible, its columns, which are exactly the vectors $\vec{v}_i = P\vec{e}_i$, are linearly independent, i.e. $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a linearly independent set of vectors.
($\Leftarrow$) Suppose that $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a linearly independent set of eigenvectors of A with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Set

$$P := \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}. \tag{21.13}$$

Then define the matrix D by $D := P^{-1}AP$. Notice that the columns of D are given by

$$D\vec{e}_i = P^{-1}AP\vec{e}_i = P^1 A\vec{v}_i = P^{-1}\lambda_i \vec{v}_i = \lambda_i P^{-1}\vec{v}_i = \lambda_i P^{-1}P\vec{e}_i = \lambda_i \vec{e}_i. \tag{21.14}$$

Therefore, D is a diagonal matrix and is of the form given in (21.10). ■

Theorem 21.15. *If an $n \times n$ matrix A is similar to an $n \times n$ matrix B, then the set of eigenvalues of A is the same as the set of eigenvalues of B and their multiplicities will also be the same.*

Proof. Let P be an invertible matrix such that $B = PAP^{-1}$. Let $\vec{v}$ be an eigenvector of A with eigenvalue λ. Then $P\vec{v}$ is an eigenvector of B with eigenvalue λ because

$$BP\vec{v} = PAP^{-1}P\vec{v} = PA\vec{v} = P\lambda\vec{v} = \lambda P\vec{v}. \tag{21.16}$$

A similar calculation shows that if $\vec{w}$ is an eigenvector of B with eigenvalue μ, then $P^{-1}\vec{w}$ is an eigenvector of A with eigenvalue μ. ■

Remark 21.17. The converse of this Theorem is not true! For example, the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad \& \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{21.18}$$

both have the same eigenvalues (1 with multiplicity 2) but are not similar. To see this, one could suppose, to the contrary, that there exists an invertible matrix

$$P = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{21.19}$$

such that $PA = BP^{-1}$ and show that this leads to some contradiction (such as showing that this system of equations does not have a solution). Another argument avoiding any calculations is to notice that similarity of matrices is an equivalence relation on the set of square matrices and diagonalizability is preserved under this equivalence relation. Namely,

i. A is similar to itself,

ii. if A is similar to B, then B is similar to A,

iii. if A is similar to B and B is similar to C, then A is similar to C.

Then, if A is diagonalizable and A is similar to B, then B is diagonalizable. Now, to apply this to our problem, since A has a basis of eigenvectors, we know that it is diagonalizable. We also know that B does not have two linearly independent eigenvectors. All of its eigenvectors are scalar multiples of the vector $\vec{e}_1$. Therefore, B is not diagonablizable and hence A cannot be similar to B.

Example 21.20. In Example 20.26, we found that the two eigenvalues with their corresponding eigenvectors of the matrix

$$A := \begin{bmatrix} 0 & -2 \\ -1 & 1 \end{bmatrix} \tag{21.21}$$

are

$$\left(\lambda_1 = -1, \vec{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) \qquad \& \qquad \left(\lambda_2 = 2, \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) \tag{21.22}$$

As in Theorem 21.8, set

$$P := \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \tag{21.23}$$

from which we can calculate the inverse as

$$P^{-1} = \frac{1}{3}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}. \tag{21.24}$$

We verify the claim of the theorem

$$\begin{aligned} PDP^{-1} &= \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}\frac{1}{3}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ 2 & -4 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} 0 & -6 \\ -3 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -2 \\ -1 & 1 \end{bmatrix} \\ &= A. \end{aligned} \tag{21.25}$$

The following is a more complicated example but illustrates how to actually solve cubic polynomials when calculating eigenvalues.

Example 21.26. Let A be the matrix[56]

$$A := \begin{bmatrix} 0 & 1 & 1 \\ 2 & 1 & 2 \\ 3 & 3 & 2 \end{bmatrix}. \tag{21.27}$$

The eigenvalues of A are obtained from solving the polynomial equation

$$\begin{aligned} \det(A - \lambda \mathbb{1}_3) &= \det \begin{bmatrix} -\lambda & 1 & 1 \\ 2 & 1-\lambda & 2 \\ 3 & 3 & 2-\lambda \end{bmatrix} \\ &= -\lambda \det \begin{bmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{bmatrix} - 1 \det \begin{bmatrix} 2 & 2 \\ 3 & 2-\lambda \end{bmatrix} + 1 \det \begin{bmatrix} 2 & 1-\lambda \\ 3 & 3 \end{bmatrix} \\ &= -\lambda\Big((1-\lambda)(2-\lambda) - 6\Big) - \Big(2(2-\lambda) - 6\Big) + \Big(6 - 3(1-\lambda)\Big) \\ &= -\lambda^3 + 3\lambda^2 + 9\lambda + 5 \\ &= 0. \end{aligned} \tag{21.28}$$

Solving cubic equations is not impossible, but it is difficult. A good strategy is to guess one solution and then factor out that term to reduce it to a quadratic polynomial. By inspection, $\lambda = -1$ is one such solution. Knowing this, we can use long division of polynomials

$$(\lambda + 1)\overline{)-\lambda^3 + 3\lambda^2 + 9\lambda + 5}$$

to figure out the other factors. To compute this division, one does what one is used to from ordinary long division by calculating remainders

$$\begin{array}{r} -\lambda^2 \\ (\lambda+1)\overline{)-\lambda^3 + 3\lambda^2 + 9\lambda + 5} \\ \underline{-(-\lambda^3 - \lambda^2)} \\ 4\lambda^2 + 9\lambda + 5 \end{array}$$

Proceeding, we get

$$\begin{array}{r} -\lambda^2 + 4\lambda \\ (\lambda+1)\overline{)-\lambda^3 + 3\lambda^2 + 9\lambda + 5} \\ \underline{-(-\lambda^3 - \lambda^2)} \\ 4\lambda^2 + 9\lambda + 5 \\ \underline{-(4\lambda^2 + 4\lambda)} \\ 5\lambda + 5 \end{array}$$

until finally

[56] This is exercise 5.3.11 in [Lay] but without knowledge of any of the eigenvalues.

$$
\begin{array}{r}
-\lambda^2+4\lambda+5 \\
(\lambda+1)\overline{)-\lambda^3+3\lambda^2+9\lambda+5} \\
\underline{-(-\lambda^3-\lambda^2)} \\
4\lambda^2+9\lambda+5 \\
\underline{-(4\lambda^2+4\lambda)} \\
5\lambda+5 \\
\underline{-(5\lambda+5)} \\
0
\end{array}
$$

From this, we can factor our polynomial equation to the form

$$-\lambda^3+3\lambda^2+9\lambda+5=(\lambda+1)(-\lambda^2+4\lambda+5). \tag{21.29}$$

Now we can apply the quadratic formula to find the roots of $-\lambda^2+4\lambda+5$ and we get

$$\lambda=\frac{-4\pm\sqrt{4^2-4(-1)(5)}}{2(-1)}=2\mp 3. \tag{21.30}$$

Thus, the polynomial factors into

$$-\lambda^3+3\lambda^2+9\lambda+5=(\lambda+1)(\lambda+1)(-\lambda+5)=(\lambda+1)^2(-\lambda+5) \tag{21.31}$$

and the roots of the polynomial, i.e. the eigenvalues of the matrix A, are $\lambda_1=-1,\lambda_2=-1,\lambda_3=5$.

Problem 21.32. From Example 21.26, calculate the eigenvectors $\vec{v}_1,\vec{v}_2,\vec{v}_3$ of A associated to the eigenvalues $\lambda_1,\lambda_2,\lambda_3$, respectively. Construct the matrix P. Calculate P^{-1} and verify explicitly that $A=PDP^{-1}$.

Answer. We first find the eigenvectors for eigenvalue $\lambda_1=\lambda_2=-1$. We must solve

$$\begin{bmatrix}0&1&1\\2&1&2\\3&3&2\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=-1\begin{bmatrix}x\\y\\z\end{bmatrix} \tag{21.33}$$

i.e.

$$\begin{bmatrix}0-(-1)&1&1\\2&1-(-1)&2\\3&3&2-(-1)\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}1&1&1\\2&2&2\\3&3&3\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}. \tag{21.34}$$

The solutions to this can be written in parametric form as

$$\begin{bmatrix}x\\y\\z\end{bmatrix}=y\begin{bmatrix}-1\\1\\0\end{bmatrix}+z\begin{bmatrix}-1\\0\\1\end{bmatrix} \tag{21.35}$$

with y and z free variables. Hence, we can take

$$\vec{v}_1:=\begin{bmatrix}-1\\1\\0\end{bmatrix}\qquad\&\qquad\vec{v}_2:=\begin{bmatrix}-1\\0\\1\end{bmatrix} \tag{21.36}$$

(in fact, you can set $\vec{v}_1$ and $\vec{v}_2$ to be any two linearly independent vectors in this kernel because the corresponding eigenvalue is the same). For $\lambda_3 = 5$, the corresponding linear system to solve is given by

$$\begin{bmatrix} 0-5 & 1 & 1 \\ 2 & 1-5 & 2 \\ 3 & 3 & 2-5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -5 & 1 & 1 \\ 2 & -4 & 2 \\ 3 & 3 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{21.37}$$

The solutions to this are written in parametric form as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = z \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{21.38}$$

(I've removed an overall factor of $-\frac{1}{3}$ which one obtains when row reducing). Hence, we can take

$$\vec{v}_3 := \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{21.39}$$

as an eigenvector of A with eigenvalue $\lambda_3 = 5$. Therefore,

$$P = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} \qquad \& \qquad P^{-1} = \frac{1}{6} \begin{bmatrix} -2 & 4 & -2 \\ -3 & -3 & 3 \\ 1 & 1 & 1 \end{bmatrix}. \tag{21.40}$$

Furthermore, the equality $A = PDP^{-1}$, which reads

$$\begin{bmatrix} 0 & 1 & 1 \\ 2 & 1 & 2 \\ 3 & 3 & 2 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix} \left(\frac{1}{6} \begin{bmatrix} -2 & 4 & -2 \\ -3 & -3 & 3 \\ 1 & 1 & 1 \end{bmatrix} \right), \tag{21.41}$$

holds.

Warning: not all matrices are diagonalizable.

Example 21.42. The vertical shear matrix

$$S_1^{|} := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{21.43}$$

is not diagonalizable. One way to see this is to use Theorem 21.8. We have already calculated the eigenvectors of this matrix in Example 20.3. We only found a *single* eigenvector

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{21.44}$$

with eigenvalue 1. Since a single vector in $\mathbb{R}^2$ cannot span all of $\mathbb{R}^2$, this means there aren't 2 linearly independent eigenvectors associated to this matrix. Therefore, $S_1^{|}$ is not diagonalizable.

Remark 21.45. Even though $S_1^|$ is not diagonalizable, it can be put into what is called *Jordan normal form*, which is the next best thing after diagonal matrices. Namely, every matrix is similar to a matrix of the form $D + N$, where D is diagonal and N has all zero entries except possibly a few 1's directly above the diagonal (and satisfies other properties). However, there is a caveat, which is that one must allow for the matrices to be expressed in terms of complex numbers and to allow for the diagonal matrix D to contain complex numbers as well.

Here's a theorem that gives a sufficient condition for a matrix to be diagonalizable.

Theorem 21.46. *Let A be an $n \times n$ matrix with n* distinct *real eigenvalues. Then A is diagonalizable.*

Proof. Let $\{\lambda_1, \ldots, \lambda_n\}$ be these eigenvalues with corresponding eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$. By Theorem 20.58, the set of vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is linearly independent (in fact, they form a basis of $\mathbb{R}^n$ since there are n of them). By Theorem 21.8, A is therefore diagonalizable. ■

It is *not necessary* for a matrix A to have distinct eigenvalues to be diagonalizable. A simple example is a matrix that is already in diagonal form. Once it is known that a matrix A is similar to a diagonal matrix, it becomes easy to calculate "polynomials" in the matrix A. This is because if P is an invertible matrix such that $A = PDP^{-1}$, then

$$A^2 = \left(PDP^{-1}\right)^2 = PD\underbrace{P^{-1}P}_{=\mathbb{1}_n}DP^{-1} = PDDP^{-1} = PD^2P^{-1} \tag{21.47}$$

and similarly for higher orders

$$A^k = \left(PDP^{-1}\right)^k = PD\underbrace{P^{-1}P}_{=\mathbb{1}_n}D\underbrace{P^{-1}P}_{=\mathbb{1}_n}D\underbrace{P^{-1}P}_{=\mathbb{1}_n}DP^{-1}\cdots PDP^{-1} = PD^kP^{-1}. \tag{21.48}$$

And to compute the power of a diagonal matrix is a piece of cake:

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n^k \end{bmatrix}. \tag{21.49}$$

Example 21.50. Let A be as in Example 21.20. Then

$$\begin{aligned} A^5 &= PD^5P^{-1} \\ &= \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}^5 \frac{1}{3}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 32 \end{bmatrix}\frac{1}{3}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} -2 & 32 \\ -1 & -32 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} 30 & -66 \\ -33 & 63 \end{bmatrix} \\ &= \begin{bmatrix} 10 & -22 \\ -11 & 21 \end{bmatrix}. \end{aligned} \tag{21.51}$$

It is also instructive to write this formula for arbitrary powers. Let $\lambda_1 = -1$ and $\lambda_2 = 2$ and fix $n \in \mathbb{N}$. Then

$$\begin{aligned} A^n &= PD^nP^{-1} \\ &= \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2\lambda_1^n & \lambda_2^n \\ \lambda_1^n & -\lambda_2^n \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2\lambda_1^n + \lambda_2^n & 2\lambda_1^n - 2\lambda_2^n \\ \lambda_1^n - \lambda_2^n & \lambda_1^n + 2\lambda_2^n \end{bmatrix}. \end{aligned} \tag{21.52}$$

Definition 21.53. Let A be an $n \times n$ matrix. Let p be a positive integer. A *degree p polynomial in A* is an expression of the form

$$t_0\mathbb{1}_n + t_1A + t_2A^2 + \cdots + t_pA^p, \tag{21.54}$$

where the t's are real numbers.

Theorem 21.55. *Let A be a diagonalizable $n \times n$ matrix with a decomposition of the form*

$$A = PDP^{-1}, \tag{21.56}$$

where D is diagonalizable and P is a matrix of eigenvectors of A. Then

$$t_0\mathbb{1}_n + t_1A + t_2A^2 + \cdots + t_pA^p = P\Big(t_0\mathbb{1}_n + t_1D + t_2D^2 + \cdots + t_pD^p\Big)P^{-1}. \tag{21.57}$$

Proof. This follows from the previous discussion since

$$\begin{aligned} t_0\mathbb{1}_n + t_1A + t_2A^2 + \cdots + t_pA^p &= t_0\mathbb{1}_n + t_1PDP^{-1} + t_2(PDP^{-1})^2 + \cdots + t_p(PDP^{-1})^p \\ &= t_0\mathbb{1}_n + t_1PDP^{-1} + t_2PD^2P^{-1} + \cdots + t_pPD^pP^{-1} \\ &= P\Big(t_0\mathbb{1}_n + t_1D + t_2D^2 + \cdots + t_pD^p\Big)P^{-1}. \end{aligned} \tag{21.58}$$

■

As we saw in earlier lectures, some $n \times n$ matrices do not have n eigenvalues or a basis of n eigenvectors. For example, shears in two dimensions only have a one-dimensional basis of eigenvectors and rotations in two dimensions have none! Each of these problems has a resolution, the former in terms of what are called *generalized eigenvectors*, and the latter in terms of *complex eigenvectors*. The former is typically studied in a more advanced course on linear algebra.

Recommended Exercises. Please check HuskyCT for the homework. Please show your work! Do *not* use calculators or computer programs to solve any problems! In this lecture, we covered Section 5.3.

Terminology checklist

similar matrix	
diagonalizable	
polynomial long division	
matrix polynomial	

22 Spectral decomposition and the Stern-Gerlach experiment*

We will now learn about sufficient conditions for a matrix to be diagonalizable as well as a useful decomposition for any such matrix.

Definition 22.1. An $m \times n$ matrix A is *symmetric* whenever $A^T = A$.

Notice that $m = n$ for symmetric $m \times n$ matrices.

A similar definition can be made for complex matrices, but one takes into account the complex conjugation. Why this is so will be made clear shortly.

Definition 22.2. Let A be an $m \times n$ matrix with possibly complex entries

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}. \tag{22.3}$$

The *conjugate transpose* of A is the $n \times m$ matrix

$$A^\dagger := \begin{bmatrix} \overline{a_{11}} & \overline{a_{21}} & \cdots & \overline{a_{m1}} \\ \overline{a_{12}} & \overline{a_{22}} & \cdots & \overline{a_{m2}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \cdots & \overline{a_{mn}} \end{bmatrix}. \tag{22.4}$$

The superscript † is pronounced "dagger."

Definition 22.5. An $m \times n$ matrix A is *Hermitian* whenever $A^\dagger = A$.

Again, an $m \times n$ matrix A is Hermitian implies $m = n$. Projections are defined similarly as before but with this complex conjugation taken into account.

Definition 22.6. A complex $m\times$m matrix P is an *orthogonal projection* iff $P^2 = P$ and $P^\dagger = P$.

Example 22.7. The three matrices

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \& \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{22.8}$$

are all Hermitian. Notice that the eigenvalues of these three matrices are all ± 1, which are real. We will say something about this phenomenon soon.

Definition 22.9. The *inner product* on $\mathbb{C}^n$ is the function

$$\langle \cdot\,,\, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C} \tag{22.10}$$

defined by

$$\mathbb{C}^n \times \mathbb{C}^n \ni (\vec{u}, \vec{v}) \mapsto \langle \vec{u}, \vec{v} \rangle := \vec{u}^\dagger \vec{v}. \tag{22.11}$$

In terms of components, this says

$$\left\langle \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \right\rangle = \overline{u_1}v_1 + \cdots + \overline{u_n}v_n. \tag{22.12}$$

Notice that for any complex number c and for any two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{C}^n$,

$$\langle \vec{u}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{u} \rangle} \tag{22.13}$$

and

$$\langle c\vec{u}, \vec{v} \rangle = \overline{c}\langle \vec{u}, \vec{v} \rangle. \tag{22.14}$$

Theorem 22.15. *Let A be a complex $m \times n$ matrix. Then*

$$\langle \vec{v}, A\vec{w} \rangle = \langle A^\dagger \vec{v}, \vec{w} \rangle \tag{22.16}$$

for all vectors $\vec{w}$ in $\mathbb{C}^n$ and $\vec{v}$ in $\mathbb{C}^m$.

Proof. A calculation easily proves this. ■

Theorem 22.17. *The eigenvalues of a Hermitian matrix are real.*

Proof. Let A be a Hermitian matrix and let λ be an eigenvalue with a corresponding eigenvector $\vec{v}$. Then

$$\lambda\langle \vec{v}, \vec{v} \rangle = \langle \vec{v}, \lambda\vec{v} \rangle = \langle \vec{v}, A\vec{v} \rangle = \langle A^\dagger \vec{v}, \vec{v} \rangle = \langle A\vec{v}, \vec{v} \rangle = \langle \lambda\vec{v}, \vec{v} \rangle = \overline{\lambda}\langle \vec{v}, \vec{v} \rangle \tag{22.18}$$

Since $\vec{v}$ is nonzero, $\langle \vec{v}, \vec{v} \rangle$ is nonzero as well, and this shows that $\lambda = \overline{\lambda}$, i.e. λ is real. ■

Theorem 22.19. *Let A be a Hermitian matrix. Then eigenvectors corresponding to different eigenvalues are orthogonal. An analogous statement holds for symmetric real matrices provided one considers real eigenvalues.*

Proof. Let λ_1 and λ_2 be two distinct eigenvalues of A and let $\vec{v}_1$ and $\vec{v}_2$ be corresponding eigenvectors. Then

$$\begin{aligned} \lambda_2\langle \vec{v}_1, \vec{v}_2 \rangle &= \langle \vec{v}_1, \lambda_2\vec{v}_2 \rangle \\ &= \langle \vec{v}_1, A\vec{v}_2 \rangle \\ &= \langle A^\dagger \vec{v}_1, \vec{v}_2 \rangle \\ &= \langle A\vec{v}_1, \vec{v}_2 \rangle \\ &= \langle \lambda_1\vec{v}_1, \vec{v}_2 \rangle \\ &= \overline{\lambda_1}\langle \vec{v}_1, \vec{v}_2 \rangle \\ &= \lambda_1\langle \vec{v}_1, \vec{v}_2 \rangle, \end{aligned} \tag{22.20}$$

where in the last step we used the fact that eigenvalues of Hermitian matrices are real. This calculation shows that

$$(\lambda_2 - \lambda_1)\langle \vec{v}_1, \vec{v}_2 \rangle = 0, \tag{22.21}$$

which is only possible if either $\lambda_2 - \lambda_1 = 0$ or $\langle \vec{v}_1, \vec{v}_2 \rangle = 0$. However, by assumption, since $\lambda_1 \neq \lambda_2$, it follows that $\lambda_2 - \lambda_1 \neq 0$. Hence $\langle \vec{v}_1, \vec{v}_2 \rangle = 0$ which means that $\vec{v}_1$ is orthogonal to $\vec{v}_2$. ■

Theorem 22.22. *A matrix A is Hermitian if and only if there exists a diagonal matrix D and a matrix P all of whose columns are orthogonal such that $A = PDP^{-1}$.*

In fact, let A be an $n \times n$ Hermitian matrix, let

$$D := \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \tag{22.23}$$

be a diagonal matrix of its eigenvalues, and let

$$P := \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_n \\ | & & | \end{bmatrix} \tag{22.24}$$

be the matrix of orthonormal eigenvectors (given a matrix P that initially might have all orthogonal eigenvectors, one can simply scale them so that they have unit length). Then, a quick calculation shows that

$$P^\dagger = P^{-1}. \tag{22.25}$$

Using this, the previous theorem says that

$$A = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_n \\ | & & | \end{bmatrix}^\dagger = \sum_{k=1}^{n} \lambda_k \vec{u}_k \vec{u}_k^\dagger. \tag{22.26}$$

Notice that $P_k := \vec{u}_k \vec{u}_k^\dagger$ is an $n \times n$ matrix satisfying $P_k^2 = P_k$and $P_k^\dagger = P_k$. In fact, this operator is the orthogonal projection operator onto the subspace span$\{\vec{u}_k\}$. Hence,

$$A = \sum_{k=1}^{n} \lambda_k P_k \tag{22.27}$$

provides a formula for the Hermitian matrix A as a weighted sum of projection operators onto orthogonal one-dimensional subspaces of A generated by the eigenvectors of A. This decomposition is referred to as the *spectral decomposition* of A.

Example 22.28. Consider the matrix

$$\sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \tag{22.29}$$

from Example 22.7. The eigenvalues are $\lambda_y^\uparrow = +1$ and $\lambda_y^\downarrow = -1$ with corresponding eigenvectors given by

$$\vec{v}_y^\uparrow = \begin{bmatrix} 1 \\ i \end{bmatrix} \qquad \& \qquad \vec{v}_y^\downarrow = \begin{bmatrix} i \\ 1 \end{bmatrix} \tag{22.30}$$

respectively. Associated normalized eigenvectors are

$$\vec{u}_y^{\uparrow} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\ i\end{bmatrix}. \qquad \& \qquad \vec{u}_y^{\downarrow} = \frac{1}{\sqrt{2}}\begin{bmatrix}i\\ 1\end{bmatrix} \tag{22.31}$$

Then, the orthogonal matrix P_y that diagonalizes σ_y is given by

$$P_y = \begin{bmatrix}1/\sqrt{2} & i/\sqrt{2}\\ i/\sqrt{2} & 1/\sqrt{2}\end{bmatrix} \tag{22.32}$$

as we can check[57]

$$P_y \begin{bmatrix}\lambda_y^{\uparrow} & 0\\ 0 & \lambda_y^{-}\end{bmatrix} P_y^{\dagger} = \frac{1}{2}\begin{bmatrix}1 & i\\ i & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}1 & -i\\ -i & 1\end{bmatrix} = \begin{bmatrix}0 & -i\\ i & 0\end{bmatrix} = \sigma_y \tag{22.33}$$

and the projection matrices $P_y^{\uparrow}$ and $P_y^{\downarrow}$ that project onto the eigenspaces $\text{span}\big(\{\vec{u}_y^{\uparrow}\}\big)$ and $\text{span}\big(\{\vec{u}_y^{\downarrow}\}\big)$, respectively, are given by (each calculated in two different ways to illustrate the possible methods)[58]

$$P_y^{\uparrow} = \vec{u}_y^{\uparrow}\vec{u}_y^{\uparrow\dagger} = \left(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ i\end{bmatrix}\right)\left(\frac{1}{\sqrt{2}}\begin{bmatrix}1 & -i\end{bmatrix}\right) = \begin{bmatrix}1/2 & -i/2\\ i/2 & 1/2\end{bmatrix} \tag{22.34}$$

and[59]

$$P_y^{\downarrow} = \begin{bmatrix}\langle \vec{u}_y^{\downarrow}, \vec{e}_1\rangle\, \vec{u}_y^{\downarrow} & \langle \vec{u}_y^{\downarrow}, \vec{e}_2\rangle\, \vec{u}_y^{\downarrow}\end{bmatrix} = \frac{1}{2}\begin{bmatrix}\left\langle \begin{bmatrix}i\\ 1\end{bmatrix}, \begin{bmatrix}1\\ 0\end{bmatrix}\right\rangle \begin{bmatrix}i\\ 1\end{bmatrix} & \left\langle \begin{bmatrix}i\\ 1\end{bmatrix}, \begin{bmatrix}0\\ 1\end{bmatrix}\right\rangle \begin{bmatrix}i\\ 1\end{bmatrix}\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1 & i\\ -i & 1\end{bmatrix}. \tag{22.35}$$

Therefore, the matrix σ_y decomposes as

$$\sigma_y = \lambda_y^{\uparrow}P_y^{\uparrow} + \lambda_y^{\downarrow}P_y^{\downarrow} = 1\begin{bmatrix}1/2 & -i/2\\ i/2 & 1/2\end{bmatrix} - 1\begin{bmatrix}1/2 & i/2\\ -i/2 & 1/2\end{bmatrix}. \tag{22.36}$$

Example 22.37. For the record, consider the matrix σ_z from Example 22.7. The eigenvalues are $\lambda_z^{\uparrow} = +1$ and $\lambda_z^{\downarrow} = -1$ with corresponding normalized eigenvectors given by

$$\vec{u}_z^{\uparrow} = \begin{bmatrix}1\\ 0\end{bmatrix} \qquad \& \qquad \vec{u}_z^{\downarrow} = \begin{bmatrix}0\\ 1\end{bmatrix} \tag{22.38}$$

respectively. The projection operators onto these eigenspaces are easy to read off because the matrix σ_z is already in diagonal form. These projections are

$$P_z^{\uparrow} = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \qquad \& \qquad P_z^{\downarrow} = \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}. \tag{22.39}$$

[57] For a complex matrix P with orthogonal columns, the inverse P^{-1} is given by $P^{\dagger}$ instead of P^T, which is what happens when P is real.

[58] In Dirac bra-ket notation, this reads $P_y^{\uparrow} = |\vec{u}_y^{\uparrow}\rangle\langle\vec{u}_y^{\uparrow}|$.

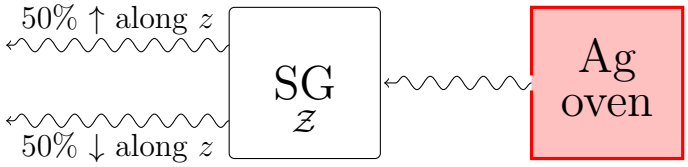
[59] Remember, the matrix A associated to a linear transformation $\mathbb{C}^2 \xleftarrow{T} \mathbb{C}^2$ is given by

$$A = \begin{bmatrix}| & |\\ T(\vec{e}_1) & T(\vec{e}_2)\\ | & |\end{bmatrix}.$$

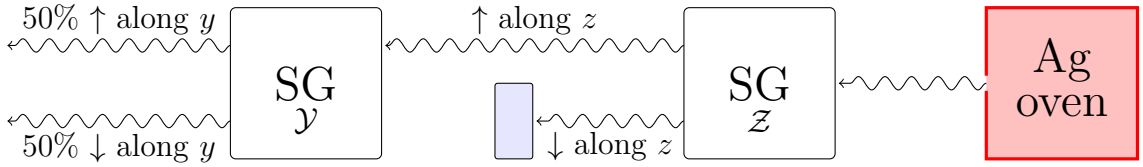
Example 22.40 (The Stern-Gerlach experiment and quantum mechanics)**.** Consider the following experiment where (light) classical magnets are sent through a specific type of magnetic field (fixed throughout the experiment). Depending on the orientation of the magnet, the deflection will be distributed continuously according to this orientation. However, if a silver atom is sent through the same apparatus, its deflection is more discrete. It is either only sent up or down. The silver atoms are show out of an oven so that their internal properties are distributed as uniformly as possible.

Watch video at: https://upload.wikimedia.org/wikipedia/commons/9/9e/Quantum_spin_and_the_Stern-Gerlach_experiment.ogv

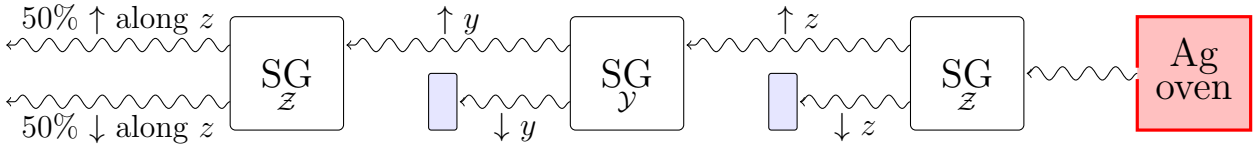
Let us visualize this Stern-Gerlach experiment by the following cartoon (read from right to left).



Now, imagine a second experiment where we isolate the silver atoms that were deflected up along the z axis and we send those atoms through a Stern-Gerlach experiment oriented along the y axis. Experimentally, we find that on average 50% of the atoms are deflected in the positive y direction and 50% in the negative y direction with the same magnitude as the deflection in the z direction from the first experiment.



What do you think happens if we take the atoms that were deflected in the positive z direction first, then deflected in the positive y direction, and we send these through yet another Stern-Gerlach experiment oriented again back along the z direction?



It turns out that not all of them will still deflect in the positive z direction. Instead, 50% of these atoms will be deflected in the positive z direction and 50% in the negative z direction! Preposterous! How can we possibly explain this phenomenon?

Because we only see two possibilities in the deflection of the silver atoms, we postulate that the state corresponding to these two possibilities is described by a normalized complex 2-component vector.

$$\vec{\psi} := \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} \tag{22.41}$$

We postulate that an atom that gets deflected in the positive z direction corresponds to an eigenvector of the σ_z operator with eigenvalue $+1$ and an atom that gets deflected in the negative z direction corresponds to an eigenvector of the σ_z operator with eigenvalue -1. This is a weird postulate (what do vectors in complex space have anything to do with reality!?), but let us see where it takes us. Furthermore, we postulate that blocking out states that get deflected in a positive or negative direction corresponds to projecting the state onto the eigenspace that is allowed to pass through. Finally, we postulate that the probability of a state $\vec{\psi}$ as above to be measured in another state

$$\vec{\phi} := \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \tag{22.42}$$

is given by

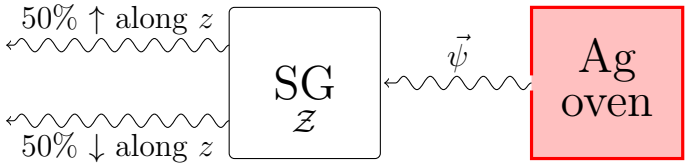
$$\left| \left\langle \vec{\phi}, \vec{\psi} \right\rangle \right|^2. \tag{22.43}$$

We now rely on our analysis from the previous examples and use the notation set therein. We interpret $\vec{u}_z^{\uparrow}$ and $\vec{u}_z^{\downarrow}$ as the states (silver atoms) that are deflected up and down along the z axis, respectively. Similarly, $\vec{u}_y^{\uparrow}$ and $\vec{u}_y^{\downarrow}$ are the states that are deflected up and down along the y axis, respectively. Because initially the oven provides the experimentally observed fact that half of the particles get deflected (up) along the positive z direction and half of them get deflected (down) along the negative z direction, any vector of the form

$$\vec{\psi} = \frac{e^{i\theta}}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{22.44}$$

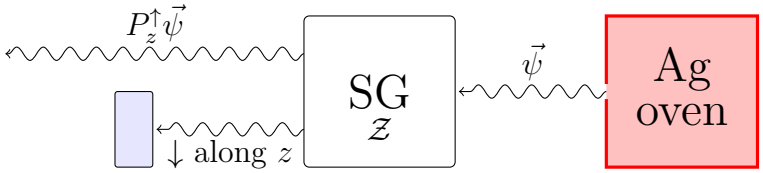
suffices to describe the initial state (in fact, we can ignore the $e^{i\theta}$ and set $\theta = 0$). This is because

$$\left| \left\langle \vec{u}_z^{\uparrow}, \vec{\psi} \right\rangle \right|^2 = \frac{1}{2} \qquad \& \qquad \left| \left\langle \vec{u}_z^{\downarrow}, \vec{\psi} \right\rangle \right|^2 = \frac{1}{2}. \tag{22.45}$$



By selecting only the silver atoms that are deflect up along the z axis, we project our initial state onto this state

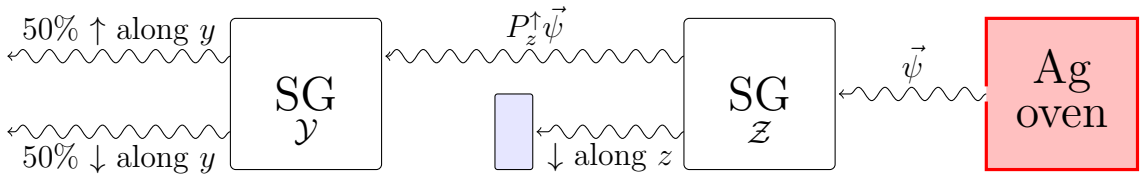
$$P_z^{\uparrow}\vec{\psi} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \vec{u}_z^{\uparrow} \tag{22.46}$$



and then we normalize this state to $\vec{u}_z^{\uparrow}$ so that our probabilistic interpretation will still hold. In the second experiment, this state gets sent through the Stern-Gerlach apparatus oriented along

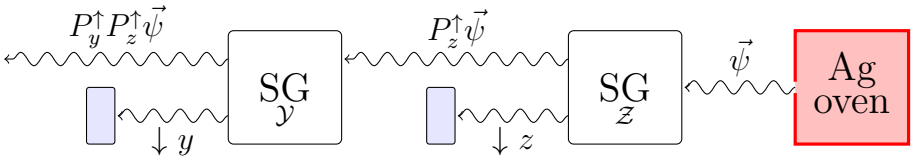
the y direction. Let us check that indeed half of the particles are deflected up along the y direction and the other half are deflected down.

$$\left|\langle \vec{u}_y^{\uparrow}, \vec{u}_z^{\uparrow} \rangle\right|^2 = \frac{1}{2} \qquad \& \qquad \left|\langle \vec{u}_y^{\downarrow}, \vec{u}_z^{\uparrow} \rangle\right|^2 = \frac{1}{2}. \tag{22.47}$$



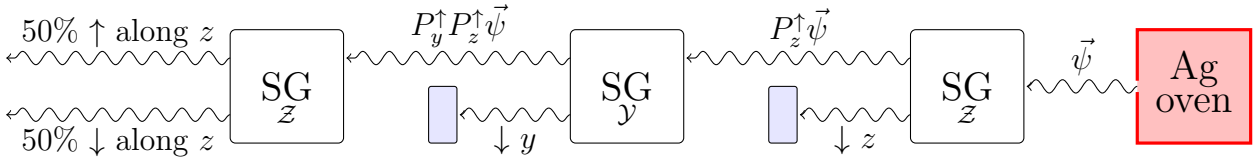
So far so good. Now let's put these postulates to the ultimate test of projecting our resulting state onto the up state along the y direction to obtain

$$P_y^{\uparrow}\vec{u}_z^{\uparrow} = \frac{1}{\sqrt{2}}\vec{u}_y^{\uparrow} \tag{22.48}$$



and again normalize to obtain just $\vec{u}_y^{\uparrow}$. Now let's send these silver atoms back to another Stern-Gerlach apparatus oriented along the z direction. So what does the math tell us? The probabilities of this resulting state deflecting either up or down along the z axis after going through all these experiments is

$$\left|\langle \vec{u}_z^{\uparrow}, \vec{u}_y^{\uparrow} \rangle\right|^2 = \frac{1}{2} \qquad \& \qquad \left|\langle \vec{u}_z^{\downarrow}, \vec{u}_y^{\uparrow} \rangle\right|^2 = \frac{1}{2} \tag{22.49}$$



which beautifully agrees with the experiment! Is all of this a coincidence? We think not. And it turns out that the study of abstract vector spaces is intimately related to other phenomena centered around the subject called quantum mechanics. The effect described in this example is due to the *spin* of an electron (in the case of silver, this is due to the single valence electron), a feature inherent of particles and described adequately in the theory of quantum mechanics.

Recommended Exercises. Please check HuskyCT for the homework. Please show your work! Do *not* use calculators or computer programs to solve any problems! In this lecture, we covered Section 7.1 and some additional material based on Chapter 1 of [5].

23 Solving ordinary differential equations

Example 23.1. Consider the third order linear differential equation (ODE) given by

$$f''' - 2f'' - f' + 2f = 0, \tag{23.2}$$

where f is a smooth function of the variable x and each prime denotes applying a derivative once. One often solves such an ODE by replacing f with a constant, say λ, and replacing primes with powers so that it becomes

$$\lambda^3 - 2\lambda^2 - \lambda + 2 = 0. \tag{23.3}$$

This comes from making the ansatz $f(x) = e^{\lambda x}$, substituting into (23.2)

$$\frac{d^3}{dx^3}(e^{\lambda x}) - 2\frac{d^2}{dx^2}(e^{\lambda x}) - \frac{d}{dx}(e^{\lambda x}) + 2e^{\lambda} = 0 \tag{23.4}$$

and cancelling the common factor of $e^{\lambda x}$ (which is allowed because $e^{\lambda x}$ is non-zero for all $x \in \mathbb{R}$ and all $\lambda \in \mathbb{C}$). One then finds the roots of this polynomial. In this case, the polynomial factors into

$$\lambda^3 - 2\lambda^2 - \lambda + 2 = (\lambda + 1)(\lambda - 1)(\lambda - 2) \tag{23.5}$$

so that the roots are $\lambda_1 = -1, \lambda_2 = 1$, and $\lambda_3 = 2$. The linearly independent solutions are therefore e^{-x}, e^{x}, and e^{2x}. But where did the ansatz $f(x) = e^{\lambda x}$ come from? Is there another way, perhaps a more naive way, to obtain these functions? There actually is. If we set $g := f'$ and $h := g' = f''$, the third order ODE becomes

$$\begin{aligned} f' &= g \\ g' &= h \\ h' &= -2f + g + 2h \end{aligned} \tag{23.6}$$

which is a system of *first order* linear differential equations and can therefore be written as

$$\frac{d}{dx}\begin{bmatrix} f \\ g \\ h \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 2 \end{bmatrix}\begin{bmatrix} f \\ g \\ h \end{bmatrix}. \tag{23.7}$$

This gives us an example of how to reduce an m-th order linear ODE to a system of m first order linear ODEs.

To continue our study of m-th order linear ODEs, we should therefore study first order ODEs.

Example 23.8. Let $a \in \mathbb{R}$ and consider the first order linear ODE

$$f' = af, \tag{23.9}$$

or equivalently

$$\frac{d}{dx}f = af. \tag{23.10}$$

The solution to this ODE is fairly straightforward and can even be accomplished by separating variables so that it is of the form

$$\frac{df}{f} = a dx \tag{23.11}$$

Integrating both sides gives

$$\int \frac{df}{f} = a \int dx \qquad \Rightarrow \qquad \ln(f) = ax + c \qquad \Rightarrow \qquad f = Ce^{ax}, \tag{23.12}$$

where c is a constant of integration and $C := e^c$. Therefore, the solutions of $f' = af$ are all scalar multiples of $f(x) = e^{ax}$.

But what if a was replaced by a matrix? If A was a diagonal matrix, perhaps we might still be able to solve such an ODE.

Example 23.13. Let $a, b \in \mathbb{R}$ and consider the system of first order linear ODEs given by

$$\frac{d}{dx}\begin{bmatrix} f \\ g \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\begin{bmatrix} f \\ g \end{bmatrix}. \tag{23.14}$$

This describes the two ODEs

$$\begin{aligned} f' &= af \\ g' &= bg \end{aligned} \tag{23.15}$$

which is also solvable individually with solutions given by $f(x) = Ce^{ax}$ and $g(x) = De^{bx}$ with C and D integration constants. It seems like we have not used any linear algebra to solve this ODE. However, notice that we can express the solution in matrix form as

$$\begin{bmatrix} f(x) \\ g(x) \end{bmatrix} = \begin{bmatrix} e^{ax} & 0 \\ 0 & e^{bx} \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix}. \tag{23.16}$$

What is the general theory behind this result?

If $\vec{x}$ is a vector in $\mathbb{R}^m$ that is a function of time t and A is an $m \times m$ matrix, then

$$\frac{d}{dt}\vec{x} = A\vec{x} \tag{23.17}$$

describes an m-th order linear ordinary differential equation (ODE) and if the initial condition is $\vec{x}_0 := \vec{x}(0)$, then the solution is given by

$$\vec{x}(t) = \exp\big(tA\big)\vec{x}_0. \tag{23.18}$$

Remember, the exponential of a square matrix B is defined to be

$$\exp(B) := \sum_{n=0}^{\infty} \frac{B^n}{n!}. \tag{23.19}$$

Let us check this claim informally by differentiating[60]

$$\begin{aligned}\frac{d}{dt}\vec{x} &= \frac{d}{dt}\Big(\exp\big(tA\big)\vec{x}_0\Big)\\ &= \frac{d}{dt}\left(\sum_{n=0}^{\infty}\frac{t^nA^n}{n!}\vec{x}_0\right)\\ &= \sum_{n=0}^{\infty}\frac{d}{dt}\left(\frac{t^nA^n}{n!}\vec{x}_0\right)\\ &= \sum_{n=1}^{\infty}\frac{t^{n-1}A^n}{(n-1)!}\vec{x}_0\\ &= A\sum_{n=1}^{\infty}\frac{t^{n-1}A^{n-1}}{(n-1)!}\vec{x}_0\\ &= A\exp\big(tA\big)\vec{x}_0\\ &= A\vec{x}\end{aligned} \tag{23.20}$$

and

$$\vec{x}(0) = \exp\big(0A\big)\vec{x}_0 = \mathbb{1}_m\vec{x}_0 = \vec{x}_0 \tag{23.21}$$

so that this is indeed the solution. However, as we know, if we could diagonalize A via some matrix P as $A = PDP^{-1}$, then computing the exponential would be vastly simplified. This is because

$$\begin{aligned}\exp\big(tPDP^{-1}\big) &= \sum_{n=0}^{\infty}\frac{(tPDP^{-1})^n}{n!}\\ &= \sum_{n=0}^{\infty}\frac{t^nPD^nP^{-1}}{n!}\\ &= P\left(\sum_{n=0}^{\infty}\frac{t^nD^n}{n!}\right)P^{-1}\\ &= P\exp\big(tD\big)P^{-1}.\end{aligned} \tag{23.22}$$

So, if A is diagonalizable, and let's say the matrix D is of the form

$$D = \begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix} \tag{23.23}$$

with λ_1 and λ_2 the eigenvalues of A, then the solution to our ODE is given simply by

$$\vec{x}(t) = \exp\big(tA\big)\vec{x}_0 = P\exp\big(tD\big)P^{-1}\vec{x}_0 = P\begin{bmatrix}e^{\lambda_1 t} & 0\\ 0 & e^{\lambda_2 t}\end{bmatrix}P^{-1}\vec{x}_0. \tag{23.24}$$

Notice how much simpler the solution has become. Rather than calculating an infinite sum of increasingly complicated products, we only have to calculate three matrix products and then act on the vector to get the equation for $\vec{x}$ as a function of time.

[60] One subtle point is regarding convergence, but let us take for granted that this exponential function acting on matrices converges uniformly on any compact interval in time. This fact allows us to bring in the derivative with respect to t into the summation in the third line.

Example 23.25. Let us go back to Example 23.1 and solve it. Because the eigenvalues are all distinct, the matrix

$$A := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 2 \end{bmatrix} \tag{23.26}$$

is diagonalizable by using a basis of eigenvectors. They are

$$\lambda_1 = -1 \quad : \quad \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ -2 & 1 & 3 & 0 \end{array}\right] \mapsto \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \Rightarrow \quad \vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \tag{23.27}$$

$$\lambda_2 = 1 \quad : \quad \left[\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -2 & 1 & 1 & 0 \end{array}\right] \mapsto \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \Rightarrow \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \tag{23.28}$$

and

$$\lambda_3 = 2 \quad : \quad \left[\begin{array}{ccc|c} -2 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ -2 & 1 & 0 & 0 \end{array}\right] \mapsto \left[\begin{array}{ccc|c} 1 & 0 & -1/4 & 0 \\ 0 & 1 & -1/2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \Rightarrow \quad \vec{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}. \tag{23.29}$$

The matrix P that diagonalizes the matrix A is therefore

$$P = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \\ 1 & 1 & 4 \end{bmatrix}. \tag{23.30}$$

We don't really need to compute this all to see what the result is. The solution to the ODE is

$$\begin{bmatrix} f(x) \\ g(x) \\ h(x) \end{bmatrix} = \exp\left(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 2 \end{bmatrix} x \right) \begin{bmatrix} f_0 \\ g_0 \\ h_0 \end{bmatrix} = P \begin{bmatrix} e^{-x} & 0 & 0 \\ 0 & e^{x} & 0 \\ 0 & 0 & e^{2x} \end{bmatrix} P^{-1} \begin{bmatrix} f_0 \\ g_0 \\ h_0 \end{bmatrix} \tag{23.31}$$

where f_0, g_0, h_0 are the initial conditions (and appear as constants of integration). From this expression, we see that $f(x)$ is an arbitrary linear combination of the functions in the set $\{e^{-x}, e^{x}, e^{2x}\}$. This is what gives us a basis of solutions to the third order ODE we started with.

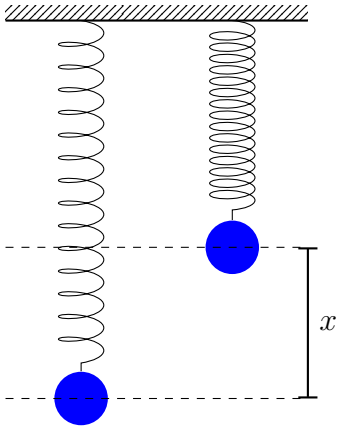
Often, we are not just interested in solving for a basis of solutions to an m-th order ODE but we might actually have some initial conditions and wish to know how the system evolves in time given those initial conditions.

Example 23.32. The position of a mass $m > 0$ on an ideal spring with spring constant $k > 0$ can be modeled by the differential equation[61]

$$m\frac{d^2x}{dt^2} = -kx \tag{23.33}$$

[61]The TikZ code for the image for the oscillators was obtained from https://tex.stackexchange.com/questions/41608/draw-mechanical-springs-in-tikz?utm_medium=organic&utm_source=google_rich_qa&utm_campaign=google_rich_qa by the author percusse. It has been slightly modified for our purposes.

This equation can come from considering many situations of a spring on a horizontal surface or a spring hanging vertically on a wall with x signifying the displacement from the equilibrium position of the spring (the effects due to gravity simply displace this equilibrium position but otherwise have no effect on the motion).



Although we can solve this differential equation by inspection, let us apply the techniques of linear algebra. Set $v := \frac{dx}{dt}$ so that the second order ODE becomes

$$\frac{d}{dt}\begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{bmatrix}\begin{bmatrix} x \\ v \end{bmatrix}. \tag{23.34}$$

The characteristic polynomial of the above matrix is

$$\lambda^2 + \frac{k}{m}. \tag{23.35}$$

The roots are

$$\lambda_1 = i\omega \qquad \& \qquad \lambda_2 = -i\omega, \tag{23.36}$$

where

$$\omega := \sqrt{\frac{k}{m}}. \tag{23.37}$$

From this, we know that the solutions are of the form $\{e^{i\omega t}, e^{-i\omega t}\}$ but these are complex functions. We started with a real physical problem so it is unreasonable to have complex solutions. We could say that only the real combinations of these are solutions, but we should see this coming directly out of the math. Let us suppose that the initial configuration of the system is $x(0) = x_0$ and $v(0) = v_0$. This corresponds to the mass having an initial displacement of x_0 from its equilibrium position and an initial velocity v_0. In this case, we need to compute the eigenvectors to solve this system fully. They are given by

$$\lambda_1 = i\omega \quad : \quad \left[\begin{array}{cc|c} -i\omega & 1 & 0 \\ -\omega^2 & -i\omega & 0 \end{array}\right] \mapsto \left[\begin{array}{cc|c} 1 & \frac{i}{\omega} & 0 \\ 0 & 0 & 0 \end{array}\right] \qquad \Rightarrow \qquad \vec{v}_1 = \begin{bmatrix} -i \\ \omega \end{bmatrix} \tag{23.38}$$

and

$$\lambda_2 = -i\omega \quad : \quad \left[\begin{array}{cc|c} i\omega & 1 & 0 \\ -\omega^2 & i\omega & 0 \end{array}\right] \mapsto \left[\begin{array}{cc|c} 1 & -\frac{i}{\omega} & 0 \\ 0 & 0 & 0 \end{array}\right] \qquad \Rightarrow \qquad \vec{v}_2 = \begin{bmatrix} i \\ \omega \end{bmatrix}. \tag{23.39}$$

The matrix P and its inverse that diagonalize our matrix are given by

$$P = \begin{bmatrix} -i & i \\ \omega & \omega \end{bmatrix} \qquad \& \qquad P^{-1} = \frac{1}{-2i\omega} \begin{bmatrix} \omega & -i \\ -\omega & -i \end{bmatrix} = \frac{1}{2\omega} \begin{bmatrix} i\omega & 1 \\ -i\omega & 1 \end{bmatrix}. \tag{23.40}$$

Therefore, the solution to the initial value problem is

$$\begin{aligned} \begin{bmatrix} x(t) \\ v(t) \end{bmatrix} &= \frac{1}{2\omega} \begin{bmatrix} -i & i \\ \omega & \omega \end{bmatrix} \begin{bmatrix} e^{i\omega t} & 0 \\ 0 & e^{-i\omega t} \end{bmatrix} \begin{bmatrix} i\omega & 1 \\ -i\omega & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ v_0 \end{bmatrix} \\ &= \frac{1}{2\omega} \begin{bmatrix} -ie^{i\omega t} & ie^{-i\omega t} \\ \omega e^{i\omega t} & \omega e^{-i\omega t} \end{bmatrix} \begin{bmatrix} i\omega & 1 \\ -i\omega & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ v_0 \end{bmatrix} \\ &= \frac{1}{2\omega} \begin{bmatrix} \omega(e^{i\omega t} + e^{-i\omega t}) & -i(e^{i\omega t} - e^{-i\omega t}) \\ i\omega^2(e^{i\omega t} - e^{-i\omega t}) & \omega(e^{i\omega t} + e^{-i\omega t}) \end{bmatrix} \begin{bmatrix} x_0 \\ v_0 \end{bmatrix} \\ &= \frac{1}{2\omega} \begin{bmatrix} 2\omega\cos(\omega t) & -2i^2\sin(\omega t) \\ 2\omega^2 i^2 \sin(\omega t) & 2\omega\cos(\omega t) \end{bmatrix} \begin{bmatrix} x_0 \\ v_0 \end{bmatrix} \\ &= \begin{bmatrix} \cos(\omega t) & \frac{1}{\omega}\sin(\omega t) \\ -\omega\sin(\omega t) & \cos(\omega t) \end{bmatrix} \begin{bmatrix} x_0 \\ v_0 \end{bmatrix} \\ &= \begin{bmatrix} x_0\cos(\omega t) - \frac{v_0}{\omega}\sin(\omega t) \\ -x_0\omega\sin(\omega t) + v_0\cos(\omega t) \end{bmatrix} \end{aligned} \tag{23.41}$$

so that our solution is the first component, which is

$$x(t) = x_0\cos(\omega t) - \frac{v_0}{\omega}\sin(\omega t). \tag{23.42}$$

Example 23.43. What about differential equations that have the same root? The characteristic polynomial associated to the differential equation

$$f'' - 2f' + f \tag{23.44}$$

is

$$\lambda^2 - 2\lambda + 1 = (\lambda - 1)^2. \tag{23.45}$$

One often solves this system by postulating that a linearly independent set of solutions is of the form $\{e^x, xe^x\}$, but why? To answer this, let us rewrite this second order ODE as a system of first order ODEs

$$\frac{d}{dx}\begin{bmatrix} f \\ g \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix} \tag{23.46}$$

where $g := f'$. The characteristic polynomial here is $(\lambda - 1)^2$ so the only eigenvalue is $\lambda_1 = 1$ and it has multiplicity 2. We can compute the eigenspace

$$\lambda_1 = 1 \quad : \quad \left[\begin{array}{cc|c} -1 & 1 & 0 \\ -1 & 1 & 0 \end{array}\right] \mapsto \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right] \qquad \Rightarrow \qquad \vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{23.47}$$

and we find that it is one-dimensional. Therefore, we cannot construct a matrix P that diagonalizes the system. Nevertheless, there *does* exist a matrix S such that

$$S\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} S^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}. \tag{23.48}$$

For example, one such matrix is

$$S = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \tag{23.49}$$

as you can check.[62] The usefulness of this is that we still have a rather simple formula for computing powers of our original matrix

$$\begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}^n = S \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^n S^{-1}. \tag{23.50}$$

To compute the n-th power of the matrix in the middle, it suffices to compute it for small values of n and the pattern immediately emerges.

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^3 = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^4 = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}. \tag{23.51}$$

Therefore, the exponential of this matrix times an arbitrary coefficient x is

$$\begin{aligned} \exp\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x\right) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x + \frac{1}{2!}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^2 x^2 + \frac{1}{3!}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^3 x^3 + \cdots \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} x & x \\ 0 & x \end{bmatrix} + \begin{bmatrix} \frac{x^2}{2!} & \frac{2x^2}{2!} \\ 0 & \frac{x^2}{2!} \end{bmatrix} + \begin{bmatrix} \frac{x^3}{3!} & \frac{3x^3}{3!} \\ 0 & \frac{x^3}{3!} \end{bmatrix} + \cdots \\ &= \begin{bmatrix} 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots & x + \frac{2x^3}{2!} + \frac{3x^2}{3!} + \cdots \\ 0 & 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \end{bmatrix} \\ &= \begin{bmatrix} e^x & x\left(1 + x + \frac{x^2}{2!} + \cdots\right) \\ 0 & e^x \end{bmatrix} \\ &= \begin{bmatrix} e^x & xe^x \\ 0 & e^x \end{bmatrix} \end{aligned} \tag{23.52}$$

and this is the first glimpse of xe^x appearing! Going back to our ODE, the general solution is given by

$$\begin{bmatrix} f(x) \\ g(x) \end{bmatrix} = \exp\left(\begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix} x\right) \begin{bmatrix} f_0 \\ g_0 \end{bmatrix} = S \begin{bmatrix} e^x & xe^x \\ 0 & e^x \end{bmatrix} S^{-1} \begin{bmatrix} f_0 \\ g_0 \end{bmatrix} \tag{23.53}$$

with f_0, g_0 arbitrary (integration) constants. We do not have to work out the matrix product here to see that our expression for $f(x)$ will be an arbitrary linear combination of e^x and xe^x.

We apply the previous results to completely solve a model describing a harmonic oscillator with friction (aka damping).

Example 23.54. Consider the ordinary differential equation in $\mathbb{R}^2$ of the form

$$\begin{aligned} \frac{dx}{dt} &= v \\ \frac{dv}{dt} &= -\frac{k}{m}x - \gamma v \end{aligned} \tag{23.55}$$

[62] Where this matrix comes from is a bit beyond the scope of this course. It is obtained by using what are called generalized eigenvectors. In this case, $\vec{v}_1$ is an eigenvector of A because it satisfies $(A - \lambda_1 \mathbb{1})\vec{v}_1 = \vec{0}$. Although a second linearly independent eigenvector does not exist, there does exist a generalized eigenvector $\vec{v}_2$ that satisfies $(A - \lambda_1 \mathbb{1})\vec{v}_2 = \vec{v}_1$ and the matrix S is obtained from $S = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$.

with initial condition (x_0, v_0) and $m > 0$ and $k, \gamma \geq 0$. This corresponds to the second order linear differential equation given by[63]

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \frac{k}{m}x = 0. \tag{23.56}$$

This system describe a one-dimensional oscillator with friction. The x variable is interpreted as the position and the v variable as the velocity, k is a constant dictating the strength of the spring, m is the mass of the particle that is affected by the spring force and the frictional force, and γ is a constant related to the strength of the frictional force. $\gamma = 0$ corresponds to no friction and $k = 0$ corresponds to having no spring. The name for this model is the *damped harmonic oscillator*. We will not plug in numbers for k, m, and γ because we want to study what happens to the motion depending on how strong the frictional constant is compared to the spring constant and mass. The matrix associated to this system is

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\gamma \end{bmatrix} \tag{23.57}$$

because then the differential equation (23.55) can be expressed as a vector equation

$$\frac{d}{dt}\begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\gamma \end{bmatrix}\begin{bmatrix} x \\ v \end{bmatrix}. \tag{23.58}$$

We will now work out this general theory for the example ODE given by (23.55). The eigenvalues of this system are therefore given by solving

$$0 = \det \begin{bmatrix} -\lambda & 1 \\ -\frac{k}{m} & -\gamma - \lambda \end{bmatrix} = \lambda(\gamma + \lambda) + \frac{k}{m} = \frac{k}{m} + \gamma\lambda + \lambda^2, \tag{23.59}$$

which has solutions

$$\lambda = -\frac{\gamma}{2} \pm \frac{1}{2}\sqrt{\gamma^2 - \frac{4k}{m}}. \tag{23.60}$$

Let's use the notation

$$\lambda_1 := \frac{1}{2}\left(-\gamma - \sqrt{\gamma^2 - \frac{4k}{m}}\right) \qquad \& \qquad \lambda_2 := \frac{1}{2}\left(-\gamma + \sqrt{\gamma^2 - \frac{4k}{m}}\right) \tag{23.61}$$

for these two possibilities for the eigenvalues. The corresponding eigenvectors would therefore be solutions to the augmented matrix problem

$$\left[\begin{array}{cc|c} -\lambda & 1 & 0 \\ -\frac{k}{m} & -\gamma - \lambda & 0 \end{array}\right] \tag{23.62}$$

for each λ. Rather than plugging in these values right away, we can first simplify. Using the first row, we find that

$$-\lambda x + y = 0 \iff y = \lambda x \tag{23.63}$$

[63] To see this, set $v := \frac{dx}{dt}$, which is the first line of (23.55), and notice that this second order differential equation reads $\frac{dv}{dt} = -\frac{k}{m}x - \gamma v$, which is the second line of (23.55).

so that we can set $x = 1$ to obtain a pair of eigenvectors of the form

$$\vec{v}_1 = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} \qquad \& \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix} \tag{23.64}$$

for λ_1 and λ_2, respectively. Note that if $\lambda_1 = \lambda_2$, we will not be able to find a second eigenvector. We will discuss this case later. In the case that $\lambda_1 \neq \lambda_2$, the similarity transformation matrix corresponding to this choice of eigenvectors is given by

$$P = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} \tag{23.65}$$

and the inverse of this matrix is

$$P^{-1} = \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix}. \tag{23.66}$$

For what follows, it helps to note that

$$\lambda_1 \lambda_2 = \frac{1}{4}\left(\gamma + \sqrt{\gamma^2 - \frac{4k}{m}}\right)\left(\gamma - \sqrt{\gamma^2 - \frac{4k}{m}}\right) = \frac{k}{m} \tag{23.67}$$

and

$$\lambda_1 + \lambda_2 = -\gamma. \tag{23.68}$$

With these facts, you can check (do it!) that this similarity transformation enables us to express A as

$$\begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\gamma \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix}. \tag{23.69}$$

Therefore, the exponential of this matrix with the variable t appended to it is given by

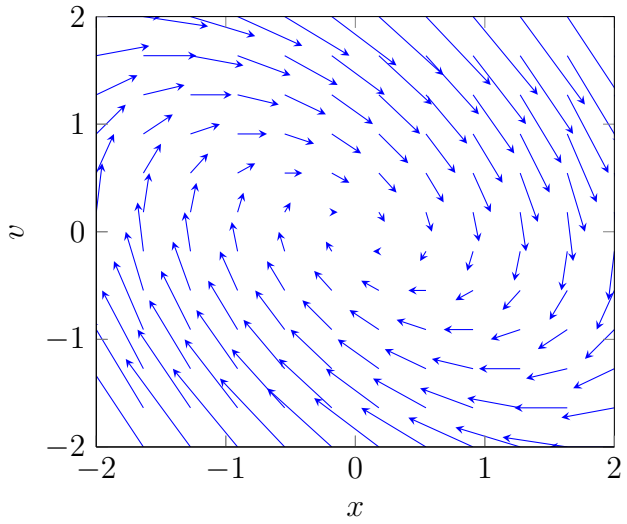
$$\exp(tA) = \exp(tPDP^{-1}) = P\exp(tD)P^{-1} = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix} \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix} \tag{23.70}$$

Multiplying this expression out, we get

$$\begin{aligned} \exp(tA) &= \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} e^{\lambda_1 t} & e^{\lambda_2 t} \\ \lambda_1 e^{\lambda_1 t} & \lambda_2 e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix} \\ &= \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t} & e^{\lambda_2 t} - e^{\lambda_1 t} \\ \lambda_1 \lambda_2 (e^{\lambda_1 t} - e^{\lambda_2 t}) & \lambda_2 e^{\lambda_2 t} - \lambda_1 e^{\lambda_1 t} \end{bmatrix} \\ &= \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t} & e^{\lambda_2 t} - e^{\lambda_1 t} \\ \frac{k}{m}\left(e^{\lambda_1 t} - e^{\lambda_2 t}\right) & \lambda_2 e^{\lambda_2 t} - \lambda_1 e^{\lambda_1 t} \end{bmatrix}. \end{aligned} \tag{23.71}$$

If $\gamma > 0$, there are three cases to consider depending on the value of $\gamma^2 - \frac{4k}{m}$, which appears inside a squareroot. Either this is negative (complex eigenvalues), positive (real eigenvalues), or zero (degenerate case when $\lambda_1 = \lambda_2$). We will discuss these three different cases one at a time in order.

i. ($\gamma^2 < \frac{4k}{m}$ underdamped case) When $\gamma = 1$ and $k = m$, the vector field associated to this system looks like (after rescaling)



In this graph, we are plotting the vector $\frac{d}{dt}\begin{bmatrix}x(t)\\v(t)\end{bmatrix}$ at each point $\begin{bmatrix}x(t)\\v(t)\end{bmatrix}$. In this case, the eigenvalues are complex and distinct. It is convenient to set

$$\alpha := \frac{\gamma}{2} \qquad \& \qquad \omega := \frac{1}{2}\sqrt{\frac{4k}{m} - \gamma^2} \tag{23.72}$$

so that

$$\lambda_1 = -\alpha - i\omega \qquad \& \qquad \lambda_2 = -\alpha + i\omega. \tag{23.73}$$

With these substitutions, the solution becomes

$$\begin{aligned}
\exp(tA) &= \frac{1}{2i\omega}\begin{bmatrix} e^{-\alpha t}\left((i\omega-\alpha)e^{-i\omega t} + (i\omega+\alpha)e^{i\omega t}\right) & e^{-\alpha t}\left(e^{i\omega t} - e^{-i\omega t}\right) \\ \frac{k}{m}e^{-\alpha t}\left(e^{-i\omega t} - e^{i\omega t}\right) & e^{-\alpha t}\left((i\omega-\alpha)e^{i\omega t} + (i\omega+\alpha)e^{-i\omega t}\right)\end{bmatrix} \\
&= \frac{1}{2i\omega}\begin{bmatrix} e^{-\alpha t}\left(2i\omega\cos(\omega t) + 2i\alpha\sin(\omega t)\right) & 2ie^{-\alpha t}\sin(\omega t) \\ -\frac{2ik}{m}e^{-\alpha t}\sin(\omega t) & e^{-\alpha t}\left(2i\omega\cos(\omega t) - 2i\alpha\sin(\omega t)\right)\end{bmatrix} \\
&= \begin{bmatrix} e^{-\alpha t}\left(\cos(\omega t) + \frac{\alpha}{\omega}\sin(\omega t)\right) & \frac{1}{\omega}e^{-\alpha t}\sin(\omega t) \\ -\frac{k}{m\omega}e^{-\alpha t}\sin(\omega t) & e^{-\alpha t}\left(\cos(\omega t) - \frac{\alpha}{\omega}\sin(\omega t)\right)\end{bmatrix}
\end{aligned} \tag{23.74}$$

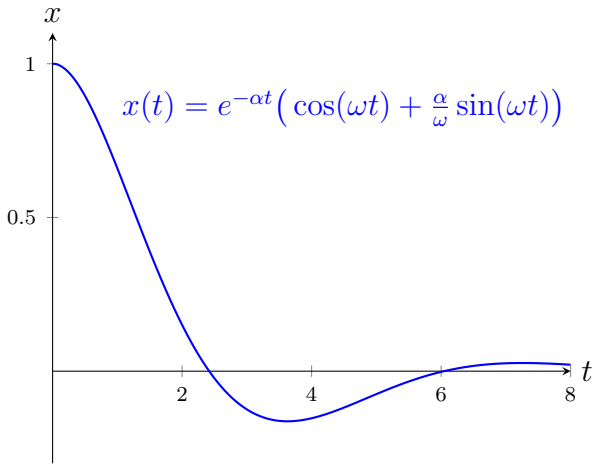
In this derivation, we have used the fact that

$$e^{i\omega t} = \cos(\omega t) + i\sin(\omega t). \tag{23.75}$$

Therefore, plugging in this result to our initial condition gives the complete trajectory both in position and velocity

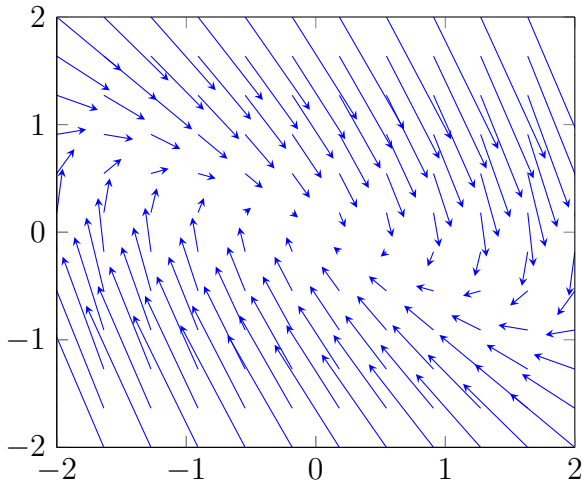
$$\begin{aligned}
\begin{bmatrix}x(t)\\v(t)\end{bmatrix} &= \begin{bmatrix} e^{-\alpha t}\left(\cos(\omega t) + \frac{\alpha}{\omega}\sin(\omega t)\right) & \frac{1}{\omega}e^{-\alpha t}\sin(\omega t) \\ -\frac{k}{m\omega}e^{-\alpha t}\sin(\omega t) & e^{-\alpha t}\left(\cos(\omega t) - \frac{\alpha}{\omega}\sin(\omega t)\right)\end{bmatrix}\begin{bmatrix}x_0\\v_0\end{bmatrix} \\
&= \begin{bmatrix} x_0 e^{-\alpha t}\left(\cos(\omega t) + \frac{\alpha}{\omega}\sin(\omega t)\right) + \frac{v_0}{\omega}e^{-\alpha t}\sin(\omega t) \\ -\frac{kx_0}{m\omega}e^{-\alpha t}\sin(\omega t) + v_0 e^{-\alpha t}\left(\cos(\omega t) - \frac{\alpha}{\omega}\sin(\omega t)\right).\end{bmatrix}
\end{aligned} \tag{23.76}$$

For example, with initial conditions $x_0 = 1$ and $v_0 = 0$, which corresponds to letting the oscillator go without giving it an initial velocity, the position as a function of time is given by ($\gamma = 2$ and $k = 2m$ so that $\alpha = 1$ and $\omega = 1$ in this graph)



ii. ($\gamma^2 > \frac{4k}{m}$ overdamped case) This case is left as an exercise.

iii. ($\gamma^2 = \frac{4k}{m}$ critically damped) This is the degenerate case where the matrix for the system is not diagonalizable. Let us again use the notation $\alpha := \frac{\gamma}{2}$. The vector field associated to this system looks like (after rescaling)



Nevertheless, it can be put into Jordan normal form,[64] and a matrix that accomplishes this is

$$P = \begin{bmatrix} -\alpha^{-1} & -\alpha^{-2} \\ 1 & 0 \end{bmatrix} \tag{23.77}$$

with a Jordan matrix given by

$$J = \begin{bmatrix} -\alpha & 1 \\ 0 & -\alpha \end{bmatrix}. \tag{23.78}$$

[64] You will not be responsible for knowing how to find a Jordan normal decomposition. This case is merely worked out to illustrate that the degenerate case can still be solved using matrix methods.

A matrix J is in Jordan normal form iff it is of the form

$$J = \begin{bmatrix} \lambda_1 & * & & & & & \\ & \ddots & * & & & & \\ & & \lambda_1 & & & & \\ & & & \ddots & & & \\ & & & & \lambda_k & * & \\ & & & & & \ddots & * \\ & & & & & & \lambda_k \end{bmatrix} \tag{23.79}$$

where $*$ are either 0 or 1 and every other off-diagonal term is 0. The inverse of P is given by

$$P^{-1} = \alpha^2 \begin{bmatrix} 0 & \alpha^{-2} \\ -1 & -\alpha^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\alpha^2 & -\alpha \end{bmatrix}. \tag{23.80}$$

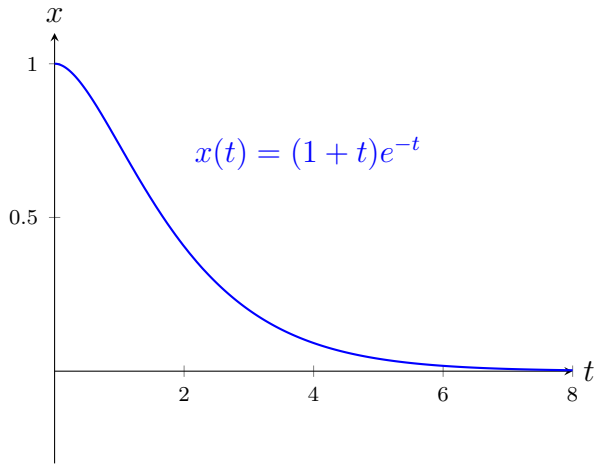
Indeed, you can check that $A = PJP^{-1}$. The exponential of tJ is given by

$$\begin{aligned} \exp(tJ) &= \exp\left(\begin{bmatrix} -\alpha t & t \\ 0 & -\alpha t \end{bmatrix}\right) \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -\alpha t & t \\ 0 & -\alpha t \end{bmatrix} + \frac{1}{2!}\begin{bmatrix} (-\alpha t)^2 & -2\alpha t \\ 0 & (-\alpha t)^2 \end{bmatrix} + \frac{1}{3!}\begin{bmatrix} (-\alpha t)^3 & 3\alpha^2 t^3 \\ 0 & (-\alpha t)^3 \end{bmatrix} + \cdots \\ &= \begin{bmatrix} 1 + (-\alpha t) + \frac{(-\alpha t)^2}{2!} + \frac{(-\alpha t)^3}{3!} + \cdots & t\left(1 + (-\alpha t) + \frac{(-\alpha t)^2}{2!} + \cdots\right) \\ 0 & 1 + (-\alpha t) + \frac{(-\alpha t)^2}{2!} + \frac{(-\alpha t)^3}{3!} + \cdots \end{bmatrix} \\ &= \begin{bmatrix} e^{-\alpha t} & te^{-\alpha t} \\ 0 & e^{-\alpha t} \end{bmatrix} \end{aligned} \tag{23.81}$$

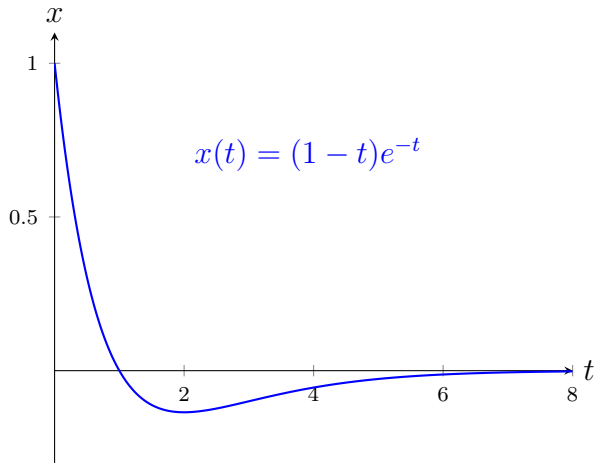
for all $t \in \mathbb{R}$. Therefore,

$$\begin{aligned} \exp(tA) &= \begin{bmatrix} -\alpha^{-1} & -\alpha^{-2} \\ 1 & 0 \end{bmatrix}\begin{bmatrix} e^{-\alpha t} & te^{-\alpha t} \\ 0 & e^{-\alpha t} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -\alpha^2 & -\alpha \end{bmatrix} \\ &= \begin{bmatrix} -\alpha^{-1} & -\alpha^{-2} \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -\alpha^2 te^{-\alpha t} & (1-\alpha t)e^{-\alpha t} \\ -\alpha^2 e^{-\alpha t} & -\alpha e^{-\alpha t} \end{bmatrix} \\ &= \begin{bmatrix} (1+\alpha t)e^{-\alpha t} & te^{-\alpha t} \\ -\alpha^2 te^{-\alpha t} & (1-\alpha t)e^{-\alpha t} \end{bmatrix}. \end{aligned} \tag{23.82}$$

Notice that there are no oscillatory terms and any initial configuration with zero velocity asymptotically approaches 0 without passing through 0. When $\alpha = 1$, and $x_0 = 1$ and $v_0 = 0$, the trajectory looks like



However, if there is an initial sufficiently strong "kick," the trajectory will pass through the origin once (but only once!). Such a trajectory occurs, for instance, when $x_0 = 1$ and $v_0 = -2$ (with $k = m$ and $\gamma = 2$) and is depicted here



Recommended Exercises. Please check HuskyCT for the homework. Please show your work! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we covered Section 5.7 in [Lay] and additional topics not covered in [Lay].

24 Vector spaces and linear transformations

The applicability of the subject matter of linear algebra goes far beyond the analysis of vectors in finite-dimensional real (and complex) Euclidean space. Many of the constructions, definitions, and arguments we gave had nothing to do with the specific structure associated with Euclidean space. Take for example the definition of a linear transformation. A function $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ is a linear transformation iff it satisfies

$$T(\vec{u}+\vec{v}) = T(\vec{v}) + T(\vec{w}) \qquad \& \qquad T(c\vec{v}) = cT(\vec{v}) \tag{24.1}$$

for all numbers c and vectors $\vec{u}, \vec{v}$. The definition looks the same whether we use a function $\mathbb{R}^5 \xleftarrow{T} \mathbb{R}^8$ or a function $\mathbb{R}^{19} \xleftarrow{T} \mathbb{R}^{34}$. Furthermore, the only structure needed to write down the above condition is that of being able to add vectors and the ability to multiply vectors by scalars. Before giving the abstract definitions of vector spaces, linear transformations, etc., an example should help motivate why we might want to do this.

Example 24.2. A real *degree m polynomial* is a function $p : \mathbb{R} \to \mathbb{R}$ of a single real variable of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m, \tag{24.3}$$

where $a_0, a_1, a_2, \ldots, a_m$ are real numbers. We do not place any restriction on these numbers (therefore, a degree n polynomial is also a degree $n+1$ polynomial but not vice-versa). A degree zero polynomial is a constant, a degree one polynomial is a linear function, and a degree two polynomial is a quadratic function. For example, $p(x) = 5x^4 - 3x^3 + 2x - 4$ is a degree 4 polynomial in the variable x. The sum of a degree m polynomial p as above and a degree n polynomial q given by

$$q(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n \tag{24.4}$$

is the degree $M := \max\{m, n\}$ polynomial $p + q$, which is given by

$$(p+q)(x) := p(x) + q(x) \equiv (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_M + b_M)x^M. \tag{24.5}$$

In this notation, if $n > m$, then $a_{m+1}, \ldots, a_M$ are defined to be zero (and vice versa if $m > n$). For example, if $q(x) = 7x^6 - 3x^4 + x^3 - 2x$, then $(p+q)(x) = 7x^6 + 2x^4 - 2x^3 - 4$. If c is a real number, then cp is the degree m polynomial given by

$$(cp)(x) := cp(x) \equiv ca_0 + ca_1 x + ca_2 x^2 + \cdots + ca_m x^m. \tag{24.6}$$

For example, if $c = 3$, then $(cp)(x) = 25x^4 - 15x^3 + 6x - 12$ for the example we have been using. Therefore, polynomials can be added and scaled just as vectors in Euclidean space. Moreover, the zero polynomial, which is just the polynomial whose coefficients are all zero, i.e. it's the constant function whose value is always zero, satisfies the same conditions as the zero vector does. Namely, if you add it to any polynomial, you get that polynomial back. If you wanted to make sense of a basis for all polynomials, you would want to be sure that any polynomial can be written as a unique linear combination of the polynomials of said basis. One such basis could be the set of monomials, $\{p_0, p_1, p_2, p_3, \ldots\}$, where

$$p_k(x) := x^k. \tag{24.7}$$

Notice that this set of monomials has infinitely many polynomials. Such a set of monomials spans the set of all polynomials because every polynomial is a (finite) linear combination of such monomials. Furthermore, the set of these monomials is linearly independent because no monomial is a (finite) linear combination of other monomials. An example of a linear transformation on the set of polynomials is the derivative operation. Recall, for $k > 0$,

$$p_k'(x) \equiv \frac{d}{dx}\big(p_k(x)\big) \equiv \frac{d}{dx}\big(x^k\big) = kx^{k-1}. \tag{24.8}$$

What about

$$\frac{d}{dx}\big(x^k + x^l\big) \qquad \& \qquad \frac{d}{dx}\big(cx^k\big)? \tag{24.9}$$

This is something usually covered in a course on calculus, but from the definition of the derivative (in terms of limits), you can prove that

$$\frac{d}{dx}\big(x^k + x^l\big) = \frac{d}{dx}\big(x^k\big) + \frac{d}{dx}\big(x^l\big) \qquad \& \qquad \frac{d}{dx}\big(cx^k\big) = c\frac{d}{dx}\big(x^k\big). \tag{24.10}$$

This is true no matter what polynomial you plug in:

$$\frac{d}{dx}\big(p(x) + q(x)\big) = \frac{d}{dx}p(x) + \frac{d}{dx}q(x) \qquad \& \qquad \frac{d}{dx}\big(cp(x)\big) = c\frac{d}{dx}\big(p(x)\big). \tag{24.11}$$

This just means that taking the derivative is a linear transformation. But to make sense of it being a linear transformation, we have to specify its source and target. What are we allowed to take derivatives of in this case? If we write $\mathbb{P}$ as the set of all polynomials, then we can think of the derivative as a function $\frac{d}{dx} : \mathbb{P} \to \mathbb{P}$ on polynomials. We could analyze this linear transformation in many ways. For instance, what is its kernel? The analogue of the definition of kernel would say that

$$\ker\left(\frac{d}{dx}\right) = \left\{p \in \mathbb{P} \ : \ \frac{d}{dx}p(x) = 0\right\}, \tag{24.12}$$

so the set of all polynomials whose first derivatives are the zero function. The only polynomials whose derivatives are the zero function are exactly the constant functions. Therefore,[65]

$$\ker\left(\frac{d}{dx}\right) = \text{span}\{p_0\} \tag{24.13}$$

since p_0 is the function representing $p_0(x) = x^0 = 1$. What is the range/image of $\frac{d}{dx}$?[66] By definition, the range/image would be

$$\text{image}\left(\frac{d}{dx}\right) = \left\{\frac{d}{dx}q(x) \ : \ q \in \mathbb{P}\right\}, \tag{24.14}$$

[65] It is more precise to write $\text{span}\{p_0\}$ instead of $\text{span}\{1\}$ because we want to make sure we understand that an entire *function* is being represented inside the span and not just a number. For example, $\text{span}\{x\}$ could be interpreted as cx for all numbers $c \in \mathbb{R}$ but if x is a fixed number (and not a variable), then this represents just all possible numbers. However, $\text{span}\{p_1\}$, where p_1 is the *function* defined by $p_1(x) = x$ for *all* x, is more accurate and less ambiguous. It might seem strange to write it this way if you are not used to thinking of functions as rules. The main point is that $f(x)$ is *not* a function—$f(x)$ is the *value* of the *function* f *at* x. I might be sloppy on occasion, however, and call $f(x)$ the function even though it's really f because this tends to confuse most students (even though it confuses me!).

[66] Notice, by the way, that we can't use the terminology "column space' because we do not have a matrix anymore.

i.e. the derivatives of all possible polynomials. Is $\frac{d}{dx}$ surjective/onto? Before we answer this, it helps to understand what this question is asking. $\frac{d}{dx}$ is surjective iff for each polynomial p there exists a polynomial q such that $\frac{d}{dx}q(x) = p(x)$. Hence, to answer this question if $\frac{d}{dx}$ is surjective, we would try to solve the following problem. If $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$ is some polynomial, can we find a polynomial q whose derivative is p? As an equation, we want to find a q such that

$$\frac{d}{dx}q(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m. \tag{24.15}$$

Finding such a polynomial can be obtained by integrating p. Namely, take q to be

$$q(x) = a_0x + \frac{a_1}{2}x^2 + \frac{a_2}{3}x^3 + \cdots + \frac{a_m}{m+1}x^{m+1}. \tag{24.16}$$

Then, you can check that $\frac{d}{dx}q(x) = p(x)$ for all $x \in \mathbb{R}$. Of course, there are many other possibilities, one for each integration constant. Therefore, because we can always integrate polynomials,

$$\text{image}\left(\frac{d}{dx}\right) = \mathbb{P}, \tag{24.17}$$

the image of $\frac{d}{dx}$ is actually all of $\mathbb{P}$, i.e. $\frac{d}{dx}$ is onto/surjective. The set of polynomials with their usual notion of addition and scalar multiplication form what is known as a vector space, a mathematical object that formalizes the structure of of vector addition and scalar multiplication along with its properties resembling those of Euclidean vectors and polynomials.

We will continue giving several examples, but let us introduce the abstract concepts one at a time.

Definition 24.18. A real (complex) *vector space* consists of a set V, the elements of which are called *vectors*, together with

i) a binary operation $+ : V \times V \to V$, called *addition* and written as $\vec{u} + \vec{v}$

ii) a vector $\vec{0}$ in V, called the *zero vector*

iii) a function $\mathbb{R} \times V \to V$ ($\mathbb{C} \times V \to V$) called *scalar multiplication* and written as $c\vec{u}$

satisfying

(a) addition is commutative: $\vec{u} + \vec{v} = \vec{v} + \vec{u}$

(b) addition is associative: $\vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w}$

(c) the zero vector is an identity for addition: $\vec{u} + \vec{0} = \vec{u}$

(d) addition is invertible: for each vector $\vec{u}$, there is a vector $\vec{v}$ such that $\vec{u} + \vec{v} = \vec{0}$

(e) scalar multiplication is distributive over vector addition: $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$

(f) scalar multiplication is distributive over scalar addition: $(c + d)\vec{u} = c\vec{u} + d\vec{u}$

(g) scalar multiplication is distributive over itself: $c(d\vec{u}) = (cd)\vec{u}$

(h) the scalar unit axiom: $1\vec{u} = \vec{u}$

If in a particular definition or statement it does not matter whether the real or complex numbers are used, the terminology "vector space" will be used more generically.[67] Depending on the context, we might not write arrows over our vectors. For example, we usually do not write arrows over polynomials and other functions, which can be viewed as vectors in some vector space.

To understand why definitions are the way they are written, imagine trying to define a piece of chalk.[68] What is a piece of chalk other than the properties and structure that characterize it? For example, chalk is made mostly of calcium, but this is not enough to define it. It serves the function of a writing utensil on certain surfaces such as blackboards. It can come in a variety of colors. It is often in cylindrical shape. To really specify what chalk is, we must keep describing its characterizing features. However, we do not want to specify so much that we identify only one particular chalk in the universe. Instead, we wish to identify the characterizing features so that any chalk can be placed into this set, but so that nothing else besides different pieces of chalk are in this set. Identifying these characterizing features is what goes behind setting up a mathematical definition (and also what goes behind image recognition software). Features such as "color" might not be so relevant when one is merely interested in a writing utensil for blackboards. Hence, we would not include these in a definition of chalk—we would instead save that for the definition of chalk of a particular color.

You might also ask: why do we need such an abstract definition? What do we gain? The pay-off is actually phenomenal. As we will see shortly, there are many examples of vector spaces. Rather than proving something about each and every one of these examples, if we can prove or discover something for general vector spaces, then all of these results will hold for every single example. The reason is because most of the concepts we have learned about vectors and linear transformations in Euclidean space have completely natural analogues for vector spaces.

For example, here is a list of some concepts/definitions that have natural analogues for arbitrary vector spaces: linear combinations, linear independence, span, linear transformations, subspace, basis, dimension, image, kernel, composing linear transformations (in succession), projections,[69] inverses,[70] eigenvectors, eigenvalues, and eigenspaces. We will discuss all of these definitions in their generality in the next few sections.

There are some other concepts/definitions that we have studied that do *not* have immediate (or straightforward) analogues for arbitrary vector spaces and linear transformations. These include, but are not limited to: the determinant (and therefore the characteristic polynomial).

Remark 24.19. Besides these concepts, there are also some definitions that do not make sense without additional structure on the vector space. For example, it is not clear how to define:

[67] In the example of Hamming's error correcting codes, we manipulated the numbers $\{0, 1\}$ in much the same way as real or complex numbers but with a different rule for addition. All our vectors and matrices had entries that were all either 0 or 1. In fact, one can extend the definition of a real (complex) vector space to the notion of a vector space over a field. $\mathbb{Z}_2$ is an example of a field. The axioms for a field are comparable in complication to the definition of a vector space and will therefore be omitted since most of the examples of vector spaces that we will be dealing with from now are real or complex.

[68] I learned this analogy from Prof. Balakrishnan at IIT Madras.

[69] But not orthogonal projections—keep reading.

[70] The notion of an inverse exists, but it is a tiny bit more subtle in infinite dimensions.

orthogonality, orthogonal projections, the transpose of a linear transformation, least squares approximations, diagonalization, and spectral decomposition. To define these, we will need the notion of an inner product as well. As this requires its own section, we will not be able to cover it here.

Rather than spending all of our time and wasting space redefining all of the concepts that do have analogues for arbitrary vector spaces, it seems more appropriate to give examples to illustrate the broad scope.

Example 24.20. Any subspace of $\mathbb{R}^n$ is a real vector space with the induced addition, zero vector, and scalar multiplication. For example, if $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ is a linear transformation, then $\ker(T)$ and $\text{image}(T)$ are both real vector spaces. We are very familiar with these kinds of examples, but notice that almost none of the subspace of $\mathbb{R}^n$ are $\mathbb{R}^k$ for some k. For example, the xy-plane in $\mathbb{R}^3$ is a 2-dimensional vector space, but we would not say it is $\mathbb{R}^2$. This is because we describe it using three coordinates with the third one being zero.

Here is a rather strange example of a vector space.

Example 24.21. The set of real $m \times n$ matrices is a real vector space with addition given by component-wise addition, the zero vector is the zero matrix, and the scalar multiplication is given by distributing the scalar to each component. This vector space is denoted by ${}_m\mathcal{M}_n$. Taking the transpose of an $m \times n$ matrix defines a function ${}_n\mathcal{M}_m \xleftarrow{T} {}_m\mathcal{M}_n$ defined by

$$ {}_m\mathcal{M}_n \ni A \mapsto A^T \in {}_n\mathcal{M}_m. \tag{24.22}$$

Is this function linear? Linearity would say

$$ (A+B)^T = A^T + B^T \qquad \& \qquad (cA)^T = cA^T \tag{24.23}$$

for all $A, B \in {}_m\mathcal{M}_n$ and for all $c \in \mathbb{R}$. A quick check shows that this is true.

Exercise 24.24. Let

$$ Q := \begin{bmatrix} 1 & -2 \\ 0 & 3 \\ -1 & 1 \end{bmatrix} \tag{24.25}$$

and let ${}_3\mathcal{M}_4 \xleftarrow{T} {}_2\mathcal{M}_4$ be the function defined by

$$ T(A) = QA \tag{24.26}$$

for all 2×4 matrices A.

(a) Prove that T is a linear transformation.

(b) Find the kernel of T.

(c) Find the image of T.

In fact, if Q is any $m \times n$ matrix, then the function ${}_m\mathcal{M}_k \xleftarrow{T} {}_n\mathcal{M}_k$ defined by sending $A \in {}_n\mathcal{M}_k$ to $T(A) := QA$ is a linear transformation for all $k \in \mathbb{N}$.

The example of polynomials given at the beginning is another example of a vector space. The following example generalizes this example quite a bit.

Example 24.27. More generally, functions from $\mathbb{R}$ to $\mathbb{R}$ form a vector space in the following way. Let f and g be two functions. Then $f + g$ is the function defined by

$$(f + g)(x) := f(x) + g(x). \tag{24.28}$$

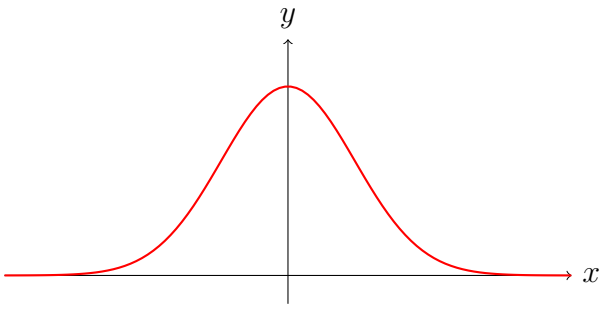
If c is a real number, then cf is the function defined by

$$(cf)(x) := cf(x). \tag{24.29}$$

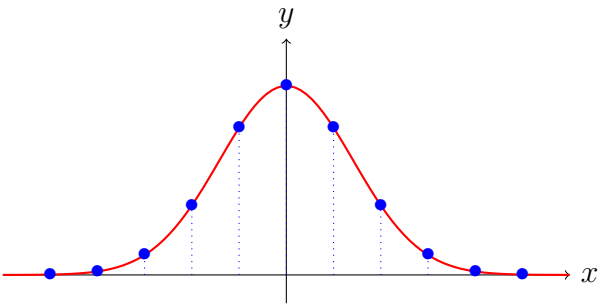
The zero function is the function 0 defined by

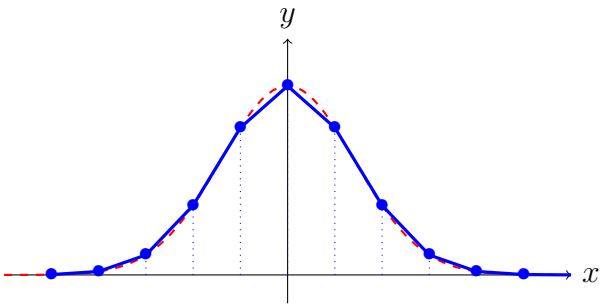
$$0(x) := 0. \tag{24.30}$$

It might seem complicated to think of functions as vectors. In fact, we can think of functions as vectors with *infinitely many components*. To see this, imagine taking a function, such as the Gaussian e^{-x^2}.

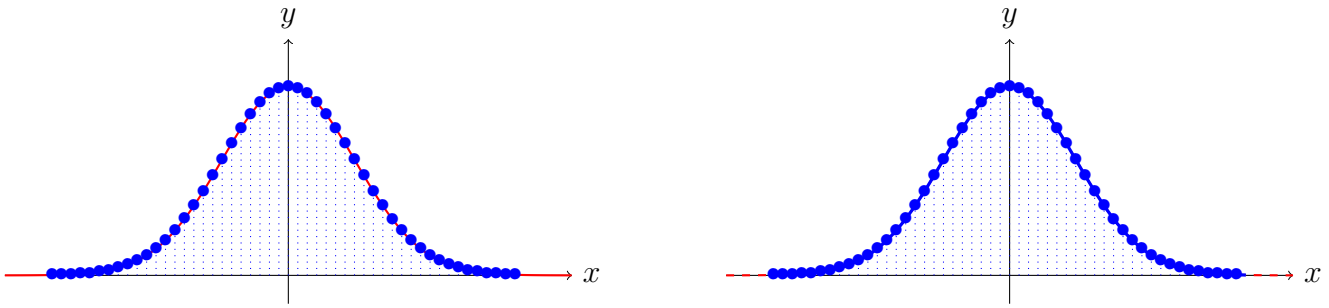


If you wanted to plug in the data for this function on a computer for instance, you wouldn't give the computer an infinite number of values since that just wouldn't be possible. You could specify certain values of this function at certain positions such as

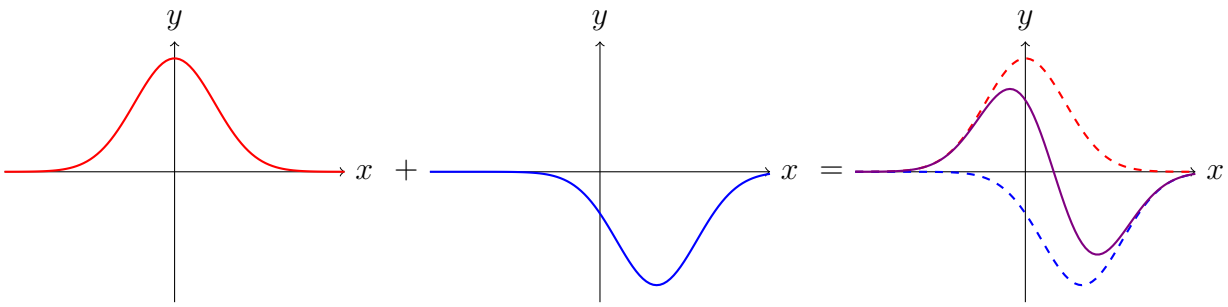




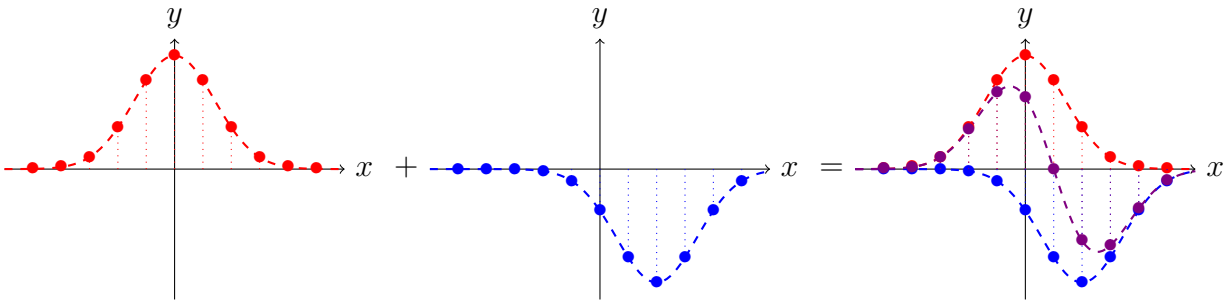
For example, the Gaussian in this picture on the left could be represented by a vector with 11 components (since there are 11 values chosen). Then you can piece them together to get a rough image of the function by linear extrapolation as shown above on the right. The more values you keep, the better your approximation.



In this case, we have a vector with 51 components. As you can see, you will never quite reach the *exact* Gaussian because you will always be using a finite approximation. This is why we view the set of functions as a vector space instead. We can visualize adding functions as follows. For example, adding the Gaussian centered at 0 to a negative Gaussian centered at 1 might look something like the following



where the sum is drawn in purple. We could view this as summing components of vectors by using an approximation as above



Example 24.31. The set $\mathbb{P}_2$ of all degree 2 polynomials, i.e. functions of the form

$$a_0 + a_1 x + a_2 x^2 \tag{24.32}$$

is a subspace of the vector space $\mathbb{P}_3$ of all degree 3 polynomials, i.e. functions of the form

$$b_0 + b_1 x + b_2 x^2 + b_3 x^3. \tag{24.33}$$

In fact, $\mathbb{P}_2$ is a subspace of $\mathbb{P}_n$ for every $n \geq 2$. Even more generally, $\mathbb{P}_k$ is a subspace of $\mathbb{P}$ for all natural numbers k.

Example 24.34. For each natural number n, let f_n and g_n be the functions on $[0,1]$ given by

$$f_n(x) := \cos(2\pi n x) \qquad \text{and} \qquad g_n(x) := \sin(2\pi n x). \tag{24.35}$$

Then, the set of all (finite) linear combinations of such sines and cosines is a subspace of the set of all continuous functions on $[0,1]$. The zero vector is g_0 because $\sin 0 = 0$. Such functions are useful in expressing periodic functions. For example, the linear combination

$$\mathfrak{h} := \frac{4}{\pi}\left(g_1 + \frac{1}{3}g_3 + \frac{1}{5}g_5 + \frac{1}{7}g_7\right) \tag{24.36}$$

given at x by

$$\mathfrak{h}(x) := \frac{4}{\pi}\left(\sin(2\pi x) + \frac{1}{3}\sin(6\pi x) + \frac{1}{5}\sin(10\pi x) + \frac{1}{7}\sin(14\pi x)\right) \tag{24.37}$$

is a decent approximation to the "square wave" piece-wise function given by

$$h(x) := \begin{cases} 1 & \text{for } 0 \le x \le \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} \le x \le 1 \end{cases}. \tag{24.38}$$

See Figure 14 for an illustration of this. We denote the set of all linear combinations of the $\{f_n\}$

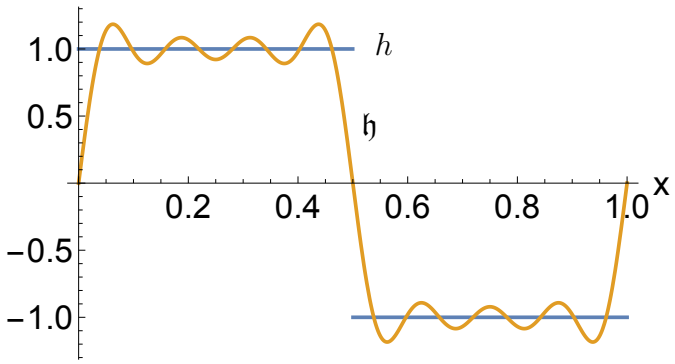


Figure 14: Plots of the step function h (in blue) and an approximation $\mathfrak{h}$ (in orange).

and $\{g_n\}$ by $\mathcal{F}$ (for Fourier),

$$\mathcal{F} := \left\{ \sum_{n=0}^{\text{finite}} a_n f_n + \sum_{m=0}^{\text{finite}} b_m g_m \ : \ a_n, b_m \in \mathbb{R} \right\}. \tag{24.39}$$

Example 24.40. Let $(a_1, a_2, a_3, \dots)$ be a sequence of real numbers satisfying

$$\sum_{n=1}^{\infty} |a_n| < \infty, \tag{24.41}$$

(recall, this means $\sum_{n=1}^{\infty} a_n$ is an absolutely convergent series). Given another sequence of real numbers $(b_1, b_2, b_3, \dots)$ satisfying

$$\sum_{n=1}^{\infty} |b_n| < \infty, \tag{24.42}$$

we define the sum of the two to be given by

$$(a_1, a_2, a_3, \dots) + (b_1, b_2, b_3, \dots) := (a_1 + b_1, a_2 + b_2, a_3 + b_3, \dots). \tag{24.43}$$

The zero vector is the sequence $(0, 0, 0, \dots)$. If c is a real number, then the scalar multiplication is defined to be

$$c(a_1, a_2, a_3, \dots) := (ca_1, ca_2, ca_3, \dots). \tag{24.44}$$

The sum of two sequences still satisfies the absolutely convergent condition because

$$\sum_{n=1}^{\infty} |a_n + b_n| \leq \sum_{n=1}^{\infty} \Big(|a_n| + |b_n| \Big) = \sum_{n=1}^{\infty} |a_n| + \sum_{n=1}^{\infty} |b_n| < \infty \tag{24.45}$$

by the triangle inequality. Furthermore, the scalar multiplication of an absolutely convergent series is still absolutely convergent

$$\sum_{n=1}^{\infty} |ca_n| = \sum_{n=1}^{\infty} |c||a_n| = |c| \sum_{n=1}^{\infty} |a_n| < \infty. \tag{24.46}$$

The set of such sequences whose associated series are absolutely convergent is denoted by ℓ^1 (read "little ell one").

Exercise 24.47. A closely related example of a vector space that shows up in quantum mechanics is the set of sequence of *complex* numbers $(a_1, a_2, a_3, \dots)$ such that

$$\sum_{n=1}^{\infty} |a_n|^2 < \infty. \tag{24.48}$$

The zero vector, sum, and scalar multiplication are defined just as in Example 24.40. Check that this is a complex vector space (this vector space is denoted by ℓ^2 and is read "little ell two").

Exercise 24.49. A generalization of the vector spaces ℓ^1 and ℓ^2 is the following. Let $p \geq 1$. Check that the set of sequences $(a_1, a_2, a_3, \dots)$ satisfying

$$\sum_{n=1}^{\infty} |a_n|^p < \infty, \tag{24.50}$$

with similar structure as in the previous examples is a vector space. This space is denoted by ℓ^p and is read "little ell p."

Exercise 24.51. Show that ℓ^1 is a subspace of ℓ^2. Note that ℓ^1 was defined in Example 24.40 and ℓ^2 was defined in Example 24.47 (let both of the sequences be either real or complex). [Warning: this exercise is not trivial and requires a solid understanding of infinite series from calculus.]

Exercise 24.52. Show that ℓ^2 is not a subspace of ℓ^1. [Hint: give an example of a sequence $(a_1, a_2, a_3, \dots)$ such that $\sum_{n=1}^{\infty} |a_n|^2 < \infty$ but $\sum_{n=1}^{\infty} |a_n|$ does not converge.]

Recommended Exercises. Please check HuskyCT for the homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we covered parts of Chapter 4, most notably Section 4.1 and 4.2 of [Lay]. We have also covered additional topics not all of which are covered in the textbook giving more context and utility for the notion of vector spaces.

Terminology checklist

degree n polynomial	
vector space	
linear transformation between vector spaces	
kernel and image of a linear transformation	
function spaces	

25 Differential operators

Let us go through examples of eigenvectors and eigenvalues for linear transformations between more general vector spaces. We will focus on vector spaces of sufficiently differentiable functions and our linear transformations will often be differential operators.

Example 25.1. Let $\mathbb{P}_n$ be the set of degree n polynomials. The derivative operator $\frac{d}{dx}$ is a linear transformation $\mathbb{P}_n \xleftarrow{\frac{d}{dx}} \mathbb{P}_n$ that takes a degree n polynomial and differentiates it

$$\frac{d}{dx}\Big(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n\Big) = a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1}. \tag{25.2}$$

Does $\frac{d}{dx}$ have any eigenvectors? If so, what are the associated eigenvalues? To answer this question, it helps to draw a table and relate it to our earlier understanding of eigenvectors for matrices. Recall, a vector $\vec{v}$ is an eigenvector for a matrix A with eigenvalue λ iff $A\vec{v} = \lambda\vec{v}$. Three mathematical objects are used in this definition. It is our job to find the three corresponding mathematical objects used in our current problem. The analogue of the matrix A is the linear transformation $\frac{d}{dx}$. The analogue of the vector $\vec{v}$ is a *polynomial* p. The analogue of the number λ is still just a number λ.

A	$\frac{d}{dx}$
$\vec{v}$	p
λ	λ

Therefore, the eigenvalue equation reads

$$\frac{d}{dx}p(x) = \lambda p(x). \tag{25.3}$$

How do we solve this? It helps to read what this equation is saying out loud. Does there exist a polynomial whose derivative is itself up to a scalar? What does $\frac{d}{dx}$ do to the monomials $\{p_0, p_1, \ldots, p_n\}$? Notice that

$$\frac{d}{dx}p_k = kp_{k-1} \tag{25.4}$$

for all $k \in \{0, 1, \ldots, n\}$. In terms of the variable x, this reads

$$\frac{d}{dx}p_k(x) = \frac{d}{dx}x^k = kx^{k-1} = kp_{k-1}(x). \tag{25.5}$$

If we used prime notation, this equation would read

$$p_k'(x) = kp_{k-1}(x). \tag{25.6}$$

We will occasionally avoid the prime notation often because we want to emphasize the perspective that the derivative is a linear transformation and should be viewed as a mathematical object in its own right just like a matrix. From this calculation, we notice that the derivative always lowers the degree of a polynomial. Hence, it seems impossible for there to be any eigenvectors! However, what about a polynomial of degree 0? If we take the derivative of a constant c, we get 0:

$$\frac{d}{dx}c = 0. \tag{25.7}$$

0 is still a constant, so we can interpret this as

$$\frac{d}{dx}c = 0c. \tag{25.8}$$

This tells us that every (non-zero) constant c is an eigenvector of $\frac{d}{dx}$ and its eigenvalue is 0. A basis of eigenvectors for the eigenvalue 0 would be $\{1\}$ or the monomial $\{p_0\}$, which is just another way of writing the constant function 1 because $p_0(x) = x^0 = 1$.

The fact that there are no other eigenvectors should surprise you if you've taken calculus. In case you're not sure why, the following example should illustrate.

Example 25.9. Let $\mathcal{A}(\mathbb{R})$ denote the set of all analytic functions. Recall, these are all infinitely differentiable functions f whose associated Taylor series expansions agree with the original function, i.e. for all real numbers a,

$$f(x+a) = f(x) + a\frac{df}{dx}(x) + \frac{a^2}{2!}\frac{d^2f}{dx^2}(x) + \frac{a^3}{3!}\frac{d^3f}{dx^3}(x) + \cdots . \tag{25.10}$$

Then $\mathcal{A}(\mathbb{R})$, with the addition and scalar multiplication for functions, is a real vector space. Just as the derivative on polynomials is a linear transformation, $\frac{d}{dx}$ is also a linear transformation on $\mathcal{A}(\mathbb{R})$.[71] Does $\frac{d}{dx}$ have any eigenvectors whose eigenvalue is not zero? Again, to answer this question, it helps to draw and fill in a table. The analogue of the matrix A is the linear transformation $\frac{d}{dx}$. The analogue of the vector $\vec{v}$ is an analytic *function* f. The analogue of the number λ is still just a number λ.

A	$\frac{d}{dx}$
$\vec{v}$	f
λ	λ

Therefore, the eigenvalue problem looks like

$$\frac{d}{dx}f = \lambda f. \tag{25.11}$$

Namely, for what real number λ and for what analytic functions f does

$$\frac{d}{dx}f(x) = \lambda f(x) \tag{25.12}$$

for all $x \in \mathbb{R}$? We cannot actually write down a basis for $\mathcal{A}(\mathbb{R})$—nobody (and I mean *nobody*) knows even one basis for $\mathcal{A}(\mathbb{R})$ so we cannot write down any matrices here to help us. However, we can still answer this question from a more conceptual point of view by using the definitions. To solve

$$\frac{df}{dx} = \lambda f, \tag{25.13}$$

[71] There is some abuse of notation here. In the previous example, we viewed $\frac{d}{dx}$ as a linear transformation whose source/domain was $\mathbb{P}_n$ and whose target/codomain was $\mathbb{P}_n$. Here, we are viewing $\frac{d}{dx}$ as a linear transformation whose source/domain is $\mathcal{A}(\mathbb{R})$ and whose target/codomain is $\mathcal{A}(\mathbb{R})$ even though we are using the same notation for it. The only reason we do this is because $\mathbb{P}_n$ is a vector subspace of $\mathcal{A}(\mathbb{R})$ and the linear transformation acts in the same way on this subspace.

we separate the variables as in

$$\frac{df}{f} = \lambda dx \tag{25.14}$$

and integrate

$$\int \frac{1}{f} df = \int \lambda dx \qquad \Rightarrow \qquad \ln(f) = \lambda x + c \qquad \Rightarrow \qquad f(x) = Ce^{\lambda x} \tag{25.15}$$

for some constant c and $C := e^c$. Let e_λ be the function $e_\lambda(x) = e^{\lambda x}$. Hence, if $\frac{d}{dx}$ were to have any eigenvector, it should be e_λ for each real number λ. But is e_λ analytic, i.e. is e_λ really a vector in $\mathcal{A}(\mathbb{R})$? This is true and it follows from the definition of the exponential (as well as a theorem that says the exponential series uniformly converges on any compact interval). The eigenvalue of e_λ is λ. Hence, $\frac{d}{dx}$ has infinitely (uncountably!) many (linearly independent!) eigenvectors, with each eigenspace being spanned by an exponential function. Because there is only one linearly independent set of functions for each eigenvalue λ, the multiplicity of λ is 1.

Example 25.16. Let $\mathbb{P}_n$ be the vector space of degree n polynomials (in the variable x). The function $\mathbb{P}_{n+1} \xleftarrow{T} \mathbb{P}_n$, given by

$$\big(T(p)\big)(x) := xp(x) \tag{25.17}$$

for any degree n polynomial p, is a linear transformation. Here p is a degree n polynomial and $T(p)$ is a degree $n+1$ polynomial. For example, if $n = 1$ and p is of the form $p(x) = mx + b$ with m and b real numbers, then $T(p)$ is the quadratic polynomial given by $mx^2 + bx$. Let us check that T is indeed a linear transformation. We must show two things.

(a) Let p and q be two degree n polynomials, which are of the form

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \qquad \& \qquad q(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n. \tag{25.18}$$

Then $p + q$ is the polynomial given by

$$(p+q)(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n. \tag{25.19}$$

Denote the polynomial $T(p + q)$ by r. By the definition of T, r is given by

$$\begin{aligned} r(x) &= x\Big((a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n\Big) \\ &= (a_0 + b_0)x + (a_1 + b_1)x^2 + (a_2 + b_2)x^3 + \cdots + (a_n + b_n)x^{n+1} \\ &= \Big(a_0x + a_1x^2 + a_2x^3 + \cdots + a_nx^{n+1}\Big) + \Big(b_0x + b_1x^2 + b_2x^3 + \cdots + b_nx^{n+1}\Big) \\ &= xp(x) + xq(x) \\ &= \big(T(p)\big)(x) + \big(T(q)\big)(x). \end{aligned} \tag{25.20}$$

This shows that $T(p + q) = T(p) + T(q)$.

(b) Now let λ be a real number. Then $T(\lambda p)$ is the polynomial given by

$$\begin{aligned}\big(T(\lambda p)\big)(x) &= x\Big(\lambda a_0 + \lambda a_1 x + \lambda a_2 x^2 + \cdots + \lambda a_n x^n\Big)\\ &= \lambda a_0 x + \lambda a_1 x^2 + \lambda a_2 x^3 + \cdots + \lambda a_n x^{n+1}\\ &= \lambda\Big(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n\Big)\\ &= \lambda\big(xp(x)\big)\\ &= \lambda\big(T(p)\big)(x)\end{aligned} \tag{25.21}$$

which shows that $T(\lambda p) = \lambda T(p)$.

These two calculations prove that T is a linear transformation.

Problem 25.22. Let $\mathcal{A}(\mathbb{R}) \xleftarrow{\frac{d}{dx}} \mathcal{A}(\mathbb{R})$ be the derivative operator on real-valued analytic functions on $\mathbb{R}$. Is cos in the image of $\frac{d}{dx}$ (here, cos is the cosine function). If so, prove it.

Answer. It helps to compare this problem to those we have solved before. If A is a matrix and $\vec{b}$ is a vector, to show that $\vec{b}$ is in the image of A, we would have to find an $\vec{x}$ such that $A\vec{x} = \vec{b}$. Setting up a table to compare that to our current problem gives

A	$\frac{d}{dx}$
$\vec{x}$	f
$\vec{b}$	cos

Therefore, we want to find an analytic function f such that

$$\frac{d}{dx} f(x) = \cos(x). \tag{25.23}$$

One such function is $f = \sin$ because its derivative is cos .

Problem 25.24. Find a basis of eigenvectors for the linear transformation $\mathcal{A}(\mathbb{R}) \xleftarrow{\frac{d^2}{dx^2}} \mathcal{A}(\mathbb{R})$ with eigenvalue 4.

Answer. Writing the table for eigenvectors and eigenvalues gives

A	$\frac{d^2}{dx^2}$
$\vec{x}$	f
λ	$\lambda = 4$

which tells us that we must solve

$$\frac{d^2}{dx^2} f(x) = 4f(x), \tag{25.25}$$

i.e. a function whose second derivative is 4 times itself. The derivative of e^{ax} is ae^{ax} so that the second derivative of e^{ax} is $a^2 e^{ax}$. Plugging this in as our ansatz gives

$$\frac{d^2}{dx^2} e^{ax} = a^2 e^{ax} = 4e^{ax}. \tag{25.26}$$

Solving for a gives two solutions $a = \pm 2$. Therefore, since the set of functions $\{e^{2x}, e^{-2x}\}$ are linearly independent, it forms a basis for the eigenspace of $\frac{d^2}{dx^2}$ with eigenvalue 4.

If you are uncomfortable guessing solutions as we have, there is a sure way to go about solving this by turning the second order differential equation into a *system* of first order differential equations. The second order equation we want to solve is

$$f'' - 4f = 0. \tag{25.27}$$

Let $g := f'$. Then the associated linear system of first order differential equations is

$$\begin{aligned} f' &= g \\ g' &= 4f. \end{aligned} \tag{25.28}$$

The first line is by definition of g. The second line is precisely the equation $f'' = 4f$. However, notice that this system can be expressed as

$$\frac{d}{dx}\begin{bmatrix} f \\ g \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix}\begin{bmatrix} f \\ g \end{bmatrix}. \tag{25.29}$$

This relates back to our discussion on solving ordinary differential equations. The eigenvalues of the matrix $\begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix}$ are $\lambda_1 = 2$ and $\lambda_2 = -2$ because the characteristic polynomial is

$$\lambda^2 - 4. \tag{25.30}$$

Corresponding eigenvectors are

$$\lambda_1 = 2 \;\; : \;\; \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \& \qquad \lambda_2 = -2 \;\; : \;\; \begin{bmatrix} 1 \\ -2 \end{bmatrix}. \tag{25.31}$$

Therefore, the matrix P that diagonalizes our matrix is

$$P = \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix} \tag{25.32}$$

so that

$$\begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}^{-1}. \tag{25.33}$$

Setting $f(0) =: f_0$ and $g(0) =: g_0$, the solutions of our first order system are therefore

$$\begin{aligned} \begin{bmatrix} f(x) \\ g(x) \end{bmatrix} &= \exp\left\{\begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix} x\right\}\begin{bmatrix} f_0 \\ g_0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} e^{2x} & 0 \\ 0 & e^{-2x} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}^{-1}\begin{bmatrix} f_0 \\ g_0 \end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix} e^{2x} & e^{-2x} \\ 2e^{2x} & -2e^{-2x} \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} f_0 \\ g_0 \end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix} 2e^{2x} + 2e^{-2x} & e^{2x} - e^{-2x} \\ 4e^{2x} - 4e^{-2x} & 2e^{2x} + 2e^{-2x} \end{bmatrix}\begin{bmatrix} f_0 \\ g_0 \end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix} \big(2f_0 + g_0\big)e^{2x} + \big(2f_0 - g_0\big)e^{-2x} \\ \big(4f_0 + 2g_0\big)e^{2x} + \big(2g_0 - 4f_0\big)e^{-2x} \end{bmatrix} \end{aligned} \tag{25.34}$$

This says that

$$f(x) = \left(\frac{2f_0 + g_0}{4}\right) e^{2x} + \left(\frac{2f_0 - g_0}{4}\right) e^{-2x}. \tag{25.35}$$

Since f_0 and g_0 can be chosen arbitrarily, this means that $f(x)$ is an arbitrary linear combination of e^{2x} and e^{-2x}. We didn't have to do this entire calculation to see that our basis for the eigenspace is $\{e^{2x}, e^{-2x}\}$. We could have stopped immediately when we found the eigenvalues. This is because the matrix exponential is expressed in terms of the exponential of the eigenvalues with the variable x in the exponent in front of these eigenvalues. The usefulness of the form given in 25.34 is if we were given *initial conditions* $f(0)$ and $f'(0) \equiv g(0)$ when x is viewed as time.

Many familiar theorems are still true for arbitrary vector spaces and linear transformations.

Theorem 25.36. *The kernel of a linear transformation $W \xleftarrow{T} V$ is a subspace of V. The image of a linear transformation $W \xleftarrow{T} V$ is a subspace of W.*

Example 25.37. Let $\mathbb{P}_{n+1} \xleftarrow{T} \mathbb{P}_n$ be the linear transformation from Example 25.16. The image of this linear transformation is the set of degree $n+1$ polynomials of the form

$$a_1 x + a_2 x^2 + \cdots + a_n x^n + a_{n+1} x^{n+1}. \tag{25.38}$$

In other words, it is the set of all polynomials with no constant term. Mathematically, as a set, this would be written as

$$\left\{a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + a_{n+1} x^{n+1} \ : \ a_0 = 0,\ a_1, a_2, \ldots, a_n \in \mathbb{R}\right\}. \tag{25.39}$$

The kernel of T consists of only the constant 0 polynomial because if p is a degree n polynomial of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{25.40}$$

then $T(p)$ is the polynomial

$$\big(T(p)\big)(x) = a_0 x + a_1 x^2 + a_2 x^3 + \cdots + a_n x^{n+1} \tag{25.41}$$

and this is zero for all $x \in \mathbb{R}$ if and only if $a_0 = a_1 = a_2 = \cdots = a_n = 0$. This linear transformation is also defined on all polynomials $\mathbb{P} \xleftarrow{T} \mathbb{P}$. Does *this* linear transformation have any eigenvalues with corresponding eigenvectors? If λ is such an eigenvalue and p is an eigenvector (a polynomial), then this would required $xp(x) = \lambda p(x)$ for all input values of x. Since p must be nonzero (in order for it to be an eigenvector), for any value of x for which $p(x) \neq 0$, this equation demands that $\lambda = x$, which is impossible since $p(x) \neq 0$ for at least two distinct values of x (in fact, infinitely many). Therefore, $\mathbb{P} \xleftarrow{T} \mathbb{P}$ has no eigenvalues either.

Problem 25.42. Solve the third order linear differential equation

$$f''' - 4f'' + 5f' - 2f = 0. \tag{25.43}$$

Answer. Set $g := f'$ and $h := g'$. Then this third order differential equation becomes

$$\begin{aligned} f' &= g \\ g' &= h \\ h' &= 2f - 5g + 4h \end{aligned} \tag{25.44}$$

or in matrix form

$$\frac{d}{dx}\begin{bmatrix} f \\ g \\ h \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{bmatrix}\begin{bmatrix} f \\ g \\ h \end{bmatrix}. \tag{25.45}$$

The characteristic polynomial of the determinant inside is

$$\lambda^3 - 4\lambda^2 + 5\lambda - 2 \tag{25.46}$$

after multiplying by -1. This surprisingly looks a lot like the differential equation itself (this is not a coincidence as will be explained in Remark 25.48). $\lambda = 1$ is a root of this polynomial. Long division gives

$$\lambda^3 - 4\lambda^2 + 5\lambda - 2 = (\lambda - 1)(\lambda^2 - 3\lambda + 2) = (\lambda - 1)^2(\lambda - 2). \tag{25.47}$$

A basis of solutions is therefore $\{e^x, xe^x, e^{2x}\}$.

Remark 25.48. The Cayley-Hamilton Theorem states that a square matrix A satisfies $p(A) = 0$, where p is the characteristic polynomial (modified by an overall sign with respect to our earlier convention here) $p(\lambda) := \det(\lambda \mathbb{1} - A)$. Although $\frac{d}{dx}$ is not a square matrix, the differential equation $f''' - 4f'' + 5f' - 2f = 0$ is precisely

$$\left(\frac{d^3}{dx^3} - 4\frac{d^2}{dx^2} + 5\frac{d}{dx} - 2\mathbb{1}\right) f(x) = 0. \tag{25.49}$$

The associated characteristic polynomial is

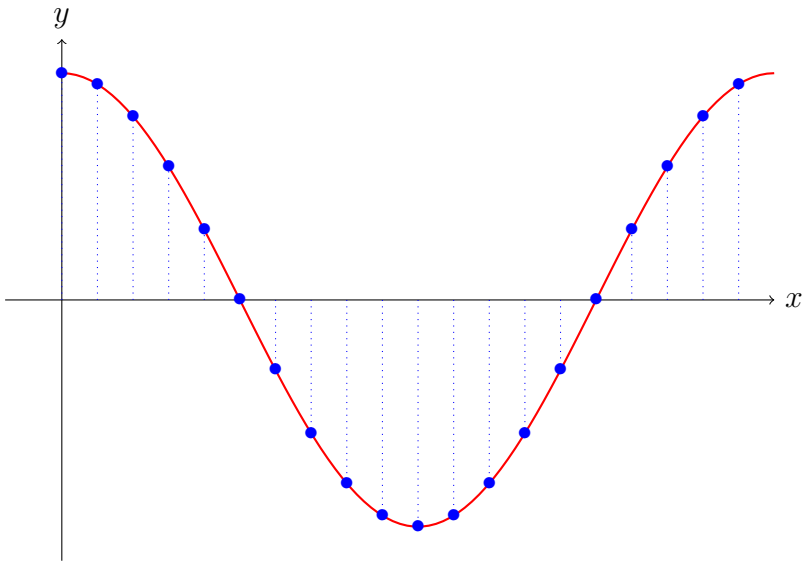
$$\lambda^3 - 4\lambda^2 + 5\lambda - 2. \tag{25.50}$$

Therefore, the reason n-th order linear homogeneous differential equations can be solved by replacing derivatives of f with powers of λ is due to a variant of the Cayley-Hamilton Theorem.

Example 25.51. Fix $N \in \mathbb{N}$ to be some large natural number and set $\Delta := \frac{1}{N}$. Consider the set of periodic functions on the interval $[0, 1]$ with values given only on the points $\frac{k}{N}$ for $k \in \{0, 1, \ldots, N-1\}$ ($k = N$ is not included because by periodicity, we assume that the value of these functions at 0 equals the value at 1). Let $T : \mathbb{R}^N \to \mathbb{R}^N$ be the function

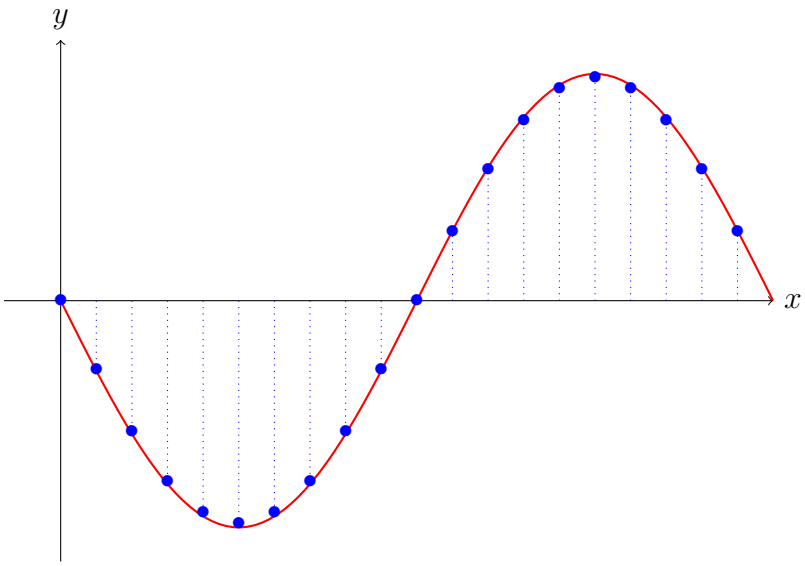
$$\mathbb{R}^N \ni e_k \mapsto T(e_k) := \frac{e_{k+1} - e_{k-1}}{2\Delta}, \tag{25.52}$$

where $-1 \equiv N - 1$ and $N = 0$. Extend this function in a linear fashion so that it is a linear transformation. T represents an approximation to the derivative function. It takes the value of the function at a point and then takes the approximate slope by using the values of the function at its nearest neighbors. For example, the cosine function with $N = 20$ looks like



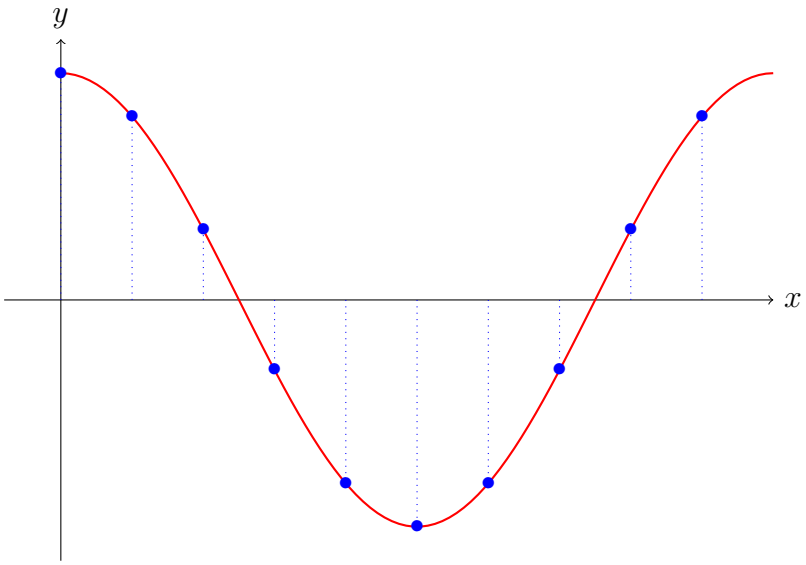
$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix}$$

Using the formula for T then gives the set of points

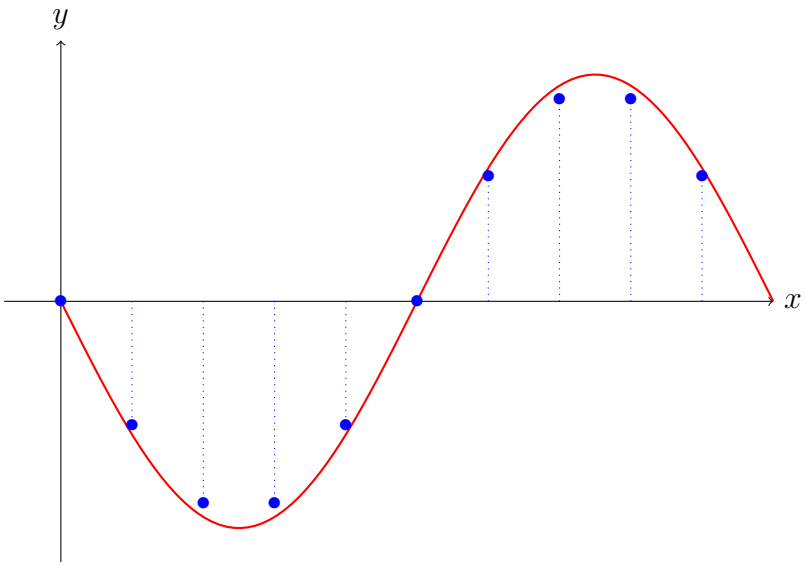


$$\frac{1}{2\Delta}\begin{bmatrix} y_1 - y_{N-1} \\ y_2 - y_0 \\ y_3 - y_1 \\ \vdots \\ y_0 - y_{N-2} \end{bmatrix}$$

which are seen to lie almost exactly along the $-\sin$ curve. If fewer points where chosen, such as 10, this approximation might not be as good. The approximation



gets mapped to



so you can see that the approximation for the derivative is not as good.

Example 25.53. The set of solutions to an n-th order homogeneous linear ordinary differential equation is a vector space. To see this, let us first write down such an ODE as

$$a_n f^{(n)} + a_{n-1} f^{(n-1)} + \cdots + a_1 f^{(1)} + a_0 f = 0. \tag{25.54}$$

Here f denotes a sufficiently smooth function of a single variable (one that admits all of these derivatives) and $f^{(k)}$ denotes the k-th derivative of f. The coefficients a_k are all constants independent of the variable input for f. An example of such an ODE is

$$f^{(2)} + f = 0 \tag{25.55}$$

whose solutions are all of the form

$$f(x) = a\cos(x) + b\sin(x), \tag{25.56}$$

with a and b real numbers. Notice that in this example, the set of solutions is given by $\text{span}\{\cos, \sin\}$. More generally, let $\mathcal{A}(\Omega)$ denote the vector space of analytic functions of a single variable on a domain $\Omega \subseteq \mathbb{R}$. Let $\mathcal{A}(\Omega) \xleftarrow{L} \mathcal{A}(\Omega)$ be the transformation defined by

$$\mathcal{A}(\Omega) \ni f \mapsto L(f) := a_n f^{(n)} + a_{n-1} f^{(n-1)} + \cdots + a_1 f^{(1)} + a_0 f. \tag{25.57}$$

Then L is a linear transformation and $\ker(L)$ is exactly the set of solutions to our general ODE. In particular, it is a vector space since every kernel of a linear transformation is. Inhomogeneous systems can also be formulated in this framework. Let $g \in \mathcal{A}(\Omega)$ be another function and let

$$a_n f^{(n)} + a_{n-1} f^{(n-1)} + \cdots + a_1 f^{(1)} + a_0 f = g \tag{25.58}$$

be an n-th order linear inhomogeneous ordinary differential equation. L is defined just as above and the solution set is actually the solution set of $Lf = g$ (which should remind you of $A\vec{x} = \vec{b}$, where A is replaced by L, $\vec{x}$ is replaced by a function f, and $\vec{b}$ is replaced by a function g). In particular, the solution set of the inhomogeneous ODE is a linear manifold in $\mathcal{A}(\Omega)$.

Theorem 3.34 from a while back tells us that the general solution to the inhomogeneous system $Lf = g$ is therefore of the form

$$f(x) = f_p(x) + f_h(x), \tag{25.59}$$

where f_p is *one* particular solution to $Lf = g$ and f_h is *any* homogeneous solution to $Lf = 0$. Therefore, many of the concepts from the theory of differential equations are special cases of the concepts from linear algebra.

Example 25.60. Let $\mathcal{A}(\mathbb{R}^2)$ be the vector space of analytic real-valued functions of two variables x and y and let $\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ be the Laplacian. Then, $\ker \Delta$ is the subspace of harmonic functions on $\mathbb{R}^2$.

Recommended Exercises. Please check HuskyCT for the homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

26 Bases and matrices for linear transformations*

It is sometimes helpful to write down matrix representations of linear transformations between abstract vector spaces. The $m \times n$ matrix associated to a linear transformation $\mathbb{R}^m \xleftarrow{T} \mathbb{R}^n$ was a convenient tool for calculating certain expressions. In fact, we were able to use the basis $\{\vec{e}_1, \ldots, \vec{e}_n\}$ in $\mathbb{R}^n$ to write down the matrix associated to T—but we didn't have to for many of the calculations we did. For example, $T(4\vec{e}_2 - 7\vec{e}_5) = 4T(\vec{e}_2) - 7T(\vec{e}_5)$ does not require writing down this matrix. If we know where a basis goes under a linear transformation T, then we know what the linear transformation does to *any* vector. For example, if $\{\vec{v}_1, \ldots, \vec{v}_n\}$ was a basis for $\mathbb{R}^n$, then any vector $\vec{u} \in \mathbb{R}^n$ can be expressed as a linear combination of these basis elements, let's say as

$$\vec{u} = u_1\vec{v}_1 + \cdots + u_n\vec{v}_n. \tag{26.1}$$

Then by linearity of T,

$$T(\vec{u}) = T\big(u_1\vec{v}_1 + \cdots + u_n\vec{v}_n\big) = u_1T(\vec{v}_1) + \cdots + u_nT(\vec{v}_n). \tag{26.2}$$

Therefore, we only need to know what the vectors $\{T(\vec{v}_1), \ldots, T(\vec{v}_n)\}$ are.

Furthermore, when we express the actual *components* of a vector such as $T(\vec{e}_2)$, we would be using the basis $\{\vec{e}_1, \ldots, \vec{e}_m\}$ for $\mathbb{R}^m$ (notice that we're now looking at $\mathbb{R}^m$ and not $\mathbb{R}^n$ because $T(\vec{e}_2)$ is a vector in $\mathbb{R}^m$). In other words, we can use this basis to express the vector $T(\vec{e}_2)$ as a linear combination of these basis vectors. But we could have also used any other basis. Therefore, the notion of a matrix associated to a linear transformation makes sense for any basis on the source and target of T.

Definition 26.3. Let V be a vector space. A set of vectors $\mathcal{S} := \{\vec{v}_1, \ldots, \vec{v}_k\}$ in V is *linearly independent* if the only values of $x_1, \ldots, x_k$ that satisfy the equation

$$\sum_{i=1}^{k} x_i\vec{v}_i \equiv x_1\vec{v}_1 + \cdots + x_k\vec{v}_k = \vec{0} \tag{26.4}$$

are

$$x_1 = 0, \quad x_2 = 0, \quad \ldots, \quad x_k = 0. \tag{26.5}$$

A set $\mathcal{S}$ of vectors as above is *linearly dependent* if there exists a solution to the above equation for which not all of the x_i's are zero. If $\mathcal{S}$ is an infinite set of vectors in V, indexed, say, by some set Λ so that $\mathcal{S} = \{\vec{v}_\alpha\}_{\alpha\in\Lambda}$, then $\mathcal{S}$ is *linearly independent* if for *every finite* subset Ω of Λ, the only solution to[72]

$$\sum_{\alpha\in\Omega} x_\alpha\vec{v}_\alpha = \vec{0} \tag{26.6}$$

is

$$x_\alpha = 0 \qquad \text{for all } \alpha \in \Omega. \tag{26.7}$$

$\mathcal{S}$ is *linearly dependent* if it is not linearly independent, i.e. if there exists a finite subset Ω of Λ with a solution to $\sum_{\alpha\in\Omega} x_\alpha\vec{v}_\alpha = \vec{0}$ in which not all of the x_α's is 0.

[72] Here, the notation $\alpha \in \Omega$ means that α is an element of Ω.

Example 26.8. Let $\mathcal{S}$ be the set of degree 7 polynomials of the form $\mathcal{S} := \{p_1, p_3, p_7\}$. Here p_k is the k-th degree monomial

$$p_k(x) = x^k. \tag{26.9}$$

The set $\mathcal{S}$ is linearly independent. This is because the only solution to

$$a_1 x + a_3 x^3 + a_7 x^7 = 0 \tag{26.10}$$

that holds for *all* x is

$$a_1 = a_3 = a_7 = 0. \tag{26.11}$$

However, the set $\{p_1, p_3, 3p_3 - 7p_7, p_7\}$ is linearly dependent because the third entry can be written as a linear combination of the other entries. The set $\{p_1 + 2p_3, p_4 - p_5, 3p_6 + 5p_0 - p_1\}$ is linearly independent. This is because the only solution to

$$a(x + 2x^3) + b(x^4 - x^5) + c(3x^6 + 5 - x) = 0 \tag{26.12}$$

is $a = b = c = 0$. To see this, rewrite the left-hand-side as

$$a(x + 2x^3) + b(x^4 - x^5) + c(3x^6 + 5 - x) = c5 + (a - c)x + a2x^3 + bx^4 - bx^5 + c3x^6. \tag{26.13}$$

The only way the right-hand-side vanishes for all values of x is when $c = 0$ which then forces $a = 0$ and $b = 0$ as well.

Example 26.14. The set of 2×2 matrices

$$A := \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \qquad B := \begin{bmatrix} -3 & 1 \\ 0 & 2 \end{bmatrix}, \quad \text{and} \quad C := \begin{bmatrix} -1 & 5 \\ 0 & 0 \end{bmatrix} \tag{26.15}$$

are linearly dependent in ${}_2\mathcal{M}_2$ because

$$2A + B = C. \tag{26.16}$$

Definition 26.17. Let V be a vector space and let $\mathcal{S} := \{\vec{v}_1, \ldots, \vec{v}_k\}$ be a set of k vectors in V. The *span* of $\mathcal{S}$ is the set of vectors in V of the form

$$\sum_{i=1}^{k} x_i \vec{v}_i \equiv x_1 \vec{v}_1 + \cdots x_k \vec{v}_k, \tag{26.18}$$

with $x_1, \ldots, x_k$ arbitrary real or complex numbers. The span of $\mathcal{S}$ is often denoted by $\text{span}(\mathcal{S})$. If $\mathcal{S}$ is an infinite set of vectors, say indexed by some set Λ, in which case $\mathcal{S}$ is written as $\mathcal{S} := \{\vec{v}_\alpha\}_{\alpha \in \Lambda}$, then the span of $\mathcal{S}$ is the set of vectors in V of the form[73]

$$\sum_{\substack{\alpha \in \Omega \subseteq \Lambda \\ \Omega \text{ is finite}}} x_\alpha \vec{v}_\alpha. \tag{26.19}$$

[73] Here, the notation $\Omega \subseteq \Lambda$ means that Ω is a subset of Λ and $\alpha \in \Omega$ means that α is an element of Ω.

The sum in the definition of span must be *finite* (even if $\mathcal{S}$ itself is infinite). At this point, it does *not* make sense to take an infinite sum of vectors because the latter requires a discussion on sequences, series, and convergence.

Example 26.20. Let $\mathcal{S} = \{p_0, p_1, p_2, p_3, \dots\}$ be the set of *all* monomials. Then $\operatorname{span}(\mathcal{S}) = \mathbb{P}$, the vector space of all polynomials. Indeed, every polynomial has some finite degree so it is a finite linear combination of monomials. A power series is *not* a polynomial.

Example 26.21. Consider the vector space $\mathcal{F}$ of Example 24.34. Recall, this is the vector space of linear combinations of the functions $f_n(x) := \cos(2\pi n x)$ and $g_m(x) := \sin(2\pi m x)$ for arbitrary natural numbers n and m. Let $\mathcal{S} := \{f_0, f_1, g_1, f_2, g_2, f_3, g_3, \dots\}$. Then the function

$$x \mapsto \sin\left(2\pi x - \frac{\pi}{4}\right) \tag{26.22}$$

is in the span of $\mathcal{S}$. Notice that this function is not *in* the set $\mathcal{S}$. The fact that this function is in the span follows from the sum angle formula for sine:

$$\sin(\theta + \phi) = \sin(\theta)\cos(\phi) + \cos(\theta)\sin(\phi), \tag{26.23}$$

which gives

$$\begin{aligned} \sin\left(2\pi x - \frac{\pi}{4}\right) &= \sin(2\pi x)\cos\left(-\frac{\pi}{4}\right) + \cos(2\pi x)\sin\left(-\frac{\pi}{4}\right) \\ &= \frac{\sqrt{2}}{2}\sin(2\pi x) - \frac{\sqrt{2}}{2}\cos(2\pi x) \\ &= \frac{\sqrt{2}}{2} g_1(x) + \left(-\frac{\sqrt{2}}{2}\right) f_1(x). \end{aligned} \tag{26.24}$$

As another example, the function

$$x \mapsto \cos^2(2\pi x) \tag{26.25}$$

is also in the span of $\mathcal{S}$. For this, recall the other angle sum formula

$$\cos(\theta + \phi) = \cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi) \tag{26.26}$$

and of course the identity

$$\cos^2(\theta) + \sin^2(\theta) = 1, \tag{26.27}$$

which can be used to rewrite

$$\cos(2\theta) = \cos^2(\theta) - \sin^2(\theta) = \cos^2(\theta) - \Big(1 - \cos^2(\theta)\Big) = 2\cos^2(\theta) - 1. \tag{26.28}$$

Using this last identity, we can write

$$\cos^2(2\pi x) = \frac{1}{2} + \frac{1}{2}\cos(4\pi x) = \frac{1}{2} f_0(x) + \frac{1}{2} f_2(x). \tag{26.29}$$

Definition 26.30. Let V be a vector space. A *basis* for V is a set $\mathcal{B}$ of vectors in V that is linearly independent *and* spans V.

Definition 26.31. The number of elements in a basis for a vector space V is the *dimension* of V and is denoted by $\dim V$. A vector space V with $\dim V < \infty$ is said to be *finite-dimensional*. A vector space V with $\dim V = \infty$ is said to be *infinite-dimensional*.

Example 26.32. A basis of $\mathbb{P}_n$ is given by the monomials $\{p_0, p_1, p_2, \dots, p_n\}$, where

$$p_k(x) := x^k. \tag{26.33}$$

Therefore, $\dim \mathbb{P}_n = n + 1$. Similarly, $\{p_0, p_1, p_2, p_3, \dots\}$ is a basis for $\mathbb{P}$. Therefore, $\dim \mathbb{P} = \infty$.

Example 26.34. A basis for $m \times n$ matrices is given by matrices of the form

$$E_{ij} := \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{26.35}$$

where the only non-zero entry is in the i-th row and j-th column, where its value is 1. In other words, E_{ij} is an $m \times n$ matrix with a 1 in the i-th row and j-th column and is zero everywhere else. Therefore, $\dim {}_m\mathcal{M}_n := mn$. For example, in ${}_2\mathcal{M}_2$, this basis looks like

$$\left\{ E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \tag{26.36}$$

Example 26.37. Let $\mathcal{F}$ be the vector space from Example 24.34. A basis for $\mathcal{F}$ is given by $\{f_0, f_1, g_1, f_2, g_2, \dots\}$. Hence, $\dim \mathcal{F} = \infty$. Notice that we have excluded g_0 because g_0 is the zero function and would render the set linearly dependent if added.

Theorem 26.38. *Let V be a vector space and let H be a subspace of V. Then* $\dim(H) \leq \dim(V)$.

Theorem 26.39. *Let V be a vector space and let $\mathcal{B}$ be a basis for V. Then every vector $\vec{v}$ in V can be written* uniquely *as a linear combination of elements of $\mathcal{B}$ (except for possibly zero weights).*

Note that this theorem is false if the word "basis" is replaced with a set that merely spans V. Also, the theorem is true even in infinite-dimensional vector spaces.

Proof. To cover the case of infinite dimensions as well, let $\mathcal{B} = \{\vec{v}_\alpha\}_{\alpha \in \Lambda}$ and suppose that the vector $\vec{u}$ can be expressed as a linear combination of the vectors in S in two ways as

$$\vec{u} = \sum_{\substack{\alpha \in \Omega \subseteq \Lambda \\ \Omega \text{ finite}}} x_\alpha \vec{v}_\alpha \qquad \text{and} \qquad \vec{u} = \sum_{\substack{\beta \in \Theta \subseteq \Lambda \\ \Omega \text{ finite}}} y_\beta \vec{v}_\beta \tag{26.40}$$

Then the difference of these two equals the zero vector and is given by the sum

$$\sum_{\alpha \in \Omega \text{ but } \alpha \notin \Theta} x_\alpha \vec{v}_\alpha + \sum_{\alpha \in \Omega \text{ and } \alpha \in \Theta} (x_\alpha - y_\alpha) \vec{v}_\alpha + \sum_{\alpha \notin \Omega \text{ but } \alpha \in \Theta} y_\alpha \vec{v}_\alpha = \vec{0}. \tag{26.41}$$

Here the notation $\alpha \notin \Omega$ means that α is not an element of Ω. Because the vectors $\vec{v}_\alpha$ are all linearly independent, this is only possible if all of the coefficients are zero. ■

In the case that the vector space was infinite-dimensional, we noticed that there is some ambiguity in the linear combinations in the sense that one could throw in more vectors and attach coefficients of 0 in front. We would like to avoid this possibility, so for the time being, we work with finite-dimensional vector spaces. This theorem motivates the following definition.

Definition 26.42. Let V be a finite-dimensional vector space and let $\mathcal{B} = \{\vec{v}_1, \cdots, \vec{v}_n\}$ be a basis for V. Let $\vec{u}$ be a vector in V. The *coordinates of $\vec{v}$ with respect to $\mathcal{B}$* is the set of coefficients for the linear combination used to express $\vec{u}$ in terms of $\mathcal{B}$, i.e. the coefficients $x_1, \ldots, x_n$ from

$$\vec{u} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n. \tag{26.43}$$

You are already familiar with this concept, but we mostly dealt with the vector space $V = \mathbb{R}^n$ before. Let us give an example of expressing vectors with respect to a basis as in Definition 26.42.

Example 26.44. The set of real 2×2 matrices

$$\left\{ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\} \tag{26.45}$$

is a basis for the subspace of all real 2×2 matrices that are *symmetric* (i.e. they are equal to their transpose). Let us *prove* this. To show that the set is linearly independent, we must show that the only solution to

$$x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + y \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + z \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{26.46}$$

is the trivial solution $x = y = z = 0$. Writing this out gives

$$\begin{bmatrix} y+z & x \\ x & y-z \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{26.47}$$

Looking at each of the components, this implies $x = 0$ immediately. The other two equations say that $y + z = 0$ and $y - z = 0$. Adding them gives $2y = 0$ so $y = 0$ and subtracting gives $2z = 0$ so $z = 0$. Thus $x = y = z = 0$ as needed. Now consider a general symmetric matrix of the form

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \tag{26.48}$$

where $a, b, c \in \mathbb{R}$ are arbitrary. We must show that our above set spans all symmetric matrices, so we must find real numbers α, β, γ such that

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \alpha \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \beta \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \gamma \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{26.49}$$

I'll leave you the task of solving this (add up the matrices on the right-hand-side as we did earlier and compare the two sides) to you as an exercise and will simply give you the answer:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \left(\frac{a+c}{2}\right) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \left(\frac{a-c}{2}\right) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{26.50}$$

Exercise 26.51. What happens to the previous example if we view the matrices with entries in $\mathbb{Z}_2$? Do they still form a basis for all symmetric 2×2 matrices with entries in $\mathbb{Z}_2$? If so, prove it. If not, remove one of the matrices so that you are left with a linearly independent set and then add to that set another matrix that is symmetric so that the resulting set is a basis (and prove that it is).

Theorem 26.52. *Let V be an n-dimensional vector space. Any linearly independent set of n vectors in V is a basis for V.*

Remark 26.53. Be careful! This theorem is *false* in infinite dimensions! For example, take ℓ^1. The sequences of the form e_m defined by

$$e_m(n) := \begin{cases} 1 & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \tag{26.54}$$

are all linearly independent. Furthermore, there are an infinite number of such elements. Nevertheless, they do not form a basis for all elements in ℓ^1. To see this, notice that the sequence $a_n := \frac{1}{n^2}$ is not expressible as a (finite) linear combination of the e_m's. Indeed, it can only be expressed as an infinite sum

$$\sum_{m=1}^{\infty} \frac{1}{m^2} e_m = \left(1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \dots\right), \tag{26.55}$$

which, as mentioned before, we have not defined.

As another example, consider the vector space $\mathcal{F}$ of finite linear combinations of the sines and cosines. Then $\{f_0, f_1, g_1, f_2, g_2, \dots\}$ is a basis with a countable number of elements. But if we remove any finite number of elements from this basis, we will still have a linearly independent set of a countable number of elements, yet they will no longer form a basis.

Example 26.56. Consider the polynomials

$$q_0(x) := x^2 + x^3, \qquad q_1(x) := 1 - x, \qquad q_2(x) := 1 + x + x^2, \qquad q_3(x) := 1 - x^3 \tag{26.57}$$

and let $\mathcal{S} := \{q_0, q_1, q_2, q_3\}$. Then $\mathcal{S}$ is a basis of degree 3 polynomials. One way to check this is to express the elements of the basis $\mathcal{B} := \{p_0, p_1, p_2, p_3\}$ (see Example 26.32) in terms of the basis $\mathcal{S}$. Then using the previous theorem, since the set $\mathcal{S}$ has 4 vectors and we know $\mathcal{B} = \{p_0, p_1, p_2, p_3\}$ is a basis, we know $\mathcal{S}$ is a basis.

The goal is to find coefficients a_{ij} such that[74]

$$\begin{aligned} p_0 &= a_{00}q_0 + a_{10}q_1 + a_{20}q_2 + a_{30}q_3 \\ p_1 &= a_{01}q_0 + a_{11}q_1 + a_{21}q_2 + a_{31}q_3 \\ p_2 &= a_{02}q_0 + a_{12}q_1 + a_{22}q_2 + a_{32}q_3 \\ p_3 &= a_{03}q_0 + a_{13}q_1 + a_{23}q_2 + a_{33}q_3 \end{aligned} \tag{26.58}$$

[74]The reason for writing the coefficients in this way will be clear soon.

which is exactly a linear system, just in a foreign vector space instead of (what looks like) $\mathbb{R}^4$. In terms of the variable x, the system (26.59) takes the form

$$\begin{aligned} 1 &= (a_{10} + a_{20} + a_{30}) + (a_{20} - a_{10})x + (a_{00} + a_{20})x^2 + (a_{00} - a_{30})x^3 \\ x &= (a_{11} + a_{21} + a_{31}) + (a_{21} - a_{11})x + (a_{01} + a_{21})x^2 + (a_{01} - a_{31})x^3 \\ x^2 &= (a_{12} + a_{22} + a_{32}) + (a_{22} - a_{12})x + (a_{02} + a_{22})x^2 + (a_{02} - a_{32})x^3 \\ x^3 &= (a_{13} + a_{23} + a_{33}) + (a_{23} - a_{13})x + (a_{03} + a_{23})x^2 + (a_{03} - a_{33})x^3 \end{aligned} \tag{26.59}$$

Solving this directly is certainly doable, but takes some time (it gives a linear system with 16 equations for all the unknowns). Another way is to express the basis $\mathcal{S}$ in terms of $\mathcal{B} = \{p_0, p_1, p_2, p_3\}$ instead (which is much easier). First, we have

$$\begin{aligned} q_0 &= p_2 + p_3 \\ q_1 &= p_0 - p_1 \\ q_2 &= p_0 + p_1 + p_2 \\ q_3 &= p_0 - p_3 \end{aligned} \tag{26.60}$$

Treating the left side as a list of vectors, we get a matrix of column vectors

$$\begin{bmatrix} q_0 & q_1 & q_2 & q_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}, \tag{26.61}$$

where the right-hand-side is a matrix with respect to the $\mathcal{B} = \{p_0, p_1, p_2, p_3\}$ basis. The inverse of this matrix will express the vectors $\mathcal{B} = \{p_0, p_1, p_2, p_3\}$ in terms of the basis $\mathcal{S} = \{q_0, q_1, q_2, q_3\}$. The inverse of this matrix is

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} -1 & -1 & 2 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & 2 & -2 \end{bmatrix} \tag{26.62}$$

The columns of this matrix should be the solution to our problem. Let us check this:

$$\begin{aligned} -q_0 + q_1 + q_2 - q_3 &= p_0 \\ -q_0 \qquad + q_2 - q_3 &= p_1 \\ 2q_0 - q_1 - q_2 + 2q_3 &= p_2 \\ -q_0 + q_1 + q_2 - 2q_3 &= p_3 \end{aligned} \tag{26.63}$$

Therefore, the columns of the inverse matrix express the basis $\{p_0, p_1, p_2, p_3\}$ in terms of the basis $\mathcal{S}$.

Definition 26.64. Let V and W be two finite-dimensional vector spaces. Let $\mathcal{V} := \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be an (ordered) basis for V and $\mathcal{W} := \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be an (ordered) basis for W. The $m \times n$

matrix associated to a linear transformation $W \xleftarrow{T} V$ with respect to the bases $\mathcal{V}$ and $\mathcal{W}$ is the $m \times n$ matrix whose ij-th entry is the unique coefficient ${}_{\mathcal{W}}[T]^{\mathcal{V}}{}_{ij}$ in front of w_i in the expansion[75]

$$T(\vec{v}_j) = \sum_{i=1}^{m} {}_{\mathcal{W}}[T]^{\mathcal{V}}{}_{ij}\vec{w}_i. \tag{26.65}$$

The same definition can be made for vector spaces of infinite dimensions provided one uses only finite linear combinations.

We will explain why ${}_{\mathcal{W}}[T]^{\mathcal{V}}{}_{ij}$ is reasonable notation after relating this concept to something more familiar.

Example 26.66. Let $\mathcal{E}$ denote the standard Euclidean bases. When $V = \mathbb{R}^m$ and $W = \mathbb{R}^n$, this reduces to what we already know. This is because the matrix associated to T is the matrix whose columns are the images of the unit vectors $\{T(\vec{e}_j)\}_{j\in\{1,\dots,n\}}$ since

$$[T] = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) & \cdots & T(\vec{e}_n) \end{bmatrix}. \tag{26.67}$$

We always wrote $T(\vec{e}_j)$ as a column matrix, but that's because we always used the standard Euclidean basis. For example,

$$T(\vec{e}_1) = T_{11}\vec{e}_1 + T_{21}\vec{e}_2 + \cdots + T_{m1}\vec{e}_m, \tag{26.68}$$

which gave us

$$[T] = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) & \cdots & T(\vec{e}_n) \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{m1} & T_{m2} & \cdots & T_{mn} \end{bmatrix} \tag{26.69}$$

which is exactly the form of the matrix from above. In other words,

$${}_{\mathcal{E}}[T]^{\mathcal{E}} = [T] \tag{26.70}$$

showing that the matrices we've been writing up to this point are matrices of a linear transformation between Euclidean spaces with respect to the Euclidean bases!

Example 26.71. Consider the derivative linear transformation $\frac{d}{dx}$ on degree n polynomials as in Example 25.1. We could express $\frac{d}{dx}$ as a matrix using the basis of monomials $\mathcal{P} := \{p_0, p_1, \dots, p_n\}$. With respect to this basis, the linear transformation $\frac{d}{dx}$ takes the form

$${}_{\mathcal{P}}\left[\frac{d}{dx}\right]^{\mathcal{P}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & n-1 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & n \\ 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \tag{26.72}$$

[75] Such and expansion exists because $\mathcal{W}$ spans W and such an expansion is unique because $\mathcal{W}$ is linearly independent.

This is an $(n+1)\times(n+1)$ matrix. For example, one can use this matrix representation to find the eigenvalues of $\frac{d}{dx}$. They are obtained by solving

$$0=\det\begin{bmatrix}-\lambda & 1 & 0 & 0 & 0 & \cdots & 0\\ 0 & -\lambda & 2 & 0 & 0 & \cdots & 0\\ 0 & 0 & -\lambda & 3 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & -\lambda & n-1 & 0\\ 0 & 0 & 0 & \cdots & \cdots & -\lambda & n\\ 0 & 0 & 0 & \cdots & \cdots & \cdots & -\lambda\end{bmatrix}=(-\lambda)^{n+1}=(-1)^{n+1}\lambda^{n+1} \tag{26.73}$$

because this is an upper triangular matrix and the determinant is therefore just the product along the diagonals. But upon inspection, the only solutions to this equation are $\lambda=0$. Therefore, the only eigenvalue of $\frac{d}{dx}$ is 0. Are there any eigenvectors? To find the eigenvectors associated to the eigenvalue 0, we would have to find polynomials p such that

$$\frac{d}{dx}p=0 \tag{26.74}$$

since $0p=0$ for all polynomials p. The only polynomial whose derivative is 0 is the constant polynomial. Hence, the set of all eigenvectors for $\frac{d}{dx}$ with eigenvalue 0 are

$$\Big\{tp_0\ :\ t\in\mathbb{R}\Big\}. \tag{26.75}$$

Example 26.76. Consider the linear transformation ${}_n\mathcal{M}_n \overset{T}{\leftarrow} {}_n\mathcal{M}_n$ defined by sending an $n\times n$ matrix A to $T(A):=A^T$, the transpose of A. What are the eigenvalues and eigenvectors of this transformation? Let us be concrete and analyze this problem for $n=2$. Then, the linear transformation acts as

$$T\left(\begin{bmatrix}a & b\\ c & d\end{bmatrix}\right)=\begin{bmatrix}a & c\\ b & d\end{bmatrix}. \tag{26.77}$$

We want to find solutions, 2×2 matrices A, together with eigenvalues λ, satisfying $T(A)=\lambda A$, i.e. $A^T=\lambda A$. Right off the bat, we can guess three eigenvectors (remember, our vectors are now 2×2 matrices!) by just looking at what T does in (26.77). These are

$$A_1=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\qquad A_2=\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix},\qquad \&\qquad A_3=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}. \tag{26.78}$$

Furthermore, their corresponding eigenvalues are all 1. This is because all of these matrices satisfy $A_i^T=A_i$ for $i=1,2,3$. Is there a fourth eigenvector?[76] For this, we could express the linear transformation T in terms of the basis $\mathcal{E}:=\{E_{11},E_{12},E_{21},E_{22}\}$. In this basis, the matrix representation of T is given by

$${}_{\mathcal{E}}[T]^{\mathcal{E}}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{26.79}$$

[76] We do not have to go through this entire calculation that follows to find this fourth eigenvector. One can think about what it should be by guessing, but we will go through this to illustrate what one would do even if it is not apparent.

because

$$\begin{aligned} T(E_{11}) &= E_{11} \\ T(E_{12}) &= E_{21} \\ T(E_{21}) &= E_{12} \\ T(E_{22}) &= E_{22}. \end{aligned} \tag{26.80}$$

The characteristic polynomial associated to this transformation is

$$\begin{aligned} \det \begin{bmatrix} 1-\lambda & 0 & 0 & 0 \\ 0 & -\lambda & 1 & 0 \\ 0 & 1 & -\lambda & 0 \\ 0 & 0 & 0 & 1-\lambda \end{bmatrix} &= (1-\lambda) \det \begin{bmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & 1-\lambda \end{bmatrix} \\ &= (1-\lambda)^2 \det \begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix} \\ &= (1-\lambda)^2(\lambda^2-1) \\ &= (1-\lambda)^3(\lambda+1). \end{aligned} \tag{26.81}$$

Hence, we see that there is another eigenvalue, namely, $\lambda_4 = -1$. The corresponding eigenvector can be solved for by solving the linear system

$$\left[\begin{array}{cccc|c} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] \tag{26.82}$$

whose solutions are all of the form

$$s \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{E}} \tag{26.83}$$

with s a free variable, which in terms of 2×2 matrices is given by

$$s \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{26.84}$$

Hence, our fourth eigenvector for T can be taken to be

$$A_4 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{26.85}$$

and its corresponding eigenvalue is $\lambda_4 = -1$.

Example 26.86. Consider the following two bases of degree 2 polynomials

$$\begin{aligned} q_0(x) &= 0 + 2x + 3x^2 & \qquad p_0(x) &= 1 \\ q_1(x) &= 1 + 1x + 3x^2 & \qquad p_1(x) &= x \\ q_2(x) &= 1 + 2x + 2x^2 & \qquad p_2(x) &= x^2 \end{aligned}$$

and the linear transformation $\mathbb{P}_2 \xleftarrow{T} \mathbb{P}_2$ satisfying $T(p_i) = q_i$ for $i = 0, 1, 2$. A matrix representation of this transformation in the $\mathcal{P} := \{p_0, p_1, p_2\}$ basis is given by

$$ {}_{\mathcal{P}}[T]^{\mathcal{P}} = \begin{bmatrix} 0 & 1 & 1 \\ 2 & 1 & 2 \\ 3 & 3 & 2 \end{bmatrix}^{\mathcal{P}}_{\mathcal{P}} . \tag{26.87} $$

Because this matrix has nonzero determinant ($\det T = 5$), it is invertible. In fact, we studied this matrix in Example 21.26. A matrix representation of the inverse of this transformation is given in the $\mathcal{Q} := \{q_0, q_1, q_2\}$ basis by

$$ {}_{\mathcal{Q}}[T^{-1}]^{\mathcal{Q}} = \frac{1}{5} \begin{bmatrix} -4 & 1 & 1 \\ 2 & -3 & 2 \\ 3 & 3 & -2 \end{bmatrix}^{\mathcal{Q}}_{\mathcal{Q}} . \tag{26.88} $$

and satisfies $T^{-1}(q_i) = p_i$ for $i = 0, 1, 2$. To check this, let us make sure that the first column of this matrix expresses the polynomial p_0 in the $\mathcal{Q}$ basis.

$$ \begin{aligned} -\frac{4}{5}q_0(x) + \frac{2}{5}q_1(x) + \frac{3}{5}q_2(x) &= -\frac{4}{5}(0 + 2x + 3x^2) \\ &\quad + \frac{2}{5}(1 - 1x + 3x^2) \\ &\quad - \frac{3}{5}(1 + 2x + 2x^2) \\ &= \quad 1(1 + 0x + 0x^2) \\ &= p_0(x). \end{aligned} $$

Anyway, we'd like to find the eigenvalues and corresponding eigenvectors of T. To do this, we can use any basis we'd like and use the matrix representation of T in this basis. Therefore, we can simply find the roots of the characteristic polynomial, which we have already done in Example 21.26. They were $\lambda_1 = -1, \lambda_2 = -1, \lambda_3 = 5$. We should now calculate the corresponding eigenvectors. For $\lambda_1 = \lambda_2 = -1$, we have to solve

$$ \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & 2 & 2 & 0 \\ 3 & 3 & 3 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{26.89} $$

which has solutions

$$ y \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{P}} + z \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}_{\mathcal{P}} \tag{26.90} $$

with two free variables y and z. Hence, a basis for such solutions, and therefore two eigenvectors for λ_1 and λ_2, is given by the two vectors

$$ \vec{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{P}} \qquad \& \qquad \vec{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}_{\mathcal{P}} . \tag{26.91} $$

(Note: your choice of eigenvectors could be different from mine!) For the eigenvalue $\lambda_3 = 5$, we must solve

$$\left[\begin{array}{ccc|c} -5 & 1 & 1 & 0 \\ 2 & -4 & 2 & 0 \\ 3 & 3 & -3 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 1 & 1 & -1 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 3 & -2 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & 0 & -1/3 & 0 \\ 0 & 1 & -2/3 & 0 \end{array}\right] \tag{26.92}$$

which has solutions

$$z \begin{bmatrix} 1/3 \\ 2/3 \\ 1 \end{bmatrix}_{\mathcal{P}} \tag{26.93}$$

with z a free variable. Thus, an eigenvector for λ_3 is

$$\vec{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_{\mathcal{P}} . \tag{26.94}$$

In terms of the polynomials, the eigenvalues together with their corresponding eigenvectors in $\mathbb{P}_2$ are given by

$$\begin{aligned} \lambda_1 : &\quad -p_0 + p_1 \leftrightarrow -1 + x \\ \lambda_2 : &\quad -p_0 + p_2 \leftrightarrow -1 + x^2 \\ \lambda_3 : &\quad p_0 + 2p_1 + 3p_2 \leftrightarrow 1 + 2x + 3x^2 \end{aligned} \tag{26.95}$$

Recommended Exercises. Please check HuskyCT for the homework. Be able to show all your work, step by step! Do *not* use calculators or computer programs to solve any problems!

27 Change of basis*

Example 27.1. In Example 26.56, we showed that $\mathcal{S} := \{q_0, q_1, q_2, q_3\}$ is a basis for $\mathbb{P}_3$. Let $\mathbb{P}_4 \xleftarrow{T} \mathbb{P}_3$ be the linear transformation that multiplies polynomials by p_0, which is the polynomial $p_0(x) := x$. Let $\mathcal{B} := \{p_0, p_1, p_2, p_3, p_4\}$ denote the monomial basis for $\mathbb{P}_4$. We also use the same notation $\mathcal{B}$ for the monomial basis for $\mathbb{P}_3$. These polynomials are just defined by $p_k(x) := x^k$ for k any non-negative integer. Notice that $T(p_k) = p_{k+1}$. Therefore, with respect to the bases $\mathcal{B}$, the linear transformation T takes the simple form

$$ {}_{\mathcal{B}}[T]^{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{27.2} $$

With respect to the basis $\mathcal{S}$ on $\mathbb{P}_3$ and the basis $\mathcal{B}$ on $\mathbb{P}_4$, it takes on a more complicated form. What we have to do is express each $T(q_k)$ (notice that now we input q) in terms of the basis $\mathcal{B}$. Fortunately, this isn't too complicated since we have already expressed the q's in terms of the p's in (26.60). The results are

$$ \begin{aligned} T(q_0) &= T(p_2) + T(p_3) = p_3 + p_4 \\ T(q_1) &= T(p_0) - T(p_1) = p_1 - p_2 \\ T(q_2) &= T(p_0) + T(p_1) + T(p_2) = p_1 + p_2 + p_3 \\ T(q_3) &= T(p_0) - T(p_3) = p_1 - p_4 \end{aligned} \tag{27.3} $$

by linearity of T. Therefore, the matrix for T with respect to $\mathcal{B}$ and $\mathcal{S}$ is

$$ {}_{\mathcal{B}}[T]^{\mathcal{S}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}. \tag{27.4} $$

Example 27.5. Now suppose that a vector space V has two bases $\mathcal{B} := \{\vec{v}_1, \ldots, \vec{v}_n\}$ and $\mathcal{C} := \{\vec{w}_1, \ldots, \vec{w}_n\}$ and let $V \xleftarrow{T} V$ be the linear transformation that sends each basis element of $\mathcal{C}$ to the corresponding basis element of $\mathcal{B}$, namely

$$ T(\vec{v}_k) = \vec{w}_k \tag{27.6} $$

for every $k \in \{1, \ldots, n\}$. Then the matrix for T with respect to these two bases takes a very simple form. In fact, it is the identity matrix!

$$ {}_{\mathcal{B}}[T]^{\mathcal{C}} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}. \tag{27.7} $$

The transformation T is called a *change of basis* linear transformation because it transforms one basis into another. In Example 26.56, the matrix in (26.61) transforms the basis of the p's into the basis of the q's and is given by a linear transformation $\mathbb{P}_3 \xleftarrow{T} \mathbb{P}_3$ defined by $T(p_k) = q_k$ for all $k \in \{0, 1, 2, 3\}$. The *matrix* written down in (26.61) is actually this transformation with respect to the basis $\{p_0, p_1, p_2, p_3\}$ so it is not the identity. Nevertheless, it is still a change of basis. The inverse in (26.62) is actually the matrix associated to the transformation that sends q_k back to p_k with respect to the basis $\{p_0, p_1, p_2, p_3\}$.

We have seen changes of bases several times before. In fact, there was a question on the practice midterm and the actual midterm! Here's the question from the practice midterm.

Problem 27.8. A linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ takes the vector $\vec{v}_1 := \frac{1}{2}\begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix}$ to $\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and takes the vector $\vec{v}_2 := \frac{1}{2}\begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}$ to $-\vec{e}_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$. Find the matrix associated to this linear transformation.

Answer. There are several ways to do this problem. Notice that because $\mathcal{B} := \{\vec{v}_1, \vec{v}_2\}$ and $\mathcal{B}' := \{\vec{e}_2, -\vec{e}_1\}$ are both bases of $\mathbb{R}^2$, the question is asking for the change of basis matrix from the basis $\mathcal{B}$ to the basis $\mathcal{B}'$ (with respect to the standard Euclidean basis). The following items list some possible methods.

i. (Brute force method) The matrix is of the general form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and is assumed to satisfy

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\left(\frac{1}{2}\begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix}\right) \qquad \& \qquad \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\left(\frac{1}{2}\begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}\right). \tag{27.9}$$

This gives a system of linear equations

$$\begin{aligned} -\frac{\sqrt{2}}{2}a + \frac{\sqrt{2}}{2}b &= 0 \\ -\frac{\sqrt{2}}{2}c + \frac{\sqrt{2}}{2}d &= 1 \\ \frac{\sqrt{2}}{2}a + \frac{\sqrt{2}}{2}b &= -1 \\ \frac{\sqrt{2}}{2}c + \frac{\sqrt{2}}{2}d &= 0 \end{aligned} \tag{27.10}$$

which should be solved for the entries a, b, c, and d.

ii. (Using the inverse) Instead of solving for the matrix itself, first solve for the inverse. In other words, find the matrix that takes the vector $\vec{e}_1$ to $-\vec{v}_2$ and the vector $\vec{e}_2$ to $\vec{v}_1$. This is just the matrix

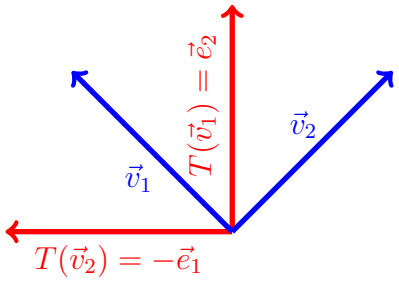
$$\frac{\sqrt{2}}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \tag{27.11}$$

and essentially (except for the fact that we switched the minus sign in front of $\vec{e}_1$) describes the change of basis in the other direction. The inverse of this matrix is

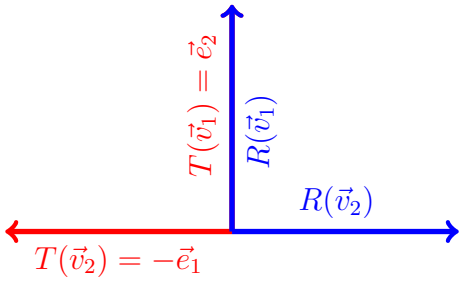
$$\frac{\sqrt{2}}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \tag{27.12}$$

and is the desired transformation. By the way, notice that this is a matrix whose inverse is itself!

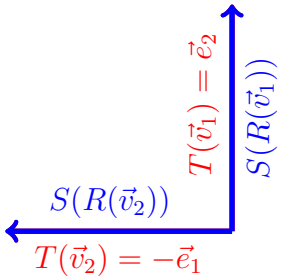
iii. (Visual method) Drawing these vectors out



we can see that if we first rotate $\vec{v}_1$ and $\vec{v}_2$ by $-45°$, let's call this transformation R, then we get (acting on just the vectors $\vec{v}_1$ and $\vec{v}_2$)



If we then reflect through the vertical direction, calling this transformation S, we get



Therefore $T = SR$ which gives

$$\left(\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\right)\left(\frac{\sqrt{2}}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\right) = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{27.13}$$

Proposition 27.14. *Let U, V, and W be three vector spaces and let $W \overset{S}{\leftarrow} V$ and $V \overset{T}{\leftarrow} U$ be two linear transformations. Then the function $W \overset{ST}{\longleftarrow} U$, defined by*

$$(ST)(\vec{u}) := S\big(T(\vec{u})\big) \tag{27.15}$$

for all vectors $\vec{u}$ in U, is a linear transformation, called the composition *of S and T (also said as "T followed by S"—notice the flip in direction). Furthermore, for any vector space V, the function $V \overset{\mathrm{id}_V}{\longleftarrow} V$ given by*

$$\mathrm{id}_V(\vec{v}) := \vec{v} \tag{27.16}$$

for all vectors $\vec{v}$ in V is a linear transformation, called the identity on V.

We now give some explanation for the notation ${}_{\mathcal{B}}[T]^{\mathcal{C}}$.

Proposition 27.17. *Let U, V, and W be finite-dimensional vector spaces, let $U \overset{S}{\leftarrow} V \overset{T}{\leftarrow} W$ be linear transformations, and let $\mathcal{A}, \mathcal{B}$, and $\mathcal{C}$ be ordered bases for U, V, and W, respectively. Then*

$${}_{\mathcal{A}}[S]^{\mathcal{B}}{}_{\mathcal{B}}[T]^{\mathcal{C}} = {}_{\mathcal{A}}[ST]^{\mathcal{C}}. \tag{27.18}$$

In other words, when we compose linear transformations, we "sum over" the intermediate basis.

Definition 27.19. Let V and W be two vector spaces. An *inverse* of a linear transformation $W \overset{T}{\leftarrow} V$ is a linear transformation $V \overset{S}{\leftarrow} W$ such that

$$ST = \mathrm{id}_V \qquad \& \qquad TS = \mathrm{id}_W. \tag{27.20}$$

When an inverse S exists for T, T is said to be *invertible* or an *isomorphism*. If V and W are two vector spaces and there exists an isomorphism from V to W, then V is said to be *isomorphic* to W.

Proposition 27.21. *Let V and W be two vector spaces. If $W \overset{T}{\leftarrow} V$ is an invertible linear transformation, then there exists a unique inverse, denoted by T^{-1}, to T.*

Proof. Let S and R be two such inverses. Then

$$ST = \mathrm{id}_V, \qquad TS = \mathrm{id}_W, \qquad RT = \mathrm{id}_V, \qquad \& \qquad TR = \mathrm{id}_W. \tag{27.22}$$

Multiplying the first equation on the right by R gives $(ST)R = \mathrm{id}_V R = R$. By associativity of composition, this becomes $S(TR) = R$. Using then the fourth equation from above gives $S = S(\mathrm{id}_W) = R$. ∎

Example 27.23. Let ${}_mM_n$ be the vector space of $m \times n$ matrices. Define a transformation $\mathbb{R}^{mn} \overset{T}{\leftarrow} {}_mM_n$ as follows (to be extra careful and to avoid confusion, I have placed a comma in between the subscripts for the matrix)

$$\begin{bmatrix} a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{1,n} \\ a_{2,1} \\ a_{2,2} \\ \vdots \\ a_{m,n} \end{bmatrix} \longleftarrow T \longmapsto \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \tag{27.24}$$

This transformation is linear (exercise). It is an isomorphism because the inverse is given by

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{mn-1} \\ a_{mn} \end{bmatrix} \longmapsto T^{-1} \longrightarrow \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ a_{n+1} & a_{n+2} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{(m-1)n+1} & a_{(m-1)n+2} & \cdots & a_{mn} \end{bmatrix}. \tag{27.25}$$

Definition 27.26. Let V and W be finite-dimensional vector spaces and let $W \xleftarrow{T} V$ be a linear transformation. The dimension of the image of T is also called the *rank* of T and is denoted by $\operatorname{rank} A$.

Example 27.27. Let $\mathbb{P}_4 \xleftarrow{T} \mathbb{P}_3$ be the linear transformation from Example 27.1 that multiplies polynomials by the polynomial p_0. The image of this transformation is by definition

$$\{T(p) \ : \ p \in \mathbb{P}_3\} \tag{27.28}$$

but what does this mean explicitly? Any polynomial $p \in \mathbb{P}_3$ is of the form

$$p = ap_0 + bp_1 + cp_2 + dp_3. \tag{27.29}$$

Hence,

$$T(p) = aT(p_0) + bT(p_1) + cT(p_2) + dT(p_3) = ap_1 + bp_2 + cp_3 + dp_4. \tag{27.30}$$

Therefore,

$$\{T(p) \ : \ p \in \mathbb{P}_3\} = \operatorname{span}\{p_1, p_2, p_3, p_4\}. \tag{27.31}$$

Since all of these vectors are linearly independent, they form a basis of the image. Since there are four elements here,

$$\operatorname{rank}(T) = 4. \tag{27.32}$$

Theorem 27.33. *Let V and W be finite-dimensional vector spaces and let $W \xleftarrow{T} V$ be a linear transformation. Then*

$$\operatorname{rank} T + \dim \ker T = \dim V. \tag{27.34}$$

Example 27.35. Let $\mathbb{P}_2$ be the vector space of degree 2 polynomials and let $\mathbb{R}^2 \xleftarrow{T} \mathbb{P}_2$ be the linear transformation given by sending a degree 2 polynomial p to the vector

$$T(p) := \begin{bmatrix} p(0) \\ p(1) \end{bmatrix} \tag{27.36}$$

We did this example for HW earlier. We found that the kernel is given by all polynomials of the form $bx - bx^2$. This kernel is spanned by the polynomial $x - x^2$ and is therefore one-dimensional. The dimension of $\mathbb{P}_2$ is 3. Therefore, by the previous theorem, this indicates that the rank of T is 2. Let us check this explicitly by finding a basis for the image of T. Every polynomial can be

expressed as a linear combination of the monomials. Let $p(x) = a + bx + cx^2$ be such a polynomial. Then the image of this vector under T is given by

$$\begin{bmatrix} a \\ a+b+c \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix} + c \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{27.37}$$

Thus, any 2-component in the span of the span of the vectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are in the image of T (since the image of T is a subspace). This subspace is two-dimensional, which agrees with our earlier calculation.

Note: Please look at the invertible matrix theorem in Section 4.6 of [Lay].

Theorem 27.38. *Let V and W be finite-dimensional vector spaces. A linear transformation $W \xleftarrow{T} V$ is invertible if and only if it is one-to-one and onto.*

Proof.
($\Rightarrow$) Suppose T is invertible. Then the equations $T^{-1}T = \mathrm{id}_V$ and $TT^{-1} = \mathrm{id}_W$ both hold. The second one holds if and only if T is onto (exercise). The first one holds if and only if T is one-to-one (exercise).
($\Leftarrow$) Suppose T is one-to-one and onto. Then, $W \xrightarrow{T^{-1}} V$ exists as a function. We need to check that T^{-1} is linear. Let $\vec{w}_1$ and $\vec{w}_2$ be two vectors in W and let c be a real number. Let $\vec{v}_1 := T^{-1}(\vec{w}_1)$ and $\vec{v}_2 := T^{-1}(\vec{w}_2)$. Then $\vec{v}_1 + \vec{v}_2$ and $T^{-1}(\vec{w}_1 + \vec{w}_2)$ are two vectors in V (we want to show they are equal). Applying T to the first one gives

$$T(\vec{v}_1 + \vec{v}_2) = T^{-1}(\vec{w}_1) + T^{-1}(\vec{w}_2) = \vec{w}_1 + \vec{w}_2 \tag{27.39}$$

by linearity of T. Applying T to the second one gives

$$T\big(T^{-1}(\vec{w}_1 + \vec{w}_2)\big) = \vec{w}_1 + \vec{w}_2 \tag{27.40}$$

because T is onto. Thus, because T is one-to-one, the vectors $T^{-1}(\vec{w}_1 + \vec{w}_2)$ and $\vec{v}_1 + \vec{v}_2$ must be equal proving that

$$T^{-1}(\vec{w}_1 + \vec{w}_2) = T^{-1}(\vec{w}_1) + T^{-1}(\vec{w}_2). \tag{27.41}$$

A similar proof shows that $T^{-1}(c\vec{w}) = cT^{-1}(\vec{w})$. ■

Finite-dimensional vector spaces are classified by their dimension.

Theorem 27.42. *Let V and W be two finite-dimensional vector spaces. Then $\dim V = \dim W$ if and only if V and W are isomorphic.*

Proof.
($\Rightarrow$) Suppose $n := \dim V = \dim W$. By definition, this means there exist a basis $\mathcal{B}_V = \{\vec{v}_1, \ldots, \vec{v}_n\}$ for V and a basis $\mathcal{B}_W = \{\vec{w}_1, \ldots, \vec{w}_n\}$ for W. Define a linear transformation $W \xleftarrow{T} V$ as follows. First, define T on the basis $\mathcal{B}_V$ by

$$T\big(\vec{v}_i\big) := \vec{w}_i \tag{27.43}$$

for all $i \in \{1, 2, \ldots, n\}$. Then, since $\mathcal{B}_V$ is a basis for V, for any vector $\vec{v}$ in V, there exists a unique expression of the form (Theorem 26.39)

$$\vec{v} = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n \tag{27.44}$$

Then, set

$$T(\vec{v}) := a_1\vec{w}_1 + \cdots + a_n\vec{w}_n. \tag{27.45}$$

By construction, T is a linear transformation. It is invertible because the inverse T^{-1} is constructed in a similar fashion by

$$T^{-1}(\vec{w}_j) := \vec{v}_j \tag{27.46}$$

and extended linearly as for T. Thus, V and W are isomorphic.
($\Leftarrow$) Let $n := \dim V$ and $m := \dim W$. Let $W \xleftarrow{T} V$ be a linear isomorphism (by assumption, one exists). Let $\mathcal{B}_V := \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for V. Because T is one-to-one, $\mathcal{B}_W := \{T(\vec{v}_1), \ldots, T(\vec{v}_n)\}$ is a linearly independent set in W. Because T is onto, every vector $\vec{w}$ can be expressed as $T(\vec{v})$ for some $\vec{v}$ in V. But since $\mathcal{B}_V$ is a basis of V, there exist unique coefficients such that

$$\vec{v} = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n. \tag{27.47}$$

By linearity of T,

$$\vec{w} = T(\vec{v}) = a_1T(\vec{v}_1) + \cdots + a_nT(\vec{v}_n) \tag{27.48}$$

showing that $\mathcal{B}_W$ spans W. Thus $\mathcal{B}_W$ is a basis for W. Since $\mathcal{B}_W$ has n elements, $m = n$. ■

The basis $\mathcal{B}_W := \{T(\vec{v}_1), \ldots, T(\vec{v}_n)\}$ in the proof of the above theorem is analogous to the column vectors for a matrix.

Example 27.49. Look back at Example 27.23. A basis of ${}_mM_n$ is given by the matrices $\{E_{ij}\}$ with $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$. Then, the linear transformation T described in that example is exactly the linear transformation taking the basis

$$\{E_{11}, E_{12}, \ldots, E_{1n}, E_{21}, E_{22}, \ldots, E_{mn}\} \tag{27.50}$$

for ${}_mM_n$ to the basis

$$\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_{mn}\} \tag{27.51}$$

for $\mathbb{R}^{mn}$ and extended linearly as in the proof of Theorem 27.42.

Proposition 27.52. *Let V and W be vector spaces, let $\mathcal{B}$ be a basis for V, and let $W \xleftarrow{T} V$ be a linear transformation. Then T is completely determined by its value on the vectors in $\mathcal{B}$.*

Proof. The proof of this is similar to the proof of Theorem 27.42. In fact, we implicitly assumed it already when working through several examples. ■

Theorem 27.42 can be applied in the following way by setting $W := V$.

Corollary 27.53. *Let V be a finite-dimensional vector space, say of dimension n, and let $\mathcal{B} := \{\vec{v}_1, \ldots, \vec{v}_n\}$ and $\mathcal{C} := \{\vec{w}_1, \ldots, \vec{w}_n\}$ be two bases for V. Then there exists a unique linear isomorphism $V \xleftarrow{T} V$ such that $T(\vec{v}_k) = \vec{v}_k$ for all $k \in \{1, \ldots, n\}$.*

Recommended Exercises. Please check HuskyCT for the homework. Please show your work! Do *not* use calculators or computer programs to solve any problems!

In this lecture, we covered Sections 4.4, 4.6, and 4.7.

References

[1] Otto Bretscher, *Linear algebra with applications*, 3rd ed., Prentice Hall, 2005.

[2] David C. Lay, *Linear algebra and its applications*, 4th ed., Pearson, 2011.

[3] David C. Lay, Steven R. Lay, and Judi J. McDonald, *Linear algebra and its applications*, 5th ed., Pearson, 2015.

[4] G. Polya, *How to solve it: A new aspect of mathematical method*, Princeton University Press, 2014.

[5] Jun John Sakurai, *Modern quantum mechanics; rev. ed.*, Addison-Wesley, Reading, MA, 1994.

[6] Jeffrey R. Weeks, *The shape of space*, Second, Monographs and Textbooks in Pure and Applied Mathematics, vol. 249, Marcel Dekker, Inc., New York, 2002.

[7] Wikipedia, *Michaelis?menten kinetics — wikipedia, the free encyclopedia*, 2017. [Online; accessed 23-October-2017].