Chapter 1

What Happened to my Algebra and why is it Linear?

This class will likely be very different from the majority of math classes you have taken in your academic career. In your previous classes, there was likely an emphasis on computations and less so on theory/concepts; in this class we will put equal emphasis on computations and theory/concepts. That is to say: you will be expected to ably compute examples while, at the same time, demonstrate a working knowledge of the theory. I do not mention this to alarm or frighten you. Rather, it is important for you to know that this class may diverge from your usual notion of a math class, and that **can** you persevere in the face of difficulty. With that being said, you might be wondering:

Question: Why Take this class?

Answer: The cop-out answer is that this depends on what you want to use linear algebra for (I suggest you think about this as the class progresses), but as this may be your first introduction to the subject, it might be hard to answer this right off the bat. In light of this, I will do my best to provide an answer that will encompass as many backgrounds as possible. You should study linear algebra because it is useful in many different areas of study, and, most importantly, you will strengthen your problem solving abilities and being comfortable with and overcoming mathematical/intellectual hurdles. I am here to help you when things get challenging; as such I encourage you to unitize office hours. In addition, I encourage you to collaborate with and befriend your classmates. **I want this to be an enjoyable and worthwhile endeavor for each of you!**

With all of that said, lets answer the question posed by the title of the chapter: What is Linear Algebra?

There are many correct answers to what is linear algebra, each of which depends strongly on what you will use linear algebra for; for now, we will keep in mind a down-to-earth answer. However, I strongly encourage you to come back periodically and think about what linear algebra means to you as you learn more about this beautiful subject.

Answer: Linear algebra is the study of n-dimensional spaces, the functions between them, and how they fit inside each other. By an n-dimensional space, I mean a 0-dimensional space is a point, a 1-dimensional space is a line, a 2-dimensional is the plane, and so on.

Now, you might be wondering:

Question: How do I succeed in this class.

Answer: To succeed in this class, you should, at minimum, do the following:

1. Study the online lecture notes and/or textbook before and after each class.
2. Come to class with questions.
3. Start homework early! It will be challenging.
4. Come to office hours. **You do not need to have questions to come to office hours; it can be a space for you to do homework with others.**
5. Talk and study with your peers.
6. Don't give up when things get hard (I can't stress this enough)!
7. Most importantly, have fun!

With all of that said, lets embark on our journey into the fantastic world of linear algebra!

Chapter 2

Systems of Linear Equations and Matrices

2.1 Systems of Linear Equations

In the last chapter, we said that linear algebra is the study of n-dimensional spaces, the functions between them, and how they fit inside of each other. Our first step toward making sense of this is through systems of linear equations. First, a couple of definitions, and then we will look at why systems of linear equations are the right thing to look at if we want to study n-dimensional spaces.

Linear Equation

Definition 2.1. A linear equation in the variables $x_1, \ldots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b,$$

where all a_i and b are complex numbers.

Example 2.1. The following are examples of linear equations:

1. $x_1 + 2x_2 = \sqrt{2}$
2. $x_2 - 8x_3 - 2x_5 = 3x_1 + 1$
3. $21x_2 + 5x_4 = 0$

Example 2.2. The following **are not** examples of linear equations:

1. $x_1^2 - x_2 = 5$
2. $e^{x_1 - x_2} + x_3 = 2$
3. $x_1x_4 - x_2x_3 = 0$
4. $\frac{1}{x_1} = 2$
5. $\sin(x_1^2 - x_2^2) = \sqrt{2}$
6. $\sqrt{x_1} + 2x_2 = \sqrt{7}$

Linear equations are very good models of n-dimensional spaces. Indeed, the graph of a linear equation in 2 variables is a line (or 1-dimensional space), the graph of a linear equation in 3 variables is a plane (or 2-dimensional space), and so on. We will want to see how these equations (spaces) interact when they are in the same ambient space; that is: we want to figure out when they intersect. This leads us to the following definition.

System of Linear Equations

Definition 2.2. A system of linear equations is a collection of one or more linear equations involving the same variables.

Example 2.3. The following are examples of systems of linear equations:

1. $\begin{cases} 4x_1 + x_2 = 7 \\ 2x_1 - 2x_2 = \sqrt{3} \\ x_2 + \sqrt{14}x_2 = 0 \end{cases}$

2. $\begin{cases} 2x_1 + x_3 = 0 \\ 2x_1 - 2x_2 = \sqrt{3} \\ x_1 - x_4 = 0 \end{cases}$

Solution

Definition 2.3. A solution to a system of linear equations is a tuple $(s_1, \dots, s_n)$ of real numbers that make each linear equation in the system a true statement.

Example 2.4. $(3, 2)$ is a solution for the following system of linear equations, but $(0, 1)$ is not a solution.

$$\begin{cases} x_1 - 2x_2 = -1 \\ -x_1 + 3x_2 = 3 \end{cases}$$

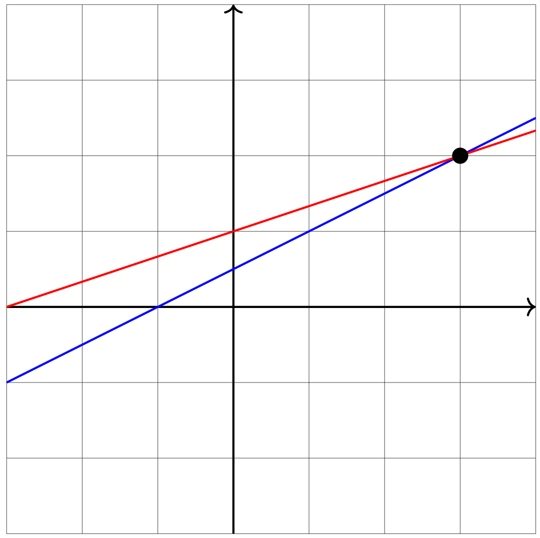
Solution Set

Definition 2.4. The set of all possible solutions to a system of linear equations is called the solution set. Two linear systems are said to be equivalent if they have the same solution sets.

Example 2.5. 1. The solution set to the system

$$\begin{cases} x_1 - 2x_2 = -1 \\ -x_1 + 3x_2 = 3 \end{cases}$$

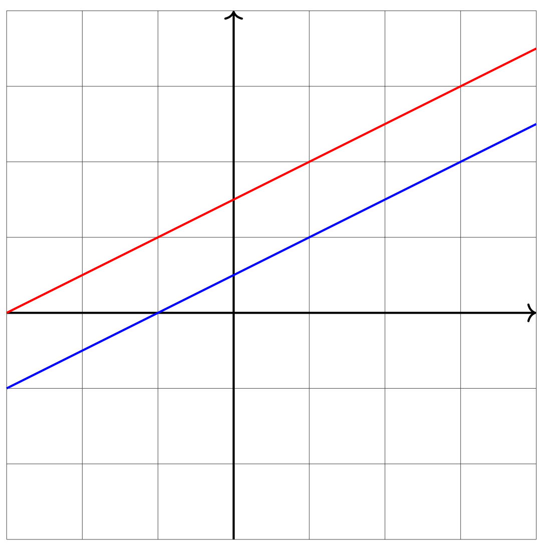
is the set $\{(3,2)\}$. Here's a geometric picture of what is going on:



2. The solution set to the system

$$\begin{cases} x_1 - 2x_2 = -1 \\ -x_1 + 2x_2 = 3 \end{cases}$$

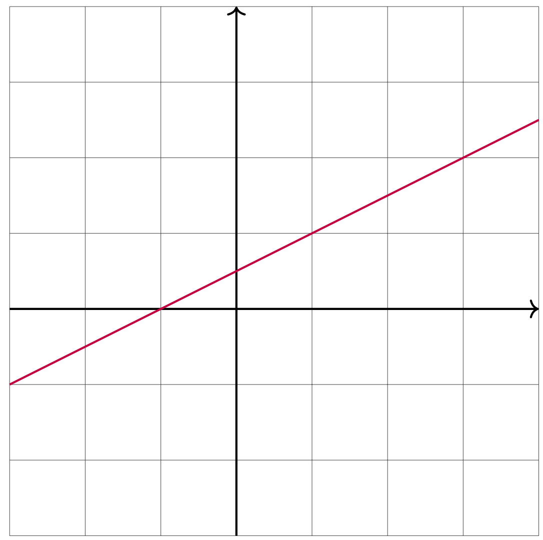
is the empty set, meaning there is no solution. Here is a geometric picture of what is going on:



3. The solution set to the system

$$\begin{cases} x_1 - 2x_2 = -1 \\ 2x_1 - 4x_2 = -2 \end{cases}$$

is the line carved out by $x_1 - 2x_2 = -1$. Here is a geometric picture of what is going on:



These three examples illustrate the following fact:

Theorem 2.1. The solution set to any system of linear equations in any number of variables is either

1. The empty set (i.e no solution).
2. One and only one point.
3. Infinitely many points.

Proof. Postponed for now. □

Consistent and Inconsistent Systems

Definition 2.5. If a system of linear equations has at least one solution, it is called indexconsistent consistent. If it has no solutions, it is called inconsistent.

Example 2.6. Using an online graphing convince yourself that Theorem 2.1 is true for systems of linear equations of three variables.

2.1.1 Matrix Notation

It is very convenient to encode information about a system of linear equations into a rectangular array called a **matrix**. There are two ways to do this, which we will demonstrate through an example.

Example 2.7. Consider the following system of linear equations

$$\begin{cases} 3x_1 + x_2 + 2x_3 = 0 \\ x_1 + x_3 = 2 \\ -2x_1 + 2x_2 = 7 \end{cases},$$

The **coefficient matrix** of the system of linear equations is made by arranging the coefficients of the system into the following matrix

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

We will find it useful to define the **augmented matrix** of a system of linear equations by also including the values on the right hand side of the equal signs:

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

Size of a Matrix

Definition 2.6. A matrix is an $m \times n$ matrix if it has m rows (horizontal) and n columns (vertical). If $m = n$, we say that the matrix is a square matrix of size m.

So, why should we look at matrices? It turns out that the augmented matrix of a system of linear equations is sufficient information to find the solution set. We will dive more into this soon. For now, we will talk about *elementary row operations* as these will very useful tools in our venture to find solution sets for any size system of linear equations!

Elementary Row Operations

Definition 2.7. The *elementary row operations* are the following:

1. Replacement: replace one row by the sum of itself and a multiple of another row.
2. Interchange: Switch the position of two rows.
3. : Multiply all entries in a row by a **nonzero** constant.

Row Equivalent Matrices

Definition 2.8. The *elementary row operations* are the following:
We say that two matrices are row equivalent if one can use elementary row operations to from one matrix to the other.

Here is why we consider these operations:

Theorem 2.2. Suppose a matrix, which we will call A, is the augmented matrix for a system of linear equations. If B is a matrix that is row equivalent to A (i.e we can use row operations to go from A to B), then the system of linear equations represented by B has the same solution set as the original system.

Proof. The proof is outlined as a bonus question on Homework 1. $\square$

This is incredibly powerful! We can simplify an augmented matrix in such a way that we can determine whether or not a system is consistent, as we shall now see:

Example 2.8. Consider the system of linear equations

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_2 - 8x_3 = 8 \\ 5x_1 - 5x_3 = 10 \end{cases}.$$

The augmented matrix of this system is

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}.$$

Since elementary row operations do not change the solution set to a system of linear equations, lets simplify the above matrix to something more tame.

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix} \xrightarrow{-5R_1+R_3 \mapsto R_3} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{bmatrix} \xrightarrow{\frac{1}{10}R_3 \mapsto R_3} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 1 & -1 & 1 \end{bmatrix}$$
$$\xrightarrow{-2R_3+R_2 \mapsto R_2} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 0 & -6 & 6 \\ 0 & 1 & -1 & 1 \end{bmatrix} \xrightarrow{\frac{1}{6}R_2 \mapsto R_2} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 1 & -1 & 1 \end{bmatrix}$$

We may stop here (although we can go further, which we will do next time). From this new matrix, we can pick out a solution to the original system of linear equations. Indeed, row 2 tells us that $x_3 = -1$. Using this, row 3 tells us that $x_2 = 0$. Finally, row 1 tells us that $x_1 = 1$. Therefore, the original system is consistent with $(1, 0, -1)$ being a solution.

Helpful Tip: It is helpful to think of matrices and systems of linear equations as the same thing under different guises. What we mean by this is that if you see a matrix you should think about it as being the augmented matrix of some system of equations, and, on the other hand, when you see a system of linear equations, you should think of its augmented matrix. **As with many aspects of mathematics, different points of view of the same thing is a power that cannot be overestimated.**

2.2 More Row Reduction and the Echelon Forms

In the last section we introduced elementary row operations and saw how they can aid us in solving a system of linear equations. It seems reasonable, at this point, to ask

Question: Can we find the solution set to any system of linear equations by writing down its augmented matrix and performing row operations?

Answer: Yes!

Not only can we find the solution set to any system of linear equations using row operations, but there are *always* two forms we can row reduce the augmented matrix into that yields valuable information about our system of linear equations called **Row Echelon Form (REF)** and **Reduced Row Echelon Form (RREF)**. These forms and the information they posses will be the content of this section, and

The Echelon Forms

Definition 2.9.

1. A matrix is said to be in row Echelon form (REF), or simply Echelon form, if it satisfies the following properties:
   (a) All nonzero rows are above any rows consisting of all zeros.
   (b) Each leading entry **(i.e the left most nonzero entry)** of a row is in a column tot he right of the leading entry of the row above it.
   (c) All entries below a leading term are 0.
2. A matrix is said to be in row reduced Echelon form (RREF) if it satisfies the following conditions:
   (a) It satisfies all properties of being in REF.
   (b) The leading entry of each row is 1.
   (c) Each leading 1 is the only nonzero entry in its column.

Remark 1. *If a matrix is in RREF, then it is in REF. However, if a matrix is in REF, then it may not be true that it is in RREF, as we shall see in examples 2.9 and 2.10.*

Example 2.9. The following matrix is in REF but not in RREF:

$$\begin{bmatrix} 2 & 0 & 7 & 9 & 1 & 10 \\ 0 & 0 & 3 & 0 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

The following matrix is not in REF:

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

Example 2.10. The following matrix is in RREF:

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

The following matrix is in REF but not RREF:

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

Theorem 2.3. Given any matrix A, there exists a series of row operations that put a matrix into a REF. Similarly, there is a series of row operations that put a matrix into a RREF. In other words, A is row equivalent to REF and RREF matrices.

Proof. We will postpone this for now. Later, we will describe an algorithm that shows this. ☐

Uniqueness of RREF

Theorem 2.4. Given any matrix A, the RREF of A is **unique**. That is to say there is only one RREF we can row reduce A to.

Proof. We won't prove this; however, I encourage you to try! ☐

The following is a homework exercise, but as it is important, we will state it here.

Exercise 2.11. *A matrix A has a unique RREF by Theorem 2.2. However, a matrix A can have many different REF's. Construct an example of a matrix with multiple REF's.*

Before we discuss an algorithm to row reduce a matrix to a REF or RREF, we will find it helpful to define a couple of terms. But first, a related meme:

Pivots

Definition 2.10. A **pivot position** in a matrix A (not necessarily in RREF) is a location in A that corresponds to a leading 1 in the RREF of A. A pivot column of A is a column of A that contains a **pivot column**.

This is, perhaps, a strange definition; so, lets do an example and identify the pivot positions and pivot columns of a matrix.

Example 2.12. We will find all of the pivot positions and pivot columns of the following matrix

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

Lets perform some row operations to put A into RREF:

$$\begin{bmatrix} 3 & 0 & 2 & 1 \\ 2 & 2 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 3 & 0 & 2 & 1 \end{bmatrix} \xrightarrow[-3R_1+R_3 \mapsto R_3]{-2R_1+R_2 \mapsto R_2} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 3 & 2 & 1 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{4}R_2 \mapsto R_2} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 3 & 2 & 1 \end{bmatrix} \xrightarrow[R_2+R_1 \mapsto R_1]{-3R_2+R_3 \mapsto R_3} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{2}R_3 \mapsto R_3} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{2} \end{bmatrix}$$

This is the RREF of A. Below are circled the leading ones, which correspond to the pivot positions of A. Also, highlighted, are the pivot columns of A.

As the above example shows **not all columns of a matrix will be pivot columns!**.

2.2.1 The Row Reduction Algorithm (Gaussian Elimination)

We will describe the algorithm of Gaussian Elimination, which yields a way to put a matrix into REF or RREF. Given a matrix A, we proceed in the following steps:

Gaussian Elimination

Algorithm 2.1.

1. Start with the left most nonzero column. Select a nonzero entry in the column to be a pivot. If necessary, interchange two rows so the pivot is at the top of the matrix.
2. Use row operations to get zeros in all positions below the pivot you found in step 1.
3. Cover the first row of the new matrix and apply 1-2 to the new matrix with the first row deleted. Keep doing this until you cannot.
4. Starting form the right most column, create zeros above each pivot.

Let's return to systems of linear equations! Given an augmented matrix, we may put it into REF or RREF, which yields information about the solution set of the system of linear equations. Before we see an example of this, we will take a brief detour through the world of solution sets. As we said in the previous section, there are systems of equations with infinitely many solutions; for these systems it would be very inefficient (and impossible) to enumerate all solutions by hand. To remedy this, we can describe all solutions using parameters (or free variables). To this end, we will find the next definition to be useful.

Free and Basic Variables

Definition 2.11. Suppose a matrix is A is the augmented matrix of a system of linear equations. The variables corresponding to pivot columns are called basic variables and the other variables are called free variables.

Example 2.13. We will find the solution set of the following system of linear equations by putting its augmented matrix into RREF and identifying pivot positions/columns:

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 0 \\ 4x_1 + 4x_3 + x_4 = 8 \\ 2x_1 + 6x_3 + 8x_4 = 4 \end{cases}.$$

Let's apply Gaussian elimination to put the augmented matrix into RREF:

$$\begin{bmatrix} 1 & 2 & 3 & 0 & 0 \\ 4 & 0 & 4 & 1 & 8 \\ 2 & 0 & 6 & 8 & 4 \end{bmatrix} \xrightarrow[-2R_1+R_3 \mapsto R_3]{-4R_1+R_2 \mapsto R_2} \begin{bmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & -8 & -8 & 1 & 8 \\ 0 & -4 & 0 & 8 & 4 \end{bmatrix} \xrightarrow{\frac{-1}{4}R_3 \mapsto R_3} \begin{bmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & -8 & -8 & 1 & 8 \\ 0 & 1 & 0 & -2 & -1 \end{bmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & -8 & -8 & 1 & 8 \end{bmatrix} \xrightarrow{8R_2+R_3 \mapsto R_3} \begin{bmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & 0 & -8 & -15 & 0 \end{bmatrix}$$

$$\xrightarrow{\frac{-1}{8}R_3 \mapsto R_3} \begin{bmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & 0 & 1 & \frac{15}{8} & 0 \end{bmatrix} \xrightarrow{-2R_3+R_1 \mapsto R_1} \begin{bmatrix} 1 & 2 & 0 & \frac{-15}{4} & 0 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & 0 & 1 & \frac{15}{8} & 0 \end{bmatrix}$$

$$\xrightarrow{-2R_2+R_1 \mapsto R_1} \begin{bmatrix} 1 & 0 & 0 & \frac{1}{4} & 2 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & 0 & 1 & \frac{15}{8} & 0 \end{bmatrix}$$

We then see that x_1, x_2, and x_3 are basic variables for our system, and x_4 is a free variable. From the RREF we found, we find that $x_3 = -\frac{15}{8}x_4$, $x_2 = 2x_2 - 1$, and $x_1 = -\frac{1}{4}x_4 + 2$. A sufficient way to write this is to say the following is a **parametric description** of the general solution:

$$\begin{cases} x_1 - \frac{1}{4}x_4 + 2 \\ x_2 = 2x_2 - 1 \\ x_3 = -\frac{15}{8}x_4 \\ x_4 \text{ is free} \end{cases}$$

Note: sometimes people replace free variables with letters like t or s; we won't do that here, but we might in the future. Here, we see that x_4 can be any number, and it will determine what x_1, x_2, and x_3 need to be to give a solution to the system (this is why we say that x_4 is a free variable). For example, set $x_4 = 0$. Then, $x_1 = 2$, $x_2 = -1$ and $x_3 = 0$. Thus, $(2, -1, 0, 0)$ is a solution to our original solution..

The number of free variables of a system is the dimension of the solution set. To make sense of this, we need to agree on a notion of dimension, which will come up later.

We will end this section with a beautiful theorem that determines whether or not a system is consistent just by analyzing the REF of a matrix!

Theorem 2.5. A system of linear equations is consistent if and only if the right most olumn of the augmented matrix is not a pivot column. In other words, the REF of the augmented matrix has *no row* of the form

$$\begin{bmatrix} 0 & \dots & 0 & b \end{bmatrix}$$

where b is any *nonzero* number.

2.3 Vectors in $\mathbb{R}^n$

In this section we will introduce the notion of vectors; however, for now, we will only focus on vectors in $\mathbb{R}^n$. As we will see later, there is a more abstract notion of vectors. We will find that vectors offer us a very convenient language to describe what is going on with systems of linear equations and the spaces that they carve out!

$\mathbb{R}^n$

Definition 2.12. The set $\mathbb{R}^n$ is define to be the set of of n-tuples of real numbers.

Example 2.14. Lately, we have been working with $\mathbb{R}^2$, which we visualize as the Cartesian plane. We've even thought about $\mathbb{R}^3$ a bit, which we visualize as 3 dimensional space. In general we may think of $\mathbb{R}^n$ as n-dimensional space. Though, I must caution you that later on n-dimensional space will encompass more than just $\mathbb{R}^n$; however, for now, we will abuse this terminology!

Vectors

Definition 2.13. A vector (or column vector) in $\mathbb{R}^n$ is a $n \times 1$ matrix of real numbers. **Caution:** later on vector will mean something that will encompass more than the elements of $\mathbb{R}^n$.

Example 2.15. The following is a vector in

$$\begin{bmatrix} 1 \\ 0 \\ 2 \\ \sqrt{6} \end{bmatrix}$$

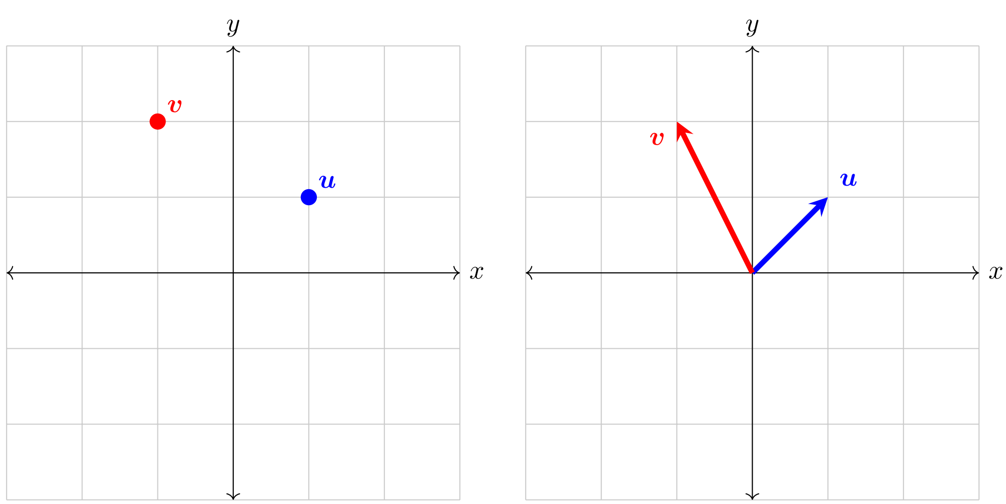
in $\mathbb{R}^4$

As we can see in the last example, sometimes it takes up a lot of space to write a vector as an $n \times 1$ matrix, so we will sometimes write a vector as an n-tuple. Precisely, we shall take

$$\begin{bmatrix} a_1 \\ \ddots \\ a_n \end{bmatrix} = (a_1, \ldots, a_n)$$

Caution: $(a_1, \ldots, a_n) \neq \begin{bmatrix} a_1 & \ldots & a_n \end{bmatrix}$. The latter is an $1 \times n$ matrix.

A vector $\mathbf{a} = (a_1, \ldots, a_n)$ can be visualized in two ways: 1) They can be thought of as points $(a_1, \ldots, a_n)$ in $\mathbb{R}^n$; 2) We can also think of the vector $\mathbf{a}$ as an arrow from the origin of $\mathbb{R}^n$ to the point $(a_1, \ldots, a_n)$. We will often take the view of 2) when thinking about vectors. Below are two figures that illustrate these two viewpoints.



Visualizing vectors as arrows is extremely useful, as we shall see shortly.

2.3.1 Operations with Vectors

While vectors **are not numbers** (rather they are an ordered list of numbers), we can still perform some operations with them, like addition, subtraction, and scaling. Though we must caution ourselves!

Vector Addition and Scalar Multiplication

Definition 2.14. We add two vectors $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$ component wise:

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, \dots, a_n + b_n) = \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix}.$$

Suppose that $c \in \mathbb{R}$ (c is a real number). Then, $\mathbf{a}$ scaled by c is

$$c\mathbf{a} = (ca_1, \dots, ca_n) = \begin{bmatrix} ca_1 \\ \vdots \\ ca_n \end{bmatrix}.$$

Example 2.16. Let $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$. We compute the following:

1. $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$

2. $-\mathbf{v} = (-1) \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -3 \end{bmatrix}$

3. $\mathbf{u} - \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$

Exercise 2.17. *Draw the vectors* $\boldsymbol{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $2\boldsymbol{u}$, $-\boldsymbol{u}$, *and* $-2\boldsymbol{u}$. *Geometrically what is the relationship between* $-\boldsymbol{u}$ *and* $\boldsymbol{u}$*?*

Zero Vector

Definition 2.15. We call the vector in $\mathbb{R}^n$ consisting of all zeros the zero vector, and denote it by $\mathbf{0}$. Explicitly,

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Proposition 2.1. Let $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ be vectors in $\mathbb{R}^n$ and $c, d \in \mathbb{R}$. Then, the following hold:

a) $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$.
b) $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$.
c) $\mathbf{a}+\mathbf{0}=\mathbf{0}+\mathbf{a}=\mathbf{a}$.
d) $\mathbf{a}+(-\mathbf{a})=-\mathbf{a}+\mathbf{a}=\mathbf{0}$.
e) $c(\mathbf{a}+\mathbf{b})=c\mathbf{a}+c\mathbf{b}=(\mathbf{a}+\mathbf{b})c$.
f) $(c+d)\mathbf{a}=c\mathbf{a}+d\mathbf{a}$.
g) $c(d\mathbf{a})=cd\mathbf{a}$.
h) $1\mathbf{a}=\mathbf{a}$.

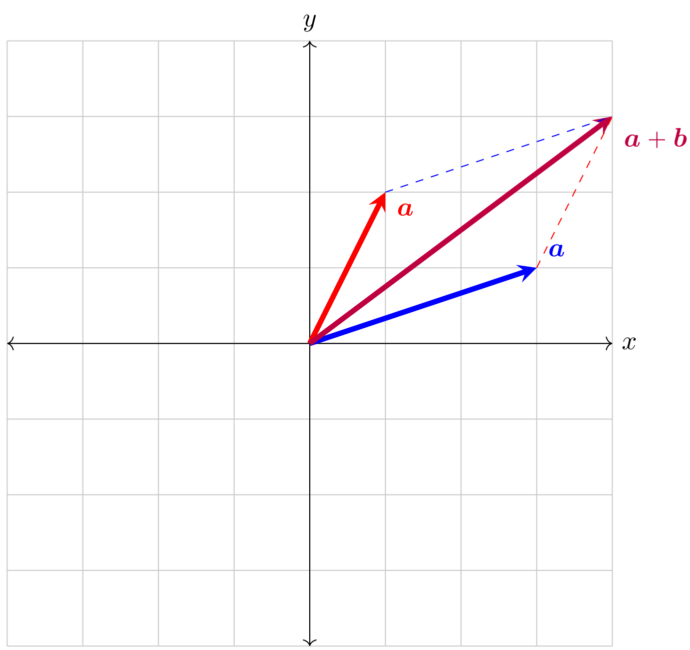
Example 2.18. We compute:

$$\frac{1}{2}\left(\begin{bmatrix}2\\4\\-6\\0\\-8\end{bmatrix}+\begin{bmatrix}4\\-2\\0\\2\\-8\end{bmatrix}\right)=\begin{bmatrix}1\\2\\-3\\0\\-4\end{bmatrix}+\begin{bmatrix}2\\-1\\0\\1\\-4\end{bmatrix}=\begin{bmatrix}3\\1\\-3\\1\\-8\end{bmatrix}.$$

Note, there are a few different ways to go about computing this. I encourage you to try and find another way!

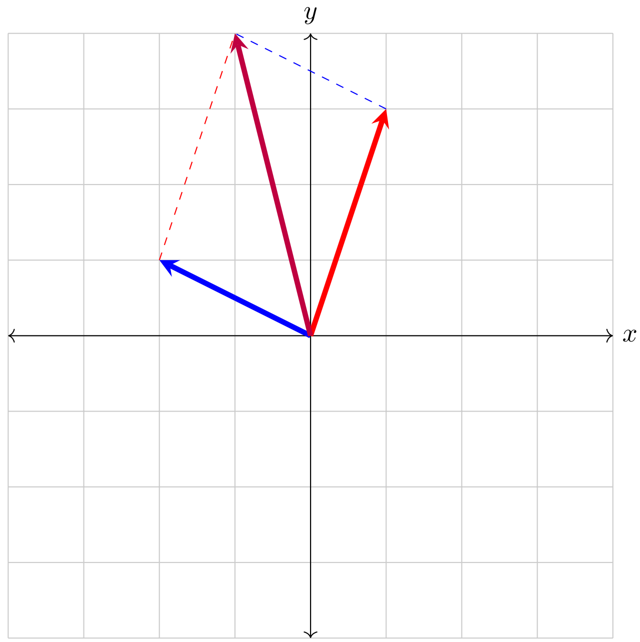
Parallelogram Law for Vectors

Proposition 2.2. The addition of two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^2$ is the fourth vertex of the parallelogram whose other three vertices are $\mathbf{0}, \mathbf{a}$, and $\mathbf{b}$.



Example 2.19. Below, we draw the vectors $\mathbf{a}=(-2,1)$ and $\mathbf{b}=(1,3)$, and then draw $\mathbf{a}+\mathbf{b}$ by using the

Parallelogram Law for Vectors.



2.3.2 Linear Combinations and Spans

Linear Combination

Definition 2.16. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be vectors in $\mathbb{R}^n$ and $c_1, \ldots, c_n \in R$. Then, we call

$$c_1\mathbf{x}_1 + \ldots + c_n\mathbf{x}_n$$

a **linear combination** of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with weights $c_1, \ldots, c_n \in R$

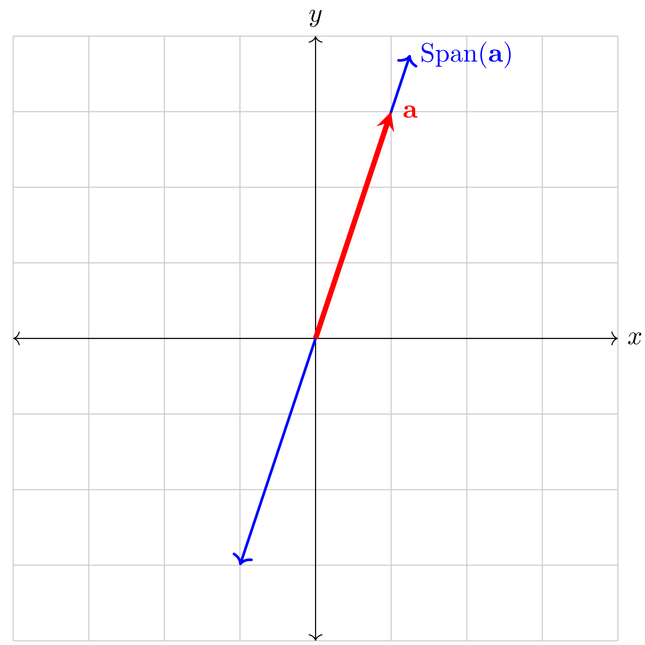
Often, we will want to consider all linear combinations of a set of vectors. To this end, we define the **span** of a set of vectors.

Span

Definition 2.17. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be vectors. We define the Span of $_\mathbf{1}, \ldots, _\mathbf{n}$ to be the set of all linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Often, we will denote this set by $\text{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$

The definition of Span seems a bit out of the blue, so lets ground ourselves in an example or two. First, lets look at the most basic example at our disposal: the span of a set of one vector in $\mathbb{R}^2$.

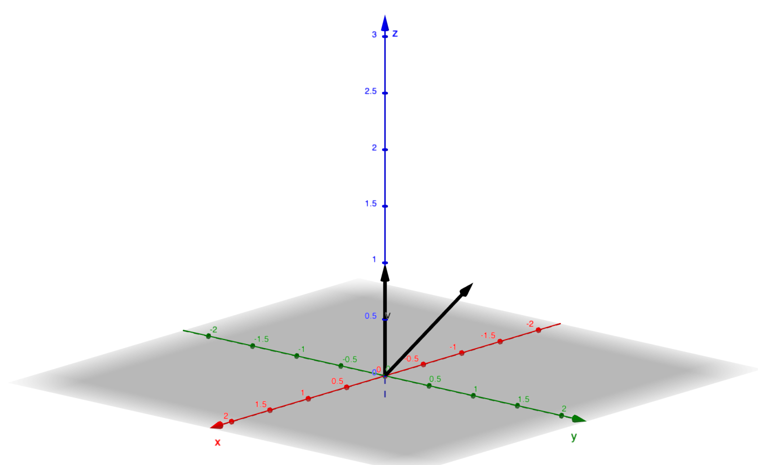
Example 2.20. Consider the vector $\mathbf{a} = (1, 3)$ in $\mathbb{R}^2$. The span of $\mathbf{a}$ is the set of all linear combinations of $\mathbf{a}$; that is the set of all things of the form $c\mathbf{a}$ with $c \in \mathbb{R}$. Geometrically, this is the line through the orgin passing through $(1, 3)$, as depicted below.



Example 2.21. Consider the vectors $\mathbf{a} = (0, 1, 1)$ and $\mathbf{b} = (0, 0, 1)$ in $\mathbb{R}^3$. What do you think Span($\mathbf{a}$) looks like? What about Span($\mathbf{a}, \mathbf{b}$)?

Using the same logic as found in the previous example, we can surmise that Span($\mathbf{a}$) is the line in $\mathbb{R}^3$ that passes through the origin passing through the point $(0, 1, 1)$.

Finding Span($\mathbf{a}, \mathbf{b}$) is slightly more challenging. It is the plane in $\mathbb{R}^3$ that contains the orgin, $\mathbf{a}$, and $\mathbf{b}$. Try and convince yourself of this! Here is a picture to help courtesy of Geogebra:



Example 2.22. Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$. We will determine whether or not the vector $\mathbf{a} = \begin{bmatrix} -1 \\ -6 \\ -2 \end{bmatrix}$ is in the span of $\mathbf{x}_1$ and $\mathbf{x}_2$. Translating from math to English, can we find weights c_1 and c_2 such that

$$c_1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -6 \\ -2 \end{bmatrix}.$$

Lets use our knowledge of addition of vectors to simplify the above equation to get:

$$\begin{bmatrix} c_1 + c_2 \\ 2c_1 + 4c_2 \\ c_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -6 \\ -2 \end{bmatrix}.$$

Thus, $\begin{bmatrix} -1 \\ -6 \\ -2 \end{bmatrix}$ is a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_n$ if and only if the system given by

$$\begin{cases} c_1 + c_2 = -1 \\ 2c_1 + 4c_2 = -6 \\ c_2 = -2 \end{cases}.$$

As if, magically, by design, we have talked about how to solve systems like this! There are a few different ways to go about it, but we will use Guassian Elimination to determine if the system is consistent or not. Skipping a few steps, of which I will leave to you to check, the REF of

$$\begin{bmatrix} 1 & 1 & -1 \\ 2 & 4 & -6 \\ 0 & 1 & -2 \end{bmatrix}$$

is the matrix

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

Thus, by Theorem the above system is consistant and so $\mathbf{a} = \begin{bmatrix} -1 \\ -6 \\ -2 \end{bmatrix}$ is a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$. Going a bit further, the REF we just found tells us that $c_1 = 1$ and $c_2 = -2$ will work (check this!)

The above example demonstrates the following theorem

The Theorem

Theorem 2.6. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and $\mathbf{b}$ be vectors. Then, the following are equivalent (i.e the following say the same thing):

1. $\mathbf{b}$ is in the span of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

2. there are $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$c_1\mathbf{x}_1 + \ldots c_n\mathbf{x}_n = \mathbf{b}.$$

3. the system, whose augmented matrix is

$$\begin{bmatrix} \mathbf{x}_1 & \ldots & \mathbf{x}_n & \mathbf{b} \end{bmatrix},$$

is consistent.

Example 2.23. Using this theorem, we can see that the span of the vector $(1,0)$ and $(0,1)$ is all of $\mathbb{R}^2$! Here, we will notice that we usually think of the dimension of $\mathbb{R}^2$ as 2, and there are two vectors that span $\mathbb{R}^2$. Hmm... maybe dimension and spanning are related. We will make this observation formal soonish!

Warning: We must be careful when we determine what the span of a collection of vectors look like. The span of a single non-zero vector will always be a line. But, the span of two vectors may not be a plane like in the exercise above; it can be a line or a plane. In fact, the span of a set of p-vectors can possibly be a 1-dimensional, 2- dimensional, 3-dimensional, ..., or p-dimensional space, though we must develop tools (not yet developed) to determine which dimension it will be.

Exercise 2.24. *Can you find an example of two vectors whose span is a line?*

2.4 Matrix Equations

As we have seen, it is often useful to interpret systems of equations as information encoded into a matrix (i.e its augmented matrix). We will continue this theme of translating ideas into expressions involving matrices! Last time we talked about linear combinations of vectors in $\mathbb{R}^n$; it turns out that we can encode this infomration into something called a matrix equation. Before we do this, we should discuss matrix operations.

2.4.1 Matrix Equations

Sometimes vectors aren't enough information. Sometimes, we might be tempted to collect vectors (in the same space) together. The usual way we go about this is through matrices!

Matrix Entry

Definition 2.18. Let A be any matrix, we define $[A]_{i,j}$ to be the (i,j)-entry of A. In other words $[A]_{i,j}$ is the entry located at the i-th row and j-th column of A.

Example 2.1. The $(2,3)$ entry of the matrix

$$A = \begin{bmatrix} 1 & 7 & 5 & 6 \\ 2 & 6 & 0 & 9 \\ 2 & 4 & 2 & 1 \end{bmatrix}$$

is $[A]_{2,3} = 0$

Square Matrix

Definition 2.19. A matrix is called a square matrix if is is of size nn for some n.

Matrix Entry

Definition 2.20. Let A be any matrix, we define $[A]_{i,j}$ to be the (i,j)-entry of A. In other words $[A]_{i,j}$ is the entry located at the i-th row and j-th column of A.

Matrix Addition

Definition 2.21. Let A and B be two matrices of the same size, say $m \times n$ with

$$A=\begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \dots & \vdots \\ a_{mn} & \dots & a_{mn} \end{bmatrix} \qquad B=\begin{bmatrix} b_{11} & \dots & a_{1n} \\ b_{21} & \dots & b_{2n} \\ \vdots & \dots & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix}.$$

We define the addition of A and B to be

$$A+B \begin{bmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & \dots & a_{2n}+b_{2n} \\ \vdots & \dots & \vdots \\ a_{1m}+b_{m1} & \dots & a_{mn}+b_{mn} \end{bmatrix}.$$

We refer to this as component wise addition.

Note that in the definition of $A+B$, where A and B are matrices of the same size, it does not make a whole lot of sense to add two matrices of different sizes!

Now lets talk about how we can multiply two matrices. Unfortunately, we cannot just multiply any two matrices we wish; the two matrices we want to multiply must complement each other in some way.

Matrix Multiplication

Definition 2.22. Let A and B be two matrices of possibly different sizes. Let A have size $m \times n$ and B have size $j \times k$. Then,

1. if the number of columns of A is the number of rows of B (or $n = j$), we define

$$AB = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \dots & \vdots \\ a_{mn} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \dots & b_{1k} \\ b_{n1} & \dots & b_{2k} \\ \vdots & \dots & \vdots \\ b_{n1} & \dots & b_{jk} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11}+a_{12}b_{21}+\dots+a_{1n}b_{n1} & a_{11}b_{12}+a_{12}b_{22}+\dots+a_{1n}b_{n2} & \dots & a_{11}b_{1k}+a_{12}b_{2k}+\dots+a_{1n}b_{nk} \\ a_{21}b_{11}+a_{22}b_{21}+\dots+a_{2n}b_{n1} & a_{21}b_{12}+a_{22}b_{22}+\dots+a_{2n}b_{n2} & \dots & a_{21}b_{1k}+a_{22}b_{2k}+\dots+a_{2n}b_{nk} \\ \vdots & \vdots & \dots & \dots \\ a_{m1}b_{11}+a_{m2}b_{21}+\dots+a_{mn}b_{n1} & a_{m1}b_{12}+a_{m2}b_{22}+\dots+a_{mn}b_{n2} & \dots & a_{m1}b_{1k}+a_{m2}b_{2k}+\dots+a_{mn}b_{nk} \end{bmatrix}$$

2. if the number of columns of A is not the number of rows of B, then AB is not defined.

Sometimes we will find it useful to point out a particular entry of a matrix product. The following proposition aids us in this endeavor

Proposition 2.3. Let A and B be two matrices of size $m \times n$ and $n \times k$, respectively. Then the (i, j) entry of the matrix AB is given by multiplying the i-th row of A against the j-th column of B. That is, using the notation of definition ,

$$[AB]_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \dots + a_{i,n}b_{n,j}$$

Exercise 2.25. ***Warning:*** *For square matrices A and B show that it is not necessarily true that $AB = BA$. In other words matrix multiplication is not commutative.*

Using this definition of matrix multiplication, we can reframe the notion of linear combinations of vectors using matrix notation.

Proposition 2.4. Let A be an $m \times n$ matrix, with columns given by the $\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}, \ldots, \mathbf{a}_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$. In other words

$$A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ a_{21} & \ldots & a_{2n} \\ \vdots & \ldots & \vdots \\ a_{mn} & \ldots & a_{mn} \end{bmatrix}.$$

Let $\mathbf{c}$ be any vector in $\mathbb{R}^n$ (note that this is the same n that occurs in the size of A). Then, the product of A and $\mathbf{c}$,

$$A\mathbf{c} = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = c_1\mathbf{a}_1 + \ldots + c_n\mathbf{a}_n$$

is the linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ with weights $c_1, \ldots, c_n$. **Warning:** $A\mathbf{c}$ makes sense, but $\mathbf{c}A$ does not! (why?)

Lets practice using this propositon!

Example 2.26. 1. $\begin{bmatrix} 2 & 7 & 8 & 11 \\ 1 & 8 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 6 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 7 \\ 8 \end{bmatrix} + 6\begin{bmatrix} 8 \\ 0 \end{bmatrix} + 0\begin{bmatrix} 11 \\ 0 \end{bmatrix}$

2. For any vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x} + 3$ in $\mathbb{R}^3$, we can write the linear combination $3\mathbf{x}_1 - 2\mathbf{x}_2 + 8\mathbf{x}_3$ as a matrix times a vector:

$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \\ 8 \end{bmatrix} = 3\mathbf{x}_1 - 2\mathbf{x}_2 + 8\mathbf{x}_3.$$

All of these equations may seem confusing and daunting, but remember: we are using matrix multiplication to translate linear combinations to an equality involving matrices (and vice versa). It is two ways of writing the same thing, and being able to fluidly go back and forth between the two is very important!

The following is an important and useful fact:

Proposition 2.5. Let A and B be $m \times n$ matrices, let $\mathbf{a}$ and $\mathbf{b}$ be vectors in $\mathbb{R}^n$ and $c \in \mathbb{R}$. Then,

1. $A(\mathbf{a} + \mathbf{b}) = A\mathbf{a} + A\mathbf{b}$
2. $(A + B)(\mathbf{a}) = A\mathbf{a} + B\mathbf{b}$
3. $A(c\mathbf{a}) = c(A\mathbf{a})$

2.4.2 Old results under new guises.

A constant theme in this class is rewriting a number of things in different ways; we've had a bit of practice with this already! So, for many sections in these notes, we will be rewriting old theorems in new terminology; in fact you will notice that the next theorem is just theorem 2.3.2 with some new notation. We will start this practice by relating systems of equations to the language of matrix equations!

The Theorem

Theorem 2.7. Let $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Then, the following are equivalent:

1. $\mathbf{x} = (x_1, \dots x_n)$ is a solution to the system of linear equations represented by the augmented matrix $\begin{bmatrix} \mathbf{a}_1 & \dots \mathbf{a}_n & \mathbf{b} \end{bmatrix}$
2. $A\mathbf{x} = \mathbf{b}$
3. $\mathbf{b}$ is a linear combination of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ with weights $x_1, \dots, x_n$. That is

$$x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n = \mathbf{b}.$$

The following corollary is a consequence of the above Theorem.

Corollary 2.1. Let $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Then, the following are equivalent:

1. For every $\mathbf{b} \in \mathbb{R}^m$, there is a solution, $\mathbf{x}$, t0 the matrix equation $A\mathbf{x} = \mathbf{b}$.
2. For every $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.
3. The columns of A span $\mathbb{R}^m$. In other words, $\text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_n) = \mathbb{R}^m$.
4. A has a pivot position in every row.

Lets get a bit of practice using these useful facts!

Example 2.27. Let $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 3 & 1 \\ 0 & 1 & -2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ be any vector in $\mathbb{R}^3$. We will determine whether or not $A\mathbf{x} = \mathbf{b}$ has a solutions (equivalently consistent) for all possible b_1, b_2, b_3.

First, lets row reduce the augmented matrix for $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 2 & 0 & b_1 \\ 2 & 3 & 1 & b_2 \\ 0 & 1 & -2 & b_3 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2 & 0 & b_1 \\ 0 & 1 & -2 & b_3 \\ 0 & 0 & 1 & \frac{-2b_1+b_2+b_3}{3} \end{bmatrix}.$$

From this, we see that A has a pivot position in each row. Hence, by our corollary above, $A\mathbf{x} = \mathbf{b}$ is consistent for any choice of $\mathbf{b}$. Note that we could have just row reduced A rather than the augmented matrix, but sometimes its useful to know what is happening to the b_i, since we can use them to solve for solutions.

2.5 Solution Sets and Applications / Worksheet 1

Homogeneous System

Definition 2.23. A system of equation is said to be homogeneous, if it can be written as $A\mathbf{x} = \mathbf{0}$. A homogeneous system always has a solution, namely $\mathbf{0}$, which we call the trivial solution. Any other solution, if it exists is called a nontrivial solution.

1. Consider the system

$$\begin{cases} 3x_1 + 5x_2 - 4x_3 = 0 \\ -3x_1 - 2x_2 - 4x_3 = 0 \\ 6x_1 + x_2 - 8x_3 - x_4 = 0 \end{cases}.$$

   (a) Does the system have a nontrivial solution?

   (b) Find a parametric description of its solution set.

(c) Think of a way to rewrite your answer in (b) as a vector equation. Hint: Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ be a solution, and use your answer in (b) to find something this vector is equal to.

Problem 1(c) inspires us to make the following definition.

Parametric Vector Equations

Definition 2.24. Suppose that $x_1, \ldots, x_n$ be the basic variables and $t_1, \ldots, t_k$ be the free variables of a system of linear equations. As we have done before, we have a parametric description of the systems solution set:

$$\begin{cases} x_1 & = a_{1,1}t_1 + \ldots + a_{1,k}t_k \\ x_2 & = a_{2,1}t_1 + \ldots + a_{2,k}t_k \\ \vdots & \vdots \\ x_n & = a_{n,1}t_1 + \ldots + a_{n,k}t_k \\ t_1 & = \text{ free} \\ \vdots & \vdots \\ t_k & = \text{free} \end{cases}$$

where all $a_{i,j}$ are real numbers. As we did in problem 1(c), we may rewrite the parametric description above as a vector equation:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} a_{1,1}t_1 + \ldots + a_{1,k}t_k \\ a_{2,1}t_1 + \ldots + a_{2,k}t_k \\ \vdots \\ a_{n,1}t_1 + \ldots + a_{n,k}t_k \\ t_1 \\ t_2 \\ \vdots \\ t_k \end{bmatrix} = t_1 \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + t_k \begin{bmatrix} a_{1,k} \\ a_{2,k} \\ \vdots \\ a_{n,k} \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

we call the vector equation above a *parametric vector equation*

Remark 2. *I would make sure that this definition makes sense and lines ups with the work we did in problem 1(c). A definition is only as good as the examples that accompany it (don't quote me on that when I forget to include examples)!*

2. Fill in the blank: the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one ________ variable. Hint: see problem 1.

Non-Homogeneous System

Definition 2.25. A system of equation is said to be non-homogeneous, if it can be written as $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} \neq \mathbf{0}$. In other words, a system of equations is said to be non-homogeneous if it is not homogeneous. **We have seen a few of these already!**

3. Come up with an example of a non-homogeneous system of equations. You don't need to solve it.

4. Come up with a non-homogeneous system of equations that does not have a solution. How is this different than homogeneous systems?

5. Consider the homogeneous system of equations

$$\begin{cases} 3x_1 + 5x_2 - 4x_3 = 1 \\ -3x_1 - 2x_2 - 4x_3 = -1 \\ 6x_1 + x_2 - 8x_3 - x_4 = 2 \end{cases}.$$

   (a) Is the system consistent?

   (b) Find a parametric description of its solution set.

(c) Think of a way to rewrite your answer in (b) as a vector equation (i.e a parametric vector equation).

Hint: Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ be a solution, and use your answer in (b) to find something this vector is equal to.

6. **Notice that the system in Problem 5 is very similar to the system we saw in problem 1.** What is the difference between the two systems?

It turns out that homogeneous systems and non-homogeneous systems with the same coefficient matrix have a nice relation between their solution sets. In short, once we know **all** solution of the homogeneous system and **one** solution of the non-homogeneous system, then we can find all solutions of the non-homogeneous system. This is nice since solving homogeneous equation is typically easier as we don't have to worry about how the row operations affect the last column of the augmented matrix (since this column is all zeros)! The method for finding these non-homogeneous solutions is described in much more detail in the next theorem:

Theorem 2.8. Suppose $A\mathbf{x} = \mathbf{b}$ is *consistent* with a solution $\mathbf{p}$ (it can be any solution you want). Then any solution, $\mathbf{w}$, to $A\mathbf{x} = \mathbf{b}$ is given by

$$\mathbf{w} = \mathbf{v}_h + \mathbf{p},$$

where $\mathbf{v}_h$ is a solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$. **Warning:** the choice of $\mathbf{v}_h$ depends on $\mathbf{w}$.

7. This exercise will outline the proof of the above theorem. There are two main parts to the proof.

   (a) Suppose that $\mathbf{w} = \mathbf{v}_h + \mathbf{p}$, where $\mathbf{v}_h$ is a solution to the homogeneous system and $\mathbf{p}$ is a solution to the non-homogeneous system. Show that $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$.

   (b) We aren't done yet! We still need to show that every solution to $A\mathbf{x} = \mathbf{b}$ has the form $\mathbf{v}_h + \mathbf{p}$ for some solution, $\mathbf{v}_h$, to the homogeneous system $A\mathbf{x} = \mathbf{0}$.

      i. Show that $\mathbf{w} - \mathbf{p}$ is a homogeneous solution to $A\mathbf{x} = \mathbf{0}$.

      ii. Set $\mathbf{v}_h := \mathbf{w} - \mathbf{p}$, and conclude $\mathbf{w} = \mathbf{v}_h + \mathbf{p}$ (this should be very short).

   (c) Briefly explain why parts (a) and (b) complete the proof of the theorem.

8. Boron Sulfide reacts with water to create boric acid and hydrogen sulfide gas. We will use linear algebra to balance the following chemical equation that illustrates this reaction:

$$x_1 B_2 S_3 + x_2 H_2 O \to x_3 H_3 B O_3 + x_4 H_2 S.$$

   To do so, find whole numbers whole numbers x_1, x_2, x_3, and x_4 such that the total number of Boron (B), Sulfur (S), Hydrogen (H), and Oxygen (0) on the left matches the number on the right. Hint: Try and set up a system of linear equations.

2.6 Linear Independence

Lets talk about independence

Linear independence is a central topic in Linear Algebra; in fact, it comes up beyond linear algebra in module theory (whatever that is). Before we define linear independence, lets discuss how it comes up based on what we did last time. If you will recall, last time we introduced *homogeneous systems of linear equation.* That is, we looked at linear systems of the form

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x} = 0,$$

where A is the systems coefficient matrix with column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$. Lets analayze how A's column vectors interact with $\mathbf{x} = (x_1, \dots, x_n)$:

$$\begin{aligned} 0 = A\mathbf{x} &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n \end{aligned}$$

This says that

$$A\mathbf{x} = 0 \qquad \text{is the same thing as} \qquad x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n = 0.$$

Therefore, if $A\mathbf{x} = 0$ has no non-trivial solution if and only if the only solution to

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n = 0$$

is $(x_1, \dots, x_n) = 0$.

On the other hand, $A\mathbf{x} = 0$ has a nontrivial solution if and only if there exist a tuple $(x_1, \dots, x_n)$ such that not all $x_i = 0$ and

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n = 0.$$

These observations lead us to make the following definitions.

Linear Independence

Definition 2.26. onsider any list of $m \times 1$ column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$. We say that the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ are *linearly independent* if the **only solution** $(x_1, \dots, x_n)$ to the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots x_n\mathbf{a}_n = 0$$

is $(x_1, x_2, \dots, x_n) = \mathbf{0}$.

Linear Dependence

Definition 2.27.
Consider any list of $m \times 1$ column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$. We say that the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ are *linearly dependent* if the vectors $\mathbf{a}_1, \dots \mathbf{a}_n$ are **not** linearly independent. In other words, there is a $n \times 1$ vector $(x_1, \dots, x_n)$ such that not every $x_i = 0$ that satisfies the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots x_n\mathbf{a}_n = 0.$$

Exercise 2.28. *Suppose a set of vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k$ contains the zero vector. Show that the vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k$ are linearly dependent.*

By our work above the two definitions just made, we see that linear independence and linear dependence of vectors translates to statements about homogeneous systems of equations. Let's spell out exactly what we mean by this.

Theorem 2.9. Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ be a list of $m \times 1$ column vectors. The following statements are equivalent (i.e say the same thing):

1. The vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ are linearly independent.
2. The homogeneous system of linear equations:

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix} \mathbf{x} = A\mathbf{x} = 0$$

   has no nontrivial solution.

The following Theorem is logically equivalent to Theorem 2.6, but we will state it anyway.

Theorem 2.10. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be a list of $m \times 1$ column vectors. The following statements are equivalent (i.e say the same thing):

1. The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly dependent.
2. The homogeneous system of linear equations:

$$\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n\end{bmatrix}\mathbf{x} = A\mathbf{x} = 0$$

   has a nontrivial solution.

I strongly encourage you to become comfortable with being able to translate between linear independence and systems of equations. These ideas will come back to haunt us time and time again.

Exercise 2.29. *If two column vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are linearly dependent, show that $\boldsymbol{a}_1$ is a scalar multiple of $\boldsymbol{a}_2$..*

Exercise 2.30. *If three column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2$, and $\boldsymbol{a}_3$ are linearly dependent, then do they need to be scalar multiples of each other?*

The next theorem is incredibly important, so I encourage you to be comforatble with the statement and why it is true.

Theorem 2.11. Consider the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ in $\mathbb{R}^n$ with $k \geqslant 2$. The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly dependent if and only if **at least one** $\mathbf{a}_i$ is a linear combination of the other $\mathbf{a}_j$. **Warning:** Not every $\mathbf{a}_i$ is a linear combination of the others! We only know there is at least one.

Proof. First, we show that if the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly dependent, then at least one of the $\mathbf{a}_i$ is a linear combination of the other $\mathbf{a}_j$'s. Since $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly dependent, there are constants $c_1, \ldots, c_k$, not all zero, such that

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \ldots + c_k\mathbf{a}_k = 0.$$

Since not all c_j are zero, there is a c_i that is not zero. Then,

$$c_1\mathbf{a}_1 + \ldots + c_{i-1}\mathbf{a}_{i-1} + c_{i+1}\mathbf{a}_{i+1} + \ldots + c_k\mathbf{a}_k = c_i\mathbf{a}_i.$$

Thus, as $c_i \neq 0$, we may divide both sides of the equation by it:

$$\frac{c_1}{c_i}\mathbf{a}_1 + \ldots + \frac{c_{i-1}}{c_i}\mathbf{a}_{i-1} + \frac{c_{i+1}}{c_i}\mathbf{a}_{i+1} + \ldots + \frac{c_k}{c_i}\mathbf{a}_k = \mathbf{a}_i.$$

Thereby proving there is at least one $\mathbf{a}_i$ that is a linear combination of the others.

On the other hand, suppose that some $\mathbf{a}_i$ is a linear combination of the others. Then, there are constants c_i, such that

$$c_1\mathbf{a}_1 + \ldots + c_{i-1}\mathbf{a}_{i-1} + c_{i+1}\mathbf{a}_{i+1} + \ldots + c_k\mathbf{a}_k = \mathbf{a}_i.$$

Therefore,

$$c_1\mathbf{a}_1 + \ldots + c_{i-1}\mathbf{a}_{i-1} - \mathbf{a}_i + c_{i+1}\mathbf{a}_{i+1} + \ldots + c_k\mathbf{a}_k =$$

□

Sometimes it is rather annoying to check whether or not a collection of vectors are linearly independent or dependent. However, occasionally we are lucky and can tell immediately through inspection. However, I encourage you to always check your answers using the definition. Nonetheless, lets talk about how we can use inspection to **sometimes** tell if a collection of vectors are linearly independent or linearly dependent. In fact, exercise 2.29 is an example of such an inspection principal; let's use it!

Example 2.2. We will determine, through inspection, if the vector $(1,2,0)$ and $(4,8,0)$ are linearly independent. Since

$$2(1,2,0) = (4,8,0),$$

exercise 2.29 tells us that $(1,2,0)$ and $(4,8,0)$ are linearly dependent. Of course, you could do this the long way, but it's nice when we can use inspection.

Often times, we will have a lot more than just two vectors for which we have to determine if they are linear independent or linear dependent. The more vectors we have, the harder to tell, at least through inspection. However, there are certain circumstances where we can tell through inspection.

Theorem 2.12. Suppose that $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are vectors in $\mathbb{R}^n$ (the n here is important)! If $k > n$, then the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly dependent.

Proof. To show that $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly dependent, by Theorem **??**, it suffices to show that the homogeneous system

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_k \end{bmatrix} \mathbf{x} = A\mathbf{x} = 0$$

has a nontrivial solution. Writing

$$\mathbf{a}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix}$$

for all i. Thus, solving $A\mathbf{x} = 0$ is the same thing as solving

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nk} \end{bmatrix} \mathbf{x} = 0.$$

By Problem 2 in Worksheet 1, this is tantamount to showing the matrix

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{bmatrix}$$

has a column with no pivot position since pivot positions correspond to free variables. As $k > n$, we have more columns than rows. Since every row can contain at most one pivot position, and we have more columns than rows, then there is at least one column with no pivot position. Hence, by Problem 2 in Worksheet 1, $A\mathbf{x} = 0$ has a nontrivial solution. Hence, the vectors $\mathbf{a}_1, \dots, \mathbf{a}_k$ are linearly dependent, as desired.

□

Remark 3. *I tend to think of Theorem 2.6 as saying a set of vectors in $\mathbb{R}^n$ is linearly dependent if the number of vectors is more than the dimension of $\mathbb{R}^n$. We are one step closer to a rigorous definition of dimension (in fact we have the language to say it now, but lets wait).*

Example 2.3. We will determine, through inspection, if the vectors

$$\begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 7 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

are linearly independent or linearly dependent. Since our vectors are in $\mathbb{R}^4$ and there are 5 vectors, by Theorem **??**, we see that the vectors are linearly dependent. Note, that this was a lot faster than checking this by hand.

Chapter 3

Even More Matrices

3.1 Arithmetic of Matrices

We have briefly spoken about matrix multiplication and matrix addition. In this section we will explore these notions and how they interact with inverses and transposes (to be defined later). Since matrix multiplication can be a bit strange the first (or even second) time we see it, lets remind ourselves how it works, and review a few examples.

Matrix Multiplication

Definition 3.1. Let A and B be two matrices of possibly different sizes. Let A have size $m \times n$ and B have size $j \times k$. Then,

1. if the number of columns of A is the number of rows of B (or $n = j$), we define

$$AB = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \dots & \vdots \\ a_{mn} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \dots & b_{1k} \\ b_{n1} & \dots & b_{2k} \\ \vdots & \dots & \vdots \\ b_{n1} & \dots & b_{jk} \end{bmatrix}$$
$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + \ldots + a_{1n}b_{n1} & a_{11}b_{12} + a_{12}b_{22} + \ldots + a_{1n}b_{n2} & \dots & a_{11}b_{1k} + a_{12}b_{2k} + \ldots + a_{1n}b_{nk} \\ a_{21}b_{11} + a_{22}b_{21} + \ldots + a_{2n}b_{n1} & a_{21}b_{12} + a_{22}b_{22} + \ldots + a_{2n}b_{n2} & \dots & a_{21}b_{1k} + a_{22}b_{2k} + \ldots + a_{2n}b_{nk} \\ \vdots & \vdots & \dots & \dots \\ a_{m1}b_{11} + a_{m2}b_{21} + \ldots + a_{mn}b_{n1} & a_{m1}b_{12} + a_{m2}b_{22} + \ldots + a_{mn}b_{n2} & \dots & a_{m1}b_{1k} + a_{m2}b_{2k} + \ldots + a_{mn}b_{nk} \end{bmatrix}$$

2. if the number of columns of A is not the number of rows of B, then AB is not defined.

Here is a good way to think about matrix multiplication (I, personally, think about it this way most often):

Proposition 3.1. Let A and B be two matrices of size $m \times n$ and $n \times k$, respectively. Then the (i, j) entry of the matrix AB is given by multiplying the i-th row of A against the j-th column of B.

Proof. This follows immediately from the definition of matrix multiplication. If it is not clear, that is okay! I encourage you to give it some more thought until you understand it better than the back of your hand. ☐

Lets do an example, just to be sure we are all on the same page!

Example 3.1. We compute the following matrix product:

$$\begin{bmatrix} 2 & 0 & 1 \\ 3 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 6 \\ 2 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2\cdot 1 + 0\cdot 2 + 1\cdot 0 & 2\cdot 6 + 0\cdot 2 + 1\cdot 1 \\ 3\cdot 1 + 0\cdot 2 + -1\cdot 0 & 3\cdot 6 + 0\cdot 2 + 1\cdot 1 \end{bmatrix} = \begin{bmatrix} 2 & 13 \\ 3 & 19 \end{bmatrix}$$

Example 3.2. The following matrix product **does not exist**

$$\begin{bmatrix} 2 & 0 & -11 \\ 8 & 8 & 1 \\ 5 & 2 & 10 \end{bmatrix} \begin{bmatrix} 2 & 9 \\ 1 & 0 \end{bmatrix}.$$

Next, we will talk about scalar multiplication of matrices; in other words, we will about how to scale a matrix by a number.

Scalar Multiplication

Definition 3.2. Let A be a matrix and c be any real number, then we define the scalar multiplication of A by c to be

$$cA = Ac = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{bmatrix}$$

It turns out that scalar multiplication, matrix multiplication, and matrix addition behave very well with each other. We describe this behavior in the next proposition. However, before we do, we define some special matrices for which it will be convinent to have notation for.

Zero Matrices and Identity Matrices

Definition 3.3. Let n be any counting number larger than zero. Consider $n \times n$ matrices

$$\begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = 0_n \qquad \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = I_n$$

.

We call 0_n the $n \times n$ zero matrix and I_n the $n \times n$ identity matrix.

The identity matrix is particularly nice looking! It only has entries of 1 along what we call the *diagonal*. We define *diagonal* more formally:

The Diagonal of a Matrix

Definition 3.4. The diagonal of a square matrix A are the entries $[A]_{i,i}$ for all possible i.

If a square matrix has zero entries outside of the diagonal, we say that the matrix is a diagonal matrix.

Warning: A non-square matrix cannot not be diagonal since there is not a nice notion of a diagonal for it.

Remark: The matrix I_n is a diagonal matrix.

Example 3.3. The following matrix

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

is a diagonal matrix. However, the matrix

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

is not a diagonal matrix since it is not a square matrix and hence there is no nice notion of a diagonal.

Properties of Matrix Arithmetic

Proposition 3.2. Let A, B, and C be matrices whose sizes are compatible for the necessary matrix operations below, and suppose that d and e are real numbers, then.

1. $A + B = B + A$
2. $A + 0 = B + 0$
3. $(A + B)C = AC + BC$
4. $C(A + B) = CA + CB$
5. $IA = AI = A$
6. $d(A + B) = dA + dB = (A + B)d$
7. $(de)A = d(eA)$
8. $(d + e)A = dA + eA$
9. $d(AB) = (dA)B = A(dB)$

These properties may or may not be surprising to you. However, it is beneficial to have be able to use these fluently. They are extremely valuable, both in a theoretic and computational point of view.

Warning: Unlike numbers, matrix multiplication cares about order! That is to say that $AB \neq BA$ in general! The following will be a homework problem

Exercise 3.1. *Find an example of two matrices A and B, such that $AB \neq BA$.*

Fun Fact: The above proposition tells us that the space of $n \times n$ matrices with real entries in $\mathbb{R}$ has a natural vector space structure (we haven't talked about what this is yet, but we will)!

Exercise 3.2. *This one is a bit challenging! Show that if*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} B = B \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

for all 2×2 matrices B, then $A = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$ for some $r \in \mathbb{R}$. Hint: set B equal to $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and see what this tells you about a, b, c or d. Keep setting B to different types of matrices like this one to widdle down what a, b, c, d are.

Example 3.4. We compute

$$\begin{aligned}
\begin{bmatrix} 1 & 2 & 7 \\ 0 & 0 & 1 \\ 1 & 0 & 6 \end{bmatrix} \left(\begin{bmatrix} 2 & 4 & 0 \\ 1 & 1 & 0 \\ 2 & 0 & 7 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \right) &= \begin{bmatrix} 1 & 2 & 7 \\ 0 & 0 & 1 \\ 1 & 0 & 6 \end{bmatrix} \begin{bmatrix} 2 & 4 & 0 \\ 1 & 1 & 0 \\ 2 & 0 & 7 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 7 \\ 0 & 0 & 1 \\ 1 & 0 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 18 & 6 & 49 \\ 2 & 0 & 7 \\ 14 & 4 & 42 \end{bmatrix} + \begin{bmatrix} 15 & 9 & 8 \\ 1 & 1 & 1 \\ 7 & 6 & 6 \end{bmatrix} \\
&= \begin{bmatrix} 33 & 15 & 57 \\ 3 & 1 & 6 \\ 21 & 10 & 48 \end{bmatrix}.
\end{aligned}$$

Can you think of another way we could have computed

$$\begin{bmatrix} 1 & 2 & 7 \\ 0 & 0 & 1 \\ 1 & 0 & 6 \end{bmatrix} \left(\begin{bmatrix} 2 & 4 & 0 \\ 1 & 1 & 0 \\ 2 & 0 & 7 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \right)?$$

Last, but certainly not least, lets talk about the powers of a square matrix!

Powers of a Square Matrix

Definition 3.5. For a square matrix A, we can take arbitrary powers of it. For notational purposes, we write

$$A^k = \underbrace{A \cdot \cdots \cdot A}_{k-\text{times}}.$$

Exercise 3.3. *Why can't we do this for non-square matrices?*

Example 3.5. Let $A = \begin{bmatrix} 1 & 2 \\ 7 & 1 \end{bmatrix}$, we will compute A^3.

$$\begin{aligned} A^3 &= \begin{bmatrix} 1 & 2 \\ 7 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 7 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 7 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 15 & 3 \\ 14 & 15 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 7 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 36 & 33 \\ 119 & 43 \end{bmatrix} \end{aligned}$$

Exercise 3.4. *Let* $\lambda \in \mathbb{R}$ *and* $A = \begin{bmatrix} \lambda & 0 & 0 & \dots & 0 \\ 0 & \lambda & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda \end{bmatrix}$. *What is* A^k *for any* k*?*

3.2 Transposes and Inverses

Sometimes, we would like to flip matrices and vectors on their sides. To this end, we make the following definitions.

Transpose of a Vector

Definition 3.6. Let $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ be a vector in $\mathbb{R}^n$. The transpose of $\mathbf{a}$ is denoted by $\mathbf{a}^T$, and we set

$$\mathbf{a}^T = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix}.$$

Transpose of a Matrix

Definition 3.7.
Let $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ be an $m \times n$ matrix. The transpose of A is denoted by A^T and is created by making the i-th column of A the i-th row of A for all columns. That is:

$$A^T = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$$

Remark 4. *Let A be an $m \times n$ matrix. The size of A^T is $n \times m$*

T

Example 3.6. e transpose of $A = \begin{bmatrix} 1 & 2 \\ 1 & 7 \end{bmatrix}$ is

$$A^T = \begin{bmatrix} 1 & 1 \\ 2 & 7 \end{bmatrix}.$$

Exercise 3.5. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *be an arbitrary* 2×2 *matrix. When does* $A^T = A$*?*

Now we will analyze how transposes play with the operation of matrices we talked about in the last section.

Proposition 3.3. Let A b have appropriate sizes for the following sums and producs. Then,

1. $(A^T)^T = A$.
2. $(A+B)^T = A^T + B^T$.
3. For any $c \in \mathbb{R}$, we have $(cA)^T = cA^T$.
4. $(AB)^T = B^T A^T$.

Proof. The proofs of parts (a), (b), and (c) will be outlined in the homework. Part (d) is a bit harder, so lets write it down here.
We prove (d). First note

$$[(AB)^T]_{i,j} = [AB]_{j,i} = \sum_k [A]_{j,k}[B]_{k,i}$$

On the other hand,

$$[B^T A^T]_{i,j} = \sum_k [B^T]_{i,k}[a^T]_{k,j} = \sum_k [B]_{k,i}[A]_{j,k}.$$

Therefore, $[(AB)^T]_{i,j} = [B^T A^T]_{i,j}$ for all i and j. Therefore $(AB)^T = B^T A^T$, as desired. □

Warning: In general $(AB)^T$ is not equal to $A^T B^T$, as we will see in the following example.

Example 3.7. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Then, $AB = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}$ with transpose $(AB)^T = \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix}$. Moreover, $A^T = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$, and $B^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. We then, see that

$$A^T B^T = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \neq (AB)^T.$$

On the other hand

$$B^T A^T = \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix} = (AB)^T.$$

3.2.1 Inverses

Now, lets turn our focus to inveres of matrices. As motivation for this, consider the set of numbers $\mathbb{R}$. For any $r \in \mathbb{R}$, there is another r^{-1} such that $r \cdot r^{-1} = 1 = r^{-1}r$. For example $5 \cdot \frac{1}{5} = 1 = \frac{1}{5} \cdot 5$. We say that r^{-1} is a *multiplicative inverse* of r.

It is natural to ask if matrices have inverses as well? Let's re-frame the question in matrix language: A **square matrix** A has and inverse and is said to be **invertible** if there is a square matrix C such that

$$CA = I_n = AC.$$

We call C the **inverse** of A and denote $C = A^{-1}$. **Warning: A priori we must check** $CA = I_n$ **and** $AC = I_n$ **since matrix multiplication cares about order.** However, it turns out in this case, $CA = I_n$ is enough to conclude that $C = A^{-1}$. However, the easiest proof I know of this fact uses determinants, which we will talk about soon. However, feel free to use the following exercise as fact. We will prove it later.

Exercise 3.6. *If A and C are square matrices of size n, and $CA = I_n$, prove that $AC = I^n$ and hence $C = A^{-1}$. Hint: if CA*

Unfortunately, while every nonzero real number has a multiplicative inverse, the same is not true for matrices.

Example 3.8. Let $A = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix}$. We will show that A has no inverse. Suppose for sake of contradiction, it did, then there is a matrix $C = \begin{bmatrix} 1 & b \\ c & d \end{bmatrix}$ such that

$$CA = I_2 = AC$$

Now,

$$CA = \begin{bmatrix} 0 & 2a + b \\ 0 & 2c + d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

This is a contradiction since $0 \neq 1$. Thus, no such matrix C exists. Hence, A is not invertible.

Example 3.9. The following matrices are invertible $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 2 & 1 \\ 1 & 7 \end{bmatrix}$, and $\begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}$. We could find their inverses in a similar way as the above example. However, lets deter that and wait until we have a more efficient method to find inverses.

Singular and Nons-singular

Definition 3.8. Let A be a square matrix. If A has an inverse, we say it is invertible (or non-singular). If A does not have an inverse, we say that it is singular.

Typically it is very hard to check if an arbitrarily large matrix is invertible. However, there is a nice test to see if a 2×2 matrix is invertible or not.

Theorem 3.1. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. A is invertible if and only if $ad - bc \neq 0$; moreover,

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

Proof. Outlined in Homework. □

Example 3.10. The matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$ is invertible since $1 \cdot 5 - 2 \cdot 3 = -1$. Moreover,

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

It turns out that invertible matrices behave well with many matrix operations that we have talked about. We gather these properties in the next theorem.

Theorem 3.2. Let A and B be invertible matrices.

1. A^{-1} is invertible and in particular $(A^{-1})^{-1} = A$.
2. AB is invertible with inverse $(AB)^{-1} = B^{-1}A^{-1}$.
3. A^T is invertible with inverse $(A^T)^{-1} = (A^{-1})^T$.

Proof. (a) and (b) are not terrible to prove, so I leave it to you as an exercise to do so. Let's prove (c). To show that $(A^T)^{-1} = (A^{-1})^T$, We need to show that $A^T \cdot A^{-1})^T = I_n$ and $A^{-1})^T \cdot A^T = I_n$. Using properteies of transpose we see that

$$A^T \cdot (A^{-1})^T = (A^{-1}A)^T = (I_n)^T = I_n.$$

Similarly

$$(A^{-1})^T \cdot A^T = (AA^{-1})^T = I_n^T = I_n.$$

Therefore, $A^{-1})^T = (A^T)^{-1}$, as desired. □

Exercise 3.7. *Come up with invertible matrices A and B, such that $A + B$ not invertible.*

3.2.2 Connection to SLE's

Why study inverses of matrices in the first place? An answer to this is the following theorem.

Theorem 3.3. Let A be the augmented matrix of some linear system and $\mathbf{b} \in \mathbb{R}^n$. If A is invertible, then $A\mathbf{x} = \mathbf{b}$ has a unique solution, namely $A^{-1}\mathbf{b}$.

Proof. Since $A\mathbf{x} = \mathbf{b}$ and A is invertible, we see that $\mathbf{x} = A^{-1}\mathbf{b}$. □

Example 3.11. In the example above we computed the inverse of the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$ to be $A^{-1} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}$. We solve $A\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$. We multiply both sides on the left by A^{-1}

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -10 \\ 6 \end{bmatrix}.$$

It turns out that invertible matrices have a role to play when it comes to row operations. In particular every row operation on an augmented matrix A (of a SLE) can be realized as EA, where E is an invertible matrix. We will go into more detail about this below. Theoretically, this is extremely powerful (we will see why); computationally, a little less so. Nonetheless having multiple view points on the same thing is a worthwhile thing!

Elementary Matrices

Definition 3.9. There are three types of elementary matrices:

1. $S_{i,j}$ is the identity matrix with the i-th row and j-th row swapped.
2. $M_{c,i}$ is the identity matrix with the i-th row multpiplied by a nonzer $c \in \mathbb{R}$.
3. $P_{c,i,j}$ is the identity matrix with the (i, j) entry replaced with a c. **Order matters here!**

I don't plan on using this notation all that much, but it's nice to have some common notation that we can all use, especially when the matrices are very large. You will notice that there is a shortcoming to this notation: no where does it indicate the size of the matrix. I spent a while trying to come up with a clean fix to this issue. Alas, it has bested me. If you have an idea for notation of elementary matrices that incorporate their size, please let me know!

Example 3.12. Lets write down some elementary matrices of size four.

$$S_{2,3} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{\sqrt{2},3} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P_{-2,3,1} = \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The next proposition explains how elementary matrices realize elementary row operations.

Proposition 3.4. Let A be any matrix of size n. Then,

1. $S_{i,j}A$ is the matrix obtained by switching the i-th and j-th rows of A.
2. $M_{c,i}A$ is the matrix obtained by multiplying the i-th column of A by a nonzero $c \in \mathbb{R}$.
3. $P_{c,i,j}A$ is the matrix obtained by adding replacing the j-th row of A with $cR_i + R_j$, where R_i is i-th row and R_j is the j-th row.

Proof. The proof is not that enlightning. Let's see some examples instead! But first, an important proposition. □

Proposition 3.5. Elementary matrices are invertible.

Proof. The inverse of an elementary matrix E is the matrix that corresponds to undoing the row operation dictated by E. Try to write the inverses down! Are they also elementary matrices? □

Example 3.13. Lets put the augmented matrix $A = \begin{bmatrix} 2 & 7 & 0 \\ 1 & 2 & 2 \end{bmatrix}$ into RREF by multiplying on the left by elementary matrices. Let's first describe the row operations we would use to compute the RREF of A; this will tell us which elementary matrices to multiply by.

I am going to write out the row operations to get A into RRER but in practice we would write the matrices to get to RREF and figure out the elementary matrices from that; I am just lazy and don't want to type out all of those matrices.

First, we will swap R_1 and R_2. Next, we will replace R_2 with $-R_1 + R_2$. Next, we divide R_2 by $\frac{1}{3}$. Lastly, we replace R_1 with $-2R_2 + R_1$. At the end of the day, when we translate this in terms of elementrary matrices, we get

$$P_{-2,2,1}M_{\frac{1}{3},2}P_{-1,1,2}S_{1,2}A = \begin{bmatrix} 1 & 0 & \frac{14}{3} \\ 0 & 1 & \frac{-4}{3} \end{bmatrix},$$

which is the RREF of A.

Doing row operations this way is rather tedious and annoying, so why would we consider these elementary matrices. Dealing with matrix multiplication in an abstract/theoretical/algorithmic frame work can be easier than describing row operations, especially in proofs, as we will now see.

Theorem 3.4. A size n matrix A is invertible if and only if A is row equivalent to I_n.

Proof. Proof deferred to next section. □

You will recall at the beginning of the inverse section, we said that finding inverses is difficult. We will talk about an algorithm to find inverses of a matrix, or to show that a matrix is not invertible.

Algorithm for Finding A^{-1}

Algorithm 3.1. Let A be an $n \times n$ matrix. We determine if A^{-1} exists, and if so, what it is as follows:

1. Place A and I_n side by side in a matrix
$$\begin{bmatrix} A & \mid & I_n. \end{bmatrix}$$
2. Row reduce $\begin{bmatrix} A & \mid & I_n. \end{bmatrix}$ to something of the form
$$\begin{bmatrix} I_n & \mid & B \end{bmatrix}.$$
3. If the second step is not possible, then A is not invertible. If it is possible, then A is invertible and $A^{-1} = A$.

This algorithm works because the identity matrix is keeping track of the row operations we are using to get the the identity. In particular, this is saying to get from A to I_n, we need to multiply A by B on the left, since B is corresponds to the row operations we performed to get A row reduced to I_n

Let's practice this!

Example 3.14. We will determine the inverse of the matrix $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 7 \\ 1 & 1 & 1 \end{bmatrix}$ by using the above algorithm. With some love and care, we can row reduce
$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 7 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$
to
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -2 & 0 \\ 0 & 1 & 0 & 0 & 1 & -7 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{array}\right].$$
Therefore $A^{-1} = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & -7 \\ -1 & -1 & 1 \end{bmatrix}$.

3.3 Inverses in Disguise

It turns out that if a square matrix A of arbitrary size is invertible, then there is a plethora of information we can extract in terms of the things we have talked about in Chapter **Warning: the matrix must be square so that we can talk about inverses.** This section will be rather short, but its implications are tremendous. So, take advantage of the lack of pages this section holds to really, and I mean really, internalize and remember what the following theorem says. Before that, we develop some notation.

Notation 1. *Lets say, for the giggles, that we have two statements a and b; for example, suppose statement a says the matrix A is invertible, and statement b says that there is a matrix B such that $BA = I$. We write*

$a \implies b$ to say statement a implies statement b. In our example, $a \implies b$ is the same thing as saying: if A is an inevitable matrix, then there is a matrix B such that $BA = I$. Leaving this example, sometimes, we will be in a situation where $a \implies b$ and $b \implies a$; rather than write this twice, we use the notation $a \iff b$.

The Inverses in Disguise Theorem

Theorem 3.5. Let A be a square matrix of size n. Then the following are equivalent.

1. A is an invertible matrix.
2. A is row equivalent to I_n.
3. A has n pivot positions.
4. The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
5. The columns of A form a linearly independent set.
6. The non-homogeneous equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$.
7. The columns of A span $\mathbb{R}^n$.
8. There is an $n \times n$ matrix C such that $CA = I_n$.
9. There is an $n \times n$ matrix D such that $AD = I_n$.
10. A^T is an invertible matrix.

Proof. Rather than a formal proof, which would take a while, lets discuss why these are all equivalent. In fact, we have seen some of these equivalences before!

1. $1 \implies 8$ and $1 \implies 9$ is immediate by definition of inverse.
2. $8 \implies 4$: If $A\mathbf{x} = \mathbf{0}$, then by multiplying both sides of the matrix equation by C on the left, we obtain $\mathbf{x} = \mathbf{0}$. Thus, $A\mathbf{x} = \mathbf{0}$ has no nontrivial solutions.
3. $4 \iff 5$ follows by Theorem 2.6.
4. $3 \iff 4$: if A has n pivot positions, then it has no free variables, and so $A\mathbf{x} = \mathbf{0}$ has no nontrivial solutions. Conversely (going backwards), if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, then A does not have a free variable. Hence A has n pivot columns.
5. $2 \iff 3$: if A and I_n are row equivalent, then A has n pivot columns since A's RREF is I_n. Conversely (going backwards) if A has n pivot columns, then there is a leading one in every columns of the REF of A. Hence, by using row operations, we can row reduce A to the identity matrix I_n.
6. $2 \implies 1$: If A is row equivalent to I_n, then there are Elementary matrices $E_1, \ldots, E_k$ such that $E_1 \cdots E_k = EA = I_n$, where we have set $E = E_1 \cdots E_k$. Now, elementrary matrices are invertible and products of invertible matrices are invertible, so E is invertible. Thus, $A = E^{-1}I_n = E^{-1}$. Since E^{-1} is invertible and $E^{-1} = A$, we conclude that A is invertible.
7. $1 \implies 6$: As A is invertible it has a left inverse A^{-1}. Thus, $A\mathbf{x} = \mathbf{b}$ implies $\mathbf{x} = A^{-1}\mathbf{b}$.
8. $6 \iff 7$: this follows by Corollary 2.4.2.
9. $8 \implies 6$: multiplying both sides of $A\mathbf{x} = \mathbf{b}$ by C on the left, we see that $I_n\mathbf{x} = C\mathbf{b}$. Since $I_n\mathbf{x} = \mathbf{x}$, this implies that $\mathbf{x} = C\mathbf{b}$. In other words $C\mathbf{b}$ is a solution to $A\mathbf{x} = \mathbf{b}$.
10. $1 \iff 10$: This follows from Theorem 3.2 and Proposition **??**.

11. $9 \implies 1$: We assume there is a matrix D such that $AD = I$. As $8 \implies 1$ and D has a multiplicative left inverse A, then D is invertible. As D is invertible, multiplying both sides of the equation $AD = I$ on the right by D^{-1} yields $A = D^{-1}$. Since D^{-1} is invertible by Theorem 3.2, we have that A is invertible, as desired.

This is enough to conclude the proof of the theorem, but lets see why using a useful diagram of implications based on what we did above:

$$
\begin{tikzcd}
& & 9 \arrow[d, Leftrightarrow] \arrow[dr, Rightarrow, dashed] & & & \\
7 \arrow[r, Leftrightarrow] & 6 \arrow[r, Leftrightarrow] & 1 \arrow[r, Rightarrow] \arrow[dl, Leftrightarrow] & 8 \arrow[dr, Rightarrow] & & \\
& 10 & 2 \arrow[u, Rightarrow] \arrow[r, Leftrightarrow] & 3 \arrow[r, Leftrightarrow] & 4 \arrow[r, Leftrightarrow] & 5
\end{tikzcd}
$$

□

Example 3.15. Determine if the following matrix is invertible by using the Inverses in Disguise Theorem:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 4 \\ 0 & 0 & 1 \end{bmatrix}.$$

We could, of course, carry out an algorithm to find an inverse (or show it does not exist); however, the beauty of the Inverses in Disguise Theorem is that we can use some of the ideas we have talked about since the beginning of class to determine if a matrix is invertible! Notice that A is in $RREF$ and has 2 pivot columns. Since A doesn't have 3 pivot columns, then A can't be invertible by the Inverses in Disguise Theorem! That was a lot faster than using an algorithm.

3.4 Linear Transformations and Matrices

Before we begin with the definition of a linear transformation (function), we define some notation and terminology, that we will find convinent.

Function

Definition 3.10. A function from a set A to a set B, denoted $f : A \to B$, is a rule that assigns every element to only one element of b. We call A the domain and B the codomain.

Range

Definition 3.11. Let $f : A \to B$ be a function. For an element $a \in A$, we say that $f(a)$ is the image of a under f. The set of all images of all elements of a is called the range of f and is denoted $\mathrm{im}(f)$

Since the beginning of the semester, we have been studying systems of linear equations in many guises ways (matrix equations, vector equations, augmented matrices). We have even used linear systems to answer questions about sets of vectors (e.g spanning, linear independence, linear dependence). This is an indication that linear systems, in particular their solutions sets, have a powerful structure behind them. The purpose of this section is to come up with a good notion of what it means to "map a solution set to another solution set". Why would we want to do this? In a very rough sense it gives us a way to compare information in one solution set to another: for example, we will use them to see whether they are the same, if one sits inside another, or if one is "larger" than another.

Let's begin with the main question: What should a function (or map) between solution sets look like? We are mathematicians, so we have the ability to declare what these maps are, but we must do so with some care in mind. We need to make sure: **a map between solution sets must preserve the "structure" of solutions sets. In particular since adding rows and scalar multiples don't change solution sets, we would like maps between solution sets to**

1. preserves addition,

2. and preserves scalar multiplication.

This leads us to make the following definition:

Linear Transformation

Definition 3.12. A map between $T : \mathbb{R}^n \to \mathbb{R}^m$ is call a *linear transformation* if for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $k \in \mathbb{R}$.

1. $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$,

2. and $T(k\mathbf{x}) = kT(\mathbf{x})$.

As with any definition, lets look at an example.

Example 3.16. Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. Then, A defines a linear transformation $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T_A\,(x_1, x_2) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

For example, $T_A(1,1) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

The following is probably one of my favorite exercises! In fact, it was a homework problem for my linear algebra class 7 years ago!

Exercise 3.8. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear transformation. Show that $T(\boldsymbol{0}) = \boldsymbol{0}$ only by using the definition of linear transformation.*

Example 3.4 can be bootstrapped to give us many examples of linear transformation. Indeed given any $n \times m$ matrix A, we can define a linear transformation $T_A : \mathbb{R}^m \to \mathbb{R}^n$ by

$$T_A((x_1, \ldots, x_n)) = A \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

It turns out that all linear transformations are determined by a matrix; isn't that neat? So, we know all the different types of linear transformations since we are familiar with matrices!

Theorem 3.6. Suppose that $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation, then T can be realized as matrix multiplication by some $n \times m$ matrix A, where A is constructed as follows: We let $\mathbf{e}_i$ denote the i-th column of the $m \times m$ identity matrix. For each i,

$$T(\mathbf{e}_i) = \begin{bmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{n,i} \end{bmatrix}.$$

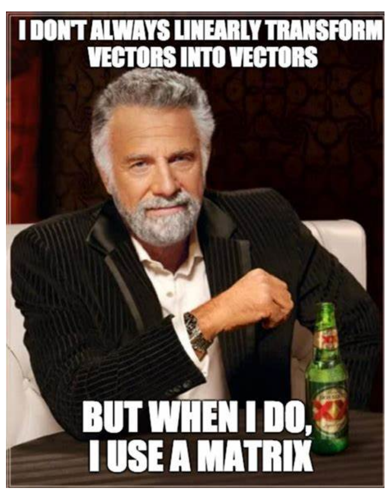
Set

$$A := \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}.$$

Then, the linear transformation T_A determined by A is exactly the linear transformation T. We call A the standard matrix of T

Proof. Deferred. This is a particular example of something involving "basis", kinda like a coordinate system, which we have yet to talk about. So, I would rather wait to prove this more generally (it's much prettier). $\square$

Remark 5. *Theorem 3.4 says that every linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ can be realized as multiplication on the left by some $n \times m$ matrix A (notice n and m switched roles!). On the other hand, every matrix gives rise to a linear transformation! Thus, to study matrices is to study linear transformations, and vice versa! Now, we can translate concepts and theorems about matrices to concepts and theorem about linear transformations! I think that's pretty neat. To sum up*



Why bring up linear transformations? Just as we have been translating statements between SLE's, Matrices, and vectors, we can do the same with linear transformations. Before, we do this though, we define a few more terms.

Injective

Definition 3.13. We say that a function (or linear transformation) $f : A \to B$ is injective if whenever $f(a) = f(b)$, then $a = b$.

Example 3.17. The functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x$ and $g(x) = x^3$ are injective.

Example 3.18. Not every function is injective! Indeed, $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$ is not injective since $f(1) = f(-1) = 1$, yet $1 \neq -1$.

Surjective

Definition 3.14. We say that a function (or linear transformation) $f : A \to B$ is surjective if for every $b \in B$, there is an $a \in A$ such that $f(a) = b$. In other words the range of f is all of B.

Example 3.19. The functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x$ and $g(x) = x^3$ are surjective.

Example 3.20. Not every function is surjective! Indeed, $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$ is not injective since the negative numbers are not in the range of f.

Bijective

Definition 3.15. We say that a function (or linear transformation) $f : A \to B$ is bijective if it is both injective and surjective.

Example 3.21. The functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x$ and $g(x) = x^3$ are bijective.

Example 3.22. There are injective functions that are not bijective (i.e not surjective). For example, the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sqrt{x}$ is injective but not surjective.
On the other hand, the function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = \sqrt{x}\ f(x) = (x-1)(x-2)x = x^3 - 3x^2 + 2$ is surjective but not injective.

Now we are ready to apply this terminology to linear transformation and connect them to systems of linear equation and matrix equations.

Theorem 3.7. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, given my multiplication by the $n \times m$ matrix A. Then, the following are equivalent:

1. T is injective.
2. The matrix equation $A\mathbf{x} = \mathbf{0}$ has no non-trivial solution.
3. The columns of A are linearly independent.

The following is really just a restatement of Theorem 3.4, but we will record it anyway.

Theorem 3.8. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, given my multiplication by the $n \times m$ matrix A. Then, the following are equivalent:

1. T is not injective.
2. The matrix equation $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.
3. The columns of A are linearly dependent.

Example 3.23. Determine if the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^4$ defined by

$$T((x_1, x_2, x_3)) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

is injective. Since the matrix has exactly three pivot columns, by a Problem 2 in Worksheet 1 and Theorem 2.6, we see that its columns are linearly independent!. Thus, by Theorem 3.4, we have that T is injective. As an aside, T being injection means that the image of T is basically of copy of $\mathbb{R}^3$ sitting inside $\mathbb{R}^4$. We will explore this "sitting inside" notion a little more in the next exercise.

Example 3.24. As we saw in Exercise 3.4, $\mathbb{R}^3$ sits inside $\mathbb{R}^4$. Intuitively, we should not expect a linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ to EVER be injective. Lets look at an explicit example to see why.
Consider the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ defined by

$$T((x_1, x_2, x_3, x_4)) = \begin{bmatrix} 2 & 3 & 9 & 1 \\ 1 & 6 & 5 & 2 \\ 0 & 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

Since there are more columns than rows, by Theorem 2.6, the columns of the matrix are not linearly independent. Thus, by Theorem 3.4, T is not injective.

Since we are mathematicians, lets bootstrap Example 3.4 to show that any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ with $n > m$ cannot be injective.

Proposition 3.6. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with $n > m$. Suppose that T is given by matrix multiplication via the $m \times n$ matrix A. Then, T is not injective.

Proof. By Theorem 3.4, we only need to show that the columns of A are linearly dependent. Since $n > m$ the number of columns of A is larger than the number of rows. Hence, by Theorem 2.6, the columns of A are linearly dependent. By Theorem 3.4, T is not injective. ☐

We can now add some things into our The Inverses in Disguise Theorem! I'll copy them here for your convenience.

Theorem 3.9. Let A be an $n \times n$ matrix

1. A is an invertible matrix.
2. A is row equivalent to I_n.
3. A has n pivot positions.
4. The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
5. The columns of A form a linearly independent set.
6. The non-homogeneous equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$.
7. The columns of A span $\mathbb{R}^n$.
8. There is an $n \times n$ matrix C such that $CA = I_n$.
9. There is an $n \times n$ matrix D such that $AD = I_n$.
10. A^T is an invertible matrix.
11. The linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$ given by multiplication by A is injective.

Let's switch gears slightly to surjective linear transformations and what they tell us about matrix equations!

Theorem 3.10. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, given by multiplication by the $n \times m$ matrix A. Then, the following are equivalent:

1. T is surjective.
2. The matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution (is consistent) for every vector $\mathbf{b} \in \mathbb{R}^n$.
3. The columns of A span $\mathbb{R}^n$.

The following theorem is really just a restatement of Theorem 3.4, but we will write it down anyway!

Theorem 3.11. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, given by multiplication by the $n \times m$ matrix A. Then, the following are equivalent:

1. T is not surjective.
2. The matrix equation $A\mathbf{x} = \mathbf{b}$ does not have a solution (is consistent) for some vector $\mathbf{b} \in \mathbb{R}^n$.
3. The columns of A do not span $\mathbb{R}^n$.

Example 3.25. Let $T : \mathbb{R}^3 \to R^2$ be the linear transformation given by matrix multiplication by A, where

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

We will show that T is surjective. Since the echelon form of the augemented matrix $[A \ \mid \mathbf{0}$ is

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

we see that the system of equations represented by A has one free variable. Hence, by Corollary 2.4.2, the columns of A span R^2. Moreover, by Theorem 3.4, we conclude that T is surjective.

Finally, we see what bijective linear transformations tell us about matrix equations! First, a definition.

Isomorphism

Definition 3.16. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation (not just any function. If T is bijective, then we say that T is an isomorphism. Note: all we are doing is giving bijectivity of linear transformations a special name

Theorem 3.12. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, given by multiplication by the $n \times m$ matrix A. Then, the following are equivalent:

1. T is an isomorphism.
2. A is a square invertible matrix (note that this says $m = n$).
3. The matrix equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for every vector $\mathbf{b} \in \mathbb{R}^n$.
4. The columns of A span $\mathbb{R}^n$ and are linearly independent.

The following is just a restatement of Theorem 3.4.

Theorem 3.13. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, given by multiplication by the $n \times m$ matrix A. Then, the following are equivalent:

1. T is not an isomorphism.
2. A is not a square invertible matrix (this includes the case if $m \neq n$).
3. The matrix equation $A\mathbf{x} = \mathbf{b}$ has does not have a unique solution for some vector $\mathbf{b} \in \mathbb{R}^n$.
4. The columns A do not span $\mathbb{R}^n$ or are linearly dependent.

Example 3.26. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation given by the matrix

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

Since columns 2 and 3 are the same, the columns of A are not linearly dependent. Hence, by Theorem 3.4, the map T is not an isomorphism!

3.5 Scaling, Reflections, Rotations, and Shears

In this section, we will continue our study of linear transformations. In particular, we will discuss special types of linear transformations that "act geometrically", e.g scaling, reflections, rotations, ect... For these we will restrict our attention to linear transformations domain and codomain are both $\mathbb{R}^2$. The physicists, engineers, and other sciences, will likely get a lot of use from this section (though I promise nothing).

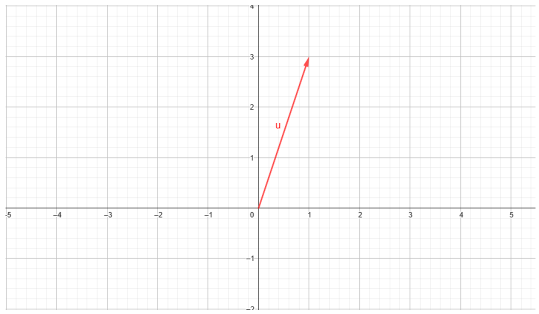
First, we focus on the case when $T : \mathbb{R}^2 \to \mathbb{R}^2$. We can (and will) think of T as acting on $\mathbb{R}^2$. In fact, there is a rigorous notion of what it means for a special type of set called a *group* to act on a space! When our space is $\mathbb{R}^n$ or even $\mathbb{C}^n$ ($\mathbb{C}$ is the complex numbers), much is known about how isomorphisms act; the study of these is called representation theory of (this is something I am currently learning about)! We won't dive that deep into linear transformation, but we will draw some pretty pictures (though if you are interested, I am happy to point you to some resources).

Scaling

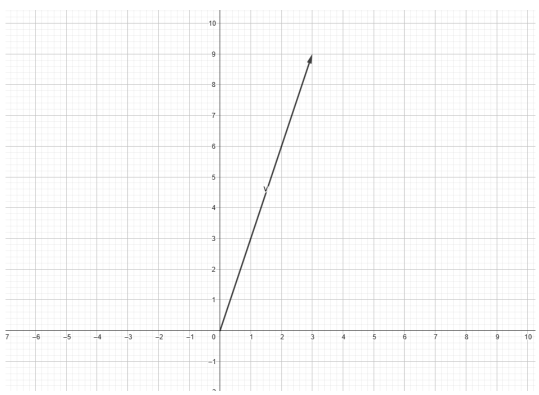
Definition 3.17. The 2×2 matrix that scales by a factor of $k > 0$ is

$$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}.$$

Example 3.27. Consider the vector $\mathbf{u} = (1, 3)$:



.
The scaling matrix $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = (2, 6)$:



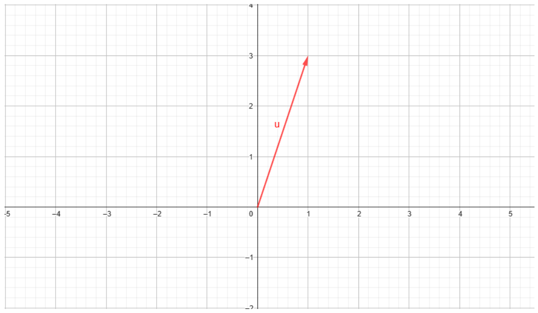
.

Scaling

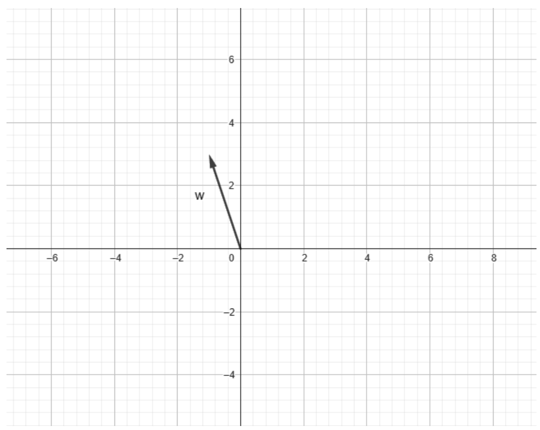
Definition 3.18. The 2×2 matrix that reflects across the x-axis is

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

Example 3.28. Consider the vector $\mathbf{u} = (1, 3)$:



.
The matrix $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \end{bmatrix} = (-1, 3)$:



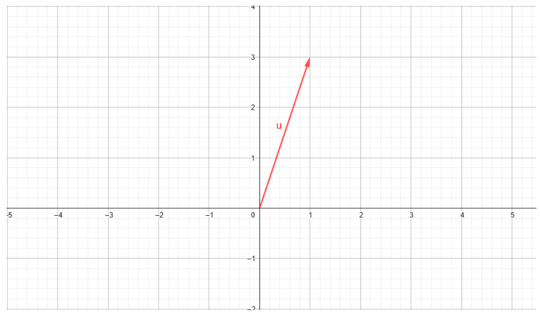
.

Reflection across y-axis

Definition 3.19. 2×2 matrix that scales by a factor of $k > 0$ is

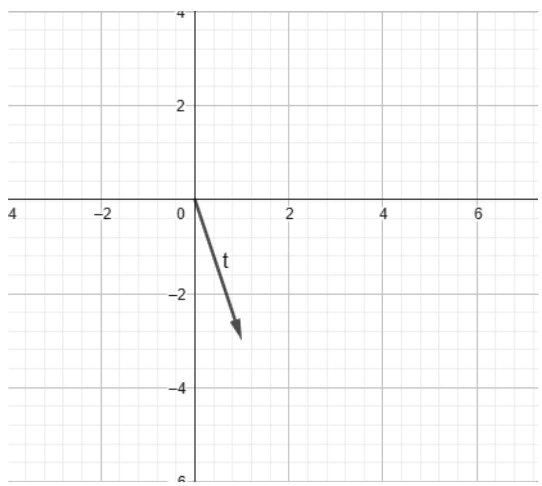
$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Example 3.29. Consider the vector $\mathbf{u} = (1, 3)$:



.

The matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (1, -3)$:



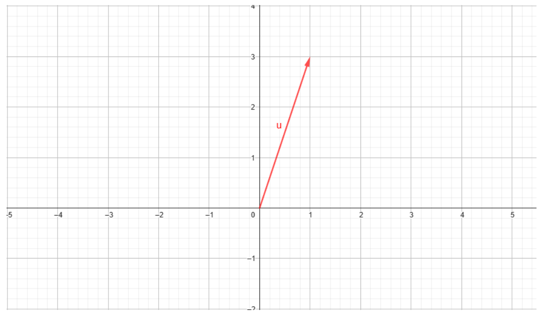
.

Horizontal Shear

Definition 3.20. The 2×2 matrix that horizontally by a factor of $k \neq 0$ is

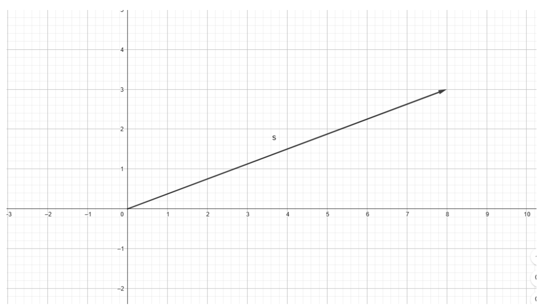
$$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}.$$

Example 3.30. Consider the vector $\mathbf{u} = (1, 3)$:



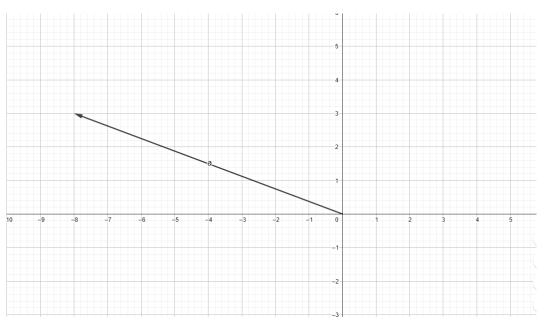
.

The matrix $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = (7, 3)$:



.

The matrix $\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = (-5, 3)$:



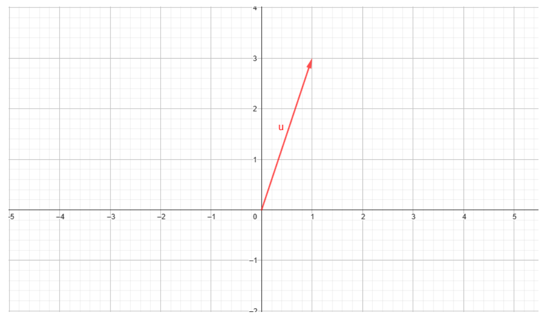
.

Vertical Shear

Definition 3.21. The 2×2 matrix that shears vertically by a factor of $k \neq 0$ is

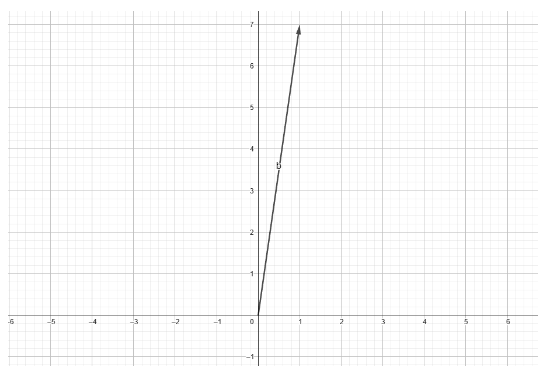
$$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}.$$

Example 3.31. Consider the vector $\mathbf{u} = (1, 3)$:



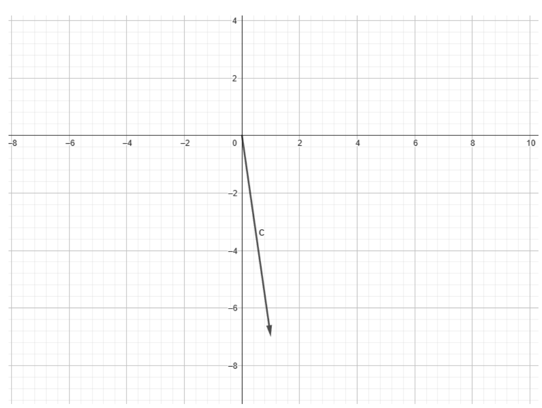
.

The matrix $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (1, 5)$:



.

The matrix $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (1, 1)$:



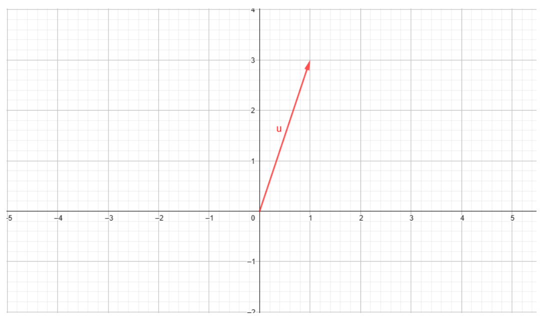
.

Rotation Matrix

Definition 3.22. 2×2 matrix that rotates by θ radians is

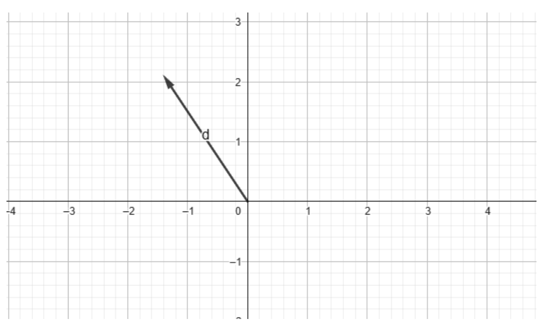
$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Example 3.32. Consider the vector $\mathbf{u} = (1, 3)$:



.

The matrix $\begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (\frac{-2}{\sqrt{2}}, \frac{3}{\sqrt{2}})$:



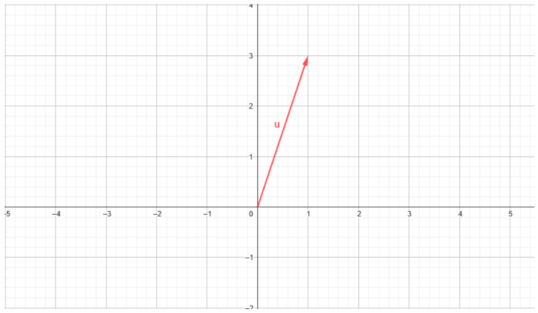
.

Reflection across $y = x$

Definition 3.23. The 2×2 matrix that reflects across the line $y = x$ is:

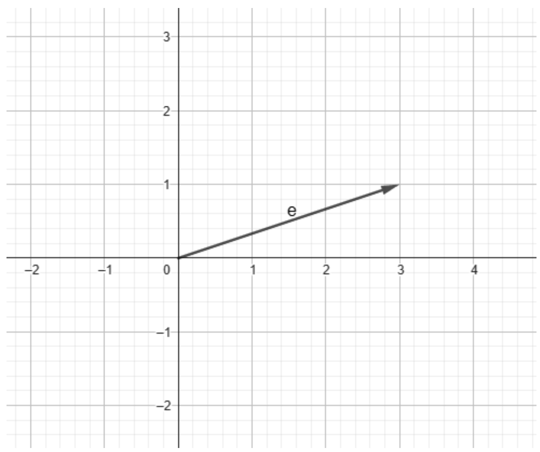
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Example 3.33. Consider the vector $\mathbf{u} = (1, 3)$:



.

The matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = (3, 1)$:



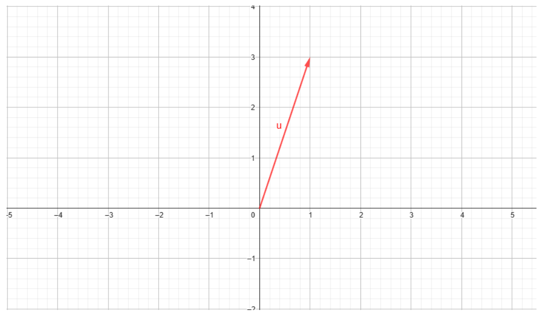
.

Reflection across $y = -x$

Definition 3.24. The 2×2 matrix that reflects across the line $y = -x$ is

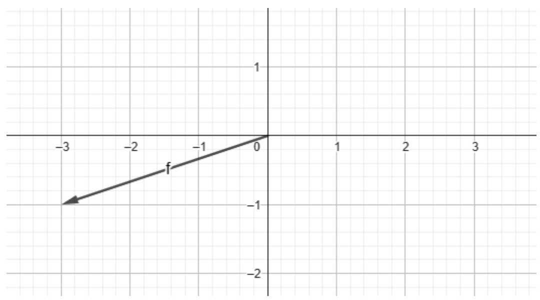
$$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Example 3.34. Consider the vector $\mathbf{u} = (1, 3)$:



.

The matrix $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ takes $\mathbf{u}$ to $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (-3, -1)$:



.

Chapter 4

Determinants

4.1 What and Why are Determinants

The purpose of this section is to, hopefully, motivate the idea behind determinants, or at least convince you that they are not coming at us from out of the blue! We will also define, in full generality, what a determinant is.

For a moment, lets think of the real numbers as square matrices of size one. That is a real number s can be thought of the matrix $[s]$. We know that the number s is invertible if and only if $s \neq 0$. So, for 1×1 matrices, we set the determinant of $[s]$ to be $\det([s]) = s$ (this number **det**ermines whether or not $[s]$ is an invertible matrix.

Next, lets explore square matrices of size two. We have seen that a square matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $ad - cb \neq 0$. As the number $ad - bc$ **det**ermines whether or not the a 2×2 matrix is invertible, lets see if we can somehow write this number as a expression of determinants of 1×1 submatrices of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Here we can see

$$a \cdot \det([d]) - b \cdot \det([c]) = ad - bc.$$

We will set

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a \cdot [d] - b \cdot [a] = ad - bc.$$

Visually, we should think that we are traveling along the first row of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$:

1. We start at a. Copy down a and multiply it by the determinate of the submatrix we get when we ignore the row and column containing a. This is where the ac is coming from.

2. Next, travel along the first row to the b. Copy down the b and multiply it by the determinate of the submatrix we get when we ignore the row and column containing b. This is where the bc is coming from, though we need to multiply it by -1 to get the $-bc$

3. Ta-da, we have $ad - bc$.

Why go through this process of rewriting $ad - bc$ in term of determinates of particular submatrices? Well, it gives us a idea of what the determinant of square $n \times n$ matrix should be! First, we develop a bit of notation.

$A_{i,j}$

Definition 4.1. Let A be any square matrix. We denote by $A_{i,j}$ the square matrix obtained by deleting the i-th row and j-th column of A. **Warning**: Do not confuse this with $[A]_{i,j}$ which is the i, j-th entry of A.

Now, we are ready to define the determinate of a square matrix!

Determinate

Definition 4.2. Let A be a square matrix of size n. We (inductively) define the determinate of A to be

$$\begin{aligned}\det(A) &= [A]_{1,1}\det(A_{1,1}) - [A]_{1,2}\det(A_{1,2}) + \dots (-1)^{n+1}[A]_{1,n}\det(A_{1,n}) \\ &= \sum_{j=1}^{n}(-1)^{j+1}[A]_{1,j}\det(A_{1,j})\end{aligned}$$

We call the expression $sum_{j=1}^{n}(-1)^{j+1}[A]_{1,j}\det(A_{1,j})$ the *cofactor expansion of A along the first row.*

If we want a determinate to **det**termine whether or not a matrix is invertible, we should see if this definition does that (otherwise it wouldn't be a good definition). We will soon see that a matrix A is invertible if and only if $\det(A) \neq 0$, but before we do that, we will have to develope some tools. For now we will content ourselves with seeing if our definition lines up with the determinant of a 2×2 matrix.

Example 4.1.
We find the determinate of

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

by using cofactor exapansion of A along the first row:

$$2 \cdot \frac{-1}{2} - (-1)(1) = 0.$$

Note that we get the same answer if we use the $ad - bc$ formula!

Let's do a more interesting example!

W

Example 4.2. find the determinate of the matrix

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

The cofactor expansion of A along the first row is

$$\begin{aligned} \det(A) &= 1 \cdot \det\left(\begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}\right) - 0 \cdot \det\left(\begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}\right) + 2 \cdot \det\left(\begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix}\right) \\ &= 1 \cdot 0 - 0 \cdot 0 + 2(-3) \\ &= -6 \end{aligned}$$

One might ask, "what is so special about the cofactor expansion along the first row? Why can't we do something similar along another row or even column?" The answer is: it doesn't matter! We just need to be a bit careful about signs though. We describe this in the nex theorem.

Theorem 4.1. Suppose that A is a square matrix of size n. Then,

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} [A]_{i,j} \det(A_{i,j})$$

for all rows i, and

$$\det(A) = \sum k = 1^n (-1)^{k+j} [A]_{k,j}$$

for all columns j.

This Theorem says that we can travel across any row and column to find $\det(A)$ like we have done for the first row! We will give these expressions a name

Cofactor Expansion

Definition 4.3. ix a row i of a square matrix A. The expression

$$\sum_{j=1}^{n} (-1)^{i+j} [A]_{i,j} \det(A_{i,j})$$

is called the cofactor expansion of A along the i-th row. Now, fix a column j of A. The expression

$$\sum k = 1^n (-1)^{k+j} [A]_{k,j}$$

is called the cofactor expansion of A along the j-th column.

The neat thing about this Theorem, is it can potentially make our lives easier when determining determinates, as long as we choose to do a cofactor expansion along a row or column with many zeros!

Example 4.3. We find the determinate of the matrix

$$A = \begin{bmatrix} 1928 & 2008 & 9039 \\ 0 & 1837 & 290 \\ 0 & 1 & 0 \end{bmatrix}.$$

This is a disgusting matrix, to find the determinate of using cofactor expansion along the first row. Instead, lets be clever about which row or column we want to do a cofactor expansion along. Let's do column 1 (row 3 is also a good choice):

$$\det(A) = 1928 \cdot \det\left(\begin{bmatrix} 1837 & 290 \\ 1 & 0 \end{bmatrix}\right) - 0\det(A_{2,0}) + 0\det(A_{3,0}) = 1928 \cdot -290.$$

We can use the same idea to find the determinants of a special class of matrices, called triangluar matrices.

Definition 4.4. A square matrix with all zero entries above the diagonal is called an lower triangular matrix. A square matrix with all zero entries below the diagonal is called a upper triangular matrix

Theorem 4.2. The determinate of a lower or upper triangular matrix is the product of the entries along the diagonal.

Proof. The proof involves a technique called induction. If you are familiar with induction, I encourage you to try it! □

Example 4.4. The determinate of the upper triangular matrix

$$\begin{bmatrix} 1 & 82 & 290 & 902 & 290 \\ 0 & 2 & 92008 & 92 & 0 \\ 0 & 0 & 3 & 92 & 2232 \\ 0 & 0 & 0 & 4 & 209 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

is $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 5! = 120$. You can also see this by doing a cofactor expansion along the first column.

Exercise 4.1. *What is the determinate of a square matrix A that has a row or column of just zeros?*

The following Theorem is very useful!

Theorem 4.3. Let A and B be $n \times n$ matrices. Then,

1. $\det(A) = \det(A^T)$
2. $\det(AB) = \det(AB)$

Proof. We will not prove these in detail, but for those who are interested, they can both be prove my induction and implementing cofactor expansion! □

Exercise 4.2. *For square matrices of the same size, say A and B, is it true that* $\det(A+B) = \det(A)+\det(B)$

4.2 Properties of Determinants

It turns out the determinants interact well with row operations, and we can use row operations on a matrix to determine its determinant. However, things will not be as simple as we would hope. Ideally, we would hope that if A and B are row equivalent, then $\det(A) = \det(B)$, but **this is not true in general**. So, when we do row operations on a matrix and want to use them to find determinants, we need to keep track of some more information. We will describe how to do this now.

Row Operations and Determinants

Theorem 4.4. et A be a square matrix.

1. If a multiple of one row of A to another row of A to produce a matrix B, then $\det(B) = \det(A)$.
2. If two rows of A are interchanged to produce B, then $\det(B) = -\det(A)$.
3. If one row of A is multiplied by K to produce B, then $\det(B) = k\det(A)$.

Proof. Each of the items above can be realized as multiplication of A on the left by an elementary matrix. In particular, we can write $B = EA$ for some elementary matrix E. As we shall see, $\det(B) = \det(EA) = \det(E)\det(A)$. If $E = P_{c,i,j}$, as this is a triangular matrix, whose diagonal consists of only 1's, we have $\det(E) = 1$. This yields part (a).

If $E = S_{i,j}$, then one can show that $\det(E) = -1$; this is a bit harder to see is true. Induction is the best way to prove it, so if you know what that is, try it! If not, that's okay; try seeing it is true for matrices of size 2 and 3. This yields (b).

If $E = C_{c,i}$, then $\det(E) = c$; this implies (c). □

Before we see an example, we develope some convinent notation. If $A = \begin{bmatrix} a_{11} & \dots & a_{1,n} \\ \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$, then we sometimes write

$$\det(A) = \begin{vmatrix} a_{11} & \dots & a_{1,n} \\ \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

Example 4.5. We compute $\det(A)$, where $A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$.

The idea is to keep track of the row operations we use to get A into REF, and then use the above theorem.

Lets multiply the top row by $\frac{1}{2}$ 2; then,

$$\frac{1}{2}\det(A) = \begin{vmatrix} 1 & -4 & 3 & 4 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{vmatrix}.$$

Next, we do row replacements to get 0's under the 1 in the first column. As, these do not change the determinant, we have

$$\frac{1}{2}\det(A) = \begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & -12 & 10 & 10 \\ 0 & 0 & -3 & 2 \end{vmatrix}$$

Next, we can add 4 times row to row 3; this does not change the determinant:

$$\frac{1}{2}\det(A) = \begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & -3 & 2 \end{vmatrix}.$$

Lastly, we multiply row 3 by $\frac{-1}{2}$ and add to row 4; again, this does not change the determinant

$$\frac{1}{2}\det(A) = \begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix}.$$

We can find the determinant of this new matrix since it is an upper triangular matrix.

$$\frac{1}{2}\det(A) = \begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = -18.$$

Hence, $\det(A) = -36$.

The following theorem utilizes how row operations interact with determinants.

Theorem 4.5. A square matrix A is invertible is if and only if $\det(A) \neq 0$.

Proof. We have that A is invertible if and only if A is row equivalent to the identity matrix. Thus, A is invertible if and only if there are elementary matrices $E_1, \ldots, E_m$ so that

$$E_1 \cdots E_m A = I_n$$

In particular, $\det(E_1 \cdots E_m A) = 1$. We see this implies $\det(E_1) \cdots \det(E_m) \det(A) = 1$. Thus, $\det(A) \neq 0$. ☐

This means that we can add another thing to our inverse in disguises theorem! Moreover, we can now say that **determin**ants determine when a square matrix is invertible!

Theorem 4.6. Let A be an $n \times n$ matrix

1. A is an invertible matrix.
2. A is row equivalent to I_n.
3. A has n pivot positions.
4. The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
5. The columns of A form a linearly independent set.
6. The non-homogeneous equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$.
7. The columns of A span $\mathbb{R}^n$.
8. There is an $n \times n$ matrix C such that $CA = I_n$.
9. There is an $n \times n$ matrix D such that $AD = I_n$.
10. A^T is an invertible matrix.
11. The linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$ given by multiplication by A is injective.
12. $\det(A) \neq 0$.

Example 4.6. The matrix $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 7 & 0 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 2 & 6 & 1 \end{bmatrix}$

is invertible since $\det(A) = 14 \neq 0$. However, the matrix $B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 7 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 2 & 6 & 1 \end{bmatrix}$

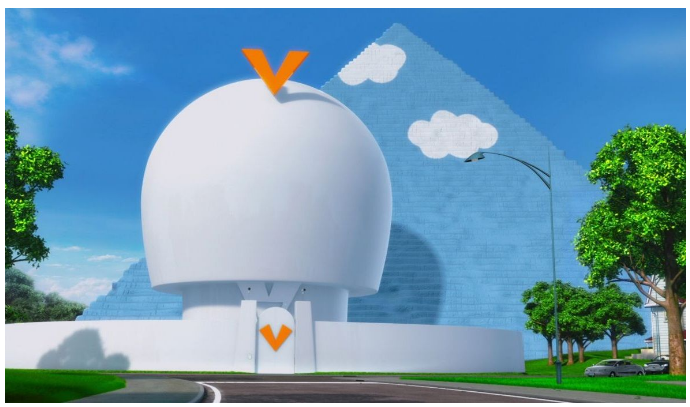
is not invertible since $\det(A) = 0$.

Chapter 5

Vector Spaces

5.1 Vector Spaces

The time has come, we now study vector spaces. I am mathematically obligated to include the following image:



Before we begin our study of vector spaces, we note that in our study of vectors in $\mathbb{R}^n$, we have encountered a vector space, namely $\mathbb{R}^n$. If you recall, we had a list of properties of vectors of $\mathbb{R}^n$, Proposition5.1, which we copy down here for your convenience.

Proposition 5.1. Let $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ be vectors in $\mathbb{R}^n$ and $c, d \in \mathbb{R}$. Then, the following hold:

a) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.
b) $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$.
c) $\mathbf{a} + \mathbf{0} = \mathbf{0} + \mathbf{a} = \mathbf{a}$.
d) $\mathbf{a} + (-\mathbf{a}) = -\mathbf{a} + \mathbf{a} = \mathbf{0}$.
e) $c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b} = (\mathbf{a} + \mathbf{b})c$.
f) $(c + d)\mathbf{a} = c\mathbf{a} + d\mathbf{a}$.
g) $c(d\mathbf{a}) = cd\mathbf{a}$.
h) $1\mathbf{a} = \mathbf{a}$.

These properties are ones that we might take for granted, or even say they are "obvious". It turns out that there are other types of spaces, that are not $\mathbb{R}^n$, that have all of the properties of Proposition 5.1; we call these **vector spaces**.

Vector Spaces

Definition 5.1. A vector space is a set V with an addition operation and scalar multiplication by $\mathbb{R}$ operation that satisfies the following conditions: for all $\mathbf{v}$, $\mathbf{w}$, $\mathbf{u} \in V$ and $c, d \in \mathbb{R}$, we **must** have

1. $\mathbf{v} + \mathbf{w} \in V$
2. $c \cdot \mathbf{v} \in V$
3. $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.
4. $(\mathbf{v} + \mathbf{w}) + \mathbf{y} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$.
5. There exists an object $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v} = \mathbf{v}$.
6. $\mathbf{v} + (-\mathbf{v}) = -\mathbf{v} + \mathbf{v} = \mathbf{0}$.
7. $c(\mathbf{v} + \mathbf{w}) = c\mathbf{v} + c\mathbf{w} = (\mathbf{v} + \mathbf{w})c$.
8. $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$.
9. $c(d\mathbf{v}) = cd\mathbf{v}$.
10. $1\mathbf{v} = \mathbf{v}$.

Vectors

Definition 5.2. Let V be a vector space. We call the elements of V vectors **Warning**: We called elements of $\mathbb{R}^n$ vectors. Not every vector space is made out of vectors from $\mathbb{R}^n$. They can be a wide variety of thing: function, polynomials, integrable functions, differentiable functions,... ect. We will see examples below.

Before we see some interesting examples, we jot down a few remarks.

Remark 6. *The notion of a vector space is quite abstract, so to keep ourselves grounded, we should think of a vector space as a way to identify spaces that are similar to $\mathbb{R}^n$. The reason that a vector space needs to satisfy the 10 rules (or axioms) above is because $\mathbb{R}^n$ satisfies them; if we want to look for spaces that behave like $\mathbb{R}^n$, we should make sure the spaces we consider share many similar properties to $\mathbb{R}^n$.* ***However, vectors in an abstract vector space do not need to look like n-tuples of real numbers, as we will see below.***

Remark 7. *The scalars we have chosen come from $\mathbb{R}$, and so the vector spaces we have defined are usually called* ***real vector spaces****. However, they do not always need to come from $\mathbb{R}$. We can have scalars coming from $\mathbb{C}$ instead, or even $\mathbb{Q}$! For now, we will only consider scalars coming from $\mathbb{R}$; perhaps, later, we will look at vector spaces with scalars from $\mathbb{C}$, which we call* ***complex vector spaces****.*

Remark 8. *This last remark can be skipped, but if you are interested keep reading! Vector spaces can have scalars coming from anything that is a* ***field*** *(roughly speaking, a field is a set with addition and multiplication that behaves like $\mathbb{R}$). There are even finite fields! We can also bootstrap the definition of vector space to include scalars that do not come from fields, but instead, they come from* ***rings*** *(roughly speaking, a ring is a set with addition and multiplication that behaves like $\mathbb{Z}$). We call such things* ***modules****. Sadly, modules over an arbitrary ring are not as nice as vector spaces over a field. However, the theory is very interesting!*

Example 5.1. The space $\mathbb{R}^n$ is a vector space! In fact, we have shown that this is indeed the case in Proposition 5.1.

Example 5.2. The set of all functions with domain $\mathbb{R}$ and codomain $\mathbb{R}$ forms a vector space, where addition is pointwise addition of functions, and scalar multiplication is scalar multiplication of function. We denote this space by $\mathrm{Fun}(\mathbb{R})$. To show that $\mathrm{Fun}(\mathbb{R})$ indeed forms a vector space, we need to see that each of the ten axioms in the definition hold.

Let $f, g, h \in \mathrm{Fun}(\mathbb{R})$ and $c, d \in \mathbb{R}$.

$f + g \in \mathrm{Fun}(\mathbb{R})$ is immediate.

$cf \in \mathrm{Fun}(\mathbb{R})$ is immediate.

$f + g = g + f$ follows because real numbers do not care about order of addition.

$(f + g) + h = f + (g_h)$ follows because real numbers do not care about the order in which we perform addition.

Let $\mathbf{0}$ be the zero function; that is $\mathbf{0}(x) = 0$ for all $x \in \mathbb{R}$. Then, $f + \mathbf{0} = f$.

For each $f \in \mathrm{Fun}(\mathbb{R})$, we define $-f$ to be the function that sends $x \in \mathbb{R}$ to $-f(x)$. It is easily verified that $f + (-f) = \mathbf{0}$.

properties 7,8,9,10 follows by properties of real numbers.

This is a **HUGE** vector space; in fact it has infinite dimension. In the next example, we consider a slightly smaller vector space.

Example 5.3. The set of all continuous function from $\mathbb{R}$ to $\mathbb{R}$ forms a vector space, where addition is pointwise addition of functions and scalar multiplication is scalar multiplication of functions. We denote this space by $C(\mathbb{R})$. Since $C(\mathbb{R})$ sits inside of $\mathrm{Fun}(\mathbb{R})$ all properties except 1,2, and 5 automatically hold by the example above. The only things we need to check is properties 1, 2 and 5. These follow, respectively, since sums of continuous functions are continuous and scalar multiplication of continuous function are continuous (thanks Calc 1!). Moreover, $\mathbf{0}$ is a continuous function, so 5 is satisfied. Hence, $C(\mathbb{R})$ is a vector space Unfortunately, this vector space has infinite dimension too.

Example 5.4. The set of all polynomials in the variable x with coefficients in $\mathbb{R}$ form a vector space, where addition and scalar multiplication are addition and scalar multiplication of polynomials, respectively. We denote this space by $\mathbb{R}[x]$. Since $\mathbb{R}[x]$ are functions from $\mathbb{R}$ to $\mathbb{R}$, they sit inside $\mathrm{Fun}(\mathbb{R})$. Therefore, all properties except 1,2, and 5 automatically hold by the example above. We need only check properties 1, 2, and 5. These follow since addition of polynomials is a polynomial and scalar multiplication of a polynomial is still a polynomial and $\mathbf{0}$ is a polynomial. Therefore, $C(\mathbb{R})$ is a vector space! Unfortunately, this vector space has infinite dimension, as well.

Example 5.5. The set of all polynomial of degree less than or equal to $n \in \mathbb{N}$ forms a vector space, where addition and scalar multiplication are addition and scalar multiplication of polynomials, respectively. We denote this space by $\mathbb{R}[x]_{\leqslant n}$. Since $\mathbb{R}[x]_{\leqslant n}$ sits inside $\mathbb{R}[x]$ properties 3 through 10 hold automatically since they hold for $\mathbb{R}[x]$. We need only check that 1, 2, and 5 hold. The addition of two polynomials of degree less than or equal to n is also a polynomial of degree less than or equal to n (there is no way this increases degree). Similarly scalar multiplication of a polynomial of degree less than or equal to n is still a polynomial of degree less than or equal to n. Lastly, $\mathbf{0}$ is a polynomial of degree $0 \leqslant n$. Therefore, $\mathbb{R}[x]_{\leqslant n}$ is a vector space. This vector space has dimension n!

As the examples above show, sometimes spaces that sit inside vector spaces are also vector spaces. We call these *subspaces*. Moreover, we saw that the only thing we needed to check if a subset of a vector space is also a vector space is conditions 1, 2, and 5. More rigorously:

Subspaces

Definition 5.3. Let V be a vector space and $W \subseteq V$ ($\subseteq$ means sits inside). We say that W is a subspace of V if the following hold:

1. The zero vector of V is also in W.
2. For all $w, w' \in W$, we have that $w + w' \in W$. We call this property closed under addition.
3. For all $w \in W$ and $c \in \mathbb{R}$, we have that $cw \in W$. We call this property closed under scalar multiplication.

To emphasize: if W is a subspace of a vector space V, then W is also a vector space.

Example 5.6. The following spaces are subspaces of each other of, as we've seen above:

$$\mathbb{R}[x]_{\leqslant n} \subseteq \mathbb{R}[x] \subseteq C(\mathbb{R}) \subseteq \mathrm{Fun}(\mathbb{R}).$$

Example 5.7. For every vector space V, we have that $\{\mathbf{0}\} \subseteq V$ is a subspace of V, where $\mathbf{0}$ is the zero vector of V. We call this the zero subspace.

Example 5.8. The solution set to a *homogenous* system of linear equations in n variables is a subspace of $\mathbb{R}^n$.

Exercise 5.1. *Show that the space of integrable functions is a subspace of Fun*$(\mathbb{R})$*.*

Exercise 5.2. *Show that the space of differentiable functions is a subspace of Fun*$(\mathbb{R})$*.*

One might be tempted to think that every subset of a vector space containing $\mathbf{0}$ is a subspace. This is not true. We demonstrate this through the examples below. Before we do this, we develop some convenient notation.

Notation 2. *Often we have found the need to describe sets of objects. There are many ways to do this. One can enumerate (or list) all the elements of a set; for example*

$$\{0, 1, 2, 3, 4, ..\}.$$

However, sometimes we cannot enumerate our sets (in fact, we can't do this for $\mathbb{R}$*)! To get around this we use set-builder notation. Here's how it works. Let* T *be a given set, we define*

$$S = \{t \in T \mid t \textit{ satisfies property } P\}$$

to be the set of elements of T *that satsify property* P*. Of course this is only useful if we have a set* T *to begin with, but this turns out to often be the case. You can think of the* $\mid$ *as a symbol for "such that". For example,*

$$\{n \in \mathbb{Z} \mid n \textit{ is even}\}$$

is the set of all even numbers in $\mathbb{Z}$*. I strongly encourage you to understand this notation. It is extremely clean and useful; it is very difficult to do modern mathematics without it!*

Example 5.9. Consider the subset $W := \{(0, n) \mid n \in \mathbb{Z}\}$ of $\mathbb{R}^2$. This is not a subspace since it is not closed under scalar multiplication. For example, $\frac{1}{2} \cdot (0, 1) = (0, \frac{1}{2})$ is not in W.

Example 5.10. The set of all polynomials of degree n is not subspace of $\mathbb{R}[x]$ since the zero polynomial is not in it.

Example 5.11. Any plane through the origin is a subspace of $\mathbb{R}^3$. However, every plane that does not go through the origin is not a subspace of $\mathbb{R}^3$ (it doesn't contain the zero vector/ origin).

Example 5.12. A*nonempty* solution set of a *non-homogeneous* system of equations in n variables is not a subspace of $\mathbb{R}^n$ since it does not contain the zero vector.

The last two examples are quite sad, because they are still interesting spaces to study, and are, in fact, very close to being vector spaces; the only condition they don't satisfy is that they don't contain the zero vector. One way to think about them is that they are shifted vector spaces. E.g a plane not through the origin can be shifted to be a plane through the orgin. This leads us to make the following definition

Definition 5.4. Let V be a vector space and W a subspace of V. Let $c \in \mathbb{R}$. The subset

$$W + c = \{w + c \mid w \in W\}$$

is called an *affine subspace.* **Warning:** $W + c$ is not a subspace in general. But it is related to a subspace, namely W.

We will likely explore affine subspaces later on in the course. For now, lets talk about how to get a subspace from a vector space that contains predetermined (or selected vectors). First, we extedn the definition of linear combination and spanning to an arbitrary vector space. The definitions are analagous to the ones we have seen before, but let's jot them down, so that we are all on the same page.

Linear Combination

Definition 5.5. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be vectors in a vector space V and $c_1, \ldots, c_n \in \mathbb{R}$. Then, we call

$$c_1\mathbf{x}_1 + \ldots + c_n\mathbf{x}_n$$

a **linear combination** of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with weights $c_1, \ldots, c_n \in R$

Often, we will want to consider all linear combinations of a set of vectors. To this end, we define the **span** of a set of vectors.

Span

Definition 5.6. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be vectors in a vector space $\mathbb{R}^n$. We define the Span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ to be the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Often, we will denote this set by $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

It turns out that for every vector space V and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$, then $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is a subspace of V. We encapsulate this fact in the following theorem.

Theorem 5.1. Let V be a vector space and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ vectors in V. Then, $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is a subspace of V

Proof. This follows because $0 \in \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is closed under addition and scalar multiplication. To see this, let $\mathbf{v} = c_1\mathbf{v}_1 + \ldots + c_1\mathbf{v}_n$ and $\mathbf{w} = d_1\mathbf{v}_1 + \ldots + d_n\mathbf{v}_n$ be elements in $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, and $r \in \mathbb{R}$. Then,

$$\mathbf{v} + \mathbf{w} = (c_1\mathbf{v}_1 + \ldots + c_1\mathbf{v}_n) + (d_1\mathbf{v}_1 + \ldots + d_n\mathbf{v}_n) = (c_1 + d_1)\mathbf{v}_1 + \ldots + (c_n + d_n)\mathbf{v}_n$$

and

$$r\mathbf{v} = r(c_1\mathbf{v}_1 + \ldots + c_1\mathbf{v}_n) = rc_1\mathbf{v}_1 + \ldots + rc_1\mathbf{v}_n$$

are in $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. $\square$

Spanning sets

Definition 5.7. For a vector space V and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ vectors in V, we call $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ the *subspace of V generated by $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.*

For a subspace W of V, we call $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ a *spanning set for W* if $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = W$.

The Theorem above can be useful in showing that certain subsets of a vector space are subspaces by finding a generating set for them.

Example 5.13. Let H be the set of all vectors of theform $(a-eb, b-a, a, b)$ where A and b are arbitary real numbers. Using our set-builder notation

$$H = \{(a-3b, b-a, a, b) \mid a, b \in \mathbb{R}\}.$$

We show that H is a subspace of $\mathbb{R}^4$, we can of course check this by verifying properties 1, 2, and 5, but instead lets show H is a subspace by finding a generating set for H. To to this, we use the skills we learned when writing parametric vector equations; any arbitrary vector in H has the form

$$\begin{bmatrix} a-3b \\ b-a \\ a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

This shows that $H = \operatorname{Span}\left(\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}\right)$. Hence, by our theorem above, H is a subspace of $\mathbb{R}^4$.

The following example is a type of problem we've done before; it is just phrased differently.

Example 5.14. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$ be vectors in $\mathbb{R}^3$.

For which values of h is $\mathbf{b}$ in the subspace of $\mathbb{R}^3$ spanned by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$? The difference here from what we've seen before is that we have used the terminology subspace rather than $\operatorname{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Since we have seen things like this before, I leave the rest of the example to you to solve.

Exercise 5.3. *Find a spanning set for $\mathbb{R}[x]_{\leqslant n}$, and show that your set indeed spans $\mathbb{R}[x]_{\leqslant n}$*

5.2 Introduction to Dimension

In this section, we finally define what the dimension of a vector space is, and show that the dimension of $\mathbb{R}^n$ is indeed n. However, to do so, we must introduce the concepts of linear independence and bases for vector spaces. We have seen linear independence before for the case of the vector space $\mathbb{R}^n$; much like spanning, the definition of linear independence for an arbitrary vector space is analagous to the definition we have for $\mathbb{R}^n$.

Linear Independence

Definition 5.8. et V be a vector space. We say that the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in V$ are *linearly independent* if the **only solution** $(c_1, \ldots, c_n)$ to the vector equation

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \ldots c_n\mathbf{a}_n = 0$$

is $(c_1, c_2, \ldots, c_n) = \mathbf{0}$.

Linear Dependence

Definition 5.9.
Let V be a vector space. We say that the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are *linearly dependent* if the vectors $\mathbf{a}_1, \ldots \mathbf{a}_n$ are **not** linearly independent. In other words, there are scalers $c_1, \ldots, c_n$ such that $c_i \neq 0$ for some i, and

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \ldots c_n\mathbf{a}_n = 0.$$

For abstract vector spaces, we cannot always use matrices (in an obvious way) to determine whether or not a set of vectors is linearly independent or dependent. The following example illustrates this.

Example 5.15. Let $f(x) = \sin(x)$ and $g(x) = \cos(x)$. Consider f and g as vectors in $C(\mathbb{R})$. The set vectors f, g are linearly dependent in $C(\mathbb{R})$ since one is not a scalar multiple of another. Here's how to show linear independence algebraically: suppose c_1, and c_2 are scalars such that

$$c_1 \sin(x) + c_2 \cos(x) = 0$$

for all $x \in \mathbb{R}$. Setting $x = 0$ yields $c_2 \cos(x) = 0$. Therefore, $c_1 \sin(x) = 0$ for all $x \in \mathbb{R}$. Set $x = \pi/2$, then $c_1 = 0$. Hence, f and g are linearly independent.

Example 5.16. Let $f(x) = \sin(2x)$ and $g(x) = \sin(x)\cos(x)$. Consider f and g as vectors in $C(\mathbb{R})$. Since $\sin(2x) = 2\sin(x)\cos(x)$ by trigonomoetry, we see that

$$\sin(2x) - 2\sin(x)\cos(x) = 0.$$

Therefore, f and g are linearly dependent.

Example 5.17. Let $f = x$ $g = 1$ and $h = 7x - 10$ be elements in the vector space $\mathbb{R}[x]$. The vectors f, g, h are linearly dependent since

$$7f - 10g = h \implies 7f - 10g - h = 0.$$

The following Theorem captures a phenomena demonstrated in the examples above (we've also seen this in the case of $\mathbb{R}^n$)

Theorem 5.2. Let V be a vector space. The vectors $v_1, \ldots, v_n$ of V are linearly dependent if and only if there is a v_k among the list $v_1, \ldots, v_n$ such that v_k is a linear combination of the remaining v_i. **Warning:** Not every vector in the list $v_1, \ldots, v_n$ satisfies the role of v_k, much like in the $\mathbb{R}^n$ case.

Basis

Definition 5.10. Let V be a vector space. A set of vectors $\mathcal{B} = \{v_1, \ldots, v_n\}$ forms a basis for V if the following are true:

1. The set of vectors $\mathcal{B}$ spans V
2. The set of vectors $\mathcal{B}$ are linearly independent.

   A good way of thinking of this definition is that a basis is a set of vectors that span the entire space in a minimal way.

Remark 9. *Given a vector space V, if $\mathcal{B}$ and $\mathcal{C}$ are bases for V, then it is not true that $\mathcal{B} = \mathcal{C}$ in general. Another way of saying this:* ***a basis for a vector space is not unique.***

Example 5.18. Let $\mathbf{v}_1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ be vectors in $\mathbb{R}^3$. We show that $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ forms a basis for $\mathbb{R}^3$. To do this, we need to shwo that this set of vectors spans $\mathbb{R}^3$ and is linearly independent. We can check each, one at a time, but a fast way to do it in this case is to use our inverses in disguise theorem (this is a square matrix). I will leave it to you to verify that the matrix

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

is row equivalent to the identity matrix. Hence, by our inverses in disguise theorem the columns of A, which are $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$, are linearly independent and span $\mathbb{R}^3$. Therefore, $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ forms a basis for $\mathbb{R}^3$.

The vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Since matrix whose columns are $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{e}_3$ is the identity matrix, which is invertible. Hence, $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are linearly independent and span $\mathbb{R}^3$, implying they form a basis for $\mathbb{R}^3$.

Definition 5.11. Consider the vector space $\mathbb{R}^n$. The vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

form a basis for $\mathbb{R}^n$. We call this basis *the standard basis for* $\mathbb{R}^n$. We also call $\mathbf{e}_i$ the *i-th standard vector of* $\mathbb{R}^n$

Theorem 5.3. Let V be a vector space. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, and set $H = \text{Span}(S)$.

1. (Cut Down): Suppose one of the vectors in S, say $\mathbf{v}_k$, is a linear combination of the remaining vectors of S. Set S' equal to the set S without $\mathbf{v}_k$. Then, $\text{Span}(S) = H = \text{Span}(S')$.
2. (A maximal cut down): If $H \neq 0$, some subset of S is a basis for H.

Proof. Omitted. $\square$

Remark 10. *Another way to interpret this theorem is that for any spanning set, we can cut it down (if necessary) into a basis.*

Theorem 5.4. Let V be a vector space and W a subspace of V with basis S. Then, there is a basis of V, say S', that contains S.

Proof. Ommited. $\square$

Remark 11. *The proofs above are omitted as I think it is better to get some intuition rather than seeing how the proof works. If you are interested in seeing a proof of these facts, feel free to come to office hours, or try proving it on your own!*

Let's see how the above theorem works!

Example 5.19. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$ be vectors in $\mathbb{R}^3$. Notice that $\mathbf{v}_1 = \mathbf{v}_2 + \mathbf{v}_3$. Then, $\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \text{Span}(\mathbf{v}_2, \mathbf{v}_3)$. We can't chop this down any further since $\mathbf{v}_2$ and $\mathbf{v}_3$ are linearly independent. Hence, $\{\mathbf{v}_1, \mathbf{v}_2\}$ forms a basis for the subspace of $\mathbb{R}^3$ generated by $\mathbf{v}_1, \mathbf{v}_2$(, and $\mathbf{v}_3$).

Example 5.20. We will find a basis for the subspace, which we will call W, of $\mathbb{R}^3$ spanned by $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix} 10 \\ 1 \\ 0 \end{bmatrix}$. First, we note that this set is linearly dependent since there are more than 3 vectors. To begin, we find a dependence relation among these vectors. One is

$$9\mathbf{v}_1 + \mathbf{v}_3 = \mathbf{v}_4.$$

Therefore, the span of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is W. One can check that $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent, as well; therefore, they also form a basis for W.

Exercise 5.4. *Show that the subspace W in the example above is $\mathbb{R}^3$. Hint: We have done things like this before but called it span instead of subspace.*

The following theorem is analogous to Theorem 2.6, except it applies to all vector spaces, not just $\mathbb{R}^n$.

Theorem 5.5. Let V be a vector space with a basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then any set in V containing more than n vectors must be linearly dependent.]

Proof. Suppose $\mathcal{C} = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is a set of vectors in V with $p > n$. Assume that

$$c_1\mathbf{u}_1 + \ldots + c_p\mathbf{u}_p = 0.$$

Since $\mathcal{B}$ is a basis for V (and hence spans), for each $i = 1, \ldots, p$, there exists scalars $s_{i,1}, \ldots, s_{i,n}$ such that

$$\mathbf{u}_i = s_{i,1}\mathbf{v}_1 + \ldots s_{i,n}\mathbf{v}_n.$$

Substituting this in for each $\mathbf{u}_i$ yields,

$$c_1(\sum_j s_{1,j}\mathbf{v}_j) + \ldots c_p(\sum_j s_{n,j}\mathbf{v}_j) = \sum_j^p c_j s_{1,j}\mathbf{v}_1 + \ldots + \sum_j^p c_j s_{n,j}\mathbf{v}_n = 0.$$

Since $\mathcal{B}$ is a basis, we have that $\sum_j^p c_j s_{i,j} = 0$ for all i. This implies we have a homogeneous matrix equation

$$\begin{bmatrix} s_{1,1} & s_{1,2} & \ldots & s_{1,p} \\ s_{2,1} & s_{2,2} & \ldots & s_{2,p} \\ \vdots & \vdots & \ldots & \vdots \\ s_{n,1} & s_{n,2} & \ldots & s_{s,p} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} = A\mathbf{c} = \mathbf{0}.$$

Since $p > n$, by Theorem 2.6, we have that the columns of the matrix above are linearly defpendent; hence there exists $(d_1, \ldots, d_p)$ with some $d_i \neq 0$, such that $A \begin{bmatrix} d_1 \\ \vdots \\ d_p \end{bmatrix} = 0$. By working backwards, we see that these same d_i yield

$$d_1\mathbf{u}_1 + \ldots + d_p\mathbf{u}_p = 0.$$

As at least one $d_i \neq 0$, we have that $\mathbf{u}_1, \ldots, \mathbf{u}_p$ are linearly dependent, as needed. □

Now we are ready to see that all basis of a vector space have the same number of elements!

Theorem 5.6. If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.

Proof. Suppose that $\mathcal{B}$ is a basis for V consisting of n vectors. Assume that $\mathcal{C}$ is another basis for V; we will show that $\mathcal{C}$ has n vectors. As $\mathcal{C}$ is linearly independent, then by Theorem 5.2, since $\mathcal{B}$ has n elements $\mathcal{C}$ has no more than n vectors. On the flip side, since $\mathcal{B}$ is linearly independent and $\mathcal{C}$ is a basis for V, then by Theorem 5.2 the number of elements of $\mathcal{B}$ cannot exceed the number of elements of $\mathcal{C}$. In other words, the number of elements of $\mathcal{C}$ is at least n. Since it is both at most and at least n, we see that $\mathcal{C}$ has n vectors. □

Theorem 5.2 allows us to make the following definition.

Dimension

Definition 5.12. Let V be a vector space. If V has a finite basis of size n, then we say V is finite-dimensional with dimension n. If V cannot be spanned by a finite set (e.g $C(\mathbb{R})$) we say that V is infinite-dimensional

Example 5.21. We have seen that the set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ in $\mathbb{R}^n$ forms a basis for $\mathbb{R}^n$. Hence, the dimension of R is n.

Example 5.22. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$ be vectors in $\mathbb{R}^3$. Notice that $\mathbf{v}_1 = \mathbf{v}_2 + \mathbf{v}_3$. We saw in Example 5.2 that $\{\mathbf{v}_1, \mathbf{v}_2\}$ forms a basis for the subspace of $\mathbb{R}^3$ generated by $\mathbf{v}_1, \mathbf{v}_2$(, and $\mathbf{v}_3$). Hence the dimension of this subspace is 2.

Exercise 5.5. *Find bases for $\mathbb{R}[x]_{\leqslant 1}$, $\mathbb{R}[x]_{\leqslant 2}$, and $\mathbb{R}[x]_{\leqslant 3}$ and determine the dimension of each of these vector spaces. Do you have a guess of what the dimension of $\mathbb{R}[x]_{\leqslant n}$ for any n?*

Proposition 5.2. Let V be a finite dimension vector space, say $\dim(V) = n$. Suppose that W is a subspace of V. Then $\dim(W) \leqslant \dim(V)$. If $W \neq V$, then $\dim(W) < \dim(V)$.

Proof. If $W = V$, then $\dim(V) = \dim(W)$. Suppose that $W \neq V$ and assume, for sake of contradiction that $\dim(W) > \dim(V)$. Then any basis $\mathcal{B}$ for W has size bigger than n. By Theorem **??**, we build a basis $\mathcal{C}$ for V that contains $\mathcal{B}$ as a subset. However, this implies that $\dim(V) \geqslant \dim(W) > \dim(V)$. This implies that $\dim(V) > \dim(V)$, which cannot happen. Thus, $\dim(W) < \dim(V)$ by way of contradiction. $\square$

We end this section by introducing another type of vector space that is closely related to $\mathbb{R}[x]_{\leqslant n}$. This vector space comes up in higher mathematics (even calc 3 a bit)! First, we define some terminology to make sure we are all on the same page.

Definition 5.13. Let $x_1, x_2, \ldots x_m$ be any number of variables (e.g x, y, z). The vector space $\mathbb{R}[x_1, x_2, \ldots, x_m]$ is the space of all polynomials in the variables $x_1, x_2, \ldots, x_m$ and with coefficients in $\mathbb{R}$.

Example 5.23. The polynomials $5x + y - 1$, $4xy + y^2x + 2z^3$, and $2x^109y^12 - z^2x - x^2z - 12$ are elements of $\mathbb{R}[x, y, z]$. Note that even though $5x + y - 1$ does not contain a z variable, it still lives inside $\mathbb{R}[x, y, z]$ since we never said all variables have to occur in the polynomial.

Definition 5.14. Let $x_1, \ldots, x_m$ be variables. We say that a polynomial in $\mathbb{R}[x_1, \ldots, x_m]$ is a monomial if it is of the form

$$rx_1^{a_1}x_2^{a_2}\cdots x_m^{a_n},$$

where each a_i is a non-negative integer, and $r \in \mathbb{R}$. Furthermore, we define the degree of a monomial $m(x_1, \ldots, x_n)rx_1^{a_1}x_2^{a_2}\cdots x_m^{a_n}$ to be $\sum_i a_i$. We denote this by

$$\deg(m(x_1, \ldots, x_n) = \sum_i a_i.$$

Example 5.24. The polynomial $3x^2 + y^2 \in \mathbb{R}[x, y]$ is not a monomial (we can't be adding terms). The polynomial $2x^4y^1z^4w^10$ in $\mathbb{R}[x, y, z, w]$ is a monomial of degree $4 + 1 + 4 + 10 = 19$.

Definition 5.15. Let $p = p(x_1, \ldots, x_m)$ be any polynomial in $\mathbb{R}[x_1, \ldots, x_n]$. We say that the degree of p is the largest degree of the monomials that make up p. We denote the degree of p by $\deg(p)$

Example 5.25. The polynomial $p = x^2 + 7y^2z^2$ in $\mathbb{R}[x, y, z]$ has degree 4. The polynomial $q = x^7 + x^2y^2z^3$ is 7.

Now we are ready to define vector spaces that are similar to $\mathbb{R}[x]_{\leqslant n}$.

Definition 5.16. Let $x_1, \ldots, x_m$ be variables. We define $\mathbb{R}[x_1, \ldots, x_m]_{\leqslant n}$ to be the set of all polynomials in $\mathbb{R}_[x_1, \ldots, x_m]$ that have degree less than or equal to n.

Proposition 5.3. The space $\mathbb{R}[x_1, \ldots, x_m]_{\leqslant n}$ is a vector space, where addition and scalar multiplication are addition and scalar multiplication of polynomials, respectively.

Proof. We only need to show it is a subspace of $\mathbb{R}[x_1, \ldots, x_m]$. Two polynomials of degree less than or equal to n sum to a polynomial of degree less than or equal to n $(\deg(f+g) \leqslant \max\{\deg(f), \deg(g)\}$. Similarly scalar multiplication of a polynomial by $r \in \mathbb{R}$ keeps the degree the same if $r \neq 0$, and takes the degree to 0 if $r = 0$. $\square$

Exercise 5.6. *Can you think of a basis for $\mathbb{R}[x_1, \ldots, x_m]_{\leqslant n}$. Hint: monomials seem to make up the polynomials...*

The dimension of $\mathbb{R}[x_1, \ldots, x_m]_{\leqslant n}$ is slightly harder to find than the dimension of $\mathbb{R}[x]_{\leqslant n}$. One first observes that the set of all monomials form a basis for $\mathbb{R}[x_1, \ldots, x_m]_{\leqslant n}$ involves a counting technique called "Stars and Bars". Since this is beyond the scope of the course (and my counting abilities), we won't talk about it. However, we will at least write down a formula for the dimension (we just won't prove it).

Proposition 5.4. Let $x_1, \ldots, x_m$ be variables and let n be any positive integer. Then,

$$\dim(\mathbb{R}[x_1, \ldots, x_m]_{\leqslant n}) = \binom{m+n}{n} = \frac{(m+n)!}{m!n!}.$$

5.3 Return of the Linear Transformations

Linear Transformation

Definition 5.17. Let V and W be vector spaces (not necessarily $\mathbb{R}^n$). A linear transformation from V to W is a function $T : V \to W$ such that

1. $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.
2. $T(k\mathbf{v}) = kT(\mathbf{v})$ for all $\mathbf{v} \in V$ and $k \in \mathbb{R}$.

Example 5.26. Let $V = \mathbb{R}^4$ and $W = R^5$. Consider the function $T : \mathbb{R}^4 \to \mathbb{R}^5$ given by

$$T(\mathbf{x}) = \begin{bmatrix} 1 & 2 & 0 & 2 \\ 1 & 2 & 1 & 6 \\ 0 & 1 & 3 & 5 \\ 1 & 1 & 1 & 1 \\ 2 & 4 & 3 & 1 \end{bmatrix} \cdot \mathbf{x}.$$

We have seen before that this is a linear transformation (in the realm of $\mathbb{R}^n$ spaces). They are also linear transformations in the sense of Definition 5.3- in fact, the definition is really not different than the one we have seen previously; this one is just more general as it incorporates any pair of vector spaces.

Example 5.27. Consider the space of differentiable function, denoted Diff($\mathbb{R}$) and the space of functions form $\mathbb{R} \to \mathbb{R}$, denoted Fun($\mathbb{R}$). We define the funtion (or map)

$$D : \text{Diff}(\mathbb{R}) \to \text{Fun}(\mathbb{R})$$

by $D(f) = \frac{df}{dx}$. The map D is a linear function. Indeed, by what we have learned in Calculus 1, we have

$$D(f + g) = D(f) + D(g),$$

and

$$D(cf) = cD(f)$$

for all $c \in \mathbb{R}$ and $f, g \in \text{Diff}(\mathbb{R})$. **Warning:** You might be wondering why the codomain of D is Fun($\mathbb{R}$) and not Diff($\mathbb{R}$). It turns out that the derivative of a differentiable function does not need to be continuous (kinda wacky, right)? Can you think of an example of a differentiable function whose derivative is not continuous?

Example 5.28. Consider the vector space of differentiable functions, which we will denote, Diff($\mathbb{R}$). Let $a \in \mathbb{R}$. Define the function

$$\left.\frac{d}{dx}\right|_{x=a} : \text{Diff}(\mathbb{R}) \to \mathbb{R}$$

by $\left.\frac{d}{dx}\right|_{x=a}(f) = f'(a)$ for all $f \in \text{Diff}(\mathbb{R})$. We show that $\left.\frac{d}{dx}\right|_{x=a}$ is a linear transformation. Let $f, g \in \text{Diff}(\mathbb{R})$ and $k \in \mathbb{R}$. Then:

$$\left.\frac{d}{dx}\right|_{x=a}(f+g) = (f+g)'(a) = f'(a) + g'(a) = \left.\frac{d}{dx}\right|_{x=a}(f) + \left.\frac{d}{dx}\right|_{x=a}(g)$$
$$\left.\frac{d}{dx}\right|_{x=a}(kf) = (kf)'(a) = kf'(a) = k\left.\frac{d}{dx}\right|_{x=a}(f).$$

Therefore $\left.\frac{d}{dx}\right|_{x=a}$ is linear transformation.

Example 5.29. Consider the vector space of integrable function on $[a, b]$, which we denote by $\mathrm{Int}([a, b])$. Define the function

$$I_{[a,b]} : \mathrm{Int}([a, b]) \to \mathbb{R}.$$

by $I_{[a,b]}(f) = \int_a^b f\ dx$. The function $I_{[a,b]}$ is a linear tranformation. Indeed from Caluclus 1/2, we know that

$$\int_a^b f + g\ dx = \int_a^b f\ dx + \int_a^b g\ dx,$$

and

$$\int_a^b cf\ dx = c\int_a^b f\ dx$$

for all $c \in \mathbb{R}$ and $f, g \in \mathrm{Int}(R)$

Example 5.30. Let $\mathbb{R}[y]$ and $\mathbb{R}[x, y]$ denote the vector space of polynomials in the variable y and variables x, y, respectively. Define the map

$$T_0 : \mathbb{R}[x, y] \to \mathbb{R}[x]$$

by $T_0(f(x, y)) = f(0, y)$. For example, $T_0(x^2y + 2y^7(x - 1)) = -2y^7$. This is a linear transfomration. Indeed, suppose that $f, g \in \mathbb{R}[x, y]$ and

As the above examples show, there are many types of linear transformations that are very interesting. Indeed, two of the ones we have seen D, and $\mathrm{Int}_{[a,b]}$ are ones that you have spent a year studying at one point. Thus, it might behoov us to study linear transformations in more detail, so that we can convert information gleaned by linear algebra to say something concrete and useful about the vector spaces we are studying. To achieve this, we begin studying the null Space and the image Space of a linear transformation. Roughly speaking, the null space tells us information about how a linear transformation interacts with our domain, and the image space tells us information about how the linear transformation interacts with the codomain.

Null Space/ Kernel

Definition 5.18. Let $T : V \to W$ be a linear transformation of vector spaces. The Null space of T is the set

$$\mathrm{Null}(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}.$$

Sometimes, people call the null space of T, the kernel of T. Sometimes, I might do this as well. This is dentoted

$$\ker(T) = \mathrm{Null}(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}.$$

In short, think of null space and kernel as synonyms.

Proposition 5.5. Let $T : V \to W$ be a linear transformation of vector spaces. Show that $T(\mathbf{0}) = \mathbf{0}$.

Proof. Since T is a linear transformation,

$$T = (\mathbf{0}) = \mathbf{T}(\mathbf{0} - \mathbf{0}) = \mathbf{T}(\mathbf{0}) - \mathbf{T}(\mathbf{0}) = \mathbf{0}.$$

□

Proposition 5.6. Let $T : V \to W$ be a linear transformation of vector spaces. The kernel of T is a subspace of V.

Proof. We will prove this by using the subspace criteria.
1) We note that $T(\mathbf{0}) = (\mathbf{0})$ by Proposition 5.3.

2) Let $\mathbf{v}$ and $\mathbf{v}'$ be in $\ker(T)$; this means that $T(\mathbf{v}) = \mathbf{0}$ and $T(\mathbf{v}') = \mathbf{0}$. We will show that $T(\mathbf{v}+\mathbf{v}') = 0$. Then, since T is a linear transformation

$$T(\mathbf{v} + \mathbf{v}') = T(\mathbf{v}) + T(\mathbf{v}') = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

Therefore, $\mathbf{v} + \mathbf{v}' \in \ker(T)$.
3) Let $\mathbf{v} \in V$ and $c \in \mathbb{R}$. This means that $T(\mathbf{v}) = \mathbf{0}$. Then, since T is alinear transformation

$$T(c\mathbf{v}) = cT(\mathbf{v}) = 0 \cdot \mathbf{0} = \mathbf{0}.$$

Hence, $c\mathbf{v} \in \ker(T)$.

By the subspace criteria, we see that $\ker(T)$ is a subspace of V. □

As the next example shows, we have worked with kernels of linear transformations before!

Example 5.31. We will find the kernel of the linear transformation $T:\mathbb{R}^3\to\mathbb{R}^2$ defined by

$$T(\mathbf{x})=\begin{bmatrix}1&2&3&0\\1&0&2&0\end{bmatrix}\cdot\mathbf{x}.$$

First, we ask ourselves, "what does it mean for $\mathbf{v}$ to be in $\ker(T)$?" It means that $T(\mathbf{v})=0$. Using our explicit linear transformation this translates to

$$T(\mathbf{v})=\begin{bmatrix}1&2&3&0\\1&0&2&0\end{bmatrix}\cdot\mathbf{v}=\mathbf{0}.$$

This is a homogeneous system! The solution set will is precisely $\ker(T)$, and we've found solution sets before. Skipping some steps, we have a row reduction:

$$\begin{bmatrix}1&2&3&0&|&0\\1&0&2&0&|&0\end{bmatrix}\to\begin{bmatrix}1&0&2&0&|&0\\0&1&\frac{1}{2}&0&|&0\end{bmatrix}.$$

The solution set can be described in parametric vector form as

$$\begin{bmatrix}\mathbf{v}_1\\\mathbf{v}_2\\\mathbf{v}_3\\\mathbf{v}4\end{bmatrix}=s\begin{bmatrix}-2\\\frac{-1}{2}\\1\\0\end{bmatrix}+t\begin{bmatrix}0\\0\\0\\1\end{bmatrix}.$$

Therefore,

$$\ker(T)=\left\{s\begin{bmatrix}-2\\\frac{-1}{2}\\1\\0\end{bmatrix}+t\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\;\middle|\;s,t\in\mathbb{R}\right\}.$$

Example 5.32. Let $a\in\mathbb{R}$. Consider the linear transformation $\frac{d}{dx}\big|_{x=a}(f)$. We will describe it's kernel. For a function in $\mathrm{Diff}(\mathbb{R})$ to be in $\ker\left(\frac{d}{dx}\big|_{x=a}\right)$ means that

$$\left.\frac{d}{dx}\right|_{x=a}(f)=f'(a)=0.$$

Therefore, $\ker\left(\frac{d}{dx}\big|_{x=a}\right)$ is the set of all differentiable functions that have a critical value at $x=a$. More precisely

$$\ker\left(\left.\frac{d}{dx}\right|_{x=a}\right)=\{f\in\mathrm{Diff}(\mathbb{R})\mid f\text{ has a critical value at }x=a\}.$$

Example 5.33. Consider the linear transformation $\frac{d}{dx}:\mathrm{Diff}(\mathbb{R})\to\mathrm{Fun}(\mathbb{R})$. We will describe it's kernel. For a function $f\in\mathrm{Diff}(\mathbb{R})$ to be in $\ker\left(\frac{d}{dx}\right)$, we have $\frac{df}{dx}=0$. Therefore,

$$\ker\left(\frac{d}{dx}\right)=\{f\in\mathrm{Diff}(\mathbb{R})\mid f'=0\}=\{f=c\mid c\in\mathbb{R}\}.$$

Image Space

Definition 5.19. Let $T : V \to W$ be a linear transformation of vector spaces. The image space of T, denoted $\text{im}(T)$, is defined to be the image of T. In other words

$$\text{im}(T) = \{T(\mathbf{v}) \in W \mid \mathbf{v} \in V\}.$$

Proposition 5.7. Let $T : V \to W$ be a linear transformation of vector spaces. Then, $\text{im}(T)$ is a subspace of W.

Proof. We will prove this by using the subspace criteria.
1) Since $T(\mathbf{0}) = \mathbf{0}$ by Proposition 5.3, we see that $\mathbf{0} \in \text{im}(\mathbf{T})$.

2) Let $\mathbf{w}, \mathbf{w}' \in \text{im}(T)$. This means that there are $\mathbf{v}$ and $\mathbf{v}'$ in V such that $T(\mathbf{v}) = \mathbf{w}$ and $T(\mathbf{v}') = \mathbf{w}'$. Since T is a linear transformation $T(\mathbf{v} + \mathbf{v}') = T(\mathbf{v}) + T(\mathbf{v}') = \mathbf{w} + \mathbf{w}'$. Therefore, $\mathbf{w} + \mathbf{w}' \in \text{im}(T)$.

3) Let $\mathbf{w} \in \text{im}(T)$ and $c \in \mathbb{R}$. This means that there is a $\mathbf{v}$ such that $T(\mathbf{v}) = \mathbf{w}$. Since T is a linear transformation, we have that $T(c\mathbf{v}) = cT(\mathbf{v}) = c\mathbf{w}$. Thus, $c\mathbf{w} \in \text{im}(T)$.

By the subspace criteria, we have that $\text{im}(T)$ is a vector space. □

Given a linear transformation $T : V \to W$ of vectors spaces, as $\ker(T)$ and $\text{im}(T)$ are vector spaces, we can talk about their dimensions. Since, these are special subspaces associated to T, we give these dimensions special names.

Definition 5.20. Let $T : V \to W$ be a linear transformation. We call $\dim(\ker(T))$ the *nullity of* T, which we denote as $\text{nullity}(T)$. We call $\dim(\text{im}(T))$ the *rank of* T, which we denote as $\text{rank}(T)$.

Given a linear transformation of vector spaces, say $T : V \to W$, there is a beautiful relationship between $\dim(V), \text{nullity}T$, and $\text{rank}T$. This relationship is dubbed The Rank-Nullity Theorem. First we prove a proposition.

Proposition 5.8. Let $T : V \to W$ be a linear transformation of finite dimensional vector spaces. Let $S' = \{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$ be a basis for $\ker(T)$. By Theorem 5.2, there is a basis $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of V that contains S'. Then, $U = \{T(\mathbf{v}_{s+1}, \ldots, \mathbf{v}_k\}$ is a basis for $\mathrm{im}(T)$.

Proof. First, we show that $T(U)$ spans $\mathrm{im}(T)$. To this end, let $\mathbf{w} \in \mathrm{im}(T)$, then there is a $\mathbf{v} \in V$ such that $T(\mathbf{v}) = w$. Since S is a basis for V, there are $c_i \in \mathbb{R}$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k.$$

Then,

$$\begin{aligned}\mathbf{w} = T(\mathbf{v}) &= T(c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k)\\ &= c_1T(\mathbf{v}_1) + \ldots + c_kT(\mathbf{v}_k)\\ &= c_1T(\mathbf{v}_1) + c_sT(\mathbf{v}_s) + c_{s+1}T(\mathbf{v}_{s+1}) + \ldots + c_kT(\mathbf{v}_k)\\ &= c_{s+1}T(\mathbf{v}_{s+1}) + \ldots + c_kT(\mathbf{v}_k)\end{aligned}$$

.

Therefore, $\mathbf{w} \in \mathrm{Span}(T(U))$. As $\mathbf{w}$ was an arbitrary element of $\mathrm{im}(T)$, we have that $\mathrm{Span}(T(U)) = \mathrm{im}(T)$.

We now show that $T(U)$ is linearly independent. Suppose

$$0 = c_{s+1}T(\mathbf{v}_{s+1}) + \ldots + c_kT(\mathbf{v}_k) = T(c_{s+1}\mathbf{v}_{s+1} + \ldots + c_l\mathbf{v}_k).$$

This implies that $c_{s+1}\mathbf{v}_{s+1} + \ldots + c_l\mathbf{v}_k$ is an element of $\ker(T)$. Since $\ker(T)$ has S' as a basis, we see that there are d_i such that

$$d_1\mathbf{v}_1 + \ldots + d_s\mathbf{v}_s = c_{s+1}\mathbf{v}_{s+1} + \ldots + c_l\mathbf{v}_k.$$

Thus,

$$d_1\mathbf{v}_1 + \ldots + d_s\mathbf{v}_s - (c_{s+1}\mathbf{v}_{s+1} + \ldots + c_l\mathbf{v}_k) = 0.$$

We recall that S is a basis for V, so all d_i and c_i are zero (linear independence). In particular, the c_i are all zero, implying $T(U)$ is linearly independent and hence completing the proof. □

Theorem 5.7. Let $T : V \to W$ be a linear transformation of vector space. Suppose V is a finite dimensional vector space. Then,

$$\mathrm{rank}(T) = \dim(V) - \mathrm{nullity}(T)$$

.

Proof. Since V is finite dimensional, so is $\ker(T)$ by Proposition 5.2. Therefore, we may select a basis $S' = \{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$for $\ker(T)$. By Theorem 5.2, there is a basis $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of V that contains S'. Then, $S = \{T(\mathbf{v}_{s+1}, \ldots, \mathbf{v}_k\}$ is a basis for $\mathrm{im}(T)$ (just like in the hypothesis of Proposition 5.3). By Proposition 5.2, we have that $\dim(V) = k \geqslant s = \dim(W)$. By Proposition 5.3, we have that $\{T(\mathbf{v}_{s+1}), \ldots, T(\mathbf{v}_k)\}$ is a basis for $\mathrm{im}(T)$. Therefore, $\mathrm{rank}(T) = k - s = \dim(V) - \mathrm{nullity}(T)$, as desired. □

Typically it is easier to calculate the nullity of a transformation than the rank of a transformation. But, the rank-nullity theorem allows us to calculate the rank of a transformation if we know the nullity. We see an example of this now.

Example 5.34. In Example 5.3 we saw that the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^2$

$$T(\mathbf{x}) = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 1 & 0 & 2 & 0 \end{bmatrix} \cdot \mathbf{x}.$$

has kernel

$$\ker(T) = \left\{ s \begin{bmatrix} -2 \\ \frac{-1}{2} \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \;\middle|\; s, t \in \mathbb{R} \right\}.$$

The dimension of $\ker(T)$ is 2 since begin $\begin{bmatrix} -2 \\ \frac{-1}{2} \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ are linearly independent and

$$\operatorname{Span}\left\{ \begin{bmatrix} -2 \\ \frac{-1}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\} = \ker(T)$$

.

By the Rank Nullity Theorem, we have that $\operatorname{rank}(T) = \dim(\mathbb{R}^4) - \operatorname{nullity}(T) = 4 - 2 = 2$. Therefore the dimension of the image of T (i.e $\operatorname{rank}(T)$) is 2.

5.4 Row Space, Column Space, and connections with with Rank-Nullity

In the last example of the last section, we applied rank-nullity to the linear tranformation $T(\mathbf{x}) = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 1 & 0 & 2 & 0 \end{bmatrix}$. We saw that $\operatorname{nullity}(T) = 2$ and $\operatorname{rank}(T) = 2$. We observe that the number of pivot columns of the matrix $\begin{bmatrix} 1 & 2 & 3 & 0 \\ 1 & 0 & 2 & 0 \end{bmatrix}$ is precisely the rank of T and the number of columns corresponding to free variables is the kernel of T. This is not a coincidence and is, in fact, a special version of the rank-nullity theorem.

Proposition 5.9. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by multiplication on the left by some $m \times n$ matrix A (that is $T(\mathbf{x}) = A\mathbf{x}$). Then,

1. $\operatorname{rank}(T) =$ number of pivot columns of A.
2. $\operatorname{nullity}(T) = n - \operatorname{rank}(T)$.

Proof. This is a consequence of the rank-nullity theorem. □

Soon we will see how to calculate rank and nullity of an linear transformation of abstract vector space $T : V \to W$ in a similar way. Until then, in particular on HW 7, we will calculate rank and nullity of linear transformations of abstract vector spaces by using just the definitions and the rank-nullity theorem. Let's put aside abstract vector spaces for a tiny bit and focus on linear transformations of the form $T : \mathbb{R}^n \to \mathbb{R}^m$.

Recall that any linear transformation of the form $T : \mathbb{R}^n \to \mathbb{R}^m$ can be realized as multiplication on the left by some $m \times n$ matrix. Proposition 5.4 tells us that to understand the rank and nullity of T, we need only understand the columns of A. For this reason, we make the following definition.

Column Space

Definition 5.21. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by left multiplication of the $m \times n$ matrix A. Suppose $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix}$. We set

$$\mathrm{Col}(T) = \mathrm{Col}(A) = \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$$

and call it the **Column Space of T (or A)**.

Example 5.35. Let $T : \mathbb{R}^3 \to \mathbb{R}^4$ be given by multiplication on the left by $A = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 2 & 3 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$.

The column space of T is

$$\mathrm{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Let's find the dimension of this space! To find it's dimension, we must find a basis for it. We do this by eliminating vectors that are not linearly independent from our spanning set. First, we calculate a REF of A to be

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

Any vector that corresponds to a pivot column we keep. Any vector that corresponds to a free variable, we toss out. After doing this, we see that

$$\mathrm{Col}(A) = \mathrm{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} = \mathrm{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

By our above work $\left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\} islinearlyindependent.Thus, \left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ is a basis for $\mathrm{Col}(A)$. Hence, $\dim(\mathrm{Col}(A)) = 3$.

Observation: By Proposition 5.4, $\mathrm{rank}(A) = 3$. Thus, $\mathrm{rank}(A) = \dim(\mathrm{Col}(A))$. This is not a coincidence!

Theorem 5.8. Let A be an $n \times m$ matrix. Then, $\mathrm{rank}(A) = \dim(\mathrm{Col}(A))$.

Proof. The proof of this follows along the same lines as the example above. Because of this, we omit it. $\square$

Proposition 5.10. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix}$ be an $n \times m$ matrix. Let B be the $n \times m$ matrix obtained by interchaning, any two $\mathbf{a}_i$, adding a multiple of one $\mathbf{a}_i$ to another $\mathbf{a}_j$, and scaling a $\mathbf{a}_i$ by some nonzero constant $k \in \mathbb{R}$ (these are called column operations). Then,

$$\mathrm{Col}(A) = \mathrm{Col}(B).$$

Proof. We prove this for when B is obtained by swappinng two columns. The other verifications are similar. We observe

$$\mathrm{Col}(A) = \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_m\} = \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_m\} = \mathrm{Col}(B).$$

□

Notice that Proposition 5.4 tells us that column operations do not affect the column space of a matrix. We might wonder if row operations change the column space of a matrix. The answer is yes!

Example 5.36. Consider the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. The column space of A is $\mathrm{Span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$. This is precisely the line $y = 0$. Now, adding row 1 to row 2 yields the matrix $B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. The column space of B is $\mathrm{Span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$, which is the precisely the line $y = x$. Thus, row operations definitely change the column space of a matrix.

While this is sad, we can turn things around (literally)!

Row Space

Definition 5.22. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by multiplication on the left by an $m \times n$ matrix A. Suppose $A = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}$. The set

$$\mathrm{Row}(T) = \mathrm{Row}(A) = \mathrm{Span}\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$$

is called the **row space of T (or A)**.

The following proposition allows us to manipulate the column space using row operations, just on a different, albeit related, matrix!

Proposition 5.11. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by multiplication on the left by an $m \times n$ matrix A. Then, $\mathrm{Col}(A) = \mathrm{Row}(A^T)$.

Proposition 5.12. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by multiplication on the left by an $m \times n$ matrix A. Let B be a matrix obtained by row reduction on A. Then, $\text{Row}(A) = \text{Row}(B)$.

Theorem 5.9. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by multiplication on the left by an $m \times n$ matrix A. Then $\text{Col}(A) = \text{Row}(REF(A^T)) = \text{Row}(REFF(A^T))$.

Proof. This is immediate by Propositions 5.4 and 5.4. □

The last proposition allows us to find a basis for the column space of a matrix quite easily! In fact, we can circumvent the method used in Example 5.4.

Example 5.37. Let $T : \mathbb{R}^3 \to \mathbb{R}^4$ be given by multiplication on the left by $A = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 2 & 3 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$.

The column space of T is

$$\text{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

We saw in Example 5.4 that $\left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ is a basis for $\text{Col}(A)$. Let's find another basis for this space using a method indicated by the theorem above. We row reduce $A^T = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & -2 & -4 \\ 1 & -1 & -5 \end{bmatrix}$.

We see that (an REF of A)

$$REF(A^T) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 24 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$RREF(A^T) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, as $\text{Row}(REF(A^T)) = \text{Row}(REFF(A^T)) = \text{Col}(A)$, the following are also bases for $\text{Col}(A)$

$$\left\{ \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 24 \end{bmatrix} \right\} \text{ and } \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Warning: Not all column spaces have such simple bases as in the example above. Sometimes free variables put in some extra entries so that we have other vectors besides the standard ones.

Summary: We have found two ways to determine a basis for the column space of a matrix (and also the row space since the row space is the column space of the transpose). Which basis that we get is better? The answer is: it depends on what you want. Sometimes we would like to figure out a basis for $\mathrm{Col}(A)$ by knocking out column vectors until you get a basis (see Example 5.4). This allows us to remember where we came from. On the flip side, sometimes all we want to know about the column space is its dimension. In this case, following Example 5.4 is quicker since we don't have to remember our original matrix. Personally I prefer the method of Exercise 5.4, but you should learn both methods since on exams/homework I will ask you to perform a specific method. Below we summarize the different algorithms to find bases for column and row spaces of a matrix

Subset of the Columns of A that forms a basis for $\mathrm{Col}(A)$

Algorithm 5.1. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$

1. Row reduce A.
2. Locate the pivot columns of A.
3. The pivot columns of A form a basis for $\mathrm{Col}(A)$.

A basis for $\mathrm{Col}(A)$ using A^T

Algorithm 5.2. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$

1. Write down A^T
2. Row reduce A^T.
3. The collection of rows with a leading 1 in A^T forms a basis for $\mathrm{Col}(A)$. **Warning:** this basis forgets the original columns of A.

Subset of the Rows of A that forms a basis for $\mathrm{Row}(A)$

Algorithm 5.3. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$

1. Write down A^T
2. Row reduce A^T.
3. Locate the pivot columns of A^T.
4. The pivot columns of A^T form a basis for $\mathrm{Row}(A)$.

A basis for $\mathrm{Row}(A)$ using A

Algorithm 5.4. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$

1. Row reduce A.
2. The collection of rows with a leading 1 in A forms a basis for $\mathrm{Row}(A)$. **Warning:** this basis forgets the original rows of A.

5.5 The Coordinate System with Respect to a Basis

In this section, we will describe how to realize a linear transformation of arbitary vector spaces as matrix multiplication. It is not immediate that we can do this; after all, vectors do not need to look line column vectors in $\mathbb{R}^n$. First, let's take a look at an example we have worked with many times. As we have mentioned before the vectors $\{\mathbf{e}_1, \mathbf{e}_2\}$ of standard unit vectors form a basis for $\mathbb{R}^2$. Visually, $\mathbf{e}_1$ spans the x-axis and $\mathbf{e}_2$ spans the y-axis. Since $\{\mathbf{e}_1, \mathbf{e}_2\}$ is a basis for $\mathbb{R}^2$, we can uniquely write every element $\mathbf{v} \in \mathbb{R}^2$ as $\mathbf{v} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2$, for some real numbers a_1 and a_2. We usually denote this by coordinate (a_1, a_2). Why not do this for any basis and for any vector space? All we used was the fact that $\{_1, \mathbf{e}_2\}$ is a basis. This leads us to the concept of *coordinate systems with respect to a basis*. First, we prove an important proposition that makes the notion of a coordinate system with respect to a basis, well defined.

Proposition 5.13. Let V be a vector space. Suppose $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is a basis for V. Then every element $\mathbf{v} \in V$ can **uniquely** be written as

$$\mathbf{v} = a_1\mathbf{w}_1 + \ldots + a_n\mathbf{w}_n,$$

for some $a_1, \ldots, a_n \in \mathbb{R}$

Proof. Let $\mathbf{v} \in V$. Then, since $\mathcal{B}$ is a basis for V, we have that $\mathcal{B}$ spans V. Thus, there exists $a_1, \ldots, a_n \in \mathbb{R}$ such that

$$\mathbf{v} = a_1\mathbf{w}_1 + \ldots + a_n\mathbf{w}_n.$$

All that is left to show is that the representation of $\mathbf{v}$ above is unique. Suppose that there exists a new set of number $\mathbf{b}_1, \ldots, \mathbf{b}_n$ such that

$$\mathbf{v} = b_1\mathbf{2_1} + \ldots \mathbf{b_n}\mathbf{w_n}.$$

We will show that $b_i = a_i$ for all i; this will show that the representation of $\mathbf{v}$ above is unique. We have

$$a_1\mathbf{w}_1 + \ldots + a_n\mathbf{w}_n = \mathbf{v} = b_1\mathbf{2_1} + \ldots \mathbf{b_n}\mathbf{w_n}.$$

Therefore,

$$(a_1 - b_1)\mathbf{w}_1 + \ldots + (a_n - b_n)\mathbf{w}_n = 0.$$

Since $\mathcal{B}$ is a basis for V, we have that $\mathcal{B}$ is linearly independent. As $(a_1 - b_1)\mathbf{w}_1 + \ldots + (a_n - b_n)\mathbf{w}_n = 0$, we have that $a_i - b_i = 0$ for all i. This implies that $a_i = b_i$ for all i, establishing the claim. □

$\mathcal{B}$-coordinate Vector

Definition 5.23. Let V be a vector space. Suppose that $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is a basis for V. For a vector $\mathbf{v} \in V$, by Proposition 5.5 we can uniquely write

$$\mathbf{v} = a_1\mathbf{w}_1 + \ldots a_n\mathbf{w}_n,$$

for some $a_1, \ldots, a_n \in \mathbb{R}$. We denote the $\mathcal{B}$-coordinate vector of v by $[\mathbf{v}]_{\mathcal{B}}$, and define it to be the $n \times 1$ column vector in $\mathbb{R}^n$

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Remark 12. *Proposition 5.5 is essential for Definition 5.5 to make sense. Take a moment to think why this is the case!*

Remark 13. *Let $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ be the standard basis for $\mathbb{R}^2$. The $\mathcal{B}$-coordinate vector of a vector in $\mathbb{R}^2$ is precisely its usually ordered pair representation as discussed in the introduction of this section.*

Example 5.38. Consider the basis $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ of $\mathbb{R}^n$ consisting of the standard unit vectors. Consider another basis of $\mathbb{R}^n$, $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ be an element of $\mathbb{R}^3$. We will find $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_{\mathcal{B}'}$. Do do, we find how to write $\mathbf{v}$ as a linear combination of the elements of $\mathcal{B}$ and then again in $\mathcal{B}'$. Luckily this is easy with the basis we have (we will look at a more difficult example shortly). We have

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Thus, $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. On the other hand

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Thus, $[\mathbf{v}]_{\mathcal{B}'} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. *Here we see that the choice of basis is important when describing $\mathcal{B}$-coordinates!*s

Example 5.39. Lets do a slightly more difficult example (difficult in the sense that finding basis representation for vectors won't be as easy). Consider the basis $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ of $\mathbb{R}^n$ consisting of the standard unit vectors. Consider another basis of $\mathbb{R}^n$, $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ be an element of $\mathbb{R}^3$. We will find $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_{\mathcal{B}'}$. Do do, we find how to write $\mathbf{v}$ as a linear combination of the elements of $\mathcal{B}$ and then again in $\mathcal{B}'$. Since,

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

Thus, $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Now, we want to find c_1, c_2, and c_3 so that

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

This is equivalent to finding the unique solution to the matrix equation

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Doing our usual Gaussian elimination we have a row equivalence

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

Therefore,

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + (-2) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Thus,

$$[\mathbf{v}]_{\mathcal{B}'} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}.$$

Example 5.40. The vector space $\mathbb{R}[x]_{\leqslant 3}$ has a basis $\mathcal{B} = \{1, x, x^2, x^3\}$. The polynomial $1 - 2x^2 + 7x^3$ has

$$[1 - 2x^2 + 7x^3]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ -2 \\ 7 \end{bmatrix}.$$

Example 5.41. We have seen that the vectors $f = \sin(x)$ and $g = \cos(x)$ in $C(\mathbb{R})$ are linear independent. Thus, $\mathcal{B} = \{f, g\}$ is a basis for $W = \text{Span} f, g$ (why is $\mathcal{B}$ not a basis for $C(\mathbb{R})$?). We have

$$[2\sin(x) - \pi \cdot cos(x)]_{\mathcal{B}} = \begin{bmatrix} 2 \\ \pi \end{bmatrix}$$

You might be thinking: why should we care about $\mathcal{B}$-coordinate vectors? As you may have realized, dealing with abstract vector spaces can be difficult- especially if we do not know what the vector space looks like. However, if we fix a basis $\mathcal{B}$ for our vector space, and consider $\mathcal{B}$-coordinates of vectors, we suddenly have a concrete way to manipulate vectors by realizing them as the $n \times 1$ vectors in $\mathbb{R}^n$ that we know and (hopefully) love. Doing this allows us to apply all of the things we have learned since the beginning of the semester to abstract vector spaces. The only catch is that this description of vectors $\mathbf{v} \in V$ depends heavily on the basis we choose. This complicates life a little bit, but the complication is well worth it; we will soon see how to deal with this complication in a very efficient and clean way. But, for now, lets explore $\mathcal{B}$-coordinate vectors a little more. First, we recall the notion of an isomorphism.

Definition 5.24. Let $T : V \to W$ be a linear transformation of vector spaces. If T is injective and surjective (i.e T is bijective), we say that T is an isomorphism.

Definition 5.25. Let V and W be vector spaces. If there is an isomorphism $T : V \to W$, we say that V is isomorphic to W

Exercise 5.7. *If V and W are isomorphic, is every linear transformation $T : V \to W$ an isomorphism.*

Exercise 5.8. *Let $T : V \to W$ be a linear transformation of vector spaces. Show that T is injective if and only if whenever $T(\boldsymbol{v}) = \boldsymbol{0}$, then $\boldsymbol{v} = \boldsymbol{0}$.*

Definition 5.26. Let V be a vector space with basis $\mathcal{B} = \{e_1, \ldots, e_n\}$. Define the coordinate map $C_{\mathcal{B}} : V \to \mathbb{R}^n$ by

$$C_{\mathcal{B}}(\mathbf{v}) = [\mathbf{v}]_{\mathcal{B}}.$$

The following theorem is absolutely amazing.

Theorem 5.10. Let V be an n-dimensional vector space with basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Then,

$$C_{\mathcal{B}} : V \to \mathbb{R}^n$$

is an isomorphism.

Proof. First we show that $C_{\mathcal{B}} : V \to \mathbb{R}^n$ is a linear transformation. To this end, suppose $\mathbf{v}, \mathbf{w} \in V$ and $r \in \mathbb{R}$. Let $\mathbf{v} = a_1\mathbf{e}_1 + \ldots + a_n\mathbf{e}_n$ and $\mathbf{v}' = b_1\mathbf{e}_1 + \ldots + b_n\mathbf{e}_n$. Then,

$$\mathbf{v} + \mathbf{w} = (a_1 + b_1)\mathbf{e}_1 + \ldots + (a_n + b_n)\mathbf{e}_n.$$

and

$$r\mathbf{v} = ra_1\mathbf{e}_1 + \ldots + ra_n\mathbf{e}_n.$$

By definition of $C_{\mathcal{B}}$, we have that $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$, $[\mathbf{w}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$,

$$[\mathbf{v} + \mathbf{w}]_{\mathcal{B}} = \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix}$$

and

$$[r\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} ra_1 \\ \vdots \\ ra_n \end{bmatrix}.$$

Therefore,

$$C_{\mathcal{B}}(\mathbf{v} + \mathbf{v}') = [\mathbf{v} + \mathbf{w}]_{\mathcal{B}} = [\mathbf{v}]_{\mathcal{B}} + [\mathbf{w}]_{\mathcal{B}} = C_{\mathcal{B}}(\mathbf{v}) + C_{\mathcal{B}}(\mathbf{w})$$

and

$$C_{\mathcal{B}}(r\mathbf{v} = [r\mathbf{v}]_{\mathcal{B}} = r[\mathbf{v}]_{\mathcal{B}} = rC_{\mathcal{B}}(\mathbf{v}).$$

Therefore, $C_{\mathcal{B}}$ is a linear transformation. We now show that $C_{\mathcal{B}}$ is a bijection. To see that $C_{\mathcal{B}}$ is injective suppose $C_{\mathcal{B}}(\mathbf{v}) = \mathcal{B}(\mathbf{w})$. Then, $[\mathbf{v}]_{\mathcal{B}} = [\mathbf{w}]_{\mathcal{B}}$. Let $\mathbf{v} = a_1\mathbf{e}_1 + \ldots + a_n\mathbf{e}_n$ and $\mathbf{v}' = b_1\mathbf{e}_1 + \ldots + b_n\mathbf{e}_n$. Then,

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

This implies $a_1 = b_1$, $a_2 = b_2$,..., $a_n = b_n$. Hence,

$$\mathbf{v} = a_1\mathbf{e}_1 + \ldots + a_n\mathbf{e}_n = b_1\mathbf{e}_1 + \ldots + b_n\mathbf{e}_n = \mathbf{w}.$$

Thus, $C_{\mathcal{B}}$ is injective. Now, we show that $C_{\mathcal{B}}$ is surjective. Suppose that $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$. Set $\mathbf{v} = c_1\mathbf{e}_1 + \ldots + c_n\mathbf{e}_n$, then

$$C_{\mathcal{B}}(\mathbf{v}) = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Hence, $C_{\mathcal{B}}$ is surjective. We conclude that $C_{\mathcal{B}}$ is an isomorphism. $\square$

Remark 14. *Theorem 5.5 says that an n-dimensional vector space is basically the same thing as $\mathbb{R}^n$. Of course, we should be careful with saying "the same". The vectors of V are technically different; however, we can associate the vectors of V to those of $\mathbb{R}^n$ that respects vector space structure. So, if all we want to understand is the vector space structure of a space, it suffices to study the vector space structure of $\mathbb{R}^n$. Sometimes, we do need to remember how V and $\mathbb{R}^n$ are isomorphic in order to translate properties of $\mathbb{R}^n$ to V.*

Remark 15. ***Warning:*** *The linear transformations $C_{\mathcal{B}}$ as in Theorem 5.5 is* ***dependent*** *on the basis $\mathcal{B}$ that we have chosen!*

The wonderful thing about the $C_{\mathcal{B}}$ is that they can be used to tell us how to realize a linear transformation as matrix multiplication. Let V be a vector space of dimension n have a basis $\mathcal{B}$ and W be a vector space of dimension m with basis $\mathcal{B}'$. Consider the following diagram

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, dashed, "S"] & \mathbb{R}^m \\
V \arrow[u, "C_{\mathcal{B}}"] \arrow[r, "T"'] & W \arrow[u, "C_{\mathcal{B}'}"']
\end{tikzcd}$$

We would like to construct a linear transformation $S : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$C_{\mathcal{B}'} \circ T = S \circ C_{\mathcal{B}}.$$

Often, when such a conditions holds, mathematicians say that the "the diagram commutes". In essence, this means that "S is essentially the same as T". Why do we want to this? We know that S can be realized as matrix multiplication. Thus, once we realize what S is, we can figure out how to realize T as matrix multiplication. Before we talk about how to do this in full generality, lets do it for a particular example.

Example 5.42. Let $V = \mathbb{R}[x]_{\leqslant 2}$, and $W = \mathbb{R}[x]_{\leqslant 1}$. Consider the bases $\mathcal{B} = \{1, x, x^2\}$ and $\mathcal{C} = \{1, x\}$ for V and W, respectively. Consider the function $\frac{d}{dx} : V \to W$ given by

$$\frac{d}{dx}(a_0 + a_1x + a_2x^2) = a_1 + 2a_2x.$$

This is a linear transformation (where have we seen this before?). Let's see how we can find a linear transformation $S : \mathbb{R}^3 \to \mathbb{R}^2$ that completes the following diagram.

$$\begin{tikzcd} \mathbb{R}^3 \arrow[r, dashed, "S"] & \mathbb{R}^2 \\ V \arrow[u, "C_{\mathcal{B}}"] \arrow[r, "\frac{d}{dx}"'] & W \arrow[u, "C_{\mathcal{C}}"'] \end{tikzcd}$$

First, we note that as $C_{\mathcal{B}}$ and $C_{\mathcal{C}}$ are isomorphisms, we have $C_{\mathcal{B}}(\mathcal{B}) = \{[1]_{\mathcal{B}}, [x]_{\mathcal{B}}, [x^2]_{\mathcal{B}}\} = \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$ is a basis for $\mathbb{R}^3$, and $C_{\mathcal{C}}(\mathcal{B}) = \{[1]_{\mathcal{B}}, [x]_{\mathcal{B}}\} = \left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$ is a basis for $\mathbb{R}^3$. Now, to describe $S : \mathbb{R}^3 \to \mathbb{R}^2$, we need only see describe what S does to the standard unit vectors of $\mathbb{R}^3$. Recall $C_{\mathcal{B}}(1) = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $C_{\mathcal{B}}(x) = \begin{bmatrix}0\\1\\0\end{bmatrix}$, and $C_{\mathcal{B}}(x^2) = \begin{bmatrix}0\\0\\1\end{bmatrix}$. In order for

$$C_{\mathcal{C}} \circ \frac{d}{dx} = S \circ C_{\mathcal{B}},$$

we must have $C_{\mathcal{C}} \circ \frac{d}{dx}(1) = S \circ C_{\mathcal{B}}(1)$, $C_{\mathcal{C}} \circ \frac{d}{dx}(x) = S \circ C_{\mathcal{B}}(x)$, and $C_{\mathcal{C}} \circ \frac{d}{dx}(x^2) = S \circ C_{\mathcal{B}}(x^2)$. Simplifying, we must have that $\begin{bmatrix}0\\0\end{bmatrix} = S\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right)$, $\begin{bmatrix}1\\0\end{bmatrix} = S\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right)$, and $\begin{bmatrix}0\\2\end{bmatrix} = S\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right)$. As we have seen before, S is matrix multiplication by

$$\begin{bmatrix}S(\mathbf{e}_1) & S(\mathbf{e}_2) & S(\mathbf{e}_3)\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\0 & 0 & 2\end{bmatrix}.$$

Hence, $S(\mathbf{v}) = \begin{bmatrix}0 & 1 & 0\\0 & 0 & 2\end{bmatrix} \cdot \mathbf{v}$.

Remark 16. *The above example tells us that the linear tranfomration $\frac{d}{dx} : V \to W$ can be thought as the linear transfomration $S : \mathbb{R}^3 \to \mathbb{R}^2$ given by multiplication by the matrix $\begin{bmatrix}0 & 1 & 0\\0 & 0 & 2\end{bmatrix}$. Thus, in order to understand $\frac{d}{dx}$, it suffices to understand $\begin{bmatrix}0 & 1 & 0\\0 & 0 & 2\end{bmatrix}$. We know how to handle the latter!*

There was nothing terribly special about the linear transformation above, and we can describe a linear transformation, in some sense, by matrix multiplication like we just did.

Algorithm 5.5. Let $T : V \to W$ be a linear transformation of vector spaces. Fix bases $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for V and $\mathcal{C} = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for W.
Draw the diagram

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, dashed, "S_{\mathcal{B},\mathcal{C}}"] & \mathbb{R}^k \\ V \arrow[u, "C_{\mathcal{B}}"] \arrow[r, "T"'] & W \arrow[u, "C_{\mathcal{C}}"'] \end{tikzcd}$$

We will determine how to find S to complete this diagram in the sense that $C_{\mathcal{B}'} \circ T = S_{\mathcal{B},\mathcal{C}} \circ C_{\mathcal{B}}$ (i.e that the diagram commutes).

1. Find $C_{\mathcal{C}}(T(\mathbf{v}_i))$ for all i.
2. Set

$$S_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} C_{\mathcal{C}}(T(\mathbf{v}_1)) & \cdots & C_{\mathcal{C}}(T(\mathbf{v}_n)) \end{bmatrix}$$

   We call $S_{\mathcal{B},\mathcal{C}}$ the matrix representing T with respect to $\mathcal{B}$ and $\mathcal{C}$.

Warning: This $S_{\mathcal{B},\mathcal{C}}$ that we have found depends heavily on our choice of bases for V and W. That is to say, that if we chose different bases, we would get a different matrix.

With this algorithm in hand, we can now understand linear transformations of abstract vector space by using methods we have developed for linear transformations of the form $S : \mathbb{R}^n \to \mathbb{R}^m$! We will have some practice with this in the homework. For now, we will use this principle to calculate the kernel of the map defined in Example 5.5.

Example 5.43. **??** Let $V = \mathbb{R}[x]_{\leqslant 2}$, and $W = \mathbb{R}[x]_{\leqslant 1}$. Consider the bases $\mathcal{B} = \{1, x, x^2\}$ and $\mathcal{C} = \{1, x\}$ for V and W, respectively. Consider the function $\frac{d}{dx} : V \to W$ given by

$$\frac{d}{dx}(a_0 + a_1 x + a_2 x^2) = a_1 + 2a_2 x.$$

We have seen that the matrix

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

makes the following diagram commute

$$\begin{tikzcd} \mathbb{R}^3 \arrow[r, dashed, "S"] & \mathbb{R}^2 \\ V \arrow[u, "C_{\mathcal{B}}"] \arrow[r, "\frac{d}{dx}"'] & W \arrow[u, "C_{\mathcal{C}}"'] \end{tikzcd}$$

We can find rank(T) and nullity(T) by finding rank(S) and nullity(S). Finding the latter things is easier since S is given by matrix multiplication, which we know how to handle very well! First, lets find the nullity of S; to do so, we find that that matrix $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ row reduces to

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

Therefore, $1 = \text{nullity}(S) = \text{nullity}(T)$. Hence, by rank-nullity

$$\text{rank}(T) = \dim(V) - \text{nullity}(T) = 3 - 1 = 2.$$

5.6 Base Change Matrices

Let us first consider the space $\mathbb{R}^n$. The standard basis $\mathcal{B}$ is for $\mathbb{R}^n$ is convenient for most problems; however, there are other times when we are working on a problem that is much easier solve with another basis $\mathcal{C}$. Since we usually view vectors, say $\mathbf{v}$, in $\mathbb{R}^n$ as its $\mathcal{B}$-coordinate, namely $[\mathbf{v}]_{\mathcal{B}}$, if we want to make things more convenient and work with $\mathcal{C}$, then we must find a convenient way to relate $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_{\mathcal{C}}$. Lets first try to tackle this through a specific example as this will help us work out the general case.

Example 5.44. Let $V = \mathbb{R}^3$ and consider the vector $\mathbf{v} = \begin{bmatrix}1\\2\\3\end{bmatrix}$. Let $\mathcal{B}$ be the standard basis of V and consider the basis $\mathcal{C} = \left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\3\end{bmatrix}\right\}$. We note that $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix}1\\2\\3\end{bmatrix}$. Let's find $[\mathbf{v}]_{\mathcal{C}}$. To do this we need to write $\mathbf{v}$ in terms of the basis $\mathcal{C}$. This is tantamount to solving the vector equation

$$\begin{bmatrix}1\\2\\3\end{bmatrix} = c_1 \begin{bmatrix}1\\2\\3\end{bmatrix} + c_2 \begin{bmatrix}1\\0\\0\end{bmatrix} + c_3 \begin{bmatrix}0\\0\\3\end{bmatrix}.$$

While we can solve this by inspection, for sake of illustration, we note that it suffices to do Gaussian Elimination on

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

Either way, we see that $c_1 = 1$, $c_2 = 0$, and $c_3 = 0$. Hence, $[\mathbf{v}]_{\mathcal{C}} = \begin{bmatrix}1\\0\\0\end{bmatrix}$.

For fun, let's see if we can come up with a systematic way to turn $[\mathbf{v}]_{\mathcal{B}}$ into $[\mathbf{v}]_{\mathcal{C}}$, for any $\mathbf{v} \in \mathbb{R}^3$. Suppose that $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}$. Then we would like to solve the vector equation

$$\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix} = c_1 \begin{bmatrix}1\\2\\3\end{bmatrix} + c_2 \begin{bmatrix}1\\0\\0\end{bmatrix} + c_3 \begin{bmatrix}0\\0\\3\end{bmatrix}.$$

Solving this vector equation is tantamount to applying Gaussian elimination to

$$\begin{bmatrix}1 & 1 & 0 & v_1\\2 & 0 & 0 & v_2\\3 & 0 & 3 & v_3\end{bmatrix}.$$

The RREF of this matrix is

$$\begin{bmatrix}1 & 0 & 0 & \frac{v_2}{2}\\0 & 1 & 0 & -\frac{v_2-2v_1}{2}\\0 & 0 & 1 & \frac{v_3}{3} - \frac{v_2}{2}\end{bmatrix}.$$

So, $[\mathbf{v}]_{\mathcal{C}} = \begin{bmatrix}\frac{v_2}{2}\\-\frac{v_2-2v_1}{2}\\\frac{v_3}{3} - \frac{v_2}{2}\end{bmatrix}$. While we do have a general answer, it can sometimes take a lot of work to get here. Below we discuss an easier, yet equivalent way of doing this.

$\mathbb{R}^n$-base change matrix from new to new

Algorithm 5.6. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be bases for $\mathbb{R}^n$ (not necessarily a standard basis here).

1. Write each $\mathbf{v}_i$ as a linear combination of the vectors in $\mathcal{C}$.
2. Use the above step to find $[\mathbf{v}_i]_{\mathcal{C}}$.
3. Construct the matrix

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix}[\mathbf{v}_1]_{\mathcal{C}} & \cdots & [\mathbf{v}_n]_{\mathcal{C}}\end{bmatrix}.$$

Then, for every $v \in \mathbb{R}^n$, we have

$$P_{\mathcal{B}\to\mathcal{C}} \cdot [\mathbf{v}]_{\mathcal{B}} = [\mathbf{v}]_{\mathcal{C}}.$$

We call $P_{\mathcal{B}\to\mathcal{C}}$ the base change matrix from $\mathcal{B}$ to $\mathcal{C}$. .

Warning: $P_{\mathcal{B}\to\mathcal{C}}$ and $P_{\mathcal{C}\to\mathcal{B}}$ are usually **not the same**.

Example 5.45. Let $\mathbf{b}_1 = \begin{bmatrix}-9\\1\end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix}-5\\-1\end{bmatrix}$, $\mathbf{c}_1 = \begin{bmatrix}1\\-4\end{bmatrix}$, and $\mathbf{c}_2 = \begin{bmatrix}3\\-5\end{bmatrix}$. Consider the bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$. We will find $P_{\mathcal{B}\to\mathcal{C}}$ and $P_{\mathcal{C}\to\mathcal{B}}$. We will use our algorithm. First, we find that

$$\mathbf{b}_1 = 6\mathbf{c}_1 - 5\mathbf{c}_2 \text{ and } \mathbf{b}_2 = 4\mathbf{c}_1 - 3\mathbf{c}_2.$$

Thus,

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix}6 & 4\\-5 & -3\end{bmatrix}.$$

Similarly, one can compute

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix}-3/2 & -2\\5/3 & 3\end{bmatrix}.$$

L

Example 5.46. t $\mathbf{b}_1 = \begin{bmatrix}1\\-3\end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix}-2\\4\end{bmatrix}$, $\mathbf{c}_1 = \begin{bmatrix}-7\\9\end{bmatrix}$, and $\mathbf{c}_2 = \begin{bmatrix}-5\\7\end{bmatrix}$. Consider the bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$. We will find $P_{\mathcal{B}\to\mathcal{C}}$ and $P_{\mathcal{C}\to\mathcal{B}}$. Using our algorithm, we find that

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix}2 & -3/2\\-3 & 5/2\end{bmatrix}$$

and

$$P_{\mathcal{C}\to\mathcal{B}} = \begin{bmatrix}5 & 3\\6 & 4\end{bmatrix}$$

inverse

Proposition 5.14. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be bases for $\mathbb{R}^n$. Then

$$P_{\mathcal{C}\to\mathcal{B}} = P_{\mathcal{B}\to\mathcal{C}}^{-1}.$$

L

Example 5.47. t $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$, $\mathbf{c}_1 = \begin{bmatrix} -7 \\ 9 \end{bmatrix}$, and $\mathbf{c}_2 = \begin{bmatrix} -5 \\ 7 \end{bmatrix}$. Consider the bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} + \{\mathbf{c}_1,$
$bfc_2\}$. We saw that

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix} 2 & -3/2 \\ -3 & 5/2 \end{bmatrix}$$

and

$$P_{\mathcal{C}\to\mathcal{B}} = \begin{bmatrix} 5 & 3 \\ 6 & 4 \end{bmatrix}$$

After a quick verification, we see that

$$P_{\mathcal{C}\to\mathcal{B}} = P_{\mathcal{B}\to\mathcal{C}}^{-1}.$$

Theorem 5.11. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be bases for $\mathbb{R}^n$. Then, we have row equivalences:

$$\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n & | & \mathbf{w}_1 & \cdots & \mathbf{w}_n\end{bmatrix} \sim\sim\sim \begin{bmatrix}I_n \mid P_{\mathcal{C}\to\mathcal{B}}\end{bmatrix}$$

and

$$\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_n & | & \mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} \sim\sim\sim \begin{bmatrix}I_n \mid P_{\mathcal{B}\to\mathcal{C}}\end{bmatrix}.$$

Caution: beware of the seemingly odd order of things.

Proof. We prove just one of these, namley the second one; the other will follow similarly. To find $P_{\mathcal{B}\to\mathcal{C}}$ we need to solve the equations

$$\begin{aligned}\mathbf{v}_1 =& c_{1,1}\mathbf{w}_1 + \cdots + c_{1,n}\mathbf{w}_n \\ \vdots & \qquad\qquad\qquad\qquad . \\ \mathbf{v}_n =& c_{n,1}\mathbf{w}_1 + \cdots + c_{n,n}\mathbf{w}_n\end{aligned}$$

To solve these equations, we row reduce the following augmented matrices

$$\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_n & | & \mathbf{v}_1\end{bmatrix}$$
$$\vdots$$
$$\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_n & | & \mathbf{v}_n\end{bmatrix}.$$

Rather than row reducing these one at a time, we can reduce

$$\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_n & | & \mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}.$$

Since $\mathcal{C}$ is a basis for W, it's vectors are linearly independent. Hence, by the Inverses in Disguises Theorem, there is a row reduction:

$$\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_n & | & \mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} \sim\sim\sim \begin{bmatrix}I_n \mid B\end{bmatrix}.$$

Here, B is precisely $P_{\mathcal{B}\to\mathcal{C}}$ by design. $\square$

Remark 17. *You can use any method you want to calculate these base-change matrices. I personally prefer attacking the equations one at a time, but I cede that using the method as illustrated in the above theorem is more computationally effective. Find out which one you like better!*

L

Example 5.48. t's convince ourselves that this theorem really works! Let $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$, $\mathbf{c}_1 = \begin{bmatrix} -7 \\ 9 \end{bmatrix}$, and $\mathbf{c}_2 = \begin{bmatrix} -5 \\ 7 \end{bmatrix}$. Consider the bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$. We will find $P_{\mathcal{B}\to\mathcal{C}}$ and $P_{\mathcal{C}\to\mathcal{B}}$. Using Theorem **??**, we row reduce

$$\begin{bmatrix} -7 & -5 & | & 1 & -2 \\ 9 & 7 & | & -3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & | & 2 & -3/2 \\ 0 & 1 & | & -3 & 5/2 \end{bmatrix}$$

So,

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix} 2 & -3/2 \\ -3 & 5/2 \end{bmatrix}$$

as verified previously.

There was nothing terribly special about doing base change matrices with $V = \mathbb{R}^n$. We can do exactly the same procedures to procure base change matrices for abstract vector spaces and their bases!

V **base change matrix from new to new**

Algorithm 5.7. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be bases for a vector space V.

1. Write each $\mathbf{v}_i$ as a linear combination of the vectors in $\mathcal{C}$.
2. Use the above step to find $[\mathbf{v}_i]_{\mathcal{C}}$.
3. Construct the matrix

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix} [\mathbf{v}_1]_{\mathcal{C}} & \cdots & [\mathbf{v}_n]_{\mathcal{C}} \end{bmatrix}.$$

   Then, for every $v \in \mathbb{R}^n$, we have

$$P_{\mathcal{B}\to\mathcal{C}} \cdot [\mathbf{v}]_{\mathcal{B}} = [\mathbf{v}]_{\mathcal{C}}.$$

   We call $P_{\mathcal{B}\to\mathcal{C}}$ the base change matrix from $\mathcal{B}$ to $\mathcal{C}$. .

A more abstract example

Example 5.49. Let $V = \mathbb{R}[x]_{\leqslant 2}$ and consider the following bases for V: $\mathcal{B} = \{1, x, x^2\}$ and $\mathcal{C} = \{1 - x, x^2 + 1, 3\}$. We will find $P_{\mathcal{B}\to\mathcal{C}}$ and $P_{\mathcal{C}\to\mathcal{B}}$. We first calculate $P_{\mathcal{B}\to\mathcal{C}}$. To this end:

$$1 = 0(1 - x) + 0(x^2 + 1) + \frac{1}{3} \cdot 3$$

$$x = -1(1 - x) + 0(x^2 + 1) + \frac{1}{3} \cdot 3$$

and

$$x^2 = 0(1 - x) + 1(x^2 + 1) - \frac{1}{3} \cdot 3.$$

Therefore,

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \end{bmatrix}.$$

Calculating the inverse of $P_{\mathcal{B}\to\mathcal{C}}$ yields

$$P_{\mathcal{C}\to\mathcal{B}} = (P_{\mathcal{B}\to\mathcal{C}})^{-1} = \begin{bmatrix} 1 & 1 & 3 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Chapter 6

Eigenvalues and Eigenvectors

6.1 Eigenvalues and Eigenvectors of Matrices

Recall Homework 5 problem 4 (included here for your convinence):

HW 5 4 Let $A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$,, and $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

1. Find all vectors in $\mathbb{R}^3$ that are fixed by A_1. In other words, find all $\mathbf{x} \in \mathbb{R}^3$ such that $A_1\mathbf{x} = \mathbf{x}$

2. Find all vectors in $\mathbb{R}^3$ that are fixed by A_2. In other words, find all $\mathbf{x} \in \mathbb{R}^3$ such that $A_2\mathbf{x} = \mathbf{x}$

3. Find all vectors in $\mathbb{R}^3$ that are fixed by A_3. In other words, find all $\mathbf{x} \in \mathbb{R}^3$ such that $A_3\mathbf{x} = \mathbf{x}$

4. Use parts (a) through (c) to find all vectors in $\mathbb{R}^3$ that are fixed by A_1, A_2, A_3.

Here we found particular vectors that satisfied $A_1\mathbf{x} = \mathbf{x}$, $A_2\mathbf{x} = \mathbf{x}$ and $A_3\mathbf{x} = \mathbf{x}$. These vectors were special in that the matrices A_1, A_2 and A_3 kept them the same. Unfortunately in general, matrices do not keep **ANY nonzero** vector the same.

Example 6.1. Let $A = \begin{bmatrix} 3 & -2 \\ 0 & 3 \end{bmatrix}$. We will show that no vector $\mathbf{b} = (b_1, b_2)$ is fixed by the matrix. That is to say, we will show that $A\mathbf{b} = \mathbf{b}$ has no nontrivial solution. This is equivalent to showing that $(A - I)\mathbf{b} = 0$. Note that

$$A - I = \begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix}.$$

Now, since this matrix is invertible $(A - I)\mathbf{b} = 0$ has no nonzero solution. Therefore A does not fix any nonzero vector.

While the above example is perhaps sad, we can look at the next best thing! Instead of hoping that a square matrix fixes a vector, we can figure out when a matrix scales a vector by a constant. That is to say: given a square matrix A, for which $\lambda \in \mathbb{R}$ does there exist a $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$? It turns out matrices always have these, so we give them a name.

Eigenvalues and Eigenvectors of a Matrix

Definition 6.1. Let A be an $n \times n$ matrix. We say that $\lambda \in \mathbb{R}$ is an eigenvalue of A if there is a **nonzero** vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. In this case we say that $\mathbf{x}$ is an **eigenvector** of A associated to λ.

Remark 18. *The word "eigen" comes from the German language. It means "self".*

Example 6.2. Let $A = \begin{bmatrix} 3 & -2 \\ 0 & 3 \end{bmatrix}$, like in the previous example. Let's see if we can find some eigenvectors for A and their corresponding eigenvalues. Consider the vector $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$. We see that

$$\begin{bmatrix} 3 & -2 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} = 3\begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Therefore, $(2, 0)$ is an eigenvector of A with eigenvalue 3.

Example 6.3. Let $A = \begin{bmatrix} 3 & -2 \\ 0 & 3 \end{bmatrix}$, like in the previous example. We will show that $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is not an eigenvector. We see that

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

Since $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ is not a multiple of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, the latter is not an eigenvector.

Example 6.4. Let $A = \begin{bmatrix} 3 & -2 \\ 0 & 3 \end{bmatrix}$, like in the previous example. We will show that 2 is not an eigenvalue for A. To determine if 2 is an eigenvalue for A, we must see if

$$A\mathbf{x} = 2\mathbf{x}$$

has a nontrivial solution. This is tantamount to seeing if

$$(A - 2I)\mathbf{x} = 0$$

has a nontrivial solution. Since $A - 2I = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$, we have that $A - 2I$ is invertible, so

$$(A - 2I)\mathbf{x} = 0$$

has no nontrivial solution. Therefore, 2 is not an eigenvalue for A.

Guessing and checking eigenvalues/eigenvectors is a very inefficient way of going about things. The following Theorem will give us a way to search for eigenvalues (and hence eigenvectors) is a better way.

Theorem 6.1. Let A be an $n \times n$ matrix. The following are equivalent

(a) $\lambda \in \mathbb{R}$ is an eigenvalue for A

(b) The equation $(A - \lambda I)\mathbf{x} = 0$ has a nontrivial solution

(c) $\det(A - \lambda I) = 0$.

Proof. Rather than give a technical proof, we will point out that $\det(A - \lambda I) = 0$ implies that $(A - \lambda I)\mathbf{x} = 0$ has a nontrivial solution by the Inverses in Disguise Theorem. □

Example 6.5. Let $A = \begin{bmatrix} 3 & -2 \\ 0 & 3 \end{bmatrix}$, like in the previous example. We will find all eigenvalues of A. To do so, we need to figure out which $\lambda \in \mathbb{R}$ satsify

$$\det(A - \lambda I) = 0$$

Now, $A - \lambda I = \begin{bmatrix} 3-\lambda & -2 \\ 0 & 3-\lambda \end{bmatrix}$. Thus, $\det(A - \lambda I) = 0$ if and only if $\det\left(\begin{bmatrix} 3-\lambda & -2 \\ 0 & 3-\lambda \end{bmatrix}\right) = 0$. Now

$$\det\left(\begin{bmatrix} 3-\lambda & -2 \\ 0 & 3-\lambda \end{bmatrix}\right) = (3-\lambda)(3-\lambda) = 0$$

if and only if $\lambda = 3$. Thus, 3 is the only eigenvalue of A.

As we can see, the condition that $\det(A - \lambda I) = 0$ is particularly useful, so lets give it a name!

The Characteristic Equation

Definition 6.2. Let A be an $n \times n$ matrix. The characteristic equation of A is

$$\det(A - \lambda I) = 0$$

Remark 19. *The characteristic equation of a (square) matrix is always a polynomial with variable λ. In fact the degree of the characteristic equation of a matrix is equal to the size of the matrix!*

Example 6.6. Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 3 & 3 \end{bmatrix}$. We will use the characteristic equation of A to find the eigenvalues of A. To this end, we have

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

By cofactor expansion along the first column, we have that

$$\det(A - \lambda I) = (1-\lambda)((2-\lambda)(3-\lambda) - 6).$$

Now, $\det(A - \lambda) = 0$ if and only if $\lambda = 1$ or $(2-\lambda)(3-\lambda) - 6 = 0$. The latter happens if and only if $\lambda = 0, 5$. Thus, the eigenvalues of A are $1, 0$, and 5.

Proposition 6.1. Let A be an upper or lower triangular matrix. Then, the eigenvalues of A are precisely the entries along the diagonal of A.

Proof. If A is an upper or lower triangluar matrix, then so is $A - \lambda I$. Therefore, $\det(A - \lambda I) = (a_{1,1} - \lambda) \cdots (a_{n,n} - \lambda) = 0$ if and only if $\lambda = a_{1,1}, \ldots, a_{n,n}$, as desired. □

Example 6.7. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 6 \\ 0 & 0 & 7 \end{bmatrix}$. Then the eigenvalues of A are $1, 4$ and 7.

Theorem 6.2. A matrix A is invertible if and only if 0 is not an eigenvector for A.

Proof. A is invertible if and only if $(A - 0I) = A$ is invertible. Hence A is invertible if and only if $\det(A - 0I) = \det(A) \neq 0$. Thus, A is invertible if and only if 0 is not an eigenvalue of A. □

Eigenspaces

Definition 6.3. Let A be an square matrix of size n. Suppose that $\lambda \in \mathbb{R}$ is a eigenvalue of A. We call the collection of all eigenvectors of A associated to λ the eigenspace of A associated to λ. We will denote this by $E_\lambda(A)$

Proposition 6.2. Let A be an square matrix of size n. Suppose that $\lambda \in \mathbb{R}$ is a eigenvalue of A. Then, the eigenspace of A associated to λ is a subspace of $\mathbb{R}^n$.

Proof. This is a quick application of the subspace criteria. Certainly, $0 \in E_\lambda(A)$ since $A0 = \lambda 0$. Now, suppose that $\mathbf{x}, \mathbf{y} \in E_\lambda(A)$, then by definition, we have that $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \lambda\mathbf{y}$. Then,

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \lambda\mathbf{x} + \lambda\mathbf{y} = \lambda(\mathbf{x} + \mathbf{y}).$$

Hence, $\mathbf{x} + \mathbf{y} \in E_\lambda(A)$. Lastly, assume that $x \in E_\lambda(A)$ and $k \in \mathbb{R}$. Then,

$$A(k\mathbf{x}) = kA\mathbf{x} = k\lambda\mathbf{x} = (k\mathbf{x}),$$

so $k\mathbf{x} \in E_\lambda(A)$. □

Example 6.8. Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 3 & 3 \end{bmatrix}$ as in the example above. We saw that 0 is an eigenvalue of A. Let's describe $E_0(A)$ concretely. To find $E_0(A)$, we need to find all eigenvectors associated to 0 of A. Such vectors must satisfy

$$0 = (A - 0I)\mathbf{x} = A\mathbf{x}.$$

Hey! The set of such vectors is precisely the kernel of A. Using row operations on $\begin{bmatrix} A & | & 0 \end{bmatrix}$, we find the kernel of A to be

$$\{(0, s, s) = s(0, 1, 1) \mid s \in \mathbb{R}\} = E_0(A).$$

Next, lets find $E_1(A)$. To find $E_1(A)$, we need to figure out which $\mathbf{x}$ satisfy

$$(A - I)\mathbf{x} = 0.$$

Hey! This is the kernel of $(A - I)$. Using row operations on $\begin{bmatrix} A - I & | & 0 \end{bmatrix}$ tells us $\ker(A - I)$ is

$$\{s(1, 0, 0) \mid s \in \mathbb{R}\} = E_1(A).$$

The above example leads us the the following propostion.

Proposition 6.3. Let A be a square matrix with eigenvalue λ. Then,

$$E_\lambda(A) = \ker(A - \lambda I).$$

The following theorem is super cool! We won't prove it since I don't think the proof is super enlightening. However, we should think of the next theorem as telling us that eigenvectors corresponding to different eigenvalues are linearly independent (and hence very different from on another).

Theorem 6.3. Let A be a square matrix of size p with distint eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$, respectively. Then the set

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$

is a linearly independent subset of $\mathbb{R}^p$

6.2 Eigenvalues and Eigenvectors of Linear Transformations

There is nothing stopping us in defining eigenvalues and eigenvectors of an abstract linear transformation! The only difference is that we don't have a matrix to work with, at least at first glance. In this section, we will talk about how to find eigenvalues and eigenvectors of linear transformations of abstract vector spaces. Note: the definition that follows is more or less the same as the once in the previous section!

Eigenvalues and Eigenvectors of a Linear Transformation

Definition 6.4. Let V be a vector space and $T : V \to V$ be a linear transformation. An eigenvalue $\lambda \in \mathbb{R}$ for T is a real number such that there exists a **nonzero** $\mathbf{v} \in V$ satisfying

$$T(\mathbf{v}) = \lambda\mathbf{v}.$$

We call the vector $\mathbf{v}$ an eigenvector associated to λ.

Finding eigenvalues and eigenvectors of an abstract linear transformation $T : V \to V$ can be intimidating since what we are working with is likely not $\mathbb{R}^n$. However, as we saw in chapter 5, we can always take an abstract linear transformation and lift it to one that looks like a matrix! We will do this to find eigenvalues and eigenvectors for $T : V \to V$, where V is any vector space. First recall Algorithm 6.2, copied here for your convenience.

Algorithm 6.1. Let $T : V \to W$ be a linear transformation of vector spaces. Fix bases $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for V and $\mathcal{C} = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for W.
Draw the diagram

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, dashed, "S_{\mathcal{B},\mathcal{C}}"] & \mathbb{R}^k \\ V \arrow[u, "C_{\mathcal{B}}"] \arrow[r, "T"'] & W \arrow[u, "C_{\mathcal{C}}"'] \end{tikzcd}$$

We will determine how to find S to complete this diagram in the sense that $C_{\mathcal{B}'} \circ T = S_{\mathcal{B},\mathcal{C}} \circ C_{\mathcal{B}}$ (i.e that the diagram commutes).

1. Find $C_{\mathcal{C}}(T(\mathbf{v}_i))$ for all i.
2. Set

$$S_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} C_{\mathcal{C}}(T(\mathbf{v}_1)) & \cdots & C_{\mathcal{C}}(T(\mathbf{v}_n)) \end{bmatrix}$$

   We call $S_{\mathcal{B},\mathcal{C}}$ the matrix representing T with respect to $\mathcal{B}$ and $\mathcal{C}$.

Warning: This $S_{\mathcal{B},\mathcal{C}}$ that we have found depends heavily on our choice of bases for V and W. That is to say, that if we chose different bases, we would get a different matrix.

We will use Algorithm 6.2 to find eigenvalues and eigenvectors of abstract linear transformations, as described in next theorem.

Theorem 6.4. Let V be a vector space of dimension n and $T : V \to V$ be a linear transformation. Let $\mathcal{B}$ be a basis for V and $S_{\mathcal{B},\mathcal{B}} : \mathbb{R}^n \to \mathbb{R}^n$ the matrix representing T with respect to $\mathcal{B}$ and $\mathcal{B}$. Then, $\lambda \in \mathbb{R}$ is an eigenvalue for T if and only if λ is an eigenvalue for $S_{\mathcal{B},\mathcal{B}}$. In other words the eigenvalues for T and $S_{\mathcal{B},\mathcal{B}}$ are exactly the same.

Proof. We omit a proof of this fact. $\square$

Remark 20. *Theorem 6.2 says that the eigenvalues of a matrix are independent of the basis we choose for V. So, no matter which basis for V we choose, we will get the same eigenvalues.*

Warning: Theorem 6.2 does not say that the eigenvectors of $S_{\mathcal{B},\mathcal{B}}$ are exactly the same as the eigenvectors of T. This is because $S_{\mathcal{B},\mathcal{B}}$ and T has different domains. Though, one can use the coordinate maps to find out what the eigenvalues of T with respect to a particular eigenvalue are; we won't do this. If you are interested, I am more than happy to talk about it with you!

Let's practice Theorem 6.2!

Example 6.9. Let $V = \mathbb{R}[x]_{\leqslant 2}$. Consider the basis $\mathcal{B} = \{1, x, x^2\}$ for V. Consider the function $\frac{d}{dx} : V \to W$ given by

$$\frac{d}{dx}(a_0 + a_1 x + a_2 x^2) = a_1 + 2a_2 x.$$

The matrix representation of $\frac{d}{dx}$ relative to $\mathcal{B}$ is:

$$S_{\mathcal{B},\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

The eigenvalues of the matrix $S_{\mathcal{B},\mathcal{B}}$ are 0, 1 and 2.

You might be wondering, what happens if we choose a different basis for V in the above example? Nothing changes except for the matrix we get! The eigenvalues will still be the same- let's check this.

Example 6.10. Let $V = \mathbb{R}[x]_{\leqslant 2}$. Consider the basis $\mathcal{B} = \{1, x-1, x^2-1\}$ for V. Consider the function $\frac{d}{dx} : V \to W$ given by

$$\frac{d}{dx}(a_0 + a_1 x + a_2 x^2) = a_1 + 2a_2 x.$$

We will find the matrix representation of $\frac{d}{dx}$ relative to $\mathcal{B}$. First, $T(1) = 0$, $T(x) = 1$ and $T(x^2) = 2x = 2 \cdot 1 + 2(x-1)$. Hence,

$$S_{\mathcal{B},\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}.$$

The eigenvalues of the matrix $S_{\mathcal{B},\mathcal{B}}$ are 0, 1 and 2 (which can be seen since it is an upper triangular matrix).

The following example is for fun and a connection to differential equations! It shows that weird things happen with infinite dimensional vector space (where our theorem does not apply)!

Example 6.11. Let $V = \text{Diff}(\mathbb{R})$ and let $D_x : V \to V$ be the derivative transformation (that is $D_x(f) = \frac{df}{dx}$). Let's find its eigenvalues and some eigenvectors! Let $\lambda \in \mathbb{R}$ be ANY real number, we will show that λ is an eigenvalue of D_x and λe^x is an eigenvector associated to λ. Indeed,

$$D_x(\lambda e^x) = \lambda e^x.$$

So, D_x has an infinite number of eigenvalues!

6.3 Diagonalization and Similarity of Matrices

Motivation for Similarity: In the previous section, we saw that to understand linear transformations of abstract vector spaces and their eigenvalues, one must complete a diagram like so:

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, dashed, "S_{\mathcal{B},\mathcal{B}}"] & \mathbb{R}^n \\
V \arrow[u, "C_{\mathcal{B}}"] \arrow[r, "T"'] & V \arrow[u, "C_{\mathcal{B}}"']
\end{tikzcd}$$

where $T : V \to V$ is a linear transformation of vector spaces and $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a fixed basis for V. In other words $C_{\mathcal{B}} \circ S_{\mathcal{B},\mathcal{B}} = C_{\mathcal{B}} \circ T$. Since $C_{\mathcal{B}}$ is an isomorphism (we showed this before), it follows that it has an inverse $C_{\mathcal{B}}^{-1}$ **(a function f has an inverse, which we denote f^{-1}, if $f \circ f^{-1} = 1$ and $f^{-1} \circ f = 1$.)** Thus,

$$T = C_{\mathcal{B}} \circ S_{\mathcal{B},\mathcal{B}} \circ C_{\mathcal{B}}^{-1}.$$

From this equation, it appears that "T is similar to $S_{\mathcal{B},\mathcal{B}}$". This motivates us to make the following definition

Definition 6.5. We say that two $n \times n$ matrices A and B are similar if there is an invertible $n \times n$ matrix P such that

$$B = PAP^{-1}$$

For ease of notation if A and B are similar, we will write $A \sim B$.

Proposition 6.4. Let A and B be $n \times n$ matrices. Then, $A \sim B$ if and only if $B \sim A$.

Proof. If $A \sim B$, then there is an invertible matrix P such that $B = PAP^{-1}$. This implies that $P^{-1}BP = A$, so $B \sim A$. The proof of the reverse direction is analogous. □

Let's check out a few examples!

Example 6.12. The matrices $\begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}$ and $\begin{bmatrix}-3 & 2\\-8 & 5\end{bmatrix}$ are similar since

$$\begin{bmatrix}1 & 0\\2 & 1\end{bmatrix} 1201 \begin{bmatrix}1 & 0\\-2 & 1\end{bmatrix} = \begin{bmatrix}-3 & 2\\-8 & 5\end{bmatrix}$$

Example 6.13. The matrices $\begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}$ and $\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$ are not similar. I leave it to you as an exercise to check this.

There is a very nice test to see whether two matrices are similar:

Proposition 6.5. Let A and B be two $n \times n$ matrices.

1. If $A \sim B$, then $\det(A) = \det(B)$.
2. If $A \sim B$, then A and B have the same characteristic polynomial.

Proof. First we prove part (a). If $A \sim B$, then there is an invertible P such that $PAP^{-1} = B$; thus, $\det(A) = \det(P)\det(A)\det(P^{-1}) = \det(PAP^{-1}) = \det(B)$.

For the proof of B, suppose that $PAP^{-1} = B$. Then,

$$\det(B - \lambda I) = \det(PAP^{-1} - P\lambda IP^{-1}) = \det(P(A - \lambda I)P^{-1}) = \det(A - \lambda I).$$

□

Warning: If $\det(A) = \det(B)$ or A and B have the same characteristic equation, this does not imply that $A \sim B$. For example, as we have seen The matrices $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are not similar, yet they have the same characteristic polynomial and the same determinate.

6.3.1 Diagonalization

What is your favorite type of square matrix? I don't know about you, but I think diagonal matrices are pretty nice! For example:

Example 6.14. Let $A = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}$, where α and β are real numbers. Then $A^t = \begin{bmatrix} \alpha^t & 0 \\ 0 & \beta^t \end{bmatrix}$, for all $t \in \mathbb{N}$. For any other non-diagonal matrix, we may not have such a nice way to calculate it's power.

Hopefully, the last example convinces you that diagonal matrices are computationally friendly; if not, then the next proposition should!

Proposition 6.6. If $D = \begin{bmatrix} \alpha_1 & 0 & 0 & \cdots & 0 \\ 0 & \alpha_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & \\ 0 & 0 & 0 \cdots \alpha_n & & \end{bmatrix}$ is a diagonal matrix, then

$$D^k = \begin{bmatrix} \alpha_1^k & 0 & 0 & \cdots & 0 \\ 0 & \alpha_2^k & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & 0 \\ 0 & 0 & 0 & \cdots & \alpha_n^k \end{bmatrix}.$$

Now, we might ask, if a matrix that is similar to diagonal matrix can have its powers easily computed. The answer is yes!

Proposition 6.7. Let $A \sim D$, where D is a diagonal matrix. Say, $PDP^{-1} = A$. Then

$$A^k = PD^kP^{-1}.$$

Proof. Since $A = PDP^{-1}$, we have

$$A^k = \underbrace{(PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})}_{k-\text{times}} = PD^kP^{-1}$$

□

Definition 6.6. We say that a matrix A is diagonalizable if it is similar to a diagonal matrix.

Proposition 6.8. An $n \times n$ matrix A is diagonalizable if and only if it has n linear independent eigenvectors.

Proof. Before proving this proposition, we observe that if P is an invertible matrix, A is a matrix, and D is a diagonal matrix, then

$$AP = A\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} = \begin{bmatrix}A\mathbf{v}_1 & \cdots & A\mathbf{v}_n\end{bmatrix}, \tag{6.1}$$

and

$$PD = P\begin{bmatrix}\alpha_1 & 0 & 0 & \cdots & 0\\ 0 & \alpha_2 & 0 & \cdots & 0\\ 0 & 0 & \ddots & 0 & \\ 0 & 0 & 0 & \cdots & \alpha_n\end{bmatrix} = \begin{bmatrix}\alpha_1\mathbf{v}_1 & \cdots & \alpha_n\mathbf{v}_n\end{bmatrix}. \tag{6.2}$$

Assume that $A = PDP^{-1}$. Then Equations 6.1 and 6.2 imply that $A\mathbf{v}_i = \alpha_i\mathbf{v}_i$ for all i. Since P is invertible, its columns must be linearly independent. Hence, the $\mathbf{v}_i$ are linearly independent eigenvectors.

Now, assume that A has n linearly indpendent eigenvectors, with

$$A\mathbf{v}_1 = \alpha_1\mathbf{v}_1, \ldots, A\mathbf{v}_n = \alpha_n\mathbf{v}_n.$$

Set $P = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$. Then, reversing Equations 6.1 and 6.2 and noting the previous string of equations, we have that

$$A = PDP^{-1}.$$

So, A is diagonalizable.

□

Proposition 6.9. If an $n \times n$ matrix A has n eigenvalues, then A is diagonalizable.

Proof. This is a direct consequence of the previous theorem. □

Lets talk about how to determine if a matrix is diagonalizable, and if it is how to find a similarity equation for it, namely how to find the P matrix.

How to Determine if A Matrix is Diagonalizable

Algorithm 6.2. Let A be a matrix of size n. Here is how we determine if A is diagonalizable and if it is, how to find P and diagonal matrix D, such that $A = PDP^{-1}$.

1. Find eigenvalues of A.
2. Find n linearly independent eigenvectors for A. If they do not exist, then A is not diagonalizable. The way to go about this is to find a basis for each eigenspace, which is $\text{Null}(A - I\lambda))$, for each eigenvalue λ.
3. Construct P by putting the vectors from step 2 into a matrix.
4. Construct D from the corresponding eigenvalues. I.e if the eigenspace of the eigenvalue λ has dimension 2, then λ will occur twice in D.

Example 6.15. We will determine if the following matrix is diagonalizable:

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

Step 1: We find

$$= \det(A - \lambda I) = -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2.$$

So, the eigenvalues of A are 1 and -2.
Step 2: Since A is a matrix of size 3, we must find three linearly independent eigenvectors. If we cannot, then A is not diagonalizable. To this end, we calculate each eigenspace. Let's start with $E_1(A)$. Recall, $E_1(A) = \ker(A - I)$, so we need to row reduce

$$\begin{bmatrix} 0 & 3 & 3 & \mid & 0 \\ -3 & -6 & -3 & \mid & 0 \\ 3 & 3 & 0 & \mid & 0 \end{bmatrix}.$$

This matrix row reduces to

$$\begin{bmatrix} 1 & 0 & -1 & \mid & 0 \\ 0 & 1 & 1 & \mid & 0 \\ 0 & 0 & 0 & \mid & 0 \end{bmatrix}.$$

From this we see that

$$\ker(A - I) = \left\{ s \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \mid s \in \mathbb{R} \right\}.$$

So, we pick $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. This is as best as we can do for $E_1(A)$ since it is one dimensional.

Next, we find a basis for $E_2(A)$. To this end

$$[A + 2I \mid 0] = \begin{bmatrix} 3 & 3 & 3 & \mid & 0 \\ -3 & -3 & -3 & \mid & 0 \\ 3 & 3 & 3 & \mid & 0 \end{bmatrix}$$

row reduces to

$$\begin{bmatrix} 1 & 1 & 1 & \mid & 0 \\ 0 & 0 & 0 & \mid & 0 \\ 0 & 0 & 0 & \mid & 0 \end{bmatrix}.$$

Therefore,

$$\ker(A + 2I) = \left\{ s \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \mid s, t \in \mathbb{R} \right\}.$$

Hence, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ are linearly independent eigenvalues. Since $\mathbf{v}_1$ is associated to another eigenvalue, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent. Thus, as there are three of them, A is diagonalizable!

Step 3: We set

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ 01 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Example 6.16. We will determine if the following matrix is diagonalizable

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Step 1: The eigenvalues for A is 1.

Step 2: We find that the matrix $A - I$ row reduces to $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Therefore, $\text{nullity}(A - I) = \dim(E_1(A)) = 1$. This is bad since as there are no other eigenvalues there are no more possibilities for linearly independent eigenvectors besides the one corresponding to 1. Therefore, A is not diagonalizable.

Index