Linear Algebra (MTH 141)

Adam Szava

April 2021

Introduction

This document contains a compilation of my notes regarding content in MTH 141: Linear Algebra from Ryerson University. We cover almost all of the sections from the textbook; more than what the course itself covers.

The information for these notes comes from my professor's lectures, online resources, and the textbook *A First Course in Linear Algebra* originally by *K. Kuttler* which is a freely available open text document. In this text, definitions are boxed in red, theorems are boxed in blue, and any examples are boxed in green.

This text follows the course textbook very closely, however all exposition is my own. Many examples come directly from the textbook, however all solutions are my own.

Contents

1 Chapter 1: Systems of Equations

1.1 Systems of Equations, Geometry

This section introduces elementary concepts relating to systems of linear equations, and demonstrates the meaning of a solution to these systems visually.

To begin, consider the linear equation as follows as an example:

$$y = 2x - 1$$

This linear equation can be plotted on the xy-plane which visually describes the relationship between the variables (x and y) as described by the equation.

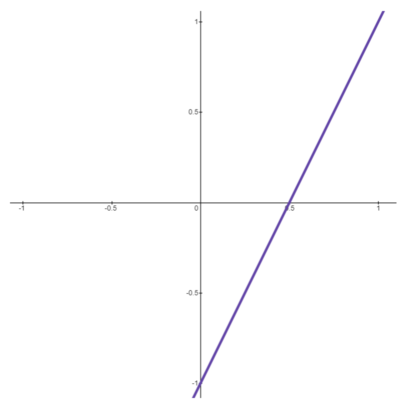

Figure 1: Plot of $y = 2x - 1$ using Desmos.

The meaning of the visualization is that any point on that line is a solution to the equation, as in if you choose some random point (a, b) on the line, then when you take the original equation and substitute $x = a$ and $y = b$ the left hand side of the equation is equal to the right hand side ($LHS = RHS$).

The example equation can be written in the following form using basic algebra:

$$2x - y = 1$$

This is can example of a linear equation in *standard form*. More precisely, this is an example of a linear equation of two variables (x and y) with coefficients of 2, -1, and a constant term of 1.

The following is also a linear equation in standard form:

$$3x - 2y + 3z = -3$$

... except this time the equation describes a plane in 3D space. As you can see the term *linear* does not mean that it forms a line when graphed, it is a more general concept that applies to higher dimensions. The following graph is a plot of that plane.

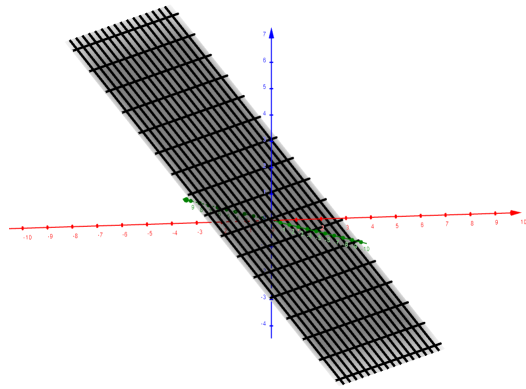

Figure 2: Plot of $3x - 2y + 3z = 1$ using GeoGebra.

We will learn in later sections how to describe any plane as a linear equation, for now just understand that the meaning of the plot is the same, every point on the plane is a solution to the equation given. As in if you choose some random point (a, b, c) that is **on** the plane, then when you take the original equation and substitute $x = a$, $y = b$, and $z = c$ then $LHS = RHS$.

More generally, the following is the definition of a linear equation.

Definition 1.1.1 Linear Equations

A **linear equation** is any equation in the form:

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

... where each $a_n \in \mathbb{R}$, $b \in \mathbb{R}$, and each x_n is a variable. We can also call these real numbers: *scalars*.

If we set $n = 2$, $x_1 = x$, and $x_2 = y$, we get something familiar, namely a line on the xy-plane like we saw before:

$$a_1x + a_2y = b \implies y = -\frac{a_1}{a_2}x + \frac{b}{a_2}$$

Note that this is the equation for *any* line on the xy-plane, and so this is more general than the specific example. This section began with a specific example of a line, but most definitions and discussions in this text (and in this course) are done in general terms, using things like the above definition.

A definition of a **system of linear equations** (SoLE) (which will be defined formally in the next section) is a set containing two or more linear equations with the same variables . Consider two lines on the xy-plane called ℓ_1 and ℓ_2:

$$\ell_1 : a_{11}x + a_{12}y = b_1$$

$$\ell_2 : a_{21}x + a_{22}y = b_2$$

These two lines form a SoLE as they have the same variables (x and y). The notation being used is a_{bc} which represents the coefficient (a) of the c^{th} variable on the b^{th} linear equation in the system. This is common notation in this course. The name for this would be a *system of two equations in two variables.* Graphically they can be plotted on the same pair of axis, and will intersect in one of the following three ways:

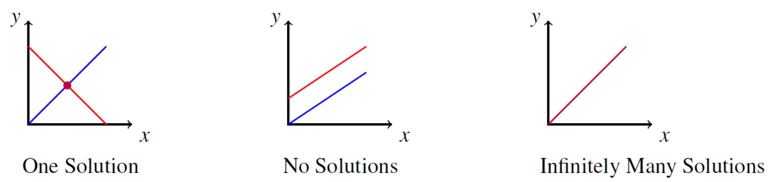

Figure 3: The 3 ways lines can collide on a 2D plane, and their respective number of solutions (From course textbook).

Graphically, a **solution** to a SoLE is the point (or set of points) where the lines intersect. Take the first instance for example, here the lines intersect at a unique point and so the system has one unique solution. Then take the second instance, here the lines are parallel and so they do not intersect, this means the system has no solutions. Finally consider the third instance, here the lines overlap each other as they are the same lines, for this reason the system has infinitely many solutions as the lines intersect at every point on either line (since they are the same line).

Generally, any system of linear equations will have one solution, no solutions, or infinitely many solutions. The definitions for these concepts will be formalized in the following section, for now we just aim to get a graphical understanding. Visualizing concepts in Linear Algebra is a very crucial part of the learning process, especially in this course, and so extra care will be taken in this text to emphasize that whenever possible.

Carrying on, consider now adding a third line to the plot, call it ℓ_3, as in our system is now:

$$\ell_1 : a_{11}x + a_{12}y = b_1$$

$$\ell_2 : a_{21}x + a_{22}y = b_2$$

$$\ell_3 : a_{31}x + a_{32}y = b_3$$

Given that the first two lines have at least one solution, there are essentially two ways the third line can be incorporated in the system to form a solution:

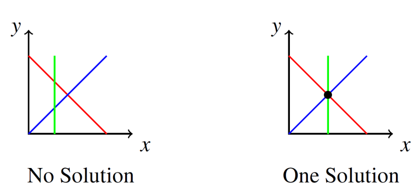

Figure 4: The 2 ways 3 different lines can collide on a 2D plane starting with a unique solution, and their respective new number of solutions (From course textbook).

It is of course possible for ℓ_3 to be equal to either of the other lines, in this case there is still just one solution as there is only one point where **all 3 lines** intersect (it will be the same as the solution to the first two equations).

If you begin with a system that has no solutions, adding a third equation (green line in the following figure) will never introduce a solution.

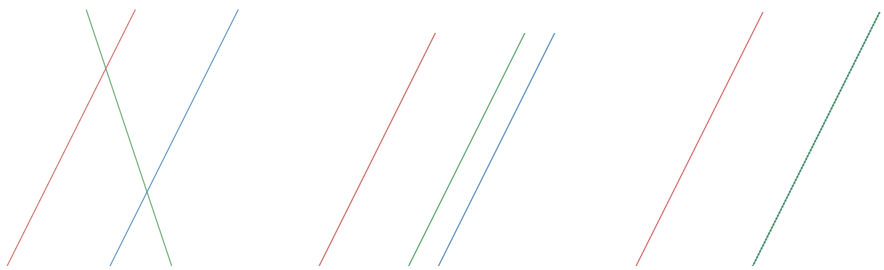

Figure 5: The 3 ways a system of 3 equations can have no solutions using Desmos.

In all of the above cases, the system of 3 equations has no solutions, as there is no unique point where the three lines intersect.

The following is an important remark: Including more equations never increases the number of solutions of a SoLE.

Now let's shift our focus from lines to planes, before we move onto the next section where we will have a more analytical approach to defining and solving these systems.

Given some system of two equations in three variables defined as follows (by convention π is used to identify a plane just like ℓ was used for lines. In this context π has nothing to do with the number $3.14159\ldots$):

$$\pi_1 : a_{11}x + a_{12}y + a_{13}z = b_1$$

$$\pi_2 : a_{21}x + a_{22}y + a_{23}z = b_2$$

There are three ways these two planes could intersect:

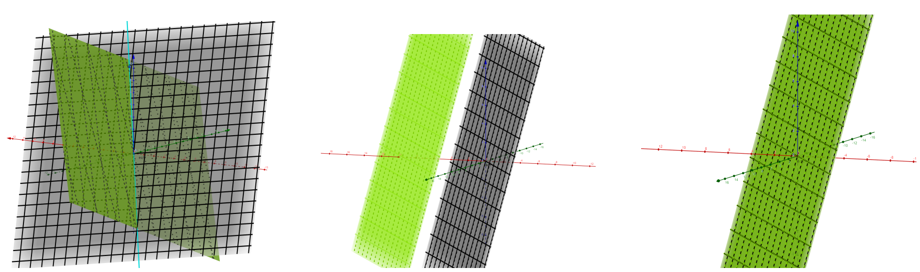

Figure 6: The 3 ways 2 planes can intersect using GeoGebra.

In this first case, the planes intersect on a line, and so there are *infinite solutions* to this system, any point on that line. Importantly note that while the planes appear to end in the images, that is only for comprehension, in reality the planes extend forever. In the second case, the planes do not intersect as they are parallel, here there are *no solutions*. In the third case, the planes are co-incident meaning they are the same plane, here there are *infinite solutions*.

Finally, let's consider a system of 3 planes, which would be called a system of 3 equations in 3 variables. There are again essentially 3 ways the planes can intersect, as follows:

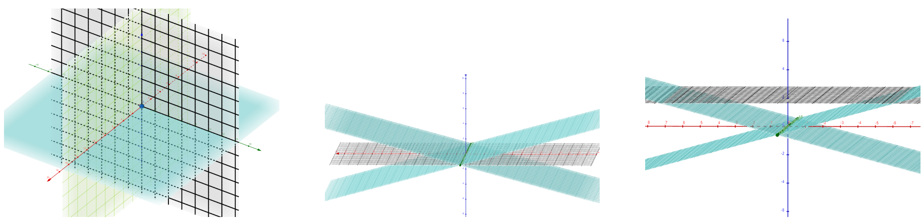

Figure 7: The 3 ways 3 planes can intersect using GeoGebra.

In the first case, the 3 planes intersect at a point, and so the system has a unique solution. In the second case all 3 planes intersect on a line, and so the system has infinite solutions. Finally in the third case the planes do not **all** intersect at any point and so the system has no solutions.

Before we move onto the next section here is one small topic I am choosing to cover now rather than later. This course uses the concept of $\mathbb{R}^n$ very often. To understand what this is, consider first the real number line. You could think of this line as a set of numbers, namely the set of all real numbers assuming the line goes infinitely in each direction. Mathematicians call this $\mathbb{R}$. Now consider a 2D plane, this plane is also a set, namely the set of all pairs of points (x, y) of which individually each $x \in \mathbb{R}$ and $y \in \mathbb{R}$ (note that $\in$ can be read as "in"). This would be called $\mathbb{R}^2$. This extends to as many dimensions as you'd like. $\mathbb{R}^{40}$ is the set of all points $(x_1, x_2, \cdots x_{40})$ of which individually each $x_i \in \mathbb{R}$ for each $1 \leq i \leq 40$. You will become very comfortable with this notation throughout the course, and $\mathbb{R}^n$ will be formally defined in *Chapter 4*. For now when I say $\mathbb{R}^n$ you should imagine n-dimensional space.

Bouncing off this last point we can now define lines and planes more precisely.

Definition 1.1.2 Lines, Planes, and Hyperplanes

In $\mathbb{R}^2$ each linear equation represents a line.
In $\mathbb{R}^3$ each linear equation represents a plane.
In $\mathbb{R}^n$ each linear equation represents an unbounded flat object called a *hyperplane* which is $(n-1)$-dimensional.

We will spend a lot of time discussing things like *hyperplanes* in *Chapter 4*, for now just keep it at the back of your mind.

This essentially concludes the content in this section of the course. I would just like to note here that much of linear algebra concerns itself with finding solutions to these systems of linear equations, however this mathematical exploration will bring you much further than you would imagine it could. Linear Algebra is a huge field of math and it has deep reaching roots into many other related fields.

1.2 Systems of Equations, Algebraic Procedures

This section of the course covers a rigorous introduction to solving these systems of linear equations of any size, and comprehending the meaning of the solutions, analytically. By the end of this chapter you will be able to solve a system of any number of equations in any number of variables. This section is broken up into 4 sub-sub-sections.

Before we begin however, we must formally define a few terms used beginning with the formal definition of a system of linear equations:

Definition 1.2.1 Systems of Linear Equations

A system of linear equations (SoLE) is a list of equations in the form:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

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

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

... where each $a_{ij} \in \mathbb{R}$ and each $b_j \in \mathbb{R}$. You would call this a *system of m equations in n* variables.

Definition 1.2.2 Homogenous Systems of Equations

A system of linear equations is homogenous if each $b_j = 0$, as in:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0$$
$$\vdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0$$

... where each $a_{ij} \in \mathbb{R}$.

Definition 1.2.3 Solution Set of a SoLE

Suppose we have a SoLE of m equations in n variables, say that the real numbers $c_1, c_2, \ldots c_n$ are solutions to the SoLE if they make all the equations true simultanously when you sub in $x_1 = c_1, x_2 = c_2, \cdots x_n = c_n$. The **solution set** of a SoLE is the set of all groups of c_n.

Recall that visually the **solution** to a SoLE was where the lines/planes/hyperplanes intersected. Algebraically the solution to a SoLE is the values that can be substituted in for each variable (as in x, y, and z) that make each equation in the system true at the same time. This is a different perspective on the same concept. Recall that if you plug in a point on a plane into the equation of the plane, then the equation is true. In that sense we can connect the ideas of intersecting planes (or lines or hyperplanes) with this algebraic definition as the intersection points are points that are shared between all the equations in the system (if it is a solution).

Definition 1.2.4 Consistent and Inconsistent Systems

Say a SoLE is **consistent** if the solution set is not empty, as in there is at least one solution to the system. Say a SoLE is **inconsistent** if there is no solution.

1.2.1 Elementary Operations

The elementary operations used in Linear Algebra form the basis for almost the entirety of the course, for this reason extra care will be taken to define the elementary operation and to prove them to you to a degree usually not done in the course.

Definition 1.2.5 Elementary Operations

Elementary operations are any of the three following operations that can be applied to a system of linear equations:

1. Interchange the order in which the equations are listed.
2. Multiply any equation by a nonzero number.
3. Replace any equation with itself added to a multiple of another equation.

What's so special about these operations? You could also add 3 to the left and right side of any of the equations, couldn't you? Yes you could, but that would change the solution of the SoLE.

The reason these *elementary operations* are useful are because they **do not** change the solution set of the SoLE. This idea forms the basis of the course. The following theorem explicitly states this idea using a system of two equations, however the idea applies to any number of equations.

Theorem 1.2.1 Elementary Operations and Solutions

Suppose you have the following system of linear equations in n variables (for simplicity E_1 and E_2 are used to descibe the non-constant part of the linear equations):

$$E_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

$$E_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$

This system has some solution set S. Then the following systems also have a solution set of S:

1.
$$E_2 = b_2$$
$$E_1 = b_1$$

2.
$$E_1 = b_1$$
$$k(E_2) = k(b_2)$$

... for some non-zero scalar k.

3.
$$E_1 = b_1$$
$$E_2 + kE_1 = b_2 + kb_1$$

... for any scalar k.

We will take the time to provide deep intuition for this theorem in this text, as I believe it is necessary to fully believe in what we are going to use as foundation. Before that, lets do a through example to understand what it means to use the operations.

Example 1.2.1 Elementary Operations

Suppose we have the following SoLE:

$$x + 3y + 6z = 25$$

$$2x + 7y + 14z = 58$$

$$2y + 5z = 19$$

Find the solution to this SoLE by using each of the three Elementary Operations.

To begin this problem, which will be described in detail, let us visualize the problem by looking at a plot of the three planes:

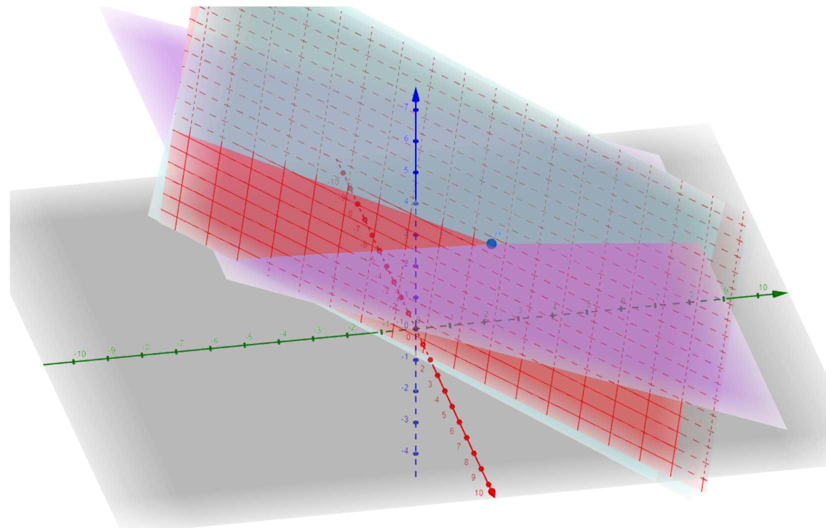

Figure 8: The 3 planes in the above example and their point of intersection using GeoGebra.

Now let's take a moment to consider what happens when we do each of the elementary row operations. Then lets combine them to find the solution (note that this system has a unique solution at the point in the graph above).

The first elementary row operation changing the order of the equations may seem useless, however it will prove to be useful to know later in this chapter. We can see this elementary row operation in action:

$$\boxed{R_1 \Updownarrow R_2}$$

$$2x + 7y + 14z = 58$$

$$x + 3y + 6z = 25$$

$$2y + 5z = 19$$

Notice the notation used: $\Updownarrow$ means switch the positions of the equations on either side of it. The letter R is used to mean *row* which can essentially be used interchangeably with the word *equation*. Unsurprisingly the graph of this new system is unchanged when compared to the original.

Now let's consider what happens when we do the second elementary row operation. This operation entails multiplying both sides of one equation by some non-zero scalar. For the sake of example, let's say we want to multiply the third equation by 2, we would get the following system (applying this operation to the original system in the example):

$$\boxed{2R_3}$$

$$2x + 7y + 14z = 58$$
$$x + 3y + 6z = 25$$
$$4y + 10z = 38$$

It may come as a surprise to hear that this operation also does not change the graph of the planes in figure 8. Since it doesn't change the graph, it should be reasonable to say that this operation does not change the solution set. But why does this operation not change the graph of the planes? Consider first an algebraic explanation.

Given the following equation:

$$x + 2 = 4$$

It should be evident that the solution to this equation is $x = 2$. Consider then the related equation:

$$2(x + 2) = 2(4)$$

The solution to this equation is *still* $x = 2$. Multiplying by a scalar does not change the solution of an equation with a single variable, but does it change something more complex, say for example a line?

Take the following line for example:

$$y = 2x + 4$$

... this line has a slope of 2 and a y-intercept of 4, this means the $\frac{rise}{run} = \frac{2}{1}$ Now consider the following line:

$$2(y) = 2(2x + 4) \implies y = \frac{4}{2}x + 4$$

This new line is notably still the same line, with the same slope and intercept. The new slope however is written in the form $\frac{4}{2}$ instead of $\frac{2}{1}$ which has the effect of stretching out our line by a factor of 2. It's a different line, but it has been transformed in a way that made it appear to not change. We will discuss transformations in detail in *Chapter 5.*

In the exact same sense, when we multiply a plane by some constant k, we are simply stretching out all of the points on the plane by that same factor (imagine spinning an infinitely thin pizza dough with infinite area), and so the effect is to not change any of the points on the graph, and most importantly it does not change the position of the solution.

Finally, let's now consider the effect of the third elementary row operation. This operation entails *replacing* an equation with itself added to some multiple of another equation in the system. For this example, say we want to replace the second equation with itself plus -2 times the third equation, we would get the following system (note that in *Figure 8* the red plane is R_2):

$$\boxed{R_2 \to R_2 + (-2)R_3}$$

$$x + 3y + 6z = 25$$
$$2x + 7y + (-2 \cdot 2)y + 14z + (-2 \cdot 5)z = 58 + (2 \cdot 19)$$
$$2y + 5z = 19$$
$$\boxed{\text{Simplify}}$$
$$x + 3y + 6z = 25$$
$$2x + 3y + 4z = 20$$
$$2y + 5z = 19$$

Plotting this new set of planes we get the following graph:

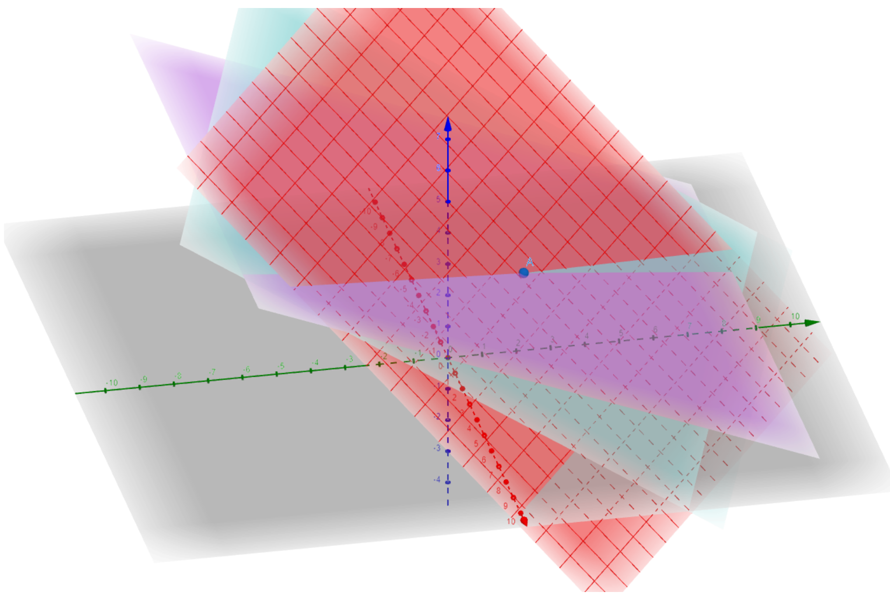

Figure 9: The 3 planes in the above example and their point of intersection after an elementary row operation of type 3 using GeoGebra.

There are important features of this plot, the red plane (R_2) has changed to a new plane, the other two planes did not change, and the new plane (R_2) still passes through the point which is the solution to the original system of equations.

When we do the elementary operation, our goal is not to find the solution to the system without changing the system, our goal is to simply find the solution. For this reason, even though the third elementary operation changes one of the planes, the only thing that matters is that it does not change the solution to the system.

Naturally the next question is: "Why doesn't it change the solution to the system?" Let's try to reason it out. Our goal is to show that if we have solutions to a system like:

$$E_1 = b_1$$

$$E_2 = b_2$$

Then the solutions to this system are the same as the solutions to the following:

$$E_1 = b_1$$

$$E_2 + kE_1 = b_2 + kb_1$$

Recall now that a solution to any one of the equations would make the E_n equal to it's respective b_n. This means that after plugging in a solution to the original system we would have:

$$b_1 = b_1$$

$$b_2 = b_2$$

Say we have some solution set S to our original system, as in when you plug it into our original system it makes $E_1 \to b_1$ and $E_2 \to b_2$. Then we have to show that this same solution makes $E_1 \to b_1$ and $E_2 + kE_1 \to b_1 + kb_2$. The first of these is easy to prove, by the definition of the solution $E_1 \to b_1$ when any of the solutions in the solution set is applied. The next part (that $E_2 + kE_1 \to b_1 + kb_2$ when a solution from the solution set is applied) we can show is true since we know that when you apply a solution from the solution set $E_1 \to b_1$ and $E_2 \to b_2$ which means that:

$$E_1 = b_1$$

$$E_2 + kE_1 = b_2 + kb_1$$

... becomes:

$$b_1 = b_1$$

$$b_2 + kb_1 = b_2 + kb_1$$

... when you apply a solution from the solution set, which is obviously always true for any k and any two linear equations E_1 and E_2.

This final point is a little harder to reason so if you're having trouble understanding then read it over a few times and try to understand each line. This extends very easily to systems of m equations (the system is already in n unknowns), as really only the line that is being replaced needs to be shown to not change the solution set.

Now that we have hopefully proved to you that all of the three elementary operations do not change the solution set, we can use the elementary operations to solve the system. We can effectively use the operations to eliminate variables from one of the equations to a point where we can just solve for a variable, and then *back substitute* to solve for the rest of the variables. Here is the process in detail:

$$\boxed{\text{Original System}}$$

$$x + 3y + 6z = 25$$

$$2x + 7y + 14z = 58$$

$$2y + 5z = 19$$

$$\boxed{R_2 \to R_2 + (-2)R_1}$$

This step eliminates the x term in the second equation.

$$x + 3y + 6z = 25$$

$$2x + (-2 \cdot 1)x + 7y + (-2 \cdot 3)y + 14z + (-2 \cdot 6)z = 58 + (-2 \cdot 25)$$
$$2y + 5z = 19$$

$$\boxed{\text{Simplify}}$$

$$x + 3y + 6z = 25$$
$$y + 2z = 8$$
$$2y + 5z = 19$$

$$\boxed{R_3 \to R_3 + (-2)R_2}$$

This step eliminates the y term in the third equation.

$$x + 3y + 6z = 25$$

$$y + 2z = 8$$
$$(-2 \cdot 0)x + 2y + (-2 \cdot 1)y + 5z + (-2 \cdot 2) = 19 + (-2 \cdot 8)$$

$$\boxed{\text{Simplify}}$$

$$x + 3y + 6z = 25$$
$$y + 2z = 8$$
$$z = 3$$

We have now learned that in the solution to our system $z = 3$, we can now substitute this value into the first and second equations:

$$x + 3y + 6(3) = 25$$

$$y + 2(3) = 8$$
$$z = 3$$

$$\boxed{\text{Simplify}}$$

$$x + 3y = 7$$
$$y = 2$$
$$z = 3$$

We have now learned that in the solution to our system $y = 2$, we can now substitute this value back into our first equation:

$$x + 3(2) = 7$$

$$y = 2$$

$$z = 3$$

$$\boxed{\text{Simplify}}$$

$$x = 1$$

$$y = 2$$

$$z = 3$$

Finally we have now learned that in the solution to our system $x = 1$, we now know that the solution to the system is $x = 1$, $y = 2$, $z = 3$.

This process is known as elimination method, and then back substitution. The back substitution part is when you know one of the components to the solution, and then you use that to go *back* and fill in earlier equations.

Understanding the process of this solution is crucial, however actually practising solving systems like this is not important. This is because the next subsection will discuss *Gaussian Elimination* which is a form of the elimination method that is much more structured and algorithmic. Nevertheless understanding this whole subsection will let you easily pick up what is to come.

1.2.2 Gaussian Elimination

In this section we formulate a systematic method to solve a system of m equations in n that is less cumbersome than the standard elimination method. To be clear, the standard elimination method will always work, small changes in perspective can make the whole process a lot easier. We will also address what happens if you try to solve for a solution to a system that has either infinite solutions or no solutions, which is something we glossed over in the last section.

The essence of this method, called *Gaussian-Elimination*, is that we can choose to write the following system (for example):

$$x + 3y + 6z = 25$$

$$2x + 7y + 14z = 58$$

$$2y + 5z = 19$$

As:

$$\left[\begin{array}{ccc|c} 1 & 3 & 6 & 25 \\ 2 & 7 & 14 & 58 \\ 0 & 2 & 5 & 19 \end{array}\right]$$

This is called the **augmented-matrix** of the SoLE. Each row corresponds to an equation in the SoLE. The left side of the bar is called the **coefficient-matrix**, and the right side is called the **constants column**. Notice that since we have three variables, on each row there are three numbers to the left of the bar. This includes the last row, which corresponds to the last equation in which x is missing. This is represented as a 0 in the matrix.

This may be a seemingly useless change of notation, however as we explore more concepts in this course (especially *Chapter 2* and *Chapter 5*) you will see that this change sheds light on a whole new way of thinking about more than just SoLEs.

Before we move on, you may be wondering what is a matrix? While the entirety of *Chapter 2* is dedicated to working with matrices, the general idea is that a matrix is an ordered form of a list of numbers. The numbers are ordered into rows and columns, and the number of those are called the dimensions of the matrix. In general a m by n matrix has m rows and n columns. The first row is the top one, and the first column is the leftmost one. Again, we will spend the entirety of the next chapter talking about matrices, so for now just focus on *augmented-matrices*. The following is the formal definition of an *augmented-matrix*:

Definition 1.2.6 Augmented Matrix

Given the SoLE:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

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

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

The augmented matrix of the system is:

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

The elementary operations can be applied to an augmented matrix, it is then called the *elementary row operations*, however as the matrix always has a direct correspondence to the original SoLE the operations are really the same and all reasoning we have developed for them are equivalent.

Definition 1.2.7 Elementary Row Operations

Elementary row operations are any of the three following operations that can be applied to an augmented matrix:

1. Switch two rows.
2. Multiply a row by a non-zero number.
3. Replace a row by any multiple of another row added to it.

Evidently, these are almost identical to the elementary operations. Let's solve a SoLE using an augmented matrix.

Example 1.2.2 Augmented Matrix

Suppose we have the following SoLE:

$$x - 2y + z = 0$$

$$2x + y - 3z = 5$$

$$4x - 7y + z = -1$$

Find the solution to this SoLE by converting it to an augmented matrix.

Right away let's convert this SoLE to augmented matrix form:

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

... and then use row operations to solve:

$$\boxed{R_2 \to R_2 - 2R_1}$$

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

$$\boxed{R_3 \to R_3 - 4R_1}$$

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

$$\boxed{R_2 \Updownarrow R_3}$$

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

$$\boxed{R_3 \to R_3 - 5R_2}$$

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -3 & -1 \\ 0 & 0 & 10 & 10 \end{array}\right]$$

$$\boxed{\frac{1}{10}R_3}$$

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

We can now convert this back in to equation form:

$$x - 2y + z = 0$$

$$y - 3z = -1$$

$$z = 1$$

From here we can simply back substitute since we know $z = 1$, as in:

$$x - 2y = -1$$

$$y = 2$$

$$z = 1$$

... and again since we know $y = 2$:

$$x = 3$$

$$y = 2$$

$$z = 1$$

This is the process by which most SoLEs are solved. The exact choices of steps used may have seemed to come out of no where, when in reality there were exactly following a process called *Gaussian Elimination*, which is a algorithm to solve any SoLE. Gaussian Elimination takes in an augmented matrix, and outputs the matrix in what is called **row-echelon form** (REF), which is the form of the augmented matrix right before I switched the SoLE back to equations, or formally that is (quickly note that **leading entry** of a row is the first non-zero value in that row from the left):

Definition 1.2.8 Row-Echelon Form (REF)

An augmented matrix is in row-echelon form if:

- All nonzero rows are above any all-zero rows.
- Each leading entry of a row is in a column to the right of the leading entries of all rows above it.
- Each leading entry of a row is equal to 1.

The following augmented matrices are in REF:

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

The following augmented matrices **are not** in REF:

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

Finally we are ready for one of the two main algorithms for solving SoLEs:

Definition 1.2.9 Gaussian Elimination

Begin with an augmented matrix:

1. Starting from the left, find the first non-zero column.
2. If necessary, switch rows so that the top entry of the column is non-zero. Call that entry the pivot position.
3. Use row operations to make the entries below the pivot position all equal to 0.
4. Completely ignoring the row which contains your pivot position, repeat the process for the remaining rows.
5. Divide each non-zero row by the value of the leading entry. The matrix will now be in row-echelon form.

Go back to the last example and see how we used *Gaussian Elimination* to solve that problem.

There is however a second algorithm, which is almost identical to the *Gaussian Elimination* algorithm, and it is called **Gauss-Jordan Elimination**.

To understand this second algorithm, first lets take a look at the REF form of the augmented matrix from the last example:

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

We could continue doing row operations on this augmented matrix to get it in a new form that would directly give us all three components instead of needing to do back substitution. The process is as follows:

$$\boxed{R_2 \to R_2 + 3R_3}$$

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

$$\boxed{R_1 \to R_1 - R_3}$$

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

$$\boxed{R_1 \to R_1 + 2R_2}$$

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

This augmented matrix is in **reduced** *row echelon form* (RREF). From here, you can essentially just read off the solutions by converting it back to equation form:

$$x = 3$$

$$y = 2$$

$$z = 1$$

No back substitution was needed here, that is the benefit of reducing an augmented matrix to RREF.

Definition 1.2.10 Reduced Row-Echelon Form (REF)

An augmented matrix is in row-echelon form if:

- All nonzero rows are above any all-zero rows.
- Each leading entry of a row is in a column to the right of the leading entries of all rows above it.
- Each leading entry of a row is equal to 1.
- All entries in a column above and below a leading entry are zero.

This definition of RREF is almost identical to the definition of REF but with an added point. The following matrices are in RREF:

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

Now that we have defined RREF, *Gauss-Jordan Elimination* is the second algorithm which takes in an augmented matrix and returns that matrix in RREF form. Note that it is almost identical to *Gaussian Elimination* except for the addition of the last line:

Definition 1.2.11 Gauss-Jordan Elimination

Begin with an augmented matrix:

1. Starting from the left, find the first non-zero column.
2. If necessary, switch rows so that the top entry of the column is non-zero. Call that entry the pivot position.
3. Use row operations to make the entries below the pivot position all equal to 0.
4. Completely ignoring the row which contains your pivot position, repeat the process for the remaining rows.
5. Divide each non-zero row by the value of the leading entry. The matrix will now be in row-echelon form.
6. Moving from right to left, use row operations to create zeros in the entires above all the leading entries. The matrix will now be in reduced row-echelon form.

From now on, we will do *Gauss-Jordan Elimination* when solving a SoLE since RREF is more useful than REF.

For a quick recap:

Gaussian Elimination

Augmented Matrix $\rightarrow$ REF

Gauss-Jordan Elimination

Augmented Matrix $\rightarrow$ RREF

We will now discuss what happens when you try to solve a SoLE which has either no solutions or infinite solutions. Beginning with no solutions, let's take a look at the following example:

Example 1.2.3 SoLE with no Solutions

Suppose we have the following SoLE:

$$2x + 4y - 3z = -1$$

$$5x + 10y - 7z = -2$$

$$3x + 6y + 5z = 9$$

Find the solution(s) to this SoLE if it has any.

Let's begin by graphing the planes so you can be convinced it has no solutions:

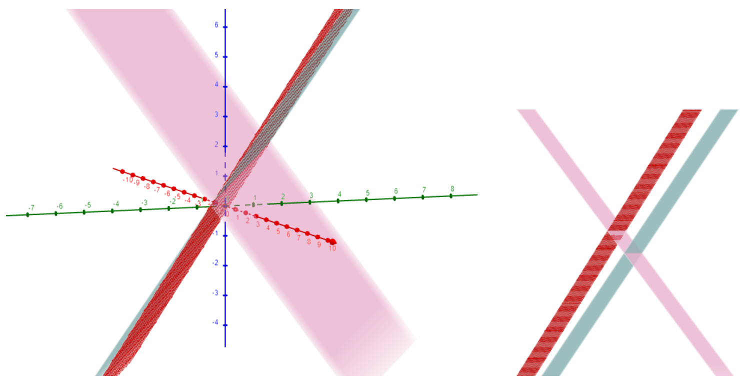

Figure 10: The system of 3 equations in 3 unknowns that has no solutions.

Let's solve this SoLE using *Gaussian/Gauss-Jordan Elimination*, and see what happens:

$$\left[\begin{array}{ccc|c} 2 & 4 & -3 & -1 \\ 5 & 10 & -7 & -2 \\ 3 & 6 & 5 & 9 \end{array}\right]$$

$$\vdots$$

$$\left[\begin{array}{ccc|c} 1 & 2 & -\frac{3}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 20 \end{array}\right]$$

There is a problem with the last row of this REF augmented matrix, if you were to convert it back to equation form you would have:

$$0x + 0y + 0z = 20 \implies 0 = 20$$

... which is never true, and so the system is *inconsistent*, meaning it has no solutions.

Generally if you see a row like the following while doing *Gaussian/Gauss-Jordan Elimination*:

$$\left[\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & a \end{array}\right] \text{ for some } a \in \mathbb{R}$$

... then we know the system is inconsistent.
Let's now look at a system that has infinite solutions, as in the following example:

Example 1.2.4 SoLE with infinite Solutions

Suppose we have the following SoLE:

$$3x - y - 5z = 9$$

$$y - 10z = 0$$

$$-2x + y = -6$$

Find the solution(s) to this SoLE if it has any.

Let's begin by graphing the planes so you can be convinced it has infinite solutions:

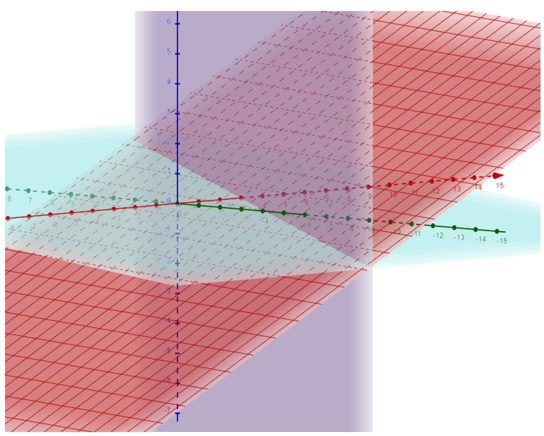

Figure 11: The system of 3 equations in 3 unknowns that has infinite solutions.

Let's solve this SoLE using *Gaussian/Gauss-Jordan Elimination*, and see what happens:

$$\left[\begin{array}{ccc|c} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ -2 & 1 & 0 & -6 \end{array}\right]$$

$$\vdots$$

$$\left[\begin{array}{ccc|c} 1 & 0 & -5 & 3 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This matrix has a column which does not contain a leading entry, this means that the variable which corresponds to that column (z) is called a **free variable** and can take on any value (which is where the infinite solutions come from). To understand what this means better, look at the equation from of the last augmented matrix:

$$x - 5z = 3$$

$$y - 10z = 0$$

$$0x + 0y + 0z = 0$$

That last row is true regardless of any values given to it, this is what leads to the infinite solutions. These equations can be rewritten as:

$$x = 3 + 5z$$

$$y = 10z$$

As you can see z has no restriction, as in it can take on the value of any $t \in \mathbb{R}$, for this reason lets set $z = t$. Which gives us our final solution set of:

$$x = 3 + 5t$$

$$y = 10t$$

$$z = t$$

... for any $t \in \mathbb{R}$. This is what is meant by z is a free variable, also called a parameter. It is just as possible for a system to have any number of free variables in its solution set. We will discuss more in *Chapter 4* what is the geometric meaning of these free variables, but for now they are just an algebraic tool to solve these systems. To get any particular solution to the system, pick some t, for example $t = 0$, which would give you the particular solution:

$$x = 3$$

$$y = 0$$

$$z = 0$$

In this last example, x and y are **basic variables**, and z is a **free variable**.

The solutions to this system could also be written in the following form using *vector* notation (something we will define in detail in the following chapters):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 5 \\ 10 \\ 1 \end{bmatrix}$$

Take a moment to process this notation, notice that the n^{th} element of each vector makes up the n^{th} equation in the solution. The left hand side of the equation is just a vector containing our variables. The first vector on the right is the constants column from our RREF matrix, and the last vector in the equation is what is related to the free variable, and it is called the system's **basic solution**, a system can have a whole list of basic solutions usually notated X_n.

In general, any system that has a row of entirely zeros in its RREF form has infinite solutions.

Using all the skills developed up to this point, you are now able to solve a system of m equations in n unknowns.

1.2.3 The Uniqueness of RREF

This sub-sub-section just contains two theorems relating to solving SoLEs, all of which are important to understand.

Theorem 1.2.2 Equivalent Augmented Matrices

If A and B are augmented matrcies, and you can begin with A, apply the elementary row operations and end with B, then the solutions of system A and B are the same, and vice-versa. Call A and B **equivalent**.

Theorem 1.2.3 Uniqueness of RREF

Every matrix A is equivlent to a unique matrix B in RREF.

What this means is that regardless of what order of row operations you perform on a system, you will always arrive at the same RREF form of the augmented matrix.

1.2.4 Rank and Homogeneous Systems

This sub-sub-section discusses some concepts relating to SoLEs that help us understand their solutions.

Recall that a homogeneous SoLE is one where all the constants are equal to 0. Every homogeneous is consistent, since $x_1 = x_2 = \cdots = x_n = 0$ is always a solution to the system of n unknowns. This solution is called the **trivial solution**, all other solutions are called **non-trivial solutions**.

Definition 1.2.12 Rank of a Matrix

Let A be a matrix and consider any REF form of A, or the RREF form of A. Then, the number of leading entries in the REF is the same in all the REF forms, as well as the RREF form, and is called the **rank** of A. Usually written rank(A).

Theorem 1.2.4 Rank, Parameters, and Columns

If A is an $m \times n$ matrix, $rank(A) = r$, and A is the coefficient matrix of a homogenous SoLE, then the number of parameters in the general solution of that system is $n - r$, as in:

$$rank(A) + \text{number of parameters} = \text{number of columns}$$

This previous theorem will show up many times in this course under many different names, and it is a fundamental result.

What this tells us is that if the rank of the coefficient matrix is less than its number of columns, it will have infinite solutions.

Theorem 1.2.4 Rank and Unique Solutions

Suppose we have a SoLE with an $m \times n$ coefficient matrix A, suppose further that the system is consistent, then the solution is unique if and only if $rank(A) = n$.

To understand this notion consider the following two augmented matrices in $RREF$:

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

This system has a unique solution since its coefficient matrix has 3 columns and also has a rank of 3.

Now consider this matrix:

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

Here the system does not have a unique solution since the coefficient matrix has 3 columns, but its rank is 2.

Theorem 1.2.5 Rank and Consistency

Suppose A is the coefficient matrix of some SoLE, and $\vec{b}$ is the column matrix which holds the constants of the same SoLE, then the system is consistent if and only if:

$$rank([A]) = rank([A \mid \vec{b}\ ])$$

Note that $\vec{b}$ just means a column matrix which contains the coefficients, the arrow on top just signifies it is a column. Take the following augmented matrix for example:

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

The rank of A is 2, but the rank of $[A \mid \vec{b}\ ]$ is 3 and so the system is inconsistent.

This concludes the content in *Chapter 1* of this text. You are now able to solve a system of m equations in n unknowns, and understand its solution set. The logic used in this chapter will be used throughout the course so spend some time here to make sure you understand everything.

2 Chapter 2: Matrices

2.1 Matrix Arithmetic

This subsection is concerned with defining the basic arithmetic used with matrices, including summation, scalar multiplication, multiplication of matrices, and inverse matrices. The final parts of this subsection are concerned with elementary matrices which relate this chapter back to the content in *Chapter 1*.

Definition 2.1.1 Matrices

A **matrix** (plr. matrices) is a 2D rectangular list of numbers. It has columns and rows, where each of the rows have the same number of entries, and each column has the same number of entries.
An $m \times n$ matrix has m rows and n columns. Each individual **entry** is denoted by its ij position, as in a_{ij} (the (i, j)-entry) is the entry that lays in the i-th row and j-th column of a matrix A. So the general form of a matrix would look like the following:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

... or more succinctly, for a matrix A:

$$A = [a_{ij}]$$

The following notation is also used for matrices:

$$\begin{bmatrix} | & | & & | \\ A_1 & A_2 & \cdots & A_n \\ | & | & & | \end{bmatrix}$$

... where each A_j is a column matrix.

Definition 2.1.2 Square Matrices

A **square matrix** is a matrix of size $n \times n$, as in the number of rows equals the number of columns.

Square matrices end up being used more often than non-square matrices, and have their own set of special properties which we will spend a lot of time discussing.

Definition 2.1.3 Vectors

A **column vector** (usually just called a **vector**) is a $m \times 1$ column matrix, as in a vector $\vec{v}$:

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

x_i is used to denote the i^{th} element of a vector.

Definition 2.1.4 The Zero Matrix

The **zero matrix** (usually just denoted as 0) is the $m \times n$ where each $0_{ij} = 0$. The size of the zero matrix can usually be decuded from other parts of the expression. The zero matrix is also called the additive identity.

Definition 2.1.5 Equality of Matrices

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $m \times n$ matrices. Then say $A = B$ if $a_{ij} = b_{ij}$ for *all* $1 \leq i \leq m$ and $1 \leq j \leq n$.

For two matrices to be equal they need to have the same dimensions and the same values in each entry.

2.1.1 Addition of Matrices

Definition 2.1.6 Addition of Matrices

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $m \times n$ matrices. Then say $A + B = C$ where $C = [c_{ij}]$ is an $m \times n$ matrix where:

$$c_{ij} = a_{ij} + b_{ij}$$

... alternatively:

$$(A + B)_{ij} = (A)_{ij} + (B)_{ij}$$

... for *each* $1 \leq i \leq m$ and $1 \leq j \leq n$.

To sum two matrices, they must be the same size. From there just add together each corresponding position's value to get the value of the resulting matrix's entry in that same position.

Example 2.1.1 Adding Matrices

Add the following matrices:

$$\begin{bmatrix} 1 & 4 & -4 \\ 5 & -2 & 0 \\ 2 & -1 & -10 \end{bmatrix} + \begin{bmatrix} 5 & -2 & 10 \\ 4 & 0 & 1 \\ -3 & -1 & 5 \end{bmatrix}$$

Just apply *Definition 2.1.4*:

$$\begin{bmatrix} 1 & 4 & -4 \\ 5 & -2 & 0 \\ 2 & -1 & -10 \end{bmatrix} + \begin{bmatrix} 5 & -2 & 10 \\ 4 & 0 & 1 \\ -3 & -1 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 1+5 & 4-2 & -4+10 \\ 5+4 & -2+0 & 0+1 \\ 2-3 & -1-1 & -10+5 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 2 & 6 \\ 9 & -2 & 1 \\ -1 & -2 & -5 \end{bmatrix}$$

Addition of matrices has very similar properties to addition of regular real numbers, mostly because the computational process to add matrices uses exclusively real number addition.

Theorem 2.1.1 Properties of Matrix Addition

Let $A, B,$ and C be $m \times n$ matrices, then:

- Commutativity:
$$A + B = B + A$$
- Associativity:
$$(A + B) + C = A + (B + C)$$
- Existency of an Additive Identity:
$$\text{There exists some } 0 \text{ such that: } A + 0 = A$$
- Existence of an Additive Inverse:
$$\text{There exists some } -A \text{ such that: } A + (-A) = 0$$

This may be your first experience with seeing the properties of addition written out formally, while it may seem overwhelming, these properties are very simple to understand and are in most cases obvious. This will not always be true however, and so it is important to formally define all the properties of the mathematical concepts we will be working with in this course.

You may also be wondering what is meant by $-A$. Is this the same as $-1 \times A$? How do you multiply a matrix by a scalar? These questions will be answered in the next subsection.

2.1.2 Scalar Multiplication of Matrices

Scalar multiplication of matrices is multiplying a matrix by a real number, which effectively *scales* up or down the values inside the matrix.

Definition 2.1.7 Scalar Multiplication of Matrices

Let $A = [a_{ij}]$ be an $m \times n$ matrix and $k \in \mathbb{R}$ be a scalar. Then say $kA = B$ where $B = [b_{ij}]$ is an $m \times n$ matrix where:

$$b_{ij} = ka_{ij}$$

... alternatively:

$$(kA)_{ij} = k(A)_{ij}$$

... for *each* $1 \le i \le m$ and $1 \le j \le n$.

Essentially, to scale a matrix, just multiply every entry in the matrix by that scalar.

Example 2.1.2 Scaling Matrices

Find kA where:

$$A = \begin{bmatrix} 1 & 4 & -4 \\ 5 & -2 & 0 \\ 2 & -1 & -10 \end{bmatrix}$$

$$k = 3$$

Applying *Definition 2.1.7*:

$$3 \begin{bmatrix} 1 & 4 & -4 \\ 5 & -2 & 0 \\ 2 & -1 & -10 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 12 & -12 \\ 15 & -6 & 0 \\ 6 & -3 & -30 \end{bmatrix}$$

Now let's look at the properties of scalar multiplication of matrices:

Theorem 2.1.2 Properties of Matrix Scaling

Let A and B be $m \times n$ matrices, and $k, p \in \mathbb{R}$ then:

- Distributive Law over Matrix Addition:

$$k(A + B) = kA + kB$$

- Distributive Law over Scalar Addition:

$$(k + p)A = kA + kA$$

- Associative Law for Scalar Multiplication:

$$k(pA) = (kp)A$$

- Multiplicative Identity:

$$1A = A$$

2.1.3 Multiplication of Matrices

This subsection covers multiplying matrices, to begin we learn how to multiply a matrix by a vector, and then how to multiply a matrix by a matrix, which can be simplified to repeated multiplication by vectors. The geometric reasoning for this definition of matrix multiplication will be covered in detail in *Chapter 5*, for now we are just trying to learn the computation.

To begin, note that matrix multiplication has an important restriction, that is when multiplying an $m \times n$ matrix with a $p \times q$ matrix, $n = p$ must be true. That is the number of columns of the first matrix must be equal to the number of rows of the second matrix. The resulting matrix will then be $m \times q$. Succinctly:

$$A[m \times n]B[n \times p] = C[m \times p]$$

or:

$$(\mathbf{m} \times n) \times (n \times \mathbf{p}) = (\mathbf{m} \times \mathbf{p})$$

If two matrices pass this criteria, then say they are **conformable**.

Definition 2.1.8 Matrix-Vector Multiplication

Let $A = [a_{ij}]$ be an $m \times n$ matrix and let X be an $n \times 1$ vector, as in:

$$\begin{bmatrix} | & | & & | \\ A_1 & A_2 & \cdots & A_n \\ | & | & & | \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Then AX is $m \times 1$ and:

$$AX = x_1 \begin{bmatrix} | \\ A_1 \\ | \end{bmatrix} + x_2 \begin{bmatrix} | \\ A_2 \\ | \end{bmatrix} + \cdots + x_n \begin{bmatrix} | \\ A_n \\ | \end{bmatrix}$$

... or alternatively:

$$AX = \sum_{j=1}^{n} \left(x_j \begin{bmatrix} | \\ A_j \\ | \end{bmatrix} \right)$$

Essentially, to multiply a matrix with a vector, scale each column of the matrix with its corresponding entry from the vector, and then add all those scaled columns back up into a single vector.

Very importantly, the result of a matrix-vector multiplication is *always* also a vector itself.

Example 2.1.3 Matrix-Vector Multiplication

Find AX where:

$$A = \begin{bmatrix} 1 & 4 & -4 \\ 5 & -2 & 0 \\ 2 & -1 & -10 \end{bmatrix}$$

$$X = \begin{bmatrix} 5 \\ 0 \\ 2 \end{bmatrix}$$

Applying *Definition 2.1.8*:

$$AX = \begin{bmatrix} 1 & 4 & -4 \\ 5 & -2 & 0 \\ 2 & -1 & -10 \end{bmatrix} \begin{bmatrix} 5 \\ 0 \\ 2 \end{bmatrix}$$

$$= 5 \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 4 \\ -2 \\ -1 \end{bmatrix} + 2 \begin{bmatrix} -4 \\ 0 \\ -10 \end{bmatrix}$$

$$= 5 \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} -4 \\ 0 \\ -10 \end{bmatrix}$$

Now applying *Definition 2.1.7*

$$= \begin{bmatrix} 5 \\ 25 \\ 10 \end{bmatrix} + \begin{bmatrix} -8 \\ 0 \\ -20 \end{bmatrix}$$

Now applying *Definition 2.1.6*

$$= \begin{bmatrix} -3 \\ 25 \\ -10 \end{bmatrix}$$

Before we move on to matrix-matrix multiplication, lets investigate the following example:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Let's multiply out the left side of this equation:

$$x \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + y \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\begin{bmatrix} a_{11}x + a_{12}y \\ a_{21}x + a_{22}y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Which is equivalent to:

$$a_{11}x + a_{12}y = b_1$$

$$a_{21}x + a_{22}y = b_2$$

Evidently, this is a system of 2 equations in 2 unknowns in equation form! Something we are very familiar with. At every point in this process the equation was a representation of the SoLE.

The following is the **Vector Form of a SoLE**:

$$x \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + y \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

The following is the **Matrix Form of a SoLE**:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Also written as:

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

... or more generally the definition for these forms are on the next page.

We now have a new perspective on a SoLE, that is we are looking for some $\vec{x}$ that when multiplied by the matrix A, you get $\vec{b}$ as a result. All the possible $\vec{x}$ that satisfy this comprise the solution set of our SoLE. While *Gauss-Jordan Elimination* may still be a faster way to actually compute the solutions to these systems, broadening our perspective will allow us to solve them in clever ways.

Definition 2.1.9 Matrix & Vector Form of a SoLE

Suppose we have a SoLE given by:

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

... then we can express this system in **vector form** as follows:

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

... and we can express this system in **matrix form** as follows:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

... which is also written as:

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

Now we begin matrix-matrix multiplication with it's definition:

Definition 2.1.10 Matrix-Matrix Multiplication

Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix, in the form:

$$A = \begin{bmatrix} | & | & & | \\ A_1 & A_2 & \cdots & A_n \\ | & | & & | \end{bmatrix} \text{ and } B = \begin{bmatrix} | & | & & | \\ B_1 & B_2 & \cdots & B_p \\ | & | & & | \end{bmatrix}$$

... then the $m \times p$ matrix AB is defined as follows:

$$AB = \begin{bmatrix} | & | & & | \\ AB_1 & AB_2 & \cdots & AB_p \\ | & | & & | \end{bmatrix}$$

Essentially, to multiply two matrices, multiply the first column of the second matrix with the first matrix, that resulting vector is the first column of the product, repeat for all the columns in the second matrix.

Example 2.1.4 Matrix-Matrix Multiplication

Find AB where:

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

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

First of all these matrices are conformable as A has 3 columns and B has 3 rows. Applying *Definition 2.1.10* we get:

$$AB = U = \begin{bmatrix} | & | & | \\ U_1 & U_2 & U_3 \\ | & | & | \end{bmatrix}$$

Where:

$$U_1 = AB_1 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

$$U_2 = AB_2 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \end{bmatrix}$$

$$U_3 = AB_3 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

Therefore:

$$AB = \begin{bmatrix} -1 & 9 & 3 \\ -2 & 7 & 3 \end{bmatrix}$$

2.1.4 The ij-th Entry of a Matrix Product

Addition and scalar multiplication were both defined in terms of their ij-th entry, and the same can be done for multiplication however it takes more work to do so.
In order to derive the formula for the ij-th entry, we will use two 3×3 matrices, however this process would work for the product of any two conformable matrices. Consider the following product:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix}$$

Let's multiply out the left side to find an expression for any entry on the right:

$$U_1 = AB_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = b_{11} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + b_{21} \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} + b_{31} \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} \end{bmatrix} = \begin{bmatrix} u_{11} \\ u_{21} \\ u_{31} \end{bmatrix}$$

$$U_2 = AB_2 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} = b_{12} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + b_{22} \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} + b_{32} \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \\ a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} \end{bmatrix} = \begin{bmatrix} u_{12} \\ u_{22} \\ u_{32} \end{bmatrix}$$

$$U_3 = AB_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{13} \\ b_{23} \\ b_{33} \end{bmatrix} = b_{13} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + b_{23} \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} + b_{33} \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} \\ a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{bmatrix} = \begin{bmatrix} u_{13} \\ u_{23} \\ u_{33} \end{bmatrix}$$

From this we learn the value of each entry in the product:

$$\begin{bmatrix} a_{11}b_{11}+a_{12}b_{21}+a_{13}b_{31} & a_{11}b_{12}+a_{12}b_{22}+a_{13}b_{32} & a_{11}b_{13}+a_{12}b_{23}+a_{13}b_{33} \\ a_{21}b_{11}+a_{22}b_{21}+a_{23}b_{31} & a_{21}b_{12}+a_{22}b_{22}+a_{23}b_{32} & a_{21}b_{13}+a_{22}b_{23}+a_{23}b_{33} \\ a_{31}b_{11}+a_{32}b_{21}+a_{33}b_{31} & a_{31}b_{12}+a_{32}b_{22}+a_{33}b_{32} & a_{31}b_{13}+a_{32}b_{23}+a_{33}b_{33} \end{bmatrix}$$

$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix}$$

Let's focus in on the $1, 2$ entry for example:

$$u_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}$$

Notice the middle indexes of each term are just incrementing by one each time, so:

$$u_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} = \sum_{k=1}^{3} a_{1k}b_{k2}$$

... here, 3 comes from the fact that A has 3 columns, equivalently B has 3 rows. More generally for u_{ij} in an $(m \times n) \times (n \times p)$ product that would be:

$$u_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

Definition 2.1.11 The ij-th Entry of a Product

Let $A = [a_{ij}]$ be an $m \times n$ matrix and let $B = [b_{ij}]$ be an $n \times p$ matrix, then AB is an $m \times p$ matrix with ij-th entry of:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{ki}$$

... alternatively:

$$(AB)_{ij} = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}$$

Essentially, to find the ij-th entry of the product take the i-th row of the left matrix, the j-th row of the right matrix, and then multiply the first entries, the second entries, and so forth, and then add up all the results.
Let's use this to actually compute an entire matrix product in the next example.

Example 2.1.5 *ij-th Entry*

Find AB where:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \end{bmatrix}$$

... using only the ij-th product.

Applying *Definition 2.1.11*:

$$\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot 7 & 1 \cdot 3 + 2 \cdot 6 & 1 \cdot 1 + 2 \cdot 2 \\ 3 \cdot 2 + 1 \cdot 7 & 3 \cdot 3 + 1 \cdot 6 & 3 \cdot 1 + 1 \cdot 2 \\ 2 \cdot 2 + 6 \cdot 7 & 2 \cdot 3 + 6 \cdot 6 & 2 \cdot 1 + 6 \cdot 2 \end{bmatrix}$$

$$= \begin{bmatrix} 16 & 15 & 4 \\ 13 & 15 & 5 \\ 46 & 42 & 14 \end{bmatrix}$$

2.1.5 Properties of Matrix Multiplication

One of the most important differences between matrices and regular real number multiplication is that while $xy = yx$ for $x, y \in \mathbb{R}$, $AB \neq BA$ for two matrices A and B (note here that I could have written for $A, B \in \mathbb{M}_{m \times n}$ as $\mathbb{M}_{m \times n}$ is the set of all $m \times n$ matrices). The geometric reasoning for this will be explored in *Chapter 5*, for now just take it as a fact and try it out with some matrices. While this is technically not a *property*, it is more of the absence of the property of commutativity, I will include it in the formal properties of matrices in these notes, as in the following theorem.

Theorem 2.1.3 Properties of Matrix Multiplication

Let A, B and C be conformable matrices, and $k, p \in \mathbb{R}$ then:

- Non-Commutativity:

$$AB \neq BA$$

Except in special cases.

- Associative Law for Multiplication:

$$A(BC) = (AB)C$$

- Distributive Law over Matrix Addition:

$$A(kB + pC) = kAB + pAC$$

$$(kB + pC)A = kBA + pCA$$

2.1.6 The Transpose

The transpose is another operation you can perform on a matrix that is unique to matrices, and has its own set of properties.

Definition 2.1.12 The Transpose of a Matrix

Let $A = [a_{ij}]$ be an $m \times n$ matrix. Then A^T called the **transpose** of A is an $n \times m$ matrix given by:

$$A^T = [a_{ij}]^T = [a_{ji}]$$

Essentially, row i from A becomes column i in A^T, and column j in A becomes row j in A^T. Another way to think of this is you reflect the matrix along its main diagonal that goes from the top left to the bottom right. By this definition, the value on the main diagonal are unaffected by the transpose.

Example 2.1.6 The Transpose

Find A^T where:

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

Applying *Definition 2.1.12*:

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

Theorem 2.1.4 Properties of Transposition

Let A and B be conformable matrices, and $k, p \in \mathbb{R}$ then:

- Transpose of a Transpose:

$$(A^T)^T = A$$

- Transposition of a Product:

$$(AB)^T = B^T A^T$$

- Transposition of a Sum:

$$(rA + sB)^T = rA^T + sB^T$$

Definition 2.1.13 Symmetric Matrices

Let $A = [a_{ij}]$ be an $m \times n$ matrix. Then say A is **symmetric** if

$$A = A^T$$

The following matrix is symmetric:

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

Definition 2.1.14 Skew-Symmetric Matrices

Let $A = [a_{ij}]$ be an $m \times n$ matrix. Then say A is **skew-symmetric** if

$$A = -A^T$$

... or equivalently:

$$-A = A^T$$

The following matrix is skew-symmetric:

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

2.1.7 The Identity and Inverses

Definition 2.1.15 Identity Matrices

The $n \times n$ **identity matrix** usually denoted I_n is the square matrix with ones along the main diagonal, and zeros everywhere else. For example:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The *identity matrix* is also called the **multiplicative identity.**

Let's see what happens if we multiply by the identity matrix. First we multiply on the right (recall that it matters what side you multiply on for matrices) by the appropriate sized *identity matrix* to make them conformable.

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + 0 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} & 0 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + 1 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

As a result we get our original matrix back! Let's try to multiply now on the other side by its appropriate identity matrix.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + a_{21} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + a_{31} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} & a_{12} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + a_{22} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + a_{32} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

We once again get our original matrix back out to us as the product. This leads to the following theorem.

Theorem 2.1.5 Multiplication by I_n

Let $A = [a_{ij}]$ be an $m \times n$ matrix. Then:

$$AI_n = A \text{ and } I_m A = A$$

The identity matrix works a lot like the number 1 in regular real number multiplication.

We will now discuss matrix inverses, which is essentially how you divide matrices. Take the real numbers for example, consider:

$$2 \cdot \frac{1}{2} = 2 \cdot 2^{-1} = 1$$

We can say that the multiplicative inverse of 2 is $\frac{1}{2}$ since when you multiply by it the result is the *multiplicative identity* (1). By this logic, dividing by some number is the same as multiplying by its inverse. This is how we divide with matrices, we multiply by a matrix inverse.

Definition 2.1.16 Inverse Matrices

Let A be an $n \times n$ matrix. A is said to have an **inverse** A^{-1} if and only if:

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

... if A^{-1} exists, say A is **invertible**. Furthermore, given that A^{-1} exists, there is only one such matrix which satisies the definition for a matrix A.

Importantly, only square matrices can have inverses. There are a few reasons for this, one important one comes from *Chapter 3*, however just consider the fact that the inverse must be conformable on *both* sides with A, this is only possible if A is a square matrix. It is possible for non-square matrices to have one sided inverses, as in a matrix B^{-1} that when multiplied by B (a non-square matrix) one one side results in I_n but when multiplied on the other side is non-conformable. These one sided inverses are **not** inverses, and are usually not considered at all.

Example 2.1.7 Matrix Inverse

Solve for the matrix A in:

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

Given that:

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

To solve this problem, we are going to multiply both sides of the equation by the inverse of the matrix behind the A. Just like you would do if you were solving $2x = 4$, where you would multiply both sides by the inverse of 2. It *does* matter if left multiply or right multiply, we have to multiply on the right in order for us to get $BB^{-1} = I_2$. Another important note is that not all matrices are invertible, and just because a matrix is square does **not** mean it's invertible.

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

In this question we were given the inverse of that matrix, and so we can substitute that in as follows (also note once again that $BB^{-1} = I_2$):

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

The left side will always multiply out to be A, just like $x \cdot 1$ will always multiply out to be x regardless of x, and the right side just needs to be multiplied out.

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

2.1.8 Finding the Inverse of a Matrix

This subsection is focused defining a method to solve for the inverse of a matrix, as well as properties of the inverse operation and how it interacts with other operations. I will begin this section with the definition of the procedure, and then motivate it with an explanation entirely in the following subsection.

Definition 2.1.17 Inverse Matrix Algorithm

Let A be an $n \times n$ matrix. To find A^{-1} create an augmented matrix in the form:

$$[A \mid I_n]$$

If possible, through row operations transform the matrix to the form:

$$[I_n \mid B]$$

If this is possible, then $B = A^{-1}$, and call A **invertible**. If it is not possible to transform the matrix to the final form then A has no inverse.

Essentially, make the augmented matrix as first defined, then do *Gauss-Jordan Elimination* to get the matrix into *RREF* form. If the *RREF* form of the matrix has a row of zeros, this would mean the matrix is non-invertible. If the RREF form of the matrix is just the identity matrix on the left, then the matrix after the augmentation bar is the inverse you are looking for.

This is also a good time to formally try to understand what the augmentation bar is. We used it in *Chapter 1* as just a notational trick and it should be clear that that is exactly that it is. The matrix:

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

... is essentially equivalent to:

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

By putting the augmentation bar, we are essentially denoting one side as the *left* side and one side as the *right* side, and that's really it. The exact meaning of the left side and right side are defined based on context. Like how in this algorithm the left side of the result is the identity matrix, while the right side is some matrix B.

Once again, the reasoning for this method will be explained at the end of the next section, for now here is an example of it in use:

Example 2.1.8 Matrix Inverse

Find A^{-1} where:

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

By using the method from *Definition 2.1.17*:

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

$\Downarrow$ Through Elementary Row Operations... $\Downarrow$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ 0 & 1 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{array}\right]$$

... which tells us that:

$$A^{-1} = \begin{bmatrix} -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{bmatrix}$$

We can find the solutions to a SoLE using an inverse matrix. This is because in general the matrix form of a SoLE looks like:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

... which is also written as:

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

... multiplying on the left by A^{-1} we get:

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

$$I_n\vec{x} = A^{-1}\vec{b}$$

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

Which tells us that to find our solutions to the linear system ($\vec{x}$) we just invert the coefficient matrix (A^{-1}) and multiply it on the left by the constant column vector ($\vec{b}$). This is particularly useful if you need to solve the same system multiple times changing only the constant column, in this situation you can reuse the inverse coefficient matrix. Let's illustrate this method with the following example:

Example 2.1.9 Solving a SoLE Using Matrix Inverse

Find the solution(s) to the following consistent SoLE:

$$x + z = 1$$

$$x - y + z = 3$$

$$x + y - z = 2$$

We can write this system in the form:

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

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

We can then multiply on the left by the inverse of the coefficient matrix on both sides:

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

Now let's calculate A^{-1} using the method described earlier in this section:

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

$\Downarrow$ Through Elementary Row Operations... $\Downarrow$

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

... which tells us that:

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

So finally the can compute the following:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{5}{2} \\ -2 \\ -\frac{3}{2} \end{bmatrix}$$

Therefore this system has a unique solution $x = \frac{5}{2}$, $y = -2$, $z = -\frac{3}{2}$.

Now consider some theorems and properties of the inverse operation:

Theorem 2.1.6 Properties of the Matrix Inverse

Let A and B be $n \times n$ matrices.

- I is invertible and $I^{-1} = I$
- If A is invertible then so is A^{-1} and $(A^{-1})^{-1} = A$
- If A is invertible than so is A^k, and $(A^k)^{-1} = (A^{-1})^k$
- If A is invertible and p is a nonzero real number, then pA is invertible and $(pA)^{-1} = \frac{1}{p}A^{-1}$
- If A is invertible, then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$
- If A and B are both invertible, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$. This can be extended to the product of n matrices, as in the next point.
- If $A_1, A_2, \cdots, A_k$ are invertible, then the product $A_1A_2 \cdots A_k$ is invertible and:

$$(A_1A_2 \cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1} \cdots A_2^{-1}A_1^{-1}$$

Let's take a moment to discuss the last two points in these properties. Let's say we have the following product:

$$ABC$$

... where $A, B,$ and C are all $n \times n$ invertible matrices. Then we are looking for some matrix D that will make the following two equations true:

$$DABC = I$$

$$ABCD = I$$

Since we now our original three equations are invertible, namely we know A and C are invertible we can make the following changes:

$$DABCC^{-1} = IC^{-1}$$

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

... which simplifies to (only showing this step once):

$$DABI_n = C^{-1}$$

$$I_nBCD = A^{-1}$$

... or equivalently:

$$DAB = C^{-1}$$

$$BCD = A^{-1}$$

We can repeat this for the other two matrices:

$$DABB^{-1} = C^{-1}B^{-1}$$

$$B^{-1}BCD = B^{-1}A^{-1}$$

... which becomes:

$$DA = C^{-1}B^{-1}$$

$$CD = B^{-1}A^{-1}$$

... then one last time:

$$DAA^{-1} = C^{-1}B^{-1}A^{-1}$$

$$C^{-1}CD = C^{-1}B^{-1}A^{-1}$$

... which finally gives us:

$$D = C^{-1}B^{-1}A^{-1}$$

$$D = C^{-1}B^{-1}A^{-1}$$

Evidently there is only one solution for D and this can be easily extended to a product of n equations. Since multiplying by D on either side of ABC results in I_n, by definition $D = (ABC)^{-1}$.

One final note before we end this subsection is the following formula for the inverse of a 2×2 matrix. The general form of this formula for $n \times n$ matrices comes in *Chapter 3* as it takes quite a bit of motivation, for now this is a worthwhile formula to memorize:

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

Note that this matrix is invertible so long as $ad - bc \neq 0$.

2.1.9 Elementary Matrices

This subsection connects what we were doing in *Chapter 1* to the new computation and notation we know from *Chapter 2*.

Definition 2.1.18 Elementary Matrices

Let E be an $n \times n$ matrix. Then say E is an **elementary matrix** if it is the result of applying a single row operations to I_n. Performing any row operation defined by E on some matrix A is equivalent to taking the product EA.

Let P^{ij} denote the elementary matrix which involves switching the i^{th} and j^{th} rows of I_n. Then P^{ij} is also called a permutation matrix and:

$$P^{ij}A = B$$

where B is obtained from A by switching the i^{th} and j^{th} rows.

Let $E(k,i)$ denote the elementary matrix corresponding to the row operation in which the i^{th} row of I_n is multiplied by the nonzero scalar k. Then

$$E(k,i)A = B$$

where B is obtained from A by multiplying the i^{th} row of A by k.

Let $E(k \times i + j)$ denote the elementary matrix obtained from I_n by replacing the j^{th} row with itself plus k times the i^{th} row. Then

$$E(k \times i + j)A = B$$

where B is obtained from A by replacing the j^{th} for with itself plus k times the i^{th} row.

Let's look at some examples of elementary matrices (for $n = 3$):

$$P^{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

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

$$E(-2 \times 3 + 1) = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Very importantly, doing a row operation on a matrix is the same as multiplying by the corresponding elementary matrix. Look at the following for an example let:

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

Recall $E(-2\times 3+1)$ is the elementary matrix where the first row is replaced by -2 times the third row, this is a type 3 row operation. Doing this row operation on A which I will call A' would give us:

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

So let's verify that:

$$E(-2 \times 3 + 1)A = A'$$

$$\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 9 \\ 0 & 1 & -1 \\ 5 & 0 & 2 \end{bmatrix} = A'$$

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

Which is true, so we can see that left-multiplying by an elementary matrix has the same affect on the matrix it is being multiplied with as doing the corresponding row operation.

Another important fact is about the inverses of elementary matrices:

Theorem 2.1.7 Elementary Matrix Inverses

Every elementary matrix (E) is invertible and its inverse (E^{-1}) is also an elementary matrix obtained by doing the elementary row operation which undoes the effect of E.

The following discussion will allow us to use elementary matrices to help our understanding of linear algebra as a whole.

Theorem 2.1.8 $B = UA$ Form

Let A be and $m \times n$ matrix and let B be the RREF form of A. Then we can say $B = UA$ where U is the product of all the elementary matrices representing the row operations done to A to obtain B.

This form encodes all the information about the original matrix A, what it's RREF form looks like in B, and all the row operations needed to get to the RREF form in U. One way to think about this form is instead of $B = UA$, is to think of an equivalent expression:

$$A = U^{-1}B$$

This states that a matrix A can be written (essentially factored into) its RREF form B and the inverse of all the steps it takes to get it into that RREF form U^{-1}. In this form the matrix U is really best thought of as a whole string of elementary matrices multiplying together. Let's take a look at the following example to try to understand this form of a matrix.

Example 2.1.10 $B = UA$ form of a Matrix

For the following matrix A, write A in the form $B = UA$:

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

To do this, we need to keep track of the row operations required to bring this matrix (A) to its RREF form (B), then at the end convert all of those row operations into elementary matrices, and then compute their ordered product (U). For the sake of clarity I will go through all the steps of *Gauss-Jordan Elimination*:

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

$$\boxed{R_2 \to R_2 - 2R_1}$$

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

$$\boxed{R_3 \to R_3 - 3R_1}$$

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

$$\boxed{\left(\frac{1}{3}\right) R_2}$$

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

$$\boxed{R_1 \to R_1 - 2R_2}$$

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

$$\boxed{R_3 \to R_3 + 6R_2}$$

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

Now that we have the matrix in RREF form, the following product would get us U:

$$U = E(6 \times 2 + 3)\ E(-2 \times 2 + 1)\ E\left(\frac{1}{3}, 2\right)\ E(-3 \times 1 + 3)\ E(-2 \times 1 + 2)$$

Which is just the list of row operations we did in the form of elementary matrices, with the first one on the far right since it is the one we want to apply first. We can write out all these matrices:

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

We can verify that this is truly the matrix U in $B = UA$ my left-multiplying A by it and checking if we get the RREF form of A as a product.

$$B = UA$$

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

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

We can see that our solution is correct and so:

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

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

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

This process is fairly long and tedious, and can be simplified. Essentially it would be nice if at the same time as we did the row operations to the original matrix to get it into RREF we could automatically keep an updated product of elementary matrices to get U. We can achieve this by augmenting the matrix A with I_n, written as $[A \mid I_n]$. When we do a elementary row operation to this matrix it essentially does it to both A and I_n. We can use operations to reduce A to it's RREF form (B), and then after all those operations the matrix I_n will effectively be the collection of all row operation that it took to get from A to be B, this is what we want in U. To summarize:

$$[A \mid I_n] \to [B \mid U]$$

Notice that this process is very similar to *Definition 2.1.17: Inverse Matrix Algorithm*. The difference in that procedure is that we change the name of U to A^{-1}, assuming that the *RREF* form of A (B) is I_n (the identity matrix). This works because of the following:

$$B = UA$$

If A is invertible, then $B = I_n$:

$$I_n = UA$$

$$U^{-1} I_n = A$$

$$U^{-1} = A$$

$$(U^{-1})^{-1} = A^{-1}$$

By property 2 of *Theorem 2.1.6*:

$$U = A^{-1}$$

Which tells us that if $B = I_n$, then $A^{-1} = U$ which is why we use the same process to find the inverse of a matrix. If $B \neq I_n$, then A is not invertible.

This particularly works for finding the inverse because in making the statement:

$$DA = I$$

... we know that at least by multiplying on the left, D acts as A^{-1} (more on this soon). Essentially we want to know what string of elementary row operations will get A to its RREF form (B). This sounds like the $B = UA$ form but in this particular case $B = I_n$ and $U = A^{-1}$ by definition since left-multiplying A by it results in I_n. Importantly by definition the inverse needs to be able to be multiplied on *either* side and result in I_n to be the inverse matrix, however we don't have to check that the other side's product results in I_n due to the following theorem:

Theorem 2.1.9 Uniqueness of the Inverse

Suppose A and B are square $n \times n$ matrices such that $AB = I_n$. Then it follows that $BA = I_n$.

This theorem essentially states that if you find a matrix that acts as the inverse on one side, and both matrices are square, then you do not need to check the other side's product to verify it also results in I_n, it will with certainty.

This concludes this section of *Chapter 2*, you can now perform most operations on matrices of any size, and compute things like the inverse and $B = UA$ form.

2.2 LU Factorization

This section introduces the factorization of a matrix into its *LU* form. Not every matrix A has a factorization in this form, but when it does, the system $A\vec{x} = \vec{b}$ can be solved using the *LU* factorization in half as many steps as *Gauss-Jordan Elimination*. *LU* factorization is the most common way that computers solve SoLEs. Before we define the *factors LU*, we need to define two types of matrices:

Definition 2.2.1 Triangular Matrices

A square matrix is called **lower triangular** if all the entries *above* the main diagonal are 0, as in the only non-zero entries are on the *lower* half of the matrix including the diagonal. A square matrix is called **upper triangular** if all the entries *below* the main diagonal are 0, as in the only non-zero entries are on the *upper* half of the matrix including the diagonal.

Definition 2.2.2 *LU* Factorization

An $m \times n$ matrix A has an LU factorization if the REF form of A can be acheived from A without switching any rows. The LU factorization of a matrix A involves writing the given matrix as the product of a lower triangular matrix (L) which has a main diagonal consisting entirely of ones, and an upper triangular matrix (U). The output of LU factorization is the following equation:

$$A = LU$$

As in for some 2×2 matrix for example, we are looking to see if we can write it in the form:

$$\begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$$

... which would be it's respective LU form. This idea of course extends to an $n \times n$ matrix. You can effectively do this factorization by taking the product of the general form of the LU factorization,as in:

$$\begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} = \begin{bmatrix} a & b \\ xa & xb + c \end{bmatrix} = LU$$

... and setting this equal to the given matrix, let's use a general 2×2 matrix:

$$M = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix}$$

... from here we can match terms:

$$\begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} = \begin{bmatrix} a & b \\ xa & xb + c \end{bmatrix}$$

Matching the corresponding elements we get the following:

$$m_1 = a$$

$$m_2 = b$$

$$m_3 = xa$$

$$m_4 = xb + c$$

... which is a system of equations in 5 variables (remember the m_i values are all constants). You can then solve this system for the values we need using substitution, or any other technique (note this is **not** a SoLE).

Now this process is lengthy for any matrix with dimensions bigger than 2×2 and also ends in a non-linear system. For this reason we almost never find the LU factorization by using this method, instead we use the *Multiplier Method*.

To begin deriving the multiplier method, note that the REF form of any matrix A is already upper triangular, and so naturally it will be the U in the LU factorization. Say you have some matrix A then:

$$A = IA$$

Here, I can take on whatever size it needs to be *conformable* with the product. Then the series of k elementary operations $(E_1 E_2 \cdots E_k)$ needed to get from A to its REF form can be written in the from:

$$E = E_1 E_2 \cdots E_k$$

... and we know that:

$$EA = B$$

... where B is the REF form of A. Except in this factorization we do not particularly need the REF form, we want almost REF form except you do not multiply the leading entries by their multiplicative inverses to make all the leading entries equal to 1. This means that the *only* elementary row operation used in this method will be of type 3. I will still be referring to this matrix as B. Then,

$$A = IA$$
$$E^{-1}EA = E^{-1}EIA$$

... where E^{-1} is the string of opposite row operations that invert E. The identity matrix can commute, and so:

$$IA = E^{-1}IEA$$

$$A = E^{-1}B$$

Note that B will be upper triangular as it is *almost* in RREF form, and E^{-1} will be lower triangular with a main diagonal of ones since we only used type 3 elementary row operations. Therefore $A = LU$ where $L = E^{-1}$ and $U = B$.

Let's do an example of this to solidify this process:

Example 2.1.11 LU Factorization

Find the LU factorization of the following matrix:

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

To do this process, begin with:

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

Now we take the right matrix to its almost-REF form using only type 3 row operations. Each row operation can be represented as E, and thus to keep the equality true we apply the *inverse* of the same row operation to the identity matrix.

$$\boxed{R_2 \to R_2 - 4R_1}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 1 & 2 & 3 & 0 \end{bmatrix}$$

$$\boxed{R_3 \to R_3 - R_1}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

... and so we have the original matrix equal to a lower triangular matrix with a main diagonal of all zeros, and an upper triangular matrix. I want to note here that you can use type 2 row operations (multiplying by a scalar) in this process, as long as in the end you invert all those type 2 operations so that the main diagonal of the L matrix is all ones.

Now let's see how we can use this to solve systems of equations. Given some system in the form:

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

Recall our goal is to solve for $\vec{x}$. If we can can find the LU factorization of A then:

$$LU\vec{x} = \vec{b}$$

... and so typically we denote:

$$U\vec{x} = \vec{y}$$

... which is also a SoLE. First we need to compute $\vec{y}$ using another new system created by the substitution:

$$L\vec{y} = \vec{b}$$

... whose solutions for $\vec{y}$ can be computed very fast due to the form of L. Note that $\vec{y}$ will always have a unique solution. That vector $\vec{y}$ can be substituted back into the equation $U\vec{x} = \vec{y}$, and we can essentially just read off the results for $\vec{x}$ due to the form of U. Let's use the matrix from our last example as the A matrix of a SoLE.

Example 2.1.12 Solving a SoLE using LU

Solve the system using LU factorization:

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Note that we know:

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

... and so the system can be changed into:

$$LU\vec{x} = \vec{b}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

We can break this system up using the following substitution which results in two systems:

$$U\vec{x} = \vec{y}$$

... which gives us:

$$L\vec{y} = \vec{b}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Which can be solved very quickly using an augmented matrix:

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

$\Downarrow$ Through Elementary Row Operations... $\Downarrow$

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

Which tells us that:

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

We can now solve for $\vec{x}$ in the system:

$$U\vec{x} = \vec{y}$$

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

... solving this system is already half-way complete since the matrix is already almost in REF. This yields us the solution:

$$x = -\frac{3}{5} + \frac{7}{5}t$$

$$y = \frac{9}{5} - \frac{11}{5}t$$

$$z = t$$

$$w = -1$$

Essentially the idea here is that we reduce the *Gaussian-Elimination* into two simpler cases of systems to solve, both of which are already partially solved.

This concludes the entirety of *Chapter 2*. After this chapter you have the skills to do all the main matrix operations, and you are also able to find the *LU* factorization of a matrix. This chapter also introduces important ideas such as relating the algebraic methods used to solve SoLEs in *Chapter 1* to matrix operations studied in *Chapter 2*.

3 Chapter 3: Determinants