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 $(s_1, \ldots, s_n)$ 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.

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

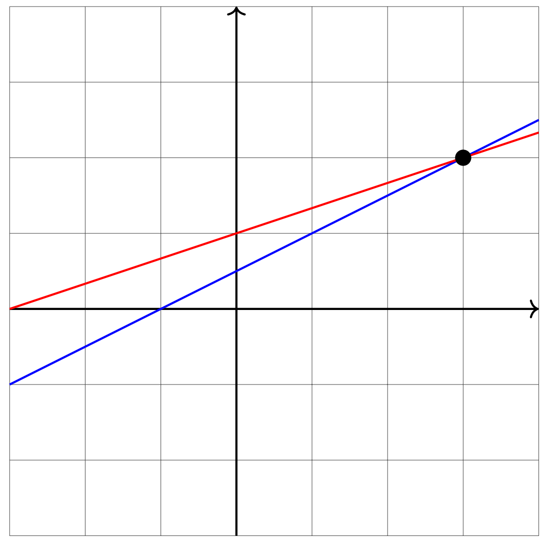
Solution Set

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

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

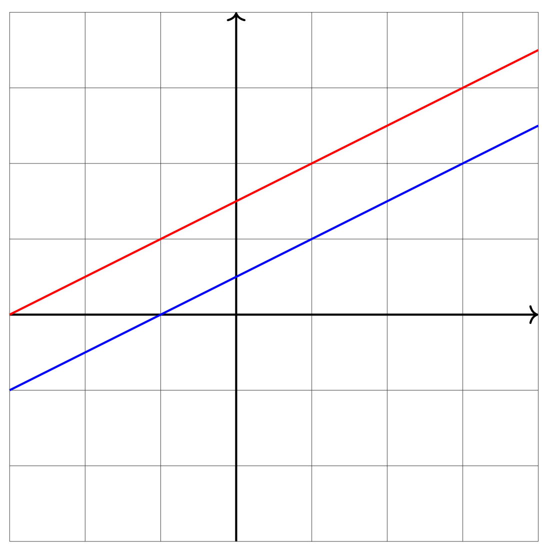
is the set $\{(3,2)\}$. Here's a geometric picture of what is going on:



2. The solution set to the system

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

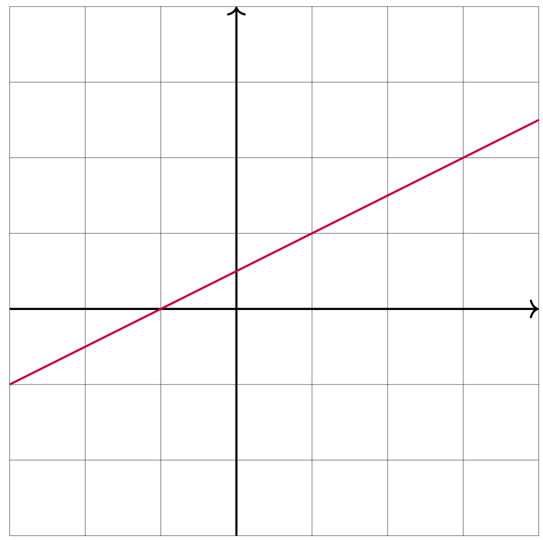
is the empty set, meaning there is no solution. Here is a geometric picture of what is going on:



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. □

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

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

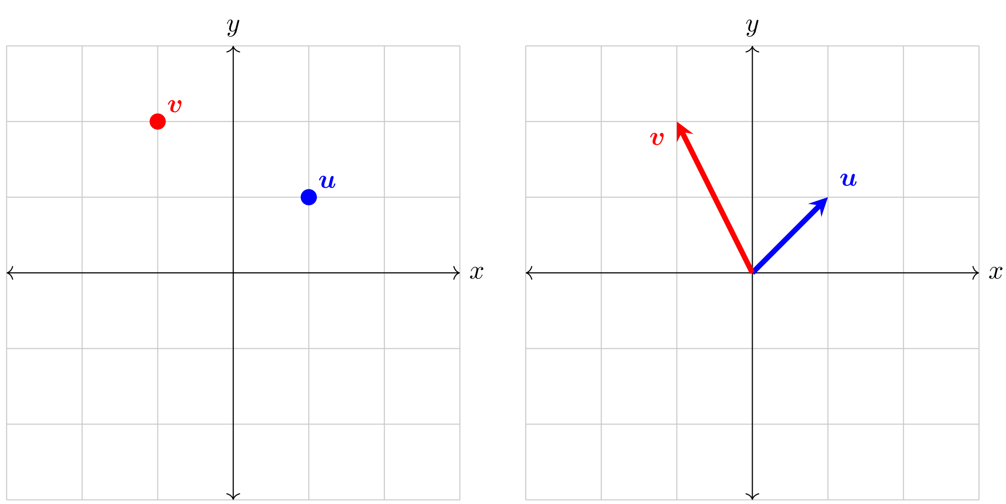
in $\mathbb{R}^4$

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

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

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

A vector $\mathbf{a} = (a_1, \dots, a_n)$ can be visualized in two ways: 1) They can be thought of as points $(a_1, \dots, a_n)$ 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 $(a_1, \dots, a_n)$. 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 $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ component wise:

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

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

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

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) $1\mathbf{a}=\mathbf{a}$.

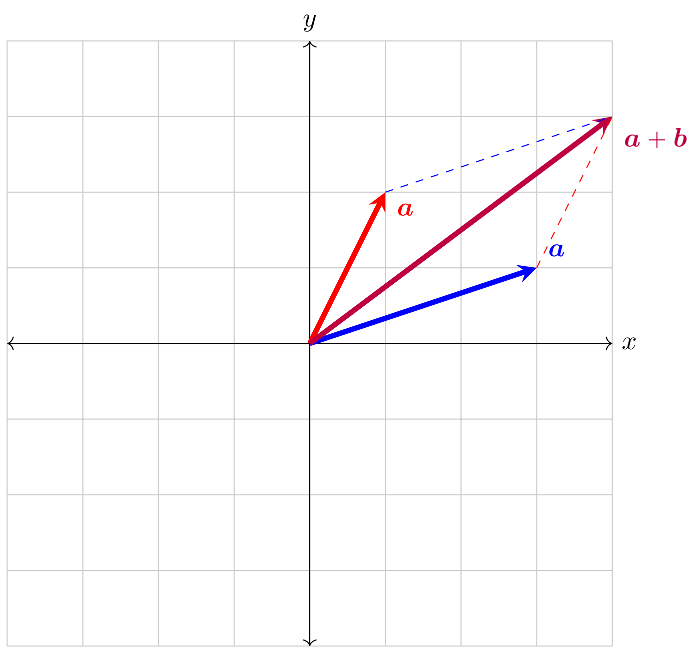
Example 2.18. We compute:

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

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

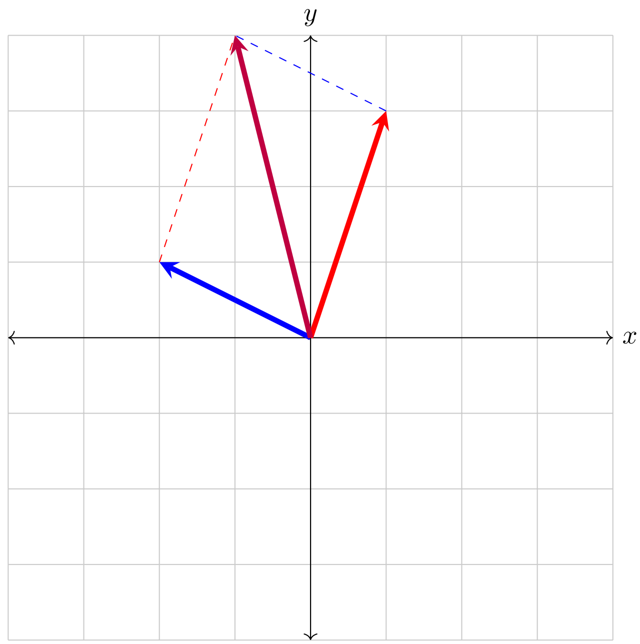
Parallelogram Law for Vectors

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 $c_1, \ldots, c_n \in R$

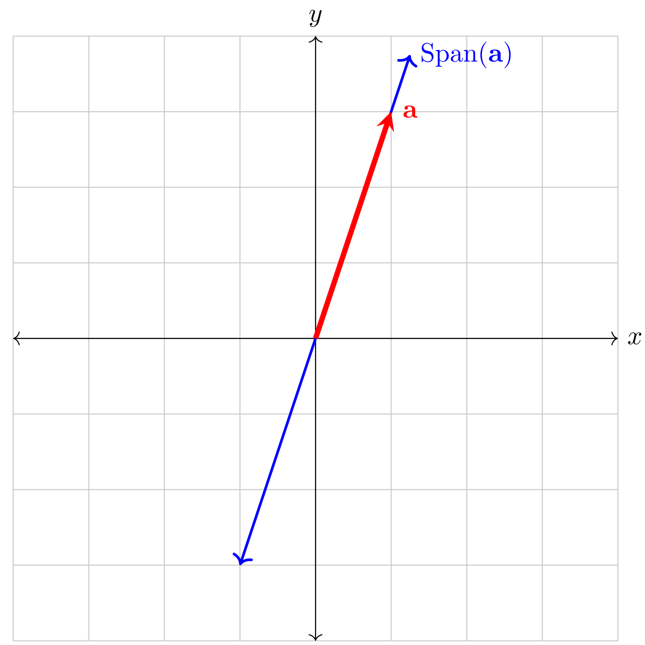
Often, we will want to consider all linear combinations of a set of vectors. To this end, we define the **span** of a set of vectors.

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 $\mathbb{R}^2$.

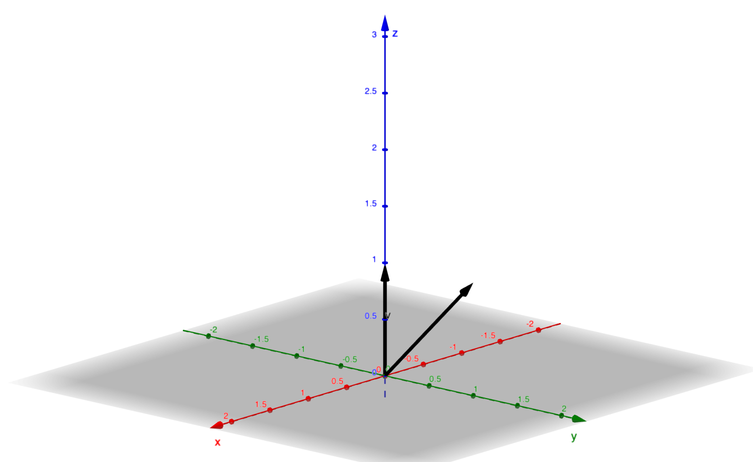
Example 2.20. Consider the vector $\mathbf{a} = (1, 3)$ in $\mathbb{R}^2$. The span of $\mathbf{a}$ is the set of all linear combinations of $\mathbf{a}$; that is the set of all things of the form $c\mathbf{a}$ with $c \in \mathbb{R}$. Geometrically, this is the line through the orgin passing through $(1, 3)$, 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 $\mathbf{x}_1, \dots, \mathbf{x}_n$, 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 $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$
2. there are $c_1, \dots, c_n \in \mathbb{R}$ such that
$$c_1\mathbf{x}_1 + \dots c_n\mathbf{x}_n = \mathbf{b}.$$
3. the system, whose augmented matrix is
$$\begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_n & \mathbf{b} \end{bmatrix},$$
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

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

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

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

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

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 ,

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

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 $\mathbf{a}_1, \ldots, \mathbf{a}_n$. We say that the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are *linearly independent* if the **only solution** $(x_1, \ldots, x_n)$ to the vector equation

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

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

Linear Dependence

Definition 2.27.
Consider any list of $m \times 1$ column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$. 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 is a $n \times 1$ vector $(x_1, \ldots, x_n)$ such that not every $x_i = 0$ that satisfies the vector equation

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

Exercise 2.28. *Suppose a set of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ contains the zero vector. Show that the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ 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 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 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 independent.
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 no nontrivial solution.

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

Theorem 2.10. Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ 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 dependent.
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 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 $\mathbf{a}_1, \dots, \mathbf{a}_k$ in $\mathbb{R}^n$ with $k \geqslant 2$. The vectors $\mathbf{a}_1, \dots, \mathbf{a}_k$ 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 $\mathbf{a}_1, \dots, \mathbf{a}_k$ are linearly dependent, then at least one of the $\mathbf{a}_i$ is a linear combination of the other $\mathbf{a}_j$'s. Since $\mathbf{a}_1, \dots, \mathbf{a}_k$ are linearly dependent, there are constants $c_1, \dots, c_k$, not all zero, such that

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

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

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

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

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

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

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

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

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

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. $\square$

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. $\square$

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 mulltiplication 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
$$\left[A \quad \middle| \quad I_n.\right]$$
2. Row reduce $\left[A \quad \middle| \quad I_n.\right]$ to something of the form
$$\left[I_n \quad \middle| \quad B\right].$$
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 (5 should be in here, but I do not have the energy to put it in

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

□

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 $\operatorname{im}(f)$

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 $T(\mathbf{0}) = \mathbf{0}$ 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,

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

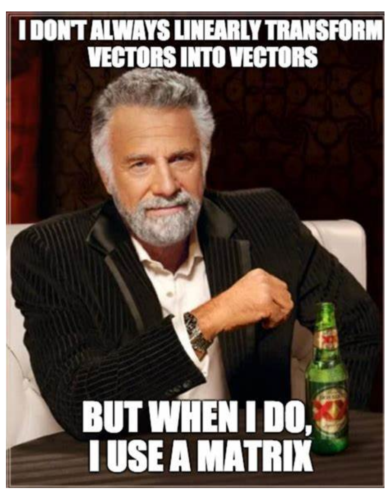
Set

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

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). □

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).

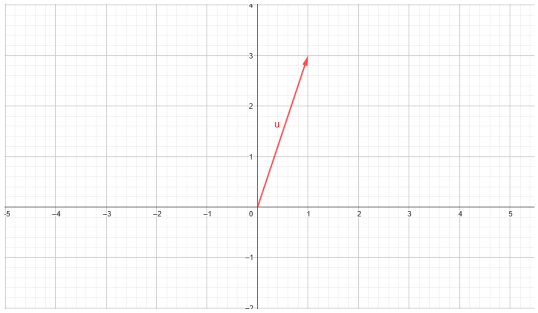
First, we focus on the case when $T : \mathbb{R}^2 \to \mathbb{R}^2$. We can (and will) think of T as acting on $\mathbb{R}^2$. In fact, there is a rigorous notion of what it means for a special type of set called a *group* to act on a space! When our space is $\mathbb{R}^n$ or even $\mathbb{C}^n$ ($\mathbb{C}$ 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

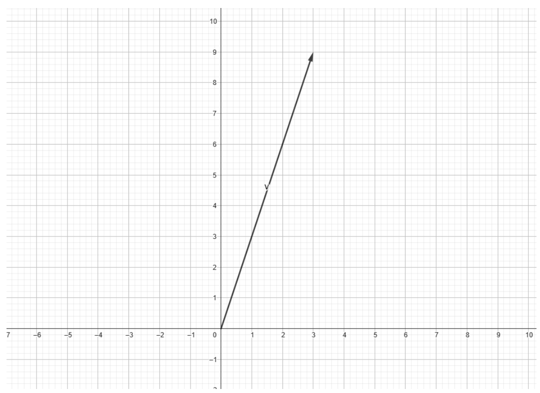
$$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}.$$

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 $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = (2, 6)$:



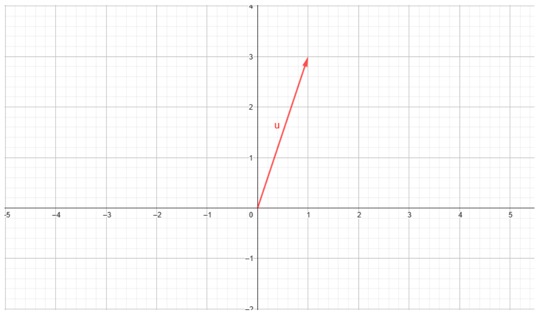
.

Scaling

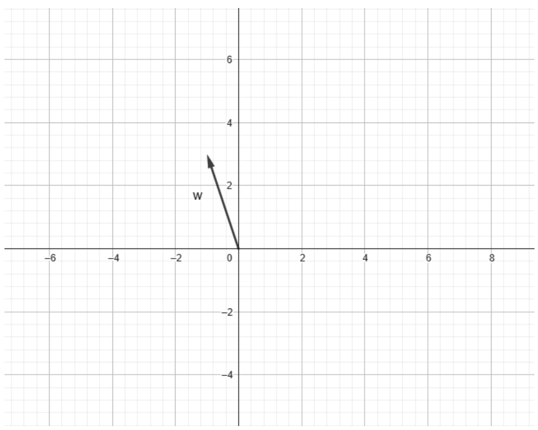
Definition 3.18. The 2×2 matrix that reflects across the x-axis is

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

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



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



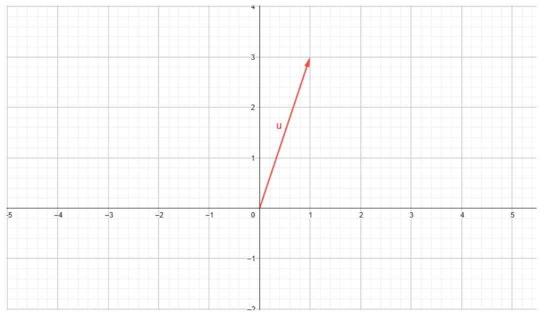
.

Reflection across y-axis

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

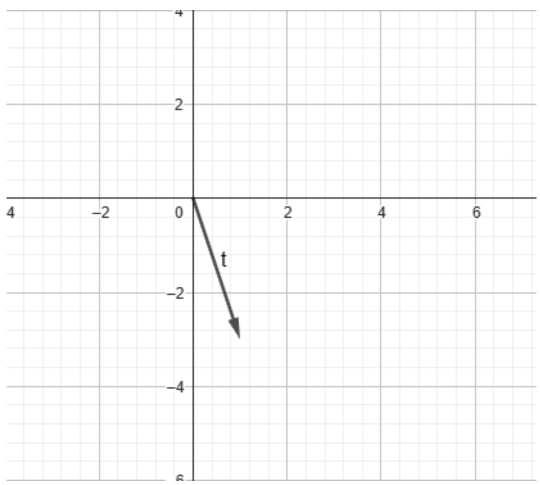
$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

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



.

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



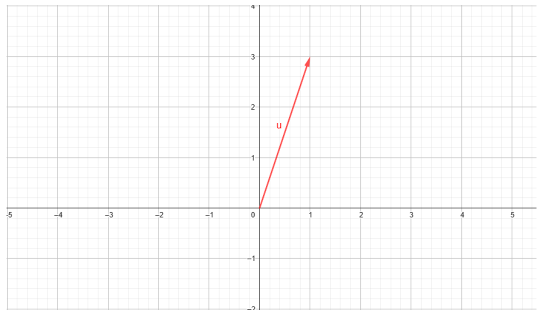
.

Horizontal Shear

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

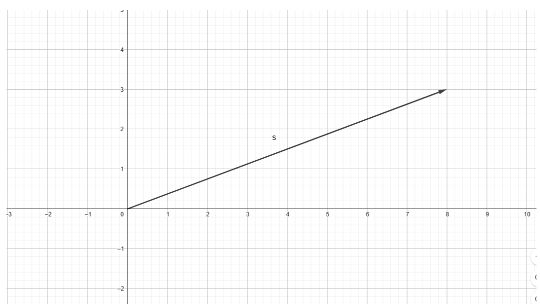
$$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}.$$

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



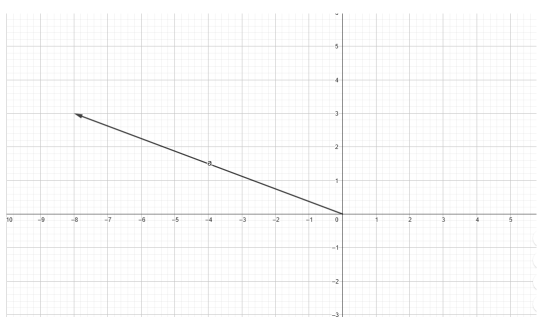
.

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 $\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (-5, 3)$:



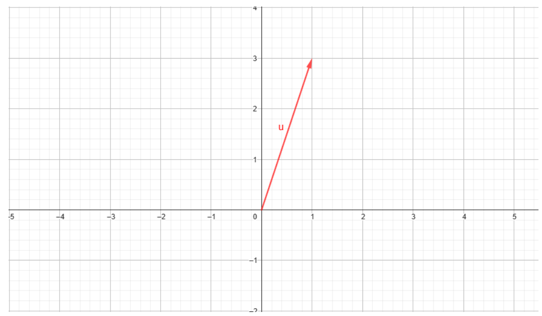
.

Vertical Shear

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

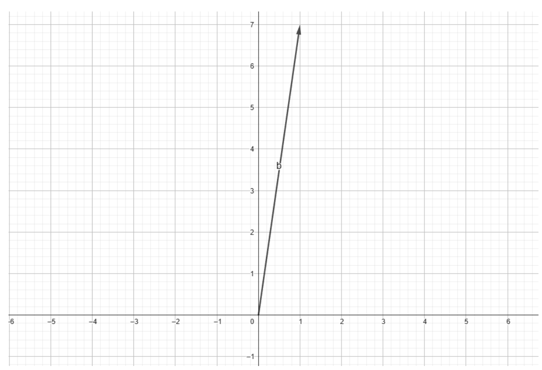
$$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}.$$

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



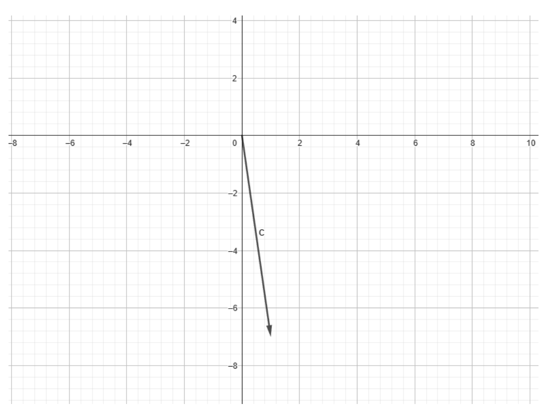
.

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 $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (1, 1)$:



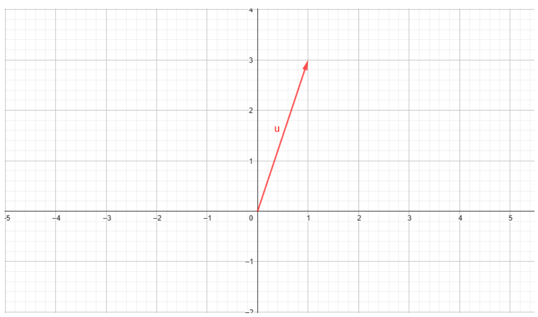
.

Rotation Matrix

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

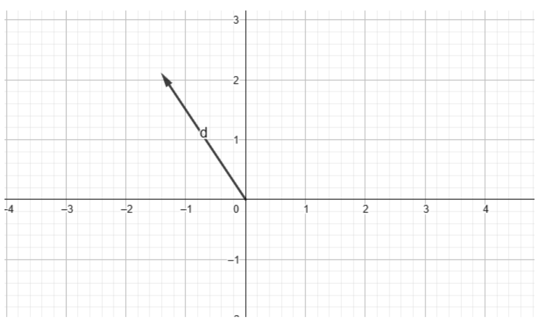
$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

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



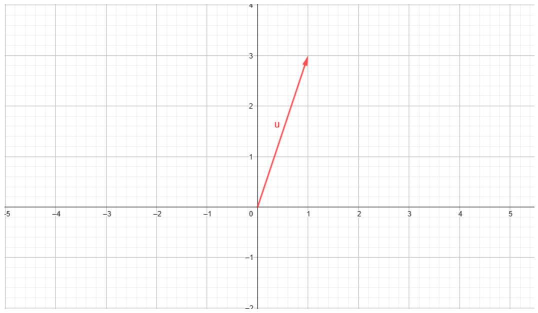
.

Reflection across $y = x$

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

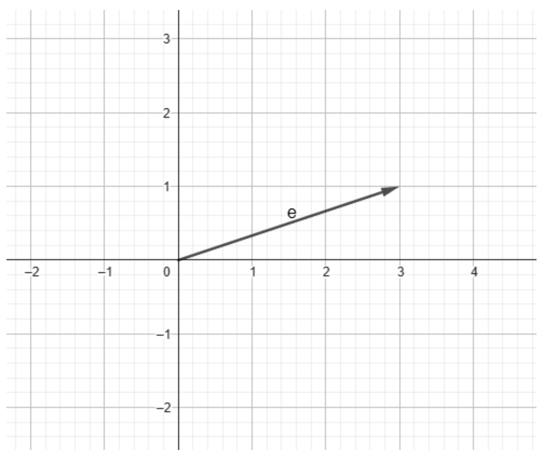
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

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



.

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



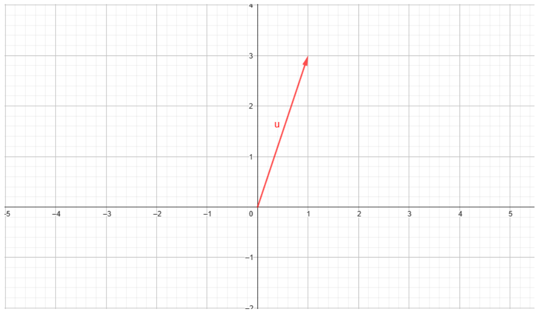
.

Reflection across $y = -x$

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

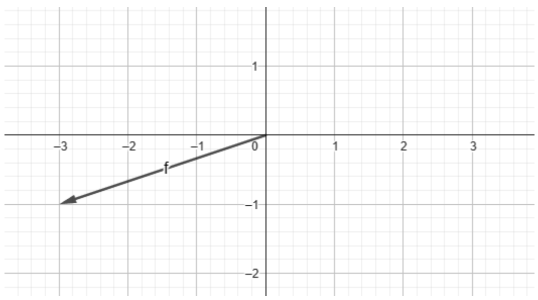
$$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

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! $\square$

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?*

Index