

calc,3d,arrows

tdplot<sub>screen</sub>coords/.style = x = (1cm,0cm),y = (0cm,1cm),z = (-1cm,-1cm)



dofile("quad.lua")

[Scale=5.5]Paint the Sky Applied Mathematics

[Scale=3.0]Arial Peter Staab

[Scale=2.0]Arial Fitchburg State University

January 2, 2024

This edition was last edited on January 2, 2024.

Copyright by Peter Staab, Jamaica Plain, Massachusetts.

Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States  
License.

# Contents



# Chapter 1

## Basics of Linear Algebra

### 1.1 Solving Linear Systems

We start this section with three examples of linear systems. One is about a biathlon with running and biking legs, a traffic flow model and finally an example from Chemistry. Examples of linear systems are found from every mathematical subfield and every science.

Travis runs 6 mph and bikes 18 mph in a race with two events. If the course is 29 miles long and it takes him 2 hours and 10 minutes to complete the race, how long is each segment?

In this case, we need to know how long each segment is. There are two legs to the race, so we will let

$$\begin{aligned}x_1 &= \text{number of miles for the run,} \\x_2 &= \text{number of miles for the bike.}\end{aligned}$$

The equations come from three statements in the problem above. First, we know that the total course is 29 miles long or

$$x_1 + x_2 = 29.$$

The remaining equation arises from the time it takes Travis to complete the race. In general recall the relationship between speed (a rate), distance and time is

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

or solving for time,

$$\text{time} = \frac{\text{distance}}{\text{rate}}$$

For example, the time it takes Travis to finish the running leg is  $x_1/6$ . The total time it takes Travis to finish the race is

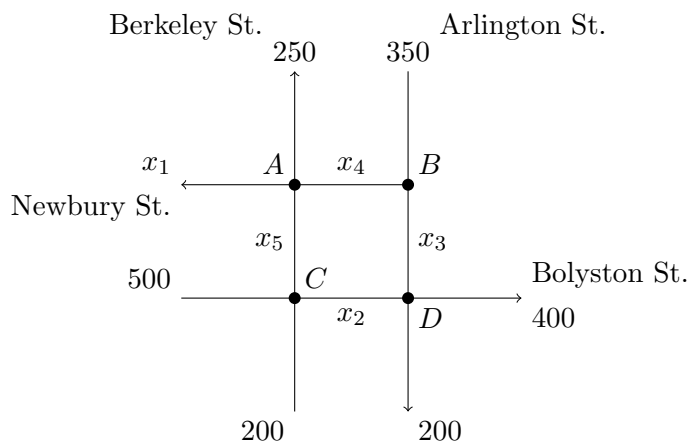
$$\frac{x_1}{6} + \frac{x_2}{18} = 2 + \frac{10}{60} = \frac{13}{6},$$

so the linear system is

$$\begin{aligned}x_1 + x_2 &= 29, \\ \frac{x_1}{6} + \frac{x_2}{18} &= \frac{13}{6}.\end{aligned}$$

The next example examines the traffic flow in a very limited part of a city (Boston in this case).

A simple model of traffic flow can be represented by the following graph:



where the arrows denote the direction of traffic flow (all of these streets are one-way) and the numbers represent the numbers of cars driving down the street in a given time period. The letters  $A$  through  $D$  will be the names of the intersections.

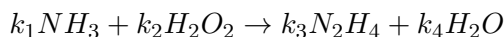
And a mathematical model can be written down by making sure that the total number of cars that go into an intersection balances with the number of cars exiting. In this case, we have:

$$\begin{aligned}x_4 + x_5 &= x_1 + 250 \\ 350 &= x_3 + x_4 \\ 700 &= x_2 + x_5 \\ x_3 + x_2 &= 600\end{aligned}$$

for intersections  $A$ ,  $B$ ,  $C$  and  $D$  respectively.

Linear equations exist in many different fields, especially the sciences. The following comes from Chemistry in which a chemical equation is found by balancing the number of atoms in the equation on both sides.

Hydrazine ( $N_2H_4$ ) is an important nitrogen-based compound. A chemical reaction to produce it from ammonia and hydrogen peroxide is given by





The values of  $k_1, k_2, k_3$  and  $k_4$  can be found by solving the following linear system which balances the number of nitrogen (N), hydrogen (H) and oxygen (O) atoms in the reaction or

$$\begin{aligned}k_1 &= 2k_3, \\3k_1 + 2k_2 &= 4k_3 + 2k_4, \\2k_2 &= k_4.\end{aligned}$$

With examples of linear algebra presented, we next move to a method of solution.

### 1.1.1 Gauss's Method

We now seek a method to solve the above linear systems (and any linear system). Gauss' method and variations of it are the standard ways that all large linear systems are solved presently. We first start with defining a few important terms:

- A **linear combination** of  $x_1, x_2, x_3, \dots, x_n$  has the form

$$a_1x_1 + a_2x_2 + \dots + a_nx_n,$$

where the numbers  $a_1, a_2, \dots, a_n \in \mathbb{R}$  are the combinations coefficients.

- A **linear equation** has the form

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b \tag{1.1}$$

where  $b \in \mathbb{R}$  is a constant.

- The  $n$ -tuple  $(s_1, s_2, \dots, s_n)$  **satisfies** or is a solution of (??) if this point satisfies (??) or

$$a_1s_1 + a_2s_2 + \dots + a_ns_n = b$$

- A **system of linear equations** or linear system is a set of linear equations:

$$\begin{aligned}a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,n}x_n &= b_1, \\a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,n}x_n &= b_2, \\&\vdots \\a_{m,1}x_1 + a_{m,2}x_2 + \dots + a_{m,n}x_n &= b_m,\end{aligned} \tag{1.2}$$

and this linear system has  $m$  equations and  $n$  unknowns (variables).

- The  $n$ -tuple  $(s_1, s_2, \dots, s_n)$  **satisfies** or is a solution of (??) if this point satisfies every equation of (??).

The following are linear equations:

$$\begin{array}{ll} 2x_1 + 3x_2 = 6, & x_1 - x_2 + x_3 - x_4 = 10, \\ 10x_1 - x_3 + 5x_5 = 9, & \sum_{i=1}^{10} ix_i = 0 \end{array}$$

and the following are not:

$$\begin{array}{ll} x_1^2 + x_2 = 6, & x_1x_2 + x_3 = 5, \\ \frac{x_1 + x_2}{x_3} = 6, & \sin(x + y) = z \end{array}$$

The next two examples give a way to determine if a point or  $n$ -tuple is a solution to a linear system.

Show that the point  $(2, 3)$  is a solution of the linear system:

$$\begin{array}{l} 3x_1 - x_2 = 3 \\ 2x_1 + 4x_2 = 16 \end{array}$$

Substitute  $x_1 = 2$  and  $x_2 = 3$  into both equations and check.

$$\begin{array}{l} 3(2) - 3 = 3, \\ 2(2) + 4(3) = 16. \end{array}$$

Since each equation is satisfied at the point  $(2, 3)$  is a solution to the linear system.

Recall that the linear system in Example ?? can be written:

$$\begin{array}{ll} k_1 & - 2k_3 = 0, \\ 2k_2 + 2k_3 - 2k_4 & = 0, \\ 2k_2 & - k_4 = 0, \end{array}$$

Is  $(8, 4, 4, 8)$  a solution to this linear system? Is  $(4, 2, 2, 4)$ ?  
Is  $(6, 1, 3, 2)$ ?

For  $(8, 4, 4, 8)$ , we need to substitute this point in and check all the equations:

$$\begin{array}{l} 8 - 2(4) = 0 \\ 2(4) + 2(4) - 2(8) = 8 + 8 - 16 = 0 \\ 2(4) - 8 = 8 - 8 = 0 \end{array}$$

so it is a solution.

For  $(4, 2, 2, 4)$ , again check all of the equations:

$$\begin{aligned} 4 - 2(2) &= 0, \\ 3(4) + 2(2) - 4(2) - 2(4) &= 12 + 4 - 8 - 8 = 0, \\ 2(2) - 4 &= 0 \end{aligned}$$

so this is also a solution.

For  $(6, 1, 3, 2)$ , we substitute this into the equation:

$$\begin{aligned} 6 - 2(3) &= 0 \\ 3(6) + 2(1) - 4(3) - 2(2) &= 18 + 2 - 12 - 4 = 4 \neq 0. \\ 2(1) - 2 &= 0 \end{aligned}$$

and it satisfies 2 of the equations, but not all, so this is not a solution.

### 1.1.2 Solving a Linear System with Gauss' Method

The following example shows how we can solve a linear system with a technique called **Gauss' Method**.

$$\begin{aligned} x - 2y + 3z &= 6, \\ 2x + 2y &= 6, \\ -3x &+ z = 0. \end{aligned}$$

If one equation had only one variable, it would be quite easy to find the solution to that equation. The following steps will result in such a system. For example, if the first equation is multiplied by  $-2$  and added to the second equation and the second equation is replaced (which we denote with  $-2R_1 + R_2 \rightarrow R_2$ ), then the above equations are replaced with

$$\begin{aligned} -2R_1 + R_2 \rightarrow R_2 \quad \begin{aligned} x - 2y + 3z &= 6, \\ 6y - 6z &= -6, \\ -3x &+ z = 0. \end{aligned} \end{aligned}$$

Next, we'd like to get rid of the  $x$  term in the 3rd equation. We can do that by multiplying the first equation by 3 and adding to the 3rd and writing down this operation as well we get:

$$\begin{aligned} 3R_1 + R_3 \rightarrow R_3 \quad \begin{aligned} x - 2y + 3z &= 6, \\ 6y - 6z &= -6, \\ -6y + 10z &= 18. \end{aligned} \end{aligned}$$

Next we will eliminate the  $y$  term in the 3rd equation. We can do that by adding equations 2 and 3 and putting the result in equation 3.

$$\begin{array}{rcl} & x - 2y + 3z = 6, & \\ R_2 + R_3 \rightarrow R_3 & 6y - 6z = -6, & (1.3) \\ & 4z = 12. & \end{array}$$

At this point we can solve for  $z$  in the third equation of (??) to get  $z = 3$ . From this, plug in  $z = 3$  into the second equation of (??) to get

$$\begin{aligned} 6y - 6(3) &= -6 \\ 6y &= 12 \\ y &= 2 \end{aligned}$$

and finally we can use  $z = 3$  and  $y = 2$  to substitute into the first equation of (??) to get

$$\begin{aligned} x - 2(2) + 3(3) &= 6 \\ x - 4 + 9 &= 6 \\ x &= 1 \end{aligned}$$

therefore,  $x = 1$ . The point  $(1, 2, 3)$  is a solution to linear system. Because this is the only point that satisfies all equations, this point is **unique**. We will also use the terminology triple to describe  $(1, 2, 3)$ .

[Gauss's Method] If a linear system  $(A)$  is changed into a second linear system  $(B)$  by one of the following operations:

1. Two equations of the linear system are swapped,
2. An equation is multiplied by a nonzero number,
3. An equation is replaced by the sum of itself and a multiple of another equation,

then linear systems  $(A)$  and  $(B)$  have the same set of solutions.

*Proof.* We will consider only the first operation in this proof. Let's assume that we swap equations  $i$  and  $j$ , thus system  $(A)$ ,

$$\begin{array}{l} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1, \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2, \\ \vdots \\ a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n = b_i, \\ \vdots \\ a_{j,1}x_1 + a_{j,2}x_2 + \cdots + a_{j,n}x_n = b_j, \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m, \end{array}$$

is transformed to

$$\begin{array}{rcl}
 a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n & = & b_1, \\
 a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n & = & b_2, \\
 & \vdots & \vdots \\
 a_{j,1}x_1 + a_{j,2}x_2 + \cdots + a_{j,n}x_n & = & b_j, \\
 & \vdots & \vdots \\
 a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n & = & b_i, \\
 & \vdots & \vdots \\
 a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n & = & b_m,
 \end{array}$$

Let  $(s_1, s_2, \dots, s_n)$  be a solution of  $(A)$  if it exists and note it may be one of many  $n$ -tuples in the solution. Thus it satisfies each equation of linear system  $(A)$ . Since the exact same equations are in  $(B)$  in just a different order,  $(s_1, s_2, \dots, s_n)$  is a solution to  $(B)$ . If there is more than one  $n$ -tuple in the solution to  $(A)$ , repeat this for every one. If there is no solution to  $(A)$ , then there will be no solution to  $(B)$  since it is the same set of equations.

Proof of #2 and #3 above are quite similar and are not shown. □

The three operations in theorem ?? are called the **elementary reduction operations** or **elementary row operations** or **Gaussian operations** of a linear system.

Although Gauss' Method is very flexible, generally, one tries to eliminate all  $x_1$  terms in all equations below equation 1,  $x_2$  terms in all equations below equation 2 and so on.

### A fourth Row Operation

Another handy row operation is that of replacing a row with linear combination of itself and another row. In general this is

$$c_i R_i + c_j R_j \rightarrow R_j$$

Since this is a combination of two row operations together, we will use it, however, not consider it a separate (fourth) row operation.

Use the strategy above to solve:

$$\begin{aligned} 4x_1 - x_3 &= 0, \\ x_1 + 3x_2 + 2x_3 &= 3, \\ 3x_2 + 5x_3 &= 14. \end{aligned}$$

We first try to eliminate the  $x_1$  term in the 2nd row. We can accomplish this with

$$\begin{aligned} -R_1 + 4R_2 &\rightarrow R_2 & \begin{aligned} 4x_1 - x_3 &= 0, \\ 12x_2 + 9x_3 &= 12, \\ 3x_2 + 5x_3 &= 14. \end{aligned} \end{aligned}$$

Lastly, we try to eliminate the  $x_2$  term in the 3rd equation with

$$\begin{aligned} -R_2 + 4R_3 &\rightarrow R_3 & \begin{aligned} 4x_1 - x_3 &= 0, \\ 12x_2 + 9x_3 &= 12, \\ 11x_3 &= 44 \end{aligned} \end{aligned}$$

Now at this point, we can solve for  $x_3$ , then  $x_2$ , then  $x_1$  in a method called **back substitution**.

$$\begin{aligned} 11x_3 &= 44 \\ x_3 &= 4 \end{aligned}$$

then substitute into the second equation,

$$\begin{aligned} 12x_2 + 9(4) &= 12 \\ 12x_2 &= 12 - 36 = -24 \\ x_2 &= -2 \end{aligned}$$

and finally into the first equation,

$$\begin{aligned} 4x_1 - 4 &= 0 \\ 4x_1 &= 4 \\ x_1 &= 1. \end{aligned}$$

So the solution is  $(1, -2, 4)$ . Again, because there is only one solution to the equation  $11x_3 = 44$ , there is only one value for  $x_3$ , hence only one value for  $x_2$  and only one values for  $x_1$ , so this solution is unique.

Solve the linear system for the biathlon race in example ??

$$\begin{aligned} x_1 + x_2 &= 29, \\ \frac{x_1}{6} + \frac{x_2}{18} &= \frac{13}{6}, \end{aligned}$$

First to make life easier, let's multiple the 2nd equation by 18 to rid the fractions.

$$\begin{aligned} 18R_2 \rightarrow R_2 \qquad \qquad \qquad x_1 + x_2 &= 29, \\ 3x_1 + x_2 &= 39, \end{aligned}$$

Then we eliminate the  $x$  in the 2nd equation by doing the following step,

$$\begin{aligned} -3R_1 + R_2 \rightarrow R_2 \qquad \qquad \qquad x_1 + x_2 &= 29, \\ -2x_2 &= -48, \end{aligned}$$

and at this point, we can solve for  $x_2$  in the last equation using back substitution.  $x_2 = 24$  and then

$$x_1 + 24 = 29$$

therefore,

$$x_1 = 5$$

So the running part of the race is 5 miles and the biking is 24 miles.

### 1.1.3 Echelon Form

Notice that the strategy laid out above gets the linear system in a nice form for solving for all variables using back substitution. We will now explicitly state what this nice form is.

In each equation of a linear system, the first variable with a non-zero coefficient is called the **leading variable**. A system is in **echelon form** if each leading variable is to the right of the leading variable of the equation above it (except the first equation).

In the definition above, it's important that there is some order to the variables. If the variables are  $x_1, x_2, \dots$ , then “to the right” means that there is a larger subscript. If the variables are  $x, y, z, \dots$ , the to the right is just after lexicographically.

Show that the linear system in the example above:

$$\begin{aligned}4x_1 & \quad - x_3 = 0, \\12x_2 + 9x_3 & = 12, \\11x_3 & = 44\end{aligned}$$

is in echelon form.

Note that the leading variable in equation 1 is  $x_1$ , the leading variable in equation 2 is  $x_2$  and the leading variable in equation 3 is  $x_3$ . Each leading variable in each successive row is to the right of the one above.

The next example determines if the system is in echelon form.



Is

$$\begin{aligned} x_1 + 3x_3 - 9x_4 + 11x_5 &= 14, \\ 2x_3 + 4x_5 &= 10, \\ 3x_5 &= 27, \end{aligned}$$

in echelon form?

Yes, the leading variable in equation 1 is  $x_1$ , the leading variable in equation 2 is  $x_3$  which is to the right of  $x_1$  and the leading variable in equation 3 is  $x_5$ , which is to the right of  $x_3$ .

Is the system

$$\begin{aligned} 3x_1 + 4x_2 &= 7, \\ 2x_1 - x_2 &= -10, \\ 2x_2 &= 8 \end{aligned}$$

in echelon form:

No, the leading variables in both equation 1 and 2 are  $x_1$ .

Let's look more carefully at the last example. It is not in echelon form, but let's put it in echelon form and write down a solution.

$$\begin{aligned} 3x_1 + 4x_2 &= 7, \\ 2x_1 - x_2 &= -10, \\ 2x_2 &= 8 \end{aligned}$$

Eliminate  $x_1$  in the 2nd equation by doing the following:

$$\begin{aligned} 3x_1 + 4x_2 &= 7, \\ -2R_1 + 3R_2 &\rightarrow R_2, & -11x_2 &= -44, \\ 2x_2 &= 8 \end{aligned}$$

then eliminate  $x_2$  in the third equation

$$\begin{aligned} 3x_1 + 4x_2 &= 7, \\ 2R_2 + 11R_3 &\rightarrow R_3 & -11x_2 &= -44, \\ 0x_2 &= 0, \end{aligned}$$

Note that the last equation in the last linear system contains no information. zero equals zero. We all knew that before. We ignore that equation. The 2nd equation can be solved to yield  $x_2 = 4$  and substituting into the first equation,

$$\begin{aligned} 3x_1 + 4(4) &= 7, & \text{or} \\ 3x_1 &= -9, & \text{or} \\ x_1 &= -3. \end{aligned}$$

so the solution is  $(-3, 4)$ .

In comparison to the last example, consider the following very similar example:

$$\begin{array}{rcl}
 3x_1 + 4x_2 & = & 7, \\
 2x_1 - x_2 & = & -10, \\
 3x_2 & = & 6 \\
 \\ 
 -2R_1 + 3R_2 & \rightarrow & R_2, \quad \begin{array}{rcl} 3x_1 + 4x_2 & = & 7, \\ -11x_2 & = & -44, \\ 3x_2 & = & 6, \end{array} \\
 \\ 
 3R_2 + 11R_3 & \rightarrow & R_3, \quad \begin{array}{rcl} 3x_1 + 4x_2 & = & 7, \\ -11x_2 & = & -44, \\ 0x_2 & = & -66, \end{array}
 \end{array}$$

And note in this case, that the last equation says  $0 = -66$ , which is not true. Therefore, this means that there is no solution. This is in contrast to the the previous example, which did have a solution.

### Are there only unique and no solutions?

In the last example of this section, we will see that we can have many solutions to a linear system. If we have the following linear system already in echelon form:

$$\begin{array}{rcl}
 x_1 & - & 2x_3 = 0, \\
 2x_2 + 2x_3 - 2x_4 & = & 0, \\
 -2x_3 + x_4 & = & 0,
 \end{array}$$

Notice that since there are 4 variables and 3 equations, that there is no way to get a solution for  $x_4$  by itself (or any of the other 3 variables). This has many (actually infinitely many) solutions. For example  $(0, 0, 0, 0)$  is one as well as  $(2, 1, 1, 2)$  and  $(-5, -5/2, -5/2, -5)$ . We will explore this in much more detail later.

#### 1.1.4 Describing the Solution Set

So far we have seen linear system with solution sets with different number of elements. The *no solution* means that the solution set is empty, the *unique solution* means that there is one element in the set. But we haven't seen the *many solutions* and how do we represent that. We explore that here in this section.

For example, if we reduce the linear system:

$$\begin{array}{rcl}
 x & + 2z & = -6 \\
 2x - y & & = 1 \\
 & y + 4z & = -13 \\
 \\ 
 -2R_1 + R_2 \rightarrow R_2 & & \\
 \begin{array}{rcl}
 x & + 2z & = -6 \\
 & -y - 4z & = 13 \\
 & y + 4z & = -13
 \end{array} \\
 \\ 
 R_2 + 2R_3 \rightarrow R_3 & & \\
 \begin{array}{rcl}
 x & + 2z & = -6 \\
 & -y - 4z & = 13 \\
 & 0 & = 0
 \end{array}
 \end{array}$$

As discussed above, this linear system has more than one solution. Therefore it is the set of points that satisfy every one of the original three equations or

$$\{(x, y, z) \mid x + 2z = -6, 2x - y = 1, y + 4z = -13\}$$

However (and this is the main point of Gauss' method), using row operations, any points that satisfy the 3rd system above is also a solution, that is

$$\{(x, y, z) \mid x + 2z = -6, -y - 4z = 13\}$$

This may be an improvement because there are only 2 equations instead of 3, but it's still difficult to write down a point that is in the solution set.

Alternatively, we return to the last linear system and solve the 2nd equation for  $y$  or

$$y = -4z - 13$$

and solving the first equation for  $x$  in the linear system above to get

$$x + 2z = -6$$

or

$$x = -2z - 6$$

And now if we know  $z$ , we can find  $x$  and  $y$ , thus we can write the solution set as

$$\{(x, y, z) = (-2z - 6, -4z - 13, z) \mid z \in \mathbb{R}\}$$

and now we have a lot of information about the solution set. For example, if  $z = 0$ , then  $x = -6, y = -13$ , so the point  $(-6, -13, 0)$  is in the solution set. Similarly if  $z = -3$ , then  $(0, -1, -3)$  is also a point in the solution set and any point given a value of  $z$ .

We also have a sense of the number of solutions and since  $z$  can take on any real number, there are an infinite number of triples 3-tuples in this solution set.

Also, the form of the solution set above is in **set builder** form. That is it shows how to create the set of points in the solution.

The non-leading variables in a linear system in echelon form are called the **free variables** or **parameters**.

The next example shows that a linear system can have more than one free variables.

In Example ??, we saw the following linear system already in echelon form:

$$\begin{aligned} x_1 + 3x_3 - 9x_4 + 11x_5 &= 14, \\ 2x_3 + 4x_5 &= 10, \\ 3x_5 &= 27, \end{aligned}$$

Write the solution set in terms of the free variables.

First, we use back substitution to solve for the *leading variables*. The third equation states that  $x_5 = 9$ , and then substitute this into the 2nd equation:

$$2x_3 + 4(9) = 10,$$

or

$$x_3 = -13,$$

and then substitute into the first equation

$$x_1 + 3(-13) - 9x_4 + 11(9) = 14$$

or

$$x_1 = 9x_4 - 46$$

Lastly, we write the solution in terms of  $x_2$  and  $x_4$  which are the two free variables.

$$\{(x_1, x_2, x_3, x_4, x_5) = (9x_4 - 46, x_2, -13, x_4, 9) \mid x_2, x_4 \in \mathbb{R}\}$$

We now find the solution to the Hydrazine chemistry problem in Example ??.

Find the solution set of the linear system in Example ??:

$$\begin{aligned} k_1 - 2k_3 &= 0, \\ 3k_1 + 2k_2 - 4k_3 - 2k_4 &= 0, \\ 2k_2 - k_4 &= 0, \end{aligned}$$

In this case, we will use Gauss's method to put the linear system in echelon form:

$$\begin{array}{rcl} & k_1 & - 2k_3 = 0, \\ -3R_1 + R_2 \rightarrow R_2, & 2k_2 + 2k_3 - 2k_4 = 0, \\ & 2k_2 & - k_4 = 0, \end{array}$$

$$\begin{array}{rcl} & k_1 & - 2k_3 = 0 \\ -R_2 + R_3 \rightarrow R_3, & 2k_2 + 2k_3 - 2k_4 = 0, \\ & -2k_3 + k_4 = 0, \end{array}$$

which is now in echelon form. We now use back substitution to solve for  $k_1$ ,  $k_2$  and  $k_3$ , the leading variables:

$$k_3 = \frac{1}{2}k_4$$

then

$$k_2 = -k_3 + k_4 = -\frac{1}{2}k_4 + k_4 = \frac{1}{2}k_4,$$

and finally,

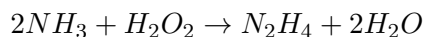
$$k_1 = 2k_3 = 2\left(\frac{1}{2}k_4\right) = k_4$$

Since  $k_1$ ,  $k_2$  and  $k_3$  all depend on  $k_4$ , a free variable, there are an infinite number of solutions to this linear system. We can write this as

$$\{(k_4, \frac{1}{2}k_4, \frac{1}{2}k_4, k_4) \mid k_4 \in \mathbb{R}\},$$

Solutions include  $(0, 0, 0, 0)$ ,  $(2, 1, 1, 2)$ ,  $(-5, -5/2, -5/2, -5)$ .

Since the above example comes from Chemical equations, it is desirable to find a solution of all positive integers, since the number represent the number of atoms in the reaction. In addition, typically the solution with smallest integers is desired. In this case, the solution  $(2, 1, 1, 2)$  results in the following equation:



### 1.1.5 Matrices

An  $m$  by  $n$  **matrix** is a rectangular grid of numbers with  $m$  rows and  $n$  columns. The **size** of a matrix is the pair of number  $m$  and  $n$  and is typically written  $m$  by  $n$  or  $m \times n$ . Each number of the matrix is called the **entry** or **element**.

The following are examples of matrices:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \begin{bmatrix} \pi & 2 \\ -1 & 4 \\ 2.5 & 7/3 \\ -11 & 1000 \end{bmatrix} \quad \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \\ 9 \end{bmatrix} \quad [8 \quad -3 \quad 2]$$

A common notation for a matrix is to use upper case letters (and often bold), for example  $A$ ,  $B$  and  $C$  are common matrices. The entry or element in the  $i$ th row and  $j$ th column of  $A$  is  $a_{i,j}$ . Also, the set of all  $m$  by  $n$  matrices is denoted  $\mathcal{M}_{m \times n}$ .

A matrix with only 1 row is called a **row vector**. A matrix with only 1 column is called a **column vector**. The **size** of a vector is the number of rows (for a column vector) or the number of columns (for a row vector).

The 3rd matrix in Example ?? above is a column vector of size 5. The 4th matrix is a row vector of size 3.

## Matrices and Linear Systems

Let's look at the following linear system.

$$\begin{aligned} x + 2y &= 7, \\ 4x - 3y &= 6. \end{aligned}$$

When we solved it above, the variables didn't play much of a role, the coefficients changed and the variables stayed in the same place as we solved for them. Because of this, we will adopt that technique to only works with the coefficients—the numbers themselves—as a grid or as it is called an **augmented coefficient matrix**. The following is this matrix for the linear system above.

$$\left[ \begin{array}{cc|c} 1 & 2 & 7 \\ 3 & -1 & 6 \end{array} \right]$$

and the two rows of the matrix represent the two equations. The first column represents the coefficient of the  $x$  variable, and the second column represents the coefficient of the  $y$  variable. The last column is the right hand sides of the linear system, and the vertical line occurs where the equal sign is in the equations.

Write down the following linear system as an augmented matrix:

$$\begin{aligned} 3x - 2y + z &= 11, \\ 2x + 4y - 3z &= 2, \\ -4x + y + 4z &= 0. \end{aligned}$$

$$\left[ \begin{array}{ccc|c} 3 & -2 & 1 & 11 \\ 2 & 4 & -3 & 2 \\ -4 & 1 & 4 & 0 \end{array} \right]$$

And the next example shows how to rewrite a matrix as a linear system.

Write down the following augmented matrix as a linear system:

$$\left[ \begin{array}{ccc|c} 3 & 0 & -1 & 7 \\ -2 & 1 & 5 & 6 \\ 0 & 3 & 4 & 8 \end{array} \right]$$

In this case, we are doing the opposite of the step above. That is, take a matrix and find the linear system.

Since the matrix has 4 columns, there are 3 variables (the last column is the right hand side). Let's use  $x$ ,  $y$  and  $z$ .

$$\begin{aligned} 3x + 0y - z &= 7 \\ -2x + y + 5z &= 6 \\ 0x + 3y + 4z &= 8 \end{aligned}$$

(We could have used  $x_1, x_2$  and  $x_3$  alternatively.)

The main advantage of matrices is to simplify the solving of linear systems. We will see this with the example above:

$$\begin{aligned} 4x_1 - x_3 &= 0, \\ x_1 + 3x_2 + 2x_3 &= 3, \\ 3x_2 + 5x_3 &= 14. \end{aligned}$$

We will write this as an augmented coefficient matrix, then use row operations as we saw for the linear system:

$$\begin{aligned} & \left[ \begin{array}{ccc|c} 4 & 0 & -1 & 0 \\ 1 & 3 & 2 & 3 \\ 0 & 3 & 5 & 14 \end{array} \right] \\ -R_1 + 4R_2 \rightarrow R_2, & \left[ \begin{array}{ccc|c} 4 & 0 & -1 & 0 \\ 0 & 12 & 9 & 12 \\ 0 & 3 & 5 & 14 \end{array} \right] \\ -R_2 + 4R_3 \rightarrow R_3, & \left[ \begin{array}{ccc|c} 4 & 0 & -1 & 0 \\ 0 & 12 & 9 & 12 \\ 0 & 0 & 11 & 44 \end{array} \right] \end{aligned}$$

and this matrix is now in echelon form, so we can use back substitution to find the solution  $(1, -2, 4)$ .

Solve the following linear system use matrices and Gauss' method.

$$\begin{aligned}x - 2y + 4z &= -15 \\ -3x + 0y + 2z &= -7 \\ x - 2y + 4z &= -15\end{aligned}$$

First write the linear system as a matrix

$$\begin{aligned}& \begin{bmatrix} -1 & -4 & 10 & | & -37 \\ -3 & 0 & 2 & | & -7 \\ 1 & -2 & 4 & | & -15 \end{bmatrix} \\ & \begin{aligned} -3R_1 + R_2 &\rightarrow R_2 \\ R_1 + R_3 &\rightarrow R_3, \end{aligned} \begin{bmatrix} -1 & -4 & 10 & | & -37 \\ 0 & 12 & -28 & | & 104 \\ 0 & -6 & 14 & | & -52 \end{bmatrix} \\ & \begin{aligned} R_2 + 2R_3 &\rightarrow R_3, \end{aligned} \begin{bmatrix} -1 & -4 & 10 & | & -37 \\ 0 & 12 & -28 & | & 104 \\ 0 & 0 & 0 & | & 0 \end{bmatrix} \end{aligned} \tag{1.4}$$

The matrix in (??) is in echelon form, so we use back substitution to solve for  $x_1$  and  $x_2$  in terms of  $x_3$ , the free variable. The 2nd row of (??) can be written:

$$12x_2 = 104 + 28x_3$$

or

$$x_2 = 17 + \frac{7}{3}x_3$$

and using this in the first equation of (??)

$$-x_1 - 4\left(17 + \frac{7}{3}x_3\right) + 10x_3 = -37$$

and solving for  $x_1$

$$x_1 = -31 + \frac{2}{3}x_3$$



### Vectors as Solutions of Linear Systems

The vector

$$\vec{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

with  $s_1, s_2, \dots, s_n$  real numbers is a solution to the linear equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

if

$$a_1s_1 + a_2s_2 + \cdots + a_ns_n = b.$$

### Addition and Scalar Multiplication of Vectors

The **sum of two vectors**  $\vec{u}$  and  $\vec{v}$  is defined as

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}$$

and the multiplication of a vector  $\vec{u}$  by a scalar  $r \in \mathbb{R}$  is

$$r\vec{u} = r \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} ru_1 \\ ru_2 \\ \vdots \\ ru_n \end{bmatrix}$$

Let

$$\vec{u} = \begin{bmatrix} 3 \\ -3 \\ 4 \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} 2 \\ 3/2 \\ 9 \end{bmatrix}.$$

Find  $\vec{u} + \vec{v}$  and  $4\vec{v}$ .

$$\vec{u} + \vec{v} = \begin{bmatrix} 3+2 \\ -3+3/2 \\ 4+9 \end{bmatrix} = \begin{bmatrix} 5 \\ -3/2 \\ 13 \end{bmatrix}, \quad 4\vec{v} = \begin{bmatrix} 4 \cdot 2, \\ 4 \cdot \frac{3}{2}, \\ 4 \cdot 9 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \\ 36 \end{bmatrix}$$

### Writing Solutions in Vector Form

There are a couple of advantages to writing a solution in vector form. We will see some of them later in the course, however, right now, we can write down specific points that are in the solution set quite easily. There's little advantage to using vectors for unique and no solution linear systems, but when you have multiple (infinite number of) solutions, then vectors shine through.

Consider the linear system from page ?? in echelon form:

$$\begin{aligned}x + 2z &= -6 \\ -y - 4z &= 13 \\ 0 &= 0\end{aligned}$$

Recall that the solution in terms of  $z$  as

$$\{(-2z - 6, -4z - 13, z) \mid z \in \mathbb{R}\}$$

Write the solution in terms of a vector.

There is one parameter in this case  $z$  and thus the solution can be written as

$$\begin{aligned}\begin{bmatrix} -2z - 6, \\ -4z - 13 \\ z \end{bmatrix} &= \begin{bmatrix} -2z \\ -4z \\ z \end{bmatrix} + \begin{bmatrix} -6 \\ -13 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} -2 \\ -4 \\ 1 \end{bmatrix} z + \begin{bmatrix} -6 \\ -13 \\ 0 \end{bmatrix}\end{aligned}$$

And this form makes it quite easy to find individual points. If we let  $z = 0$ , then  $(-6, -13, 0)$  is a point in the solution set, if let  $z = -3$ , then  $(0, -1, -3)$  is a point.

Note that it takes a little work to get a linear system in echelon form into a solution in vector form. We will see in the next section that there is another form that the linear system (matrix) can be put into for an easier transition to this form.

Write the solution to the linear system that we saw in Example ??

$$\begin{aligned}x_1 + 3x_3 - 9x_4 + 11x_5 &= 14, \\ 2x_3 + 4x_5 &= 10, \\ 3x_5 &= 27,\end{aligned}$$

in vector form.

In example ??, the solution set was found to be

$$\{(x_1, x_2, x_3, x_4, x_5) = (9x_4 - 46, x_2, -13, x_4, 9) \mid x_2, x_4 \in \mathbb{R}\}$$

with the free variables  $x_2$  and  $x_4$ . This can be written as a vector as

$$\begin{bmatrix} 9x_4 - 46 \\ x_2 \\ -13 \\ x_4 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 9 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} x_4 + \begin{bmatrix} -46 \\ 0 \\ -13 \\ 0 \\ 9 \end{bmatrix}$$

Thus the solution can be written:

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 9 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} x_4 + \begin{bmatrix} -46 \\ 0 \\ -13 \\ 0 \\ 9 \end{bmatrix} \mid x_2, x_4 \in \mathbb{R} \right\}.$$

This is the most general form of the solution to the linear system in this example and as before one can write down solutions with specific values of free variables. For example, if  $x_2$  and  $x_4$  are both 0, the the point  $(-46, 0, -13, 0, 9)$  is a solution to the linear system. (Try it!) This is an example of a particular solution as we define below.

Consider a linear system. If the point  $(s_1, s_2, \dots, s_n)$  is a solution to the system and  $s_1, s_2, \dots, s_n$  do not have any free variables, then the point  $(s_1, s_2, \dots, s_n)$  is called a **particular solution**.

A few things of note before we go on:

- After the first two sections of the text, you should know how to do Gauss's method on any linear system and be able to write down the solution set in either *set builder notation* or as vectors.
- Gauss' method states that any of the 3 elementary row operations results in the same solution. Since there is no algorithm to apply to a linear system (or matrix), if we use 2 different set of row operations, do we get the same solution (or at least the same set of free variables)?

### 1.1.6 General = Particular + Homogeneous

If we return to the solution set in Example ?? in the previous section:

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 9 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} x_4 + \begin{bmatrix} -46 \\ 0 \\ -13 \\ 0 \\ 9 \end{bmatrix} \mid x_2, x_4 \in \mathbb{R} \right\}.$$

where the third vector is a particular solution and the other two vectors are multiplied by free variables (parameters). This form will give us a lot of information about the solution set.

A linear system can be described as

$$\{\vec{p} + c_1\vec{\beta}_1 + c_2\vec{\beta}_2 + \dots + c_n\vec{\beta}_n, \mid c_1, c_2, \dots, c_n \in \mathbb{R}\}$$

where  $\vec{p}$  is any particular solution and the number of vectors  $\vec{\beta}_1, \vec{\beta}_2, \dots, \vec{\beta}_n$  equals the number of free variables that the system has after Gaussian reduction (and in echelon form).

A linear equation is **homogeneous** if the constant term (on the right hand side of each equation) is 0. A linear system is **homogeneous** if all constant terms is 0.

The next two examples show the possible results of a homogeneous system.

Find the solution to the linear system:

$$\begin{aligned} 2x - 3y &= 0, \\ 5x + 2y &= 0, \end{aligned}$$

Using the row operation,  $-5R_1 + 2R_2 \rightarrow R_2$ , we get

$$\begin{aligned} 2x - 3y &= 0, \\ 0x + 19y &= 0, \end{aligned}$$

and from back substitution, we get the solution  $(0, 0)$  and note that this is a unique solution.

Find the solution to the homogeneous linear system from Example ??:

$$\begin{aligned} k_1 &= 2k_3, \\ 3k_1 + 2k_2 &= 4k_3 + 2k_4, \\ 2k_2 &= k_4. \end{aligned}$$

In Example ??, we wrote down the solution

$$\{(k_4, \frac{1}{2}k_4, \frac{1}{2}k_4, k_4) \mid k_4 \in \mathbb{R}\},$$

In vector form the solution is

$$\left\{ \begin{bmatrix} 1 \\ 1/2 \\ 1/2 \\ 1 \end{bmatrix} k_4 \mid k_4 \in \mathbb{R} \right\}.$$

In this case, the solution set has an infinite number of points.

### Non-homogeneous Systems

If a system is not homogeneous, it is called **non-homogeneous**. For a non-homogeneous system, there is an **associated homogeneous system** found by replacing the right hand side with zeros. We now look at the following: Solve

$$\begin{aligned} x + 3y - z &= 9, \\ +y + 3z &= 6, \\ x + 4y + 2z &= 15, \end{aligned}$$

and it's associated homogenous system.

First the solution can be found by

$$\begin{array}{rcl} & x + 3y - z = 9 \\ -R_1 + R_3 \rightarrow R_3, & y + 3z = 6, \\ & y + 3z = 6, \end{array}$$

$$\begin{array}{rcl} & x + 3y - z = 9 \\ -R_2 + R_3 \rightarrow R_3, & y + 3z = 6, \\ & 0 = 0, \end{array}$$

and using back-substitution:

$$\begin{aligned} y &= 6 - 3z \\ x &= 9 - 3y + z = 9 - 3(6 - 3z) + z \\ &= -9 + 10z \end{aligned}$$

so the solution can be written in vector form as

$$\left\{ \begin{bmatrix} 10 \\ -3 \\ 1 \end{bmatrix} z + \begin{bmatrix} -9 \\ 6 \\ 0 \end{bmatrix} \mid z \in \mathbb{R} \right\}$$

and recall that  $(-9, 6, 0)$  is a particular solution.

The associated homogeneous system is

$$\begin{aligned} x + 3y - z &= 0, \\ +y + 3z &= 0, \\ x + 4y + 2z &= 0, \end{aligned}$$

and the solution can be found using the same two row operations as above,

$$\begin{array}{rcl} & x + 3y - z = 0 \\ -R_1 + R_3 \rightarrow R_3, & y + 3z = 0, \\ & y + 3z = 0, \end{array}$$

$$\begin{array}{rcl} & x + 3y - z = 0 \\ -R_2 + R_3 \rightarrow R_3, & y + 3z = 0, \\ & 0 = 0, \end{array}$$

and using backsubstitution:

$$\begin{aligned} y &= -3z \\ x &= -3y + z = -3(-3z) + z = 10z \end{aligned}$$

which can be written in vector form as

$$\left\{ \begin{bmatrix} 10 \\ -3 \\ 1 \end{bmatrix} z \mid z \in \mathbb{R} \right\}$$

The above example indicates that a solution to non-homogeneous system consists of a particular solution and the solution to the associated homogeneous system.

For any linear system with a particular solution  $\vec{p}$ , the solution set is

$$\{\vec{p} + \vec{h} \mid \vec{h} \text{ satisfies the associated linear system}\}$$

*Proof.* To prove this, we will first show that any solution  $\vec{s}$  of the non-homogeneous linear system,

$$a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n = b_i \quad \text{for } i = 1, 2, \dots, m \quad (1.5)$$

then  $\vec{s} - \vec{p}$  satisfies the associated homogeneous system,

$$a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n = 0 \quad \text{for } i = 1, 2, \dots, m \quad (1.6)$$

then show that if  $\vec{h}$  is a solution to the associated homogeneous system, then  $\vec{h} + \vec{p}$  satisfies the non-homogeneous system.

1. Assume that  $\vec{s} = (s_1, s_2, \dots, s_n)$  satisfies (??), that is

$$a_{i,1}s_1 + a_{i,2}s_2 + \cdots + a_{i,n}s_n = b_i \quad \text{for } i = 1, 2, \dots, m$$

then  $\vec{s} - \vec{p}$  satisfies

$$\begin{aligned} & a_{i,1}(s_1 - p_1) + a_{i,2}(s_2 - p_2) + \cdots + a_{i,n}(s_n - p_n) \\ &= a_{i,1}s_1 - a_{i,1}p_1 + a_{i,2}s_2 - a_{i,2}p_2 + \cdots + a_{i,n}s_n - a_{i,n}p_n \\ &= (a_{i,1}s_1 + a_{i,2}s_2 + \cdots + a_{i,n}s_n) - \\ & \quad (a_{i,1}p_1 + a_{i,2}p_2 + \cdots + a_{i,n}p_n) \\ &= b_i - b_i = 0 \end{aligned}$$

the associated homogeneous system.

2. Next assume that  $\vec{h}$  satisfies (??). Then we show that  $\vec{h} + \vec{p}$  satisfies:

$$\begin{aligned} & a_{i,1}(h_1 + p_1) + a_{i,2}(h_2 + p_2) + \cdots + a_{i,n}(h_n + p_n) \\ &= a_{i,1}h_1 + a_{i,2}h_2 + \cdots + a_{i,n}h_n + \\ & \quad a_{i,1}p_1 + a_{i,2}p_2 + \cdots + a_{i,n}p_n \\ &= 0 + b_i = b_i \end{aligned}$$

satisfies (??).

□

As a summary of this section, the title

$$\text{Solution} = \text{Particular} + \text{Homogeneous}$$

and if we can find any particular solution and add it to the homogeneous solution, then we have the full solution.

		number of solutions of the associated homogeneous system	
		one	infinitely many
particular solution exists?	yes	unique solution	infinitely many solutions
	no	no solutions	no solutions

### Solving homogeneous Linear Systems

A homogeneous linear system can be solved in a reasonably efficient manner. Consider the system:

$$\begin{aligned}x_1 - 2x_2 + x_3 &= 0 \\3x_1 + 0x_2 - 2x_3 &= 0 \\0x_1 + 6x_2 - 5x_3 &= 0\end{aligned}$$

and if we write down the augmented coefficient matrix we get:

$$\left[ \begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 3 & 0 & -2 & 0 \\ 0 & 6 & -5 & 0 \end{array} \right]$$

If we perform row operations on this matrix, then the 4th column (right hand side of the linear system), will remain zero, so instead of including this vector, we'll perform row operations only on the first three columns:

$$\begin{aligned} & \left[ \begin{array}{ccc} 1 & -2 & 1 \\ 3 & 0 & -2 \\ 0 & 6 & -5 \end{array} \right] \\ -3R_1 + R_2 \rightarrow R_2 & \left[ \begin{array}{ccc} 1 & -2 & 1 \\ 0 & 6 & -5 \\ 0 & 6 & -5 \end{array} \right] \\ -R_2 - R_3 \rightarrow R_3 & \left[ \begin{array}{ccc} 1 & -2 & 1 \\ 0 & 6 & -5 \\ 0 & 0 & 0 \end{array} \right] \end{aligned}$$

which is now in echelon form. To find the solution, we will write down the top two equations, recalling that the right hand side is 0.

$$\begin{aligned}x_1 - 2x_2 + x_3 &= 0 \\6x_2 - 5x_3 &= 0\end{aligned}$$

solving the second equation for  $x_2$  results in

$$x_2 = \frac{5}{6}x_3$$

and then substitute this into the first equation:

$$\begin{aligned}x_1 - 2\left(\frac{5}{6}x_3\right) + x_3 &= 0 \\x_1 - \frac{2}{3}x_3 &= 0 \\x_1 &= \frac{2}{3}x_3\end{aligned}$$

and writing the solution in vector form is

$$\left\{ \begin{bmatrix} 2/3 \\ 5/6 \\ 1 \end{bmatrix} x_3 \mid x_3 \in \mathbb{R} \right\}$$

### 1.1.7 Singular and Nonsingular Matrices

A square matrix  $A$  is **nonsingular** if it is a matrix of coefficients of a homogeneous linear system with the unique solution  $\vec{0}$ . Otherwise it is **singular**, that is the associated homogeneous system has a solution set with an infinite number of points.

The associated linear system in Example ?? has an infinite number of solutions, therefore the matrix

$$\begin{bmatrix} 1 & 3 & -1 \\ 0 & 1 & 3 \\ 1 & 4 & 2 \end{bmatrix}$$

is singular.

The linear system

$$\begin{aligned}4x_1 - x_3 &= 0, \\x_1 + 3x_2 + 2x_3 &= 3, \\3x_2 + 5x_3 &= 14.\end{aligned}$$

was shown above to have a unique solution, thus the associated homogeneous system also has a unique solution which implies that the matrix

$$\begin{bmatrix} 4 & 0 & -1 \\ 1 & 3 & 2 \\ 0 & 3 & 5 \end{bmatrix}$$



is nonsingular.

Is the matrix

$$\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$$

singular or nonsingular?

We can consider the elements of the matrix to be the coefficients of a homogeneous linear system and find its solution. Thus we reduce

$$\begin{array}{c} \begin{bmatrix} 3 & 2 & | & 0 \\ 1 & 4 & | & 0 \end{bmatrix} \\ -R_1 + 3R_2 \rightarrow R_2 \end{array} \quad \begin{bmatrix} 3 & 2 & | & 0 \\ 0 & 10 & | & 0 \end{bmatrix}$$

which is now in echelon form. The last equation  $10y = 0$  implies  $y = 0$  and the top row can be written as  $3x + 2y = 0$  or  $3x = 0$  or  $x = 0$ , thus this has the unique solution  $(0, 0)$ . This implies that the original matrix is nonsingular.

## 1.2 Linear Geometry of $n$ -space

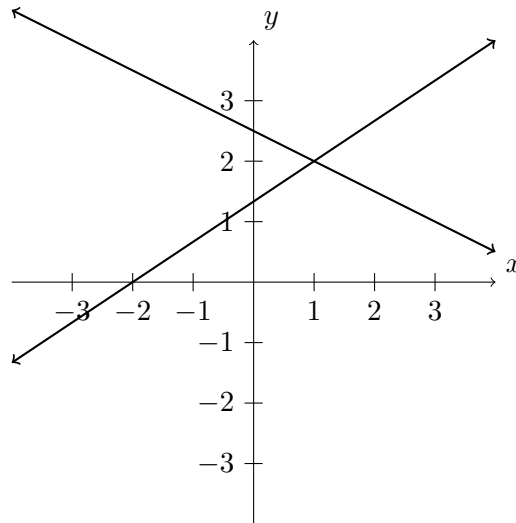
If we consider only linear systems with two variables, then each equation is just a line, which we can graph on a set of axes. We are going to examine the geometry of linear systems of two variables.

First, we are going to look at three linear systems that each have different types of solutions. We can see the solutions by looking at the graphs.

- The graph of each line in the linear system

$$\begin{cases} x + 2y = 5, \\ 2x - 3y = -4, \end{cases} \quad (1.7)$$

is

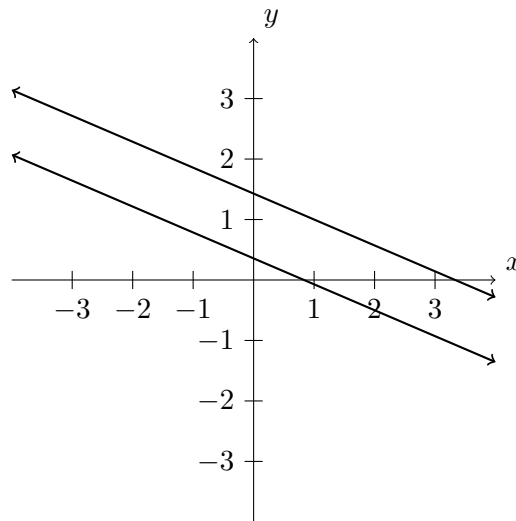


The solution of the system is the point at which they cross. In this case it looks like the point  $(1, 2)$  or  $x = 1$  and  $y = 2$ . This is an example where there is only one solution (or a **unique solution**).

- If we graph the lines in the linear system

$$\begin{cases} 3x + 7y = 10, \\ 6x + 14y = 5, \end{cases} \quad (1.8)$$

we get

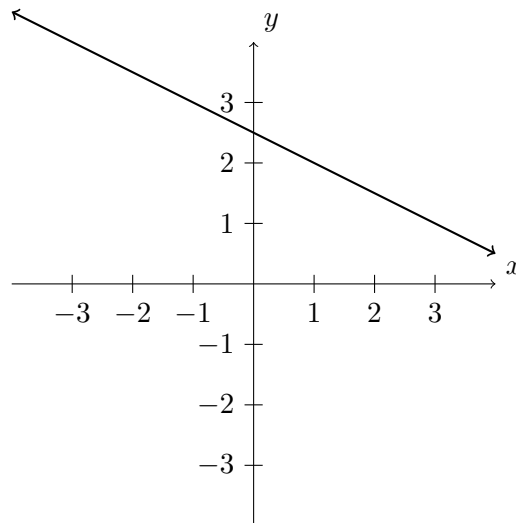


As you can see, it doesn't appear that the lines cross anywhere. In fact, they don't because the lines are parallel. This is an example of a linear system with **no solution**.

- The linear system

$$\begin{cases} x + 2y = 5, \\ 2x + 4y = 10, \end{cases} \quad (1.9)$$

has the following graph for each line



It appears that there is only one line. This is because both lines have the same graph. Each point on the line is a solution to the linear system and since there are an infinite number of such points, this is an example of a linear system with **infinite number of solutions**.

You can see if you have two lines, each with two variables, the example in (??) is what happens if the two slopes are different. In this case, there is one (or a **unique**) solution.

In the other two cases as in equations (??) (??), both sets of lines have the same slope. In the case of (??) the lines have different  $y$ -intercepts, and therefore the lines are parallel and thus there is **no solution** to the linear system. In the case of (??), both the slope and  $y$ -intercepts are equal, so the lines are equal and thus any point on the line is a solution, and this system has an **infinite number of solutions**.

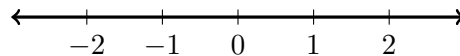
### 1.2.1 Summaries of Linear Systems

We will see more complicated linear systems (including systems with more than two equations as well as more than two variables) shortly. Even with more complicated systems, there are only three possibilities for solutions linear systems:

- One solution. Mathematically, we say this is a **unique** solution.
- No solution. We often say that the system is **inconsistent**.
- An infinite number of solutions. We often say that such a system is **dependent**.

### 1.2.2 Vectors in $\mathbb{R}^n$

Let's start simple and with something we know. The notation  $\mathbb{R}$  is the set of all real numbers. It is often drawn as

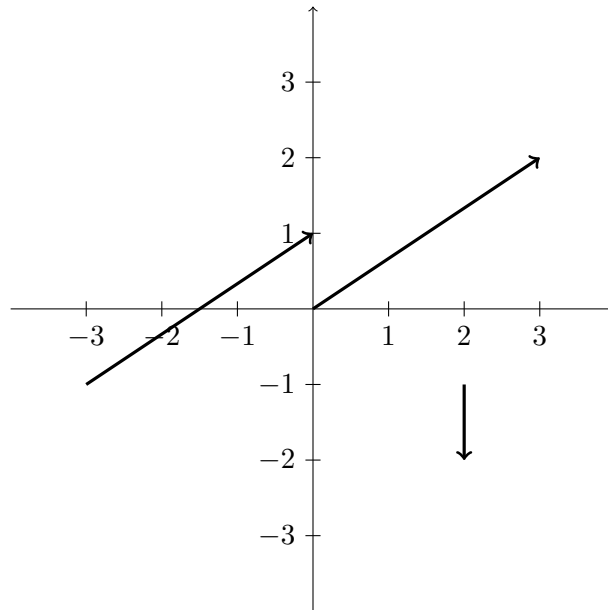


and we use the number line to help with a number of mathematical concepts. For example,  $3 + 2$  means start at the point 3 and go 2 spaces to the right (positive direction). The point of this is that there is direction and magnitude on  $\mathbb{R}$ .

[Alternative Definition of a Vector] A **vector** is a mathematical object with a magnitude and a direction.

In  $\mathbb{R}$ , for two points, the vector between them is the distance between the point with the direction (left as negative and right as positive).

In  $\mathbb{R}^2$ , vectors are often drawn as



with the magnitude being the length of the arrow and the direction as the angle (typically from the positive horizontal axis).

The left 2 vectors above are identical above because they have the same magnitude and direction. That is, the origin does not matter.

To connect these concepts with that of vectors that we saw in section ?? above, we saw that the two different vectors in the plot above are

$$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

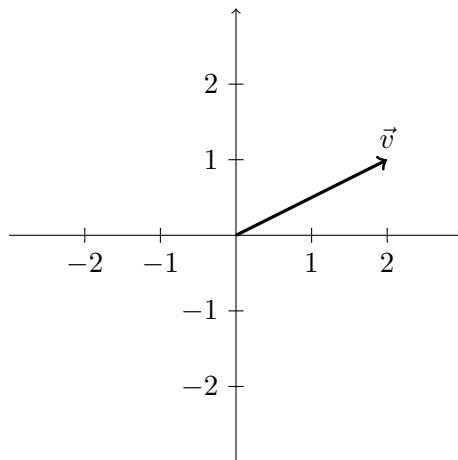
$$\begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

and as we will see that this formulation is much easier than writing down a vector as a length and an angle.

### Vectors, Free Vectors and Points

As noted above, vectors by definition have only a direction and magnitude which is why two of the vectors in the figure above are equal. Often, to clarify that the source point of the vector does not matter, the term *free vector* is used. However, generally, unless indicated a vector is a free vector.

One case where a vector has a fixed point is when the source point is the origin. For example:



The vector starts at the origin and the end point is at the point  $(2, 1)$ . This vector is

$$\vec{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

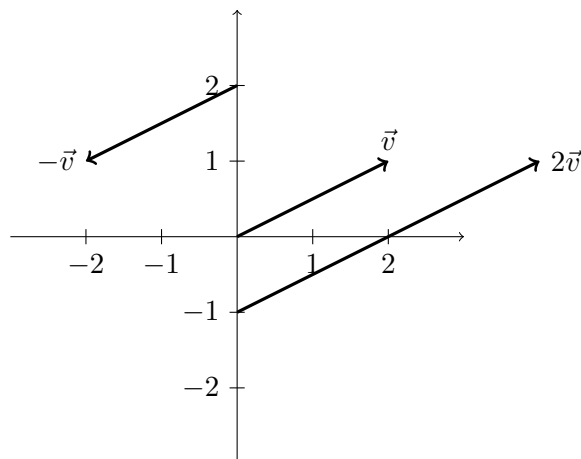
and the vector and the end point are identical. This can be very handy in many cases as we will see.

### Scalar Multiplication of Vectors

In the previous chapter, we saw the scalar multiplication of a vector. In  $\mathbb{R}^2$ , this means

$$r \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} rx_1 \\ rx_2 \end{bmatrix}$$

in that each component of the vector is multiplied by the scalar  $r$ . Geometrically, multiplication by the scalar  $r$ , *scales* the length of the vector by a factor of  $r$  and flips its direction if  $r < 0$ .

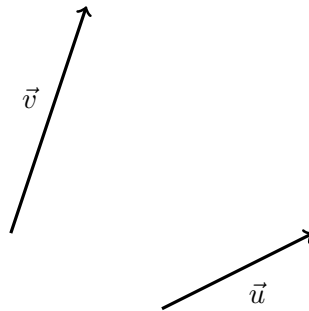


**Vector Addition**

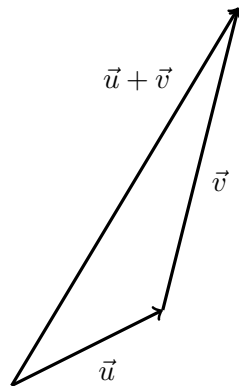
As we saw in the previous chapter, vector addition in  $\mathbb{R}^2$  is

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$$

Consider two vectors  $\vec{u}$  and  $\vec{v}$  in the plane:

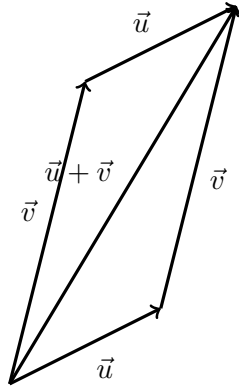


The sum of these vectors can be represented by taking the  $\vec{u}$  vector, then drawing the  $\vec{v}$  at the tail end of the  $\vec{u}$  vector. The resulting vector



starts at the beginning of  $\vec{u}$  and ends at the end of  $\vec{v}$  as seen above.

Another way to think about this is to use  $\vec{u}$  and  $\vec{v}$  as the sides of the parallelogram. The vector  $\vec{u} + \vec{v}$  is diagonal from the starting point of both  $\vec{u}$  and  $\vec{v}$  extending to the ending point of both.



### Geometry of Addition and Scalar Multiplication in $\mathbb{R}^n$

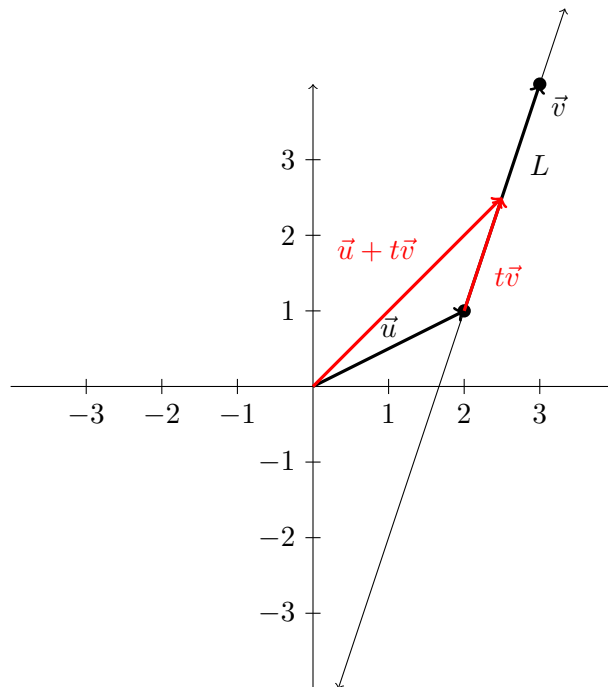
As we saw in the previous section, we know how to add and scalar multiply vectors in  $\mathbb{R}^n$ . The geometry of these operations are similar to that in  $\mathbb{R}^2$ . For example, in  $\mathbb{R}^3$ , a vector connects two points in 3-dimensional space. Scalar multiplication results in scaling that vector by a factor of  $r$ . Addition works the same way: make the ending point of the first vector, the starting point of the second vector. The result is the vector from the starting point of the first vector to the ending point of the 2nd vector.

And this extends to any dimension,  $\mathbb{R}^n$ . Although this is difficult to visualize, it still works the same way. Typically there is no need to draw any vectors in dimensions above 3.

### Lines in Vector form in $\mathbb{R}^2$ and $\mathbb{R}^3$

First, let's look at a line in  $\mathbb{R}^2$  and for example, passes through  $(2, 1)$  and  $(3, 4)$  and denote the line  $L$ . This would look like:





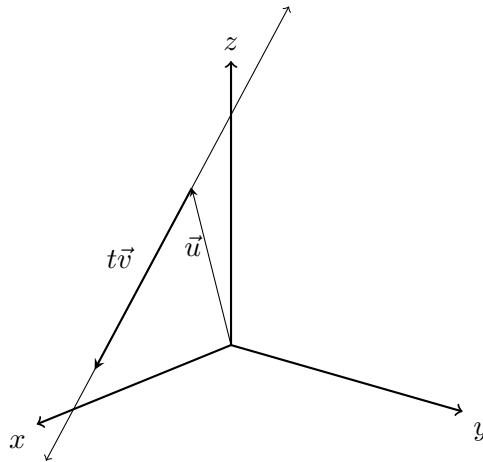
Let's make the vector  $\vec{v}$  the vector between the points  $(2,1)$  and  $(3,4)$  as shown on the figure above. Recall that a vector and a point is synonymous if the vector starts at the origin. Call  $\vec{u}$  the vector from  $(0,0)$  to  $(2,1)$ .

Next, any point on the line can be written as a vector by addition of  $\vec{u}$  and a scale of  $\vec{v}$  or  $\vec{u} + t\vec{v}$ . Thus the line between these two points can be written as

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 3 \end{bmatrix} \mid t \in \mathbb{R} \right\}$$

This notion extend easily to  $\mathbb{R}^3$ . The set of points

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \mid t \in \mathbb{R} \right\}$$



### Planes in $\mathbb{R}^3$

A plane in  $\mathbb{R}^3$  can also be written using vectors although perhaps harder to visualize.

A plane in  $\mathbb{R}^3$  is the set of points

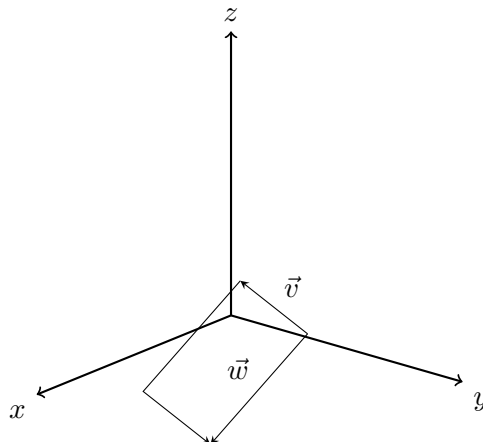
$$\{\vec{p} + \vec{u}t + \vec{v}s \mid t, s \in \mathbb{R}\}$$

for nonzero vectors  $\vec{u}$  and  $\vec{v}$  and  $\vec{p}, \vec{u}$  and  $\vec{v}$  are vectors in  $\mathbb{R}^3$ .

Consider the following example in which a parallelogram is drawn in the plane. The point  $(2, 3, 1)$  is in one corner of the parallelogram and the two sides are

$$\vec{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$$

The parallelogram in  $\mathbb{R}^3$  would look like:



The set of all points on the plane can then be written as

$$\left\{ \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} t + \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} s \mid t, s \in \mathbb{R} \right\}$$

Extending this idea to linear surfaces in  $\mathbb{R}^n$  is a natural generalization of planes. A  **$k$ -dimensional linear surface** in  $\mathbb{R}^n$  is the set:

$$\{\vec{p} + t_1\vec{v}_1 + t_2\vec{v}_2 + \cdots + t_k\vec{v}_k \mid t_1, t_2, \dots, t_k \in \mathbb{R}\}$$

If  $k = n - 1$ , then the surface is called a **hyperplane**.

### Geometry of Linear Systems

You should have noticed the the  $k$ -dimensional linear surface above has the same form as the solution to the general linear system.

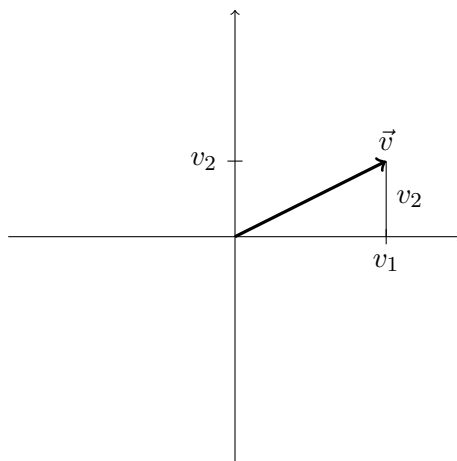
- If the linear system has one free variable, the solution is a line.
- If the linear system has two free variables, the solution is a plane (or hyperplane).

### 1.2.3 Length and Angle Measures

Two of the basic ideas of geometry are the notions of length and angles. These are well-defined in  $\mathbb{R}^2$  and with a relatively easy extension to  $\mathbb{R}^3$  and in this section we generalize to  $\mathbb{R}^n$ . We'll start with the notion of distance as the length of a vector. Consider first a vector

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

in the plane ( $\mathbb{R}^2$ ).



and using plane geometry, the length of  $\vec{v}$ , denoted  $\|\vec{v}\|$  is the hypotenuse of the triangle or

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2}$$

and if  $\vec{v}$  is in  $\mathbb{R}^3$ , the length would include a square of the third component inside the square root. Thus, we defined the length of any vector in  $\mathbb{R}^n$  to be the following.

The **length** of a vector  $\vec{v} \in \mathbb{R}^n$  is given by

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots v_n^2}$$

which fits our expectations in  $\mathbb{R}^2$  and  $\mathbb{R}^3$ .

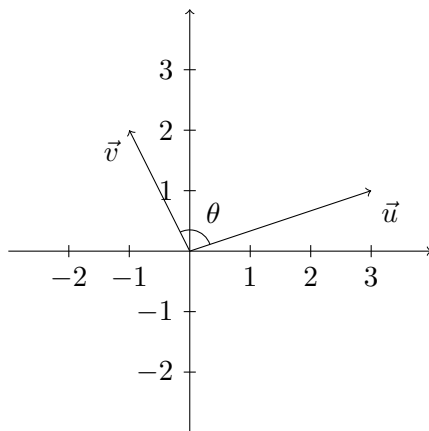
Find the length of the vector:

$$\vec{v} = \begin{bmatrix} 3 \\ 1 \\ 0 \\ -5 \end{bmatrix}$$

$$\|\vec{v}\| = \sqrt{9 + 1 + 0 + 25} = \sqrt{35}$$

### Angles of vectors in $\mathbb{R}^2$

Consider first 2 vectors in  $\mathbb{R}^2$ . The angle between them would be the angle (measured in the counterclockwise direction) between the vectors if the starting point is anchored at the same place. For example:



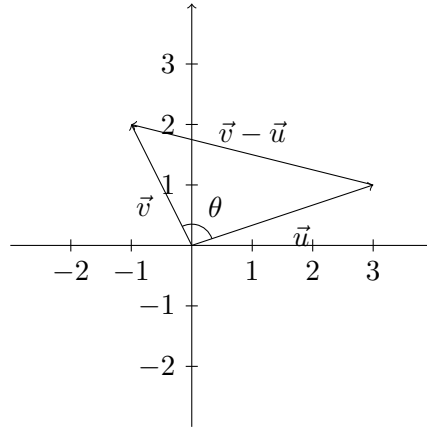
where

$$\vec{u} = \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

$$\vec{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

You can find the angle using knowledge of geometry. In this case, you can make a triangle by connecting the end points of each vector. Note that the third leg of the triangle can be written as

$$\vec{v} - \vec{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$$



and the lengths of the three sides are:

$$\|\vec{u}\| = \sqrt{10}, \quad \|\vec{v}\| = \sqrt{5}, \quad \|\vec{v} - \vec{u}\| = \sqrt{17}$$

and then using the law of cosines:

$$\begin{aligned} \|\vec{v} - \vec{u}\|^2 &= \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\|\cos\theta \\ 17 &= 10 + 5 - 2\sqrt{10}\sqrt{5}\cos\theta \\ \frac{2}{-2\sqrt{10}\sqrt{5}} &= \cos\theta \end{aligned}$$

which in this case  $\theta \approx 98.13^\circ$ .

Note: since the range of  $\cos^{-1}\theta$  is  $[0, \pi]$  (or  $[0, 180^\circ]$ ), the angle will always be the angle between the vectors with this range. If you need the proper angle, you may need to subtract the result from  $360^\circ$ .

### Angles of vectors in $\mathbb{R}^n$

Understanding the above section allows us to extend the notion of vectors in  $n$  dimensions, with the key being the law of cosines:

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\|\cos\theta$$

and if we expand this out for

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Then

$$\begin{aligned} \|\vec{u} - \vec{v}\|^2 &= (u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2, \\ \|\vec{u}\|^2 &= u_1^2 + u_2^2 + \cdots + u_n^2, \\ \|\vec{v}\|^2 &= v_1^2 + v_2^2 + \cdots + v_n^2, \end{aligned}$$

Expanding the top equation and subtracting the two below:

$$\|\vec{u} - \vec{v}\|^2 - \|\vec{u}\|^2 - \|\vec{v}\|^2 = -u_1v_1 - u_2v_2 - \cdots - u_nv_n$$

The negative of this expression appears often throughout linear algebra and is called the **dot product**

The **dot product** (or **inner product**) of the vectors  $\vec{u}$  and  $\vec{v}$  is defined as

$$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

Note: the dot product between two vectors results in a number (a scalar). Also, the dot product is only defined between two vectors of the same length. Also, for any vector  $\vec{u}$ , there is a nice relationship between the length and the dot product:

$$\vec{u} \cdot \vec{u} = \|\vec{u}\|^2$$

Again, returning to law of cosines and solving for  $\cos \theta$ :

$$\begin{aligned} \cos \theta &= \frac{\|\vec{u} - \vec{v}\|^2 - \|\vec{u}\|^2 - \|\vec{v}\|^2}{-2\|\vec{u}\|\|\vec{v}\|} \\ &= \frac{-\vec{u} \cdot \vec{v}}{-2\|\vec{u}\|\|\vec{v}\|} \\ &= \frac{\vec{u} \cdot \vec{v}}{2\|\vec{u}\|\|\vec{v}\|} \end{aligned}$$

### Properties of The Dot Product

As we will see, the dot product is an extremely important concept. Before going on, we show the properties of the dot product.

**(Commutative)**  $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$ ,

**(Distributive)**  $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$ .

**(Associative)**  $\vec{u} \cdot (r\vec{v}) = (r\vec{u}) \cdot \vec{v} = r(\vec{u} \cdot \vec{v})$ .

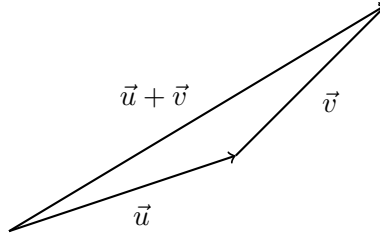
These properties can be shown using the Definition above.

[The Triangle Inequality] For any  $\vec{u}$  and  $\vec{v}$  both in  $\mathbb{R}^n$ , then

$$\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$$

Equality is only reached if one of the vectors is a nonnegative scalar multiple of the other.

This can be visualized by considering the plane in which  $\vec{u}$  and  $\vec{v}$  lie (and note that regardless of the value of  $n$ , they will lie in a plane or planar subset of  $\mathbb{R}^n$ .)



The vector  $\vec{u} + \vec{v}$  is one side of the triangle and we know that any one side must always be no larger than the sum of the other two.

*Proof.* We'll assume that the triangle inequality is true and show this results in another true statement.

$$\begin{aligned} \|\vec{u} + \vec{v}\|^2 &\leq (\|\vec{u}\| + \|\vec{v}\|)^2 \\ (\vec{u} + \vec{v}) \cdot (\vec{u} + \vec{v}) &\leq \|\vec{u}\|^2 + 2\|\vec{u}\|\|\vec{v}\| + \|\vec{v}\|^2 \\ \vec{u} \cdot \vec{u} + 2\vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{v} &\leq \vec{u} \cdot \vec{u} + 2\|\vec{u}\|\|\vec{v}\| + \vec{v} \cdot \vec{v} \\ 2\vec{u} \cdot \vec{v} &\leq 2\|\vec{u}\|\|\vec{v}\| \end{aligned} \tag{1.10}$$

multiply both sides by  $\|\vec{u}\|\|\vec{v}\|$

$$\begin{aligned} 2(\|\vec{v}\|\vec{u}) \cdot (\|\vec{u}\|\vec{v}) &\leq 2\|\vec{u}\|^2\|\vec{v}\|^2 \\ 0 &\leq 2\|\vec{u}\|^2\|\vec{v}\|^2 - 2(\|\vec{v}\|\vec{u}) \cdot (\|\vec{u}\|\vec{v}) \\ 0 &\leq \|\vec{u}\|^2\|\vec{v}\|^2 - 2(\|\vec{v}\|\vec{u}) \cdot (\|\vec{u}\|\vec{v}) + \|\vec{u}\|^2\|\vec{v}\|^2 \\ 0 &\leq \|\vec{u}\|^2(\vec{v} \cdot \vec{v}) - 2(\|\vec{v}\|\vec{u}) \cdot (\|\vec{u}\|\vec{v}) + \|\vec{v}\|^2(\vec{u} \cdot \vec{u}) \\ 0 &\leq (\|\vec{u}\|\vec{v}) \cdot (\|\vec{u}\|\vec{v}) - 2(\|\vec{v}\|\vec{u}) \cdot (\|\vec{u}\|\vec{v}) + (\|\vec{v}\|\vec{u}) \cdot (\|\vec{u}\|\vec{v}) \\ 0 &\leq (\|\vec{u}\|\vec{v} - \|\vec{v}\|\vec{u}) \cdot (\|\vec{u}\|\vec{v} - \|\vec{v}\|\vec{u}) \\ 0 &\leq \|(\|\vec{u}\|\vec{v} - \|\vec{v}\|\vec{u})\|^2 \end{aligned}$$

Since this statement is true and all steps are reversible, the triangle inequality is true.

Now since the left hand side is a length, by definition it is greater than or equal to zero. To show equality, assume that  $||\vec{v}|| \neq 0$ ,

$$||\vec{u}||\vec{v} - ||\vec{v}||\vec{u} = 0$$

or

$$\vec{u} = \frac{||\vec{u}||}{||\vec{v}||}\vec{v}$$

therefore  $\vec{u}$  is a scalar multiple of  $\vec{v}$ . □

[Cauchy-Swartz Inequality]

$$|\vec{u} \cdot \vec{v}| \leq ||\vec{u}|| ||\vec{v}||$$

with equality if and only if  $\vec{u}$  and  $\vec{v}$  are scalar multiples of each other.

*Proof.* A consequence of the Triangle Inequality is (??), thus the Cauchy-Swartz Inequality holds if  $\vec{u} \cdot \vec{v} > 0$ . Assume that  $\vec{u} \cdot \vec{v} < 0$ . Then

$$\vec{u} \cdot \vec{v} = -(\vec{u} \cdot \vec{v}) = (-\vec{u} \cdot \vec{v}) \leq ||-\vec{u}|| ||\vec{v}||$$

and since  $||-\vec{u}|| = ||\vec{u}||$

$$= ||\vec{u}|| ||\vec{v}||$$

The equality condition is not shown. □

We now use the dot product to define the angle between two vectors in  $\mathbb{R}^n$ .

The **angle between vectors**  $\vec{u}$  and  $\vec{v}$  is the value of  $\theta$  that satisfies:

$$\cos \theta = \frac{\vec{u} \cdot \vec{v}}{||\vec{u}|| ||\vec{v}||}$$

So now with any two vectors, the above expression is a way to define the angle between them.

Find the angle between

$$\vec{u} = \begin{bmatrix} 2 \\ -4 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{v} = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 5 \end{bmatrix}$$

$$||\vec{u}|| = \sqrt{22},$$

$$||\vec{v}|| = \sqrt{38}$$



$$\vec{u} \cdot \vec{v} = 6 - 8 + 0 + 5 = 3$$

and therefore the angle can be found by

$$\cos \theta = \frac{3}{\sqrt{22}\sqrt{38}} \approx 0.1037571696$$

and  $\theta \approx 84.04^\circ$ . And although it's difficult to visualize these vectors, we can imagine the angle between them.

One of the most important angles in geometry is  $\theta = 90^\circ$ , which occurs in right triangles and perpendicular lines. In terms of vectors, we use the dot product to define this.

Two vectors  $\vec{u}$  and  $\vec{v}$  are **perpendicular** or **orthogonal** if their dot product is 0.

### 1.3 Reduced Row Echelon Form

As we have seen in section ??, there are numerous ways to reduce a matrix (or a linear system) to echelon form using Gauss's method. For example:

$$\begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}$$

Can be reduced to any of the following:

$$\begin{bmatrix} 3 & 4 \\ 0 & -2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$$

What row operations get us to these matrices?

A question arises with this differences in the echelon forms of the matrices. Are the solution sets equivalent? The answer is yes. Are the number, free variables the same? The way we will answer this last question is by putting the matrix in **reduced row echelon form**.

#### 1.3.1 Gauss-Jordan Reduction

We solved the linear system from Example ??.

$$\begin{aligned} 4x_1 - x_3 &= 0, \\ x_1 + 3x_2 + 2x_3 &= 3, \\ 3x_2 + 5x_3 &= 14. \end{aligned}$$

by putting the array in echelon form using elementary row operations,

$$\begin{aligned} & \begin{bmatrix} 4 & 0 & -1 & | & 0 \\ 1 & 3 & 2 & | & 3 \\ 0 & 3 & 5 & | & 14 \end{bmatrix} \\ -R_1 + 4R_2 \rightarrow R_2, & \begin{bmatrix} 4 & 0 & -1 & | & 0 \\ 0 & 12 & 9 & | & 12 \\ 0 & 3 & 5 & | & 14 \end{bmatrix} \\ -R_2 + 4R_3 \rightarrow R_3, & \begin{bmatrix} 4 & 0 & -1 & | & 0 \\ 0 & 12 & 9 & | & 12 \\ 0 & 0 & 11 & | & 44 \end{bmatrix} \end{aligned}$$

then using back-substitution.

Note that there is still work to be done to get the solution using back-substitution. Instead, we can continue to use Gauss' method to further reduce the matrix into a nicer form:

$$\begin{array}{ll}
\frac{1}{11}R_3 \rightarrow R_3, & \left[ \begin{array}{ccc|c} 4 & 0 & -1 & 0 \\ 0 & 12 & 9 & 12 \\ 0 & 0 & 1 & 4 \end{array} \right] \\
-9R_3 + R_2 \rightarrow R_2, & \left[ \begin{array}{ccc|c} 4 & 0 & 0 & 4 \\ 0 & 12 & 0 & -24 \\ 0 & 0 & 1 & 4 \end{array} \right] \\
R_3 + R_1 \rightarrow R_1, & \\
\frac{1}{12}R_2 \rightarrow R_2, & \left[ \begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 4 \end{array} \right] \\
\frac{1}{4}R_1 \rightarrow R_1, &
\end{array}$$

Now this is a nicer form, because if the linear system is written from the matrix, then:

$$\begin{aligned}
x &= 1, \\
y &= -2, \\
z &= 4.
\end{aligned}$$

or in vector form

$$\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}$$

which is the last column of the matrix.

A matrix is in **reduced row-echelon form** if in addition to being in echelon form, that each leading entry is a 1, and the leading entry is the only non-zero entry in the column.

The following matrices are in reduced row echelon form.

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 0 & 0 & 5 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & 2 & 9 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Explain why the following are not in reduced row echelon form:

$$\begin{bmatrix} 2 & 0 & 4 \\ 0 & 1 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & -1 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 3 & 3 \\ 1 & 0 & 2 & -1 \end{bmatrix}$$

- The first matrix is not in row reduced echelon form because the leading term in row 1 needs to be a 1, not a 2.
- The second matrix is not in row reduced echelon form because the column with the leading term in row 2 (which is column 2) should be 0 except the 1 in the second row.

- The third matrix is not in row reduced echelon form because the matrix is not in echelon form.

The reduction of a matrix to reduced row echelon form using the three standard row operations is called *Gauss-Jordan reduction* or the *Gauss-Jordan method*.

As seen above, reduced row echelon form makes linear systems with a unique solution very easy to write down—the solution is in the last row. We will also see that this form makes solutions with many solutions easier to write down as well.

Put the following linear system:

$$\begin{aligned} x_1 + 3x_3 - 9x_4 + 11x_5 &= 14, \\ 2x_3 + 4x_5 &= 10, \\ 3x_5 &= 27, \end{aligned}$$

into reduced row echelon form.

First, this is already in echelon form, so we only need to do the steps to get to reduced row echelon form:

$$\begin{aligned} & \begin{bmatrix} 1 & 0 & 3 & -9 & 11 & | & 14 \\ 0 & 0 & 2 & 0 & 4 & | & 9 \\ 0 & 0 & 0 & 0 & 3 & | & 27 \end{bmatrix} \\ \frac{1}{3}R_3 \rightarrow R_3, & \begin{bmatrix} 1 & 0 & 3 & -9 & 11 & | & 14 \\ 0 & 0 & 2 & 0 & 4 & | & 9 \\ 0 & 0 & 0 & 0 & 1 & | & 9 \end{bmatrix} \\ \begin{aligned} -4R_3 + R_2 &\rightarrow R_2, \\ -11R_3 + R_1 &\rightarrow R_1, \end{aligned} & \begin{bmatrix} 1 & 0 & 3 & -9 & 0 & | & -85 \\ 0 & 0 & 2 & 0 & 0 & | & -26 \\ 0 & 0 & 0 & 0 & 1 & | & 9 \end{bmatrix} \\ \frac{1}{2}R_2 \rightarrow R_2, & \begin{bmatrix} 1 & 0 & 3 & -9 & 0 & | & -85 \\ 0 & 0 & 1 & 0 & 0 & | & -13 \\ 0 & 0 & 0 & 0 & 1 & | & 9 \end{bmatrix} \\ -3R_2 + R_1 \rightarrow R_1, & \begin{bmatrix} 1 & 0 & 0 & -9 & 0 & | & -46 \\ 0 & 0 & 1 & 0 & 0 & | & -13 \\ 0 & 0 & 0 & 0 & 1 & | & 9 \end{bmatrix} \end{aligned}$$

And now this is in reduced row echelon form. The linear system from this equation can be written by solving for the leading variables:

$$\begin{aligned} x_1 &= -46 + 9x_4, \\ x_3 &= -13 \\ x_5 &= 9 \end{aligned}$$

and then the solution can be written in vector form as

$$\left\{ \begin{bmatrix} -46 \\ 0 \\ -13 \\ 0 \\ 9 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 9 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} x_4 \mid x_2, x_4 \in \mathbb{R}, \right\}$$

which is the same solution as that from that in Example ??.

An alternative way to read off the solution from the reduced row echelon matrix is to add rows of zeros where the free variables would be. In this case in row 1 and 4 and pushing others down.

$$\left[ \begin{array}{cccccc|c} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -9 & 0 & -\frac{89}{2} \\ 0 & 0 & 1 & 0 & 0 & -\frac{27}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 9 \end{array} \right]$$

The particular solution is now the last column. The solution also contains  $x_1$  terms which is the first column with a 1 instead of a 0 in the first row and also the fourth column with a 1 instead of a 0 in the fourth row times  $x_4$ .

Write down the solution set of the following linear system given in the reduced row echelon form of the matrix:

$$\left[ \begin{array}{cccccc|c} 1 & 0 & 3 & 0 & 0 & 5 & 11 \\ 0 & 1 & -4 & 0 & 0 & 9 & -4 \\ 0 & 0 & 0 & 1 & 0 & 3 & 5 \\ 0 & 0 & 0 & 0 & 1 & 0 & -2 \end{array} \right]$$

If we add a row between the 2nd and 3rd rows as well as after the fourth row, the matrix is:

$$\left[ \begin{array}{cccccc|c} 1 & 0 & 3 & 0 & 0 & 5 & 11 \\ 0 & 1 & -4 & 0 & 0 & 9 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 3 & 5 \\ 0 & 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

The the solution set is

$$\left\{ \begin{bmatrix} 11 \\ -4 \\ 0 \\ 5 \\ -2 \\ 0 \end{bmatrix} + \begin{bmatrix} -3 \\ 4 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} x_3 + \begin{bmatrix} -5 \\ -9 \\ 0 \\ -3 \\ 0 \\ 1 \end{bmatrix} x_6 \mid x_3, x_6 \in \mathbb{R} \right\}$$

Recalling that you need to replace the 0 in the 3rd row of the  $x_3$  vector with a 1 and replace the 0 in the 6th row of the  $x_6$  vector with a 1.

### Matrix Reduction is Matrix Equivalence

We have used the term *matrix reduction* in converting a matrix to either echelon or reduced row echelon form.

Two matrices that are irreducible by the elementary row operations are **row equivalent**.

#### 1.3.2 Row Equivalence

A linear combination of linear combinations is a linear combination.

*Proof.* Consider the  $m$  linear combinations  $c_{1,1}x_1 + \cdots + c_{1,n}x_n$  through  $c_{m,1}x_1 + \cdots + c_{m,n}x_n$ . A linear combination of these is

$$\begin{aligned} & d_1(c_{1,1}x_1 + \cdots + c_{1,n}x_n) + \cdots + d_m(c_{m,1}x_1 + \cdots + c_{m,n}x_n) \\ &= d_1c_{1,1}x_1 + \cdots + d_1c_{1,n}x_n + \cdots + d_mc_{m,1}x_1 + \cdots + d_mc_{m,n}x_n \\ &= (d_1c_{1,1} + \cdots + d_mc_{m,1})x_1 + \cdots + (d_1c_{1,n} + \cdots + d_mc_{m,n})x_n \end{aligned}$$

□

We will use the same convention as in the text. That is for a matrix named with a capital roman letter, the  $i$ th row will be denoted with the equivalent lower-case greek with subscript  $i$ , that is

$$A = \begin{bmatrix} \cdots & \alpha_1 & \cdots \\ \cdots & \alpha_2 & \cdots \\ & \vdots & \\ \cdots & \alpha_n & \cdots \end{bmatrix} \qquad B = \begin{bmatrix} \cdots & \beta_1 & \cdots \\ \cdots & \beta_2 & \cdots \\ & \vdots & \\ \cdots & \beta_n & \cdots \end{bmatrix}$$

When one matrix is transformed to another matrix, each row of the second is a linear combination of the first.

*Proof.* The base step, we prove that zero row operations that transforms  $A$  to  $B$ , satisfies the corollary. In this case,

$$\vec{\beta}_i = 0 \cdot \alpha_1 + 0\alpha_2 + \cdots + 1 \cdot \alpha_i + \cdots + 0 \cdot \alpha_n \quad \text{for each } i = 1, 2, \dots, n.$$

so each row of  $B$  is a linear combination of  $A$ .

For the inductive step, assume that transforming  $A$  to  $B$  in  $t + 1$  steps. Also assume that the matrix in the step before  $B$  is called  $G$ , or

$$A \rightarrow \cdots \rightarrow G \rightarrow B$$

with each arrow denoting a row operation. Now we proceed showing that each of the 3 elementary row operations are linear combinations. If  $G \rightarrow B$  is given by

(i)  $cR_i \rightarrow R_i$ , then

$$\vec{\beta}_j = \begin{cases} c\vec{\gamma}_j & \text{if } i = j \\ \vec{\gamma}_j & \text{otherwise} \end{cases}$$

each of which is a linear combination of the rows of  $G$ ,

(ii)  $R_i \leftrightarrow R_j$ , for  $i \neq j$ , then

$$\vec{\beta}_k = \begin{cases} \vec{\gamma}_i & \text{if } k = j \\ \vec{\gamma}_j & \text{if } k = i \\ \vec{\gamma}_k & \text{otherwise} \end{cases}$$

again, each of which is a linear combination of the rows of  $G$ .

(iii)  $cR_i + R_j \rightarrow R_j$  for  $i \neq j$ , then

$$\vec{\beta}_k = \begin{cases} c\vec{\gamma}_k + \vec{\gamma}_j & \text{if } k = i \\ \vec{\gamma}_k & \text{otherwise} \end{cases}$$

again, each of which is a linear combination of the rows of  $G$ .

Now this proves that for the step  $G \rightarrow B$ , each row of  $B$  is a linear combination of rows of  $G$ .

Now, using induction, since the first  $t$  steps transforming  $A$  to  $G$  is a linear combination, the steps from  $A$  to  $B$ , each row of  $B$  is a linear combination of rows of  $A$  because linear combinations of linear combinations are linear combinations.

□

Show that the steps that transform:

$$\begin{array}{rcl}
 & & \begin{bmatrix} 0 & 3 \\ 1 & 2 \end{bmatrix} \\
 R_1 \leftrightarrow R_2 & & \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \\
 \frac{1}{3}R_2 \rightarrow R_2 & & \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \\
 -2R_2 + R_1 \rightarrow R_1 & & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
 \end{array}$$

are each linear combinations of row operations.

Call the 4 matrices,  $A$ ,  $D$ ,  $G$  and  $B$ , then

$$\begin{array}{lll}
 \delta_1 = \alpha_2 & \gamma_1 = \alpha_1 & \beta_1 = -2\gamma_2 + \gamma_1 \\
 \delta_2 = \alpha_1 & \gamma_2 = \frac{1}{3}\alpha_2 & \beta_2 = \gamma_2
 \end{array}$$

### Echelon Form and Linear Combinations of Rows

Let's return to echelon form. Why is this a nice form? Well we can use back substitution to solve for the leading variables. The reason this works is the following lemma

In an echelon form matrix, no nonzero row is a linear combination of the other rows.

In lieu of a proof, let's take a look at an example. Consider the echelon form matrix of a the linear system from Example ??:

$$\left[ \begin{array}{ccccc|c} 0 & 1 & 3 & -9 & 11 & 14 \\ 0 & 0 & 2 & 0 & 4 & 9 \\ 0 & 0 & 0 & 0 & 3 & 27 \end{array} \right]$$

We look and determine if the second row (as an example) is a linear combination of the other two rows. Thus:

$$[0 \ 0 \ 2 \ 0 \ 4 \ 9] = c_1 [0 \ 1 \ 3 \ -9 \ 11 \ 14] + c_2 [0 \ 0 \ 0 \ 0 \ 3 \ 27]$$

Each component then becomes an equation for the linear system:

$$\begin{array}{l}
 0c_1 + 0c_2 = 0 \\
 c_1 + 0c_2 = 0 \\
 3c_1 + 0c_2 = 2 \\
 -9c_1 + 0c_2 = 0 \\
 11c_1 + 3c_2 = 4 \\
 14c_1 + 27c_2 = 9
 \end{array}$$



which can be written as the augmented matrix:

$$\left[ \begin{array}{cc|c} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 3 & 0 & 2 \\ -9 & 0 & 0 \\ 11 & 3 & 4 \\ 14 & 27 & 9 \end{array} \right]$$

if we transform this to row echelon form we get:

$$\left[ \begin{array}{cc|c} 14 & 27 & 9 \\ 0 & -3 & -1 \\ 0 & 0 & 28 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right]$$

and the 3rd row says “no solution”, so that the 2nd row is not a linear combination of the other two rows. One can show this is also true for the other two rows.

The notation of rows not being linear combinations of other rows is called **independence** and will be defined more-explicitly later in the course, however this example should give you a sense of what independence means.

## 1.4 Matrix Operations

As we have seen, a matrix is a group of numbers arranged in a rectangular block. An  $m$  by  $n$  matrix has  $m$  rows and  $n$  columns.

For example,

$$\begin{bmatrix} 1 & 2 & 11 & -1 \\ 3 & -2 & 3 & 4 \end{bmatrix}, \quad (1.11)$$

is a 2 by 4 matrix.

The **size** of a matrix is the number of rows and columns in the matrix. The number of rows is listed first. The size of the example above is 2 by 4.

The numbers in a matrix are called the **entries** or **elements** of the matrix. For example, for the matrix in (1.11), the entry on the first row and third column is 11. Often we will use the notation  $a_{1,3} = 11$ , where the subscript 1 is the row number and 3 is the column number.

### 1.4.1 Matrix, Vector and Scalar Notation

A **scalar** is a fancy term for a number. Mathematicians use this term to distinguish them from matrices and vectors, which are not scalars. Whenever variables are used for scalars, then lower case letters will be used. For example,  $a, n$  and  $x$  are scalar variables.

When we use variable names for matrices, we will use capital letters. For example,  $A, B$  and  $X$  are matrix (or vector) variables.

### 1.4.2 Addition and Subtraction of Matrices

In this section, we learn how to perform some basic operations between matrices. First, we will look at adding and subtracting two matrices and later we will look at multiplying a matrix by a scalar (a number).

In this case, let  $k \in \mathbb{R}$  and  $A$  and  $B$  are  $m \times n$  matrices. Then

- $kA$  is the matrix formed by each entry is  $ka_{i,j}$ . That is multiply each entry by  $k$ .
- $A + B$  is the matrix form by each entry is  $a_{i,j} + b_{i,j}$ . That is to add matrices, we add element by element.
- $A - B$  is the matrix form by each entry is  $a_{i,j} - b_{i,j}$ . That is to add matrices, we subtract element by element.

If  $A$  and  $B$  are not the same size then  $A + B$  and  $A - B$  does not exist.

Note: an alternative definition of matrix subtraction,  $A - B = A + (-B)$ , which is consistent with subtraction of real numbers.

Let

$$A = \begin{bmatrix} 1 & -1 & 3 \\ 2 & 7 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 2 & -7 \\ 11 & 2 & 0 \end{bmatrix}$$

Then the sum  $A + B$  is found by adding the individual elements.

$$A + B = \begin{bmatrix} 1+3 & -2+2 & 3-7 \\ 2+11 & 7+2 & 3+0 \end{bmatrix} = \begin{bmatrix} 4 & 1 & -4 \\ 13 & 9 & 3 \end{bmatrix}$$

similarly we can subtract in the same way

$$A - B = \begin{bmatrix} 1-3 & -1-2 & 3-(-7) \\ 2-11 & 7-2 & 3-0 \end{bmatrix} = \begin{bmatrix} -2 & -3 & 10 \\ -9 & 5 & 3 \end{bmatrix}$$

Find the following result or state that it does not exist.

$$\begin{bmatrix} 1 & 0 \\ 3 & -2 \end{bmatrix} + \begin{bmatrix} 3 & -2 \\ 5 & 0 \\ -3 & 4 \end{bmatrix}$$

Since the matrices are not the same size, this operation is not valid.

### Scalar Multiplication of Vectors and Matrices

Scalar multiplication of vectors and matrices is a simple operation where the result is the scalar times each element of the vector or matrix.

Evaluate

(a)

$$3 \begin{bmatrix} 5 & 5 \\ -4 & 2 \end{bmatrix}.$$

(b)

$$-7 \begin{bmatrix} 3 & -2 & 0 & 5 \end{bmatrix}$$

(a)

$$\begin{bmatrix} 15 & 15 \\ -12 & 6 \end{bmatrix}$$

(b)

$$\begin{bmatrix} -21 & 14 & 0 & -35 \end{bmatrix}$$

### 1.4.3 Transpose

The last basic matrix operation that we will cover here is the transpose of a matrix. As an example, if

$$A = \begin{bmatrix} 3 & -1 & 0 \\ 2 & 7 & 5 \end{bmatrix}$$

then the transpose is the matrix,  $A^T$  given by

$$A^T = \begin{bmatrix} 3 & 2 \\ -1 & 7 \\ 0 & 5 \end{bmatrix}$$

where the row and column of each element is flipped.

The **transpose** of an  $m \times n$  matrix  $A$ , denoted  $A^T$  is  $A$  flipped over the diagonal. In particular, the element in the  $i$ th row and  $j$ th column of  $A^T$  is given by the  $j$ th row and  $i$  column of  $A$ .

### Properties of Transposes

We will see that transposes play a big role in linear algebra.

The following are true:

1.  $(A + B)^T = A^T + B^T$
2.  $(cA)^T = cA^T$
3.  $(A^T)^T = A$ .
4.  $(AB)^T = B^T A^T$

### 1.4.4 Multiplication of Vectors and Matrices

In an perhaps backwards way of thinking, we will start with the product of two matrices. We will then see that the product of matrices and vectors and then vectors and vectors can be seen in the same manner.

Let  $A$  and  $B$  be matrices. The product of  $A$  and  $B$ , or  $C = AB$  can be found in the following manner. Element  $c_{ij}$  is the dot product of the  $i$ th row of  $A$  and the  $j$ th column of  $B$ .

Note: In order for the dot product to be defined, the number columns of  $A$  must equal the number of rows of  $B$ . If not the product is not defined.

Also if the product  $AB$  is defined, the result,  $C$  has the same number of rows of  $A$  and the same number of columns as  $B$ .

Find  $C = AB$  if

$$A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 1 \\ 3 & -2 \\ 7 & 5 \end{bmatrix}$$

First, since the number of rows of  $A$  is 2 and the number of columns of  $B$  is 2, the size of  $C$  is 2.

- The 1st row and 1st column of  $C$  is the 1st row of  $A$  times the 1st column of  $B$ .
- The 1st row and 2nd column of  $C$  is the 1st row of  $A$  times the 2nd column of  $B$ .
- The 2nd row and 1st column of  $C$  is the 2nd row of  $A$  times the 1st column of  $B$ .
- The 2nd row and 2nd column of  $C$  is the 2nd row of  $A$  times the 2nd column of  $B$ .

We now explicitly show the dot products.

$$\begin{aligned}
 C &= \begin{bmatrix} [2 & -1 & 3] \begin{bmatrix} 2 \\ 3 \\ 7 \end{bmatrix} & [2 & -1 & 3] \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix} \\ [1 & 1 & 2] \begin{bmatrix} 2 \\ 3 \\ 7 \end{bmatrix} & [1 & 1 & 2] \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix} \end{bmatrix} \\
 &= \begin{bmatrix} 2(2) + (-1)(3) + 3(7) & (2)(1) + (-1)(-2) + (3)(5) \\ (1)(2) + (1)(3) + (2)(7) & (1)(1) + (1)(-2) + (2)(5) \end{bmatrix} \\
 &= \begin{bmatrix} 22 & 19 \\ 19 & 9 \end{bmatrix}
 \end{aligned}$$

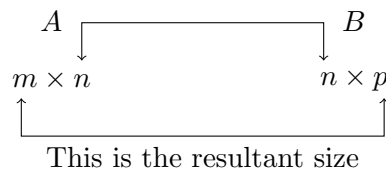
### 1.4.5 Size Restrictions on Matrices

Note that for each element in the resulting matrix, there is a vector-vector product, coming from each row of the first matrix and each column of the second matrix:

The number of columns of the first matrix must equal the number of rows of the second matrix.

The size of the resulting matrix can also be found by the sizes of the input matrices.

If the matrices  $A$  and  $B$  are multiplied then this diagram helps with a valid product as well as the size of the result.



That is if  $A$  is  $m$  by  $n$  and  $B$  is  $n$  by  $p$ , then  $C = AB$  has size  $m$  by  $p$ .

Let  $A$  be a matrix that is 2 by 3,  $B$  be a matrix that is 3 by 3 and  $C$  be a matrix that is 2 by 2. Determine which of the following products are valid and if the product is valid, list its size.

- (a)  $AB$                       (b)  $AC$                       (c)  $BC$   
 (d)  $BA$                       (e)  $CA$                       (f)  $CAB$

- (a) Since  $A$  is 2 by 3 and  $B$  is 3 by 3, the inner numbers are equal and thus multiplication is valid and the result is 2 by 3.  
 (b) Since  $A$  is 2 by 3 and  $C$  is 2 by 2, the inner numbers are not equal, so multiplication is not valid.  
 (c) Since  $B$  is 3 by 3 and  $C$  is 2 by 2, the inner numbers are not equal, so multiplication is not valid.  
 (d) Since  $B$  is 3 by 3 and  $A$  is 2 by 3, the inner numbers are not equal, so multiplication is not valid.  
 (e) Since  $C$  is 2 by 2 and  $A$  is 2 by 3, the inner numbers are equal so multiplication is valid and the result are the outer numbers or 2 by 3.  
 (f) In this case, from (e), the size of  $CA$  is 2 by 3 and  $B$  is 3 by 3, so the inner numbers are equal so multiplication is valid and the result is 2 by 3.

Let

$$A = \begin{bmatrix} 2 & 0 \\ 3 & -1 \\ -3 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & -1 \\ 2 & 2 \end{bmatrix}$$

find  $AB$  if this is a valid operation.

The size of  $A$  is 3 by 2 and the size of  $B$  is 2 by 2, so the inner numbers are equal and thus this operations is valid. The size of the result is 3 by 2.

$$\begin{aligned} AB &= \begin{bmatrix} [2 \ 0] \begin{bmatrix} 3 \\ 2 \end{bmatrix} & [2 \ 0] \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ [3 \ -1] \begin{bmatrix} 3 \\ 2 \end{bmatrix} & [3 \ -1] \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ [-3 \ 5] \begin{bmatrix} 3 \\ 2 \end{bmatrix} & [-3 \ 5] \begin{bmatrix} -1 \\ 2 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} (2)(3) + (0)(2) & (2)(-1) + (0)(2) \\ (3)(3) + (-1)(2) & (3)(-1) + (-1)(2) \\ (-3)(3) + (5)(2) & (-3)(-1) + (5)(2) \end{bmatrix} = \begin{bmatrix} 6 & -2 \\ 7 & -5 \\ 1 & 13 \end{bmatrix} \end{aligned}$$

#### 1.4.6 Matrix-Vector Multiplication

Since a vector can be thought of as a matrix, the matrix-vector product (and vector-matrix products) exist if they are compatible. If  $A$  is a  $m \times n$  matrix and  $\vec{u}$  is a column vector of size  $n$ , then the result is valid.

Let

$$A = \begin{bmatrix} 3 & -2 & 7 \\ 0 & 3 & 5 \\ 2 & -1 & 3 \end{bmatrix} \qquad \vec{u} = \begin{bmatrix} 2 \\ 4 \\ 9 \end{bmatrix}$$

Find  $A\vec{u}$ .

This product is found by taking the dot product of rows of  $A$  and  $\vec{u}$ .

$$A\vec{u} = \begin{bmatrix} 3 & -2 & 7 \\ 0 & 3 & 5 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 3(2) + (-2)(4) + 7(9) \\ 0(2) + (3)(4) + 5(9) \\ 2(2) + (-1)(4) + 3(9) \end{bmatrix} = \begin{bmatrix} 61 \\ 57 \\ 27 \end{bmatrix}$$

### 1.4.7 Reexamination of the Dot Product

Recall that the dot product of two vectors  $\vec{u}$  and  $\vec{v}$  of the same length discussed in Section ?? is the sum of the products of the individual elements.

We can see that the dot product can also be defined as

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

or alternatively using summation notation

$$\vec{u}^T \vec{v} = \sum_{i=1}^n u_i v_i$$

### 1.4.8 Matrix Multiplication and Linear Systems

In this section, we will learn an alternative way to look at a linear system and in the next section, we will solve it. Consider the following linear system.

$$\begin{aligned} 3x + 2y - z &= 10, \\ x - 2y + z &= 11, \\ 2x - 2y + 4z &= 5. \end{aligned}$$

This can be written as:

$$\begin{bmatrix} 3 & 2 & -1 \\ 1 & -2 & 1 \\ 2 & -2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 10 \\ 11 \\ 5 \end{bmatrix}.$$

The first matrix is simply the coefficient matrix without the last column. The column vector on the right hand side is that last column in the coefficient matrix.

To show that this matrix equation is identical to the linear system above, we use matrix multiplication to get:

$$\begin{bmatrix} 3x + 2y - z \\ x - 2y + z \\ 2x - 2y + 4z \end{bmatrix} = \begin{bmatrix} 10 \\ 11 \\ 5 \end{bmatrix}$$

and since two vectors are equal if each of the corresponding elements are equal, this is the same as the linear system above.

If we let

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -2 & 1 \\ 2 & -2 & 4 \end{bmatrix} \quad \vec{u} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \vec{b} = \begin{bmatrix} 10 \\ 11 \\ 5 \end{bmatrix}$$

then the linear system can be written as

$$A\vec{u} = \vec{b}$$

In the next section we will learn a way to solve a linear system in this way using an alternative method to Gauss' method.



## 1.5 The Inverse of a Matrix

At the end of the previous section, we learned how to write a linear system in the form  $A\vec{u} = \vec{b}$ . In this section, we will learn how to find an inverse and then use it to solve the matrix equation  $A\vec{u} = \vec{b}$ .

### 1.5.1 The Identity Matrix

The *identity matrix* of size  $n$  is an  $n$  by  $n$  matrix of all zeros except ones on the diagonal (from upper left to lower right).

The following are identity matrices of sizes 2, 3, and 4

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The identity matrix is a matrix such that  $IA = AI = A$ .

### 1.5.2 The Inverse

Recall that we showed at the end of the previous section that a linear system can be written as:

$$A\vec{u} = \vec{b},$$

where  $A$  is the coefficient matrix of the unknowns,  $\vec{u}$  is a column vector of unknowns, and  $\vec{b}$  is the right hand sides of each equation.

We possibly would like to say that:

$$\vec{u} = \frac{\vec{b}}{A},$$

but how do we divide through by  $A$ ? We can't because there is no matrix division. However, we will see that there is a matrix which will let us multiply through to give the answer.

### 1.5.3 Solving a simple linear equation

To motivate solving a matrix equation, let's look at solving  $ax = b$ , where each is simply a number (or technically a scalar). The solution is  $x = b/a$ , but we can also write it as  $x = a^{-1}b$ , where  $a^{-1} = \frac{1}{a}$  is the reciprocal of  $a$ , which has the property

$$a^{-1}a = aa^{-1} = 1$$

for any value of  $a \neq 0$ . And the important part of this is that the solution  $x = a^{-1}b$  uses only multiplication for the solution.

We wish to do the same for matrices, which means that we'd like to find a matrix  $A^{-1}$  (called  $A$  inverse) such that

$$A^{-1}A = AA^{-1} = I$$

And later we will see how to solve  $A\vec{u} = \vec{b}$ . First, let's determine how to find an inverse matrix and we'll start with the inverse of a 2 by 2 matrix.

### 1.5.4 The Inverse of a 2 by 2 matrix

The inverse of a 2 by 2 matrix is given by

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad (1.12)$$

if  $ad - bc \neq 0$ . If  $ad - bc = 0$ , the matrix does not have an inverse.

We will prove this, by showing that if  $A^{-1}$  is an inverse then  $AA^{-1} = I$  (or  $A^{-1}A = I$ ).

*Proof.*

$$AA^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Although we haven't shown this, the scalar  $1/ad - bc$  can be written out front.

$$= \frac{1}{ad - bc} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Multiply the two matrices like we did in the previous section.

$$= \frac{1}{ad - bc} \begin{bmatrix} ad - bc & -ac + ac \\ da - da & ad - bc \end{bmatrix}$$

simplify and multiply through by  $ad - bc = ad - bc$ .

$$\begin{aligned} &= \begin{bmatrix} \frac{ad-bc}{ad-bc} & 0 \\ 0 & \frac{ad-bc}{ad-bc} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I. \end{aligned}$$

And since we have shown that  $AA^{-1} = I$ , this proves the formula for the inverse in (??).

□

Let's now use the formula in (??) to find the inverse matrix of a 2 by 2 matrix.

Find the inverse of

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$$

Since (?? has the determinant of  $A$  as part of it, first, we will find  $ad-bc = 3(1)-(2)(2) = -1$ . Then apply the formula,

$$A^{-1} = \frac{1}{-1} \begin{bmatrix} 1 & -2 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 2 & -3 \end{bmatrix}$$

The next example shows that not every matrix has an inverse. Since a matrix inverse is similar to the reciprocal of a number, a matrix with no inverse is similar to a number with no reciprocal. And this occurs for the number 0 because  $1/0$  is undefined.

Show that the matrix

$$A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$$

does not have an inverse.

Again we use (??), but because  $ad-bc = 2(2)-4(1) = 0$ , the matrix does not have an inverse.

### 1.5.5 Solving $A\vec{u} = \vec{b}$

At the end of section ??, we learned how to take a linear system and write it as the matrix equation  $A\vec{u} = \vec{b}$ . In this section we will learn how to solve a system written in this form. If we start with the equation,

$$\begin{array}{ll} A\vec{u} = \vec{b} & \text{Multiply this by } A^{-1} \\ A^{-1}A\vec{u} = A^{-1}\vec{b} & \text{Use the fact that } A^{-1}A = I \\ I\vec{u} = A^{-1}\vec{b} & \text{And } I\vec{u} = \vec{u} \text{ for any matrix } \vec{u}. \\ \vec{u} = A^{-1}\vec{b} & \end{array}$$

so if we can find the inverse of a matrix, then solving linear systems becomes matrix multiplication.

If you are trying to solve the matrix equation

$$A\vec{u} = \vec{b}$$

for a square matrix  $A$  and the inverse matrix  $A^{-1}$  exists, then

$$\vec{u} = A^{-1}\vec{b}$$

Solve the system:

$$\begin{array}{l} 3x + 4y = 1, \\ 2x - y = -3, \end{array}$$

by writing the system as  $A\vec{u} = \vec{b}$ , then finding  $A^{-1}$  and finally by writing  $\vec{u} = A^{-1}\vec{b}$ .

First, write down  $A$ ,  $\vec{u}$  and  $\vec{b}$ . (We learned how to do this in section ??.)

$$A = \begin{bmatrix} 3 & 4 \\ 2 & -1 \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}.$$

Then find the inverse of  $A$ , which is a 2 by 2 matrix, so we will use (??). Since  $ad - bc = -3 - 8 = -11$ .

$$A^{-1} = \frac{1}{-11} \begin{bmatrix} -1 & -4 \\ -2 & 3 \end{bmatrix}$$

Then write down the solution:

$$\begin{aligned} \vec{u} &= A^{-1}\vec{b} \\ &= \frac{1}{-11} \begin{bmatrix} -1 & -4 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix} \\ &= \frac{1}{-11} \begin{bmatrix} -1 + 12 \\ -2 - 9 \end{bmatrix} = \frac{1}{-11} \begin{bmatrix} 11 \\ -11 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned}$$

So the solution is  $\vec{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$  or  $x = -1$  and  $y = 1$ .

The next example shows how to perform the same steps for a 3 by 3 matrix, although we don't know how to find an inverse of a 3 by 3 matrix yet.

We will later show that if

$$A = \begin{bmatrix} -4 & -5 & -1 \\ 2 & 1 & -3 \\ 0 & 2 & 5 \end{bmatrix}$$

then

$$A^{-1} = \begin{bmatrix} \frac{11}{2} & \frac{23}{2} & 8 \\ -5 & -10 & -7 \\ 2 & 4 & 3 \end{bmatrix}$$

Use this to find the solution to:

$$\begin{aligned} -4x - 5y - z &= -12, \\ 2x + y - 3z &= 6, \\ 2y + 5z &= 10. \end{aligned}$$

First, the linear system can be written as  $A\vec{u} = \vec{b}$  if

$$A = \begin{bmatrix} -4 & -5 & -1 \\ 2 & 1 & -3 \\ 0 & 2 & 5 \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -12 \\ 6 \\ 10 \end{bmatrix}.$$

Then we write the solution to this systems as

$$\begin{aligned}\vec{u} &= A^{-1}\vec{b}, \\ &= \begin{bmatrix} \frac{11}{2} & \frac{23}{2} & 8 \\ -5 & -10 & -7 \\ 2 & 4 & 3 \end{bmatrix} \begin{bmatrix} -12 \\ 6 \\ 10 \end{bmatrix}, \\ &= \begin{bmatrix} 83 \\ -70 \\ 30 \end{bmatrix},\end{aligned}$$

therefore the solution is  $x = 83, y = -70$ , and  $z = 30$ .

### 1.5.6 the Gauss-Jordan Method for Calculating Inverses

As we have seen, often matrices are larger than 2 by 2 and thus far we have no method to find an inverse of a matrix of this size. In this section, we will see that the Gauss-Jordan method can help us do this.

If we write a matrix  $A$  of the form

$$[A \mid I]$$

where the matrix  $A$  is on the left and the identity matrix is on the right and apply Gauss-Jordan to get the identity matrix on the left, then the matrix will look like,

$$[I \mid A^{-1}]$$

and the matrix on the right half is the inverse matrix,  $A^{-1}$ .

Use the Gauss-Jordan Method to find the inverse of

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$$

First, write the matrix  $[A \mid I]$  and then perform Gaussian Elimination.

$$\left[ \begin{array}{cc|cc} 3 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{array} \right]$$

We desire a 1 in the upper right. Instead of dividing the first row by 3, we will do the following,

$$-R_2 + R_1 \rightarrow R_1 \quad \left[ \begin{array}{cc|cc} 1 & 1 & 1 & -1 \\ 2 & 1 & 0 & 1 \end{array} \right]$$

Then eliminate the lower right element on the left side,

$$-2R_1 + R_2 \rightarrow R_2 \quad \left[ \begin{array}{cc|cc} 1 & 1 & 1 & -1 \\ 0 & -1 & -2 & 3 \end{array} \right]$$

Multiply through by  $-1$  to get a 1 in the lower right of the left side

$$-R_2 \rightarrow R_2 \quad \left[ \begin{array}{cc|cc} 1 & 1 & 1 & -1 \\ 0 & 1 & 2 & -3 \end{array} \right]$$

And finally zero out the 2nd column, 1st row.

$$-R_2 + R_1 \rightarrow R_1 \quad \left[ \begin{array}{cc|cc} 1 & 0 & -1 & 2 \\ 0 & 1 & 2 & -3 \end{array} \right]$$

We stop here because the left half of the matrix is the identity matrix. The right half is now the inverse. Therefore:

$$A^{-1} = \begin{bmatrix} -1 & 2 \\ 2 & -3 \end{bmatrix}$$

This example showed how to find the inverse of a 2 by 2 matrix using the Gauss-Jordan method. We saw the formula in (??), and generally it is easier to compute the inverse using that formula. However, the formula in (??) only works for 2 by 2 matrices. The next example shows how to use the Gauss-Jordan method to find the inverse of a 3 by 3 matrix.

Find the inverse of

$$A = \begin{bmatrix} 3 & 3 & -1 \\ 2 & 1 & -3 \\ 0 & 2 & 5 \end{bmatrix}$$

First, we write  $[A \mid I]$  and then use the Gauss-Jordan Method.

$$\left[ \begin{array}{ccc|ccc} 3 & 3 & -1 & 1 & 0 & 0 \\ 2 & 1 & -3 & 0 & 1 & 0 \\ 0 & 2 & 5 & 0 & 0 & 1 \end{array} \right]$$

Since we want a 1 in the upper right and don't want fractions, try

$$-R_2 + R_1 \rightarrow R_2 \quad \left[ \begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & -1 & 0 \\ 2 & 1 & -3 & 0 & 1 & 0 \\ 0 & 2 & 5 & 0 & 0 & 1 \end{array} \right]$$

And zero out the rest of the first column.

$$-2R_1 + R_2 \rightarrow R_2 \quad \left[ \begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & -1 & 0 \\ 0 & -3 & -7 & -2 & 3 & 0 \\ 0 & 2 & 5 & 0 & 0 & 1 \end{array} \right]$$

Again, to avoid fractions, try

$$2R_3 + R_2 \rightarrow R_2 \quad \left[ \begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & -1 & 0 \\ 0 & 1 & 3 & -2 & 3 & 2 \\ 0 & 2 & 5 & 0 & 0 & 1 \end{array} \right]$$

And zero out the other elements in the 2nd column.

$$\begin{array}{l} -2R_2 + R_3 \rightarrow R_3 \\ -2R_2 + R_1 \rightarrow R_1 \end{array} \quad \left[ \begin{array}{ccc|ccc} 1 & 0 & -4 & 5 & -7 & -4 \\ 0 & 1 & 3 & -2 & 3 & 2 \\ 0 & 0 & -1 & 4 & -6 & -3 \end{array} \right]$$

Next, we need a 1 where the  $-1$  is.

$$-R_3 \rightarrow R_3 \quad \left[ \begin{array}{ccc|ccc} 1 & 0 & -4 & 5 & -7 & -4 \\ 0 & 1 & 3 & -2 & 3 & 2 \\ 0 & 0 & 1 & -4 & 6 & 3 \end{array} \right]$$

And lastly zero out the rest of the 3rd column.

$$\begin{array}{l} -3R_3 + R_2 \rightarrow R_2 \\ 4R_3 + R_1 \rightarrow R_1 \end{array} \quad \left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & -11 & 17 & 8 \\ 0 & 1 & 0 & 10 & -15 & -7 \\ 0 & 0 & 1 & -4 & 6 & 3 \end{array} \right]$$

Since we have found the 3 by 3 identity on the left, the inverse is on the right or

$$A^{-1} = \begin{bmatrix} -11 & 17 & 8 \\ 10 & -15 & -7 \\ -4 & 6 & 3 \end{bmatrix}$$

As we saw with some 2 by 2 matrices (as in example ??), their inverse does not exist. We see how using the Gauss-Jordan method results in this situation for a 3 by 3 example. Show that the matrix

$$\begin{bmatrix} 1 & -5 & 3 \\ -2 & 3 & 7 \\ -1 & -2 & 10 \end{bmatrix}$$

does not have an inverse.

First, write  $[A \mid I]$  and then use the Gauss-Jordan Method to alter the matrix.

$$\begin{array}{l} 2R_1 + R_2 \rightarrow R_2 \\ R_1 + R_3 \rightarrow R_3 \\ \\ -R_2 + R_3 \rightarrow R_3 \end{array} \quad \left[ \begin{array}{ccc|ccc} 1 & -5 & 3 & 1 & 0 & 0 \\ -2 & 3 & 7 & 0 & 1 & 0 \\ -1 & -2 & 10 & 0 & 0 & 1 \end{array} \right]$$

$$\left[ \begin{array}{ccc|ccc} 1 & -5 & 3 & 1 & 0 & 0 \\ 0 & -7 & 13 & 2 & 1 & 0 \\ 0 & -7 & 13 & 1 & 0 & 1 \end{array} \right]$$

$$\left[ \begin{array}{ccc|ccc} 1 & -5 & 3 & 1 & 0 & 0 \\ 0 & -7 & 13 & 2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{array} \right]$$

And we stop since the row of zeros on the bottom of the left half of the matrix indicates that we cannot get an identity matrix on the left half of the matrix. This shows that the matrix is not invertible.

Consider the linear system

$$\begin{aligned}x + 2y &= 3, \\y - z &= 0, \\x + y &= 5.\end{aligned}$$

Write this system as  $A\vec{u} = \vec{b}$  and then solve it use the inverse matrix.

First, let

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{bmatrix} \quad \vec{u} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \vec{b} = \begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix}$$

We need to find  $A^{-1}$  and will use Gauss-Jordan:

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

$$-R_1 + R_3 \rightarrow R_3 \quad \left[ \begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \end{array} \right]$$

$$\begin{aligned} -2R_2 + R_1 &\rightarrow R_1 \\ R_2 + R_3 &\rightarrow R_3 \end{aligned} \quad \left[ \begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & -2 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & -1 & 1 & 1 \end{array} \right]$$

$$-R_3 \rightarrow R_3 \quad \left[ \begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & -2 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{array} \right]$$

$$\begin{aligned} R_3 + R_2 &\rightarrow R_2 \\ -2R_3 + R_1 &\rightarrow R_1 \end{aligned} \quad \left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 0 & 2 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{array} \right]$$

Thus the inverse is:



$$A^{-1} = \begin{bmatrix} -1 & 0 & 2 \\ 1 & 0 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

Next, we solve the linear system by writing

$$\begin{aligned} \vec{u} &= A^{-1}\vec{b} \\ &= \begin{bmatrix} -1 & 0 & 2 \\ 1 & 0 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix} \\ &= \begin{bmatrix} 7 \\ -2 \\ -2 \end{bmatrix} \end{aligned}$$

and therefore the solution to the linear system is  $x = 7, y = -2$ , and  $z = -2$ .

The last example shows how to perform a linear system solve using the inverse matrix technique of a linear system with 4 equations and 4 unknowns. The technique is exactly the same as the previous example, however there are quite a few steps to find the inverse.

Use Matrix Inversion to solve:

$$\begin{aligned} 2x + y &= 5, \\ x + 2y + z &= 0, \\ y + 2z + w &= 5, \\ z + 2w &= 0. \end{aligned}$$

First, we need to find the inverse. Use the Gauss-Jordan method:

$$\begin{array}{l} R_1 \leftrightarrow R_2, \\ R_2 \leftrightarrow R_3, \\ R_3 \leftrightarrow R_4, \end{array} \begin{bmatrix} 2 & 1 & 0 & 0 & | & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & | & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & | & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & | & 0 & 0 & 0 & 1 \\ 1 & 2 & 1 & 0 & | & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & | & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & | & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 & | & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{array}{l}
-2R_1 + R_4 \rightarrow R_4, \\
3R_2 + R_4 \rightarrow R_4 \\
-2R_2 + R_1 \rightarrow R_1 \\
-4R_3 + R_4 \rightarrow R_4 \\
-2R_3 + R_2 \rightarrow R_2 \\
3R_3 + R_1 \rightarrow R_1
\end{array}
\begin{array}{c}
\left[ \begin{array}{cccc|cccc}
1 & 2 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 2 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 2 & 0 & 0 & 0 & 1 \\
0 & -3 & -2 & 0 & 1 & -2 & 0 & 0
\end{array} \right] \\
\left[ \begin{array}{cccc|cccc}
1 & 0 & -3 & -2 & 0 & 1 & -2 & 0 \\
0 & 1 & 2 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 2 & 0 & 0 & 0 & 1 \\
0 & 0 & 4 & 3 & 1 & -2 & 3 & 0
\end{array} \right] \\
\left[ \begin{array}{cccc|cccc}
1 & 0 & 0 & 4 & 0 & 1 & -2 & 3 \\
0 & 1 & 0 & -3 & 0 & 0 & 1 & -2 \\
0 & 0 & 1 & 2 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & -5 & 1 & -2 & 3 & -4
\end{array} \right]
\end{array}$$
  

$$\begin{array}{l}
2R_4 + 5R_3 \rightarrow R_3, \\
-3R_4 + 5R_2 \rightarrow R_2, \\
4R_4 + 5R_1 \rightarrow R_1, \\
\frac{1}{5}R_1 \rightarrow R_1, \\
\frac{1}{5}R_2 \rightarrow R_2, \\
\frac{1}{5}R_3 \rightarrow R_3, \\
-\frac{1}{5}R_4 \rightarrow R_4,
\end{array}
\begin{array}{c}
\left[ \begin{array}{cccc|cccc}
5 & 0 & 0 & 0 & 4 & -3 & 2 & -1 \\
0 & 5 & 0 & 0 & -3 & 6 & -4 & 2 \\
0 & 0 & 5 & 0 & 2 & -4 & 6 & -3 \\
0 & 0 & 0 & -5 & 1 & -2 & 3 & -4
\end{array} \right] \\
\left[ \begin{array}{cccc|cccc}
1 & 0 & 0 & 0 & 4/5 & -3/5 & 2/5 & -1/5 \\
0 & 1 & 0 & 0 & -3/5 & 6/5 & -4/5 & 2/5 \\
0 & 0 & 1 & 0 & 2/5 & -4/5 & 6/5 & -3/5 \\
0 & 0 & 0 & 1 & -1/5 & 2/5 & -3/5 & 4/5
\end{array} \right]
\end{array}$$

Therefore:

$$A^{-1} = \begin{bmatrix} 4/5 & -3/5 & 2/5 & -1/5 \\ -3/5 & 6/5 & -4/5 & 2/5 \\ 2/5 & -4/5 & 6/5 & -3/5 \\ -1/5 & 2/5 & -3/5 & 4/5 \end{bmatrix}$$

and then the solution to  $AX = B$ , where  $B = [5 \ 0 \ 5 \ 0]^T$  is

$$X = \begin{bmatrix} 4/5 & -3/5 & 2/5 & -1/5 \\ -3/5 & 6/5 & -4/5 & 2/5 \\ 2/5 & -4/5 & 6/5 & -3/5 \\ -1/5 & 2/5 & -3/5 & 4/5 \end{bmatrix} \begin{bmatrix} 5 \\ 0 \\ 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ -7 \\ 8 \\ -4 \end{bmatrix}$$

Note that since the matrix  $A$  is symmetric (in many ways), the matrix  $A^{-1}$  is also symmetric.

## 1.6 Linear Combinations and the Span of a Set of Vectors

In section ??, we discussed a linear combination of rows in a matrix. We can find a linear combination of any vector (or as we will see more generally other things) as well. For example, a linear combination of the vectors  $\begin{bmatrix} 1 & 2 \end{bmatrix}$  and  $\begin{bmatrix} 3 & 4 \end{bmatrix}$  is

$$c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

and these can be any number of vectors and the vectors can be any length.

A **linear combination** of a set of vectors  $\{\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n\}$  is

$$c_1 \vec{s}_1 + c_2 \vec{s}_2 + \dots + c_n \vec{s}_n$$

for some  $c_1, c_2, \dots, c_n \in \mathbb{R}$ .

For example, the solutions of homogeneous systems is an excellent example of a linear combination. Example ?? shows a linear system. The associated homogenous system is

$$\begin{aligned} x_2 + 3x_3 - 9x_4 + 11x_5 &= 14, \\ 2x_3 &+ 4x_5 = 9, \\ 3x_5 &= 27, \end{aligned}$$

and the solution in vector form is

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 9 \\ 0 \\ 1 \\ 0 \end{bmatrix} x_4 \mid x_1, x_4 \in \mathbb{R} \right\}.$$

and thus this solution is a linear combination of the vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 9 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

The solution is the set of all linear combinations and this plays an important role, so it defined as the span of a set of vectors.

The **span** of a nonempty set subset  $S = \{\vec{s}_1, \dots, \vec{s}_n\}$  of a vector space is the set of all linear combinations of the vectors in  $S$ . In particular,

$$\text{span}(S) = \{c_1 \vec{s}_1 + \dots + c_n \vec{s}_n \mid c_1, c_2, \dots, c_n \in \mathbb{R}\}.$$

Understanding the span in a more general way is important. Often the span of a set of vectors is something we already know as the next example shows.

Show that

$$\text{span}\left(\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}\right)$$

can be written as  $\left\{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R}\right\}$ .

In order to show this, we show that for any vector of the form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

there exist  $c_1, c_2$  such that

$$c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

The constants  $c_1$  and  $c_2$  can be found by the following linear system:

$$\begin{aligned} c_1 + c_2 &= x_1 \\ c_2 &= x_2 \end{aligned}$$

with the solution  $c_1 = x_1 - x_2$  and  $c_2 = x_2$ .

Since we found by  $c_1$  and  $c_2$ , a linear combination of these two vectors can produce any vector in  $\mathbb{R}^2$ .

We saw that the span of the two vectors in the example above was  $\mathbb{R}^2$ . What if we add a third vector? The following example answers this question.

Find

$$\text{span}\left(\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}\right)$$

In the case, it is the set formed by the linear combination or specifically

$$\left\{c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mid c_1, c_2, c_3 \in \mathbb{R}\right\}$$

and it seems that this set is also  $\mathbb{R}^2$ . In order to show this, we'll take  $x_1, x_2 \in \mathbb{R}$ , we will solve

$$c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and this can be solved with the linear system

$$\begin{aligned} c_1 + c_2 &= x_1 \\ c_2 + c_3 &= x_2 \end{aligned}$$

If we take  $c_3$  as the free variable, we get

$$c_2 = x_2 - c_3$$

$$c_1 = x_1 - c_2 = x_1 - (x_2 - c_3) = x_1 - x_2 + c_3$$

and since there is a solution, these vectors span  $\mathbb{R}^2$ . It appears that adding another vector does not change the span. This actually does not depend on this vector, but we'll need information in the next section to show this.

And if we continue to investigate this, what if we have two vectors in  $\mathbb{R}^3$ . Is the span equal to  $\mathbb{R}^3$ ? The next example looks as this.

Is

$$\text{span}\left(\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right\}\right)$$

equal to  $\mathbb{R}^3$ ?

In order to check this, then we try to find  $c_1$  and  $c_2$  such that

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

or the linear system

$$c_1 + c_2 = x_1$$

$$c_2 = x_2$$

$$c_1 = x_3$$

however from the latter two equations, there is a contradiction with the first equation which would say  $x_2 + x_3 = x_1$  therefore there is no solution. This means that the two given vectors do not span  $\mathbb{R}^3$ .

If we consider the above example, one of the reasons that this happened is that we don't have enough vectors. The span of two vectors in  $\mathbb{R}^3$  form a plane as we saw in section ???. One way to look at  $\mathbb{R}^3$  is that is the set of all points in  $\mathbb{R}^3$ , so there is no way a plane can equal all points.

## 1.7 Linear Independence

As we saw in Examples ?? and ?? two vectors seem to span  $\mathbb{R}^2$  and a third vector does not contribute anything new (which is why there was a free variable in Example ??). We will be able to explain this with the notion of linear independence and dependence.

Let  $S$  be a set of vectors from  $\mathbb{R}^n$ . The set is **linearly independent** if none of the elements of  $S$  can be written in terms of the other elements of  $S$ . If it is not linearly independent, then  $S$  is said to be **linearly dependent**.

Show that

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

is linearly independent.

In this case, there are only two vectors and it is fairly easy to see that the second cannot be written as a linear combination of the first (or vice versa). Therefore, this set is linearly independent.

Is the set of vectors

$$\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right\}$$

linearly independent or dependent?

In this case, since the second vector is twice the first, these vectors are linearly dependent.

The example above had only two vectors in the set and two vectors are linear independent if and only if they are not constant multipliers of one another. When the set gets larger, it is more difficult to determine this, however the next lemma helps this.

A subset  $S = \{\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n\} \subset \mathbb{R}^n$  is linearly independent if and only if the only relationship of

$$c_1 \vec{s}_1 + c_2 \vec{s}_2 + \dots + c_n \vec{s}_n = \vec{0}$$

is the trivial solution,  $c_1 = c_2 = \dots = c_n = 0$ .

*Proof.*  $\implies$  Assume first that the set  $S$  is linearly independent. This means that no vector can be written as a linear combination of the other vectors so the only relationship is  $c_1 = c_2 = \dots = c_n = 0$ .

$\Leftarrow$  Next, assume that  $S$  is not linearly independent. This means that we can write some vector  $\vec{s}_i$  as a linear combination of the others or

$$\vec{s}_i = c_1 \vec{s}_1 + c_2 \vec{s}_2 + \dots + c_{i-1} \vec{s}_{i-1} + c_{i+1} \vec{s}_{i+1} + \dots + c_n \vec{s}_n$$

which means that there is a solution to the  $c$ 's other than  $c_1 = c_2 = \dots = c_n = 0$ , however in proving the  $\Leftarrow$ , we assumed that this was the only solution, therefore, by contradiction,  $S$  must be linear independent.  $\square$

Show using the lemma above that the set

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

is linearly independent.

We will show that only solution to

$$c_1 \vec{s}_1 + c_2 \vec{s}_2 + c_3 \vec{s}_3 = \vec{0} \quad \text{or} \quad c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

is the trivial solution. Writing down an equation for each component leads to

$$\begin{aligned} c_1 + c_2 &= 0 \\ 2c_2 &= 0, \\ c_3 &= 0. \end{aligned}$$

The latter two equation show that  $c_2 = c_3 = 0$  and this can be used in the first equation to show that  $c_1 = 0$ , thus this is the trivial solution. From lemma ??, this set is linearly independent.

The last example in this section will be a review of why we got the results in Example ??.

Is the set of vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

linearly independent or dependent?

Again, we will use Lemma ?.?. We solve a linear combination of these vectors and set them equal to the zero vector.

$$c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or the linear system in terms of the  $c$ 's

$$\begin{aligned} c_1 + c_2 &= 0 \\ c_2 + c_3 &= 0 \end{aligned}$$

and in this case, let  $c_3$  be a free variable, then the solution

$$\begin{aligned} c_2 &= -c_3 \\ c_1 &= -c_2 = c_3 \end{aligned}$$

Since this is not the trivial solution, then these vectors are linearly dependent. The relationship between the constants shows the dependence. That is

$$c_3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - c_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or as a simple relationship

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



## 1.8 Row Space, Column Space and Rank of a Matrix

If we examine a matrix, we can think about the number of linearly independent row or columns in a matrix or also the span of the set of rows and columns in a matrix. We will see how these concepts are connected to other concepts from this chapter.

### 1.8.1 Row Space of a Matrix

The **row space** of a matrix is the span of the set of its rows. The **row rank** is the number of linearly independent rows of the matrix.

The row space of

$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$$

is the set of all 2-component row vectors with the form:

$$\{c_1 \cdot \begin{bmatrix} 2 & 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 & 0 \end{bmatrix} \mid c_1, c_2 \in \mathbb{R}\}$$

and this is the set of all 2-component row vectors. The row rank of this matrix is 2.

The row space of

$$\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$$

can be written as

$$\{c_1 \cdot \begin{bmatrix} 2 & 1 \end{bmatrix} + c_2 \begin{bmatrix} 4 & 2 \end{bmatrix} \mid c_1, c_2 \in \mathbb{R}\}$$

However, since the two vectors are not linearly independent, the row space can be written as

$$\{c_1 \cdot \begin{bmatrix} 2 & 1 \end{bmatrix} \mid c_1 \in \mathbb{R}\}$$

Therefore the row rank is 1.

If the matrices  $A$  and  $B$  are related by transformation by an elementary row operation, then their row spaces are equivalent.

*Proof.* By the Corollary ??, each row of  $B$  is a linear combination of each row of  $A$ , therefore it is in the row space of  $A$  or  $\text{rowspace}(B) \subset \text{rowspace}(A)$ .

Similarly, each row of  $A$  is a linear combination of  $B$ , so using the same argument,  $\text{rowspace}(A) \subset \text{rowspace}(B)$ , therefore  $\text{rowspace}(A) = \text{rowspace}(B)$ .  $\square$

The nonzero rows of an echelon matrix make up a linearly independent set of rows.

*Proof.* Lemma ?? states that in an echelon matrix, no nonzero row is a linear combination of the other rows.  $\square$

Although this proof is almost trivial, the main point of this is to reframe what we saw in the first chapter in terms of this newer framework.

Gaussian Reduction to echelon form eliminates linear dependence between the rows, leaves the row space unchanged, and results in a set of linearly independent rows whose span is the row space.

Consider the linear system (written as a matrix) from Example ??:

$$\begin{bmatrix} 4 & 0 & -1 & 0 \\ 1 & 3 & 2 & 3 \\ 0 & 3 & 5 & 14 \end{bmatrix}$$

Perform row operations to put it in echelon form and find the row rank.

The row operations are

$$\begin{aligned} -R_1 + 4R_2 &\rightarrow R_2, \\ -R_2 + 4R_3 &\rightarrow R_3 \end{aligned} \quad \begin{bmatrix} 4 & 0 & -1 & 0 \\ 0 & 12 & 9 & 12 \\ 0 & 0 & 11 & 44 \end{bmatrix}$$

Since this is in echelon form now, there are three linearly independent rows, so the row rank is 3.

### 1.8.2 Column Space of a Matrix

The **column space** of a matrix is the span of the set of its columns. The **column rank** is the number of linearly independent columns of the matrix.

Find the column space of

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 0 & -4 \\ 0 & 1 & 1 \\ 3 & 4 & -2 \end{bmatrix}$$

In this case, we look for a span of

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ -4 \\ 1 \\ -2 \end{bmatrix} \right\}$$

Instead of trying to eliminate any linear dependence as written, note that if we treat the columns as rows. Recall that this operation is called the transpose of the matrix.

To find the column space of a matrix  $A$ , find the row space of  $A^T$  and transpose the resulting row space.

Find the column space and column rank of

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 0 & -4 \\ 0 & 1 & 1 \\ 3 & 4 & -2 \end{bmatrix}$$

1. First take the transpose

$$A^T = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 3 & 0 & 1 & 4 \\ 1 & -4 & 1 & -2 \end{bmatrix}$$

2. Now row reduce the matrix:

$$\begin{array}{l} -3R_1 + R_2 \rightarrow R_2, \\ -R_1 + R_3 \rightarrow R_3, \\ \\ -R_2 + R_3 \rightarrow R_3 \end{array} \quad \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & -6 & 1 & -5 \\ 0 & -6 & 1 & -5 \\ \\ 1 & 2 & 0 & 3 \\ 0 & -6 & 1 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which is now in echelon form, so we know that the non-zero rows are linearly independent.

3. Take the transpose of the first two rows. The span of this is the column space:

$$\text{colspace}(A) = \text{span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ -6 \\ 1 \\ -5 \end{bmatrix} \right\}\right)$$

And since there are two linearly independent columns, the column rank is 2.

The following theorem explains the relationship between the row rank and column rank.

The row rank of a matrix equals its column rank.

And because of this fact (which we won't prove here), the rank of a matrix is typically the interesting property.

The **rank** of its matrix equals the row or column rank of the matrix.

### 1.8.3 Null Space of a Matrix

Let  $\vec{x}$  be any vector that satisfies  $A\vec{x} = \vec{0}$ . The **null space** of the matrix  $A$  is the span of all vectors  $\vec{x}$ . The number of independent vectors in the null space is called the **nullity** of  $A$ .

Find the null space of the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & -1 \end{bmatrix}$$

We can solve this by solving the homogeneous matrix equation.

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

and performing the row operation  $-2R_1 + R_2 \rightarrow R_2$

$$\begin{array}{l} 2R_2 + R_1 \rightarrow R_1 \\ -R_2 \rightarrow R_2 \end{array} \quad \left[ \begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -1 & -7 & 0 \\ 1 & 0 & -11 & 0 \\ 0 & -1 & -7 & 0 \end{array} \right]$$

$$\left[ \begin{array}{ccc|c} 1 & 0 & -11 & 0 \\ 0 & 1 & 7 & 0 \end{array} \right]$$

results in the reduced row-echelon form. The resulting equations are

$$\begin{aligned} x_1 - 11x_3 &= 0 \\ x_2 + 7x_3 &= 0 \end{aligned}$$

and the solution set can be written as in vector form:

$$\left\{ \begin{bmatrix} -11 \\ 7 \\ 1 \end{bmatrix} x_3 \mid x_3 \in \mathbb{R} \right\}$$

and this set is the null space of the matrix. The null space is

$$\text{span}\left(\left\{ \begin{bmatrix} -11 \\ 7 \\ 1 \end{bmatrix} \right\}\right)$$

and therefore the dimension of the null space is 1, and thus the nullity of  $A$  is 1.

To find the null space of a matrix  $A$ , find the row reduced form the matrix, and find the homogeneous solution. The solution set is the null space.

The basis of the homogeneous solution set is the basis of the null space.

The following example shows a larger example.

Find the null space of

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 2 \\ 2 & 1 & 5 & 2 \\ 0 & -2 & -2 & -4 \end{bmatrix}$$

Again, we solve the homogeneous linear system  $A\vec{x} = \vec{0}$  by finding the reduced row echelon form of  $A$

$$\begin{array}{l} -2R_1 + R_2 \rightarrow R_2 \\ \\ -R_2 + R_3 \rightarrow R_3, \\ 2R_2 + R_4 \rightarrow R_4, \end{array} \quad \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & -2 & -2 & -4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and the corresponding equations are:

$$\begin{aligned} x_1 + 2x_3 &= 0 \\ x_2 + x_3 + 2x_4 &= 0 \end{aligned}$$

which has the solution set:

$$\left\{ \begin{bmatrix} -2 \\ -1 \\ 1 \\ 0 \end{bmatrix} x_3 + \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} x_4 \mid x_3, x_4 \in \mathbb{R} \right\}$$

and this is the null space of the matrix  $A$ . Since there are two linearly independent vectors that span the space, the dimension of the null space of the matrix is 2. Thus the nullity of  $A$  is 2.

#### 1.8.4 Relationship between the Rank and Nullity

The relationship between the rank and the nullity of a  $m \times n$  matrix  $A$  is summed up in the following theorem.

[Rank-Nullity Theorem] Let  $A$  be an  $m \times n$  matrix. The sum of the rank of  $A$  and the nullity of  $A$  is  $n$ .

Show that the Rank-Nullity Theorem holds for Examples ?? and ??.

In Example ??, the  $2 \times 3$  matrix is reduced to reduced row echelon form and it shows that there are 2 non-zero rows, hence the rank of the matrix is 2. Also in Example ??, the nullity was shown to be 1 and the sum is 3, the number of columns of  $A$ .

In Example ??, the  $4 \times 4$  matrix is reduced to reduced row echelon form and it shows that there are 2 non-zero rows, and again the rank of the matrix is 2. The example also shows that the nullity is 2 and the sum is 4, the number of columns of  $A$ .

## 1.9 Determinants of Square Matrices

The determinant is a useful function for specifying when matrix is invertible or singular. For example, in section ??, we saw that for

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

that the inverse matrix exists if and only if  $ad - bc \neq 0$ . The determinant of this matrix will be the function  $ad - bc$ . We will see two different methods for finding the determinant of an  $n \times n$  matrix.

### 1.9.1 Properties of the Determinant

An  $n \times n$  **determinant** is a function  $\det : \mathcal{M}_{n \times n} \rightarrow \mathbb{R}$  such that

1.  $\det(\vec{\rho}_1, \dots, k\vec{\rho}_i + \vec{\rho}_j, \dots, \vec{\rho}_n) = \det(\vec{\rho}_1, \dots, \vec{\rho}_j, \dots, \vec{\rho}_n)$  for  $i \neq j$ ,
2.  $\det(\vec{\rho}_1, \dots, \vec{\rho}_i, \dots, \vec{\rho}_j, \dots, \vec{\rho}_n) = -\det(\vec{\rho}_1, \dots, \vec{\rho}_j, \dots, \vec{\rho}_i, \dots, \vec{\rho}_n)$  for  $i \neq j$ ,
3.  $\det(\vec{\rho}_1, \dots, k\vec{\rho}_i, \dots, \vec{\rho}_n) = k \det(\vec{\rho}_1, \dots, \vec{\rho}_i, \dots, \vec{\rho}_n)$ ,
4.  $\det(I) = 1$ , where  $I$  is the identity matrix.

The  $\vec{\rho}_i$ 's are the rows of the matrix and often the notation  $|T| = \det(T)$  is used.

In more colloquial terms, these properties are:

1. if one multiplies a row by a constant and adds to a second row that the determinant is unchanged.
2. If one swaps rows in a matrix, the determinant of the new matrix is the opposite sign of the original matrix.
3. If one multiplies any row of a matrix by  $k$  then the determinant of the new matrix is  $k$  times the determinant of the original matrix.
4. The determinant of the identity matrix is 1.

It appears for the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

that the function  $f(A) = ad - bc$  satisfies the properties of the determinant for a  $2 \times 2$  matrix. Show this is true for arbitrary rows.

We know  $f(A) = ad - bc$ . We will find  $A'$  which is a row operation applied to  $A$  and then find  $f(A')$  for each of these.

1.

$$kR_1 + R_2 \rightarrow R_2 \quad A' = \begin{bmatrix} a & b \\ ka + c & kb + d \end{bmatrix}$$

and

$$\begin{aligned} \det(A') &= a(kb + d) - b(ka + c) \\ &= kab + ad - kab - bc = ad - bc = \det(A) \end{aligned}$$

2.

$$R_1 \leftrightarrow R_2 \quad A' = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$$

and

$$\det(A') = cb - ad = -\det(A)$$

3.

$$kR_1 \rightarrow R_1 \quad A' = \begin{bmatrix} ka & kb \\ c & d \end{bmatrix}$$

and

$$\det(A') = kad - kbc = k(ad - bc) = k \det(A)$$

4. And finally if

$$A' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

then

$$\det(A') = 1(1) - 0(0) = 1$$

This shows that the function  $ad - bc$  satisfies all of the properties of the determinant, so we can use the formula

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

1. A matrix with two identical rows has a zero determinant.
2. A matrix with a row of zeros has a zero determinant.



3. A matrix is nonsingular if and only if its determinant is nonzero.
4. The determinant of a matrix in echelon form is the product of its diagonal elements.

*Proof.* 1. Consider a matrix with two identical rows. Swapping the two rows changes the sign of the determinant, however leaves the matrix unchanged. This implied that the determinant must be zero.

2. If the matrix is size  $1 \times 1$  with its only element of 0, then the statement is true. Assume it has at least two rows. Let  $j$  be the row with all zeros. The operation  $R_i + R_j \rightarrow R_j$  for any  $i \neq j$  results in a matrix with two identical rows, which has a zero determinant from statement 1.
3. Recall that for a matrix to be nonsingular, that a linear system with coefficients from the matrix has a unique system or the row reduced echelon form of the matrix has no row of zeros. Let  $T$  be the original matrix and  $\hat{T}$  is the matrix in echelon form. If the  $\hat{T}$  is nonsingular, then it is the identity matrix, which has the nonzero determinant 1. A singular matrix  $T$  reduces to a matrix  $\hat{T}$  with a row of zeros. By #2, this determinant is 0.

4. Consider a matrix in echelon form:

$$|\hat{T}| = \begin{vmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,n} \\ 0 & t_{2,2} & & t_{2,n} \\ \vdots & & \ddots & \\ 0 & 0 & \vdots & t_{n,n} \end{vmatrix}$$

Using property (3) of the determinant definition,

$$= t_{1,1}t_{2,2} \cdots t_{n,n} \begin{bmatrix} 1 & t_{1,2}/t_{1,1} & \cdots & t_{1,n}/t_{1,1} \\ 0 & 1 & & t_{2,n}/t_{2,2} \\ \vdots & & \ddots & \\ 0 & 0 & \vdots & 1 \end{bmatrix}$$

The matrix can be reduced to the identity matrix, with determinant 1

$$= t_{1,1}t_{2,2} \cdots t_{n,n}$$

□

The 1st, 2nd and 4th statements of the lemma are very helpful for calculating the determinant. The 4th statement gives a very general way to calculate a determinant. Use the elementary row operations to put the matrix in echelon form (recall to keep track of multiplying any row by a constant and row swaps). The determinant will then be the product of the diagonal elements.

The third statement of the lemma above gives a function (the determinant) that determines whether or not a matrix is singular.

### 1.9.2 Calculating Determinants

Find the determinant of the following matrix using a) the formula for  $2 \times 2$  determinants and b) using Gauss' method.

$$T = \begin{bmatrix} 3 & 2 \\ 1 & -2 \end{bmatrix}$$

Using the formula  $|T| = ad - bc = -6 - 2 = -8$ .

Using Gauss's method,

$$\begin{aligned} |T| &= \begin{vmatrix} 3 & 2 \\ 1 & -2 \end{vmatrix} \\ R_1 \leftrightarrow R_2 & \quad -|T| = \begin{vmatrix} 1 & -2 \\ 3 & 2 \end{vmatrix} \\ -3R_1 + R_2 \rightarrow R_2 & \quad -|T| = \begin{vmatrix} 1 & -2 \\ 0 & 8 \end{vmatrix} = 8 \end{aligned}$$

So  $|T| = -8$ .

This shows that although Gauss' method succeeds in finding the determinant, it takes more operations than the simple formula.

Using Gauss's method to find the determinants of the following matrices:

$$T = \begin{bmatrix} 3 & 0 & 2 \\ 1 & 4 & 0 \\ 0 & 2 & 5 \end{bmatrix} \quad S = \begin{bmatrix} 0 & 1 & 3 & -4 \\ 2 & 0 & 2 & 7 \\ 0 & 0 & 6 & 8 \\ 1 & 0 & 10 & 6 \end{bmatrix}$$

$$\begin{aligned} |T| &= \begin{vmatrix} 3 & 0 & 2 \\ 1 & 4 & 0 \\ 0 & 2 & 5 \end{vmatrix} \\ R_1 \leftrightarrow R_2 & \quad -|T| = \begin{vmatrix} 1 & 4 & 0 \\ 3 & 0 & 2 \\ 0 & 2 & 5 \end{vmatrix} \\ -3R_1 + R_2 \rightarrow R_2 & \quad -|T| = \begin{vmatrix} 1 & 4 & 0 \\ 0 & -12 & 2 \\ 0 & 2 & 5 \end{vmatrix} \\ R_2 \leftrightarrow R_3 & \quad |T| = \begin{vmatrix} 1 & 4 & 0 \\ 0 & 2 & 5 \\ 0 & -12 & 2 \end{vmatrix} \\ 6R_2 + R_3 \rightarrow R_3 & \quad |T| = \begin{vmatrix} 1 & 4 & 0 \\ 0 & 2 & 5 \\ 0 & 0 & 32 \end{vmatrix} = 64 \end{aligned}$$

$$\begin{aligned}
|S| &= \begin{vmatrix} 0 & 1 & 3 & -4 \\ 2 & 0 & 2 & 7 \\ 0 & 0 & 6 & 8 \\ 1 & 0 & 10 & 6 \end{vmatrix} \\
R_1 \leftrightarrow R_4 \quad -|S| &= \begin{vmatrix} 1 & 0 & 10 & 6 \\ 2 & 0 & 2 & 7 \\ 0 & 0 & 6 & 8 \\ 0 & 1 & 3 & -4 \end{vmatrix} \\
-2R_1 + R_2 \rightarrow R_2 \quad |S| &= - \begin{vmatrix} 1 & 0 & 10 & 6 \\ 0 & 0 & -18 & -5 \\ 0 & 0 & 6 & 8 \\ 0 & 1 & 3 & -4 \end{vmatrix} \\
R_2 \leftrightarrow R_4 \quad |S| &= \begin{vmatrix} 1 & 0 & 10 & 6 \\ 0 & 1 & 3 & -4 \\ 0 & 0 & 6 & 8 \\ 0 & 0 & -18 & -5 \end{vmatrix} \\
3R_3 + R_4 \rightarrow R_4 \quad |S| &= \begin{vmatrix} 1 & 0 & 10 & 6 \\ 0 & 1 & 3 & -4 \\ 0 & 0 & 6 & 8 \\ 0 & 0 & 0 & 19 \end{vmatrix} = 6(19) = 114
\end{aligned}$$

### 1.9.3 Expansion Method for finding the Determinant

Although Gauss' method is a very robust and in general efficient method for finding determinants, a method called the Laplace Expansion method can be quite helpful at times as well. Before defining this, we need to know a matrix minor and cofactor first.

For any  $n \times n$  matrix  $T$ , the  $(n-1) \times (n-1)$  matrix formed by deleting row  $i$  and column  $j$  of  $T$  is the  $i, j$  **minor** of  $T$ . The  $i, j$  **cofactor**  $T_{i,j}$  of  $T$  is  $(-1)^{i+j}$  times the determinant of the  $i, j$  minor of  $T$  and denoted  $(-1)^{i+j}|T_{i,j}|$ .

Find the  $T_{1,1}$  and  $T_{2,3}$  minors of the matrix

$$T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Recall that  $T_{i,j}$  minor is found by removing the  $i$ th row and  $j$ th column or

$$T_{1,1} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} \qquad T_{2,3} = \begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix}$$

and the cofactors are the determinants of each of these matrices times  $(-1)^{i+j}$  or

$$\begin{aligned}
(-1)^{1+1}|T_{1,1}| &= (1)(45 - 24) = 21 \\
(-1)^{2+3}|T_{2,3}| &= (-1)(8 - 14) = 6
\end{aligned}$$

Now that we have the prerequisites, the following is the Laplace Expansion method for finding a determinant.

[Laplace Expansion of Determinants] The determinant of an  $n \times n$  matrix  $T$  can be found by expanding by cofactors on row  $i$  or column  $j$  of  $T$ . That is

$$|T| = t_{i,1}T_{i,1} + t_{i,2}T_{i,2} + \cdots + t_{i,n}T_{i,n}$$

for any  $i$  satisfying  $1 \leq i \leq n$  or

$$|T| = t_{1,j}T_{1,j} + t_{2,j}T_{2,j} + \cdots + t_{n,j}T_{n,j}$$

for any  $j$  satisfying  $1 \leq j \leq n$ .

Use the expansion formula to find the determinants of the matrices in Example ??, namely

$$T = \begin{bmatrix} 3 & 0 & 2 \\ 1 & 4 & 0 \\ 0 & 2 & 5 \end{bmatrix} \qquad S = \begin{bmatrix} 0 & 1 & 3 & -4 \\ 2 & 0 & 2 & 7 \\ 0 & 0 & 6 & 8 \\ 1 & 0 & 10 & 6 \end{bmatrix}$$

In the case of  $T$ , we will expand across the first row and use the formula for the  $2 \times 2$  determinant.

$$\begin{aligned} |T| &= (-1)^{1+1}(3) \begin{vmatrix} 4 & 0 \\ 2 & 5 \end{vmatrix} + (-1)^{1+2}(0) \begin{vmatrix} 1 & 0 \\ 0 & 5 \end{vmatrix} + (-1)^{1+3}(2) \begin{vmatrix} 1 & 4 \\ 0 & 2 \end{vmatrix} \\ &= 3(20) + (2)(2 - 0) = 64 \end{aligned}$$

and for  $S$ , we'll expand down the 2nd column because all but one is zero. And because of this, I won't show the cofactors of  $T_{1,2}$ ,  $T_{1,3}$  and  $T_{1,4}$ .

$$|S| = (-1)^{1+2}(1) \begin{vmatrix} 2 & 2 & 7 \\ 0 & 6 & 8 \\ 1 & 10 & 6 \end{vmatrix} + 0 + 0 + 0$$

and now to find this  $3 \times 3$  determinant, expand about the 2nd row

$$|S| = (-1)((-1)^{2+2}(6) \begin{vmatrix} 2 & 7 \\ 1 & 6 \end{vmatrix} + (-1)^{2+3}(8) \begin{vmatrix} 2 & 2 \\ 1 & 10 \end{vmatrix})$$

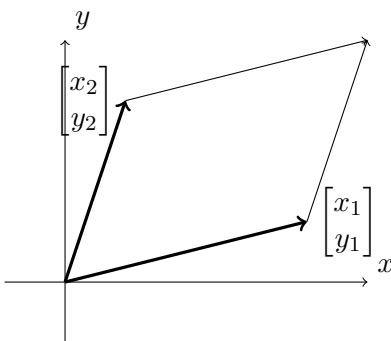
and now use the formula for  $2 \times 2$  determinants.

$$|S| = -(6(12 - 7) - 8(20 - 2)) = -(30 - 144) = 114$$

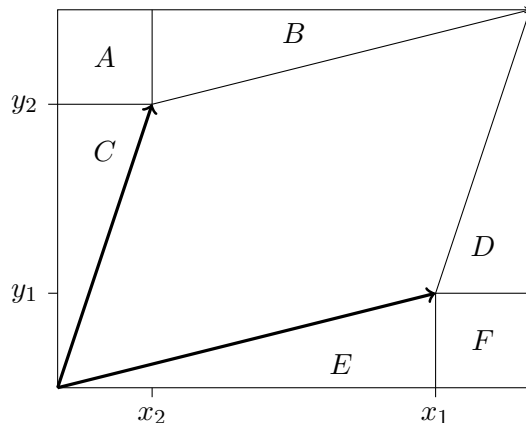
### 1.9.4 Geometry of Determinants

In the previous section, the determinant was introduced as a function that determines whether or not a matrix was singular due to whether or not the function was 0. In this section, we will look at a geometric approach to the determinant and show that it can be used to determine areas (and volumes) of regions bounded by vectors. We will show that this geometric approach is identical (in the two-dimensional case) as the properties in Definition ??.

Consider the parallelogram formed by two vectors. In the argument below, it is important that the vector  $\langle x_1, y_1 \rangle$  is below and to the right of the vector  $\langle x_2, y_2 \rangle$ .



The area of the parallelogram can be determined by taking the area of the enclosing rectangle and subtracting out the rectangles  $A$  and  $F$  and triangles  $B, C, D$  and  $E$  as shown below:



$$\begin{aligned}
&\text{area of parallelogram} = \text{area of enclosing rect} \\
&\quad - \text{area of rectangle } A - \text{area of triangle } B \\
&\quad \cdots - \text{area of rectangle } F \\
&= (x_1 + x_2)(y_1 + y_2) - x_2y_1 - \frac{1}{2}x_1y_1 \\
&\quad - \frac{1}{2}x_2y_2 - \frac{1}{2}x_2y_2 - \frac{1}{2}x_1y_1 - x_2y_1 \\
&= x_1y_2 - x_2y_1
\end{aligned}$$

and note that

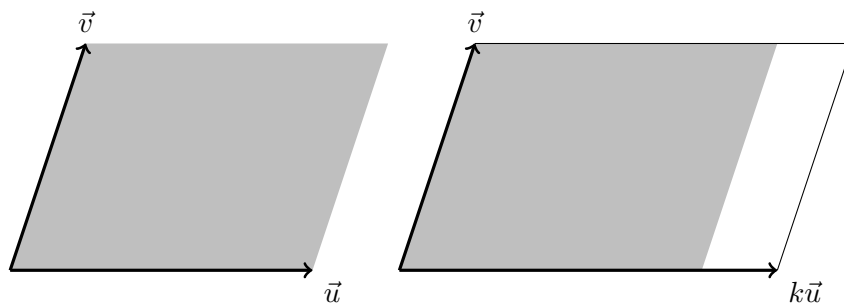
$$\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = x_1y_2 - x_2y_1$$

And this result is identical to the determinant seen above. Again, as noted, the vectors were set up to have a positive area, however in general, one can define the area as the absolute value of the determinant.

### Transformation of the Vectors and the size of the Parallelogram

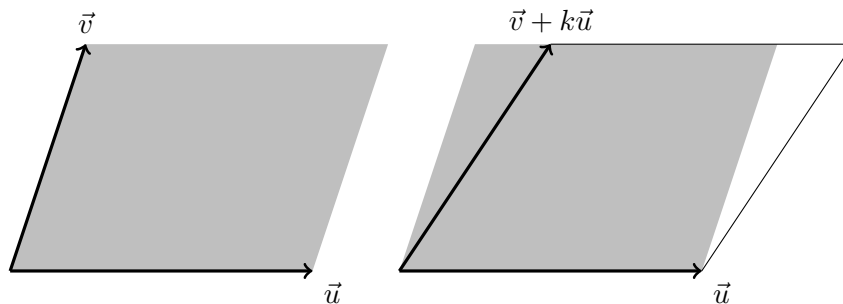
From above, the area of the parallelogram is the

Consider two vectors in  $\mathbb{R}^2$  and rotate them so one is on the  $x$ -axis. Also take  $\vec{u}$  and multiply it by a factor of  $k$



From this geometric argument, the area of the parallelogram formed by the vectors  $\vec{v}$  and  $k\vec{u}$  appears to be  $k$  times larger. This is property 3 of Definition ??.

Next, let's look at transformation  $\vec{u} + k\vec{v}$ . The picture on the left is the original two vectors and that on the right is the transformed vectors (with  $k$  about 0.2 in this picture). The original area and the transformed area are identical in this case since neither the height of the parallelogram nor its width has changed.



This property shows that replacing a row with a constant times another row plus the current row results in an unchanged area is consistent with property 1 of Definition ??.

The other transformation related to the determinant is property 2 of Definition ?? or in other words, if one switched the order of the vectors (row swaps), that the determinant changes sign. The area does not change because the area is the absolute value of the determinant.

In  $\mathbb{R}^n$ , the **parallelepiped** formed by  $\langle \vec{v}_1, \vec{v}_2, \dots, \vec{v}_n \rangle$  includes the set

$$\{t_1\vec{v}_1 + t_2\vec{v}_2 + \dots + t_n\vec{v}_n \mid t_1, t_2, \dots, t_n \in [0, 1]\}$$

The **volume** of the parallelepiped is the absolute value of the determinant of the matrix whose columns are  $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ .

Find the volume of the parallelepiped formed by the vectors:

$$\begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

The volume is the absolute value of the determinant of the matrix with these three columns. We'll use Gauss' method to find the determinant.

$$\begin{aligned} |A| &= \begin{vmatrix} 3 & -1 & 2 \\ 0 & 2 & 3 \\ 2 & 0 & 1 \end{vmatrix} \\ 3R_3 \rightarrow R_3 \quad 3|A| &= \begin{vmatrix} 3 & -1 & 2 \\ 0 & 2 & 3 \\ 6 & 0 & 3 \end{vmatrix} \\ -2R_1 + R_3 \rightarrow R_3 \quad 3|A| &= \begin{vmatrix} 3 & -1 & 2 \\ 0 & 2 & 3 \\ 0 & 2 & -1 \end{vmatrix} \\ -R_2 + R_3 \rightarrow R_3 \quad 3|A| &= \begin{vmatrix} 3 & -1 & 2 \\ 0 & 2 & 3 \\ 0 & 0 & -4 \end{vmatrix} \end{aligned}$$

and multiplying down the diagonal,  $3|A| = -24$ , so  $|A| = -8$ . This means that the volume is 8 units.

## 1.10 Summary of the basics of linear algebra

We finish this chapter with a summarizing theorem about linear algebra. This covers how much of all of the above concepts are related.

Let  $A$  be an  $n$  by  $n$  matrix of real numbers. The following statements are equivalent:

1.  $|A| \neq 0$
2. The inverse matrix,  $A^{-1}$  exists.
3. For every vector  $\vec{b}$ , the equation  $A\vec{x} = \vec{b}$  has a unique solution  $\vec{x} = A^{-1}\vec{b}$ .
4. The matrix equation  $A\vec{x} = \vec{0}$  has the unique solution  $\vec{x} = \vec{0}$ .
5. The columns of  $A$  are linearly independent.
6. The rows of  $A$  are linearly independent.
7. The rank of  $A$  is  $n$ .
8. The column space of  $A$  is  $\mathbb{R}^n$
9. The row space of  $A$  is  $\mathbb{R}^n$ .
10. The null space of  $A$  is  $\{\vec{0}\}$  and the nullity is 0.
11. The reduced row echelon form of  $A$  is  $I$ , the identity matrix.

Examine Theorem ?? using the matrix

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & -1 \\ 0 & 3 & -1 \end{bmatrix}$$

In this example, we will show all of the equivalent properties directly on the matrix  $A$ .

1. First find the determinant. Using the Laplace expansion method and expanding down the first column

$$|A| = 1 \begin{vmatrix} 2 & -1 \\ 3 & -1 \end{vmatrix} = 1(2(-1) - (-1)3) = 1$$

which is nonzero.



2. Next, we'll find the inverse matrix:

$$\begin{array}{l}
 \begin{bmatrix} 1 & 0 & 3 & | & 1 & 0 & 0 \\ 0 & 2 & -1 & | & 0 & 1 & 0 \\ 0 & 3 & -1 & | & 0 & 0 & 1 \end{bmatrix} \\
 -3R_2 + 2R_3 \rightarrow R_3 \quad \begin{bmatrix} 1 & 0 & 3 & | & 1 & 0 & 0 \\ 0 & 2 & -1 & | & 0 & 1 & 0 \\ 0 & 0 & 1 & | & 0 & -3 & 2 \end{bmatrix} \\
 \begin{array}{l} R_3 + R_2 \rightarrow R_2, \\ -3R_3 + R_1 \rightarrow R_1 \end{array} \quad \begin{bmatrix} 1 & 0 & 0 & | & 1 & 9 & -6 \\ 0 & 2 & 0 & | & 0 & -2 & 2 \\ 0 & 0 & 1 & | & 0 & -3 & 2 \end{bmatrix} \\
 \frac{1}{2}R_2 \rightarrow R_2 \quad \begin{bmatrix} 1 & 0 & 0 & | & 1 & 9 & -6 \\ 0 & 1 & 0 & | & 0 & -1 & 1 \\ 0 & 0 & 1 & | & 0 & -3 & 2 \end{bmatrix}
 \end{array}$$

and this shows that the inverse matrix is

$$A^{-1} = \begin{bmatrix} 1 & 9 & -6 \\ 0 & -1 & 1 \\ 0 & -3 & 2 \end{bmatrix}$$

3. Since the inverse matrix exists, then a unique solution to  $A\vec{x} = \vec{b}$  can be found by  $\vec{x} = A^{-1}\vec{b}$ .

4. Again, since  $A^{-1}$  exists, then

$$\vec{x} = A^{-1}\vec{0} = \vec{0}$$

5. See #8 below.

6. See #9 below.

7. See #8 and #9 below.

8. The column space is found by row reducing  $A^T$ .

$$\begin{array}{l}
 A^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 3 & -1 & -1 \end{bmatrix} \\
 -3R_1 + R_3 \rightarrow R_3 \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & -1 & -1 \end{bmatrix} \\
 R_2 + 2R_3 \rightarrow R_3 \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}
 \end{array}$$

which is now in echelon form and since all nonzero row in a echelon form matrix are linearly independent, this shows #5.

Continuing to put this is reduced row echelon form:

$$\begin{array}{l} -3R_3 + R_2 \rightarrow R_2 \\ \frac{1}{2}R_2 \rightarrow R_2 \end{array} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and this shows that the column space of  $A$  is

$$\text{span}\left(\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}\right)$$

which is all of  $\mathbb{R}^3$ . This also shows that the rank of  $A$  is 3.

9. In a similar manner to #8, we put  $A$  in reduced row echelon form:

$$\begin{array}{l} -3R_2 + 2R_3 \rightarrow R_3 \\ R_3 + R_2 \rightarrow R_2, \\ -3R_3 + R_1 \rightarrow R_1 \\ \frac{1}{2}R_2 \rightarrow R_2 \end{array} \quad \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & -1 \\ 0 & 3 & -1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and thus the row space is

$$\text{span}(\{[1 \ 0 \ 0], [0 \ 1 \ 0], [0 \ 0 \ 1]\})$$

and this is  $\mathbb{R}^3$ . In addition, this shows that the rank of  $A$  is 3.

10. The null space is the set of all  $\vec{x}$  such that  $A\vec{x} = \vec{0}$ , but from #4, we showed that the only solution to this is  $\vec{x} = \vec{0}$ . The nullity is the number of linearly independent vectors in this set which is 0, by definition.

11. From #9, we showed that the reduced row echelon form of  $A$  is  $I$ .

Let  $A$  be an  $n \times n$  real matrix. The following statements are equivalent

1.  $|A| = 0$
2. The matrix equation  $A\vec{x} = \vec{b}$  has no solution or an infinite number of solutions.
3. The matrix equation  $A\vec{x} = \vec{0}$  has a nontrivial solution.
4. The rank of  $A$  is less than  $n$ .

This theorem will be extremely helpful in finding a certain type of scalar and vector called an eigenvalue and eigenvector. The following example shows its usefulness.

Let

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -2 & 1 \\ 2 & -2 & 5 \end{bmatrix}$$

Show that  $A\vec{x} = \vec{0}$  has a nontrivial solution first by using Theorem ?? then by directly finding solutions.

First, find the determinant by expansion:

$$\begin{aligned} |A| &= 1 \begin{vmatrix} -2 & 1 \\ -2 & 5 \end{vmatrix} + 2 \begin{vmatrix} 0 & -2 \\ 2 & -2 \end{vmatrix} \\ &= (-10 - (-2)) + 2(0 - (-4)) = -8 + 8 = 0 \end{aligned}$$

and therefore by Theorem ??, there is a nontrivial solution to  $A\vec{x} = \vec{0}$

Next, we'll solve the matrix equation by Gauss' method.

$$\begin{array}{l} \begin{bmatrix} 1 & 0 & 2 & | & 0 \\ 0 & -2 & 1 & | & 0 \\ 2 & -2 & 5 & | & 0 \end{bmatrix} \\ -2R_1 + R_3 \rightarrow R_3 \quad \begin{bmatrix} 1 & 0 & 2 & | & 0 \\ 0 & -2 & 1 & | & 0 \\ 0 & -2 & 1 & | & 0 \end{bmatrix} \\ -R_2 + R_3 \rightarrow R_3 \quad \begin{bmatrix} 1 & 0 & 2 & | & 0 \\ 0 & -2 & 1 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix} \end{array}$$

The resulting equations are

$$\begin{aligned} x_1 + 2x_3 &= 0 \\ -2x_2 + x_3 &= 0 \end{aligned}$$

or

$$\begin{aligned}x_1 &= -2x_3 \\x_2 &= \frac{1}{2}x_3\end{aligned}$$

and the solution set is

$$\left\{ \begin{bmatrix} -2 \\ 1/2 \\ 1 \end{bmatrix} x_3 \mid x_3 \in \mathbb{R} \right\}$$

This shows directly that the matrix equation  $A\vec{x} = \vec{0}$  does not have a trivial solution.

Also notice that the last matrix of Gauss' method showed that the rank was 2 (since there were only 2 nonzero rows). This was another statement in the theorem.

## Chapter 2

# Vector Spaces, Inner Product Spaces and Linear Transformations

This chapter explores linear systems in a more abstract manner. We will see that framing vector and matrix addition and scalar multiplication in a more general way will allow us to make connections between seemingly disparate parts of mathematics and then discover new things because of this.

We first investigate vector spaces, which at first seem like they are just sets of vectors, but are actually much more. Then we will see inner product spaces, another way of looking at vectors and finally linear transformations.

### 2.1 Vector Spaces

A vector space as we will see is a set of vectors with a certain collection of properties. This notion arose from vectors of  $\mathbb{R}^n$ , however, we will see that they generalize.

Let  $V$  be a nonempty set of elements with  $\vec{x}, \vec{y}, \vec{z} \in V$  and  $c, k \in \mathbb{R}$ . The set  $V$  is called a **real vector space** and the elements of  $V$  are called **vectors** if in  $V$  there are defined two algebraic operations (called *vector addition*,  $+$  and *scalar multiplication*,  $\cdot$ ) that satisfy the following:

**Vector Addition:** The operation  $+$  satisfies:

1. (Closure):  $\vec{x} + \vec{y} \in V$ .
2. (Commutativity):  $\vec{x} + \vec{y} = \vec{y} + \vec{x}$
3. (Associativity):  $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$
4. (Identity Element): There exist an element  $\vec{0} \in V$  that satisfies:  $\vec{x} + \vec{0} = \vec{x}$ .
5. (Inverse Elements): For every  $\vec{x} \in V$ , there exists an element denoted  $-\vec{x}$  such that  $\vec{x} + (-\vec{x}) = \vec{0}$ .

**Scalar Multiplication:** The real numbers  $c$  and  $k$  are called **scalars**. The operations  $+$  and  $\cdot$  also satisfy:

6. (Closure:)  $c \cdot \vec{x} \in V$ .
7. (Distributivity:)  $c \cdot (\vec{x} + \vec{y}) = c \cdot \vec{x} + c \cdot \vec{y}$
8. (Distributivity:)  $(c + k) \cdot \vec{x} = c \cdot \vec{x} + k \cdot \vec{x}$
9. (Associativity:)  $c \cdot (k \cdot \vec{x}) = (ck) \cdot \vec{x}$
10. (Identity:) For every  $\vec{x} \in V$ ,  $1 \cdot \vec{x} = \vec{x}$ .

From its name, it's clear that sets of vectors are vector spaces. We will prove this soon. However, it is interesting that there are many other familiar sets of mathematical objects that are also vector spaces. We will now see these.

### 2.1.1 Examples of Vector Spaces

Let's begin with the set of all vectors of length 2, which we typically write as  $\mathbb{R}^2$ .

Show that  $\mathbb{R}^2$  together with the standard notion of vector addition and scalar multiplication is a vector space.

*Proof.* Let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \vec{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

and  $c, k \in \mathbb{R}$  as well as vector addition and scalar multiplication as

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix} \quad c \cdot \vec{x} = \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix}$$

We now show that all 10 properties of a vector space hold:

1. (Closure) Since

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

are elements of  $\mathbb{R}^2$ , the closure by addition holds.

2. (Commutativity) For all  $\vec{x}, \vec{y}, \vec{z} \in V$ ,

$$\begin{aligned} \vec{x} + \vec{y} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix} \\ &= \begin{bmatrix} y_1 + x_1 \\ y_2 + x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{y} + \vec{x} \end{aligned}$$

## 3. (Associativity)

$$\begin{aligned}
(\vec{x} + \vec{y}) + \vec{z} &= \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\
&= \begin{bmatrix} (x_1 + y_1) + z_1 \\ (x_2 + y_2) + z_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1 + (y_1 + z_1) \\ x_2 + (y_2 + z_2) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 + z_1 \\ y_2 + z_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \left( \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right) = \vec{x} + (\vec{y} + \vec{z})
\end{aligned}$$

4. (Identity Element) The zero element  $\vec{0}$  is

$$\vec{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and it satisfies,

$$\vec{0} + \vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 + x_1 \\ 0 + x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{x}$$

5. (Inverse Element) The inverse element  $-\vec{x}$  is given as

$$-\vec{x} = \begin{bmatrix} -x_1 \\ -x_2 \end{bmatrix}$$

and the following satisfies:

$$\vec{x} + (-\vec{x}) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -x_1 \\ -x_2 \end{bmatrix} = \begin{bmatrix} x_1 + (-x_1) \\ x_2 + (-x_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}$$

6. (Closure) For  $c \in \mathbb{R}$ ,

$$c \cdot \vec{x} = c \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix}$$

which is in  $\mathbb{R}^2$ , so scalar multiplication is closed.

7. (Distributivity) Let  $c \in \mathbb{R}$  and

$$\begin{aligned}
c \cdot (\vec{x} + \vec{y}) &= c \cdot \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) = c \cdot \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix} = \begin{bmatrix} c(x_1 + y_1) \\ c(x_2 + y_2) \end{bmatrix} \\
&= \begin{bmatrix} cx_1 + cy_1 \\ cx_2 + cy_2 \end{bmatrix} = \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix} + \begin{bmatrix} cy_1 \\ cy_2 \end{bmatrix} = c \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + c \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\
&= c \cdot \vec{x} + c \cdot \vec{y}
\end{aligned}$$

8. (Distributivity) Let  $c$  and  $k$  be elements of  $\mathbb{R}$ .

$$\begin{aligned}(c+k) \cdot \vec{x} &= (c+k) \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (c+k)x_1 \\ (c+k)x_2 \end{bmatrix} = \begin{bmatrix} cx_1 + kx_1 \\ cx_2 + kx_2 \end{bmatrix} \\ &= \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix} + \begin{bmatrix} kx_1 \\ kx_2 \end{bmatrix} = c \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + k \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = c \cdot \vec{x} + k \cdot \vec{x}\end{aligned}$$

9. (Associativity) Let  $c$  and  $k$  be elements of  $\mathbb{R}$ .

$$\begin{aligned}c \cdot (k \cdot \vec{x}) &= c \cdot \left( k \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = c \cdot \begin{bmatrix} kx_1 \\ kx_2 \end{bmatrix} = \begin{bmatrix} c(kx_1) \\ c(kx_2) \end{bmatrix} = \begin{bmatrix} (ck)x_1 \\ (ck)x_2 \end{bmatrix} \\ &= (ck) \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (ck) \cdot \vec{x}\end{aligned}$$

10. (Identity element)

$$1 \cdot \vec{x} = 1 \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{x}$$

□

The following are also vector spaces. There are not proofs associated with these.

- The set of vectors that consist of  $\mathbb{R}^3$  with the standard definition of vector addition and scalar multiplication.
- The set of vectors that consist of  $\mathbb{R}^n$  with the standard definition of vector addition and scalar multiplication.
- The set of all  $m \times n$  matrices,  $\mathcal{M}_{m \times n}$  with standard definition of matrix addition and scalar multiplication.

In the next few examples, we will also show some other interesting sets are vector spaces. To save a bunch of time and effort, in some cases, we don't need to show all 10 properties. We can skip most of properties, if the set  $V$  is a subset of a known vector space.

The reason for this is that properties 2–5 and 7–10 will automatically hold because the vectors are already in a vector space. We will do this formally in the next section.

Show that all lines that pass through the origin, or the set

$$V = \{ax + by = 0 \mid x, y \in \mathbb{R} \text{ and both } a \text{ and } b \text{ are not both } 0.\}$$

is a vector space.



First, let's rewrite the vector space as

$$V = \left\{ \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \mid au_1 + bu_2 = 0, \text{ with } a \text{ and } b \text{ both not } 0. \right\}$$

because this is the same set of points, but it is written in terms of a vector, which makes things easier to show.

As stated above, since  $V$  is a subset of  $\mathbb{R}^2$ , we don't need to prove properties 2–5, 7–10, which hold because we already showed that these properties show. Instead, to save time and effort, we only prove properties 1 and 6 of definition of a vector space in Definition ??.

1. Let  $\vec{u}, \vec{v} \in V$ . Therefore the two equations  $au_1 + bu_2 = 0$  and  $av_1 + bv_2 = 0$  are satisfied. Now check the sum,

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$$

and need to show that this is in  $V$ . Since

$$\begin{aligned} a(u_1 + v_1) + b(u_2 + v_2) &= au_1 + av_1 + bu_2 + bv_2 \\ &= (au_1 + bu_2) + (av_1 + bv_2) = 0 + 0 = 0 \end{aligned}$$

then  $\vec{u} + \vec{v} \in V$ .

6. We need to show that  $r \cdot \vec{u} \in V$ , so that the vector  $r \cdot \vec{u}$  satisfies the property that the line passes through the origin.

$$r \cdot \vec{u} = \begin{bmatrix} ru_1 \\ ru_2 \end{bmatrix}$$

and since  $au_1 + bu_2 = 0$ , then

$$r \cdot 0 = r(au_1 + bu_2) = a(ru_1) + b(ru_2) = 0$$

therefore  $r \cdot \vec{u} \in V$ .

So this is a vector space.

The next example shows that polynomials (specifically quadratics) are vector spaces. Since we haven't shown anything about polynomials yet, we need to show all 10 properties of Definition ?? hold.

Show that the set,  $\mathcal{P}_2$ , the set of all polynomials of degree 2 is a vector space with vector addition defined as the sum of two polynomials and scalar multiplication as the multiplication of a polynomial by a constant.

In this case scalars are elements in  $\mathbb{R}$  and the vector space  $V$  is the set of all polynomials of degree at most 2. Let

$$\begin{aligned} \vec{x} &= a_0 + a_1x + a_2x^2 \\ \vec{y} &= b_0 + b_1x + b_2x^2 \\ \vec{z} &= c_0 + c_1x + c_2x^2 \end{aligned}$$

Vector addition is simply polynomial addition for example:

$$\begin{aligned}\vec{x} + \vec{y} &= (a_0 + a_1x + a_2x^2) + (b_0 + b_1x + b_2x^2) \\ &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2.\end{aligned}$$

and scalar addition is given by

$$r\vec{x} = r(a_0 + a_1x + a_2x^2) = (ra_0) + (ra_1)x + (ra_2)x^2$$

1. (Closure) Show that  $\vec{x} + \vec{y}$  is in  $\mathcal{P}_2$ , the set of polynomials of degree 2.

$$\begin{aligned}\vec{x} + \vec{y} &= (a_0 + a_1x + a_2x^2) + (b_0 + b_1x + b_2x^2) \\ &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2\end{aligned}$$

which is a polynomial of degree 2, so this is in  $\mathcal{P}_2$ .

2. (Commutativity)

$$\begin{aligned}\vec{x} + \vec{y} &= (a_0 + a_1x + a_2x^2) + (b_0 + b_1x + b_2x^2) \\ &= (b_0 + b_1x + b_2x^2) + (a_0 + a_1x + a_2x^2) \\ &= \vec{y} + \vec{x}\end{aligned}$$

3. (Associativity)

$$\begin{aligned}(\vec{x} + \vec{y}) + \vec{z} &= (a_0 + a_1x + a_2x^2 + b_0 + b_1x + b_2x^2) \\ &\quad + (c_0 + c_1x + c_2x^2) \\ &= a_0 + a_1x + a_2x^2 \\ &\quad + (b_0 + b_1x + b_2x^2) + (c_0 + c_1x + c_2x^2) \\ &= \vec{x} + (\vec{y} + \vec{z})\end{aligned}$$

4. (Additive Identity) The zero element  $\vec{0}$  is the function 0.

$$\begin{aligned}\vec{0} + \vec{x} &= 0 + a_0 + a_1x + a_2x^2 \\ &= a_0 + a_1x + a_2x^2 = \vec{x}\end{aligned}$$

5. (Additive Inverse) The inverse element  $-\vec{x}$  is  $-a_0 - a_1x - a_2x^2$ , and

$$\begin{aligned}\vec{x} + (-\vec{x}) &= (a_0 + a_1x + a_2x^2) + (-a_0 - a_1x - a_2x^2) \\ &= (a_0 - a_0) + (a_1 - a_1)x + (a_2 - a_2)x^2 = 0 = \vec{0}.\end{aligned}$$

6. (Closure under scalar multiplication)

$$r\vec{x} = r(a_0 + a_1x + a_2x^2) = (ra_0) + (ra_1)x + (ra_2)x^2$$

since this is a polynomial of degree 2, then  $r\vec{x} \in \mathcal{P}_2$ .

7. (Distributivity) Let  $r, s \in \mathbb{R}$ ,

$$\begin{aligned}(r + s)\vec{x} &= (r + s)(a_0 + a_1x + a_2x^2) \\ &= r(a_0 + a_1x + a_2x^2) + s(a_0 + a_1x + a_2x^2) = r\vec{x} + s\vec{x}\end{aligned}$$

8. (Distributivity) Let  $r \in \mathbb{R}$ ,

$$\begin{aligned}r(\vec{x} + \vec{y}) &= r((a_0 + a_1x + a_2x^2) + (b_0 + b_1x + b_2x^2)) \\ &= r(a_0 + a_1x + a_2x^2) + r(b_0 + b_1x + b_2x^2) \\ &= r\vec{x} + r\vec{y}\end{aligned}$$

9. (Associativity) Let  $r, s \in \mathbb{R}$ ,

$$\begin{aligned}(rs)\vec{x} &= (rs)(a_0 + a_1x + a_2x^2) = r(s(a_0 + a_1x + a_2x^2)) \\ &= r(s\vec{x})\end{aligned}$$

10. (Multiplicative Identity)

$$1 \cdot \vec{x} = 1(a_0 + a_1x + a_2x^2) = a_0 + a_1x + a_2x^2 = \vec{x}.$$

The above example shows that there are examples of vectors and vector spaces that don't seem like vectors in  $\mathbb{R}^n$ . The following are also vector spaces:

- The set  $\{a \sin x + b \cos x \mid a, b \in \mathbb{R}\}$ .
- The set  $\{f \mid f : \mathbb{R} \rightarrow \mathbb{R}\}$  of functions of a single variable.

The “vector addition” is given by  $(f + g)(x) = f(x) + g(x)$  and the “scalar multiplication” is  $(r \cdot f)(x) = r \cdot f(x)$ .

We won't cover all of the properties of this set mainly because the real-valued function derive their properties from the reals, which is a vector space.

- The set of functions that satisfy  $f' + f = 0$ .
- Solutions of homogeneous differential equations, which is a generalization of functions that satisfy  $f' + f = 0$ .

### Is every set a vector space?

In short, no. For example, consider the following:

$$V = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1 \geq 0, x_2 \geq 0 \right\}$$

which is the first quadrant of the plane. To check if this is a vector space, we don't need to check all 10 properties. This is because since  $V$  is a subset of  $\mathbb{R}^2$ , most of the properties (commutative, associative, distributive, for examples) will work just as they did in  $\mathbb{R}^2$ . The only two that we do need to check are the closure properties.

1. If  $\vec{u}, \vec{v} \in V$ , then

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$$

and since  $u_1, u_2, v_1, v_2$  all satisfy  $\geq 0$ , then this vector is in  $V$ .

6. Let  $r \in \mathbb{R}$ , then

$$r\vec{u} = \begin{bmatrix} ru_1 \\ ru_2 \end{bmatrix}$$

but if  $r < 0$  then  $r\vec{u} \notin V$ , so  $V$  is not a vector space.

Is the set  $U = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1^2 + x_2^2 = 1 \right\}$  a vector space? If not what properties does it violate?

This set is a circle of radius one centered at the origin. Again, like above,  $U$  is a subset of  $\mathbb{R}^2$  and thus the properties to check are the closure properties.

1. If we let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

both be in  $U$ , then

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

Is this in  $U$ . To check sum the squares of the individual elements.

$$\begin{aligned} (x_1 + y_1)^2 + (x_2 + y_2)^2 &= x_1^2 + 2x_1y_1 + y_1^2 + x_2^2 + 2x_2y_2 + y_2^2 \\ &= 2 + 2x_1y_1 + 2x_2y_2 \end{aligned}$$

which is only 1 in special circumstances and not in general, so this is not a vector space.

4. The additive identity requires that the zero vector is an element of the vector space. However, since

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

is not in  $U$ , then this property is also not satisfied.

6. This property is also not satisfied for the same reason as #1. Geometrically if  $\vec{x}$  is in  $U$  then  $r\vec{x}$  is on a circle of radius  $r$ , so won't be in  $U$  in general.

## 2.2 The Span and Basis of a Subspace

In section ??, we saw the span of vectors in  $\mathbb{R}^n$ . We now extend this example to the span of any subspace. In addition, the notion of a basis of the subspace is introduced.

For any vector space,  $V$  a **subspace**,  $U$  is a subset of  $V$  that itself is a vector space.

We showed in example ?? that the set of all lines in  $\mathbb{R}^2$  that pass through the origin is a vector space. Since the set is a subset of  $\mathbb{R}^2$ , it is a subspace of  $\mathbb{R}^2$  as well.

Show that  $\mathbb{R}^2$  is a subspace of  $\mathbb{R}^3$ .

Since  $\mathbb{R}^2$  is itself a vector space and a subset of  $\mathbb{R}^3$ , then  $\mathbb{R}^2$  is a subspace.

Recall that the set

$$\mathcal{P}_2 = \{a_0 + a_1x + a_2x^2 \mid a_0, a_1, a_2 \in \mathbb{R}\}$$

is the set of all quadratic functions.

The set  $\mathcal{P}_1 = \{a_0 + a_1x \mid a_0, a_1 \in \mathbb{R}\}$  of all linear functions is itself a vector space as well as a subset of  $\mathcal{P}_2$ , therefore  $\mathcal{P}_1$  is a subspace of  $\mathcal{P}_2$ .

In addition, the set  $\{ax^2 \mid a \in \mathbb{R}\}$  is a vector space as well as a subset of  $\mathcal{P}_2$ , therefore it is a subspace.

The above examples show that there are many already known subspaces. There are many cases though that aren't evident or to show it is a subspace, we would need to prove all 10 properties that it is a vector space. The next lemma, however, shows that isn't the case.

Let  $S$  be a nonempty subset  $S$  of a vector space  $V$ , under the inherited operations. If for all  $\vec{s}_1$  and  $\vec{s}_2$  in  $S$  and  $r_1, r_2 \in \mathbb{R}$

$$r_1\vec{s}_1 + r_2\vec{s}_2 \in S$$

then  $S$  is a subspace.

This means that if  $S$  is a subset of  $V$ , that we know is a vector space, to prove that  $S$  is a subspace, we only need to check if  $r_1\vec{s}_1 + r_2\vec{s}_2 \in S$ .

*Proof.* Since  $S$  is a subspace of  $V$ , properties (2)–(5) and (7)–(10) of Definition ?? hold for  $S$ . Thus we only need to prove closure under addition and scalar multiplication.

Property 1: Because  $r_1\vec{s}_1 + r_2\vec{s}_2 \in S$ , let  $r_1 = r_2 = 1$ , thus  $\vec{s}_1 + \vec{s}_2 \in S$ .

Property 6: Because  $r_1\vec{s}_1 + r_2\vec{s}_2 \in S$ , let  $r_2 = 0$ , thus  $r_1\vec{s}_1 \in S$ .

□

Show that

$$V = \left\{ \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \mid v_2 = kv_1 \right\}$$

(that is, all vectors on a line of slope  $k$ ) is a subspace of  $\mathbb{R}^2$ .

We will use lemma ??. Let

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

be elements of  $V$ . That is  $v_2 = kv_1$  and  $u_2 = ku_1$ . Then

$$\begin{aligned} r_1\vec{u} + r_2\vec{v} &= r_1 \begin{bmatrix} u_1 \\ ku_1 \end{bmatrix} + r_2 \begin{bmatrix} v_1 \\ kv_1 \end{bmatrix} = \begin{bmatrix} r_1u_1 + r_2v_1 \\ r_1ku_1 + r_2kv_1 \end{bmatrix} \\ &= \begin{bmatrix} r_1u_1 + r_2v_1 \\ k(r_1u_1 + r_2v_1) \end{bmatrix} \end{aligned}$$

which is an element of  $V$  because the second component is  $k$  times the first one. Thus  $V$  is a subspace of  $\mathbb{R}^2$ .

Show using Lemma ?? that

$$V = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \mid a, b \in \mathbb{R} \right\}$$

the set of all diagonal matrices is a subspace of  $\mathcal{M}_{2 \times 2}$ .

In this case, if we show that for any two matrices

$$A = \begin{bmatrix} a_1 & 0 \\ 0 & b_1 \end{bmatrix}, \quad B = \begin{bmatrix} a_2 & 0 \\ 0 & b_2 \end{bmatrix}$$

and scalars  $r_1, r_2 \in \mathbb{R}$  that  $r_1A + r_2B$  is in the set.

$$\begin{aligned} r_1A + r_2B &= r_1 \begin{bmatrix} a_1 & 0 \\ 0 & b_1 \end{bmatrix} + r_2 \begin{bmatrix} a_2 & 0 \\ 0 & b_2 \end{bmatrix} \\ &= \begin{bmatrix} r_1a_1 + r_2a_2 & 0 \\ 0 & r_1b_1 + r_2b_2 \end{bmatrix} \end{aligned}$$

which is a diagonal matrix, therefore in  $V$ , thus this is a subspace.

Let  $A$  be an  $m$  by  $n$  matrix. The null space of  $A$  is a subspace of  $\mathbb{R}^n$ .

*Proof.* We will use lemma ?? to solve this. Let both  $\vec{x}$  and  $\vec{y}$  be in the null space of  $A$ . This means that  $A\vec{x} = \vec{0}$  and  $A\vec{y} = \vec{0}$ . We need to show that  $r_1\vec{x} + r_2\vec{y}$  is in the null space of  $A$ .

$$\begin{aligned} A(r_1\vec{x} + r_2\vec{y}) &= r_1A\vec{x} + r_2A\vec{y} \\ &= r_1(0) + r_2(0) = 0 \end{aligned}$$

Vectors in the null space are vectors of length  $n$ , so the null space is a subset of  $\mathbb{R}^n$  and since  $r_1\vec{x} + r_2\vec{y}$  is in the null space of  $A$ , then the null space is a subspace of  $\mathbb{R}^n$ . □

This is an important result that we will see in eigenvalues in the next chapter.

### 2.2.1 The Span of a set of vectors

We saw in Section ?? the span of a set of vectors in  $\mathbb{R}^n$ . We now generalize this to any vector space.

The **span** of a nonempty set subset  $S$  of a vector space is the set of all linear combinations of the vectors in  $S$ .

$$\text{span}(S) = \{c_1\vec{s}_1 + c_2\vec{s}_2 + \cdots + c_n\vec{s}_n \mid c_1, c_2, \dots, c_n \in \mathbb{R}, \vec{s}_1, \vec{s}_2, \dots, \vec{s}_n \in S\}.$$

To show that a subset of vectors span a subspace  $S$ , we need to show that any vector in  $S$  can be written as a linear combination of the spanning vectors.

Show that the set  $\{2+x, 1, x+x^2\}$  spans  $\mathcal{P}_2$ .

In this case, we need show that a general polynomial in  $\mathcal{P}_2$  can be written as a linear combination of elements of the given set. That is

$$c_1(2+x) + c_2(1) + c_3(x+x^2) = a_0 + a_1x + a_2x^2$$

and if there is a solution for the  $c$ 's, then that shows the the set spans  $\mathcal{P}_2$ . To find the solution, use the technique of equating coefficients. Write down the coefficients for the constant terms,  $x$  terms and  $x^2$  terms respectively.

$$2c_1 + c_2 = a_0$$

$$c_1 + c_3 = a_1$$

$$c_3 = a_2$$

This has a solution  $c_3 = a_2, c_1 = a_1 - a_2$  and  $c_2 = a_0 - 2(a_1 - a_2)$ , so the given set spans  $\mathcal{P}_2$ .

The span of any subset of a vector space is a subspace.

*Proof.* Let  $S$  be the subset and  $\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n$  be the elements of  $S$ . Using Lemma ??, we need to check that  $\text{span}(S)$  is closed under linear combinations. Let

$$\vec{v} = c_1\vec{s}_1 + c_2\vec{s}_2 + \cdots + c_n\vec{s}_n,$$

$$\vec{w} = k_1\vec{s}_1 + k_2\vec{s}_2 + \cdots + k_n\vec{s}_n$$

then

$$\begin{aligned} r_1\vec{v} + r_2\vec{w} &= r_1(c_1\vec{s}_1 + c_2\vec{s}_2 + \cdots + c_n\vec{s}_n) + r_2(k_1\vec{s}_1 + k_2\vec{s}_2 + \cdots + k_n\vec{s}_n) \\ &= (r_1c_1 + r_2k_1)\vec{s}_1 + (r_1c_2 + r_2k_2)\vec{s}_2 + \cdots + (r_1c_n + r_2k_n)\vec{s}_n \end{aligned}$$

Since this shows that  $r_1\vec{v} + r_2\vec{w}$  is in  $S$ , then  $S$  is a subspace. □

This lemma allows us to talk about a vector space in terms of the vectors that span it. For example, instead of thinking of  $\mathcal{P}_2$ , we think of the span of  $\{2+x, 1, x+x^2\}$  (in this case, it may not be more helpful, but other cases it is).

Show that the following vectors span  $\mathbb{R}^3$ :

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Because the vector  $\vec{x} = [x_1, x_2, x_3]^T$  can be written  $\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3$ , then these vectors span  $\mathbb{R}^3$ .

Does  $\{2 + x, x^2\}$  span  $\mathcal{P}_2$ ?

To determine this, we will try write a general polynomial in  $\mathcal{P}_2$ ,

$$a_0 + a_1x + a_2x^2$$

in term of the set or

$$a_0 + a_1x + a_2x^2 = c_1(2 + x) + c_2x^2$$

and equating coefficients,

$$a_0 = 2c_1$$

$$a_1 = c_1$$

$$a_2 = c_2$$

There's no solution to this because  $c_1$  can't simultaneously equal  $a_1$  and  $a_0/2$ , so  $\{2 + x, x^2\}$  does not span  $\mathcal{P}_2$ .

### 2.2.2 The Basis and Dimension of a Vector Space

The **basis** of a vector space is a sequence of vectors of the vector space that form a linearly independent set that spans the vector space.

Note: the basis will be a sequence of vectors because the order of the vectors will be important. We will denote the sequence with parentheses,  $(\vec{s}_1, \vec{s}_2, \dots)$ .

We showed in Example ?? that the set  $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$  span  $\mathbb{R}^2$  and since the second is not a multiple of the first, they are linearly independent. Therefore the sequence

$$\left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$$

form a basis for  $\mathbb{R}^2$ .

The sequence of vectors

$$\left( \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)$$

is a different basis of  $\mathbb{R}^2$  because of the order of vectors. The fact that they span  $\mathbb{R}^2$  and are linearly independent do not depend on the order.



There are many bases of a vector space. For example,

$$\left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)$$

also spans  $\mathbb{R}^2$  and are linearly independent.

Does

$$\left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right)$$

form a basis of  $\mathbb{R}^2$ ?

These three vectors are not linearly independent. Although one can show this in general, note that

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

and since they are not linearly independent, then they cannot form a basis of  $\mathbb{R}^2$ .

Although there are lots of different bases for a given subspace, there are some that are more useful than others. There is a basis for more subspaces called a standard basis.

The sequence

$$\mathcal{E}_n = \left( \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right)$$

is called the **standard basis** or **natural basis** of  $\mathbb{R}^n$ . The vectors in the basis are called  $\vec{e}_1, \vec{e}_2, \dots$

The natural basis of  $\mathcal{P}_3$  is  $(1, x, x^2, x^3)$ .

We saw bases of vector spaces (or subspaces) at the beginning of this course. In Example ??, we solved a linear system. It's associated homogeneous system is

$$\begin{aligned} x_2 + 3x_3 - 9x_4 + 11x_5 &= 0, \\ 2x_3 + 4x_5 &= 0, \\ 3x_5 &= 0, \end{aligned}$$

The solution (which is a subspace of  $\mathbb{R}^5$ ) can be written as

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 9 \\ 0 \\ 1 \\ 0 \end{bmatrix} x_4 \mid x_1, x_4 \in \mathbb{R}, \right\}$$

The two vectors in the solution are a basis of the solution space. Since there are only two vectors and they are not constant multiples of each other, it's easy to see that they are linearly independent. Also because of the form of the solution set, you can also see that that span the space.

In a vector space with basis  $B$ , the **representation of a vector  $\vec{v}$  with respect to the basis  $B$**  is the column vector of the coefficients used to express  $\vec{v}$  as a linear combination of the basis vectors:

$$\text{Rep}_B(\vec{v}) = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

where  $B = (\vec{\beta}_1, \vec{\beta}_2, \dots, \vec{\beta}_n)$  and

$$\vec{v} = c_1\vec{\beta}_1 + c_2\vec{\beta}_2 + \dots + c_n\vec{\beta}_n$$

Consider the space  $\mathcal{P}_2$ , the space of quadratic functions. Let  $B = (1, 1+x, 1+x+x^2,)$  be a basis of  $\mathcal{P}_2$  and  $\vec{v} = 2x + x^2$ . To find the representation, we need to find  $c_1, c_2$  and  $c_3$  such that

$$c_1 \cdot 1 + c_2 \cdot (1+x) + c_3 \cdot (1+x+x^2) = 2x + x^2$$

by equating coefficients this is same as solving the linear system:

$$c_1 + c_2 + c_3 = 0$$

$$c_2 + c_3 = 2$$

$$c_3 = 1$$

resulting in  $c_1 = -2, c_2 = 1, c_3 = 1$ , therefore

$$\text{Rep}_B(\vec{v}) = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

If instead the basis is given as  $D = (2, 2x, x^2)$ , then

$$\text{Rep}_D(\vec{v}) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

### Representations in the natural basis

As we saw above, finding representations in a basis requires solving another linear system. However, representations in the natural basis are simple calculations. If we used the natural basis  $E = (1, x, x^2)$  for the quadratic example above, then

$$\text{Rep}_E(\vec{v}) = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

are just the coefficients of  $x^n$  terms of the vector  $2x + x^2$ . The following example shows that the representation of a vector in  $\mathbb{R}^3$  is what we expect, itself.

Find the Representation of the vector

$$\vec{v} = \begin{bmatrix} -3 \\ 2 \\ 4 \end{bmatrix}$$

in the natural basis

$$\mathcal{E}_3 = \left( \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right)$$

We seek the vector  $\vec{c} = [c_1, c_2, c_3]^\top$  such that

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \\ 4 \end{bmatrix}$$

which is just that  $c_1 = -3$ ,  $c_2 = 2$ , and  $c_3 = 4$  so the representation of the vector in the basis  $\mathcal{E}_3$  is

$$\text{Rep}_{\mathcal{E}}(\vec{v}) = \begin{bmatrix} -3 \\ 2 \\ 4 \end{bmatrix}$$

which is just the original vector.

The last example in this section uses matrices. The natural basis for  $\mathcal{M}_{2 \times 2}$  is

$$B = \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

Find

$$\text{Rep}_B \left( \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \right)$$

Formally, one needs to find  $c_1, c_2, c_3$  and  $c_4$  such that

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

but since the nice structure of the basis  $c_1 = 1$ ,  $c_2 = 2$ ,  $c_3 = 3$ , and  $c_4 = 4$ , so

$$\text{Rep}_B\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

One can generalize to show that

$$\text{Rep}_B\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

and this shows that matrices (which are vectors in the formal sense of vector spaces) can be represented by vectors by reshaping the matrix as a vector.

### 2.2.3 Dimension

We have been talking about a few big topics in this chapter. One of those is the spanning set of a vector space. We noted that many different sets can span a vector space. This brought in the notation of linear independence and a basis. However for a vector space there can be many different bases.

Although we did introduce a natural basis, this works well for some spaces, like  $\mathcal{P}_2$  and  $\mathbb{R}^3$ , however what is the natural basis for a solution of homogeneous linear system.

Perhaps if two people argue over the basis of a vector space, one thing they will agree on is the number of vectors in a basis as we will see. We noted earlier that disregarding extra vectors is generally a good thing to result in a basis, but there is a unique thing about bases and that is the number of vectors in any basis.

A vector space is **finite dimensional** if it has a basis with only finitely-many vectors.

In any finite-dimensional vector space, all of the bases have the same number of elements.

Because of this theorem, we define the dimension in following manner.

The **dimension** of a finite dimensional vector space is the number of vectors in any of its bases.

- The dimension of  $\mathbb{R}^n$  is  $n$ . Although there are many bases, consider  $\mathcal{E}_n$ , the natural basis, which has  $n$  elements.
- The dimension of  $\mathcal{P}_n$  is  $n + 1$ . The natural basis of  $\mathcal{P}_n$  is  $(1, x, x^2, \dots, x^n)$  with  $n + 1$  elements.
- The dimension of the vector space of all 2 by 2 matrices is 4. A natural basis for this is:

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

and since there are 4 elements, the dimension is 4.

### Bases of Subspaces

There were a number of important ideas in this section, so a summary is necessary. The basis of a space or subspace is useful for writing down elements in the space. That is, if we know the basis, then we know what's in the space. Additionally, the representation of an element are the coefficients in terms of the basis.

This means that any vector in a finite-dimensional vector space can be represented as a vector and as we will see this will be helpful in that we can use many of the nice techniques from chapter 1 to help. We will start to see that since we can write any polynomial as a vector, that many operations that we do to polynomials (such as multiplication, differentiation, and integration) can be done using matrices and vectors.

## 2.3 Inner Product Spaces

In section ??, we examined the dot product between two vectors. Recall that we can write the dot product between two vectors  $\vec{u}$  and  $\vec{v}$  in  $\mathbb{R}^n$  is

$$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

We used the dot product for finding the angle between two vectors and we'll show that there are many other applications that the dot product later in the text. In this section, we extend that to any vector space and call this an inner product space.

The vector space  $V$  is called an **inner product space** if for every pair of vectors  $\vec{u}$  and  $\vec{v}$  in  $V$  there is a unique number  $\langle \vec{u}, \vec{v} \rangle$ , called the **inner product** of  $\vec{u}$  and  $\vec{v}$  such that for all  $\vec{u}, \vec{v}$  and  $\vec{w}$  in  $V$  and  $a, b \in \mathbb{R}$  then the following properties are satisfied:

1.  $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$
2.  $\langle a\vec{u} + b\vec{v}, \vec{w} \rangle = a\langle \vec{u}, \vec{w} \rangle + b\langle \vec{v}, \vec{w} \rangle$ .
3.  $\langle \vec{u}, \vec{u} \rangle \geq 0$  and equals 0 if and only if  $\vec{u} = \vec{0}$ .

Show that  $\mathbb{R}^n$  is an inner product space with

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

and  $\langle \vec{u}, \vec{v} \rangle = \vec{u}^T \vec{v}$ .

First, the set  $\mathbb{R}^n$  is a vector space. Next, we need to show that the three properties are satisfied for  $\langle \vec{u}, \vec{v} \rangle = \vec{u}^T \vec{v}$ .

1.

$$\langle \vec{u}, \vec{v} \rangle = \vec{u}^T \vec{v}$$

since the transpose of a number is itself

$$= (\vec{u}^T \vec{v})^T$$

using the properties of transposes in Section ??.

$$= \vec{v}^T (\vec{u}^T)^T = \vec{v}^T \vec{u} = \langle \vec{v}, \vec{u} \rangle.$$

so the first property is satisfied.

2.

$$\begin{aligned} \langle a\vec{u} + b\vec{v}, \vec{w} \rangle &= (a\vec{u} + b\vec{v})^T \vec{w} = (a\vec{u}^T + b\vec{v}^T) \vec{w} \\ &= a\vec{u}^T \vec{w} + b\vec{v}^T \vec{w} = a\langle \vec{u}, \vec{w} \rangle + b\langle \vec{v}, \vec{w} \rangle. \end{aligned}$$

3.

$$\langle \vec{u}, \vec{u} \rangle = \vec{u}^T \vec{u} = u_1^2 + u_2^2 + u_3^2 + \cdots + u_n^2$$

which satisfies  $\geq 0$ . The only time that this quantity equals 0, is when  $u_1 = u_2 = \cdots = u_n$  or  $\vec{u} = \vec{0}$ .

The next example shows that the notion of an inner product is not limited to the dot product of vectors in  $\mathbb{R}^n$ .

Show that  $\mathcal{P}_n$  on the interval  $[0, 1]$  is a inner product space with

$$\langle p, q \rangle = \int_0^1 pq \, dx$$

for any  $p, q \in \mathcal{P}_n$ .

We have seen above that  $\mathcal{P}_n$  is a vector space and limiting this to the interval  $[0, 1]$  does not change that fact. We need to show that the inner product defined as an integral satisfies the three properties of inner products.

1.

$$\langle p, q \rangle = \int_0^1 pq \, dx = \int_0^1 qp \, dx = \langle q, p \rangle.$$

2. Let  $r$  also be in  $\mathcal{P}_n$  and  $a, b \in \mathbb{R}$ .

$$\begin{aligned} \langle ap + bq, r \rangle &= \int_0^1 (ap + bq)r \, dx = \int_0^1 (apr + bqr) \, dx \\ &= a \int_0^1 pr \, dx + b \int_0^1 qr \, dx = a\langle p, r \rangle + b\langle q, r \rangle \end{aligned}$$

3.

$$\langle p, p \rangle = \int_0^1 p^2 \, dx$$

and this is greater than or equal to zero and can be shown that equals 0 if and only if  $p(x) \equiv 0$ .

### 2.3.1 Vector Norms and distance

Let  $\vec{u}$  be an element of an inner product space. The norm of the vector  $\vec{u}$  is given by

$$\|\vec{u}\| = \sqrt{\langle \vec{u}, \vec{u} \rangle}$$

Note: if  $\vec{u} \in \mathbb{R}^n$  then the vector norm is the length of the vector. In addition, the notion of norm of a vector is also often generalized and the norm above is generally called the 2-norm.

The distance between vectors  $\vec{u}$  and  $\vec{v}$  denoted  $d(\vec{u}, \vec{v})$  and is defined as

$$d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|.$$

Note: if  $\vec{u}$  and  $\vec{v}$  are in  $\mathbb{R}^n$ , then the distance function is the standard distance function where the vectors are considered to be points.

The distance  $\|\vec{u} - \vec{v}\| = 0$  if and only if  $\vec{u} = \vec{v}$ .

*Proof.*  $\implies$  Let  $\|\vec{u} - \vec{v}\| = 0$ , then  $\langle \vec{u} - \vec{v}, \vec{u} - \vec{v} \rangle = 0$  By definition, this is only zero if  $\vec{u} - \vec{v} = \vec{0}$ , therefore  $\vec{u} = \vec{v}$ .

$\impliedby$  If  $\vec{u} = \vec{v}$ , then  $\|\vec{u} - \vec{v}\| = \|\vec{0}\| = 0$ . □

Except for the example of the inner product as an integral, all of the other discussion in this section has been about vectors in  $\mathbb{R}^n$ , and this is mainly because it is the canonical vector and inner product space and some of the interpretations of length of vectors don't make a lot of sense when talking about polynomials or other functions.

However, distance does have some interpretation with functions. Lemma ?? can be used to show when two functions are equal (at least on the interval of the integration), but can also be useful to determine when they are close to equal (when the integrals are close to one another). This will be useful in future chapters.

### 2.3.2 Angles between vectors

If  $\vec{u}$  and  $\vec{v}$  are vectors in  $\mathbb{R}^2$  then we saw in section ?? that

$$\cos \theta = \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \|\vec{v}\|}$$

and thus the angle between the vectors can be found.

This notion generalizes to any vectors in an inner product space,  $V$ . Most helpful, two vectors meet at a right angle if  $\langle \vec{u}, \vec{v} \rangle = 0$  and it is said that if this holds then  $\vec{u}$  and  $\vec{v}$  are orthogonal. The next section talks about how an entire set of vectors can be orthogonal.

### 2.3.3 Orthonormal sets of vectors

Let  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  each be elements of an inner product space,  $V$ . The set is called an **orthonormal set** if

$$\langle \vec{v}_i, \vec{v}_j \rangle = 0 \quad \text{if } i \neq j$$

and

$$\|\vec{v}_i\| = \sqrt{\langle \vec{v}_i, \vec{v}_i \rangle} = 1$$

If only the first condition holds, the set is called **orthogonal**.

Show that

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

is an orthonormal set where each vector is an element of  $\mathbb{R}^3$ .

We need to show that the three inner products between the vectors are zero

$$\begin{aligned} \langle \vec{v}_1, \vec{v}_2 \rangle &= (1)(0) + (0)(1/\sqrt{2}) + (0)(1/\sqrt{2}) = 0 \\ \langle \vec{v}_1, \vec{v}_3 \rangle &= (1)(0) + (0)(1/\sqrt{2}) + (0)(-1/\sqrt{2}) = 0 \\ \langle \vec{v}_2, \vec{v}_3 \rangle &= (0)(0) + (1/\sqrt{2})(1/\sqrt{2}) + (1/\sqrt{2})(-1/\sqrt{2}) = \frac{1}{2} - \frac{1}{2} = 0 \end{aligned}$$

and that the norms of each vector is 1.

$$\begin{aligned} \|\vec{v}_1\| &= \sqrt{(1)^2 + (0)^2 + (0)^2} = 1 \\ \|\vec{v}_2\| &= \sqrt{(0)^2 + (1/\sqrt{2})^2 + (1/\sqrt{2})^2} = \sqrt{0 + \frac{1}{2} + \frac{1}{2}} = 1 \\ \|\vec{v}_3\| &= \sqrt{(0)^2 + (1/\sqrt{2})^2 + (-1/\sqrt{2})^2} = \sqrt{0 + \frac{1}{2} + \frac{1}{2}} = 1 \end{aligned}$$

Note: this is also a basis for  $\mathbb{R}^3$ .

Show that

$$\{\sin x, \sin 2x, \sin 3x\}$$

form an orthogonal set on  $[-\pi, \pi]$  with the inner product taken to be the definite integral on  $[-\pi, \pi]$ .

Note: the following integrals will come up in the next few chapters a number of times and the technique shown below is helpful for such integrals. This also shows that even though in most cases in this text so far, that using complex numbers helps out tremendously.



Recall that

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i} \qquad \cos x = \frac{e^{ix} + e^{-ix}}{2}$$

and using this we find the integrals of all pairs of functions on  $[-\pi, \pi]$

$$\begin{aligned} \langle \sin x, \sin 2x \rangle &= \int_{-\pi}^{\pi} \sin x \sin 2x \, dx = \int_{-\pi}^{\pi} \frac{e^{ix} - e^{-ix}}{2i} \frac{e^{2ix} - e^{-2ix}}{2i} \, dx \\ &= \frac{1}{-4} \int_{-\pi}^{\pi} (e^{3ix} - e^{ix} - e^{-ix} + e^{-3ix}) \, dx \\ &= -\frac{1}{4} \left( \frac{e^{3ix}}{3i} - \frac{e^{ix}}{i} - \frac{e^{-ix}}{-i} + \frac{e^{-3ix}}{-3i} \right) \Big|_{-\pi}^{\pi} \\ &= -\frac{1}{4} \left( \frac{e^{3i\pi}}{3i} - \frac{e^{i\pi}}{i} - \frac{e^{-i\pi}}{-i} + \frac{e^{-3i\pi}}{-3i} \right) \\ &\quad + \frac{1}{4} \left( \frac{e^{-3i\pi}}{3i} - \frac{e^{-i\pi}}{i} - \frac{e^{i\pi}}{-i} + \frac{e^{3i\pi}}{-3i} \right) = 0 \end{aligned}$$

and as it can be seen that all term made with the substitution  $x = \pi$  cancel with terms made with the substitution  $x = -\pi$ .

The other two inner products are similar. In fact it can be shown (and we will later) that

$$\langle \sin kx, \sin \ell x \rangle = \int_{-\pi}^{\pi} \sin kx \sin \ell x \, dx = 0$$

if  $k \neq \ell$ .

We saw the following theorem applied to vectors in  $\mathbb{R}^n$  in section ???. It was a consequence of the triangle inequality. We again generalize this to the inner product.

[Cauchy-Swartz Inequality] If  $\vec{u}$  and  $\vec{v}$  are elements of an inner product space then

$$|\langle \vec{u}, \vec{v} \rangle| \leq \|\vec{u}\| \|\vec{v}\|$$

*Proof.* If  $\vec{v} = \vec{0}$ , then the inequality is satisfied. Assume  $\vec{v} \neq \vec{0}$  and define

$$\lambda = \langle \vec{v}, \vec{v} \rangle^{-1} \langle \vec{u}, \vec{v} \rangle.$$

By definition of the inner product,

$$0 \leq \langle \vec{u} - \lambda \vec{v}, \vec{u} - \lambda \vec{v} \rangle$$

Using properties of the inner product,

$$\begin{aligned} 0 &\leq \langle \vec{u}, \vec{u} \rangle - 2\lambda \langle \vec{u}, \vec{v} \rangle + \lambda^2 \langle \vec{v}, \vec{v} \rangle \\ &= \langle \vec{u}, \vec{u} \rangle - 2\langle \vec{v}, \vec{v} \rangle^{-1} \langle \vec{u}, \vec{v} \rangle \langle \vec{u}, \vec{v} \rangle + (\langle \vec{v}, \vec{v} \rangle^{-1} \langle \vec{u}, \vec{v} \rangle)^2 \langle \vec{v}, \vec{v} \rangle. \\ &= \langle \vec{u}, \vec{u} \rangle - \langle \vec{v}, \vec{v} \rangle^{-1} \langle \vec{u}, \vec{v} \rangle^2 \end{aligned}$$

Multiply through by  $\langle \vec{v}, \vec{v} \rangle$

$$= \langle \vec{u}, \vec{u} \rangle \langle \vec{v}, \vec{v} \rangle - \langle \vec{u}, \vec{v} \rangle^2$$

This can be rearranged to get:

$$\langle \vec{u}, \vec{v} \rangle^2 \leq \langle \vec{u}, \vec{u} \rangle \langle \vec{v}, \vec{v} \rangle$$

and taking the square root, you get the desired result.  $\square$

The Cauchy-Swartz inequality has numerous uses (which is not unexpected for a theorem named after such famous mathematicians). One such is to show that the angle between two vectors in  $\mathbb{R}^n$  and actually in any inner product space is defined.

Recall that above in this section, the angle between two vectors was defined as the angle  $\theta$  that satisfies

$$\cos \theta = \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \|\vec{v}\|}$$

and solving for the inner product,

$$\langle \vec{u}, \vec{v} \rangle = \|\vec{u}\| \|\vec{v}\| \cos \theta$$

and using the Cauchy-Swartz Inequality,

$$|\langle \vec{u}, \vec{v} \rangle| = \|\vec{u}\| \|\vec{v}\| |\cos \theta| \leq \|\vec{u}\| \|\vec{v}\|$$

which results in

$$|\cos \theta| \leq 1$$

which is defined for all  $\theta$ .

### 2.3.4 The Gram-Schmidt Orthogonalization Process

We saw above that an orthonormal set of vectors has the property that any two different vectors in the set are orthogonal (inner product of zero) and have norm of one. Often, if we have a set of vectors (say a basis), they aren't orthonormal, but it is advantageous to have another set that is. The **Gram-Schmidt** algorithm creates an orthonormal set of vectors from a set of linearly independent vectors.

We first see the idea from a pair of vectors in the following example.

Find a orthonormal set of vectors that span the same set as  $\{\vec{u}_1, \vec{u}_2\}$ , where

$$\vec{u}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

There are many ways of doing this, but the following will always work. First, let's start by finding a unit vector in the same direction as  $\vec{u}_1$ . We denote this as  $\hat{v}_1$  and this can be found by dividing  $\vec{u}_1$  by its length:

$$\hat{v}_1 = \frac{1}{\|\vec{u}_1\|} \vec{u}_1 = \frac{1}{2} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Next, we will form a new vector  $\vec{v}_2$  which is a linear combination of  $\hat{v}_1$  and  $\vec{u}_2$  and has the property that it is orthogonal to  $\vec{v}_1$

$$\vec{v}_2 = c_1 \hat{v}_1 + c_2 \vec{u}_2$$

and we can take  $c_1 = 1$  without loss of generality. with

$$\begin{aligned} 0 &= \langle \vec{v}_2, \hat{v}_1 \rangle \\ &= \langle \hat{v}_1 + c_2 \vec{u}_2, \hat{v}_1 \rangle \\ &= \langle \hat{v}_1, \hat{v}_1 \rangle + c_2 \langle \vec{u}_2, \hat{v}_1 \rangle \end{aligned}$$

and solving for  $c_2$ ,

$$c_2 = -\frac{\langle \vec{u}_2, \hat{v}_1 \rangle}{\langle \hat{v}_1, \hat{v}_1 \rangle}$$

and since  $\vec{v}_1$  is a unit vector

$$= -\langle \vec{u}_2, \hat{v}_1 \rangle = -1$$

So

$$\vec{v}_2 = \hat{v}_1 - \vec{u}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Now since  $\vec{v}_2$  is already a unit vector, normalization is not needed. If it were, then dividing by its length would create a unit vector. A orthonormal set that spans the set is

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

If there are more than two vectors in the set, then the process is similar in that iteratively, one creates a new vector that is orthogonal to all previous vectors and then normalizes each one.

The **Gram-Schmidt Orthogonalization** algorithm is the following. Consider a set of vectors  $\{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n\}$  which is the basis for some vector space.

1. Let  $\vec{v}_1 = \vec{u}_1$ .

2. Let  $\vec{v}_2 = \vec{u}_2 - \frac{\langle \vec{v}_1, \vec{u}_2 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1$
3. Let  $\vec{v}_3 = \vec{u}_3 - \frac{\langle \vec{v}_2, \vec{u}_3 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2 - \frac{\langle \vec{v}_1, \vec{u}_3 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1$ .
4. And so on with

$$\begin{aligned} \vec{v}_k &= \vec{u}_k - \frac{\langle \vec{v}_{k-1}, \vec{u}_k \rangle}{\langle \vec{v}_{k-1}, \vec{v}_{k-1} \rangle} \vec{v}_{k-1} - \frac{\langle \vec{v}_{k-2}, \vec{u}_k \rangle}{\langle \vec{v}_{k-2}, \vec{v}_{k-2} \rangle} \vec{v}_{k-2} \\ &\quad \dots - \frac{\langle \vec{v}_2, \vec{u}_k \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2 - \frac{\langle \vec{v}_1, \vec{u}_k \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 \\ &= \vec{u}_k - \sum_{i=1}^{k-1} \frac{\langle \vec{v}_i, \vec{u}_k \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \vec{v}_i \end{aligned}$$

The resulting set of vectors  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  are orthogonal. If the set also needs to be orthonormal, for each vector  $\vec{v}_k$  divide through by its length.

*Proof.* To prove Gram-Schmidt orthogonalization works, we will prove that  $\vec{v}_k$  is orthogonal to  $\vec{v}_i$  for all  $i < k$ . We will show this by induction. First show that  $\vec{v}_2$  is orthogonal to  $\vec{v}_1$ :

$$\begin{aligned} \langle \vec{v}_2, \vec{v}_1 \rangle &= \left\langle \vec{u}_2 - \frac{\langle \vec{v}_1, \vec{u}_2 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1, \vec{v}_1 \right\rangle \\ &= \langle \vec{u}_2, \vec{v}_1 \rangle - \frac{\langle \vec{v}_1, \vec{u}_2 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \langle \vec{v}_1, \vec{v}_1 \rangle \\ &= \langle \vec{u}_2, \vec{v}_1 \rangle - \langle \vec{v}_1, \vec{u}_2 \rangle = 0 \end{aligned}$$

Now assume that  $\vec{v}_k$  is orthogonal to all other  $\vec{v}$ 's or  $\langle \vec{v}_k, \vec{v}_i \rangle = 0$  for  $i = k-1, k-2, \dots, 2, 1$ . Now we will prove that  $\vec{v}_{k+1}$ , given by Gram-Schmidt is orthogonal to all other  $\vec{v}$ 's. For all  $j = 1, 2, \dots, k$ ,

$$\begin{aligned} \langle \vec{v}_{k+1}, \vec{v}_j \rangle &= \left\langle \vec{u}_{k+1} - \sum_{i=1}^k \frac{\langle \vec{v}_i, \vec{u}_{k+1} \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \vec{v}_i, \vec{v}_j \right\rangle, \\ &= \langle \vec{u}_{k+1}, \vec{v}_j \rangle - \sum_{i=1}^k \frac{\langle \vec{v}_i, \vec{u}_{k+1} \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \langle \vec{v}_i, \vec{v}_j \rangle \end{aligned}$$

The only term that is nonzero in the sum is when  $i = j$

$$\begin{aligned} &= \langle \vec{u}_{k+1}, \vec{v}_j \rangle - \frac{\langle \vec{v}_j, \vec{u}_{k+1} \rangle}{\langle \vec{v}_j, \vec{v}_j \rangle} \langle \vec{v}_j, \vec{v}_j \rangle \\ &= \langle \vec{u}_{k+1}, \vec{v}_j \rangle - \langle \vec{v}_j, \vec{u}_{k+1} \rangle = 0 \end{aligned}$$

Therefore  $\vec{v}_{k+1}$  is orthogonal to all other  $\vec{v}$ 's and thus the basis consisting of  $\vec{v}$  form an orthonormal set.  $\square$

Find a orthogonal set of vectors than spans the same subspace as

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Let  $\vec{u}_1, \vec{u}_2$  and  $\vec{u}_3$  be these three vectors. We now use Gram-Schmidt orthogonalization to find an orthogonal set.

1.

$$\vec{v}_1 = \vec{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

2.

$$\vec{v}_2 = \vec{u}_2 - \frac{\langle \vec{u}_2, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \end{bmatrix}$$

3.

$$\begin{aligned} \vec{v}_3 &= \vec{u}_3 - \frac{\langle \vec{u}_3, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2 - \frac{\langle \vec{u}_3, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 \\ &= \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1/2}{3/2} \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 2/3 \\ 2/3 \end{bmatrix} \end{aligned}$$

And the set  $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$  is now orthogonal (check it).

Note: there are many other orthogonal sets that span this space subspace. For example, the standard basis does this or if we would have shuffled the original three vectors, so this is not unique.

Although the example shown here uses vectors in  $\mathbb{R}^3$ , Gram-Schmidt can be done on any set of vectors from an inner product space. This is very common for polynomials and will show an example with this next. Also, in Chapter ??, we will also show this with trigonometric functions.

Use Gram-Schmidt orthogonalization to find a orthogonal set that spans  $\mathcal{P}_3[-1, 1]$ , the set of all cubic functions on the interval  $[-1, 1]$  with inner product

$$\langle p, q \rangle = \int_{-1}^1 p(x)q(x) dx$$

To begin, we need a set of vectors (polynomials) on the set and we can take the standard basis  $\{1, x, x^2, x^3\}$ . Call these  $p_0, p_1, p_2, p_3$  and we'll use Gram-Schmidt to find a set  $\{q_0, q_1, q_2, q_3\}$  that is orthogonal.

It is very helpful to recall that  $x^n$  is an odd function if  $n$  is odd. Also  $\int_{-1}^1 x^n dx = 0$  if  $n$  is odd.

$$\begin{aligned} q_0 &= p_0 = 1, \\ q_1 &= p_1 - \frac{\langle p_1, q_0 \rangle}{\langle q_0, q_0 \rangle} q_0 \\ &= x - \frac{\int_{-1}^1 x \cdot 1 dx}{\int_{-1}^1 1 \cdot 1 dx} 1 \end{aligned}$$

and since  $x \cdot 1 = x$  is odd, the integral in the numerator is 0,

$$\begin{aligned} &= x \\ q_2 &= p_2 - \frac{\langle p_2, q_1 \rangle}{\langle q_1, q_1 \rangle} q_1 - \frac{\langle p_2, q_0 \rangle}{\langle q_0, q_0 \rangle} q_0 \\ &= x^2 - \frac{\int_{-1}^1 x^2 \cdot x dx}{\int_{-1}^1 x \cdot x dx} - \frac{\int_{-1}^1 x^2 \cdot 1 dx}{\int_{-1}^1 1 \cdot 1 dx} 1 \end{aligned}$$

and  $x^2 \cdot x = x^3$  is odd, so the first integral is 0 and also  $\int_{-1}^1 1 dx = 2$ ,

$$\begin{aligned} &= x^2 - \frac{x^3/3|_{-1}^1}{2} 1 = x^2 - \frac{1}{3} \\ q_3 &= p_3 - \frac{\langle p_3, q_2 \rangle}{\langle q_2, q_2 \rangle} q_2 - \frac{\langle p_3, q_1 \rangle}{\langle q_1, q_1 \rangle} q_1 - \frac{\langle p_3, q_0 \rangle}{\langle q_0, q_0 \rangle} q_0 \\ &= x^3 - \frac{\int_{-1}^1 x^3(x^2 - 1/3) dx}{\int_{-1}^1 (x^2 - 1/3)^2 dx} (x^2 - \frac{1}{3}) - \frac{\int_{-1}^1 x^3 \cdot x dx}{\int_{-1}^1 x \cdot x dx} x - \frac{\int_{-1}^1 x^3 \cdot 1 dx}{\int_{-1}^1 1 \cdot 1 dx} 1 \end{aligned}$$

and the integrals in the numerators of the 2nd and 4th terms are 0 due to the functions being odd, so

$$\begin{aligned} &= x^3 - \frac{x^5/5|_{-1}^1}{x^3/3|_{-1}^1} x \\ &= x^3 - \frac{3}{5} x \end{aligned}$$

There an orthogonal set of cubic polynomials that span  $\mathcal{P}_3[-1, 1]$  is

$$\{1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x\}$$

and these are the first four Legendre Polynomials. Also, they can be used for form a basis for  $\mathcal{P}_3[-1, 1]$ .

## 2.4 Linear Transformations

This section discusses linear transformations. In short, such a transformation will map vectors to vectors in a linear way. The definition of a linear transformation (or linear map) is the following:

Let  $V$  and  $W$  be vector spaces. A **linear transformation** or **linear map**  $T$  from  $V$  to  $W$  is a function that assigns to each vector  $\vec{v} \in V$  a unique vector  $T\vec{v} \in W$  and that satisfies for each  $\vec{u}$  and  $\vec{v}$  in  $V$  and each scalar  $\alpha$ ,

$$T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}), \quad (2.1)$$

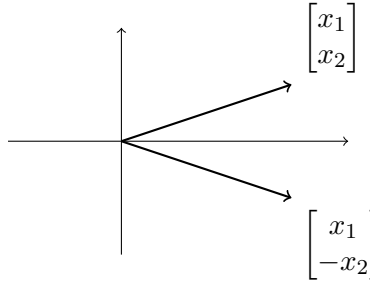
$$T(\alpha\vec{v}) = \alpha T(\vec{v}), \quad (2.2)$$

These are also called *homomorphisms* and the notation explaining that a map  $T$  goes from  $V$  to  $W$  is  $T : V \rightarrow W$ .

The reflection of any vector in  $\mathbb{R}^2$  across the horizontal axis is a linear map. Specifically this is given as

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$$

and geometrically you can see this as:



Show that this is a linear transformation.

Specifically, we need to show that  $T$  defined above satisfies (??) and (??).

Let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

and  $\alpha \in \mathbb{R}$ , then

$$\begin{aligned} T(\vec{x} + \vec{y}) &= T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} x_1 + y_1 \\ -(x_2 + y_2) \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ -x_2 - y_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ -y_2 \end{bmatrix} = T(\vec{x}) + T(\vec{y}). \end{aligned}$$

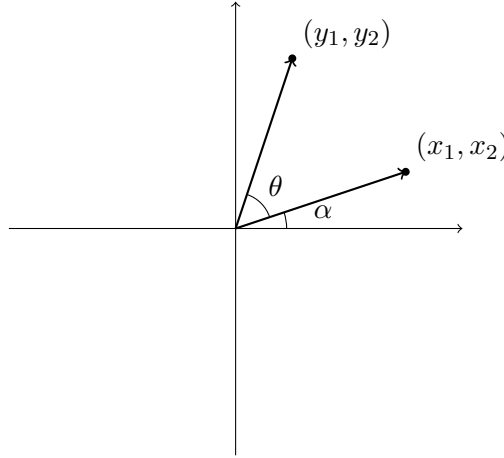
so (??) is satisfied. Next,

$$\begin{aligned} T(\alpha \vec{x}) &= T\left(\alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} \alpha x_1 \\ \alpha x_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} \alpha x_1 \\ -\alpha x_2 \end{bmatrix} = \alpha \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix} = \alpha T(\vec{x}). \end{aligned}$$

so ??) is satisfied.

A very important linear map in  $\mathbb{R}^2$  is the rotational map that takes any vector in the plane and rotates it a given angle. The next example, derives this map.

Let  $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  be a vector in  $\mathbb{R}^2$ . Let the function  $T$  take the vector  $\vec{x}$  and rotate it by  $\theta$  radians in the counterclockwise direction. Call the new vector  $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ .



Derive the formula for the rotational map.

Let  $r = \|\vec{x}\|$  and since  $\vec{y}$  is the rotated version of  $\vec{x}$  it has the same length, therefore  $r = \|\vec{y}\|$ . The values  $x_1, x_2, y_1, y_2$  can be written in terms of  $\alpha$ , the angle that the vector  $\vec{x}$  makes with the positive horizontal axis, and  $\theta$  the angle between the vectors as follows.

$$\begin{aligned} x_1 &= r \cos \alpha, & x_2 &= r \sin \alpha, \\ y_1 &= r \cos(\alpha + \theta), & y_2 &= r \sin(\alpha + \theta) \\ &= r \cos \alpha \cos \theta - r \sin \theta \sin \alpha, & y_2 &= r \sin \theta \cos \alpha + r \sin \alpha \cos \theta \\ &= x_1 \cos \theta - x_2 \sin \theta, & &= x_1 \sin \theta + x_2 \cos \theta \end{aligned}$$

and note that these can be written:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

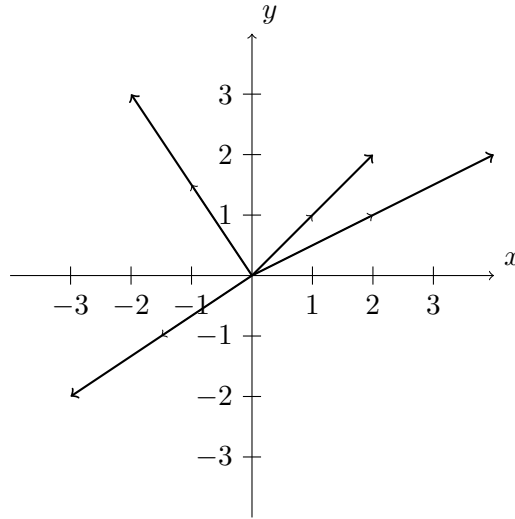
This is a linear transformation (as we will explain later) and is called a **rotational transformation**.



Consider the map  $S : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  that scales any vector in the plane by a factor of  $k$  given by

$$S(\vec{x}) = k\vec{x}$$

which can be visualized in the following diagram where  $k = 2$ :



where every vector under the map results in a new vector that is twice as long as the original. In general, the scale  $k$  will scale the vector by a factor of  $k$  and recall that if  $k < 0$ , then the direction changes. Show that this is a linear map.

Again, we show that (??) and (??) are satisfied. Let  $\vec{x}$  and  $\vec{y}$  be elements of  $\mathbb{R}^2$ .

$$S(\vec{x} + \vec{y}) = k(\vec{x} + \vec{y}) = k\vec{x} + k\vec{y} = S(\vec{x}) + S(\vec{y})$$

so (??) is satisfied and

$$S(\alpha\vec{x}) = k(\alpha\vec{x}) = \alpha(k\vec{x}) = \alpha S(\vec{x})$$

so (??) is satisfied so  $S$  is a linear map.

If a transformation map  $T$  is written as a matrix or  $T(\vec{x}) = A\vec{x}$ , then  $T$  is a linear transformation.

*Proof.* This is a consequence of matrix operations.

$$\begin{aligned} T(\vec{x} + \vec{y}) &= A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} = T(\vec{x}) + T(\vec{y}) \\ T(\alpha\vec{x}) &= A(\alpha\vec{x}) = \alpha A\vec{x} = \alpha T(\vec{x}). \end{aligned}$$

□

Note that the rotational transformation that was defined in Example ?? is easily shown to be a linear transformation because from Theorem ??, any transformation shown as a matrix, is a linear transformation.

The next theorem shows the counter direction to Theorem ??, that is that any linear transformation can be written as a matrix.

Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear transformation. Then there exists a unique  $m$  by  $n$  matrix  $A_T$  such that

$$T\vec{x} = A_T\vec{x} \quad \text{for every } \vec{x} \in \mathbb{R}^n.$$

We won't prove this here, but instead will motivate this below. In short, if  $T$  is a linear transformation, then the matrix  $A_T$  corresponding to the linear transformation is called the **transformation matrix**.

### 2.4.1 Finding the Matrix Form of a Linear Transformation

The theorem above shows that any linear transformation,  $T : V \rightarrow W$  can be written in matrix form. This section explains how to find it. Let  $B_V = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m)$  be a basis of  $V$  and  $B_W = (\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n)$  be a basis of  $W$ . Any vector  $\vec{x}$  in  $V$  can be written

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_m\vec{v}_m = \sum_{j=1}^m c_j\vec{v}_j$$

or in other words  $\text{Rep}_{B_V}(\vec{x}) = \vec{c}$ . Applying the map to  $\vec{x}$  is

$$T(\vec{x}) = T\left(\sum_{j=1}^m c_j\vec{v}_j\right)$$

and since it is a linear map

$$= \sum_{j=1}^m c_j T(\vec{v}_j) \tag{2.3}$$

Next, we write the transformation in terms of the basis vectors of  $W$  or

$$T(\vec{v}_j) = a_{1,j}\vec{w}_1 + a_{2,j}\vec{w}_2 + \dots + a_{n,j}\vec{w}_n = \sum_{i=1}^n a_{i,j}\vec{w}_i \tag{2.4}$$

Substituting (??) into (??) results in

$$T(\vec{x}) = \sum_{j=1}^m c_j \sum_{i=1}^n a_{i,j}\vec{w}_i = \sum_{i=1}^n \sum_{j=1}^m a_{i,j}c_j\vec{w}_i$$

and letting  $A_T$  be the  $m \times n$  matrix with entries  $a_{i,j}$  then

$$\text{Rep}_{B_W}(T(\vec{x})) = A_T\vec{c}$$

or in other words, the matrix performs the map on the coefficients.

Equation (??) also shows how the matrix  $A_T$  can be created from the linear map. That equation can also be thought of as a representation of the basis vectors or

$$\text{Rep}_{B_W}(T(\vec{v}_j)) = \vec{a}_j$$

where  $a_j$  is the  $j$ th column of  $A$ . The following summarizes how to find the matrix.

If  $A_T$  is the matrix representation of the map  $T$ , then the  $j$ th column of  $A_T$  is the vector  $\text{Rep}_{B_W}(T(\vec{v}_j))$ , the map applied to the  $j$ th basis vector of  $V$  written in terms of the basis of  $W$ .

We now show many examples on how to apply this.

Find the matrix representation of the reflection map from Example ?? given by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$$

where the representation will be in terms of the standard basis vectors.

In this case, we need to determine how the standard basis vectors map under the reflection. Thus

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and

$$T(\vec{v}_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad T(\vec{v}_2) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Because we are using the standard basis vectors, the representations of these vectors are themselves therefore,

$$A_T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

and just to verify,

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$$

Find the matrix representation of the scale map in Example ??.

We need to map the basis vectors  $[1 \ 0]^\top$  and  $[0 \ 1]^\top$  to determine the columns of the matrix representation.

$$S\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} k \\ 0 \end{bmatrix} \qquad S\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ k \end{bmatrix}$$

so the matrix representation is

$$A_S = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$$

Let  $T : \mathcal{M}_{2 \times 2} \rightarrow \mathbb{R}$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \rightarrow a + d$$

which is the trace of a 2 by 2 matrix. Show that the trace is a linear map and find the matrix representation of the trace.

First, to prove that the trace is a linear map, we need to show that it satisfies (??) and (??). Let

$$A_1 = \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}$$

Then

$$\begin{aligned} T(A_1 + A_2) &= T\left(\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} + \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{bmatrix}\right) \\ &= (a_1 + a_2) + (d_1 + d_2) \\ &= (a_1 + d_1) + (a_2 + d_2) = T(A_1) + T(A_2) \end{aligned}$$

Similarly,

$$\begin{aligned} T(\alpha A_1) &= T\left(\alpha \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}\right) = T\left(\begin{bmatrix} \alpha a_1 & \alpha b_1 \\ \alpha c_1 & \alpha d_1 \end{bmatrix}\right) \\ &= \alpha a_1 + \alpha d_1 = \alpha(a_1 + d_1) = \alpha T(A_1) \end{aligned}$$

Next, we want to find the matrix representation of the trace. To do this, we need to determine how the map affects the basis of the vector space  $\mathcal{M}_{2 \times 2}$ , which is

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

and since

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} &\mapsto 1, & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} &\mapsto 0, \\ \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} &\mapsto 0, & \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} &\mapsto 1, \end{aligned}$$

The matrix representation is

$$A_T = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}$$

To verify the above results, recall that from Example ?? that the representation of a matrix is the unfolded matrix or in the  $2 \times 2$  case, that

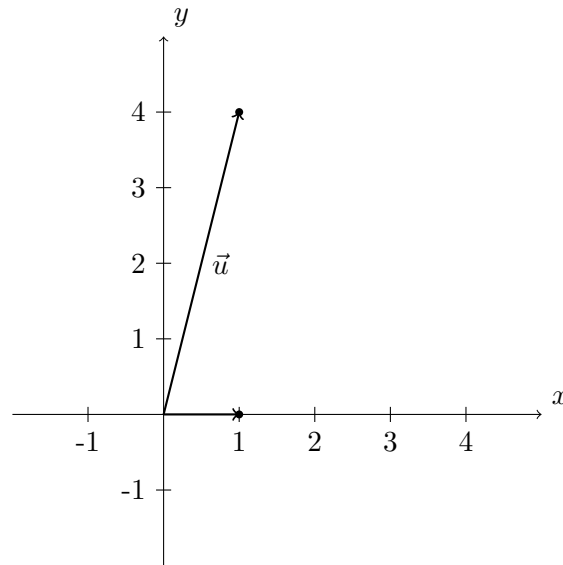
$$\text{Rep}_B \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

where  $B$  is the natural basis of  $\mathcal{M}_{2 \times 2}$ . So the matrix trace can be written as

$$\begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} \text{Rep}_B(A) = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = a + d$$

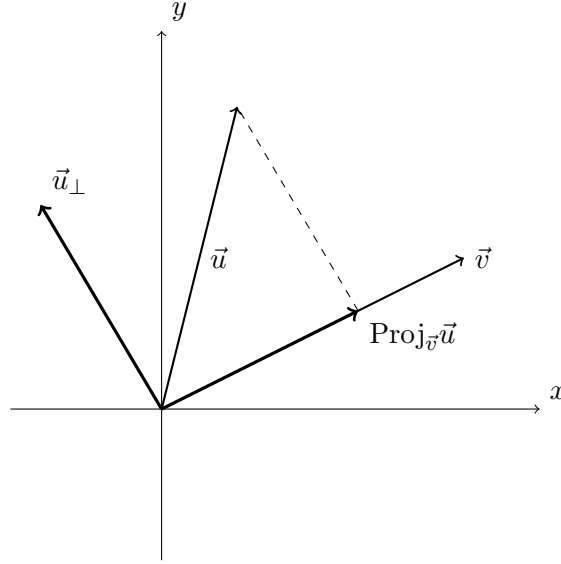
## 2.5 Projection Maps

There is a class of linear transformations that are very important and have a nice geometric interpretation called projection maps. Let's look at an example in the  $xy$ -plane. Consider the point  $(1, 4)$  as shown below.



If the point (or the vector) is projected onto the  $x$ -axis, then the result is the point  $(1, 0)$  found by taking the  $x$ -component of the point. In term of the vector, this is  $\begin{bmatrix} 1 & 0 \end{bmatrix}^T$ .

We now look at projecting a vector  $\vec{u}$  onto a line (or a vector  $\vec{v}$ ). We will derive this in the  $xy$ -plane, however the result will work for any vector space.



Let  $\vec{u}$  and  $\vec{v}$  be two vectors in  $\mathbb{R}^2$  as shown above. We seek the projection of  $\vec{u}$  onto  $\vec{v}$  and denote this as  $\text{Proj}_{\vec{v}}\vec{u}$ . The projection is a vector that is in the same direction as  $\vec{v}$ . The length will be explained below.

Any vector  $\vec{u}$  can be split into two parts,  $\vec{u}_{||}$  a vector in a direction parallel to  $\vec{v}$  and a vector  $\vec{u}_{\perp}$  which satisfies  $\langle \vec{u}_{\perp}, \vec{v} \rangle = 0$ .

$$\vec{u} = \vec{u}_{||} + \vec{u}_{\perp} \quad (2.5)$$

where  $\vec{u}_{||} = \alpha \vec{v}$ .

Take the inner product of both sides of (2.5) with the vector  $\vec{v}$

$$\begin{aligned} \langle \vec{u}, \vec{v} \rangle &= \langle \vec{u}_{||} + \vec{u}_{\perp}, \vec{v} \rangle \\ &= \langle \alpha \vec{v}, \vec{v} \rangle + \langle \vec{u}_{\perp}, \vec{v} \rangle \\ &= \alpha \langle \vec{v}, \vec{v} \rangle + 0 \end{aligned}$$

and solving for  $\alpha$ ,

$$\alpha = \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle}$$

Therefore the projection vector

$$\text{Proj}_{\vec{v}}\vec{u} = \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \vec{v} \quad (2.6)$$

and if needed the vector  $\vec{u}_{\perp}$  is

$$\vec{u}_{\perp} = \vec{u} - \text{Proj}_{\vec{v}}\vec{u} = \vec{u} - \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \vec{v}$$

Although we derived this projection in  $\mathbb{R}^2$ , there is nothing about the projection formula to indicate it's confined to  $\mathbb{R}^2$ . In fact, as we will see, the projection mapping is a linear transformation and applies to vectors in an inner product space.

The projection of any vector  $\vec{u} \in \mathbb{R}^n$  onto a vector  $\vec{v} \in \mathbb{R}^n$  is given by

$$\text{Proj}_{\vec{v}}\vec{u} = \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \vec{v}$$

and the vector perpendicular to  $\text{Proj}_{\vec{v}}\vec{u}$  that satisfies  $\vec{u} = \text{Proj}_{\vec{v}}\vec{u} + \vec{u}_{\perp}$  is given by

$$\vec{u}_{\perp} = \vec{u} - \text{Proj}_{\vec{v}}\vec{u} = \vec{u} - \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \vec{v}$$

The example below finds the projection of vectors in  $\mathbb{R}^3$ .

Find the projection of  $\vec{u} = [1 \ 2 \ 3]^T$  onto the vector  $\vec{v} = [-2 \ 0 \ 4]$ .

Since

$$\langle \vec{u}, \vec{v} \rangle = -2 + 0 + 12 = 10 \qquad \langle \vec{v}, \vec{v} \rangle = (-2)^2 + 0^2 + 4^2 = 20.$$

$$\text{Proj}_{\vec{v}}\vec{u} = \frac{10}{20}\vec{v} = \frac{1}{2} \begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

Given a nonzero vector  $\vec{v}$  and a vector  $\vec{u}$  both in  $\mathbb{R}^n$ , the projection map in (??) is a linear transformation.

*Proof.* We need to prove the two properties of linear transformations in Definition ??.

$$\begin{aligned} \text{Proj}_{\vec{v}}(\vec{u} + \vec{w}) &= \frac{\langle \vec{u} + \vec{w}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \\ &= \frac{\langle \vec{u}, \vec{v} \rangle + \langle \vec{w}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \\ &= \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} + \frac{\langle \vec{w}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \\ &= \text{Proj}_{\vec{v}}\vec{u} + \text{Proj}_{\vec{v}}\vec{w} \end{aligned}$$

$$\begin{aligned} \text{Proj}_{\vec{v}}(\alpha\vec{u}) &= \frac{\langle \alpha\vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} \\ &= \alpha \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} = \alpha \text{Proj}_{\vec{v}}\vec{u} \end{aligned}$$

□

The example shown so far for projections have been onto a vector (which can be thought of as the line through the origin with that direction), but in general, one can project onto any linear space (more technically, any inner product space). Before showing the general projection, let's consider the projection of a vector onto a plane. The following example shows this.

Find the projection of vector  $[6, -2, 4]^T$  onto plane given by  $x + 2y + 3z = 0$ .

First, let's find two vectors in the plane, called  $P$ , that form an orthogonal basis (we will see why it's helpful to have an orthogonal basis later). Recall that a normal vector to the plane is given by the vector  $[1, 2, 3]^T$ . Two vectors that are normal to it is  $\vec{u}_1 = [3, 0, -1]$  and  $\vec{u}_2 = [0, -3, 2]^T$ . (Note: these are two points in the plane). To form an orthonormal basis, we need to perform Gram-Schmidt.

$$\begin{aligned}\vec{v}_1 &= \vec{u}_1 \\ &= \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} \\ \vec{v}_2 &= \vec{u}_2 - \frac{\langle \vec{v}_1, \vec{u}_2 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 \\ &= \begin{bmatrix} 0 \\ -3 \\ 2 \end{bmatrix} - \left(-\frac{1}{5}\right) \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 3/5 \\ -3 \\ 9/5 \end{bmatrix}\end{aligned}$$

and to make the second vector nicer, we'll multiply by 5/3. An orthogonal set in the plane is

$$\left\{ \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -5 \\ 3 \end{bmatrix} \right\}$$

To consider how to project the vector  $\vec{u} = [1, 2, 3]^T$  into the plane we will write

$$\vec{u} = \vec{u}_\perp + \vec{u}_\parallel$$

where  $\vec{u}_\perp$  is orthogonal to both  $\vec{v}_1$  and  $\vec{v}_2$  and we can write

$$\text{Proj}_P \vec{u} = \vec{u}_\parallel = c_1 \vec{v}_1 + c_2 \vec{v}_2$$

So we next take the inner product of  $\vec{u}$  with both  $\vec{v}_1$  and  $\vec{v}_2$ .

$$\begin{aligned}\langle \vec{u}, \vec{v}_1 \rangle &= \langle \vec{u}_\perp + \vec{u}_\parallel, \vec{v}_1 \rangle & \langle \vec{u}, \vec{v}_2 \rangle &= \langle \vec{u}_\perp + \vec{u}_\parallel, \vec{v}_2 \rangle \\ &= \langle \vec{u}_\perp, \vec{v}_1 \rangle + \langle \vec{u}_\parallel, \vec{v}_1 \rangle & &= \langle \vec{u}_\perp, \vec{v}_2 \rangle + \langle \vec{u}_\parallel, \vec{v}_2 \rangle \\ &= 0 + \langle c_1 \vec{v}_1 + c_2 \vec{v}_2, \vec{v}_1 \rangle & &= 0 + \langle c_1 \vec{v}_1 + c_2 \vec{v}_2, \vec{v}_2 \rangle \\ &= c_1 \langle \vec{v}_1, \vec{v}_1 \rangle + c_2 \langle \vec{v}_2, \vec{v}_1 \rangle & &= c_1 \langle \vec{v}_1, \vec{v}_2 \rangle + c_2 \langle \vec{v}_2, \vec{v}_2 \rangle \\ &= c_1 \langle \vec{v}_1, \vec{v}_1 \rangle + 0 & &= 0 + c_2 \langle \vec{v}_2, \vec{v}_1 \rangle\end{aligned}$$



Therefore

$$c_1 = \frac{\langle \vec{u}, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \qquad c_2 = \frac{\langle \vec{u}, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle}$$

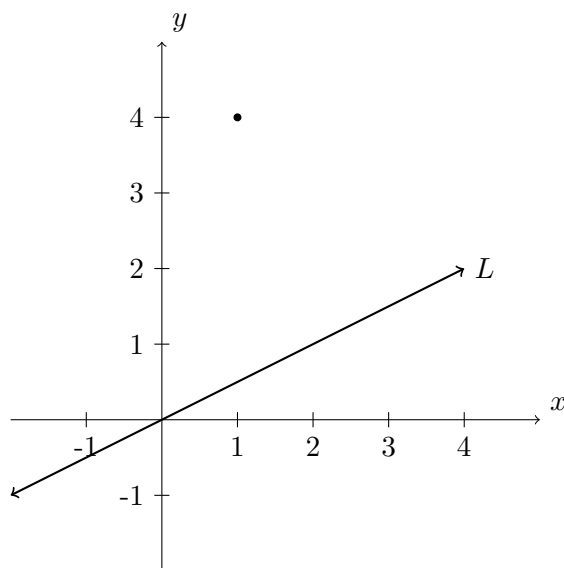
So in the case of projecting a vector onto a plane,  $P$ ,

$$\begin{aligned} \text{Proj}_P \vec{u} &= c_1 \vec{v}_1 + c_2 \vec{v}_2 \\ &= \frac{\langle \vec{u}, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 + \frac{\langle \vec{u}, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2 \\ &= \frac{14}{10} \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} + \frac{28}{35} \begin{bmatrix} 1 \\ -5 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix} \end{aligned}$$

So the vector  $[5, -4, 1]^\top$  is the projection of the vector  $[1, 2, 3]^\top$  onto the plane  $x+2y+3z = 0$ .

### 2.5.1 Using Projections to solve Least-Squares Problems

A useful problem is to find the minimum distance between a point and line. For example consider the point  $(1, 4)$  and the line  $L : y = \frac{1}{2}x$  as shown in the figure below.

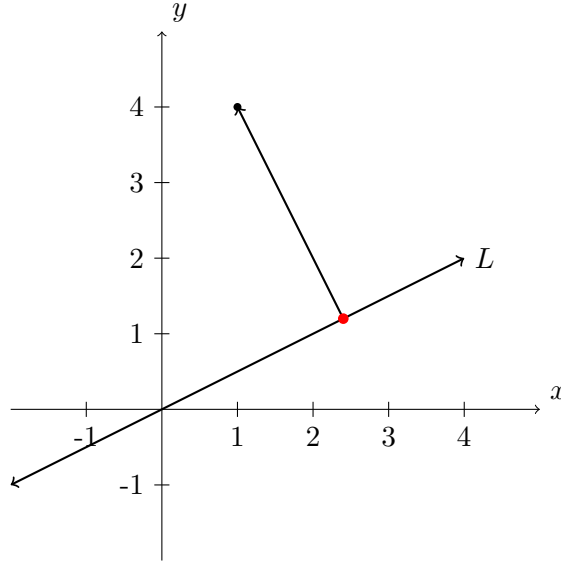


If we want to minimize the distance from the point  $P$  to the line  $L$ , we often minimize the square<sup>1</sup> of the distance from the point to the line or

$$g(x) = (x - 1)^2 + \left(\frac{1}{2}x - 4\right)^2$$

<sup>1</sup>Recall that the point that minimizes a function is the same point that to minimizes the square of the function. The reason for doing this it it gets rid of the square root and makes calculations easier to do.

and using techniques from calculus, this function is minimized when  $x = 12/5$ . Looking at the plot, the  $y$ -value is  $12/5$  and the vector from  $(12/5, 6/5)$  to  $(1, 4)$  is perpendicular to the vector in the direction of  $L$  as shown below:



In light of projections, we can reframe this problem. We are seeking the point on  $L$  closest to  $P$ . This can be found by projecting the vector  $\overrightarrow{OP}$  to the line  $L$  or

$$\text{proj}_{\vec{v}} \overrightarrow{OP} = \frac{\langle \vec{v}, \overrightarrow{OP} \rangle}{\langle \vec{v}, \vec{v} \rangle} \vec{v}$$

and in this case, with  $\vec{v} = [2 \ 1]^\top$  and  $\overrightarrow{OP} = [1 \ 4]^\top$ ,

$$\text{proj}_{\vec{v}} \overrightarrow{OP} = \frac{\langle \vec{v}, \overrightarrow{OP} \rangle}{\langle \vec{v}, \vec{v} \rangle} = \frac{6}{5} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 12/5 \\ 6/5 \end{bmatrix}$$

which is the same result as from Calculus.

### 2.5.2 Projecting onto a Vector Space

We again generalize from projecting onto a single vector or set of vectors to a general vector space,  $V$ . First, to make things easier, we will use an orthogonal basis for  $V$ , call it,  $(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$ .

Recall that a projection onto a vector is found by writing the original vector,  $\vec{u} = \vec{u}_{\parallel} + \vec{u}_{\perp}$ , where  $\vec{u}_{\parallel}$  is in  $V$  or

$$\vec{u} = \vec{u}_{\perp} + \sum_{i=1}^n c_i \vec{v}_i \quad (2.7)$$

Finding the projection is analogous to finding the constants  $c_i$ . Take the inner product of (??) with  $v_k$ , the  $k$ th basis vector of  $V$  or

$$\langle \vec{u}, \vec{v}_k \rangle = \left\langle \vec{u}_\perp + \sum_{i=1}^n c_i \vec{v}_i, \vec{v}_k \right\rangle$$

using properties of inner products

$$= \langle \vec{u}_\perp, \vec{v}_k \rangle + \sum_{i=1}^n c_i \langle \vec{v}_i, \vec{v}_k \rangle$$

since  $\vec{u}_\perp$  is chosen to be orthogonal to the vector space and since the basis of  $V$  is orthogonal,

$$= c_k \langle \vec{v}_k, \vec{v}_k \rangle$$

Therefore

$$c_k = \frac{\langle \vec{u}, \vec{v}_k \rangle}{\langle \vec{v}_k, \vec{v}_k \rangle}$$

and thus we can make the following statement about projecting any element onto a subspace:

Let  $\vec{u}$  be an element in a vector space  $S$  and  $(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$  be an orthogonal basis of a subspace  $V$ . The projection of  $\vec{u}$  onto  $V$  can be written:

$$\text{Proj}_V(\vec{u}) = \sum_{i=1}^n c_i \vec{v}_i$$

where

$$c_k = \frac{\langle \vec{u}, \vec{v}_k \rangle}{\langle \vec{v}_k, \vec{v}_k \rangle}$$

Note: if the basis of  $V$  is also orthonormal, then  $c_k = \langle \vec{u}, \vec{v}_k \rangle$ .

Find the projection of  $\sin \pi x$ , an element of  $\mathcal{C}^{(0)}$  onto  $\mathcal{P}_5[-1, 1]$

First, we need a orthogonal basis of the subspace. This was found above and we can use the Legendre polynomials or

$$(1, x, 3x^2 - 1, 5x^3 - 3x, 35x^4 - 30x^2 + 3, 63x^5 - 70x^3 + 15x) \quad (2.8)$$

Next, find all of the values of  $c_k$  for  $k = 0, 1, 2, 3, 4, 5$

$$c_0 = \frac{\langle \sin \pi x, 1 \rangle}{\langle 1, 1 \rangle} = \frac{\int_{-1}^1 \sin \pi x \, dx}{\int_{-1}^1 1^2 \, dx} = 0$$

$$c_1 = \frac{\langle \sin \pi x, x \rangle}{\langle x, x \rangle} = \frac{\int_{-1}^1 x \sin \pi x \, dx}{\int_{-1}^1 x^2 \, dx} = \frac{2/\pi}{2/3} = \frac{3}{\pi}$$

Similarly, it can be shown that

$$c_2 = c_4 = 0$$

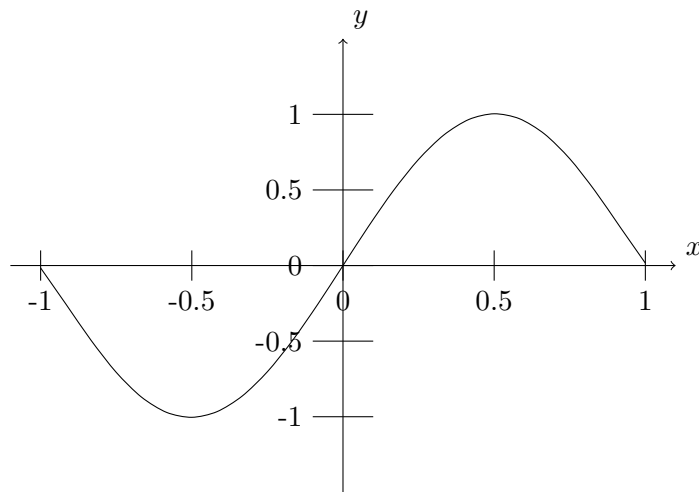
$$c_3 = \frac{7(\pi^2 - 15)}{\pi^3} \qquad c_5 = \frac{11(\pi^4 - 105\pi^2 + 945)}{\pi^5}$$

The projection then is the sum

$$\sum_{i=0}^5 c_k P_k(x)$$

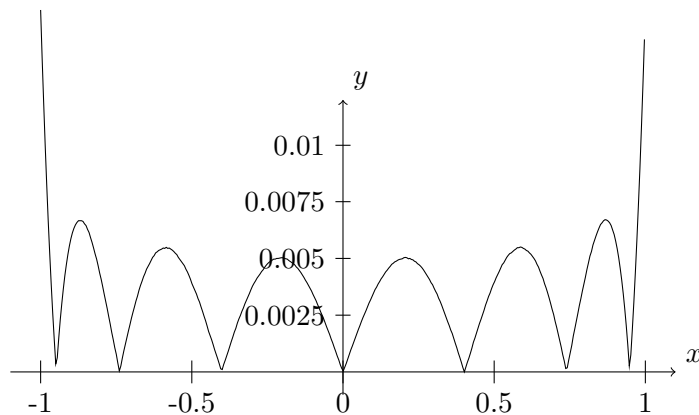
where  $P_k$  is the  $k$ th Legendre polynomial as in (??).

A plot of this is



which looks quite sine-like and is fairly close if we plot

$$|\sin \pi x - \sum_{i=0}^5 c_k P_k(x)|$$



---

This last example shows the power of projections. If we are using functions as elements of vector spaces, then a projection of a function onto a vector space (using the span of a vector space), is the closest function in the vector space to the original function. In Chapter ??, we will do this with trigonometric function and is called Fourier Series.



## Chapter 3

# Eigenvalues and Eigenvectors

There is another very important aspect to linear algebra that will play a very important role in applied mathematics. This is the notion of an eigenvector and eigenvalue. Briefly, these are a vector and scalar in an equation, however, first, there is a strong geometric role behind them and secondly, eigenvectors and eigenvalues play an important role in many parts of linear algebra. Lastly, we will see an extension of them for functions later in the text.

### 3.1 Eigenvalues and Eigenvectors

For a square matrix  $A$ , an **eigenvalue**  $\lambda$  with related **eigenvector**  $\vec{v}$  satisfy

$$A\vec{v} = \lambda\vec{v}$$

as long as  $\vec{v}$  is not the zero vector.

Recall that the matrix operation  $A\vec{v}$  is a linear transformation on  $\vec{v}$ , that is, it takes a vector  $\vec{v}$  to another vector. Thus the eigenvector of a matrix is a vector such that  $A\vec{v}$  results in a vector in the same direction as  $\vec{v}$ . The eigenvalue is the scalar transformation associated with this.

Show that  $\vec{v} = [1 \ 1]^T$  is an eigenvector of

$$A = \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix}$$

We need to show that this satisfies  $A\vec{v} = \lambda\vec{v}$  for some scalar  $\lambda$ . Since

$$A\vec{v} = \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

and since this matrix is 7 times the vector  $\vec{v}$ , this proves that  $\vec{v}$  is an eigenvector. The scalar number 7 is the eigenvalue.

### 3.1.1 Finding Eigenvalues and Eigenvectors

We will first find the eigenvalues of a given matrix. Start with the eigenvector equation:

$$A\vec{v} = \lambda\vec{v}$$

subtract  $\lambda\vec{v}$

$$A\vec{v} - \lambda\vec{v} = \vec{0},$$

introduce an identity matrix

$$A\vec{v} - \lambda I\vec{v} = \vec{0},$$

factor out the  $\vec{v}$

$$(A - \lambda I)\vec{v} = \vec{0} \tag{3.1}$$

and we have transformed this into a homogeneous matrix equation. As stated in the definition of the eigenvector, it cannot be the zero vector, and by Theorem ??, the only solution where  $\vec{v}$  is nonzero is when

$$\det(A - \lambda I) = 0 \tag{3.2}$$

This equation is called the characteristic equation.

To find an eigenvector associated with the eigenvalue, solve

$$(A - \lambda I)\vec{v} = \vec{0}$$

for  $\vec{v}$ , that is find the basis of the null space of the matrix  $A - \lambda I$ .

Note: as we will see in all of the examples in this section that the null space of  $A - \lambda I$  has at least dimension 1 and that the reduced echelon form of  $A - \lambda I$  has at least one row of zeros, indicating that the rank of  $A - \lambda I$  is less than  $n$ . The reason for this is because of Theorem ??. This is helpful for detecting errors in finding eigenvalues/eigenvectors.

Find all eigenvalues and eigenvectors of

$$\begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix}$$

First, solve  $|A - \lambda I| = 0$ ,

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 3 - \lambda & 4 \\ 2 & 5 - \lambda \end{vmatrix} = (3 - \lambda)(5 - \lambda) - 8 \\ &= \lambda^2 - 8\lambda + 7 = (\lambda - 1)(\lambda - 7) = 0 \end{aligned}$$

so  $\lambda = 1, 7$ .



Next, we need to find an eigenvector for each eigenvalue. When  $\lambda = 1$ ,

$$(A - 1I)\vec{v} = \begin{bmatrix} 2 & 4 \\ 2 & 4 \end{bmatrix} \vec{v} = \vec{0}$$

To solve this matrix equation, we'll use gaussian elimination on the augmented matrix:

$$\begin{array}{l} -R_1 + R_2 \rightarrow R_2 \\ \frac{1}{2}R_1 \rightarrow R_1 \end{array} \quad \begin{bmatrix} 2 & 4 & | & 0 \\ 2 & 4 & | & 0 \\ 2 & 4 & | & 0 \\ 0 & 0 & | & 0 \\ 1 & 2 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$$

The first row of the matrix corresponds to  $x_1 + 2x_2 = 0$  so the null space of  $A - \lambda I$  is

$$\left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} s \mid s \in \mathbb{R} \right\}$$

The eigenvector(s) associated with  $\lambda = 1$  is the basis of this space or

$$\vec{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

And to find the eigenvector associated with  $\lambda = 7$ , solve

$$(A - 7I)\vec{v} = \begin{bmatrix} -4 & 4 \\ 2 & -2 \end{bmatrix} \vec{v} = \vec{0}$$

and use Gaussian Elimination to reduce

$$\begin{array}{l} -\frac{1}{4}R_1 \rightarrow R_1 \\ -2R_1 + R_2 \rightarrow R_2 \end{array} \quad \begin{bmatrix} -4 & 4 & | & 0 \\ 2 & -2 & | & 0 \\ 1 & -1 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$$

The null space of  $A - 7I$  is

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} s \mid s \in \mathbb{R} \right\}$$

so the eigenvector associated with  $\lambda = 7$  is the basis of this space or

$$\vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Overall, there are two eigenvalues and two related eigenvectors they are

$$\lambda_1 = 1, \quad \vec{v}_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \qquad \lambda_2 = 7, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

This example showed that this  $2 \times 2$  matrix has two real eigenvalues and two corresponding eigenvectors. Also, although it seemed that these eigenvectors may be unique, we'll show that  $[-3 \ -3]^\top$  is also a eigenvector of the matrix in Example ??.

$$\begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} -3 \\ -3 \end{bmatrix} = \begin{bmatrix} -21 \\ -21 \end{bmatrix} = 7 \begin{bmatrix} -3 \\ -3 \end{bmatrix}$$

which shows directly that  $[-3 \ -3]^\top$  is an eigenvector with corresponding eigenvalue 7. Does this mean that there are other eigenvectors? Yes. The following lemma shows this.

If  $\vec{v}$  is an eigenvector of  $A$  with corresponding eigenvalue  $\lambda$ , then  $k\vec{v}$  is also an eigenvector of  $A$  with corresponding eigenvalue  $\lambda$  if  $k \neq 0$ .

*Proof.* Since  $\vec{v}$  is an eigenvector of  $A$  with corresponding eigenvalue  $\lambda$ , then  $A\vec{v} = \lambda\vec{v}$ . Multiplying this by  $k$  results in

$$\begin{aligned} k(A\vec{v}) &= k(\lambda\vec{v}) \\ A(k\vec{v}) &= \lambda(k\vec{v}) \end{aligned}$$

which shows the result. Also note, that  $k$  cannot be zero, because  $0\vec{v}$  is the zero vector, which is not an eigenvector.  $\square$

There are other possibilities of eigenvalues and eigenvectors of  $2 \times 2$  matrices. The following has complex eigenvalues.

Find the eigenvalues and eigenvectors of

$$\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$$

To find the eigenvalues, we find

$$\begin{vmatrix} -\lambda & 2 \\ -2 & -\lambda \end{vmatrix} = 0$$

or  $\lambda^2 + 4 = 0$ , which has the solutions  $\lambda = \pm 2i$ .

To find the eigenvectors, we find the null space associated with each of the two eigenvalues.

**Eigenvector for  $\lambda = 2i$**

Find the null space of  $A - 2iI$ :

$$\frac{1}{-2i}R_1 \rightarrow R_1 \qquad \begin{bmatrix} -2i & 2 \\ -2 & -2i \end{bmatrix} \qquad \begin{bmatrix} 1 & i \\ -2 & -2i \end{bmatrix}$$

where  $\frac{-2}{2i} = \frac{-1}{i} = \frac{-i}{i^2} = \frac{-i}{-1} = i$  is used

$$2R_1 + R_2 \rightarrow R_2, \quad \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}$$

and the top equation says  $x_1 = -ix_2$ , so the null space is

$$\left\{ \begin{bmatrix} -i \\ 1 \end{bmatrix} x_2 \mid x_2 \in \mathbb{R} \right\}$$

so the eigenvector is  $[-i \ 1]^\top$ .

**Eigenvector for  $\lambda = -2i$**

Find the null space of  $A - 2iI$ :

$$\begin{array}{l} \frac{1}{2i}R_1 \rightarrow R_1 \\ 2R_1 + R_2 \rightarrow R_2, \end{array} \quad \begin{array}{l} \begin{bmatrix} 2i & 2 \\ -2 & 2i \end{bmatrix} \\ \begin{bmatrix} 1 & -i \\ -2 & -2i \end{bmatrix} \\ \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix} \end{array}$$

and the top equation says  $x_1 = ix_2$ , so the null space is

$$\left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} x_2 \mid x_2 \in \mathbb{R} \right\}$$

so the eigenvector is  $[i \ 1]^\top$ .

Notice in the last example that there were complex conjugate pairs of eigenvalues and associated eigenvectors. This always occurs with a real matrix  $A$ . The following theorem summarizes this.

Let  $A$  be a real matrix. If  $\lambda$  is a complex eigenvalue with corresponding eigenvector,  $\vec{v}$ , then the complex conjugate of  $\lambda$ , denoted  $\bar{\lambda}$  is also an eigenvalue with corresponding complex conjugate eigenvector  $\bar{\vec{v}}$ .

*Proof.* If  $\lambda, \vec{v}$  is an eigenvalue-eigenvector pair of  $A$ , then

$$A\vec{v} = \lambda\vec{v}$$

And taking the complex conjugate of this equation.

$$\begin{aligned} \overline{A\vec{v}} &= \overline{\lambda\vec{v}} \\ A\bar{\vec{v}} &= \bar{\lambda}\bar{\vec{v}} \end{aligned}$$

where  $\bar{A} = A$  because  $A$  is real. This shows that  $\bar{\lambda}, \bar{\vec{v}}$  is a eigenvalue-eigenvector pair of  $A$ .

□

The result of this theorem will save work if you have a real matrix and find a complex eigenvalue. For example in Example ??, the first eigenvalue-eigenvector pairs was  $\lambda_1 = 2i$  and  $\vec{v}_1 = [-i \ 1]^\top$ . The above theorem shows that  $\lambda_2 = \overline{\lambda_1} = -2i$  and  $\vec{v}_2 = \overline{\vec{v}_1} = [i \ 1]^\top$  is another eigenvalue-eigenvector pair.

The next example has real eigenvalue and eigenvector, but has a different feature than Example ??.

Find the eigenvalues and eigenvectors of

$$\begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix}$$

The eigenvalues are found by solving the characteristic equation

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 2 - \lambda & -1 \\ 1 & 4 - \lambda \end{vmatrix} = (2 - \lambda)(4 - \lambda) + 1 \\ &= \lambda^2 - 6\lambda + 9 \end{aligned}$$

And solving  $|A - \lambda I| = 0$  has the single root  $\lambda = 3$ . To find the associated eigenvector, we solve for the null space of  $A - 3I$

$$\begin{aligned} A - 3I &= \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \\ R_1 + R_2 &\rightarrow R_2 \quad \begin{bmatrix} -1 & -1 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

which has the only equation  $-x_1 = x_2$  or  $x_1 = -x_2$  which shows that the null space is

$$\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} x_2 \mid x_2 \in \mathbb{R} \right\}$$

and the eigenvector is  $\vec{v} = [-1 \ 1]^\top$ .

And this next example has only one eigenvalue however has two linearly independent eigenvectors.

Find the eigenvalues and eigenvectors of

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

This has the characteristic equation  $(\lambda - 2)^2 = 0$  or the single root  $\lambda = 2$ .

To find the eigenvectors we find the null space of  $A - 2I$  or

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and any vector in  $\mathbb{R}^2$  is in the null space. We can write down the space using the standard basis vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} x_2 \mid x_1, x_2 \in \mathbb{R} \right\}$$

so  $\vec{v}_1 = [1 \ 0]^\top$  and  $\vec{v}_2 = [0 \ 1]^\top$  are both eigenvectors associated with  $\lambda = 2$ .

**Possible Eigenvalue/Eigenvector combinations of a  $2 \times 2$  matrix**

In the previous examples, we have seen the following:

- 2 real eigenvalues, 2 linearly independent eigenvectors
- 2 complex eigenvalues (complex conjugates) and 2 linearly independent eigenvectors (complex conjugates).
- 1 real eigenvalue, 2 linearly independent eigenvectors
- 1 real eigenvalue, 1 eigenvector.

In fact, these are the possibilities for any 2 by 2 real matrix. We now examine two examples of the eigenvalues/eigenvectors of 3 by 3 matrices.

Find the eigenvalues and eigenvectors of the 3 by 3 matrix

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

To find the eigenvalues, we solve the characteristic equation  $|A - \lambda I| = 0$  or expanding across the first row to find the determinant and we solve

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 1 - \lambda & 0 & 0 \\ 3 & -2 - \lambda & 6 \\ 2 & 1 & 3 - \lambda \end{vmatrix} = (1 - \lambda) \begin{vmatrix} -2 - \lambda & 6 \\ 1 & 3 - \lambda \end{vmatrix} \\ &= (1 - \lambda)((-2 - \lambda)(3 - \lambda) - 6) = 0 \end{aligned}$$

which has roots  $\lambda = 1, 4, -3$ .

To find the corresponding eigenvectors, we solve for the null space of  $A - \lambda I$  for each  $\lambda$

$$\begin{aligned} A - I &= \begin{bmatrix} 0 & 0 & 0 \\ 3 & -3 & 6 \\ 2 & 1 & 2 \end{bmatrix} \\ R_2 \leftrightarrow R_1 \\ R_3 \leftrightarrow R_2 & \begin{bmatrix} 3 & -3 & 6 \\ 2 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \\ \frac{1}{3}R_1 \rightarrow R_1 & \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \\ -2R_1 + R_2 \rightarrow R_2 & \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & -2 \\ 0 & 0 & 0 \end{bmatrix} \\ R_2 + 3R_1 \rightarrow R_1 & \begin{bmatrix} 3 & 0 & 4 \\ 0 & 3 & -2 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

and other that a factor of 3, this is in reduced row echelon form. The top two equations are

$$\begin{aligned}x_1 &= -\frac{4}{3}x_3 \\x_2 &= \frac{2}{3}x_3\end{aligned}$$

and the null space is

$$\begin{bmatrix} -4/3 \\ 2/3 \\ 1 \end{bmatrix} x_3 \mid x_3$$

And we could take  $[-4/3 \ 2/3 \ 1]^\top$  as an eigenvector, however, from Lemma ??, a scalar multiple of an eigenvector is also an eigenvector, so multiplying by 3 results in  $\vec{v}_1 = [-4 \ 2 \ 3]^\top$ .

$$\lambda = 4$$

$$\begin{array}{l} A - 4I = \begin{bmatrix} 3 & 0 & 0 \\ 3 & -6 & 6 \\ 2 & 1 & -1 \end{bmatrix} \\ \frac{1}{3}R_1 \rightarrow R_1 \quad \begin{bmatrix} 1 & 0 & 0 \\ 3 & -6 & 6 \\ 2 & 1 & -1 \end{bmatrix} \\ \begin{array}{l} -3R_1 + R_2 \rightarrow R_2 \\ -2R_1 + R_3 \rightarrow R_3 \end{array} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & -6 & 6 \\ 0 & 1 & -1 \end{bmatrix} \\ -\frac{1}{6}R_2 \rightarrow R_2 \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix} \\ -R_2 + R_3 \rightarrow R_3 \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

and the top two equations are  $x_1 = 0$  and  $x_2 = x_3$  so the null space can be written

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} x_3 \mid x_3 \in \mathbb{R} \right\}$$

so the eigenvectors is  $[0 \ 1 \ 1]^\top$ .

$$\lambda = -3$$

$$\begin{array}{rcl}
 A + 3I & = & \begin{bmatrix} 4 & 0 & 0 \\ 3 & 1 & 6 \\ 2 & 1 & 6 \end{bmatrix} \\
 \frac{1}{4}R_1 \rightarrow R_1 & & \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 6 \\ 2 & 1 & 6 \end{bmatrix} \\
 \begin{array}{l} -3R_1 + R_2 \rightarrow R_2 \\ -2R_1 + R_3 \rightarrow R_3 \end{array} & & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 6 \\ 0 & 1 & 6 \end{bmatrix} \\
 -R_2 + R_3 \rightarrow R_3 & & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 6 \\ 0 & 0 & 0 \end{bmatrix}
 \end{array}$$

and the top two equations are  $x_1 = 0$ ,  $x_2 = -6x_3$  so the null space is

$$\left\{ \begin{bmatrix} 0 \\ -6 \\ 1 \end{bmatrix} x_3 \mid x_3 \in \mathbb{R} \right\}$$

Therefore the eigenvector is  $\vec{v}_3 = [0 \ -6 \ 1]^T$ .

Find the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \\ -1 & 0 & -2 \end{bmatrix}$$

First, we solve for the eigenvectors by solving  $|A - \lambda I| = 0$ ,

$$\begin{aligned}
 |A - \lambda I| &= \begin{vmatrix} 1-\lambda & 0 & 2 \\ 2 & -\lambda & 4 \\ -1 & 0 & -2-\lambda \end{vmatrix} \\
 &= -\lambda \begin{vmatrix} 1-\lambda & 2 \\ -1 & -2-\lambda \end{vmatrix} = -\lambda((1-\lambda)(-2-\lambda) + 2) \\
 &= -\lambda(\lambda^2 - \lambda - 2 + 2) = -\lambda^3 - \lambda^2
 \end{aligned}$$

and this is 0, when  $\lambda = 0$  and  $\lambda = -1$ .

Next, find the eigenvectors. The eigenvectors with  $\lambda = 0$  are found by finding the null space of  $A - 0I$ :

$$\begin{array}{rcl}
 & & \begin{bmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \\ -1 & 0 & -2 \end{bmatrix} \\
 \begin{array}{l} -2R_1 + R_2 \rightarrow R_2 \\ R_1 + R_3 \rightarrow R_3 \end{array} & & \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
 \end{array}$$

and since the only equation is  $x_1 + 2x_3 = 0$ , both  $x_2$  and  $x_3$  are free variables and the solution space (therefore the null space) can be written:

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} x_3 \mid x_2, x_3 \in \mathbb{R} \right\}$$

This shows that the vectors

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

are both eigenvectors associated with  $\lambda = 0$ . The eigenvectors associated with  $\lambda = -1$  is found by seeking the null space of  $A - I$ :

$$\begin{array}{l} \frac{1}{2}R_1 \rightarrow R_1 \\ -2R_1 + R_2 \rightarrow R_2, \\ R_1 + R_3 \rightarrow R_3, \end{array} \quad \begin{array}{l} \begin{bmatrix} 2 & 0 & 2 \\ 2 & 1 & 4 \\ -1 & 0 & -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 4 \\ -1 & 0 & -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

These equations are

$$\begin{aligned} x_1 &= -x_3 \\ x_2 &= -2x_3 \end{aligned}$$

so the solution set (and the null space) is

$$\left\{ \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} x_3 \mid x_3 \in \mathbb{R} \right\}$$

so the eigenvector associated with  $\lambda_3 = -1$  is

$$\begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$$

### 3.2 Eigenvalues and Linear Maps

Recall that in section ?? that the eigenvalues and eigenvectors of a square matrix was found. We will examine how we can find eigenvalues and eigenvectors of a linear map. The definition of these are extensions of what we saw for matrices.



Let  $T : V \rightarrow V$  be a linear map and  $V$  be a finite dimensional vector space. The nonzero vector  $\vec{x}$  is an **eigenvector** with associated **eigenvalue** if

$$T(\vec{x}) = \lambda \vec{x}$$

There are a few ways to find  $\vec{x}$  and  $\lambda$ . In this section, we'll only see some examples that are relatively simple to see as well as finding the matrix representation of the map. In addition, this section only shows examples from finite dimensional vector spaces, however, in general, there is no restriction.

For the remainder of this chapter we will see examples of eigenvalues and eigenvectors of linear maps, including rotations, scales and derivative maps. We first see an example of scaling a vector in  $\mathbb{R}^2$ .

Find the eigenvalues and eigenvectors of the scale map  $S$  from Example ??.

Recall that the scale map  $S : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is given by

$$S(\vec{x}) = k\vec{x}$$

To find the eigenvalues and eigenvectors of  $S$ , we seek an  $\vec{x}$  and a  $\lambda$  such that

$$S(\vec{x}) = \lambda \vec{x}$$

but since  $S(\vec{x}) = k\vec{x}$ , then  $\lambda = k$  and any  $\vec{x}$  is an eigenvector.

Alternatively, we can write down the matrix  $A_S$  associated with the map. This was done in Example ?? and is

$$A_S = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$$

The eigenvalues and eigenvectors of this was found in Example ?? for a particular  $k$ , but generalizing that, one can see that  $\lambda = k$  will be the only eigenvalue of  $A_S$  with eigenvectors  $[1 \ 0]^\top$  and  $[0 \ 1]^\top$ .

In this case, since there are two eigenvectors associated with  $\lambda = k$ , any linear combination of the two eigenvectors is also an eigenvector, and since  $[1 \ 0]^\top$  and  $[0 \ 1]^\top$  span  $\mathbb{R}^2$ , any vector in  $\mathbb{R}^2$  is an eigenvector.

The next example shows the eigenvalues of the linear map associated with a derivative.

The set

$$V = \{ae^x + be^{-x} \mid a, b \in \mathbb{R}\}$$

is a subspace of all functions on  $\mathbb{R}$ . A basis for this subspace is  $(e^x, e^{-x})$ . In addition, the differential operation  $D : V \rightarrow V$  is a linear transformation. What are the eigenvalues and eigenvectors of  $D$ ?

There are two ways of looking at this. Since  $e^x \mapsto e^x$ , this means that  $e^x$  is an eigenvector with corresponding eigenvalue 1. Similarly, since  $e^{-x} \mapsto -e^{-x}$ , this also means that  $e^{-x}$  is an eigenvector with eigenvalue  $-1$ .

Alternatively, this can be done by first finding the matrix representation of the differential operator or

$$A_D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Recall that in the case of diagonal matrices, the eigenvalues are on the diagonal or  $\lambda_1 = 1$  and  $\lambda_2 = -1$ . One can also find that the corresponding eigenvectors are

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

These two vectors can be translated back to the functional forms  $e^x$  and  $e^{-x}$ , the same as we found above. These two functions which are in  $V$  are functions that stay the same in the subspace  $V$  up to a scalar constant.

Consider the differential map  $D : \mathcal{P}_3 \rightarrow \mathcal{P}_3$  which maps cubic polynomials to other cubic polynomials.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Find the eigenvalues of this matrix and interpret.

First, we find the determinant of

$$|A - \lambda I| = \begin{vmatrix} -\lambda & 1 & 0 & 0 \\ 0 & -\lambda & 2 & 0 \\ 0 & 0 & -\lambda & 3 \\ 0 & 0 & 0 & -\lambda \end{vmatrix} = (-\lambda)^4$$

and setting it to zero, thus  $\lambda = 0$  is the only eigenvalue. To find the eigenvectors, we find the null space of the original matrix, which after scaling the second and third row, the

matrix is the reduced row echelon form:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and the solution to the null space is

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} s \mid s \in \mathbb{R} \right\}$$

This vector is the representation of the cubic polynomial  $p(x) = c$ , a constant. Thus, the only vector that remains the same under differentiation is the constant polynomial with eigenvalue 0.

For the last example in this section, consider the map that rotates a 2 by 2 matrix  $90^\circ$  clockwise. That is  $R : \mathcal{M}_{2 \times 2} \rightarrow \mathcal{M}_{2 \times 2}$  such that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} c & a \\ d & b \end{bmatrix}$$

One can show that if we consider the vector representation in the basis:

$$B = \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

that

$$\text{Rep}_B \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

and the map  $R$  can be represented by the matrix as

$$A_R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

We now find the eigenvalues and eigenvectors of this. The eigenvalue-eigenvector pairs

are

$$\begin{array}{llll} \lambda_1 = 1 & \vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \lambda_2 = -1 & \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \\ \lambda_3 = i & \vec{v}_3 = \begin{bmatrix} 1 \\ -i \\ i \\ -1 \end{bmatrix} & \lambda_4 = -i & \vec{v}_4 = \begin{bmatrix} 1 \\ i \\ -i \\ -1 \end{bmatrix} \end{array}$$

To translate this back to the map that rotates the matrix, we translate each of the eigenvectors to the matrix that it represents. For example,  $\vec{v}_1$  is the matrix

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

and if that matrix is rotated, you get it back and the eigenvector is 1. The second eigenvector can be written as the matrix

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and if you rotate this matrix, you get the matrix

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

which is the above matrix multiplied by the eigenvalue  $\lambda_2 = -1$ . In other words:

$$R\left(\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right) = -\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

The other two work in a similar manner, however complex numbers are needed.

### 3.3 Applications of Eigenvalues and Eigenvectors

We will develop a simple model of population growth in this section and show an example where eigenvalues and eigenvectors play an important role.

Consider a population of a single species, perhaps a bacteria, which varies over time. Let  $p_n$  be the population at the  $n$ th time unit (day, month, year, ...). A simple model will be that the population at the next time will be proportional to the current population, thus  $p_{n+1} = rp_n$ .

If the number of bacteria starting is 50, then  $p_0 = 50$  and the population doubles every time unit, so  $r = 2$ , then

$$p_1 = 2p_0 = 100$$

$$p_2 = 2p_1 = 200$$

$$p_3 = 2p_2 = 400$$

and so on. In general

$$\begin{aligned} p_1 &= rp_0 \\ p_2 &= rp_1 = r^2 p_0 \\ p_3 &= rp_2 = r^3 p_0 \\ &\vdots \\ p_n &= r^n p_0 \end{aligned}$$

From this model, the population increases without bound if  $r > 1$ , drops to zero if  $r < 1$  and if  $r = 1$ ,  $p_n = p_0$ .

### 3.3.1 Leslie Population Models

The previous model actually works well for species like single-celled organisms, however, more advanced species will require a more complicated model. Consider a species, like some type of bird, that has two stages of life: juvenile and adult. We will also assume that the population of males and females are equal. The juvenile female cannot reproduce and the adult female can. Let  $p_n$  and  $q_n$  be the population of the juveniles and adults females at stage  $n$ . A simple model for growth is

$$\begin{aligned} p_{n+1} &= kq_n \\ q_{n+1} &= \alpha p_n + \beta q_n \end{aligned}$$

where at each stage the fraction  $\alpha$  of the juveniles survive to adulthood, and the population of the adults will produce a factor  $k$  juveniles. Finally, the fraction  $\beta$  of the adults survive one time unit. This linear system can be written as the matrix equation

$$\vec{p}_{n+1} = A\vec{p}_n$$

where

$$\vec{p} = \begin{bmatrix} p \\ q \end{bmatrix} \quad A = \begin{bmatrix} 0 & k \\ \alpha & \beta \end{bmatrix}$$

As a specific example, let's consider

$$A = \begin{bmatrix} 0 & 2 \\ 0.3 & 0.5 \end{bmatrix}$$

The first line means that for every adult female, two juvenile females are produced. The second line means that 0.3 (30 percent) of the juveniles survive to adulthood and at each stage 0.5 (50 percent) of the adults survive to the next stage.

Let's see what happens to this population for some initial condition. Let's say that we have 10 female adults and 0 juvenile adults. Then

$$\vec{p}_1 = A\vec{p}_0 = \begin{bmatrix} 0 & 2 \\ 0.3 & 0.5 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \end{bmatrix} = \begin{bmatrix} 20 \\ 5 \end{bmatrix}$$

In general,

$$\vec{p}_n = A^n \vec{p}_0$$

year	$p_n$	$q_n$	$T_n$	$p_n/q_n$	$T_{n+1}/T_n$
0	0	10	10	0	
1	20	5	25	4.00	2.50
2	10	8	18	1.18	0.74
3	17	7	24	2.34	1.31
4	14	8	22	1.66	0.96
5	17	8	25	2.00	1.13
$\vdots$					
10	22	12	34	1.87	1.06
11	24	12	36	1.88	1.07
12	25	13	38	1.88	1.06

We'd like to explain what is going on here. As we will show, the eigenvalues are important. The characteristic equation is

$$\lambda^2 - 0.5\lambda - 0.6 = 0$$

which has the solutions

$$\lambda = \frac{0.5 \pm \sqrt{0.25 + 2.4}}{2} \approx 1.06, -0.56$$

The eigenvectors are

$$\vec{v}_1 = \begin{bmatrix} 1.88 \\ 1 \end{bmatrix} \quad \vec{v}_2 = \begin{bmatrix} -3.54 \\ 1 \end{bmatrix}$$

Assume that  $\lambda_1$  and  $\lambda_2$  are the eigenvalues of  $A$  and  $\vec{v}_1$  and  $\vec{v}_2$  are the eigenvectors. If  $\vec{v}_1$  and  $\vec{v}_2$  are linearly independent, then we can write

$$\vec{p}_0 = a_1 \vec{v}_1 + a_2 \vec{v}_2$$

for some real numbers  $a_1$  and  $a_2$ . Let's use the matrix equations above to find  $\vec{p}_1$ ,

$$\begin{aligned} \vec{p}_1 &= A\vec{p}_0 = A(a_1 \vec{v}_1 + a_2 \vec{v}_2) \\ &= a_1 A\vec{v}_1 + a_2 A\vec{v}_2 \\ &= a_1 \lambda_1 \vec{v}_1 + a_2 \lambda_2 \vec{v}_2 \end{aligned}$$

similarly,

$$\begin{aligned} \vec{p}_2 &= A\vec{p}_1 = A(a_1 \lambda_1 \vec{v}_1 + a_2 \lambda_2 \vec{v}_2) \\ &= a_1 \lambda_1 A\vec{v}_1 + a_2 \lambda_2 A\vec{v}_2 \\ &= a_1 \lambda_1^2 \vec{v}_1 + a_2 \lambda_2^2 \vec{v}_2 \end{aligned}$$

Now assume that  $|\lambda_1| > |\lambda_2|$  and factor  $\lambda_1^n$  out from the equation above:

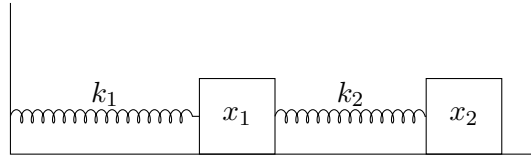
$$\vec{p}_n = \lambda_1^n \left( a_1 \vec{v}_1 + a_2 \frac{\lambda_2^n}{\lambda_1^n} \vec{v}_2 \right)$$

as  $n$  increases the ratio  $(\lambda_2/\lambda_1)^n$  goes to zero, so

$$\vec{p}_n \approx a_1 \lambda_1^n \vec{v}_1$$

### 3.3.2 Second-Order Differential Equations Systems

Consider a mass-spring system:



where  $x_1$  and  $x_2$  are the positions of the masses  $m_1$  and  $m_2$  relative to the unstretched spring positions and  $k_1$  and  $k_2$  are the spring constants of the two springs.

The mass spring system above can be modeled using the 2nd order system of differential equations.

$$\begin{aligned} m_1 x_1'' &= -k_1 x_1 + k_2 (x_2 - x_1) \\ m_2 x_2'' &= -k_2 (x_2 - x_1). \end{aligned}$$

This can be written in matrix form as

$$M \vec{x}'' = K \vec{x}$$

with

$$M = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \quad K = \begin{bmatrix} -(k_1 + k_2) & k_2 \\ -k_2 & k_1 \end{bmatrix}$$

where  $M$  is the mass matrix and  $K$  is called the stiffness matrix, which can take other forms depending on the connectedness of the spring system.

Since  $M$  is nonsingular, we can find  $M^{-1}$  and write the system above as:

$$\vec{x}'' = A \vec{x}$$

where  $A = M^{-1}K$ .

### 3.3.3 Solutions of $\vec{x}'' = A\vec{x}$

Let  $v$  be an eigenvector of  $A$ , with eigenvalue  $\lambda = \alpha^2$ , then  $\vec{x} = \vec{v}(c_1 e^{\alpha t} + c_2 e^{-\alpha t})$  is a solution.

$$\begin{aligned}\vec{x}'' &= \alpha^2 \vec{v}(c_1 e^{\alpha t} + c_2 e^{-\alpha t}) \\ &= A\vec{v}(c_1 e^{\alpha t} + c_2 e^{-\alpha t})\end{aligned}$$

if  $\lambda < 0$ , then the  $\lambda = -\omega^2$  and the solution is:

$$\vec{x} = \vec{v}(a \cos \omega t + b \sin \omega t)$$

If  $\lambda = 0$  is an eigenvalue of  $A$  with corresponding eigenvector  $\vec{v}_0$ , then the part of the solution associate with this is:

$$\vec{x}(t) = (a_0 + b_0 t)\vec{v}_0$$

If the  $n \times n$  matrix  $A$  has distinct negative eigenvalues  $-\omega_1^2, -\omega_2^2, \dots, -\omega_n^2$  with associated real eigenvectors, then a general solution to

$$\vec{x}'' = A\vec{x}$$

is given by

$$\vec{x} = \sum_{i=1}^n (a_i \cos \omega_i t + b_i \sin \omega_i t) \vec{v}_i$$

where  $a_i$  and  $b_i$  arbitrary constants.

If  $\lambda = 0$  is an eigenvalue of  $A$  with corresponding eigenvector  $\vec{v}_0$ , then the part of the solution associate with this is:

$$\vec{x}(t) = (a_0 + b_0 t)\vec{v}_0$$

Consider the mass-spring system above with  $m_1 = 1$ ,  $m_2 = 2$ ,  $k_1 = 1$ , and  $k_2 = 2$ , then the mass matrix and stiffness matrix are:

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad K = \begin{bmatrix} -3 & 2 \\ -2 & 4 \end{bmatrix}$$

resulting in the system:

$$\vec{x}'' = \begin{bmatrix} -3 & 2 \\ -1 & 2 \end{bmatrix} \vec{x}$$

The eigenvalues and eigenvectors are:

$$\begin{aligned}\lambda_1 &= -4 & \vec{v}_1 &= (-2, 1)^T \\ \lambda_2 &= -1 & \vec{v}_2 &= (1, 1)^T\end{aligned}$$



Therefore the solution is:

$$\vec{x} = (a_1 \cos 2t + b_1 \sin 2t) \begin{bmatrix} -2 \\ 1 \end{bmatrix} + (a_2 \cos t + b_2 \sin t) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (3.3)$$

Next, let's find the values of the  $a$ 's and  $b$ 's if  $x_1 = 1$  and  $x_2 = 2$  when  $t = 0$  or

$$\vec{x}(0) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

and  $\vec{x}'(0) = \vec{0}$ . First use the initial value  $\vec{x}(0)$  in (??),

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} = a_1 \begin{bmatrix} -2 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and this is true when  $a_1 = 1/3$  and  $a_2 = 5/3$ . To find  $b_1$  and  $b_2$ , differentiate (??),

$$\vec{x}'(t) = (-2a_1 \sin 2t + 2b_1 \cos 2t) \begin{bmatrix} -2 \\ 1 \end{bmatrix} + (-a_2 \sin t + b_2 \cos t) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

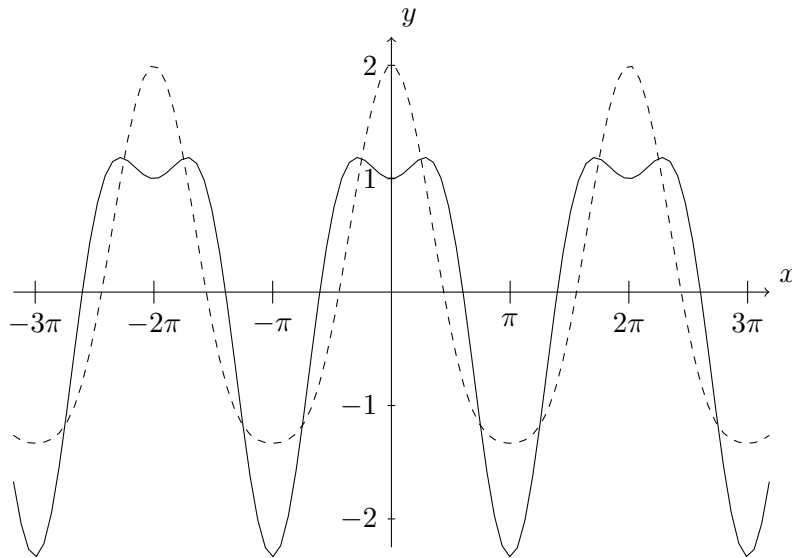
and evaluate it at  $t = 0$

$$\vec{x}'(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 2b_1 \begin{bmatrix} -2 \\ 1 \end{bmatrix} + 2b_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

or  $b_1 = 0$  and  $b_2 = 0$ . The solution with the given initial condition,

$$\vec{x}(t) = \frac{1}{3} \cos 2t \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \frac{5}{3} \cos t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The value of  $x_1$  and  $x_2$ , the position of the two masses are the first and second components. A plot of these two are:





## Chapter 4

# Matrix Factoring

Matrices have appeared throughout previous chapters and has many uses. In this chapter, we will take a matrix  $A$  and write it as the product of two or more matrices. In each case, we will show the purpose of doing this. In each case, the factored matrices will have desirable properties. The first factorization technique that we will examine is called LU factorization, which writes  $A = LU$ , where  $L$  is a lower triangular matrix and  $U$  is upper triangular. This factorization will be helpful for solving  $A\mathbf{x} = \mathbf{b}$  repeatedly and can be helpful for finding the inverse of a matrix. We will also see the Singular Value Decomposition (SVD) of a matrix which is quite useful in finding matrices that are close to the original matrix with a smaller rank. We'll explain why this is important with some nice applications.

### 4.1 Matrix Properties and Operations

Although we have spent some time with matrices earlier in the text, we summarize these and they will be important throughout this chapter.

#### 4.1.1 The Matrix Transpose

Let  $A$  be an  $m$  by  $n$  matrix. The matrix transpose of  $A$  is denoted  $A^T$  and

$$(A^T)_{i,j} = A_{j,i}$$

or in other words, the  $i$ th row of the transpose is the  $j$ th column of  $A$ .

If

$$A = \begin{bmatrix} 1 & -3 & 7 \\ 2 & 0 & 5 \end{bmatrix}$$

then

$$A^T = \begin{bmatrix} 1 & 2 \\ -3 & 0 \\ 7 & 5 \end{bmatrix}$$

The matrix  $A^\top$  is the matrix  $A$  flipped over the diagonal.

(thinking of the transpose in a more abstract manner)

[Properties of the matrix transpose]

The following are properties of the transpose. Let  $A$  be an  $m$  by  $n$  matrix (unless more restricted as noted).

1.  $(A^\top)^\top = A$ . That is, the transpose of the transpose is the original matrix back.

2.  $(A + B)^\top = A^\top + B^\top$ .

This is one of the properties of linearity.

3.  $(AB)^\top = B^\top A^\top$ .

4.  $(cA)^\top = cA^\top$ .

And this is the other linearly property.

If, in addition, the matrix  $A$  is square, then

5.  $(A^{-1})^\top = (A^\top)^{-1}$ .

6. The eigenvalues of  $A$  are the same as the eigenvalues of  $A^\top$ . Also, the characteristic polynomial of  $A$  and  $A^\top$  are the same.

#### 4.1.2 Symmetric Matrices

We have seen throughout this book examples where a matrix is symmetric in some sense. However, there is a precise definition.

A matrix is symmetric if it is square and if  $A = A^\top$ .

The following matrix is symmetric,

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 3 & -4 \\ 0 & -4 & -5 \end{bmatrix}$$

and the following is not

$$A = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 3 & 1 \\ -3 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

since it is not square.

Notice that a symmetric matrix has the property that it is symmetric over the main diagonal (running from the upper left to lower right).

### Eigenvalues of Symmetric Matrices

If we follow the techniques in ??, we can find the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

then the eigenvalues are  $\lambda = 8, -1, -1$  and the corresponding eigenvectors are

$$\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \quad \vec{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad \vec{v}_3 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$$

which are both real eigenvalues and eigenvectors. This fact is (as shown in the next lemma) that the matrix is symmetric.

Let  $A$  be a real symmetric matrix. The eigenvalues and eigenvectors of  $A$  are real.

*Proof.* Let  $\lambda$  be an eigenvalue of  $A$  with corresponding eigenvector  $\vec{v}$ , then

$$A\vec{v} = \lambda\vec{v} \tag{4.1}$$

take the complex conjugate of both sides

$$\overline{A\vec{v}} = \overline{\lambda\vec{v}}$$

and since  $A$  is real

$$A\vec{v} = \lambda\vec{v} \tag{4.2}$$

If we left-multiply (??) by  $\vec{v}^\top$ ,

$$\vec{v}^\top A\vec{v} = \vec{v}^\top (\lambda\vec{v}) = \lambda\vec{v}^\top \vec{v}$$

and solving for  $\lambda$ ,

$$\lambda = \frac{\vec{v}^\top A\vec{v}}{\vec{v}^\top \vec{v}} \tag{4.3}$$

and left-multiply (??) by  $\vec{v}^\top$  to get:

$$\vec{v}^\top A\vec{v} = \vec{v}^\top (\lambda\vec{v})$$

and if we take the transpose of both sides of this

$$\begin{aligned} (\vec{v}^\top A\vec{v})^\top &= (\vec{v}^\top \lambda\vec{v})^\top \\ \vec{v}^\top A^\top \vec{v} &= \overline{\lambda\vec{v}^\top} \vec{v} \end{aligned}$$

and since  $A$  is symmetric,  $A = A^\top$  and then solving for  $\bar{\lambda}$

$$\bar{\lambda} = \frac{\bar{\vec{v}}^\top A \vec{v}}{\bar{\vec{v}}^\top \vec{v}} \quad (4.4)$$

and since (??) is the same as (??), then  $\lambda = \bar{\lambda}$  so the eigenvalue is real.

Since the eigenvalue of a symmetric matrix is real and the matrix  $A$  is real, then the eigenvectors are also real. □

The fact that symmetric matrices have real eigenvalues will be used later in this section.

### 4.1.3 Orthogonal Matrices

An **orthogonal matrix**,  $Q$  is a square matrix that satisfies  $QQ^\top = Q^\top Q = I$ .

The properties include:

- The columns of  $Q$  are orthogonal unit vectors.
- The rows of  $Q$  are orthogonal unit vectors.
- The inverse of  $Q$  is its transpose or  $Q^{-1} = Q^\top$ .
- The determinant of  $Q$  is 1 or  $-1$ . This can be found by

$$1 = \det I = \det(QQ^\top) = \det(Q) \det(Q^\top) = \det(Q)^2$$

So the determinant is 1 or  $-1$ .

The following are examples of 2 by 2 orthogonal matrices:

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

And the following are 3 by 3 orthogonal matrices:

$$\begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ -\frac{1}{\sqrt{5}} & 0 & \frac{2}{\sqrt{5}} \\ \frac{4}{3\sqrt{5}} & \frac{5}{3\sqrt{5}} & \frac{2}{3\sqrt{5}} \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The matrix on the right is an example of a permutation matrix. Every permutation matrix is an orthogonal matrix.

#### 4.1.4 Upper- and Lower-Triangular Matrices

An *upper triangle matrix* is a matrix (not-necessarily square) such that all elements below the diagonal are 0. A *lower triangular matrix* is a matrix such that all elements above the diagonal are 0. More precisely,

A **upper-triangular matrix** is an  $m$  by  $n$  matrix  $A$ , such that  $a_{i,j} = 0$  if  $j < i$ . A **lower triangular matrix** is an  $m$  by  $n$  matrix  $A$  such that  $a_{i,j} = 0$  if  $j > i$ .

The matrix  $L$  below is a lower triangular matrix and  $U$  is an upper triangular matrix:

$$L = \begin{bmatrix} 3 & 0 & 0 \\ 2 & -1 & 0 \\ 0 & 5 & 2 \end{bmatrix} \qquad U = \begin{bmatrix} 1 & -7 & -5 \\ 0 & 3 & 2 \\ 0 & 0 & -3 \end{bmatrix}$$

Hopefully that it's clear that a matrix in echelon form defined in Definition ?? is also upper triangular.

#### Back Substitution

An advantage to having an upper- or lower-triangular matrix is that solving the matrix is quite straightforward. In the case of a upper-triangular matrix,  $U$  the steps that solve  $U\vec{x} = \vec{b}$  is called **back substitution**. We will show the steps in an example.

If

$$U = \begin{bmatrix} 1 & -7 & -5 \\ 0 & 3 & 2 \\ 0 & 0 & -3 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 2 \\ -5 \\ 6 \end{bmatrix},$$

solve  $U\vec{x} = \vec{b}$ .

To illustrate back substitution, we write out the matrix equation as the corresponding equations:

$$\begin{aligned} x_1 - 7x_2 - 5x_3 &= 2 \\ 3x_2 + 2x_3 &= -5 \\ -3x_3 &= -6 \end{aligned}$$

and solving the 3rd equation above,  $x_3 = 2$ . This is then substituted into the 2nd equation to get:

$$3x_2 + 2(2) = -5$$

and solving leads to  $x_2 = -3$ . And then lastly,  $x_2$  and  $x_3$  are substituted into the first equation to get:

$$x_1 - 7(-3) - 5(2) = 2$$

and solving this leads to  $x_1 = 13$ . Thus the solution to  $U\vec{x} = \vec{b}$  is

$$\vec{x} = \begin{bmatrix} 13 \\ -3 \\ 2 \end{bmatrix}$$

In general, back substitution works the same way.

To solve the matrix equation  $U\vec{x} = \vec{b}$ , where  $U$  is an  $n$  by  $n$ , non-singular, upper-triangular matrix, first

$$x_n = \frac{b_n}{a_{nn}}$$

and assume that  $x_{j+1}, x_{j+2}, \dots, x_n$  are known. Then,

$$x_j = \frac{1}{a_{jj}} \left( b_j - \sum_{k=j+1}^n a_{jk}x_k \right) \quad \text{for } j = n-1, n-2, \dots, 1$$

and it is important that the steps are performed in reverse order. This technique is called **back substitution**.

### Forward Substitution

Similar to back substitution, if a matrix is in lower-diagonal form, then the matrix equation  $L\vec{y} = \vec{c}$  can be solved in a straightforward manner. We first look at an example:

Solve  $L\vec{y} = \vec{c}$  if

$$L = \begin{bmatrix} 3 & 0 & 0 \\ 2 & -1 & 0 \\ 0 & 5 & 2 \end{bmatrix} \quad \vec{c} = \begin{bmatrix} 9 \\ 8 \\ 2 \end{bmatrix}$$

We first write down the equations that correspond to these:

$$\begin{aligned} 3y_1 &= 9 \\ 2y_1 - y_2 &= 8 \\ 5y_2 + 2y_3 &= 2 \end{aligned}$$

Solving the first equation is  $y_1 = 3$ . Then substitute this into the second equation to get:

$$2(3) - y_2 = 8$$

which has the solution  $y_2 = -2$ . Lastly, we substitute these into the last equation to get:

$$5(-2) + 2y_3 = 2$$



which has the solution  $y_3 = 6$ . Thus the solution to the matrix equation is

$$\vec{y} = \begin{bmatrix} 3 \\ -2 \\ 6 \end{bmatrix}$$

Notice in this example, we solved the unknown's in the  $\vec{y}$  vector in order. This is because of the structure of the lower-triangular matrix and hence why this technique is called **forward substitution**.

To solve the matrix equation  $L\vec{y} = \vec{c}$ , where  $L$  is an  $n$  by  $n$ , nonsingular, lower-triangular matrix, first,

$$y_1 = \frac{c_1}{a_{11}}$$

and then assume that  $y_1, y_2, \dots, y_{j-1}$  are known. Then

$$y_j = \frac{1}{a_{jj}} \left( b_j - \sum_{k=1}^{j-1} a_{jk} x_k \right) \quad \text{for } j = 2, 3, \dots, n$$

and note that in this case, the steps are performed in order. This technique is called **forward substitution**.

## 4.2 LU Factorization

Let  $A$  be a square matrix. We seek to write  $A = LU$ , where  $L$  is a lower triangular matrix and  $U$  is an upper triangular matrix. The reason for doing such a factorization is often to solve  $A\vec{x} = \vec{b}$  for multiple right hand sides.

Recall (or is this somewhere else), than an *upper triangle matrix* is a matrix (not-necessarily square) such that all elements below the diagonal are 0. A *lower triangular matrix* is a matrix such that all elements above the diagonal are 0. More precisely,

The next example shows that there is a lower-triangular matrix  $L$  and an upper triangular matrix  $U$  whose product is the original matrix  $A$ . Later we will show where this arises.

Let

$$L = \begin{bmatrix} 3 & 0 & 0 \\ 2 & -1 & 0 \\ 0 & 5 & 2 \end{bmatrix} \quad U = \begin{bmatrix} 1 & -7 & -5 \\ 0 & 3 & 2 \\ 0 & 0 & -3 \end{bmatrix} \quad A = \begin{bmatrix} 3 & -2 & 1 \\ 6 & 0 & 4 \\ -6 & -8 & -7 \end{bmatrix}$$

Show that  $A = LU$  if

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -2 & -3 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 3 & -2 & 1 \\ 0 & 4 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

and using standard matrix multiplication:

$$A = \begin{bmatrix} 1(3) + 0(0) + 0(0) & 1(-2) + 0(4) + 0(0) & 1(1) + 0(2) + 0(1) \\ 2(3) + 1(0) + 0(0) & 2(-2) + 1(4) + 0(0) & 2(1) + 1(2) + 0(1) \\ -2(3) - 3(0) + 1(0) & -2(-2) - 3(4) + 1(0) & -2(1) - 3(2) + 1(1) \end{bmatrix} = \begin{bmatrix} 3 & -2 & 1 \\ 6 & 0 & 4 \\ -6 & -8 & -7 \end{bmatrix}$$

This is nice, but the goal of LU factorization is to take any matrix  $A$  and find matrices  $L$  and  $U$ . We first, show how to do this for a 3 by 3 matrix and then extend it.

### LU factorization of a 3 by 3 matrix

Consider a general 3 by 3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

It is interesting to note that if we define

$$B_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

then

$$B_{21}A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} - \frac{a_{12}a_{21}}{a_{11}} & a_{23} - \frac{a_{13}a_{21}}{a_{11}} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

And if

$$B_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{a_{31}}{a_{11}} & 0 & 1 \end{bmatrix}$$

then

$$B_{31}B_{21}A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} - \frac{a_{12}a_{21}}{a_{11}} & a_{23} - \frac{a_{13}a_{21}}{a_{11}} \\ 0 & a_{32} - \frac{a_{12}a_{31}}{a_{11}} & a_{33} - \frac{a_{13}a_{31}}{a_{11}} \end{bmatrix}$$

If we define this matrix as  $A^{(1)}$  and then repeat the steps.

$$B_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{a_{32}^{(1)}}{a_{22}^{(1)}} & 1 \end{bmatrix}$$

and then

$$B_{32}A^{(1)} = \begin{bmatrix} a_{11}^{(1)} & a_{21}^{(1)} & a_{31}^{(1)} \\ 0 & a_{22}^{(1)} & a_{32}^{(1)} \\ 0 & 0 & a_{33}^{(1)} - \frac{a_{23}^{(1)}a_{32}^{(1)}}{a_{22}^{(1)}} \end{bmatrix}$$

Notice that this matrix is an upper triangular matrix. We let  $U = B_{32}A^{(1)} = B_{32}B_{31}B_{21}A$ . If  $U$  has non zeros along its diagonal, then it is invertible and

$$A = LU$$

multiply through on the right by  $U^{-1}$

$$AU^{-1} = L$$

Use the steps above to find matrices  $L$  and  $U$  such that  $A = LU$  for

$$A = \begin{bmatrix} 3 & -2 & 1 \\ 6 & 0 & 4 \\ -6 & -8 & -7 \end{bmatrix}$$

As above,

$$B_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{6}{3} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

And

$$B_{21}A = \begin{bmatrix} 3 & -2 & 1 \\ 0 & 4 & 2 \\ -6 & -8 & -7 \end{bmatrix}$$

And if

$$B_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{-6}{3} & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

and

$$B_{31}B_{21}A = \begin{bmatrix} 3 & -2 & 1 \\ 0 & 4 & 2 \\ 0 & -12 & -5 \end{bmatrix}$$

And we will call this matrix  $A^{(1)}$ .

$$B_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{12}{4} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}$$

$$B_{32}A^{(1)} = \begin{bmatrix} 3 & -2 & 1 \\ 0 & 4 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

and we will call this  $U$ . And  $L$  is found by

$$L = AU^{-1} = \begin{bmatrix} 3 & -2 & 1 \\ 6 & 0 & 4 \\ -6 & -8 & -7 \end{bmatrix} \begin{bmatrix} 1/3 & 1/6 & -2/3 \\ 0 & 1/4 & -1/2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -2 & -3 & 1 \end{bmatrix}$$

### Extending LU factorization to invertible matrices

The 3 by 3 example above can be extended to larger square matrices. For a matrix  $A = \{a_{i,j}\}$ , if we define

$$B^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -a_{21}/a_{11} & 1 & 0 & 0 & \cdots & 0 \\ -a_{31}/a_{11} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 0 & 1 & \ddots & \vdots \\ -a_{n1}/a_{11} & 0 & \cdots & & & 1 \end{bmatrix}$$

and define

$$A^{(1)} = B^{(1)}A$$

will satisfy

$$a_{i1}^{(1)} = \begin{cases} a_{11} & i = 1 \\ 0 & i > 1 \end{cases}$$

If this process is continued with

$$B^{(k)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & & \\ 0 & 0 & 0 & 1 & 0 & \cdots & \vdots \\ 0 & 0 & 0 & -a_{k+1,k}^{(k-1)}/a_{k,k}^{(k-1)} & 1 & 0 & \\ 0 & 0 & 0 & -a_{k+2,k}^{(k-1)}/a_{k,k}^{(k-1)} & 0 & 1 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & -a_{n,k}^{(k-1)}/a_{k,k}^{(k-1)} & 0 & \cdots & 0 & 1 \end{bmatrix}$$

and define

$$A^{(k)} = B^{(k)} B^{(k-1)}$$

and this will have nonzero elements on the diagonal and zeros below it through column  $k + 1$ . If this is continued up to  $A^{(n-1)}$ , then

$$U = A^{(n-1)} = B^{(n-1)} B^{(n-2)} \dots B^{(1)} A$$

is an upper diagonal matrix and as before let  $L = U^{-1}A$ .

### Further Details

A keen eye will notice some possible problems with the above formulation for the  $LU$  factorization. If any of the terms  $a_{k,k}^{(k)} = 0$ , then the matrix  $B^{(k)}$  is not defined. There are examples of invertible matrices that do not have an  $LU$  factorization, however, what can be done is to find a permutation matrix  $P$  such that  $LUP = A$ . Details of this are not provided here.

### Any Square Matrix has a LU Decomposition

Additionally, an square matrix  $A$  can be written as  $A = LUP$ , where  $L$  is a lower-diagonal,  $U$  is upper-diagonal and  $P$  is a permutation matrix. **NEED A REFERENCE**

### Applications of LU Decomposition

The main application of  $LU$  decomposition is that of solving linear systems. Consider the matrix equation  $A\vec{x} = \vec{b}$  and assume that  $A = LU$  (that is the needed permutation matrix is  $P = I$ ).

1. Solve  $L\vec{y} = \vec{b}$  for  $\vec{y}$ .
2. Then solve  $U\vec{x} = \vec{y}$  for  $\vec{x}$ .

First, this works, because if  $\vec{y} = U\vec{x}$ , then

$$\begin{aligned} L\vec{y} &= \vec{b} \\ LU\vec{x} &= \vec{b} \\ A\vec{x} &= \vec{b} \end{aligned}$$

Solve the system  $A\vec{x} = \vec{b}$  using  $LU$ -decomposition for

$$A = \begin{bmatrix} 3 & -2 & 1 \\ 6 & 0 & 4 \\ -6 & -8 & -7 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 14 \\ 22 \\ -9 \end{bmatrix}$$

Recall that in Example ?? the  $LU$ -factorizations of the matrix  $A$  was found and is

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -2 & -3 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 3 & -2 & 1 \\ 0 & 4 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

To use this to solve  $A\vec{x} = \vec{b}$ , first we solve  $L\vec{y} = \vec{b}$  via forward substitution.

$$\begin{aligned} y_1 &= 14 \\ 2y_1 + y_2 &= 22 \end{aligned}$$

so

$$y_2 = 22 - 2y_1 = 22 - 28 = -6$$

and the last equation is

$$-2y_1 - 3y_2 + y_3 = -9$$

or

$$y_3 = -9 + 2(14) + 3(-6) = 1$$

So the solution to  $L\vec{y} = \vec{b}$  is  $\vec{y} = [14, -6, 1]^T$ . Next, we solve  $U\vec{x} = \vec{y}$  by back substitution,

$$\begin{aligned} x_3 &= 1 \\ 4x_2 + 2x_3 &= -6 \end{aligned}$$

or

$$x_2 = (-6 - 2(1))/4 = -2$$

and the first equation is

$$3x_1 - 2x_2 + x_3 = 14$$

or

$$\begin{aligned} 3x_1 &= 14 + 2(-2) - x_3 = 9 \\ x_1 &= 3 \end{aligned}$$

So the solution is

$$\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

You can see from the example above that once the matrix  $A$  factored that it is relatively simple to find the solutions to  $L\vec{y} = \vec{b}$  and  $A\vec{x} = \vec{y}$  and there are relatively few computations to perform.

In fact, this is true in general. In that if one actually finds the factorization of  $A$  and then solves this in the manner show that about 1/2 of the operations are done then solving  $A\vec{x} = \vec{b}$  directly, say by reducing the matrix to reduced row echelon form.

### Inverting a Matrix

The same idea can be used to find the inverse of  $A$  by repeatedly solving  $LU\vec{x} = \vec{b}$  by repeating this for  $\vec{b}$  the columns of the identity matrix. The example below shows this without all of the details of finding the factorization:

Find the inverse of the matrix in Example ??,

$$A = \begin{bmatrix} 3 & 3 & -1 \\ 2 & 1 & -3 \\ 0 & 2 & 5 \end{bmatrix}$$

using the LU-decomposition.

Following the steps above, the LU decomposition is

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2/3 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 3 & 3 & -1 \\ 0 & -1 & -7/3 \\ 0 & 0 & 1/3 \end{bmatrix}$$

and we will solve  $A\vec{x} = LU\vec{x} = \vec{e}_j$ , where  $\vec{e}_j$  is the  $j$ th column of the 3 by 3 identity matrix.

Solving  $LU\vec{x} = \vec{e}_1$  by solving  $L\vec{y}_1 = (1, 0, 0)^\top$  or

$$\vec{y}_1 = \begin{bmatrix} 1 \\ -2/3 \\ -4/3 \end{bmatrix}$$

then solve  $U\vec{x}_1 = \vec{y}_1$  for  $\vec{x}_1$  or

$$\vec{x}_1 = \begin{bmatrix} -11 \\ 10 \\ -4 \end{bmatrix}$$

Repeating this for  $\vec{b} = \vec{e}_2$ , first solving  $L\vec{y}_2 = (0, 1, 0)^\top$  or

$$\vec{y}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

then solve  $U\vec{x}_2 = \vec{y}_2$  for  $\vec{x}_2$  or

$$\vec{x}_2 = \begin{bmatrix} 17 \\ -15 \\ 6 \end{bmatrix}$$

and lastly, solve  $L\vec{y}_3 = (0, 0, 1)^\top$  or

$$\vec{y}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and solving  $U\vec{x}_3 = \vec{y}_3$  results in

$$\vec{x}_3 = \begin{bmatrix} 8 \\ -7 \\ 3 \end{bmatrix}$$

The inverse matrix is then the matrix with  $\vec{x}_j$  as the columns or

$$A^{-1} = \begin{bmatrix} -11 & 17 & 8 \\ 10 & -15 & -7 \\ -4 & 6 & 3 \end{bmatrix}$$

### 4.3 Diagonalization of Matrices

Consider a matrix  $A$ . There are many application where the power of the matrix,  $A^n$  is helpful. One such cases is at the end of this this chapter. One way to approach this is that the  $n$ th power is just the matrix product with itself  $n$  times or

$$\begin{aligned} A^2 &= AA \\ A^3 &= AAA \end{aligned}$$

and this can be extended to any positive integer power. However, finding the 50th power may not be practical.

To make this an easier task, let's assume that we can write  $A$  in the form:

$$A = PDP^{-1} \tag{4.5}$$

where  $D$  is a diagonal matrix and  $P$  is invertible. If this is possible, then

$$\begin{aligned} A^3 &= AAA \\ &= (PDP^{-1})(PDP^{-1})(PDP^{-1}) \\ &= PD^3P^{-1} \end{aligned}$$

and  $D^3$  is easy to find because it is a diagonal matrix.

If it's possible to factor  $A$  as in (??), then we call a matrix diagonalizable. Here's a formal definition.

An  $n$  by  $n$  matrix  $A$  is said to be **diagonalizable** if it can be written

$$A = PDP^{-1}$$



where  $D$  is a diagonal matrix and  $P$  is invertible.

Before explaining how starting with a matrix  $A$ , we can find matrices  $D$  and  $P$ , let's look at one that is factored in this way.

Let

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \quad D = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}$$

Show that  $A = PDP^{-1}$ .

First, recall that

$$P^{-1} = \frac{1}{|P|} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$$

and

$$PD = \begin{bmatrix} -1 & 5 \\ 1 & 10 \end{bmatrix}$$

then

$$PDP^{-1} = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$$

which shows in this particular example that  $A = PDP^{-1}$ .

An  $n$  by  $n$  matrix  $A$  is diagonalizable if and only if it has  $n$  linearly independent eigenvectors. The matrix  $P$  is a matrix of the eigenvectors and the matrix  $D$  is the diagonal matrix of eigenvalues.

*Proof.* Thus we want to show that

$$\begin{aligned} A &= PDP^{-1} & \text{or} \\ AP &= PD \end{aligned}$$

since  $P$  is invertible.

Let

$$P = [\vec{v}_1 \quad \vec{v}_2 \quad \cdots \quad \vec{v}_n] \quad D = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

where  $\vec{v}_i$  is the eigenvector associated with the eigenvalue  $\lambda_i$ .

$$\begin{aligned} AP &= [A\vec{v}_1 \quad A\vec{v}_2 \quad \cdots \quad A\vec{v}_n] \\ &= [\lambda_1\vec{v}_1 \quad \lambda_2\vec{v}_2 \quad \cdots \quad \lambda_n\vec{v}_n] \\ &= [\vec{v}_1\lambda_1 \quad \vec{v}_2\lambda_2 \quad \cdots \quad \vec{v}_n\lambda_n] = PD \end{aligned}$$

□

Is the matrix

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}$$

diagonalizable? If so, find  $P$  and  $D$ .

To check this, we need to find the eigenvalues and eigenvectors. First, find the eigenvalues by solving  $|A - \lambda I| = 0$  or  $\lambda^2 - 5\lambda + 6 = (\lambda - 3)(\lambda - 2) = 0$  or  $\lambda_1 = 2$  and  $\lambda_2 = 3$

The associated eigenvectors are

$$\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

(The steps to find these aren't shown, but follow the steps in section ??). Since there are 2 linearly independent eigenvectors, this vector is diagonalizable and

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Is the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 3 & -2 & 6 \\ 2 & 1 & 3 \end{bmatrix}$$

diagonalizable? If so, find  $P$  and  $D$ .

From example ??, we found that  $A$  has eigenvalues  $\lambda_1 = 1, \lambda_2 = 4, \lambda_3 = -3$ . It also has eigenvectors  $\vec{v}_1 = [-4 \ 2 \ 3]^T$ ,  $\vec{v}_2 = [0 \ 1 \ 1]^T$  and  $\vec{v}_3 = [0 \ -6 \ 1]^T$ . Therefore the matrices:

$$P = \begin{bmatrix} -4 & 0 & 0 \\ 2 & 1 & -6 \\ 3 & 1 & 1 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

satisfy  $A = PDP^{-1}$

### Powers of Diagonalizable Matrices

One main reason for writing a matrix in diagonalizable form is that powers of the matrix are easy to compute. Note that if  $A = P^{-1}DP$ , then

$$\begin{aligned} A^2 &= (PDP^{-1})(PDP^{-1}) \\ &= PD^2P^{-1} \end{aligned}$$

and by induction:

$$A^k = PD^kP^{-1}$$

Note that raising the diagonal matrix  $D$  to a power is a simple process .

Use the fact that  $A$  is diagonalizable to find

$$\begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}^5$$

From Example ??,  $A$  can be written  $A = PDP^{-1}$  with

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \qquad D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Then,  $A^5$  can be written

$$\begin{aligned} A^5 &= PD^5P^{-1} \\ &= \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}^5 \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 32 & 0 \\ 0 & 243 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} -179 & -211 \\ 422 & 454 \end{bmatrix} \end{aligned}$$

#### 4.3.1 Similar Matrices

Two  $n$  by  $n$  matrices  $A$  and  $B$  are said to be **similar** if there exists an invertible matrix  $S$  such that

$$B = S^{-1}AS$$

If  $A$  and  $B$  are similar  $n$  by  $n$  matrices, then the following are identical for the two matrices:

- Rank
- Determinant
- Trace

If  $A$  and  $B$  are similar  $n$  by  $n$  matrices, then  $A$  and  $B$  have the same characteristic equation and therefore have the same eigenvalues.

*Proof.* Let  $A$  satisfy the characteristic equation

$$\det(A - \lambda I) = 0$$

Since  $A$  and  $B$  are similar, then  $A = S^{-1}BS$ ,

$$\det(S^{-1}BS - \lambda I) = 0$$

Since  $S^{-1}S = I = S^{-1}IS$

$$\det(S^{-1}BS - \lambda S^{-1}IS) = 0$$

Factoring out a  $S^{-1}$  from the left and  $S$  from the right

$$\begin{aligned}\det(S^{-1}(B - \lambda S^{-1}IS)S) &= 0 \\ \det(S^{-1}) \det(B - \lambda I) \det(S) &= 0\end{aligned}$$

since  $\det(S^{-1}) \det(S) = \det(S^{-1}S) = \det(I) = 1$

$$\det(B - \lambda I) = 0$$

so the characteristic equation is identical, thus the eigenvalues are the same.  $\square$

Note: The eigenvectors are in general not the same in  $A$  and  $B$ .  
Show that

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 2 \\ 0 & 3 \end{bmatrix}$$

are similar matrices with

$$S = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}$$

First, using the formula for the inverse of a  $2 \times 2$  matrix,

$$S^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

and

$$S^{-1}AS = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 0 & 3 \end{bmatrix} = B$$

### 4.3.2 Similar and Diagonalizable Matrices

It appears that there is a connection to similar and diagonalizable matrices through at least their near identical formulas. Notice that  $A$  and  $B$  are similar if there exists an invertible matrix  $P$  such that

$$A = S^{-1}BS$$

and  $A$  is diagonalizable if there exists a  $P_1$  such that

$$A = P_1 D_1 P_1^{-1} \tag{4.6}$$

for a diagonal matrix  $D_1$ . Similarly  $B$  is diagonalizable if there exists an invertible  $P_2$  such that

$$B = P_2 D_2 P_2^{-1} \quad (4.7)$$

for diagonal matrix  $D_2$ .

We know that the matrices  $P_1$  and  $P_2$  exist if there are  $n$  linearly independent eigenvectors, but how do you find  $S$ ?

Solving for  $D_1$  in (??) and  $D_2$  in (??),

$$D_1 = P_1^{-1} A P_1 \qquad D_2 = P_2^{-1} B P_2$$

Also, since the eigenvalues of  $A$  and  $B$  are the same, we can rearrange the eigenvectors of  $A$  and  $B$  to the same order thus without loss of generality,  $D_1 = D_2$  and

$$P_1^{-1} A P_1 = P_2^{-1} B P_2$$

left multiplying by  $P_1$  and right multiplying by  $P_1^{-1}$

$$A = P_1 P_2^{-1} B P_2 P_1^{-1}$$

thus let

$$S = P_2 P_1^{-1}. \quad (4.8)$$

Above we showed that the matrices

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 2 \\ 0 & 3 \end{bmatrix}$$

are similar. Use the above discussion to find  $S$  such that  $A = S^{-1} B S$ .

Briefly, we need to diagonalize both  $A$  and  $B$ . In example ??, we found that the eigenvalues of  $A$  are  $\lambda_1 = 2$  and  $\lambda_2 = 3$  with associated eigenvectors  $\vec{v}_1 = [1 \ -1]^\top$  and  $\vec{v}_2 = [1 \ -2]^\top$ .

Using the techniques of section ??, the eigenvalues of  $B$  are  $\lambda_1 = 2$  and  $\lambda_2 = 3$  with associated eigenvectors  $\vec{v}_1 = [1 \ 4]^\top$  and  $\vec{v}_2 = [0 \ 1]^\top$ .

Letting  $D_1$  and  $P_1$  be the matrices associated with  $A$  and  $D_2$  and  $P_2$ , those associated with  $B$ , let

$$D_1 = D_2 = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

and

$$P_1 = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$$

and using (??),

$$S = P_2 P_1^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & -1 \end{bmatrix}$$

And although this  $S$  is not the same  $S$  as Example ??, this matrix  $S$  is the negative of the inverse of that matrix in Example ??.

## 4.4 Singular Value Decomposition

The singular value decomposition, also known as the SVD, is yet another way to factor a matrix. In short, a matrix  $A$  can be written  $A = U\Sigma V^\top$ , where  $U$  and  $V$  are orthogonal matrices and  $\Sigma$  is a diagonal matrix. We will show that it is possible to factor any  $m$  by  $n$  matrix  $A$  in this form and also determine how to find the matrices  $U$ ,  $\Sigma$  and  $V$ .

Consider a real  $m$  by  $n$  matrix  $A$ . We will consider  $U$ , an  $m$  by  $m$  matrix,  $\Sigma$ , an  $m$  by  $n$  matrix and  $V$  a  $n$  by  $n$  matrix. If we write down  $AA^\top$

[Singular Value Decomposition] Let  $A$  be a general  $m$  by  $n$  matrix. The matrix can be written:

$$A = U\Sigma V^\top \quad (4.9)$$

where  $U$  and  $V$  are square orthogonal matrices of size  $m$  and  $n$  respectively and  $\Sigma$  is an  $m$  by  $n$  diagonal matrix. The columns of  $U$  are the eigenvectors of  $AA^\top$  and the columns of  $V$  are the eigenvectors of  $A^\top A$ .

The general proof is not presented here, but can be found [NEED A REFERENCE](#). Instead, assume that the matrix can be factored as in the Singular Value Decomposition, then

$$\begin{aligned} AA^\top &= U\Sigma V^\top (U\Sigma V^\top)^\top \\ &= U\Sigma V^\top V \Sigma^\top U^\top \end{aligned}$$

and since  $V^\top V = I$ , then

$$AA^\top = U\Sigma\Sigma^\top U^\top. \quad (4.10)$$

The astute reader will notice that (4.10) shows that the matrices  $AA^\top$  and  $\Sigma\Sigma^\top$  are similar and thus have the same eigenvalues. Recall also that  $U$  will be the eigenvectors of  $AA^\top$ . These properties are used to actually find  $U$  and  $\Sigma$ .

Also, if you can find either  $U$  or  $V$ , then finding the other is easier by solving (4.10). If we know  $V$ , then

$$\begin{aligned} U\Sigma V^\top &= A \\ \Sigma V^\top &= U^{-1}A \end{aligned}$$

and recall that since  $U$  is an orthogonal matrix that  $U^{-1} = U^\top$ . This is now left-multiplied by  $\Sigma^\top$  in case the original matrix is not square, then  $\Sigma$  is not square and the inverse is not defined.

$$\begin{aligned} \Sigma^\top \Sigma V^\top &= \Sigma^\top U^\top A \\ V^\top &= (\Sigma^\top \Sigma)^{-1} \Sigma^\top U^\top A \\ V &= ((\Sigma^\top \Sigma)^{-1} \Sigma^\top U^\top A)^\top \end{aligned}$$

Use a property in Theorem ??

$$V = A^\top U \Sigma ((\Sigma^\top \Sigma)^{-1})^\top \quad (4.11)$$

and similarly if one solves for  $U$ :

$$U \Sigma V^\top = A$$

and recall that since  $V$  is an orthogonal matrix that  $V^{-1} = V^\top$  and that first this equation is right-multiplied by  $V$  to get

$$\begin{aligned} U \Sigma &= AV \\ U \Sigma \Sigma^\top &= AV \Sigma^\top \\ U &= AV \Sigma^\top (\Sigma \Sigma^\top)^{-1} \end{aligned} \quad (4.12)$$

Two examples of finding the SVD are presented below. The first is for a square matrix and the second is for a non-square matrix. We first start with the SVD of a 2 by 2 matrix with the following example. Find  $U$ ,  $\Sigma$ , and  $V$  of the SVD for the matrix

$$A = \begin{bmatrix} -1 & 2 \\ 1 & 2 \end{bmatrix}$$

As explained above, the columns of  $U$  are the eigenvectors of  $AA^\top$  and

$$AA^\top = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

The eigenvalues are found by

$$\det(A - \lambda I) = \begin{vmatrix} 5 - \lambda & 3 \\ 3 & 5 - \lambda \end{vmatrix} = (5 - \lambda)^2 - 9 = 0$$

with solutions  $\lambda = 2, 8$ . If  $\lambda = 2$ , then the solution to

$$(AA^\top - 2I)\vec{v}_1 = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \vec{v}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

is  $\vec{v}_1 = [1, -1]^\top$ . And if  $\lambda = 8$ , then the solution to

$$(AA^\top - 8I)\vec{v}_2 = \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \vec{v}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

is  $\vec{v}_2 = [1, 1]^\top$ . Thus the columns of  $U$  are the two eigenvectors scaled to be unit vectors,

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

The singular values in the matrix  $\Sigma$  are the square roots of the eigenvalues of  $AA^\top$  and the matrix is

$$\Sigma = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{8} \end{bmatrix}$$

To find the matrix  $V$ , we will use (??). Since

$$\Sigma^\top \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$$

and

$$(\Sigma^\top \Sigma)^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/8 \end{bmatrix},$$

then using (??),

$$\begin{aligned} V &= A^\top U \Sigma (\Sigma^\top \Sigma)^{-1}^\top \\ &= \begin{bmatrix} -1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{8} \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 1/8 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

It can be checked that indeed  $A = U \Sigma V^\top$ .

The next example finds the SVD of a 2 by 3 matrix, showing how to handle a non-square matrix.

Find the SVD of the matrix

$$A = \begin{bmatrix} 2 & -3 & -3 \\ 2 & 3 & 3 \end{bmatrix}$$

In this example, we will first find the columns of  $V$  by the eigenvectors of  $A^\top A$ . Generally, if the original matrix is not square, one should select the larger matrix between  $AA^\top$  and  $A^\top A$ .

$$A^\top A = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 18 & 18 \\ 0 & 18 & 18 \end{bmatrix}$$

and the eigenvalues of  $A^\top A$  is found by

$$\begin{aligned} \det(A^\top A - \lambda I) &= \begin{vmatrix} 8 - \lambda & 0 & 0 \\ 0 & 18 - \lambda & 18 \\ 0 & 18 & 18 - \lambda \end{vmatrix} = (8 - \lambda)((18 - \lambda)^2 - 18^2) \\ &= (8 - \lambda)(\lambda^2 - 36\lambda) = 0 \end{aligned}$$



which has the solutions  $\lambda = 8, 36, 0$ . Thus the singular values result in

$$\Sigma = \begin{bmatrix} \sqrt{8} & 0 & 0 \\ 0 & 6 & 0 \end{bmatrix}$$

and recall that this matrix is always the same size as the original matrix,  $A$ . Also, you will get zero eigenvalues whenever the matrix is not square.

The corresponding eigenvectors are

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

And the matrix  $V$  with the columns above scaled to make them unit vectors is

$$V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

To find  $U$ , we will use (??) and note that

$$(\Sigma\Sigma^\top)^{-1} = \begin{bmatrix} 1/8 & 0 \\ 0 & 1/36 \end{bmatrix}$$

and now using (??)

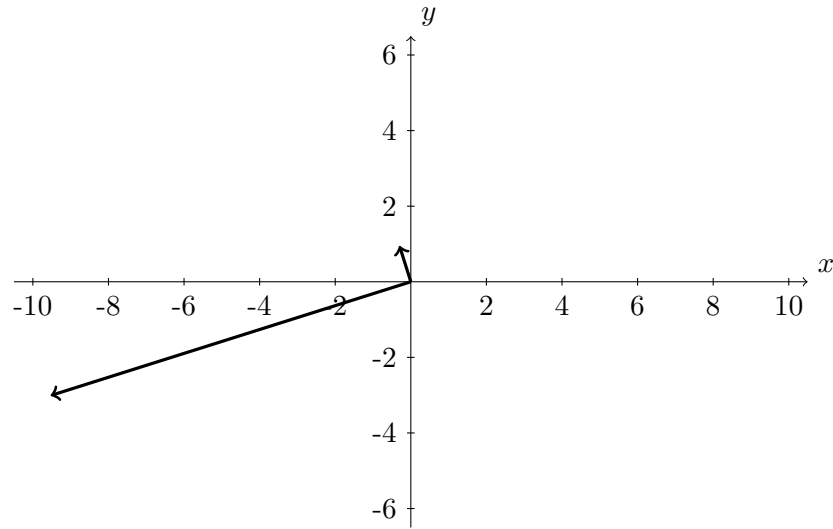
$$\begin{aligned} U &= AV\Sigma^\top(\Sigma\Sigma^\top)^{-1} \\ &= \begin{bmatrix} 2 & -3 & -3 \\ 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{8} & 0 \\ 0 & 6 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/8 & 0 \\ 0 & 1/36 \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix} \end{aligned}$$

and it can be checked that  $U\Sigma V^\top = A$ .

It is noted that the above two examples result in fairly nice matrices  $U, V$  and  $\Sigma$ . This very rarely happens. In addition, as the size of the matrices grow, the eigenvalues generally cannot be found exactly. Therefore for matrices of sizes larger than 3 by 3, numerical techniques are used to find the factors.

### The geometry of the SVD

This section will explain in geometric terms what the SVD generates for an example that can be plotted in 2D. First, let's consider 100 points that are scattered around the origin and perhaps they look like the following:



If the matrix is a 2 by 100 matrix with each column the  $x$  and  $y$  value of each point, we can perform an SVD of this. The results for  $U$  and the 2 diagonal elements of  $\Sigma$  (denoted  $\sigma_1$  and  $\sigma_2$ ) are

$$\sigma_1 = 54.51177 \quad \sigma_2 = 5.576430 \quad U = \begin{bmatrix} -0.953478 & -0.30146 \\ -0.30146 & 0.953478 \end{bmatrix}$$

where these have been done using software (not by hand).

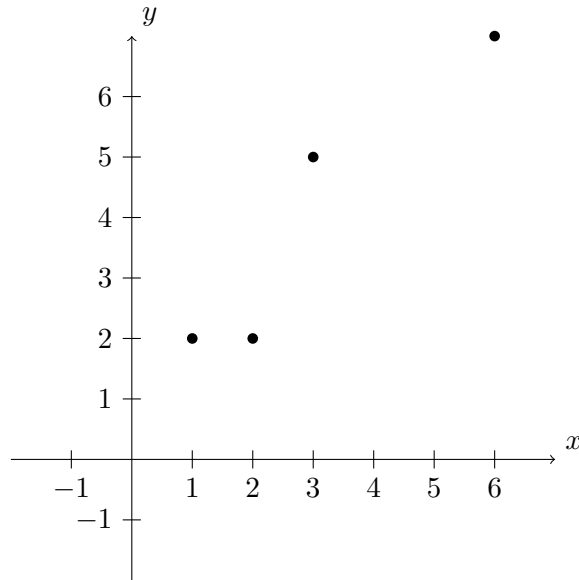
If we plot the two vectors in the columns of  $U$  with relative sizes of those found in  $\Sigma$ , you see in the plot above that the larger vector points in the direction of stretch of the points and the second vector is orthogonal. The relative lengths are the relative stretches of the points. The geometry of the columns in the  $V$  matrix is more difficult to understand in this situation because each of the vectors are length 100.

If the original matrix had 3 rows, then each column is a point in 3 dimensions. The 3 columns in  $U$  would again show the stretching of the points in 3 orthogonal directions with the stretching factor in the  $\Sigma$  matrix.

### Finding a best-fit line and plane using SVD

You might notice from above that perhaps we can use the SVD to find a line to a set of points and this is true. We will use a simpler example to show how this work and how it is similar to other fitting techniques. We can also use a similar technique to find the best-fit plane through points in 3 dimensions.

We first start with a simple example of points in the plan. Consider the points (1,2), (2,2), (3,5), (6,7). We wish to fit a line that best fits the points.



First, we will find the center of these points by just finding the average of the  $x$  and  $y$  values.

$$\bar{x} = \frac{1 + 2 + 3 + 6}{4} = 3 \qquad \bar{y} = \frac{2 + 2 + 5 + 7}{4} = 4$$

We now generate a new data set where each  $x$  and  $y$  value is adjusted by the means above. The new data set is  $\{(-2, -2), (-1, -2), (0, 1), (3, 3)\}$ . We then use these values to find

$$A = \begin{bmatrix} -2 & -1 & 0 & 3 \\ -2 & -2 & 1 & 3 \end{bmatrix}$$

As above, the singular values in  $\Sigma$  and the matrix  $U$  will be most important. As above, the matrix  $U$  have the eigenvectors of  $AA^T$  with the eigenvalues as the diagonal elements of  $\Sigma$ .

$$AA^T = \begin{bmatrix} 14 & 15 \\ 15 & 18 \end{bmatrix}$$

Again, using software, the diagonal elements of  $\Sigma$ , labelled  $\sigma_1$  and  $\sigma_2$  and the matrix  $U$  is

$$\sigma_1 = 5.5796 \qquad \sigma_2 = 0.93126 \qquad U = \begin{bmatrix} -0.658724 & -0.752384 \\ -0.752384 & 0.658725 \end{bmatrix}$$

The prominent eigenvector is the one corresponding to the larger singular value (the first one), so the eigenvalue to use is the first column of  $U$ .

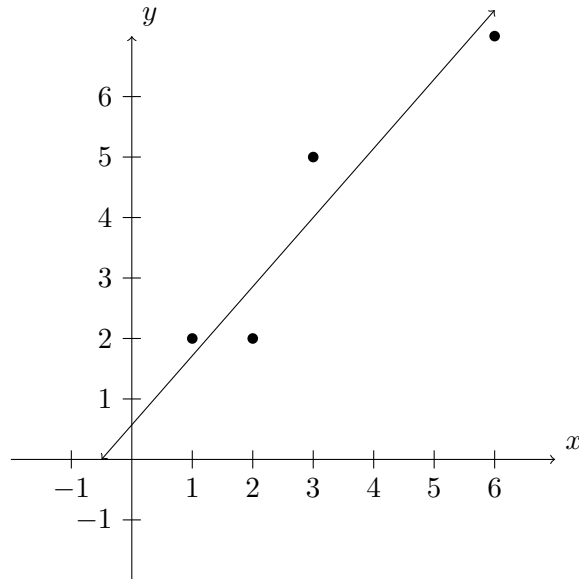
The slope of the line will be the  $y$ -component over the  $x$ -component (rise over run) or

$$m = \frac{-0.752384}{-0.658724} = 1.1421830$$

Lastly, to find the line, we will use the point-slope form with the average values (or center point) of  $(3, 4)$ :

$$\begin{aligned} y - 4 &= 1.1421830(x - 3) \\ y &= 1.1421830x + 0.571561 \end{aligned}$$

and the following is a plot of the data and the line:



And as a comparison, if we use least-squares to find the line, the result is

$$m = 1.0714x + 0.285714$$

The reason for the difference is because each method minimizes something different. The least-squares line is the one that minimizes the vertical distance between the data and the line. The SVD best-fit line produces a line that minimizes the square of the orthogonal distance to all the points. (this would be good to prove/show at some point).

## Chapter 5

# Fourier Series

### 5.1 Properties of Functions

In section ??, we reviewed inner product spaces and saw orthonormal sets of vectors (both in  $\mathbb{R}^3$  as well as polynomials). In this section, we will examine another set of functions, sines and cosines that are orthogonal. First, let's see a short review of periodic functions.

#### 5.1.1 Periodic Functions

A function is **periodic with period  $p$**  if

$$f(t) = f(t + p)$$

for all  $t$  in the domain of  $f$ . The smallest value of  $p > 0$  that makes this true is called the **period** of the function.

Show that  $f(t) = \sin t$  is periodic with period  $2\pi$ .

$$f(t + 2\pi) = \sin(t + 2\pi) = \sin t \cos 2\pi + \sin 2\pi \cos t = \sin t$$

What is the period of the function  $f(t) = \cos kt$ .

We know that the period of  $\cos t$  is also the same as  $\sin t$  or  $2\pi$ . If we let  $u = kt$ , then  $f(u) = \cos u$  has period  $p = 2\pi$  since it is the smallest value of  $p$  such that  $f(u) = f(u + p)$  for all  $u$ . The function  $f(t)$  would then had period  $p = 2\pi/k$ , since  $t = u/k$ .

The standard period functions that we will be using in this text are the sine and cosine function. We review here a few convenient identities with these functions and the complex exponential. From Euler's formula,

$$e^{i\theta} = \cos \theta + i \sin \theta$$

we can then write sine and cosine in terms of  $e^{i\theta}$

$$\sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \qquad \cos \theta = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

Euler's formula also leads to the following:

**The Most Interesting Equation in Mathematics**

$$e^{i\pi} + 1 = 0 \quad (5.1)$$

and this is often called the most interesting equation in mathematics because it arguably contains the 5 most important mathematical constants: 0, 1,  $e$ ,  $i$ ,  $\pi$ .

If  $k \in \mathbb{Z}$ , and  $x_0 \in \mathbb{R}$  then

$$\int_{x_0}^{x_0+2\pi} e^{ikx} dx = \begin{cases} 2\pi & k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

*Proof.* If  $k = 0$ , then the integral is of the constant function 1 over a interval of length  $2\pi$ , so the lemma holds, If  $k \neq 0$ ,

$$\begin{aligned} \int_{x_0}^{x_0+2\pi} e^{ikx} dx &= \left. \frac{e^{ikx}}{ik} \right|_{x_0}^{x_0+2\pi} \\ &= \frac{1}{ik} (e^{ik(x_0+2\pi)} - e^{ikx_0}) \\ &= \frac{1}{ik} e^{ikx_0} (e^{2\pi ik} - 1) \\ &= \frac{1}{ik} e^{ikx_0} ((e^{i\pi})^{2k} - 1) = 0 \end{aligned}$$

because  $e^{i\pi} = -1$  from (??), but this is raised to an even power so  $(e^{i\pi})^{2k} = 1$

□

### 5.1.2 One-Side Limits and Derivatives; Piecewise Continuous Functions

As we will see, the notion of a piecewise continuous function is a function that is continuous on subintervals. However, there are some technical details that we need before a formal definition.

A function  $f$  has a **left-hand limit** at  $x = x_0$  if

$$f(x_0^-) = \lim_{x \rightarrow x_0^-} f(x)$$

exists. In addition, a function  $f$  has a **right-hand limit** at  $x = x_0$  if

$$f(x_0^+) = \lim_{x \rightarrow x_0^+} f(x)$$

exists.

If one is talking about either a left- or right-handed limit, these are typically called one-sided limits. A function  $f$  has a **left-hand derivative** at  $x = x_0$  if

$$f'(x_0^-) = \lim_{h \rightarrow 0^-} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists. Similarly, a function  $f$  has a **right-hand derivative** at  $x = x_0$  if

$$f'(x_0^+) = \lim_{h \rightarrow 0^+} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists.

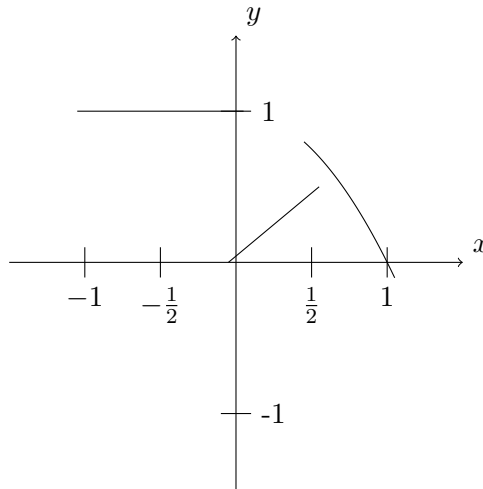
If one is talking about either a left- or right-handed derivative, these are typically called one-sided derivatives.

A function  $f$  is **piecewise continuous** on an interval  $[a, b]$  if  $f$  is continuous on all  $x \in [a, b]$  except for a finite number of points  $x_i$ . In addition for all  $x_i$ ,  $f(x_i^+)$  and  $f(x_i^-)$  exist.

The following function is piecewise continuous on  $[-1, 1]$

$$f(x) = \begin{cases} 1 & -1 \leq x < 0 \\ x & 0 \leq x < 1/2 \\ 1 - x^2 & 1/2 \leq x \leq 1 \end{cases}$$

Also, the graph of piecewise functions are helpful. These are found by finding the graphs of  $f$  on each given interval. The graph of  $f$  is shown below.



In addition, we need to show that all of the one-sided limits exist.

$$\lim_{x \rightarrow 0^-} f(x) = 1$$

$$\lim_{x \rightarrow 0^+} f(x) = 0$$

$$\lim_{x \rightarrow \frac{1}{2}^-} f(x) = \frac{1}{2}$$

$$\lim_{x \rightarrow \frac{1}{2}^+} f(x) = \frac{3}{4}$$

And since the function is continuous at all points except at 0 and  $1/2$ , but the one-sided limits are finite here, then the function  $f$  is piecewise continuous.

Find both the left- and right-handed derivatives of the function defined in example ?? at  $x = 0$  and  $x = 1/2$ .

First, consider the derivative of the function

$$f'(x) = \begin{cases} 0 & -1 < x < 0 \\ 1 & 0 < x < 1/2 \\ -2x & 1/2 < x < 1 \end{cases}$$

where the equality parts of the derivative have been removed (and explained later).

Since

$$\begin{array}{ll} \lim_{x \rightarrow 0^-} f'(x) = 0 & \lim_{x \rightarrow 0^+} = 1 \\ \lim_{x \rightarrow \frac{1}{2}^-} f'(x) = 1 & \lim_{x \rightarrow \frac{1}{2}^+} = -1 \end{array}$$

then the left-handed derivative at 0 is 0, the right-handed derivative of  $f$  at 1, the left-handed derivative at  $1/2$  is 1 and the right-handed derivative of  $f$  at  $1/2$  is -1.

Show that  $f(x) = 1/x$  is not a piecewise continuous function on  $[-1, 1]$ .

The problem in this function on the interval  $[-1, 1]$  is at  $x = 0$ . Also

$$\lim_{x \rightarrow 0^-} f(x) = -\infty \qquad \lim_{x \rightarrow 0^+} f(x) = \infty$$

and since the one-sided limits do not exist,  $f$  is not piecewise continuous on  $[-1, 1]$ .

### 5.1.3 Odd and Even Functions

A function  $f(x)$  is an **odd function** if  $f(-x) = -f(x)$  for all  $x$  in its domain.

Note: recall that an odd function is symmetric about the origin, meaning that if the graph of  $f$  is rotated a half circle about the origin, that one gets the graph back.

A function  $f(x)$  is an **even function** if  $f(-x) = f(x)$  for all  $x$  in its domain.

Recall that an even function is symmetric about the  $y$ -axis. This means that if the graph is reflected over the  $y$ -axis that one gets the same graph upon the reflection.

Here's a list of a few functions that are odd or even (without showing details):

- The following functions are odd:  $x, x^3, x^5, \sin x, \tan x$
- The following functions are even:  $1, x^2, x^4, \cos x$ .

The following theorem is helpful for finding whether or not products of functions are odd or even.

- The product of two odd functions is even.



- The product of two even functions is even.
- The product of an even and an odd function is odd.

The derivative of an even function is odd. The derivative of an odd function is even.

*Proof.* Let  $f$  be an even function, then  $f(-x) = f(x)$  for all  $x$ .

$$\frac{d}{dx}f(x) = \frac{d}{dx}f(-x) = -\frac{df(-x)}{dx}$$

by the chain rule. And the proof that the derivative of an odd function is similar.  $\square$

And as a corollary, antiderivatives work in the same way.

- Let  $f$  be an odd function. Any antiderivative of  $f$  is even.
- Let  $g$  be an even function and  $G$  be its antiderivative. The antiderivative  $G(x) + C$  such that  $C = -G(0)$  is odd.

Let  $f$  be a piecewise continuous function on the interval  $[-a, a]$  for  $a > 0$ .

- If  $f$  is an odd function then

$$\int_{-a}^a f(x) dx = 0$$

- If  $f$  is an even function then

$$\int_{-a}^a f(x) dx = 2 \int_0^a f(x) dx$$

*Proof.* First, examine the first statement. Let  $F(x)$  be an antiderivative of  $f(x)$ , an odd function. The function  $F(x)$  can be written as  $F(x) = G(x) + C$  where  $G(x)$  is an even function.

$$\int_{-a}^a f(x) dx = G(x) + C \Big|_{-a}^a = G(-a) + C - (G(a) + C) = 0 \quad (5.2)$$

since  $G(x)$  is even. The proof of the second statement is similar.  $\square$

### 5.1.4 Tabular Integration

A very handy formula for many integrations in this section is called **tabular integration**, which is just a recursive version of integration by parts that works well for integrals of a certain type. Before we show this, recall that the integration by parts formula is

$$\int u dv = uv - \int v du$$

and integration by parts is helpful for rewriting one integral (on the left) in terms of a second integral (on the right) and generally it is used to create a simpler integral. The next example shows a standard integration done with integration by parts.

Find

$$\int x e^x dx$$

In this case, we'll let  $u = x$  and  $dv = e^x dx$ , finding the differential  $u$  results in  $du = dx$  and finding an antiderivative of  $dv$  results in  $v = e^x$ , so using integration by parts to get

$$\int x e^x dx = x e^x - \int e^x dx$$

and finding the antiderivative of  $e^x$ ,

$$= x e^x - e^x + C$$

This example shows that in order to integrate with the by parts formula, one must replace one integral with another. In more difficult examples, this may need to be done multiple times until the resulting integral is able to be done without by parts. This is the case when tabular integration is useful.

The technique of **tabular integration** applied to

$$\int f(x)g(x) dx$$

where there exists an  $n$  such that  $f^{(n)}(x) = 0$ , that is eventually the derivative of  $f(x)$  is 0. Creates a table of three columns with

**Column 1:** The function  $f(x)$  and its derivatives until you reach zero.

**Column 2:** The signs  $+$  and  $-$ , starting with  $+$  and alternating signs.

**Column 3:** The function  $g(x)$  and its antiderivatives.

For columns 2 and 3, continue until you reach the same row as the 0 in the first column.

To find the antiderivative, draw arrows from each function in the first column, to a function in the third column one row below. The result is the sum of the product of each pair of functions connected by the arrows with the sign of that above the given arrow.

This is best seen with a couple of examples.

Find

$$\int x^3 e^x dx$$

using tabular integration.

First, we will build the table:

$f'(x) \pm \int g(x) dx$		
$x^3$	+	$e^x$
$3x^2$	-	$e^x$
$6x$	+	$e^x$
$6$	-	$e^x$
$0$	+	$e^x$

and then read off the result which is the product of terms connected by the arrows with the sign above each arrow.

$$\int x^3 e^x dx = x^3 e^x - 3x^2 e^x + 6x e^x - 6e^x + C$$

and don't forget the  $C$  for an indefinite integral.

And the following is an example that is similar as we will see below: Find

$$\int_0^\pi x^2 \cos 2x dx$$

using tabular integration.

First, we will build the table:

$f'(x) \pm \int g(x) dx$		
$x^2$	+	$\cos 2x$
$2x$	-	$\frac{1}{2} \sin 2x$
$2$	+	$-\frac{1}{4} \cos 2x$
$0$	-	$-\frac{1}{8} \sin 2x$

and then read off the result which is the product of terms connected by the arrows with the sign above each arrow.

$$\begin{aligned}
\int_0^\pi x^2 \cos 2x \, dx &= \left( \frac{x^2}{2} \sin 2x + \frac{x}{2} \cos 2x - \frac{1}{4} \sin 2x \right) \Big|_0^\pi \\
&= \left( \frac{\pi^2}{2} \sin 2\pi + \frac{\pi}{2} \cos 2\pi - \frac{1}{4} \sin 2\pi \right) \\
&\quad - \left( 0 + 0 - \frac{1}{4} \sin 0 \right) \\
&= \frac{\pi}{2}
\end{aligned}$$

### 5.1.5 Inner Products of Functions

As we saw in section ??, the inner product of two functions can be defined using the integral. In this section, we consider functions that are piecewise continuous on  $[a, b]$  and

$$\langle f, g \rangle = \int_a^b f(x)g(x) \, dx$$

In the standard way, the norm of  $f$  is defined as

$$\|f\| = \langle f, f \rangle^{1/2} = \left( \int_a^b f^2(x) \, dx \right)^{1/2}$$

An infinite set of continuous function  $\{f_1, f_2, \dots\}$  is said to be **orthogonal on**  $[a, b]$  if  $\langle f_n, f_m \rangle = 0$  for all  $n, m, n \neq m$ . If in addition,  $\|f_n\| = 1$  for all  $n$ , the set is said to be **orthonormal**.

In the next couple of examples, we examine a couple of the most important orthogonal and orthonormal sets of functions.

Show that  $f_n(x) = \sin nx$  for  $n = 1, 2, 3, \dots$  form an orthogonal set on  $[-\pi, \pi]$ .

In this case, we will expand on the example ?? and use the handy alternative definition of the sine function. Find the inner product if  $n \neq m$ ,

$$\begin{aligned}
\langle f_n, f_m \rangle &= \int_{-\pi}^{\pi} \sin nx \sin mx \, dx \\
&= \int_{-\pi}^{\pi} \frac{e^{inx} - e^{-inx}}{2i} \frac{e^{imx} - e^{-imx}}{2i} \, dx \\
&= -\frac{1}{4} \int_{-\pi}^{\pi} (e^{i(m+n)x} - e^{i(m-n)x} - e^{i(n-m)x} + e^{-i(m+n)x}) \, dx
\end{aligned}$$

Using Lemma ??

$$= 0$$

Show that the set of functions

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos x, \frac{1}{\sqrt{\pi}} \sin x, \frac{1}{\sqrt{\pi}} \cos 2x, \frac{1}{\sqrt{\pi}} \sin 2x, \dots \right\}$$

for an orthonormal set on  $[-\pi, \pi]$ .

We showed above that  $\langle \sin mx, \sin nx \rangle = 0$  for all  $m, n$  when  $m \neq n$ . Therefore  $\langle (\sin mx)/\sqrt{\pi}, (\sin nx)/\sqrt{\pi} \rangle$ . For simpler notation let

$$s_n(x) = \frac{1}{\sqrt{\pi}} \sin nx, \quad n \geq 0$$

$$c_n(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} & n = 0, \\ \frac{1}{\sqrt{\pi}} \cos nx & n > 0, \end{cases}$$

First, we will show that  $\langle s_m(x), c_n(x) \rangle = 0$  for all  $n, m > 0$ .

$$\begin{aligned} \langle s_m(x), c_n(x) \rangle &= \left\langle \frac{1}{\sqrt{\pi}} \sin mx, \frac{1}{\sqrt{\pi}} \cos nx \right\rangle \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \sin mx \cos nx \, dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{imx} - e^{-imx}}{2i} \frac{e^{inx} + e^{-inx}}{2} \, dx \\ &= \frac{1}{4i\pi} \int_{-\pi}^{\pi} (e^{i(m+n)x} - e^{i(n-m)x} + e^{i(m-n)x} - e^{-i(m+n)x}) \, dx \\ &= 0 \end{aligned}$$

since each integral is zero from Lemma ???. Next, we will show that  $\langle c_m(x), c_n(x) \rangle = 0$  for all  $n, m > 0$  such that  $n \neq m$ .

$$\begin{aligned} \langle c_m(x), c_n(x) \rangle &= \left\langle \frac{1}{\sqrt{\pi}} \cos mx, \frac{1}{\sqrt{\pi}} \cos nx \right\rangle \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \cos mx \cos nx \, dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{imx} + e^{-imx}}{2} \frac{e^{inx} + e^{-inx}}{2} \, dx \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} (e^{i(m+n)x} + e^{i(n-m)x} + e^{i(m-n)x} + e^{-i(m+n)x}) \, dx \\ &= 0 \end{aligned}$$

because again each integral is zero from Lemma ??.

Lastly, we need to show that the norm of each of the functions is 1.

$$\begin{aligned}\langle c_0, c_0 \rangle &= \int_{-\pi}^{\pi} \left( \frac{1}{\sqrt{2\pi}} \right)^2 dx = 1, \\ \langle c_n, c_n \rangle &= \frac{1}{\pi} \int_{-\pi}^{\pi} \cos^2 nx \, dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \left( \frac{e^{inx} + e^{-inx}}{2} \right)^2 dx \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} (e^{2inx} + 2 + e^{-2inx}) \, dx\end{aligned}$$

and the integrals of the first and third terms are zero from Lemma ??.

$$\begin{aligned}&= \frac{1}{4\pi} \int_{-\pi}^{\pi} 2 \, dx = 1 \\ \langle s_n, s_n \rangle &= \frac{1}{\pi} \int_{-\pi}^{\pi} \sin^2 nx \, dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \left( \frac{e^{inx} - e^{-inx}}{2i} \right)^2 dx \\ &= \frac{1}{(2i)^2 \pi} \int_{-\pi}^{\pi} (e^{2inx} - 2 + e^{-2inx}) \, dx\end{aligned}$$

and the integrals of the first and third terms are zero from Lemma ??.

$$= \frac{1}{-4\pi} \int_{-\pi}^{\pi} -2 \, dx = 1$$

This shows that the set of functions given above is an orthonormal set of functions.

## 5.2 Fourier Series

An infinite series of the form:

$$a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

is called the **trigonometric series**.

The trigonometric series is periodic with period at most  $2\pi$ . Consider the terms  $\cos x$  and  $\sin x$ , which each have period  $2\pi$ . All other functions have period  $2\pi/n$ , which are also periodic with period  $2\pi$ , but their fundamental period is  $2\pi/n$ .

Let  $f(x)$  be periodic of period  $2\pi$  and be piecewise continuous in  $[-\pi, \pi]$ . Suppose  $f(x)$  can be written as a trigonometric series. Then it is called a **Fourier Series** for  $f(x)$ . The

constants  $a_n$  and  $b_n$  are called the **Fourier Coefficients** of  $f(x)$  and are given by the Euler formulas:

$$a_0 = \frac{1}{2\pi} \langle f(x), 1 \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx \quad (5.3)$$

$$a_n = \frac{1}{\pi} \langle f(x), \cos nx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx \quad (5.4)$$

$$b_n = \frac{1}{\pi} \langle f(x), \sin nx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx \quad (5.5)$$

$$(5.6)$$

In this section, we're going to write period functions as Fourier Series. This is possible due to the following theorem.

Let  $f$  be a continuous function that is periodic with period  $2\pi$ . Then  $f$  can be written as a trigonometric series or

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx). \quad (5.7)$$

*Proof.* In this proof, we will take the inner product of  $f$  with each element in the set from Example ???. We will start with the constant function and use (??).

$$\begin{aligned} \left\langle \frac{1}{\sqrt{2\pi}}, f(x) \right\rangle &= \left\langle \frac{1}{\sqrt{2\pi}}, a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \right\rangle \\ \frac{1}{\sqrt{2\pi}} \langle 1, f(x) \rangle &= \frac{1}{\sqrt{2\pi}} \left( \langle a_0, 1 \rangle + \sum_{n=1}^{\infty} a_n \langle \cos nx, 1 \rangle + b_n \langle \sin nx, 1 \rangle \right) \end{aligned}$$

since the inner product of 1 with the trig functions are 0

$$= \frac{a_0}{\sqrt{2\pi}} \langle 1, 1 \rangle$$

and solving for  $a_0$ ,

$$a_0 = \frac{\langle f(x), 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx$$

Next, take the inner product of (??) with  $\frac{1}{\sqrt{\pi}} \cos nx$ .

$$\begin{aligned}\left\langle \frac{1}{\sqrt{\pi}} \cos nx, f(x) \right\rangle &= \left\langle \frac{1}{\sqrt{\pi}} \cos nx, a_0 + \sum_{m=1}^{\infty} (a_m \cos mx + b_m \sin mx) \right\rangle \\ \frac{1}{\sqrt{\pi}} \langle \cos nx, f(x) \rangle &= \frac{1}{\sqrt{\pi}} \left( a_0 \langle 1, \cos nx \rangle + \sum_{m=1}^{\infty} (a_m \langle \cos mx, \cos nx \rangle \right. \\ &\quad \left. + b_m \langle \sin mx, \cos nx \rangle) \right),\end{aligned}$$

All of the inner products on the right side are zero except when  $n = m$ . Canceling a  $\sqrt{\pi}$ , the result is

$$\langle \cos nx, f(x) \rangle = a_n \langle \cos nx, \cos nx \rangle$$

or solving for  $a_n$ ,

$$a_n = \frac{\langle f(x), \cos nx \rangle}{\langle \cos nx, \cos nx \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx$$

Lastly, take the inner product of (??) with  $\frac{1}{\sqrt{\pi}} \sin nx$ .

$$\begin{aligned}\left\langle \frac{1}{\sqrt{\pi}} \sin nx, f(x) \right\rangle &= \left\langle \frac{1}{\sqrt{\pi}} \sin nx, a_0 + \sum_{m=1}^{\infty} (a_m \cos mx + b_m \sin mx) \right\rangle \\ \frac{1}{\sqrt{\pi}} \langle \sin nx, f(x) \rangle &= \frac{1}{\sqrt{\pi}} \left( a_0 \langle 1, \sin nx \rangle + \sum_{m=1}^{\infty} (a_m \langle \cos mx, \sin nx \rangle \right. \\ &\quad \left. + b_m \langle \sin mx, \sin nx \rangle) \right),\end{aligned}$$

All of the inner products on the right side are zero except when  $n = m$ . Canceling a  $\sqrt{\pi}$ , the result is

$$\langle \sin nx, f(x) \rangle = b_n \langle \sin nx, \sin nx \rangle$$

and solving for  $b_n$ ,

$$b_n = \frac{\langle f(x), \sin nx \rangle}{\langle \sin nx, \sin nx \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx$$

□

Notice again that the statement in Theorem ?? is the  $f$  is periodic with period  $2\pi$ . This is a fairly strict requirement that we will relax over the rest of this section, however, one

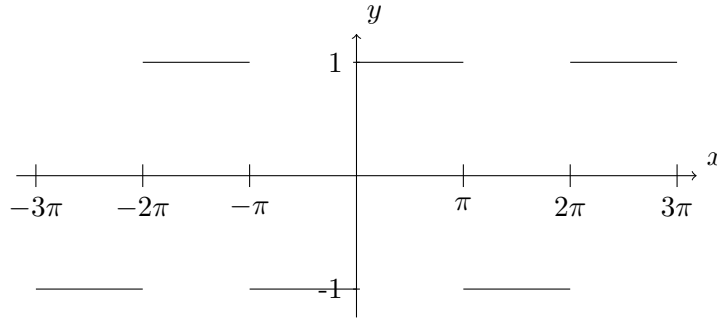


way to get a periodic function is to start with a function that is defined on  $[-\pi, \pi]$  and extend it periodically such that  $f(x) = f(x + 2\pi)$ . We do this in the following example.

Find the Fourier coefficients and the Fourier series for the periodic extension square wave:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq \pi \\ -1 & -\pi < x < 0 \end{cases}$$

and define  $f(x)$  to be its periodic extension of period  $2\pi$ . That is if  $x$  is outside of  $[-\pi, \pi]$ , then apply  $f(x) = f(x + 2\pi)$  or  $f(x) = f(x - 2\pi)$  until  $x \in [-\pi, \pi]$ . This function looks like:



To begin with, we find all of the coefficients:

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx \\ &= \frac{1}{2\pi} \left( \int_{-\pi}^0 (-1) dx + \int_0^{\pi} (1) dx \right) = \frac{1}{2\pi} (-\pi + \pi) = 0 \end{aligned}$$

$$\begin{aligned} a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx \\ &= \frac{1}{\pi} \left( \int_{-\pi}^0 (-1) \cos nx dx + \int_0^{\pi} \cos nx dx \right) \\ &= \frac{1}{\pi} \left( -\frac{1}{n} \sin nx \Big|_{-\pi}^0 + \frac{1}{n} \sin nx \Big|_0^{\pi} \right) = 0 \end{aligned}$$

$$\begin{aligned}
 b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx \\
 &= \frac{1}{\pi} \left( \int_{-\pi}^0 (-1) \sin nx \, dx + \int_0^{\pi} \sin nx \, dx \right) \\
 &= \frac{1}{\pi} \left( \frac{1}{n} \cos nx \Big|_{-\pi}^0 - \frac{1}{n} \cos nx \Big|_0^{\pi} \right) \\
 &= \frac{1}{n\pi} ((1 - \cos n(-\pi)) - (\cos(n\pi) - 1)) \\
 &= \frac{2}{n\pi} (1 - (-1)^n)
 \end{aligned}$$

So the Fourier Series can be written:

$$\begin{aligned}
 F(x) &= \sum_{n=1}^{\infty} \frac{2}{n\pi} (1 - (-1)^n) \sin nx \\
 &= \frac{4}{\pi} \left( \sin x + \frac{1}{3} \sin 3x + \frac{1}{5} \sin 5x + \frac{1}{7} \sin 7x + \frac{1}{9} \sin 9x + \dots \right)
 \end{aligned}$$

or again, writing it more compactly,

$$= \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{1}{2n-1} \sin(2n-1)x$$

The above series also shows an interesting result that you should have seen in the series chapter of Calculus. If we let  $x = \pi/2$ , then  $\sin(2n-1)\frac{\pi}{2} = (-1)^{n+1}$  and  $F(\pi/2) = 1$  from the definition of the square wave and substituting this into the Fourier series, we get:

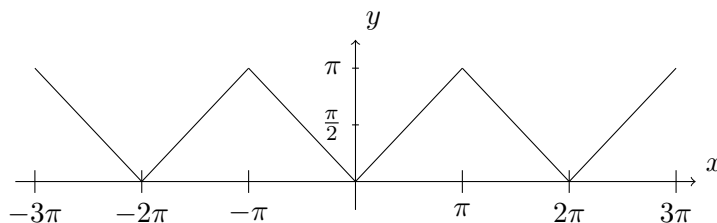
$$1 = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{2n-1}$$

or multiplying both sides by  $\pi/4$

$$\frac{\pi}{4} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{2n-1}$$

which shows that some infinite sums have closed form values. This particular series is usually found using the Taylor Series of  $\tan x$  and evaluating it at 1.

Find the Fourier series of the period sawtooth wave:



Let  $f(x)$  be the sawtooth wave defined in the picture above. We can write it as a piecewise function as

$$f(x) = \begin{cases} x & 0 \leq x \leq \pi \\ -x & -\pi < x < 0 \end{cases}$$

and extending it periodically.

Then using the formulas in (??)–(??) and we will take advantage of the fact that  $f(x)$  is an even function.

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx \\ &= \frac{2}{2\pi} \int_0^{\pi} x dx \\ &= \frac{1}{\pi} \left. \frac{x^2}{2} \right|_0^{\pi} = \frac{\pi}{2} \end{aligned}$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx$$

and since  $\cos nx$  is even and the product of even functions is even

$$= \frac{2}{\pi} \int_0^{\pi} x \cos nx dx$$

And this is a good example to use tabular integration.

$f'(x)$	$\pm$	$\int g(x) dx$
$x$	$+$	$\cos nx$
$1$	$-$	$\frac{1}{n} \sin nx$
$0$	$+$	$-\frac{1}{n^2} \cos nx$

and then using the table to find

$$\begin{aligned} a_n &= \frac{2}{\pi} \left( \frac{x}{n} \sin nx + \frac{1}{n^2} \cos nx \right) \Big|_0^{\pi} \\ &= \frac{2}{\pi} \left( \left( \frac{\pi}{n} \sin n\pi + \frac{1}{n^2} \cos n\pi \right) - \left( 0 + \frac{1}{n^2} \cos 0 \right) \right) \\ &= \frac{2}{\pi n^2} ((-1)^n - 1) \end{aligned}$$

Lastly,

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx$$

but this is a product of an even and an odd function, which is odd and integrating an odd function over a symmetric interval is 0. Therefore the Fourier series is

$$F(x) = \frac{\pi}{2} + \sum_{n=1}^{\infty} \frac{2}{\pi n^2} ((-1)^n - 1) \cos nx$$

### 5.2.1 Convergence of a Sum of a Fourier Series

Since Fourier series are infinite series, it is important to consider if it converges. As we will see, Fourier series will generally converge, to what value will depend on  $x$ . Consider the Fourier series in Example ?? . If we let

$$F(x) = \frac{\pi}{2} + \sum_{n=1}^{\infty} \frac{1}{n^2} ((-1)^n - 1) \cos nx$$

we would need to test the convergence of every value of  $x$ . In this case, this can be done by using the direct comparison test to the series  $\sum_{n=1}^{\infty} \frac{1}{n^2}$  which converges, so the series  $F(x)$  converges for all  $x$ .

This doesn't work for all series and the other difficulty is that we don't know what it converges to. Fortunately, the following theorem gives a very nice result.

Let  $f(x)$  be periodic with period  $2\pi$  and piecewise continuous in the interval  $-\pi \leq x \leq \pi$ . Let  $F(x)$  be the Fourier series of  $f(x)$  and

$$F(x) = \begin{cases} f(x) & \text{if } f \text{ is continuous at } x \\ \frac{f(x^+) + f(x^-)}{2} & \text{if } f \text{ is not continuous at } x \end{cases}$$

In other words, the Fourier series converges to the average of the left- and right-hand limits of  $f(x)$ .

Show that the Fourier series of the square wave function above converges to 0 when  $x = 0$ .

Note that average of the left- and right-handed limits of the square wave function at  $x = 0$  is  $(1 + (-1))/2 = 0$ , so using the theorem above, the function converges to 0 when  $x = 0$ .

Alternatively, we can evaluate the Fourier series of the square wave function directly. Evaluating the Fourier series at  $x = 0$  is

$$F(0) = \sum_{n=1}^{\infty} \frac{2}{n\pi} (1 - \cos(n\pi)) \sin nx \Big|_{x=0} = 0,$$

which is consistent with that above.

### 5.2.2 Fourier Series of Functions of Period $2L$

We saw that the Fourier series above applied only to functions that were periodic with period  $2\pi$ . This section covers functions with arbitrary periodicity, which we will call period  $2L$ .

If we let  $u = \frac{\pi x}{L}$ , and substitute this into (??), then

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right) \quad (5.8)$$

and then it can be shown in a similar manner to that above that (??)–(??) can be written as

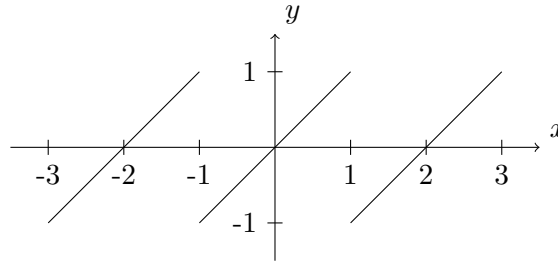
$$a_0 = \frac{1}{2L} \int_{-L}^L f(x) dx \quad (5.9)$$

$$a_n = \frac{1}{L} \int_{-L}^L f(x) \cos \left( \frac{n\pi x}{L} \right) dx \quad (5.10)$$

$$b_n = \frac{1}{L} \int_{-L}^L f(x) \sin \left( \frac{n\pi x}{L} \right) dx \quad (5.11)$$

The series is called the Fourier series of period  $2L$  with the corresponding Fourier coefficients.

Find the Fourier Series of the periodic extension (of period 2) of  $f(x) = x$  for  $x \in [-1, 1]$  as shown in the graph below:



To find the Fourier series, we first need to find the Fourier coefficients, by evaluating the integrals in (??)–(??),

$$a_0 = \frac{1}{2} \int_{-1}^1 x dx = 0$$

because  $x$  is an odd function

$$a_n = \frac{1}{1} \int_{-1}^1 x \cos n\pi x dx = 0$$

because  $x$  times  $\cos n\pi x$  is odd

$$b_n = \frac{1}{1} \int_{-1}^1 x \sin n\pi x dx$$

because  $x$  times  $\sin n\pi x$  is even

$$= 2 \int_0^1 x \sin n\pi x$$

using tabular integration,

$$\begin{aligned} b_n &= -\frac{x}{n\pi} \cos(n\pi x) + \frac{1}{n^2\pi^2} \sin(n\pi x) \Big|_{-1}^1 \\ &= -\frac{2}{n\pi} \cos(n\pi) = \frac{2}{n\pi} (-1)^{n+1} \end{aligned}$$

The Fourier Series of the function is

$$\sum_{n=1}^{\infty} \frac{2}{n\pi} (-1)^{n+1} \sin n\pi x$$

### 5.3 Even and Odd Functions; Half-Range Expansions

We saw above that the periodic extension of  $f(x) = x$  on  $[-1, 1]$  in Example ?? resulted in a odd function and that only the sine terms of the Fourier Series was left. That is, all of the Fourier coefficients for the cosine terms were 0. In this section, we use this idea to produce only even and odd extensions which results in only sine expansions or cosine expansions.

To begin, let's clearly define an even- and odd-periodic extension.

Let  $f$  be defined on  $[0, L]$  for some  $L > 0$ .

- The even periodic extension of  $f$  is called  $F(x)$  and is defined:

$$F(x) = \begin{cases} f(x) & 0 \leq x \leq L \\ f(-x) & -L \leq x < 0 \\ f(x + 2kL) & \text{otherwise for appropriate integer } k \end{cases}$$

- The odd periodic extension of  $f$  is called  $F(x)$  and is defined:

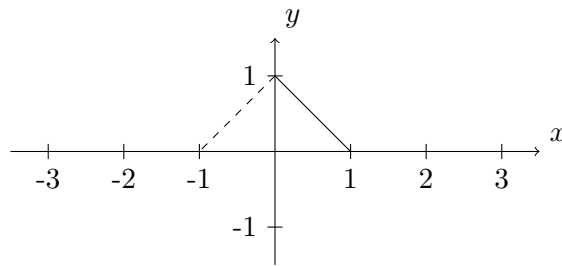
$$F(x) = \begin{cases} f(x) & 0 \leq x \leq L \\ -f(-x) & -L \leq x < 0 \\ f(x + 2kL) & \text{otherwise for appropriate integer } k \end{cases}$$

Graph the even- and odd-periodic extension of

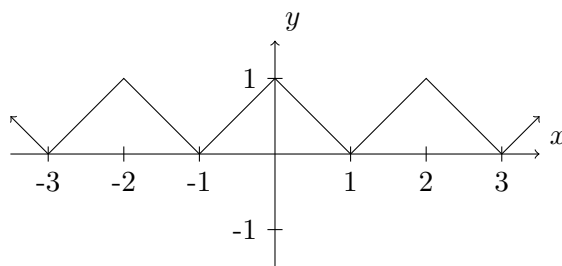
$$f(x) = 1 - x$$

defined on  $[0, 1]$ .

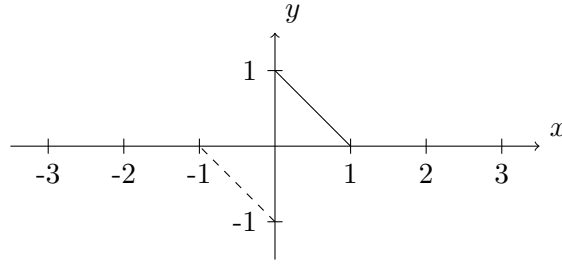
For the even extension, we first graph the function on  $[0, 1]$ , then make the even extension of it on  $[-1, 0]$ . The original function is shown below as a solid line and the even extension is dashed.



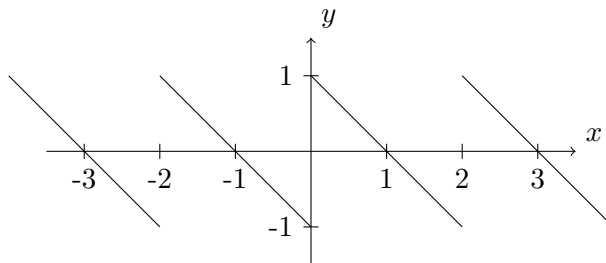
and then we produce the period extension of period 2.



To find the odd extension, flip the original function on  $[0, 1]$  around the origin to get:



Then extend the function on  $[-1, 1]$  in a periodic way.



### 5.3.1 The Fourier Sine and Cosine Series

We now address the Fourier series of the even- and odd-periodic extensions of  $f$  on  $[0, L]$ . As in Example ??, there are no cosine terms and the Fourier series of the odd periodic

extension of  $f$  can be written

$$f(x) = \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right) \quad (5.12)$$

where,

$$b_n = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx \quad (5.13)$$

and this is often called the **Fourier sine series**.

Similarly, the Fourier series of the even periodic extension of  $f$  is

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right) \quad (5.14)$$

where

$$a_0 = \frac{1}{L} \int_0^L f(x) dx \quad (5.15)$$

$$a_n = \frac{2}{L} \int_0^L f(x) \cos\left(\frac{n\pi x}{L}\right) dx \quad (5.16)$$

and is called the **Fourier cosine series**.

Find the Fourier Cosine Series of  $f(x) = x$  for  $0 \leq x \leq 1$ .

For this, we need to find the coefficients in (??) and (??),

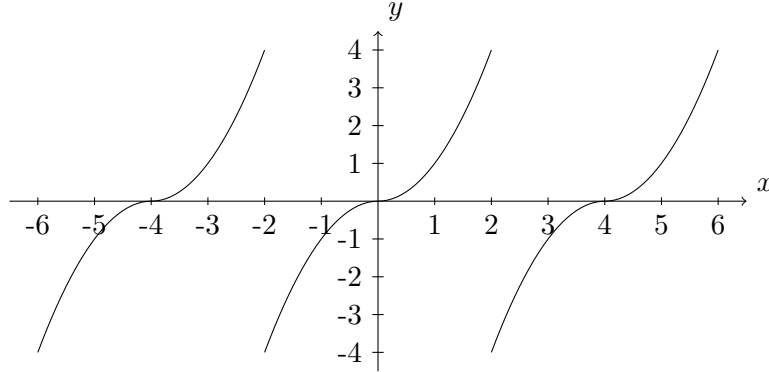
$$\begin{aligned} a_0 &= \frac{1}{1} \int_0^1 x dx = \frac{1}{2} \\ a_n &= \frac{2}{1} \int_0^1 x \cos n\pi x dx \\ &= \frac{2(\cos n\pi - 1)}{n\pi} = \frac{2}{n\pi}((-1)^n - 1) \end{aligned}$$

and use (??) to find

$$f(x) = \frac{1}{2} + \sum_{n=1}^{\infty} \frac{2}{n\pi}((-1)^n - 1) \cos n\pi x$$

Find the odd periodic extension of  $f(x) = x^2$  on  $[0, 2]$ . A graph of this is:





For this, we need to find  $b_n$  from (??),

$$b_n = \frac{2}{2} \int_0^2 x^2 \sin\left(\frac{n\pi x}{2}\right) dx$$

and again, this can be done using tabular integration,

$$\begin{aligned} &= -\frac{2x^2}{n\pi} \cos\left(\frac{n\pi x}{2}\right) + \frac{8x}{n^2\pi^2} \sin\left(\frac{n\pi x}{2}\right) + \frac{16}{n^3\pi^3} \cos\left(\frac{n\pi x}{2}\right) \Big|_0^2 \\ &= -\frac{8}{n\pi} \cos(n\pi) + \frac{16}{n^3\pi^3} \cos(n\pi) = \frac{16 - 8n^2\pi^2}{n^3\pi^3} (-1)^n \end{aligned}$$

and then use this in (??)

$$f(x) = \sum_{n=1}^{\infty} \frac{16 - 8n^2\pi^2}{n^3\pi^3} (-1)^n \sin\left(\frac{n\pi x}{2}\right)$$

## 5.4 Approximation by Trigonometric Polynomials

Consider a periodic function  $f$  of period  $2\pi$  on the interval  $[-\pi, \pi]$ . The  $N$ th partial sum of the Fourier Series of  $f$  is denoted  $f_N$ ,

$$f_N(x) = a_0 + \sum_{n=1}^N (a_n \cos nx + b_n \sin nx)$$

where  $a_0, a_n$  and  $b_n$  are the Fourier Coefficients as before. The function  $f_N$  is also called the **Trigonometric Polynomial of degree  $n$** .

Let's ask a question about approximation. Consider a function of the form:

$$F_N(x) = A_0 + \sum_{n=1}^N (A_n \cos nx + B_n \sin nx)$$

What is the best approximate for a trigonometric polynomial to another function  $F(x)$ . That is, what coefficients can be chosen  $A_0, A_n, B_n$ ?

To answer this question, we will need to know what error we are taking about. Typically the error will be some function of the two functions, called  $E(f, g)$  that outputs a number. We would like the error to have the following properties:

$$\begin{aligned} E(f, g) &= E(g, f) \\ E(f, g) &\geq 0 \\ E(f, f) &= 0 \end{aligned}$$

one such function that we know is the square of the function norm or

$$\begin{aligned} E(F, F_N) &= \langle F - F_N, F - F_N \rangle = \|F - F_N\|^2 \\ &= \int_{-\pi}^{\pi} (F(x) - F_N(x))^2 dx \\ &= \int_{-\pi}^{\pi} \left( F(x) - A_0 - \sum_{n=1}^N (A_n \cos nx + B_n \sin nx) \right)^2 dx \end{aligned}$$

To find the minimum of this, we will take the derivatives of  $E$  with respect to  $A_0, A_n$  and  $B_n$ :

$$\begin{aligned} \frac{\partial E}{\partial A_0} &= -2 \int_{-\pi}^{\pi} (F(x) - A_0 - \sum_{n=1}^N (A_n \cos nx + B_n \sin nx)) \\ 0 &= \langle F, 1 \rangle - \langle A_0, 1 \rangle - \sum_{n=1}^N (\langle A_n \cos nx, 1 \rangle + \langle B_n \sin nx, 1 \rangle) \\ 0 &= \langle F, 1 \rangle - A_0 \langle 1, 1 \rangle \end{aligned}$$

or

$$A_0 = \frac{\langle F, 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(x) dx$$

Similarly take the derivative with respect to  $A_m$ :

$$\frac{\partial E}{\partial A_m} = -2 \int_{-\pi}^{\pi} \cos mx (F(x) - A_0 - \sum_{n=1}^N (A_n \cos nx + B_n \sin nx))$$

set the derivative to 0

$$\begin{aligned} 0 &= \langle F, \cos mx \rangle - \langle A_0, \cos mx \rangle \\ &\quad - \sum_{n=1}^N (\langle A_n \cos nx, \cos mx \rangle + \langle B_n \sin nx, \cos mx \rangle) \end{aligned}$$

Note that  $\langle \cos nx, \cos mx \rangle = 0$  unless  $m = n$  and  $\langle 1, \cos mx \rangle = 0$

$$0 = \langle F, \cos mx \rangle - A_m \langle \cos mx, \cos mx \rangle$$

or

$$\begin{aligned} A_m &= \frac{\langle F, \cos mx \rangle}{\langle \cos mx, \cos mx \rangle} \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} F(x) \cos mx \, dx \end{aligned}$$

And similarly it can be shown that

$$B_m = \frac{1}{\pi} \int_{-\pi}^{\pi} F(x) \sin mx \, dx$$

Let  $f(x)$  be a piecewise continuous function on  $[-\pi, \pi]$ ,  $N > 0$  and

$$f_N(x) = a_0 + \sum_{n=1}^N (a_n \cos nx + b_n \sin nx)$$

The values of  $a_n$  and  $b_n$  that minimize  $\|f - f_N\|$  (or  $\|f - f_N\|^2$ ) is

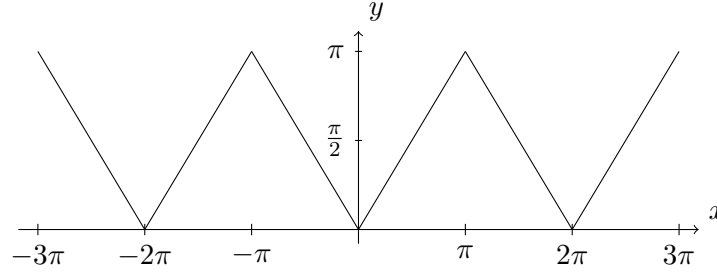
$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \, dx \\ a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx \end{aligned}$$

that is they are the Fourier Coefficients.

In Example ??, the Fourier series of the sawtooth function was found the Fourier Coefficients are:

$$\begin{aligned} a_0 &= \frac{\pi}{2} \\ a_n &= \frac{2}{\pi n^2} ((-1)^n - 1) \quad n \geq 1 \\ b_n &= 0 \end{aligned}$$

Graph the sawtooth function and  $f_9(x)$ , the 9th degree trigonometric polynomial.

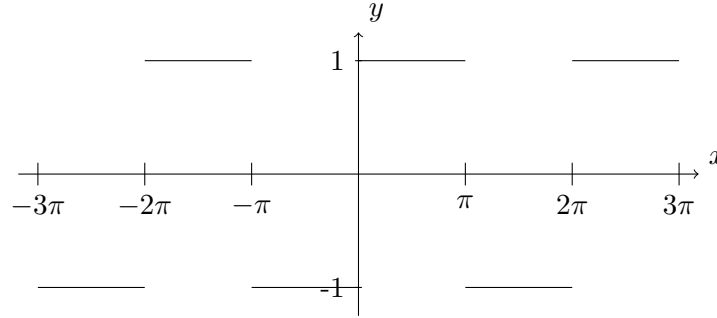


and the two plots are indistinguishable on this scale.

In Example ??, the Fourier series of the square wave function was found and the Fourier Coefficients are:

$$\begin{aligned} a_0 &= 0 \\ a_n &= 0 \\ b_n &= \frac{2}{n\pi}(1 - (-1)^n) \end{aligned}$$

Graph the square wave function and  $f_9(x)$ , the 9th degree trigonometric polynomial.



And in contrast to the previous example, the trigonometric polynomial  $f_9(x)$  and the original function  $f$  are quite different.

We will explore this example in a bit more detail after seeing some important theorems.

#### 5.4.1 Theorems Related to Fourier Series

The quantity  $\|F - F_N\|^2$  on the interval  $[-\pi, \pi]$  is the minimum if and only if the coefficients of  $F_N$  in (2) are the Fourier coefficients of  $F$ . This minimum value is

$$E^* = \int_{-\pi}^{\pi} f^2 dx - 2A_0\pi - \pi \sum_{n=1}^N (A_n^2 + B_n^2) \quad (5.17)$$

[Bessel's Inequality] Let  $a_0, a_n$  and  $b_n$  be the fourier coefficients related to the function  $f$  on  $[-\pi, \pi]$ . Then

$$2a_0^2 + \sum_{n=1}^N (a_n^2 + b_n^2) \leq \frac{1}{\pi} \int_{-\pi}^{\pi} f^2 dx$$

[Parseval's Theorem] Let  $a_0, a_n$  and  $b_n$  be the fourier coefficients related to the function  $f$  on  $[-\pi, \pi]$ . Then

$$2a_0^2 + \sum_{n=1}^{\infty} (a_n^2 + b_n^2) = \frac{1}{\pi} \int_{-\pi}^{\pi} f^2 dx$$

There are two important consequences of this theorem:

- If the integral on the right side is finite, then the series on the left converges. Functions in which the right side is finite are piecewise continuous functions.
- The error,  $E^*$  in (??) goes to zero. That is Fourier Series converge to  $f$  (using the square error).

Calculate  $E^*$  for the function:

$$f(x) = \begin{cases} \pi + x & x \in [-\pi, 0) \\ \pi - x & x \in [0, \pi] \end{cases}$$

and extended periodically and let  $N = 5, 10, 25, 50, 100, 250, 500, 1000$ .

The Fourier Series of this function is

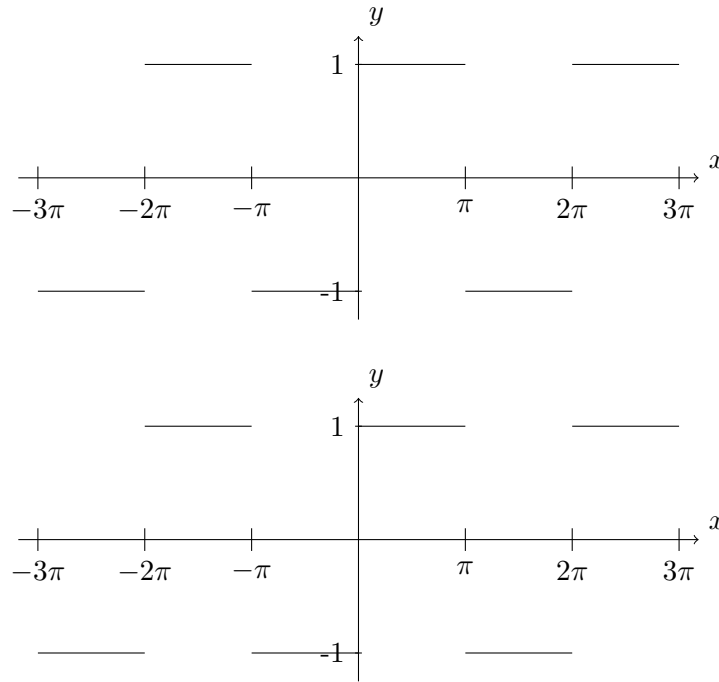
$$f(x) = \frac{\pi}{2} + \sum_{n=1}^{\infty} \frac{2}{\pi n^2} (1 - (-1)^n) \cos nx$$

or  $a_0 = \pi/2$  and  $a_n = (2/\pi n^2)(1 - (-1)^n)$

$$\begin{aligned} E^* &= \int_{-\pi}^{\pi} f^2 dx - 2A_0\pi - \pi \sum_{n=1}^N (A_n^2 + B_n^2) \\ &= 2 \int_0^{\pi} (\pi - x)^2 dx - 2\left(\frac{\pi}{2}\right) - \pi \sum_{n=1}^N \left(\frac{2}{\pi n^2} (1 - (-1)^n)\right) \end{aligned}$$

$n$	$E^*$
5	0.00372
10	0.000832
25	0.0000482
50	$6.78 \times 10^{-6}$
100	$8.49 \times 10^{-8}$
250	$5.43 \times 10^{-8}$
500	$6.79 \times 10^{-9}$
1000	$8.48 \times 10^{-10}$

A consequence of Parseval's Theorem is that for piecewise continuous functions, the Fourier Series converges as  $n \rightarrow \infty$ . So in light of the plot in Example ??, that it would appear that the plot of  $f_N$  would approach the square wave as  $N \rightarrow \infty$ . However the plots of  $f_{25}$  and  $f_{100}$  are shown below (with  $n = 25$  on top):



And despite the larger value of  $N$ ,  $f_N$  does not appear to be approaching the square wave function. The difference is pronounced near the discontinuities in the function. This is called *Gibbs Phenomena* and it can be shown in this situation that the local max near  $x = 0$  in fact grows without bound as  $N \rightarrow \infty$ , despite the fact that  $\|f_N - f\| \rightarrow 0$ .

## 5.5 The Fourier Series using the complex exponential

Recall that the Fourier Series of  $f(x)$  on  $[-\pi, \pi]$  is

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

where

$$\begin{aligned} a_0 &= \frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx \\ a_n &= \frac{\langle f, \cos nx \rangle}{\langle \cos nx, \cos nx \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx \\ b_n &= \frac{\langle f, \sin nx \rangle}{\langle \sin nx, \sin nx \rangle} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx \end{aligned}$$

and if  $f(x)$  is a real-valued function then  $a_0, a_n$  and  $b_n$  are real values.

If we use the identities:

$$\begin{aligned}\cos nx &= \frac{1}{2}(e^{inx} + e^{-inx}) \\ \sin nx &= \frac{1}{2i}(e^{inx} - e^{-inx})\end{aligned}$$

then the Fourier Series can be written:

$$\begin{aligned}f(x) &= a_0 + \sum_{n=1}^{\infty} \left( \frac{a_n}{2}(e^{inx} + e^{-inx}) + \frac{b_n}{2i}(e^{inx} - e^{-inx}) \right) \\ &= a_0 + \frac{1}{2} \sum_{n=1}^{\infty} (a_n - ib_n)e^{inx} + (a_n + ib_n)e^{-inx}\end{aligned}$$

if we let  $c_0 = a_0$  and  $c_n = (a_n - ib_n)/2$  for  $n > 0$  and  $c_n = (a_{-n} + ib_{-n})$  for  $n < 0$  then we can write the sum as

$$\begin{aligned}f(x) &= c_0 + \sum_{n=1}^{\infty} c_n e^{inx} + c_{-n} e^{-inx} \\ &= \sum_{n=-\infty}^{\infty} c_n e^{inx}\end{aligned}$$

where

$$\begin{aligned}c_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx \\ c_n &= \frac{a_n - ib_n}{2} = \frac{1}{2} \cdot \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)(\cos nx - i \sin nx) dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-inx} dx \\ c_{-n} &= \frac{a_n + ib_n}{2} = \frac{1}{2} \cdot \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)(\cos nx + i \sin nx) dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{inx} dx\end{aligned}$$

and since  $c_{-n}$  is consistent with  $c_n$ , we use  $c_n$  for all  $n \neq 0$  and actually since that is also consistent, then

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-inx} dx$$

If we take the interval from  $[-L, L]$  instead of from  $[-\pi, \pi]$  we get

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i x/L}$$

where

$$c_n = \frac{1}{2L} \int_{-L}^L f(x) e^{-2\pi i x/L} dx$$

Find the complex Fourier series on  $[-\pi, \pi]$  if  $f(x) = x$ .

$$\begin{aligned} c_n &= \frac{1}{2\pi} \int_{-\pi}^{\pi} x e^{-inx} dx \\ &= \begin{cases} 0 & n = 0 \\ \frac{-2i\pi}{n} e^{-i\pi n} = \frac{-2(-1)^n \pi}{n} & \text{otherwise} \end{cases} \end{aligned}$$

As you can see, finding the complex Fourier Series is generally easier in terms of only needing to find one integral instead of 3 for the standard Fourier Series.



## Chapter 6

# Partial Differential Equations

Partial differential equations (or PDEs) are very common in modeling many physical phenomena including understanding wave propagation, heat transport, and the motion of fluids as just a limited selection of examples. We will derive equations for some simple examples in the section using conservation from physics.

Before getting to the derivation and solution of PDEs, we need to review and introduce a few concepts.

### 6.1 Partial Derivatives and Differential Equations

A **partial derivative** of  $f(x, y, z)$  with respect to  $x$  is the derivative of  $f$  with respect to  $x$  keeping the other variables constant. Technically

$$\frac{\partial f(x, y, z)}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y, z) - f(x, y, z)}{\Delta x}$$

If  $f(x, y) = x^2 + xy + y \sin x$ , Find  $\frac{\partial f}{\partial x}$  and  $\frac{\partial f}{\partial y}$ .

To find  $\frac{\partial f}{\partial x}$  consider  $y$  a constant, so

$$\frac{\partial f}{\partial x} = 2x + y + y \cos x$$

similarly to find  $\frac{\partial f}{\partial y}$  consider  $x$  a constant,

$$\frac{\partial f}{\partial y} = x + \sin x$$

#### 6.1.1 Higher-Order Partial Derivatives

As with ordinary derivatives, we can have higher-order partial derivatives. That is we define the 2nd order partial derivative of  $f$  with respect to  $x$  as

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \frac{\partial f}{\partial x}$$

that is it is the partial derivative of the partial derivative.

Since ordinary derivatives involve only functions of one variable, the mixed derivative is a new concept. If  $f$  is a function of  $x$  and  $y$  or  $f(x, y)$ , then we can have the partial derivative of  $\frac{\partial f}{\partial x}$  with respect to  $y$  or  $\frac{\partial f}{\partial y}$  with respect to  $x$ . We write these as

$$\begin{aligned}\frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y} \frac{\partial f}{\partial x} \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x} \frac{\partial f}{\partial y}\end{aligned}$$

Find all second-order derivatives of

$$f(x, y) = x^3 + x^2y + xe^y$$

First, let's find the two first derivatives:

$$\begin{aligned}\frac{\partial f}{\partial x} &= 3x^2 + 2xy + e^y \\ \frac{\partial f}{\partial y} &= x^2 + xe^y\end{aligned}$$

And then there are four 2nd-order derivatives:

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x} \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} (3x^2 + 2xy + e^y) = 6x + 2y \\ \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y} \frac{\partial f}{\partial x} = \frac{\partial}{\partial y} (3x^2 + 2xy + e^y) = 2x + e^y \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x} \frac{\partial f}{\partial y} = \frac{\partial}{\partial x} (x^2 + xe^y) = 2x + e^y \\ \frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y} \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} (x^2 + xe^y) = xe^y\end{aligned}$$

You probably noticed that the two derivatives involving both  $x$  and  $y$  resulted in the same results. This is true for most functions as is shown in the next theorem: [Clairaut's Theorem] If  $f(x_1, x_2, \dots, x_n)$  is a real valued function with continuous second derivatives at the point  $(a_1, a_2, \dots, a_n)$ , then

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a_1, a_2, \dots, a_n) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a_1, a_2, \dots, a_n)$$

In other words, partial derivatives commute.

### 6.1.2 Notation for Partial Derivatives

In terms of notation, it's often common to use a subscript as a derivative. For example  $f_x$  can be used instead of  $\frac{\partial f}{\partial x}$  or  $f_y$  instead of  $\frac{\partial f}{\partial y}$ . This can also be extended to higher order

derivatives as the following shows:

$$\begin{array}{ll} f_{xx} = \frac{\partial^2 f}{\partial x^2} & f_{yy} = \frac{\partial^2 f}{\partial y^2} \\ f_{yx} = \frac{\partial^2 f}{\partial x \partial y} & f_{xy} = \frac{\partial^2 f}{\partial y \partial x} \\ f_{xxx} = \frac{\partial^3 f}{\partial x^3} & f_{yyx} = \frac{\partial^3 f}{\partial x \partial y^2} \end{array}$$

and note that the order on the variables switch between the two notations. This generally isn't a problem because of Clairaut's Theorem says that the derivative is independent of the order taken on the derivatives.

### 6.1.3 Differential Equations

A differential equation is an equation that contains derivatives. If the differential equation has partial derivatives, then the equation is called a **partial differential equation**, if not it is called an **ordinary differential equation**. The order of the PDE is the highest degree of any derivative. If the equation is linear in  $u$ , the dependent variable and its derivatives, then the equation is **linear**, if not, it is **nonlinear**. Lastly, a linear differential equation is called **homogeneous** if the only nonzero terms in the equation contain the dependent variable. If a differential equation is not homogeneous, then it is called **nonhomogeneous**.

The following are ordinary differential equations.

1.  $y' = x$
2.  $y' = \sin x$ .
3.  $\frac{dy}{dx} = y^2$
4.  $y'' + 3y' + y = 0$ .

Equations 1, 2, and 4 are linear and 3 is nonlinear. Note that even though  $\sin x$  is not a linear function, in order for a differential equation to be linear, it needs only be linear in its dependent variables (in all of these examples,  $y$  is dependent). Equations 1 and 2 are nonhomogeneous and equation 4 is homogeneous.

Also, the first three are first-order equations and the 4th equation is 2nd order.

The following are partial differential equations:

1.  $\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = 0$
2.  $u_t - uu_x = 0$ .
3.  $\frac{\partial u}{\partial x} + \left( \frac{\partial^2 u}{\partial t^2} \right)^2 = 0$

4.  $u_t + u_y - u_{xxx} = 0.$

And the first and fourth equations are linear with the other two being nonlinear. The first and second equations above are first order, the 3rd equation is 2nd order and the fourth is a third-order PDE.

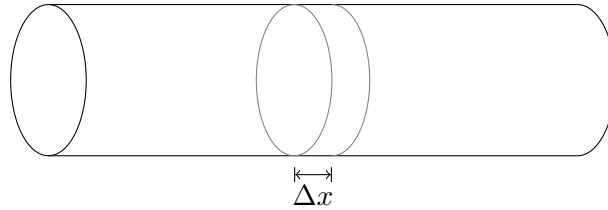
## 6.2 Modeling with PDEs

In this section, we show where two important differential equations arise based on modeling the physical situation. The first will be the heat transport or **heat equation** and the second will be the **wave equation** by modeling the motion of a taut string. In both situations we will only show this for motion in one-dimension although similar PDEs model more complex behavior.

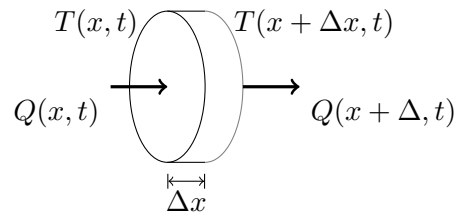
### 6.2.1 The Heat Equation

In this section, we will derive the transport of heat in a solid bar. In short, heat moves in a substance from a warmer region to a cooler region and if we know the direction of motion and the rate then we can understand heat flow.

We will consider a bar of length  $L$  with fixed density  $\rho$ , constant cross sectional area  $A$  and insulated sides. This derivation will result in flow only in the direction of the length of the bar. The shape of the cross section and material in the bar does not matter but we will show a cylindrical bar.



The small segment of rod has heat flux  $Q(x, t)$  entering in the left side and has heat flux  $Q(x + \Delta x, t)$  exiting the right side. The temperature is  $T(x, t)$  on the left and  $T(x + \Delta x, t)$  on the right.



Next, we consider the amount of heat in the segment. The heat is proportional to both the mass of the segment ( $\rho\Delta x$ ) and the temperature. The proportionality constant,  $c$  is

called the specific heat. The amount of heat in this small segment is:

$$c\rho T\Delta x$$

The temporal change in heat of this element is the derivative with respect to  $t$  and the only quantity that depends on time is the temperature  $T$ :

$$c\rho \frac{\partial T}{\partial t} \Delta x$$

and is equal to the amount of heat entering, which is  $Q(x, t)$  minus the amount of heat leaving or  $Q(x + \Delta x, t)$ .

$$c\rho \frac{\partial T}{\partial t} \Delta x = Q(x, t) - Q(x + \Delta x, t)$$

Divide both sides of this equation by  $\Delta x$  and take the limit  $\Delta x \rightarrow 0$ ,

$$c\rho \frac{\partial T}{\partial t} = \lim_{\Delta x \rightarrow 0} Q(x, t) - Q(x + \Delta x, t) = -\frac{\partial Q}{\partial x} \quad (6.1)$$

The relationship between the heat flux,  $Q$  and the temperature is called *Fourier's Law*, and states that the heat transferred across unit area is proportional to the temperature gradient or

$$Q = -k \frac{\partial T}{\partial x}$$

where  $k$  is the proportionality constant called the thermal conductivity and the negative sign is due to the fact that heat flows from hot to cold (in a negative direction). Substituting Fourier's Law into (6.1) results in

$$c\rho \frac{\partial T}{\partial t} = -\frac{\partial}{\partial x} \left( -k \frac{\partial T}{\partial x} \right)$$

and assuming that  $k$  is a constant,

$$c\rho \frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2}$$

rearranging we get

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial x^2}$$

where  $\kappa = k/c\rho$  is called the thermal diffusivity.

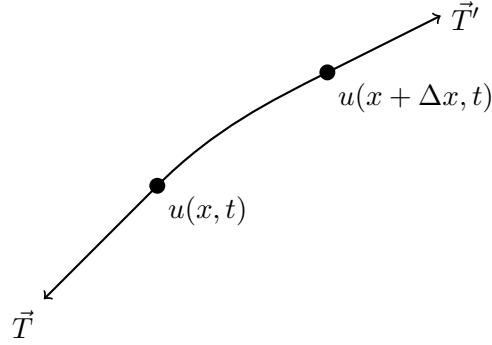
The **one-dimensional heat equation** is

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial x^2}$$

and is typically solved for a given initial condition  $T(x, 0) = f(x)$  and often with boundary conditions.

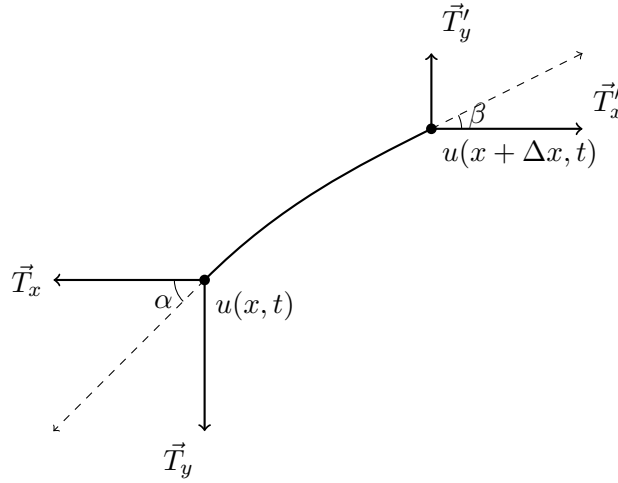
### 6.2.2 Wave Equation: Modeling a Taut String

Next, we turn to modeling the motion of a taut string fixed on both ends. A common example of this is a guitar string. We look at a small part of the string that has been displaced vertically from rest:



where  $u(x, t)$  be the vertical position of the string at position  $x$  and time  $t$  and the tension at  $x$  and  $x + \Delta x$  are the force vectors  $\vec{T}$  and  $\vec{T}'$ .

To determine an equation that describes the motion of the string, we need to balance the horizontal and vertical forces. We can write the two force vectors  $\mathbf{T}$  and  $\mathbf{T}'$  in terms of the horizontal and vertical components where the subscripts in this case are the components as shown in the figure below.



And recall that using Newton's 2nd law of motion the sum of the forces on a object is equal to the mass of the object times its acceleration in that direction or  $\sum F = ma$ . There is no horizontal motion of the string, so the horizontal acceleration is zero and therefore:

$$\begin{aligned} -\vec{T}_x + \vec{T}'_x &= 0 \\ -||\vec{T}|| \cos \alpha + ||\vec{T}'|| \cos \beta &= 0 \end{aligned}$$

and solving for  $||\vec{T}'||$

$$||\vec{T}'|| = ||\vec{T}'|| \frac{\cos \alpha}{\cos \beta} \quad (6.2)$$

Next, we examine the vertical forces. The main difference is that there is a vertical acceleration which is the 2nd derivative of the position  $u$  with respect to  $t$ .

$$\begin{aligned} \sum F &= ma \\ -\vec{T}_y + \vec{T}'_y &= ma \\ -||\vec{T}'|| \sin \alpha + ||\vec{T}'|| \sin \beta &= \rho \Delta x \frac{\partial^2 u}{\partial t^2} \end{aligned}$$

where the mass of the short piece is  $\rho$  the mass density of the string and  $\Delta x$  is the horizontal length of the piece.

Substituting (??) into the equation above,

$$-||\vec{T}'|| \sin \alpha + ||\vec{T}'|| \frac{\cos \alpha}{\cos \beta} \sin \beta = \rho \Delta x \frac{\partial^2 u}{\partial t^2}$$

and dividing through by  $||\vec{T}'|| \cos \alpha$

$$\begin{aligned} -\frac{\sin \alpha}{\cos \alpha} + \frac{\sin \beta}{\cos \beta} &= \frac{\rho \Delta x}{||\vec{T}'|| \cos \alpha} \frac{\partial^2 u}{\partial t^2} \\ -\tan \alpha + \tan \beta &= \frac{\rho \Delta x}{||\vec{T}'|| \cos \alpha} \frac{\partial^2 u}{\partial t^2} \end{aligned}$$

The tangent of the angles are the slopes of the function  $u(x, t)$  at  $x + \Delta x$  and  $x$  respectively, or  $\frac{\partial u}{\partial x}$  at  $x + \Delta x$  and  $x$  therefore the above can be written

$$\frac{\partial u}{\partial x} \Big|_{x+\Delta x} - \frac{\partial u}{\partial x} \Big|_x = \frac{\rho \Delta x}{||\vec{T}'|| \cos \alpha} \frac{\partial^2 u}{\partial t^2}$$

Divide by  $\Delta x$  and let  $T = ||\vec{T}'|| \cos \alpha$  be the horizontal tension in the string

$$\frac{1}{\Delta x} \left( \frac{\partial u}{\partial x} \Big|_{x+\Delta x} - \frac{\partial u}{\partial x} \Big|_x \right) = \frac{\rho}{T} \frac{\partial^2 u}{\partial t^2}$$

take the limit as  $\Delta x \rightarrow 0$

$$\frac{\partial^2 u}{\partial x^2} = \frac{\rho}{T} \frac{\partial^2 u}{\partial t^2}$$

The parameter  $\rho/T$  plays an important role in the solution, and we'll define  $c = \sqrt{T/\rho}$ , so the equation can be written:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \quad (6.3)$$

Typically, (6.3) is called the **wave equation** or since there is only one spatial variable, the **one-dimensional wave equation**. The remainder of this chapter discusses solutions of the wave equation and related equations.

### 6.3 Linear ordinary differential equations

As discussed in section 6.1, a differential equation is linear if it is linear in its dependent variable and derivatives of the dependent variable. Let's assume that  $y$  is the dependent variable and  $x$  is the independent variable. The most general  $n$ th order ordinary linear differential equation has the form:

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y = f(x) \quad (6.4)$$

In the most general terms, this differential equation is very difficult to solve despite being linear. However, there are cases when this has nice solutions. For the most part, we'll look at such equations in this section.

#### 6.3.1 Second-order Constant-Coefficient Homogeneous differential equations.

The most important ordinary differential equations that arise from solving the wave and heat equation are 2nd-order constant coefficient homogeneous differential equations. If  $f(x) \equiv 0$  in (6.4), then it is called homogeneous and if the coefficients of the  $y$  terms and its derivatives are not dependent on  $x$ , then the equation is constant-coefficient. A general 2nd-order ODE with these characteristics can be written:

$$ay'' + by' + cy = 0$$

where  $a, b, c$  are real constants. To solve these, if we let  $y = e^{rx}$  and substitute this into the ODE,

$$\begin{aligned} ar^2e^{rx} + bre^{rx} + ce^{rx} &= 0 \\ e^{rx}(ar^2 + br + c) &= 0 \end{aligned}$$

and since  $e^{rx}$  is never 0

$$ar^2 + br + c = 0$$

is called the **characteristic equation** related to the ODE. In general, there are two solutions to this equation,  $r_1$  and  $r_2$ . The general solution to the equation will thus be

$$y = c_1e^{r_1x} + c_2e^{r_2x}$$



Find the general solution to

$$y'' + 3y' + 2y = 0$$

The characteristic equation is

$$\begin{aligned} r^2 + 3r + 2 &= 0 \\ (r + 2)(r + 1) &= 0 \end{aligned}$$

which has solutions  $r = -2$  and  $r = -1$ . The general solution thus is

$$y = c_1 e^{-2x} + c_2 e^{-x}$$

To find the values of  $c_1$  and  $c_2$  in the example above or any other differential equation, additional information must be given. This is analogous to using a point to find the integration constant from an indefinite integral. The information is generally framed in the language of initial conditions and initial value problems as defined below.

A differential equation has **initial conditions** if at some point  $t_0$ , then  $y(t_0)$ ,  $y'(t_0)$ ,  $\dots$ ,  $y^{(n-1)}(t_0)$  are known, where  $n$  is the order of the differential equation. The differential equation together with the initial conditions is known as a **initial value problem**.

The next example shows how to fully solve an initial value problem:

Find the solution to the initial value problem

$$y'' + 3y' + 2y = 0 \quad y(0) = 3, y'(0) = -4.$$

To solve the initial value problem we first solve the differential equation, and this was done in example ???. And the general solution is

$$y = c_1 e^{-2x} + c_2 e^{-x} \tag{6.5}$$

Since the initial conditions include knowing the derivative, we will need

$$y' = -2c_1 e^{-2x} - c_2 e^{-x} \tag{6.6}$$

and then we will substitute  $y(0) = 3$  into (??) and  $y'(0) = -4$  into (??) or

$$\begin{aligned} 3 &= c_1 + c_2 \\ -4 &= -2c_1 - c_2 \end{aligned}$$

and the solution to this is  $c_1 = 1$  and  $c_2 = 2$ . Therefore the solution to the initial value problem is

$$y = e^{-2x} + 2e^{-x}$$

The next example shows a differential equation that we will see often in this chapter.

Find the general solution to  $y'' + 9y = 0$ .

The characteristic equation is

$$\begin{aligned} r^2 + 9 &= 0 \\ r &= \pm 3i \end{aligned}$$

Thus the general solution is

$$y = c_1 e^{3ix} + c_2 e^{-3ix}$$

However, we will use Euler's Formula:

$$e^{ki\theta} = \cos k\theta + i \sin k\theta$$

to write the equation as a real function:

$$\begin{aligned} y &= c_1(\cos 3x + i \sin 3x) + c_2(\cos 3x - i \sin 3x) \\ &= (c_1 + c_2) \cos 3x + (ic_1 - ic_2) \sin 3x \end{aligned}$$

letting  $C_1 = c_1 + c_2$ ,  $C_2 = ic_1 - ic_2$

$$y = C_1 \cos 3x + C_2 \sin 3x$$

The characteristic equation for this next example has complex roots.

Find the solution to  $y'' + 4y' + 5y = 0$ .

The characteristic equation for this differential equation is

$$r^2 + 4r + 5 = 0$$

and since this can't be factored, we'll use the quadratic equation to solve this:

$$r = \frac{-4 \pm \sqrt{16 - 20}}{2} = -2 \pm i$$

and using  $y = e^{rx}$  the general solution to this would be written

$$\begin{aligned} y &= c_1 e^{(-2+i)x} + c_2 e^{(-2-i)x} \\ &= e^{-2x} (c_1 e^{ix} + c_2 e^{-ix}) \\ &= e^{-2x} (c_1 (\cos x + i \sin x) + c_2 (\cos x - i \sin x)) \\ &= e^{-2x} (C_1 \cos x + C_2 \sin x) \end{aligned}$$

where  $C_1 = c_1 + c_2$  and  $C_2 = i(c_1 - c_2)$ .

Consider the general 2nd-order constant-coefficient linear homogeneous differential equation,

$$ay'' + by' + cy = 0 \tag{6.7}$$

and let  $r_1$  and  $r_2$  be the roots of the characteristic equation.

1. If  $r_1$  and  $r_2$  are real and distinct, then the general solution to (??) is

$$y = c_1 e^{r_1 x} + c_2 e^{r_2 x} \quad (6.8)$$

2. If  $r_1 = r_2$  is the only real root of the characteristic equation then the general solution to (??) is

$$y = e^{r_1 x} (c_1 + c_2 x) \quad (6.9)$$

3. If the roots of the characteristic equation are pure imaginary, that is  $r_1 = \beta i$  and  $r_2 = -\beta i$ , then the solution to (??) is

$$y = c_1 \sin \beta x + c_2 \cos \beta x \quad (6.10)$$

4. and if the roots of the characteristic equation are complex conjugates, that is  $r_1 = \alpha + \beta i$  and  $r_2 = \alpha - \beta i$ , then the solution to (??) is

$$y = e^{\alpha x} (c_1 \sin \beta x + c_2 \cos \beta x) \quad (6.11)$$

*Proof.* First, recall that the characteristic equation of (??) is

$$ar^2 + br + c = 0 \quad (6.12)$$

We will prove each statement in order.

1. The derivatives of (??) are

$$\begin{aligned} y' &= c_1 r_1 e^{r_1 x} + c_2 r_2 e^{r_2 x} \\ y'' &= c_1 r_1^2 e^{r_1 x} + c_2 r_2^2 e^{r_2 x} \end{aligned}$$

and substituting these into (??),

$$\begin{aligned} ay'' + by' + cy &= a(c_1 r_1^2 e^{r_1 x} + c_2 r_2^2 e^{r_2 x}) + b(c_1 r_1 e^{r_1 x} \\ &\quad + c_2 r_2 e^{r_2 x}) + c(c_1 e^{r_1 x} + c_2 e^{r_2 x}) \end{aligned}$$

rearranging

$$= c_1 (ar_1^2 + br_1 + c) e^{r_1 x} + c_2 (ar_2^2 + br_2 + c) e^{r_2 x}$$

and since  $r_1$  and  $r_2$  are roots of the characteristic equation, then both terms in parentheses are zero and

$$ay'' + by' + cy = 0$$

2. The derivatives of (??) are

$$\begin{aligned}y' &= e^{r_1 x}(r_1 c_1 + c_2 + r_1 c_2 x) \\y'' &= e^{r_1 x}(r_1^2 c_1 + 2r_1 c_2 + r_1^2 c_2 x)\end{aligned}$$

where the product rule has been used. In this case, recall also that since  $r_1$  is the only real root, that  $r_1 = -b/(2a)$  from the quadratic formula with the discriminant  $b^2 - 4ac = 0$ . Substituting into (??)

$$\begin{aligned}ay'' + by' + cy &= ae^{r_1 x}(r_1^2 c_1 + 2r_1 c_2 + r_1^2 c_2 x) \\&\quad + be^{r_1 x}(r_1 c_1 + c_2 + r_1 c_2 x) + c(c_1 + c_2 x)e^{r_1 x} \\&= (c_1(ar_1^2 + br_1 + c) + c_2(2ar_1 + b) \\&\quad + c_2x(ar_1^2 + br_1 + c))e^{r_1 x} = 0\end{aligned}$$

where the term with factor  $c_1$  and the term with factor  $c_2 x$  are each 0, because  $r_1$  is a root of the characteristic equation. Lastly, the term with factor  $c_2$  is 0 because as stated about  $r_1 = -b/(2a)$ .

3. Note that since the roots are pure imaginary, then then  $b = 0$  in (??) and therefore  $r_1$  and  $r_2$  satisfy  $r^2 = -c/a$  and since  $r_1 = \beta i$ , then  $\beta^2 = c/a$ . The derivatives of (??) are

$$\begin{aligned}y' &= c_1 \beta \cos \beta x - c_2 \beta \sin \beta x \\y'' &= -c_1 \beta^2 \sin \beta x - c_2 \beta^2 \cos \beta x\end{aligned}$$

and then substituting into (??) with  $b = 0$ ,

$$\begin{aligned}ay'' + cy &= a(-c_1 \beta^2 \sin \beta x - c_2 \beta^2 \cos \beta x) + c(c_1 \sin \beta x + c_2 \cos \beta x) \\&= c_1(-a\beta^2 + c) \sin \beta x + c_2(-a\beta^2 + c) \cos \beta x = 0\end{aligned}$$

since  $\beta^2 = c/a$ .

4. In this case, since the roots are complex conjugates, we can write  $r_1 = \alpha + i\beta$  and  $r_2 = \alpha - i\beta$ . The characteristic equation can be written:

$$\begin{aligned}a(r - r_1)(r - r_2) &= a(r - (\alpha + i\beta))(r - (\alpha - i\beta)) \\&= ar^2 - 2a\alpha r + a(\alpha^2 + \beta^2) = 0\end{aligned}$$

which implies that  $b = -2a\alpha$  and  $c = a(\alpha^2 + \beta^2)$ . The derivatives of (??) are

$$\begin{aligned}y' &= e^{\alpha x}((c_1 \alpha - c_2 \beta) \sin \beta x + (c_2 \alpha + c_1 \beta) \cos \beta x) \\y'' &= e^{\alpha x}((c_1 \alpha^2 - 2c_2 \alpha \beta - c_1 \beta^2) \sin \beta x \\&\quad + (c_2 \alpha^2 + 2c_1 \alpha \beta - c_2 \beta^2) \cos \beta x)\end{aligned}$$

and substituting them into (??),

$$\begin{aligned}
 ay'' + by' + cy &= ae^{\alpha x}((c_1\alpha^2 - 2c_2\alpha\beta - c_1\beta^2)\sin\beta x \\
 &\quad + (c_2\alpha^2 + 2c_1\alpha\beta - c_2\beta^2)\cos\beta x) \\
 &\quad + be^{\alpha x}((c_1\alpha - c_2\beta)\sin\beta x + (c_2\alpha + c_1\beta)\cos\beta x) \\
 &\quad + ce^{\alpha x}(c_1\sin\beta x + c_2\cos\beta x) \\
 &= e^{\alpha x}(c_1(a\alpha^2 - a\beta^2 + b\alpha + c)\sin\beta x \\
 &\quad + c_1(2a\alpha\beta + b\beta)\cos\beta x + c_2(-2a\alpha\beta - b\beta)\sin\beta x \\
 &\quad + c_2(a\alpha^2 - a\beta^2 + b\alpha + c)) \tag{6.13}
 \end{aligned}$$

And using  $b = -2a\alpha$  and  $c = a(\alpha^2 + \beta^2)$ , the right hand side of (??) is 0, therefore (??) is a solution.

□

### 6.3.2 Solutions to Ordinary Differential Equations as Vector Spaces

We studied vector spaces in section ?? and a common example of a vector space that is not  $\mathbb{R}^n$  is that of functions. It can be shown that the set related to the differential equation in Example ??

$$V = \{y \in \mathcal{C}^{(2)}(\mathbb{R}) \mid y'' + 9y = 0\}$$

is a vector space or a subspace of  $\mathcal{C}^{(2)}$ , the vector space of all functions with continuous 2nd derivatives. The basis of the subspace are the two solutions of the differential equation or  $B = (\cos 3x, \sin 3x)$ . Note that the general solution to the differential equation is the linear combination of the basis vectors or the span of the two solutions.

## 6.4 Sturm-Liouville Problems

We now turn to a class of differential equations that arise in solving partial differential equations. This class is called Sturm-Liouville problems and they satisfy boundary conditions instead of the initial conditions that we saw in the previous section.

Consider the differential equation

$$[r(x)y']' + [p(x) + \lambda q(x)]y = 0$$

for  $r(x), p(x) \in \mathcal{C}^{(1)}[a, b]$ ,  $q(x) \in \mathcal{C}^{(0)}[a, b]$ , and  $\lambda \in \mathbb{R}$ . The differential equation is also subject to the boundary conditions:

$$\begin{aligned}
 a_1y(a) + a_2y'(a) &= 0 \\
 b_1y(b) + b_2y'(b) &= 0
 \end{aligned}$$

such that both  $a_1$  and  $a_2$  cannot be zero as well as both  $b_1$  and  $b_2$ .

The differential equation with these boundary conditions are called a **Sturm-Liouville Problem**. The solution is called an **eigenfunction** of the problem and the values of  $\lambda$  are called the **eigenvalues**.

The following example shows how to solve a Sturm-Liouville problem, that is, find the eigenvalues and eigenfunctions of the problem.

Find the eigenvalue and eigenfunctions of the Sturm-Liouville problem:

$$y'' + \lambda y = 0 \qquad y(0) = y(L) = 0$$

First note that  $r(x) = q(x) \equiv 1$  and also  $a_1 = b_1 = 1$  and  $a_2 = b_2 = 0$ , which satisfies the conditions on the boundaries.

The characteristic equation for this problem is:

$$r^2 + \lambda = 0$$

which has the solutions  $r = \pm\sqrt{-\lambda}$ . The form of the equation depends on  $\lambda$ . If  $\lambda = 0$ , we get:

$$y = A + Bx \tag{6.14}$$

if  $\lambda < 0$ , then the solution is

$$y = c_1 e^{\sqrt{-\lambda}x} + c_2 e^{-\sqrt{-\lambda}x} \tag{6.15}$$

and if  $\lambda > 0$ , then we get

$$y = C_1 \cos \sqrt{\lambda}x + C_2 \sin \sqrt{\lambda}x \tag{6.16}$$

Next, we apply the boundary conditions on all three solutions. Recall that  $y(0) = y(L) = 0$ .

If  $\lambda = 0$ , substituting the boundary conditions into (??),

$$\begin{aligned} A + B(0) &= 0 & \text{so } A &= 0 \\ A + B(L) &= 0 & \text{so } B &= 0 \end{aligned}$$

so the only solution to (??) is the trivial solution  $y \equiv 0$ .

If  $\lambda > 0$ , then substituting the boundary conditions into (??) results in

$$\begin{aligned} c_1(1) + c_2(1) &= 0 \\ c_1(e^{\sqrt{-\lambda}L} + c_2 e^{-\sqrt{-\lambda}L}) &= 0 \end{aligned}$$

From the first equation,  $c_1 = -c_2$  and substituting this into the 2nd equation

$$\begin{aligned} c_1 e^{\sqrt{-\lambda}L} - c_1 e^{-\sqrt{-\lambda}L} &= 0 \\ c_1 e^{\sqrt{-\lambda}L} (1 - e^{-2\sqrt{-\lambda}L}) &= 0 \end{aligned}$$

$c_1 = 0$  is a solution, the second term is never zero and the third term is only zero if  $L = 0$ , which is not true or  $\lambda = 0$ , which is also not true, since this case is  $\lambda > 0$ . Therefore again, the only solution to (??) is the trivial solution.

If  $\lambda < 0$  then substituting the boundary conditions into (??) results in

$$\begin{aligned} C_1(1) + C_2(0) &= 0 & \text{so } C_1 &= 0 \\ (0) \cos \sqrt{\lambda}L + C_2 \sin \sqrt{\lambda}L &= 0 \end{aligned}$$

the second states that either  $C_2 = 0$ , again the trivial solution or

$$\sin \sqrt{\lambda}L = 0$$

which occurs if  $\sqrt{\lambda}L = n\pi$  for  $n = 0, \pm 1, \pm 2, \pm 3, \dots$ . Or

$$\lambda = \frac{n^2\pi^2}{L^2} \quad (6.17)$$

We now check which values of  $n$  result in valid values of  $\lambda$ .

When  $n = 0$ , we get  $\lambda = 0$  again, which violates  $\lambda < 0$  and for  $n$  both plus and minus the same number, we get the same eigenvalue, so we will discard the negative values of  $n$  and just include  $n = 1, 2, 3, \dots$ .

The eigenvalues of this problem are those in (??) for  $n = 1, 2, 3, \dots$  and the eigenfunctions of this problem are:

$$\phi_n(x) = \sin \frac{n\pi x}{L}$$

It may have appears that we were lucky that there was a solution to the Sturm-Liouville problem in the above example. However, this is not the case and any Sturm-Liouville problem has a solution as the following theorem shows.

Let  $\lambda_n$  and  $\phi_n(x)$  be any eigenvalue and eigenfunction of the Sturm Liouville problem.

1. The eigenvalue  $\lambda_n$  is real.
2. There are an infinite number of eigenvalues that can be ordered  $\lambda_1 < \lambda_2 < \lambda_3 < \dots$  and for each eigenvalue, there is only one eigenfunction.
3. Eigenfunctions  $\phi_n$  and  $\phi_m$  with  $\lambda_n \neq \lambda_m$  satisfy  $\langle \phi_n, \phi_m \rangle = 0$ .
4. Let  $f$  and  $f'$  be piecewise continuous functions on  $[a, b]$ . If

$$a_n = \frac{\langle f, \phi_n \rangle}{\langle \phi_n, \phi_n \rangle},$$

then the series:

$$\sum_{n=1}^{\infty} a_n \phi_n(x)$$

converges to  $f$  if  $f$  is continuous at  $x$  and to the value  $(f(x+) + f(x-))/2$  if  $f$  is discontinuous at  $x$  for each point in  $[a, b]$ .

Let  $f(x) = x$  on  $[0, L]$ . Find the series expansion listed in the theorem corresponding to the Sturm-Liouville problem  $y'' + \lambda y = 0$ .

$$\begin{aligned} a_n &= \frac{\langle f, \phi_n \rangle}{\langle \phi_n, \phi_n \rangle} \\ &= \frac{1}{L/2} \int_0^L f(x) \sin \frac{n\pi x}{L} dx \end{aligned}$$

which is the Fourier sine series.

Find the solution of the Sturm-Liouville problem

$$y'' + \lambda y = 0 \qquad y'(0) = 0, \qquad y'(L) = 0$$

Since this the same differential equation as in Example ??, we note that when  $\lambda > 0$ , there was no solution and the same is true here. In the case of  $\lambda = 0$ , the solution is

$$y = A + Bx$$

and the derivative is needed as well,

$$y' = B$$

and then applying the boundary condition  $y'(0) = y'(L) = 0$ , implies that  $B = 0$ , however,  $A$  is not determined and  $y = A$  is a solution.

Next, we turn to  $\lambda > 0$ , with the solution,

$$y = C_1 \cos \sqrt{\lambda}x + C_2 \sin \sqrt{\lambda}x$$

and again, we need the derivative,

$$y' = -C_1 \sqrt{\lambda} \sin \sqrt{\lambda}x + C_2 \sqrt{\lambda} \cos \sqrt{\lambda}x$$

Applying the boundary condition,  $y'(0) = 0$ , results in

$$y'(0) = C_2 \sqrt{\lambda} = 0$$

which implies that  $C_2 = 0$ . Applying the boundary condition  $y'(L) = 0$  results in

$$y'(L) = -C_1 \sqrt{\lambda} \sin \sqrt{\lambda}L = 0$$

and if  $C_1 = 0$ , this results in the trivial solution,  $\lambda$  cannot be zero, however

$$\sin \sqrt{\lambda}L = 0$$

when  $\sqrt{\lambda}L = n\pi$  or  $\lambda = n^2\pi^2/L^2$ .

The eigenvalues and eigenfunctions of this problem then are  $\lambda_0 = 0$  and  $\phi_0 = 1$ , and

$$\lambda_n = \frac{n^2\pi^2}{L^2} \qquad \phi_n = \cos \frac{n\pi x}{L}$$

There are other Sturm-Liouville problems that arise commonly and we will see others later in this chapter and solve them as they arise. We will use these solutions that we just found in solving the PDEs that we derived above.



## 6.5 Solving the 1D Wave Equation Using Separation of Variables

We derived the 1D wave equation in section ???. We will now solve this as an initial value problem. Consider

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

with boundary conditions

$$u(0, t) = u(L, t) = 0, \quad \text{for } t \geq 0$$

and initial conditions

$$\begin{aligned} u(x, 0) &= f(x) & \text{for } 0 \leq x \leq L \\ u_t(x, 0) &= 0 & \text{for } 0 \leq x \leq L \end{aligned}$$

The technique of **separation of variables** is the standard way to solve this and other linear PDEs. First start by assuming that the solution  $u(x, t)$  can be written as a product of functions solely of  $x$  and  $t$  or  $u(x, t) = X(x)T(t)$ . Then substituting this into the PDE:

$$\begin{aligned} \frac{\partial^2(X(x)T(t))}{\partial t^2} &= c^2 \frac{\partial^2(X(x)T(t))}{\partial x^2} \\ XT'' &= c^2 X''T & \text{dividing by } c^2 XT \\ \frac{1}{c^2} \frac{T''}{T} &= \frac{X''}{X} \end{aligned}$$

Now since  $T$  (and therefore  $T''$ ) depends only on  $t$  and  $X$  (and also  $X''$ ) only depends on  $x$ , each side of the PDE must be equal to a constant (independent of either  $x$  or  $t$ ). Let's say it is  $-\lambda$  or

$$\frac{1}{c^2} \frac{T''}{T} = \frac{X''}{X} = -\lambda$$

This results in two equations:

$$T'' + c^2 \lambda T = 0 \tag{6.18}$$

$$X'' + \lambda X = 0 \tag{6.19}$$

The boundary conditions are  $u(0) = u(L) = 0$  for all  $t$ , so substituting  $u = XT$  we get

$$X(0)T(t) = X(L)T(t) = 0$$

for all  $t$  or  $X(0) = X(L) = 0$ . The ODE for  $X$  is that in (??) with these boundary conditions. This is a Sturm-Liouville problem that was solved in example ??. We know

that the only form of the solution is when  $\lambda > 0$  and from the problem that we solved above we know that

$$\lambda = \frac{n^2\pi^2}{L^2} \quad (6.20)$$

and the solution for  $X$  are the eigenfunctions of the problem:

$$X_n(x) = \sin \frac{n\pi x}{L}$$

for  $n = 1, 2, 3, \dots$

Next, we need to solve (??) and using the eigenvalues in (??)

$$T'' + c^2 \frac{n^2\pi^2}{L^2} T = 0$$

The solution to this is found by assuming  $T(t) = e^{rt}$  and getting the characteristic equation

$$r^2 + c^2 \frac{n^2\pi^2}{L^2} = 0$$

and thus

$$r = \pm i \frac{nc\pi}{L}$$

and the solution to this is

$$T_n(t) = a_n \cos\left(\frac{nc\pi}{L}t\right) + b_n \sin\left(\frac{nc\pi}{L}t\right)$$

We know then that the following is a solution to the PDE:

$$u_n(x, t) = X_n(x)T_n(t) = \left(a_n \cos\left(\frac{nc\pi}{L}t\right) + b_n \sin\left(\frac{nc\pi}{L}t\right)\right) \sin \frac{n\pi x}{L}$$

The **principle of superposition** states that a sum of these solutions is also a solution to the original PDE and in fact the most general solution is

$$\begin{aligned} u(x, t) &= \sum_{n=1}^{\infty} u_n(x, t) \\ &= \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{nc\pi}{L}t\right) + b_n \sin\left(\frac{nc\pi}{L}t\right)\right) \sin \frac{n\pi x}{L} \end{aligned}$$

Lastly, we use the initial conditions  $u(x, 0) = f(x)$  and  $u_t(x, 0) = 0$  for  $0 \leq x \leq L$  to solve for  $a_n$  and  $b_n$  above. We will look at the second condition first:

$$u_t(x, t) = \sum_{n=1}^{\infty} \frac{nc\pi}{L} \left(-a_n \sin\left(\frac{nc\pi}{L}t\right) + b_n \cos\left(\frac{nc\pi}{L}t\right)\right) \sin \frac{n\pi x}{L}$$

and evaluated at  $t = 0$ :

$$u_t(x, 0) = \sum_{n=1}^{\infty} \frac{nc\pi}{L} \left(0 + b_n\right) \sin \frac{n\pi x}{L}$$

and since  $u_t(x, 0) \equiv 0$ , then  $b_n = 0$ . The first condition:

$$u(x, 0) = f(x) = \sum_{n=1}^{\infty} a_n \sin \frac{n\pi x}{L}$$

is the Fourier sine series for  $0 \leq x \leq L$ . Thus

$$a_n = \frac{2}{L} \int_0^L f(x) \sin \frac{n\pi x}{L} dx$$

We summarize the solution to the wave equation.

The solution to the wave equation

$$u_{tt} = c^2 u_{xx}$$

with boundary conditions

$$u(0, t) = u(L, t) = 0$$

and initial conditions

$$\begin{aligned} u(x, 0) &= f(x), \\ u_t(x, 0) &= 0 \end{aligned}$$

is

$$u(x, t) = \sum_{n=1}^{\infty} a_n \cos\left(\frac{nc\pi}{L}t\right) \sin \frac{n\pi x}{L}$$

where

$$a_n = \frac{2}{L} \int_0^L f(x) \sin \frac{n\pi x}{L} dx$$

The following two examples show specific solutions for given functions  $f(x)$ .

Find the specific solution to the wave equation if  $f(x) = 0.1 \sin(\pi x/L)$

Specifically, we only need to find the coefficients  $a_n$  above:

$$\begin{aligned} a_n &= \frac{2}{L} \int_0^L 0.1 \sin \frac{\pi x}{L} \sin \frac{n\pi x}{L} dx \\ &= \begin{cases} 0.1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Thus the full solution to the PDE is:

$$u(x, t) = 0.1 \sin \frac{\pi x}{L} \cos \frac{\pi t}{L}$$

Find the solution to the wave equation above if the initial shape of the string is

$$f(x) = \begin{cases} kx, & \text{if } 0 \leq x < L/2 \\ k(L - x), & \text{if } L/2 \leq x \leq L \end{cases}$$

As before, we only need to find the coefficients of  $a_n$  given by

$$\begin{aligned} a_n &= \frac{2}{L} \int_0^L f(x) \sin \frac{n\pi x}{L} dx \\ &= \frac{2}{L} \int_0^{L/2} kx \sin \frac{n\pi x}{L} dx + \frac{2}{L} \int_{L/2}^L k(L - x) \sin \frac{n\pi x}{L} dx \\ &= \begin{cases} \frac{4kL}{n^2\pi^2} (-1)^{(n-1)/2}, & \text{if } n \text{ is odd} \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

And thus the full solution to the wave equation is:

$$\begin{aligned} u(x, t) &= \sum_{n=1}^{\infty} a_n \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{n\pi t}{L}\right) \\ &= \sum_{n=1}^{\infty} \frac{4kL(-1)^{n+1}}{\pi^2(2n-1)^2} \sin\left(\frac{(2n-1)\pi x}{L}\right) \sin\left(\frac{(2n-1)\pi t}{L}\right) \end{aligned}$$

### 6.5.1 Solution to the Wave Equation with a free boundary condition

We now look at the wave equation with what is termed a free boundary condition. Consider

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

with boundary conditions

$$u(0, t) = 0 \quad u_x(L, t) = 0, \quad \text{for } t \geq 0 \quad (6.21)$$

and initial conditions

$$u(x, 0) = f(x) \quad \text{for } 0 \leq x \leq L \quad (6.22)$$

$$u_t(x, 0) = 0 \quad \text{for } 0 \leq x \leq L \quad (6.23)$$

and the difference between this and the initial value problem at the beginning of the section is the boundary condition at  $x = L$ . Above, the function  $u$  was forced to be 0 there and in this case, the derivative is 0 at  $x = L$ .

Using separation of variables by assuming that  $u(x, t) = X(x)T(t)$ , and substituting into the wave equation and dividing through by  $c^2XT$ , we again get the equation

$$\frac{1}{c^2} \frac{T''}{T} = \frac{X''}{X} = -\lambda^2$$

where again since each term only depends on  $t$  or  $x$ , it must be a constant, which we say is  $-\lambda$ . The differential equation in  $X$  is

$$X'' + \lambda X = 0 \qquad X(0) = 0, X'(L) = 0$$

where the boundary conditions came from (??). The solution to this Sturm-Liouville problem is similar to those in Examples ?? and ??, however the details of this are not shown. The eigenfunctions and eigenvalues of this are

$$X_n(x) = \sin \lambda_n x \qquad \lambda_n = \frac{(2n-1)\pi}{2L}$$

Next, we seek the solution to  $T'' + c^2\lambda^2T = 0$  which again, similar to solution of the wave equation with fixed boundary conditions, is

$$T = a \cos c\lambda t + b \sin c\lambda t$$

which has a solution for each value of  $n$  or

$$T_n = a_n \cos c \frac{(2n-1)\pi}{2L} t + b_n \sin c \frac{(2n-1)\pi}{2L} t$$

Using the principle of superposition, the sum of solutions is also a solution, so

$$u(x, t) = \sum_{n=1}^{\infty} \left( a_n \cos c \frac{(2n-1)\pi}{2L} t + b_n \sin c \frac{(2n-1)\pi}{2L} t \right) \sin \frac{(2n-1)\pi x}{2L} \quad (6.24)$$

To find the particular solution, we need to find the constants  $a_n$  and  $b_n$  which can be found using the initial conditions of the problem. Similar to the problem above with the fixed boundary conditions, we will use the initial condition in (??) first, which requires that we know the derivative. And

$$u_t = \sum_{n=1}^{\infty} c\lambda_n \left( -a_n \sin c \frac{(2n-1)\pi}{2L} t + b_n \cos c \frac{(2n-1)\pi}{2L} t \right) \sin \frac{(2n-1)\pi x}{2L}$$

Applying the initial condition in (??),

$$u_t(x, 0) = 0 = \sum_{n=1}^{\infty} b_n c\lambda_n \sin \frac{(2n-1)\pi x}{2L}$$

which implies that  $b_n = 0$  for all  $n$ . Next, if we substitute this into (??) and apply the boundary condition in (??),

$$u(x, 0) = f(x) = \sum_{n=1}^{\infty} a_n \sin \frac{(2n-1)\pi x}{2L}$$

and the coefficients can be found using Theorem ??, to be

$$\begin{aligned} a_n &= \frac{\langle f(x), X_n(x) \rangle}{\langle X_n(x), X_n(x) \rangle} \\ &= \frac{\int_0^L f(x) \sin(2n-1)\pi x / (2L) dx}{\int_0^L \sin^2(2n-1)\pi x / (2L) dx} \\ &= \frac{2}{L} \int_0^L f(x) \sin \frac{(2n-1)\pi x}{2L} dx \end{aligned}$$

To summarize, the solution to the 1D wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

with boundary conditions

$$u(0, t) = 0 \quad u_x(L, t) = 0, \quad \text{for } t \geq 0$$

and initial conditions

$$\begin{aligned} u(x, 0) &= f(x) \quad \text{for } 0 \leq x \leq L \\ u_t(x, 0) &= 0 \quad \text{for } 0 \leq x \leq L \end{aligned}$$

is

$$u(x, t) = \sum_{n=1}^{\infty} a_n \cos c \frac{(2n-1)\pi}{2L} t \sin \frac{(2n-1)\pi x}{2L} \quad (6.25)$$

where

$$a_n = \frac{2}{L} \int_0^L f(x) \sin \frac{(2n-1)\pi x}{2L} dx \quad (6.26)$$

Find the solution to the 1D wave equation with free boundary condition at  $x = 1$  if

$$f(x) = x^2 \quad \text{for } 0 \leq x \leq 1$$

In this case, we only need to find the constants  $a_n$  as defined in (??),

$$\begin{aligned} a_n &= 2 \int_0^1 x^2 \sin \frac{(2n-1)\pi x}{2} dx \\ &= \frac{(-1)^n (8\pi + 16\pi n) - 16}{(2n-1)^3 \pi^3} \end{aligned}$$

So the full solution to this wave equation is

$$u(x, t) = \sum_{n=1}^{\infty} \frac{(-1)^n (8\pi + 16\pi n) - 16}{(2n-1)^3 \pi^3} \cos c \frac{(2n-1)\pi}{2L} t \sin \frac{(2n-1)\pi x}{2L}$$

## 6.6 Solution to the 1D Heat Equation

In this section we will investigate solving the 1D heat equation

$$\frac{1}{\kappa} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

with boundary conditions:

$$\left. \frac{\partial u}{\partial x} \right|_{x=0} = 0 \qquad \qquad \qquad \left. \frac{\partial u}{\partial x} \right|_{x=L} = 0$$

indicating that the end are insulated and the initial condition is:

$$u(x, 0) = f(x), \quad \text{if } 0 < x < L$$

As with the wave equation, we use the technique of separation of variables. That is let  $u = X(x)T(t)$  to get:

$$\begin{aligned} \frac{1}{\kappa} T' X &= T X'' && \text{divide through by } XT \\ \frac{1}{\kappa} \frac{T'}{T} &= \frac{X''}{X} \end{aligned}$$

since the left side only depends on  $t$  and the right side only depends on  $X$ , it must be that these must both equal only a constant, (call it  $-\lambda$ ) therefore we get the two equations:

$$\frac{1}{\kappa} \frac{T'}{T} = -\lambda \qquad \qquad \qquad \frac{X''}{X} = -\lambda$$

or

$$T' + \kappa \lambda T = 0 \qquad \qquad \qquad X'' + \lambda X = 0$$

The boundary conditions for the second equation becomes:

$$X'(0) = X'(L) = 0$$

This is a Sturm-Liouville problem that we saw in Example ?? which has the solution:

$$\begin{aligned} \lambda_n &= \frac{n^2 \pi^2}{L^2} \quad n = 0, 1, 2, \dots \\ X_n(x) &= \cos \frac{n\pi x}{L} \end{aligned}$$

if  $n > 1$  and  $\lambda_n = 0$  and  $X_n = 1$  is also a solution.

The solution to  $T' + \kappa\lambda T = 0$  can be found by letting  $T = e^{at}$  and substituting in

$$ae^{at} + \kappa\lambda e^{at} = e^{at}(a + \kappa\lambda) = 0$$

or

$$a = -\kappa\lambda$$

therefore the solution is

$$T_n = e^{-\kappa\lambda t} = e^{-(\kappa/L^2)n^2\pi^2 t}$$

A solution  $u_n(x, t)$  to the equation is

$$\begin{aligned} u_n(x, t) &= X_n(x, t)T_n(x, t) \\ &= e^{-(\kappa/L^2)n^2\pi^2 t} \cos \frac{n\pi x}{L} \end{aligned}$$

and using the principle of superposition the general solution to the PDE with given boundary conditions is:

$$u = \sum_{n=0}^{\infty} C_n e^{-(\kappa/L^2)n^2\pi^2 t} \cos \frac{n\pi x}{L}$$

Next, the initial condition is:

$$u(x, 0) = f(x) = \sum_{n=0}^{\infty} C_n \cos \frac{n\pi x}{L}$$

and the coefficients can be found by the Sturm-Liouville theorem to get:

$$\begin{aligned} C_0 &= \frac{1}{L} \int_0^L f(x) dx \\ C_n &= \frac{2}{L} \int_0^L f(x) \cos \frac{n\pi x}{L} dx \end{aligned}$$

Find the solution to the heat equation given above if the initial condition is:

$$f(x) = 1 + \cos \frac{2\pi x}{L}$$

In this case, we need to find  $C_0$  and  $C_n$ :

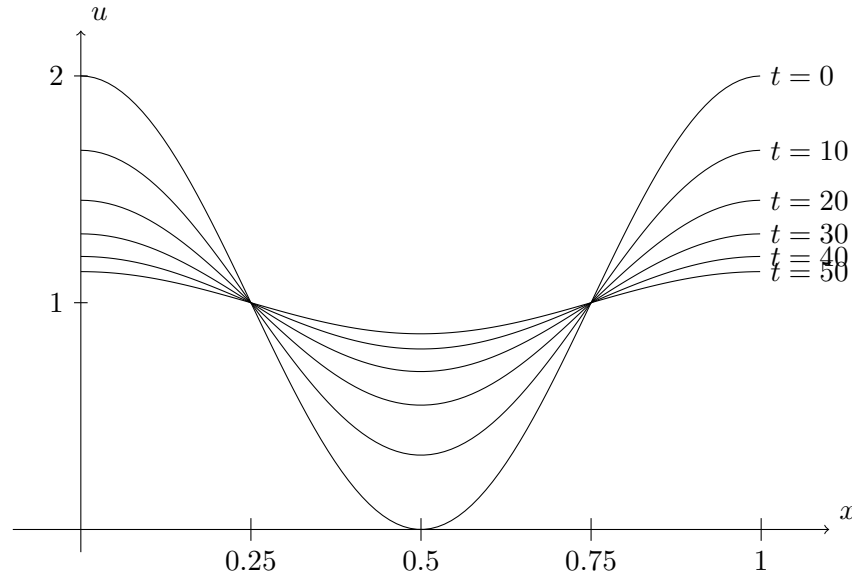
$$\begin{aligned} C_0 &= \frac{1}{L} \int_0^L (1 + \cos \frac{2\pi x}{L}) dx = 1 \\ C_n &= \frac{2}{L} \int_0^L \cos \frac{2\pi x}{L} \cos \frac{n\pi x}{L} dx \\ &= \begin{cases} 1 & \text{if } n = 2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$



so the solution to the PDE is:

$$u(x, t) = 1 + e^{-(\kappa/L^2)4\pi^2 t} \cos \frac{2\pi x}{L}$$

To get a feeling for the solution, the following is a plot when  $\kappa = 0.001$ ,  $L = 1$  for  $t = 0, 5, 10, 15$ .



The plot shows the temperature distribution for the initial case  $t = 0$  and subsequent times. The temperature evens out as time increases and in the limit the temperature would be 1 throughout, which is the average initial temperature.

## 6.7 Heat Equation in two spatial dimensions

The Heat equation in 2 spatial dimensions can be written:

$$\frac{1}{\kappa} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad (6.27)$$

In this case, let's say that we have the following boundary conditions:

$$\begin{aligned} u|_{x=0} &= 0 & \frac{\partial u}{\partial x} \Big|_{x=L} &= 0 \\ \frac{\partial u}{\partial y} \Big|_{y=0} &= 0 & \frac{\partial u}{\partial y} \Big|_{y=\ell} &= 0 \end{aligned}$$

which means that along the edge  $x = 0$ , the temperature is 0 and the other three edges are insulated. In addition, assume that the initial condition is

$$u(x, y, 0) = f(x, y)$$

In this section, we will examine how to solve this problem using the separation of variables. Since there are 3 variables, let's assume that the solution  $u$  can be written:

$$u = X(x)Y(y)T(t)$$

and substituting this into the heat equation, we get:

$$\begin{aligned}\frac{1}{\kappa} \frac{\partial XYT}{\partial t} &= \frac{\partial^2 XYT}{\partial x^2} + \frac{\partial^2 XYT}{\partial y^2} \\ \frac{1}{\kappa} XYT' &= X''YT + XY''T\end{aligned}$$

dividing through by  $XYT$

$$\frac{1}{\kappa} \frac{T'}{T} = \frac{X''}{X} + \frac{Y''}{Y}$$

Since  $T'/T$  is only a function of  $t$ ,  $X''/X$  is only a function of  $x$  and  $Y''/Y$  is only a function of  $y$ , the only option for allowing the above to be true is to assume that

$$\frac{X''}{X} = -k_1 \qquad \frac{Y''}{Y} = -k_2$$

The boundary condition can also be written in terms of  $X$  and  $Y$  as  $X(0) = 0, X'(L) = 0, Y'(0) = 0, Y'(\ell) = 0$ . Thus, in this case, we have two Sturm-Liouville problems,

$$X'' + k_1 X = 0 \qquad X(0) = 0, \quad X'(L) = 0$$

$$Y'' + k_2 Y = 0 \qquad Y'(0) = 0, \quad Y'(\ell) = 0$$

The solution to the first is

$$k_1 = \frac{(2n-1)^2 \pi^2}{4L^2} \qquad \text{for } n = 1, 2, 3, \dots$$

and

$$X = \sin \frac{(2n-1)\pi}{2L} x$$

The solution to the second is

$$k_2 = \frac{m^2 \pi^2}{\ell^2} \qquad \text{for } m = 0, 1, 2, 3, \dots$$

and

$$Y = \cos \frac{m\pi}{\ell} y$$

Next, then we need to solve

$$\frac{1}{\kappa} \frac{T'}{T} = -k_1 - k_2$$

or

$$T' + \kappa(k_1 + k_2)T = 0$$

which has the solution

$$T_{n,m} = e^{-\kappa(k_1+k_2)t}$$

Then put the solutions together:

$$\begin{aligned} u_{n,m} &= T_{n,m} X_n Y_m \\ &= e^{-\kappa(k_1+k_2)t} \sin \frac{(2n-1)\pi x}{2L} \cos \frac{m\pi y}{\ell} \end{aligned}$$

and the solution that satisfies the boundary conditions is:

$$u = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} c_{n,m} e^{-\kappa(k_1+k_2)t} \sin \frac{(2n-1)\pi x}{2L} \cos \frac{m\pi y}{\ell} dy dx$$

Finally, we apply the initial condition.

$$u(x, y, 0) = f(x, y) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} c_{n,m} \sin \frac{(2n-1)\pi x}{2L} \cos \frac{m\pi y}{\ell}$$

which results in the generalized Fourier Series:

$$c_{n,m} = \frac{4}{L\ell} \int_0^L \int_0^\ell f(x, y) \sin \frac{(2n-1)\pi x}{2L} \cos \frac{m\pi y}{\ell} dy dx$$

Find the full solution if  $L = \ell = \pi$  and

$$f(x, y) = xy(\pi - y)$$

We only need to find  $c_{m,n}$

$$\begin{aligned} c_{n,m} &= \frac{4}{\pi^2} \int_0^\pi \int_0^\pi \sin(n+1/2)x \cos my dx dy \\ &= 16(-1)^{n+1} \frac{1 + (-1)^m}{\pi(2n+1)^2 m^2} \end{aligned}$$

So the solution is

$$u = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} 16(-1)^{n+1} \frac{1 + (-1)^m}{\pi(2n+1)^2 m^2} e^{-\kappa(k_1+k_2)t} \sin(n+1/2)x \cos my$$

where  $k_1 = (n+1/2)^2$  and  $k_2 = m^2$ .

## 6.8 Bessel's equation and Bessel Functions

Bessel's equation is

$$x^2 y'' + xy' + (x^2 - \nu^2)y = 0 \quad (6.28)$$

A solution can be obtained by a power series solution and represented as

$$J_\nu(x) = \frac{x^\nu}{4^\nu} \sum_{m=0}^{\infty} \frac{(-1)^m}{4^m m! \Gamma(m + \nu + 1)} x^{2m}$$

where  $\Gamma$  is the gamma function, a generalized factorial. The function  $J_\nu(x)$  is called the **Bessel Function of the first kind**. We are often interested in solutions of (??) in which  $\nu = n$  is an integer. **If this is the case, then  $J_n(x)$  and  $J_{-n}(x)$  are two linearly independent solutions. (This statement is not true.)** The power series representation in this case is

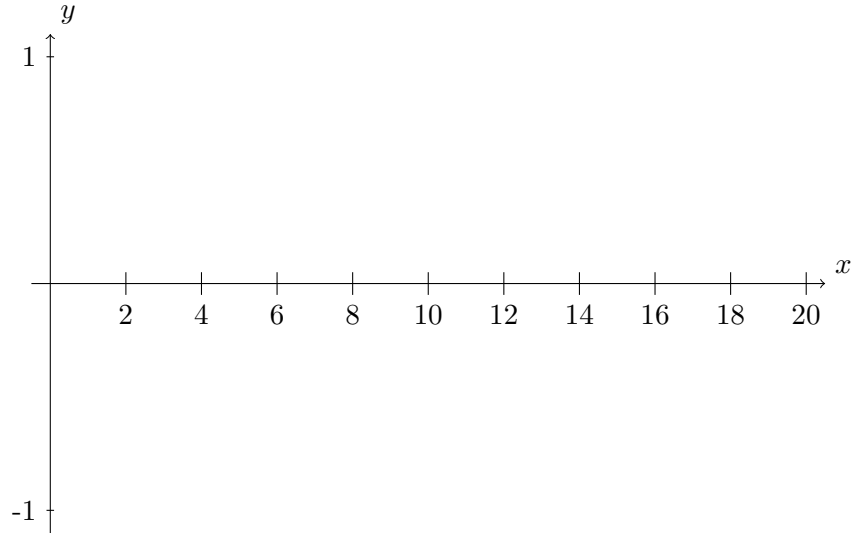
$$J_n(x) = \frac{x^n}{4^n} \sum_{m=0}^{\infty} \frac{(-1)^m}{4^m m! (m + n)!} x^{2m} \quad (6.29)$$

### 6.8.1 Properties of $J_n(x)$

The following is a plot of  $J_0(x)$  (solid line) and  $J_1(x)$  (dashed line) on  $0 \leq x \leq 10$ . Each of the Bessel functions have oscillatory behavior with decay and an infinite number of roots for  $x \geq 0$ . Also note that the roots of  $J_1$  are between the roots of  $J_0$ .

```
function besselpplot() local xmin = 0; local xmax = 20; local x = xmin; local numpoints
= 300; local y; tex.print('
draw'); for i = 0,numpoints-1 do y = bessel(0,x); tex.print('('x,',',y,') - '); x = x + (xmax-
xmin)/numpoints; end tex.print('('x,',',y,');') x=xmin tex.print('
draw[dashed]'); for i = 0,numpoints-1 do y = bessel(1,x); tex.print('('x,',',y,') - '); x = x
+ (xmax-xmin)/numpoints; end tex.print('('x,',',y,');') x=xmin tex.print('
draw[dotted]'); for i = 0,numpoints-1 do y = bessel(2,x); tex.print('('x,',',y,') - '); x = x
+ (xmax-xmin)/numpoints; end tex.print('('x,',',y,');')
```

end



Using (??), it can be shown that

$$J_n(0) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{otherwise} \end{cases}$$

In addition, using the power series representation, one can show that the other solution of (??) can be written:

$$J_{-n}(x) = (-1)^n J_n(x)$$

However for  $n < 0$ ,  $J_n(x)$  has a term  $x^{-n}$  which means that it is undefined at  $x = 0$ , which is generally why it not relevant as we will show later. Additionally, when  $n$  is an integer (again, a case we are interested in)  $J_n$  and  $J_{-n}$  are linearly independent, so a second solution is desired for this. Using techniques from differential equation a function, denoted  $Y_n(x)$

There are a number of identities that are useful for understanding Bessel functions. Two of these are shown in the follow lemma.

Consider  $n \geq 1$ , where  $n$  is an integer. Then

$$\frac{d}{dx}(x^n J_n(x)) = x^n J_{n-1}(x) \quad (6.30)$$

$$\frac{d}{dx}(x^{-n} J_n(x)) = -x^{-n} J_{n+1}(x) \quad (6.31)$$

for all  $x > 0$ .

*Proof.* First we will prove (??). Using (??), we can write

$$\begin{aligned} x^n J_n(x) &= x^n \left( \frac{x^n}{2^n} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^m m! (m+n)!} x^{2m} \right) \\ &= \sum_{m=0}^{\infty} \frac{(-1)^m}{2^{m+n} m! (m+n)!} x^{2m+2n} \end{aligned}$$

and differentiating,

$$\begin{aligned}
 \frac{d}{dx}(x^n J_n(x)) &= \sum_{m=0}^{\infty} \frac{(-1)^m (2m+2n)}{2^{m+n} m! (m+n)!} x^{2m+2n-1} \\
 &= \frac{x^{n-1}}{2^{n-1}} \sum_{m=0}^{\infty} \frac{(-1)^m 2(m+n)}{2^{m+1} m! (m+n)!} x^{2m+n} \\
 &= x^n \left( \frac{x^{n-1}}{2^{n-1}} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^m m! (m+n-1)!} x^{2m} \right) \\
 &= x^n J_{n-1}(x)
 \end{aligned}$$

The proof for (??) is very similar and is not shown.  $\square$

In addition, there are another two identities for Bessel functions that are often called recurrence relationships.

Let  $n \geq 1$  for  $n$  an integer and  $x > 0$ , then

$$x(J_{n-1} + J_{n+1}) = 2nJ_n \quad (6.32)$$

$$J_{n-1} - J_{n+1} = 2J'_n \quad (6.33)$$

*Proof.* If we use the product rule to expand (??) and (??), we get

$$\begin{aligned}
 nx^{n-1}J_n(x) + x^n J'_n(x) &= x^n J_{n-1}(x) \\
 -nx^{-n-1}J_n(x) + x^{-n} J'_n(x) &= -x^{-n} J_{n+1}(x)
 \end{aligned}$$

and multiply the first equation by  $x^{1-n}$  and the second by  $x^{1+n}$ , one gets

$$\begin{aligned}
 nJ_n + xJ'_n &= xJ_{n-1} \\
 -nJ_n + xJ'_n &= -xJ_{n+1}
 \end{aligned}$$

Adding the two above equations and dividing through by  $x$  results in (??) whereas subtracting the bottom equation from the top results in (??).  $\square$

These properties can now be used to find higher order Bessel functions, the derivatives of Bessel functions as well as the closed form of some integrals as shown in the next three examples.

Use the identities in lemmas ?? and ?? to find  $J_3$  in terms of  $J_0$  and  $J_1$ .

Let  $n = 2$  in (??) or

$$\begin{aligned}
 x(J_1 + J_3) &= 4J_2 && \text{solving for } J_3 \\
 J_3 &= \frac{4}{x}J_2 - J_1
 \end{aligned}$$

use (??) again with  $n = 1$  or  $x(J_0 + J_2) = 2J_1$  which can be written  $J_2 = (2/x)J_1 - J_0$

$$\begin{aligned} &= \frac{4}{x} \left( \frac{2}{x} J_1 - J_0 \right) - J_1 \\ J_3 &= \left( \frac{8}{x^2} - 1 \right) J_1 - \frac{2}{x} J_0 \end{aligned}$$

The above technique can be used to find  $J_n$  where  $n$  is an integer in terms of  $J_0$  and  $J_1$ , showing the importance of the first two Bessel functions.

The next example shows how to calculate the derivatives of the first two Bessel functions.

Use the identities in lemmas ?? and ?? to find  $J'_0$  and  $J'_1$  in terms of  $J_0$  and  $J_1$ .

First, differentiate (??) with  $n = 1$  to get

$$x(J'_0 + J'_2) + J_0 + J_2 = 2J'_1$$

solving for  $xJ'_0$

$$xJ'_0 = 2J'_1 - J_0 - J_2 - xJ'_2$$

using (??) with  $n = 1$  and  $n = 2$ ,

$$\begin{aligned} xJ'_0 &= (J_0 - J_2) - J_0 - J_2 - \frac{xJ_1 - xJ_3}{2} \\ &= -2J_2 - \frac{xJ_1 - xJ_3}{2} \end{aligned}$$

Using (??) with  $n = 1$

$$= \frac{-x(J_1 + J_3)}{2} - x \frac{J_1 - J_3}{2} = -xJ_1$$

and finally dividing through by  $x$

$$J'_0 = -J_1$$

Next, we'll find  $J'_1$ . Again, use (??) with  $n = 1$  to get

$$2J'_1 = J_0 - J_2$$

and use (??) with  $n = 1$  to get

$$= J_0 - \left( \frac{2}{x} J_1 - J_0 \right) = 2J_0 - \frac{2}{x} J_1$$

and dividing by 2

$$J_1' = J_0 - \frac{1}{x}J_1$$

The last example shows how to evaluate an integral.

Evaluate  $\int x^4 J_1(x) dx$ .

Integrating this by parts with  $u = x^2$  and  $du = x^2 J_1 dx$  results in

$$\int x^4 J_1 dx = \int x^2 (x^2 J_1) dx = x^2 (x^2 J_2) - \int x^2 J_2(2x) dx - 2 \int x^3 J_2 dx$$

where  $(x^2 J_2)' = x^2 J_1$  is used from (??). Next, if we again apply (??) with  $n = 3$ , to the last integral, we get

$$\int x^4 J_1 dx = x^4 J_2 - 2x^3 J_3 + C$$

## 6.8.2 Roots of the Bessel functions

As we will see in the next section that the roots of Bessel functions are important in solving Sturm-Liouville problems of whose differential equation has the form (??). There is not an analytic way to find the roots of any of the Bessel functions, so we will resort to numerical approximation. Instead of spending time determining numerical approximation to finding roots, we will only consider the use of Maple to find the roots of the Bessel Function  $J_0(x)$ . From the plot on the previous page, it appears that the first root is between 0 and 5, so we use the following Maple code to find a root on the interval  $0 \leq x \leq 5$ .

```
fsolve(BesselJ(0,x)=0,x,0,5)
```

which returns 2.404825558. In general, the  $i$ th root of  $J_0(x)$  is between  $(i-1)\pi$  and  $i\pi$  (can't find a reference for this), so the following Maple code will find the first 50.

```
 $\sigma := [\text{seq}(\text{fsolve}(\text{BesselJ}(0,x)=0,x,(i-1)\cdot\pi,i\cdot\pi))]$ 
```

If denote  $\sigma_n$  to be the  $n$ th root of  $J_0$ , then

$\sigma_1 = 2.404825558$	$\sigma_2 = 5.520078110$	$\sigma_3 = 8.653727913$
$\sigma_4 = 11.79153444$	$\sigma_5 = 14.93091771$	$\sigma_6 = 18.07106397$
$\sigma_7 = 21.21163663$	$\sigma_8 = 24.35247153$	$\sigma_9 = 27.49347913$
$\sigma_{10} = 30.63460647$	$\sigma_{11} = 33.77582021$	$\sigma_{12} = 36.91709835$

We will use this table (or actually the code to produce the roots) in the next section.



### 6.8.3 Bessel Functions of the Second Kind

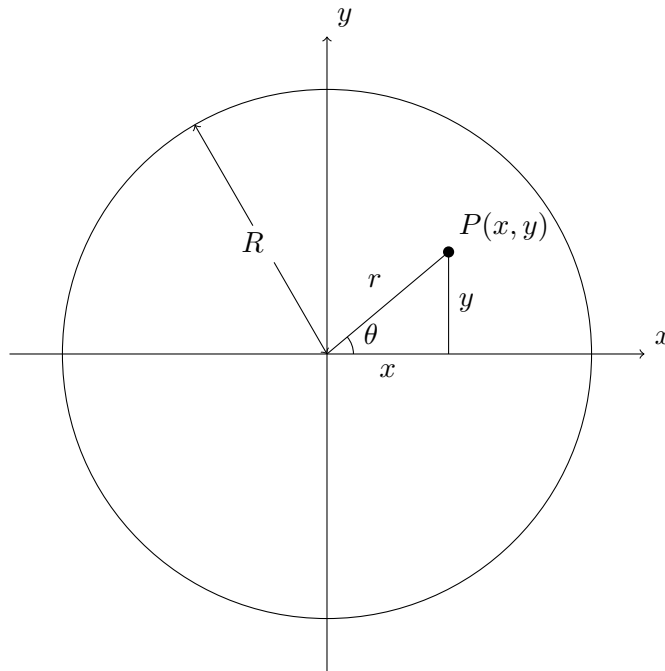
As pointed out above, the Bessel functions  $J_n(x)$  and  $J_{-n}(x)$  are linearly independent unless  $n$  is an integer and when that is true the relationship  $J_{-n}(x) = (-1)^n J_n(x)$  holds. This means that when  $n$  is an integer in the differential equation (??), that a second linearly independent solution is needed. To find such a solution there is a technique called the Frobenius method which generally gives the desired second solution. In this case, the second solution is given by

$$Y_n(x) = \frac{2}{\pi} J_n(x) \left( \ln \frac{x}{2} + \gamma \right) - \frac{1}{\pi} \sum_{k=1}^{n-1} \frac{(n-k-1)!}{k!} \left( \frac{x}{2} \right)^{2k-n} \\ + \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^{k-1} \left( \sum_{i=1}^k \frac{1}{i} + \sum_{i=1}^{k+n} \frac{1}{i} \right)}{k!(k+n)!} \left( \frac{x}{2} \right)^{2k+n}$$

and is called the **Bessel function of the second kind**. Although these functions play an important role in PDEs, we will not focus on such problems in this text. The reason for this is that the functions  $Y_n(x)$  is undefined at zero. This last fact will be used in solutions in the next few sections.

## 6.9 The Heat Equation in a Circular Region

Next, we examine how to solve the heat equation in a circular region:  $\{(x, y) | x^2 + y^2 \leq R^2\}$  as shown in the following figure.



Instead of solving the equation in cartesian coordinates, we look to write the heat equation in *polar coordinates*. A point in polar coordinates is labelled  $(r, \theta)$  where

$$r = \sqrt{x^2 + y^2} \qquad \theta = \tan^{-1} \frac{y}{x}$$

or written as  $x$  and  $y$  in terms of  $r$  and  $\theta$ ,

$$x = r \cos \theta \qquad y = r \sin \theta.$$

To convert the heat equation to polar coordinates, we need to write the right hand side of (??) or

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$$

in terms of  $r$  and  $\theta$ . This is basically an exercise in using the chain rule with multiple independent variables.

We start by finding the first partial derivatives of  $u$  with respect to  $x$  and  $y$ .

$$\begin{aligned} u_x &= \frac{\partial u}{\partial x} = \frac{\partial u}{\partial r} \frac{\partial r}{\partial x} + \frac{\partial u}{\partial \theta} \frac{\partial \theta}{\partial x} = u_r r_x + u_\theta \theta_x \\ u_y &= \frac{\partial u}{\partial y} = \frac{\partial u}{\partial r} \frac{\partial r}{\partial y} + \frac{\partial u}{\partial \theta} \frac{\partial \theta}{\partial y} = u_r r_y + u_\theta \theta_y \end{aligned}$$

where the shorthand notation for derivatives is used throughout the rest of this session. Differentiating again, we get:

$$u_{xx} = (u_r r_x)_x + (u_\theta \theta_x)_x$$

using the product rule

$$\begin{aligned} &= (u_r)_x r_x + u_r r_{xx} + (u_\theta)_x \theta_x + u_\theta \theta_{xx} \\ &= (u_{rr} r_x + u_{r\theta} \theta_x) r_x + u_r r_{xx} + (u_{\theta r} r_x + u_{\theta\theta} \theta_x) \theta_x + u_\theta \theta_{xx} \end{aligned} \tag{6.34}$$

and similarly,

$$\begin{aligned} u_{yy} &= (u_r r_y)_y + (u_\theta \theta_y)_y \\ &= (u_r)_y r_y + u_r r_{yy} + (u_\theta)_y \theta_y + u_\theta \theta_{yy} \\ &= (u_{rr} r_y + u_{r\theta} \theta_y) r_y + u_r r_{yy} + (u_{\theta r} r_y + u_{\theta\theta} \theta_y) \theta_y + u_\theta \theta_{yy} \end{aligned} \tag{6.35}$$

To complete this, we need to find  $r_x, \theta_x, r_{xx}, \theta_{xx}$

$$r_x = \frac{x}{\sqrt{x^2 + y^2}} = \frac{x}{r} \tag{6.36}$$

$$r_{xx} = \frac{\partial}{\partial x} \frac{x}{r} = \frac{r - x r_x}{r^2} = \frac{r - x(x/r)}{r^2} = \frac{r^2 - x^2}{r^3} = \frac{y^2}{r^3} \tag{6.37}$$

$$\theta_x = \frac{1}{1 + (y/x)^2} \frac{\partial}{\partial x} \left( \frac{y}{x} \right) = \frac{x^2}{x^2 + y^2} \left( -\frac{y}{x^2} \right) = -\frac{y}{r^2} \tag{6.38}$$

$$\theta_{xx} = \frac{\partial}{\partial x} \left( -\frac{y}{r^2} \right) = -y \frac{-2}{r^3} \frac{\partial r}{\partial x} = \frac{2y}{r^3} \frac{x}{r} = \frac{2xy}{r^4} \tag{6.39}$$

$$\tag{6.40}$$

as well as  $r_y, \theta_y, r_{yy}$  and  $\theta_{yy}$ .

$$r_y = \frac{y}{\sqrt{x^2 + y^2}} = \frac{y}{r} \quad (6.41)$$

$$r_{yy} = \frac{\partial}{\partial y} \left( \frac{y}{r} \right) = \frac{r - y \frac{\partial r}{\partial y}}{r^2} = \frac{r - y(y/r)}{r^2} = \frac{r^2 - y^2}{r^3} = \frac{x^2}{r^3} \quad (6.42)$$

$$\theta_y = \frac{1}{1 + (y/x)^2} \frac{\partial}{\partial y} \left( \frac{y}{x} \right) = \frac{x^2}{x^2 + y^2} \left( \frac{1}{x} \right) = \frac{x}{r^2} \quad (6.43)$$

$$\theta_{yy} = \frac{\partial}{\partial y} \left( \frac{x}{r^2} \right) = x \left( \frac{-2}{r^3} \right) \frac{\partial r}{\partial y} = \frac{-2x}{r^3} \cdot \frac{y}{r} = -\frac{2xy}{r^4} \quad (6.44)$$

Next, substitute (??)–(??) in (??) and (??) to get

$$\begin{aligned} u_{xx} &= \left( u_{rr} \frac{x}{r} + u_{r\theta} \left( -\frac{y}{r^2} \right) \right) \frac{x}{r} + u_r \frac{y^2}{r^3} \\ &\quad + \left( u_{r\theta} \frac{x}{r} + u_{\theta\theta} \left( -\frac{y}{r^2} \right) \right) \left( -\frac{y}{r^2} \right) + u_\theta \frac{2xy}{r^4}, \\ u_{yy} &= \left( u_{rr} \frac{y}{r} + u_{r\theta} \frac{x}{r^2} \right) \frac{y}{r} + u_r \frac{x^2}{r^3} + \left( u_{r\theta} \frac{y}{r} + u_{\theta\theta} \frac{x}{r^2} \right) \frac{x}{r^2} + u_\theta \frac{-2xy}{r^4}. \end{aligned}$$

Then the sum of these two is

$$\begin{aligned} u_{xx} + u_{yy} &= u_{rr} \frac{x^2 + y^2}{r^2} + u_{r\theta} \left( -\frac{xy}{r^3} + \frac{xy}{r^3} \right) + u_r \left( \frac{y^2 + x^2}{r^3} \right) \\ &\quad + u_{r\theta} \left( -\frac{xy}{r^3} + \frac{xy}{r^3} \right) + u_{\theta\theta} \left( \frac{y^2}{r^4} + \frac{x^2}{r^4} \right) + u_\theta \left( \frac{2xy}{r^4} - \frac{2xy}{r^4} \right) \\ &= u_{rr} + \frac{1}{r} u_r + \frac{1}{r^2} u_{\theta\theta} \end{aligned}$$

And then substituting this into (??) results in the heat equation in polar coordinates,

$$\frac{1}{\kappa} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} \quad (6.45)$$

Before solving this equation in general, we will examine a simpler equation that is related.

### 6.9.1 Rotationally Symmetric Solutions

The general case is a bit hard to deal with, so we first start with a *rotationally symmetric solution*, which means that there is no  $\theta$  dependence. Thus the term  $\frac{\partial^2 u}{\partial \theta^2} = 0$  and the heat equation becomes:

$$\frac{1}{\kappa} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} \quad (6.46)$$

And let's consider the solution of this with the boundary condition  $u(L, t) = 0$  and the initial condition that  $u(r, 0) = f(r)$ . In addition, we will assume that the temperature at  $r = 0$  is finite.

To solve this using separation of variables, let  $u(r, t) = T(t)R(r)$  and substitution into the PDE:

$$\begin{aligned} \frac{1}{\kappa} T' R &= T R'' + \frac{1}{r} R' T && \text{divide by } RT \\ \frac{1}{\kappa} \frac{T'}{T} &= \frac{r R'' + R'}{r R} \end{aligned}$$

As before, the only way that the left hand side can equal the right hand side is if each side only depends on a constant or

$$r R'' + R' = -\lambda r R \qquad T' + \lambda \kappa T = 0 \qquad (6.47)$$

The boundary conditions become  $R(L) = 0$  and  $R(0)$  is finite, so we get the Sturm-Liouville problem by multiplying (??) by  $r$ :

$$r^2 R'' + r R' + \lambda r^2 R = 0, \qquad R(L) = 0, \qquad R(0) \text{ is finite} \qquad (6.48)$$

which is nearly the Bessel equation with  $\nu = 0$  in (??).

Letting  $\lambda r^2 = x^2$  or  $r = x/\sqrt{\lambda}$  then we can rewrite the Sturm-Liouville problem as

$$\begin{aligned} \frac{1}{\lambda} x^2 R''(\lambda)^2 + \frac{1}{\sqrt{\lambda}} x R' \sqrt{\lambda} + x^2 R &= 0 \\ x^2 R'' + x R' + x^2 R &= 0 \end{aligned}$$

and the solution to (??) in section ?? has the solution  $R = c_1 J_0(x) + c_2 Y_0(x)$ . Returning to  $r$  as the variable, the solution of (??) is

$$R(r) = c_1 J_0(\sqrt{\lambda} r) + c_2 Y_0(\sqrt{\lambda} r).$$

If we apply the first “boundary” condition that  $R(0)$  must be finite, we take that  $c_2 = 0$  since  $Y_0(x)$  is not finite at  $x = 0$ . The next boundary condition is  $R(L) = 0$  or

$$R(L) = c_1 J_0(\sqrt{\lambda} L) = 0$$

and if  $c_1 = 0$ , this results in a trivial solution, however we know that  $J_0$  has an infinite number of roots that are called  $\sigma_n$  (see section ??), or

$$\begin{aligned} \sqrt{\lambda} L &= \sigma_n \\ \lambda &= \frac{\sigma_n^2}{L^2} \end{aligned}$$

are the eigenvalues of the problem with eigenfunctions:

$$R_n(r) = J_0\left(\frac{\sigma_n r}{L}\right)$$

Next, the solution to

$$T' + \kappa\lambda T = 0$$

is

$$T_n = e^{-\kappa\sigma_n t/L}$$

The full solution is

$$\begin{aligned} u &= \sum_{n=1}^{\infty} u_n(r, t) \\ &= \sum_{n=1}^{\infty} c_n J_0\left(\frac{\sigma_n r}{L}\right) e^{-\kappa\sigma_n t/L} \end{aligned}$$

Lastly, using the initial condition, we get:

$$u(r, 0) = f(r) = \sum_{n=1}^{\infty} c_n J_0\left(\frac{\sigma_n r}{L}\right)$$

which is a Fourier-type series with

$$c_n = \frac{\int_0^L r f(r) J_0(\sigma_n r/L) dr}{\int_0^L r J_0^2(\sigma_n r/L) dr}$$

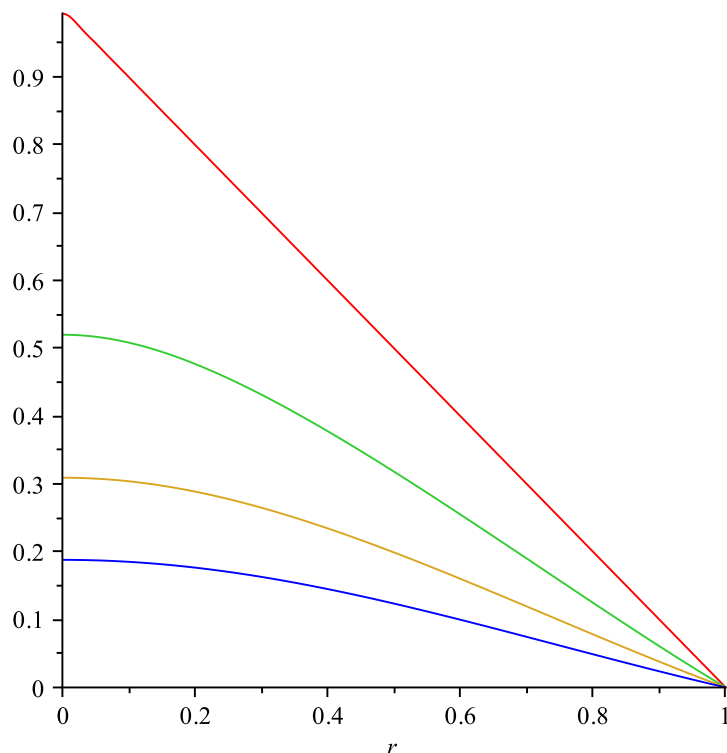
Find the solution using the initial condition  $f(r) = 1 - r$  and let  $L = 1$ . Again, we need only to find the Fourier Coefficients. The first three are:

$$\begin{aligned} c_1 &= \frac{\int_0^1 r(1-r) J_0(\sigma_1 r) dr}{\int_0^1 r J_0^2(\sigma_1 r) dr} = 0.7845194227 \\ c_2 &= \frac{\int_0^1 r(1-r) J_0(\sigma_2 r) dr}{\int_0^1 r J_0^2(\sigma_2 r) dr} = 0.06868885655 \\ c_3 &= \frac{\int_0^1 r(1-r) J_0(\sigma_3 r) dr}{\int_0^1 r J_0^2(\sigma_3 r) dr} = 0.05311413893 \end{aligned}$$

So the solution is

$$u(r, t) = \sum_{n=1}^{\infty} c_n J_0(\sigma_n r) e^{-\kappa\sigma_n t}$$

To visualize this solution, the following is a plot for  $t = 0$  (red),  $t = 5$  (green),  $t = 10$ , (amber) and  $t = 15$  (blue).



## 6.10 General Circular Heat Equation

The last section of this chapter focuses on the full heat equation in polar coordinates. Recall that in section ?? that we derived (??),

$$\frac{1}{\kappa} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2}$$

Although there are numerous boundary conditions, let's take the conditions that on the edge of the region ( $r = L$ ) that  $u_r = 0$ , that is it is insulated. When  $r = 0$ , the center of the region, the temperature  $u$  is finite. These lead to the boundary conditions

$$u_r(L, \theta, t) = 0 \quad u(0, \theta, t) \text{ is finite} \quad (6.49)$$

and then since the region is continuous, we'll assume that The boundary conditions on  $\theta$  are different from what we've seen earlier. There is no boundary, since the entire circular region is being studied. Therefore, it is usually assumed that the temperature across the line  $\theta = 0$  (which is also  $\theta = 2\pi$ ) is continuous and smooth or

$$u(r, 0, t) = u(r, 2\pi, t) \quad u_\theta(r, 0, t) = u_\theta(r, 2\pi, t) \quad (6.50)$$

that is, the temperature at  $\theta = 0$  is identical to that at  $\theta = 2\pi$  and that the derivative match there as well.

Again, as in previous chapters, we use separation of variables to solve this. We'll make the assumption that  $u = T(t)\Theta(\theta)R(r)$  and substitute this into (??) to get

$$\begin{aligned} T'\Theta R &= R''T\Theta + \frac{1}{r}R'T\Theta + \frac{1}{r^2}RT\Theta'' && \text{divide by } RT\Theta \\ \frac{T'\Theta R}{RT\Theta} &= \frac{R''T\Theta + \frac{1}{r}R'T\Theta + \frac{1}{r^2}RT\Theta''}{RT\Theta} \\ \frac{T'}{T} &= \frac{R''}{R} + \frac{1}{r}\frac{R'}{R} + \frac{1}{r^2}\frac{\Theta''}{\Theta} \end{aligned}$$

The left hand side only depends on  $t$ , therefore the right hand side must only depend on a constant. If we let

$$\frac{\Theta''}{\Theta} = -\lambda$$

or

$$\Theta'' + \lambda\Theta = 0 \tag{6.51}$$

and then the right hand side becomes:

$$\frac{R''}{R} + \frac{1}{r}\frac{R'}{R} - \frac{\lambda}{r^2} = \nu$$

multiply through by  $r^2$  and rearrange

$$r^2R'' + rR' + (r^2\nu - \lambda)R = 0, \tag{6.52}$$

which is Bessel's equation, (??). Both (??) and (??) with the boundary conditions in (??) and (??). We now study those in the next section.

### 6.10.1 Sturm-Liouville Problems from the heat equation

The first Sturm-Liouville problem from (??) is

$$\Theta'' + \lambda\Theta = 0$$

with boundary conditions  $\Theta(0) = \Theta(2\pi)$  as well as  $\Theta'(0) = \Theta'(2\pi)$ , which arise from (??).

As we've seen before in section ??, this DE only has non trivial solution when  $\lambda \geq 0$  and the solution is

$$\Theta(\theta) = c_1 \cos \sqrt{\lambda}\theta + c_2 \sin \sqrt{\lambda}\theta.$$

if  $\lambda > 0$  and  $\Theta(\theta) = 1$  if  $\lambda = 0$ . Using the boundary condition above,

$$\Theta(0) = c_1 = \Theta(2\pi) = c_1 \cos 2\pi\sqrt{\lambda} + c_2 \sin 2\pi\sqrt{\lambda}$$

which is satisfied when  $\lambda = n^2$ , for  $n = 1, 2, 3, \dots$ . The derivative of  $\Theta$  is

$$\Theta(\theta) = -nc_1 \sin n\theta + nc_2 \cos n\theta$$

The second boundary condition:

$$\Theta'(0) = nc_2 \cos(0) = \Theta'(2\pi) = -nc_1(0) + nc_2 \cos 2\pi n$$

which is satisfied. Thus

$$\Theta(\theta) = 1 \quad \Theta(\theta) = \sin n\theta \quad \text{and} \quad \Theta(\theta) = \cos n\theta$$

each satisfy the boundary condition.

The next differential equation is

$$r^2 R'' + rR' + (r^2 \nu - n^2)R = 0$$

The solution of this is

$$R(r) = c_1 J_n(\sqrt{\nu}r) + c_2 Y_n(\sqrt{\nu}r)$$

and the boundary conditions are  $R(0)$  is finite and  $R(L) = 0$ . The condition at  $r = 0$  sets  $c_2 = 0$  and the other condition:

$$0 = c_1 J_n(\sqrt{\nu}L)$$

results in

$$\sqrt{\nu}L = \sigma_{n,i}$$

where  $\sigma_{n,i}$  is the  $i$ th root of  $J_n(x)$ . Thus the eigenvalue is

$$\nu = \frac{\sigma_{n,i}^2}{L^2}$$

and the eigenfunction is

$$J_n\left(\frac{\sigma_{n,i}r}{L}\right)$$

The last DE is

$$\frac{1}{\kappa} \frac{T'}{T} = \nu = \frac{\sigma_{n,i}^2}{L^2}$$

of which the solution is:

$$T = e^{-\kappa(\sigma_{n,i}^2/L^2)t}$$



$$\begin{aligned}
u_{0,i} &= J_0\left(\frac{\sigma_{0,i}r}{L}\right)e^{-\kappa(\sigma_{0,i}^2/L^2)t} \\
u_{n,i} &= R_{n,i}(r)\Theta_n T_{n,i} \\
&= J_n\left(\frac{\sigma_{n,i}r}{L}\right)(a_n \cos n\theta + b_n \sin \theta)e^{-\kappa(\sigma_{n,i}^2/L^2)t}
\end{aligned}$$

and using the principle of superposition the full solution is:

$$\begin{aligned}
u(r, \theta, t) &= \sum_{i=1}^{\infty} a_0 J_0\left(\frac{\sigma_{0,i}r}{L}\right)e^{-\kappa(\sigma_{0,i}^2/L^2)t} + \\
&\quad \sum_{i=1}^{\infty} \sum_{n=1}^{\infty} J_n\left(\frac{\sigma_{n,i}r}{L}\right)(a_n \cos n\theta + b_n \sin \theta)e^{-\kappa(\sigma_{n,i}^2/L^2)t}
\end{aligned}$$

Finding the coefficients. In this case we use the initial condition that

$$u(r, \theta, 0) = f(r, \theta)$$

and substituting into the solution:

$$f(r, \theta) = \sum_{i=1}^{\infty} a_0 J_0\left(\frac{\sigma_{0,i}r}{L}\right) + \sum_{i=1}^{\infty} \sum_{n=1}^{\infty} J_n\left(\frac{\sigma_{n,i}r}{L}\right)(a_n \cos n\theta + b_n \sin \theta)$$

The coefficients are:

$$\begin{aligned}
a_{0,i} &= \frac{\int_0^L \int_0^{2\pi} r f(r, \theta) J_0(\sigma_{0,i}r/L) d\theta dr}{2\pi \int_0^L r J_0(\sigma_{0,i}r/L)^2 dr} \\
a_{n,i} &= \frac{\int_0^L \int_0^{2\pi} r f(r, \theta) J_n(\sigma_{n,i}r/L) \cos n\theta d\theta dr}{\pi \int_0^L r J_n(\sigma_{n,i}r/L)^2 dr} \\
b_{n,i} &= \frac{\int_0^L \int_0^{2\pi} r f(r, \theta) J_n(\sigma_{n,i}r/L) \sin n\theta d\theta dr}{\pi \int_0^L r J_n(\sigma_{n,i}r/L)^2 dr}
\end{aligned}$$

Solve the equation above when  $f(r, \theta) = J_1(\sigma_{1,1}r) \sin \theta$ . Use  $\kappa = 0.04$  and  $L = 1$ .

Again, we just need to compute the coefficients above. Use can either use Maple or make a symmetry argument to see that

$$\begin{aligned}
a_{0,i} &= 0 \\
a_{n,i} &= 0 \\
b_{1,1} &= 1 \\
a_{n,i} &= 0 \quad n \neq 1, i \neq 1.
\end{aligned}$$

So the solution is

$$u(r, \theta, t) = J_1(\sigma_{1,1}r) \sin \theta e^{-0.04\sigma_{1,1}^2 t}$$



## Chapter 7

# Differential Equations on $\mathbb{R}$ and the Fourier Transform

### 7.1 The Fourier Transform

From above, we can write the Fourier Series of  $f(x)$  on  $[-L/2, L/2]$  as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_0 x}$$

where

$$c_n = \frac{1}{L} \int_{-L/2}^{L/2} f(x) e^{-in\omega_0 x} dx$$

where  $\omega_0 = 2\pi/L$  and  $\omega_n = n\omega_0$  and let

$$\Delta\omega = \omega_{n+1} - \omega_n = \frac{2\pi}{L}$$

The series can be written:

$$f(x) = \sum_{n=-\infty}^{\infty} \left[ \frac{1}{2\pi} \int_{-L/2}^{L/2} f(x) e^{-i\omega_n x} dx \right] e^{i\omega_n x} \Delta\omega$$

and letting  $L \rightarrow \infty$  or  $\Delta\omega \rightarrow 0$ , we get:

$$f(x) = \int_{-\infty}^{\infty} \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx \right] e^{i\omega x} d\omega$$

which can be written as follows:

The **Fourier Transform** of  $f$  is :

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx$$

and using this, the function  $f(x)$  can be written:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(x) e^{i\omega x} d\omega$$

Note: not all functions have a Fourier Transform. If the integral is defined (and converges), then the transform exists.

(discussion of function space versus frequency space).

Find the Fourier Transform of

$$f(x) = \begin{cases} e^{-ax} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \hat{f}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-ax} e^{-i\omega x} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-(a+i\omega)x} dx \\ &= \frac{1}{\sqrt{2\pi}} \frac{-1}{a+i\omega} e^{-(a+i\omega)x} \Big|_0^{\infty} = \frac{1}{\sqrt{2\pi}} \frac{1}{a+i\omega} = \frac{1}{\sqrt{2\pi}} \frac{a-i\omega}{a^2+\omega^2} \end{aligned}$$

### 7.1.1 Even and Odd Functions and Fourier Transforms

if  $f(x)$  is an even function then the Fourier Transform can be written:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(x) e^{i\omega x} d\omega$$

where

$$\hat{f}(x) = \frac{1}{2\pi}$$

Fourier Transform of the function

$$f(x) = \begin{cases} 1 & |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The Fourier Transform is:

$$\begin{aligned}
 \hat{f}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-1}^1 e^{-i\omega x} dx \\
 &= \frac{1}{\sqrt{2\pi}} \left. \frac{1}{-i\omega} e^{-i\omega x} \right|_{-1}^1 \\
 &= \frac{1}{\sqrt{2\pi}} \left( \frac{e^{-i\omega}}{-i\omega} + \frac{e^{i\omega}}{i\omega} \right) \\
 &= \frac{1}{\sqrt{2\pi}} \frac{2 \sin \omega}{\omega}
 \end{aligned}$$

Also,

$$\begin{aligned}
 f(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega x} d\omega \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2 \sin \omega}{\omega} e^{i\omega x} d\omega
 \end{aligned}$$

Since  $f(0) = 1$  (from the original definition of the function), this means that

$$1 = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin \omega}{\omega} d\omega$$

or

$$\pi = \int_{-\infty}^{\infty} \frac{\sin \omega}{\omega} d\omega$$

since  $\frac{\sin \omega}{\omega}$  is an even function we can write:

$$\frac{\pi}{2} = \int_0^{\infty} \frac{\sin \omega}{\omega} d\omega$$

An important function is that on the right side

$$\text{Si}(u) = \int_0^u \frac{\sin \omega}{\omega} d\omega$$

### 7.1.2 Physical Interpretation of the Fourier Transform

The Fourier Transform of a function  $f(x)$  is a function that lists all of its frequencies, often called the **spectral representation**. The name come from understanding light in terms of its frequencies (colors) and many analyses in physics and chemistry are easier to do in term of the spectrum

The **total energy** is

$$\int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 d\omega$$

### 7.1.3 Properties of the Fourier Transform

[Linearity] The Fourier Transform is *linear*; that is, for any pair of functions  $f(x)$  and  $g(x)$  whose Fourier transforms exist and any constants  $a$  and  $b$ , then

$$\mathcal{F}(af + bg) = a\mathcal{F}(f) + b\mathcal{F}(g)$$

*Proof.*

$$\begin{aligned}\mathcal{F}(af + bg) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (af(x) + bg(x)) dx \\ &= a \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) dx + b \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) dx \\ &= a\mathcal{F}(f) + b\mathcal{F}(g)\end{aligned}$$

□

[Derivatives] Let  $f(x)$  be a continuous on the  $x$ -axis and  $f(x) \rightarrow 0$  as  $|x| \rightarrow \infty$ . Furthermore, let  $f'(x)$  be absolutely integrable on the  $x$ -axis. Then

$$\mathcal{F}\{f'(x)\} = i\omega\mathcal{F}\{f(x)\}$$

*Proof.*

$$\mathcal{F}\{f'(x)\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(x) e^{-i\omega x} dx$$

integrate by parts

$$= \frac{1}{\sqrt{2\pi}} \left( f(x) e^{-i\omega x} \Big|_{-\infty}^{\infty} - (-i\omega) \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx \right)$$

since  $f(x) \rightarrow 0$  as  $|x| \rightarrow \infty$

$$= \frac{1}{\sqrt{2\pi}} i\omega \mathcal{F}\{f(x)\}$$

□

## 7.2 Solving the Heat Equation using Fourier Transform

Now we solve:

$$\frac{1}{\kappa} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

if  $-\infty < x < \infty$  and  $u(x, 0) = f(x)$ .

Let  $\hat{u}(\omega, t)$  be the Fourier Transform of  $u(x, t)$ . Taking the Fourier Transform of the PDE above:

$$\frac{1}{\kappa} \frac{\partial \hat{u}}{\partial t} = \frac{\partial^2 \hat{u}}{\partial x^2}$$

and using the derivative of the Fourier Transform:

$$\frac{1}{\kappa} \frac{\partial \hat{u}}{\partial t} = (i\omega)^2 \hat{u} = -\omega^2 \hat{u}$$

and solving this we get:

$$\hat{u}(\omega, t) = A(\omega) e^{-\kappa \omega^2 t}$$

If we take the Fourier Transform of the initial condition:

$$\hat{u}(\omega, 0) = \mathcal{F}\{f(x)\} = \hat{f}(\omega)$$

and then evaluate the transformed solution to get:

$$\hat{u}(\omega, 0) = A(\omega)$$

so

$$\hat{u}(\omega, t) = \hat{f}(\omega) e^{-\kappa \omega^2 t}$$

and lastly, we take the inverse transform:

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{-\kappa \omega^2 t} e^{i\omega x} d\omega$$

Solve the Heat equation on the real line if the initial condition is

$$f(x) = \begin{cases} 1 & |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

From above

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{\sin \omega}{\omega}$$

so the solution is:

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{\sin \omega}{\omega} e^{-\kappa \omega^2 t + i\omega x} d\omega$$