Linear Algebra

Complete Notes and Transcript

Aaron Lengyel

Abstract

This is a combination of all of my understandings and findings in regards to linear algebra. The explanations are mine, and these are meant to act as transcripts for the video series on linear algebra I will be producing alongside. This covers everything from the basics to more advanced concepts near the end. There aren't any exercises, only examples and solutions. A **ton** is going to be covered in this series so don't be surprised if you do not capture or understand it all during one sitting. This is meant to be a comprehensive resource of as much of linear algebra as I could personally find, and there *will* be mistakes (which I hope to rectify along the way).

Contents

Introduction

Motivating Example

Amber is looking to buy some apple cider, bread, and cookies from the store. However she has limited money in her wallet and limited space in her car. She only has room in her car for 4 L of groceries, and she can only spend 13 dollars. Additionally she can't carry more than 10 kg of groceries. Given that each bottle of cider cost \$3, weighs 1 kg, and has a volume of 400 mL, each cookie costs \$2, weighs 0.5 kg, with a volume of 240 mL, each piece of bread cost \$5, has a volume of 1500 mL, and weighs 1 kg; how many groceries can she buy?

For those who know a little algebra, you might recognize that we can form a single equation for the cost of each of our groceries. Then we can write one for the volumes, and one for the weights.

$$\begin{aligned} 3a + 5b + 2c &= 13 \\ 0.4a + 1.5b + 0.25c &= 4 \\ 1a + 1b + 0.5c &= 10 \end{aligned}$$

But how would we go about solving this?

This is called a system of linear equations, and they've been known to humans for millenium. In fact the chinese wrote about how to solve systems like this in *The Nine Chapters of the Mathematical Art* written as far back as 200 BC.

What I hope to tackle in these upcoming videos is finding new ways to look at and understand linear equations and how to solve them. Later advancing to new notations and a deeper look into the extension of these ideas through the tools of **Linear Algebra**.

This is meant to be similar to 3B1B's linear algebra playlist. Except rather than solely teaching intuition I will be going through the mechanics and inner workings of matrices and vectors, discussing the algorithms and tools that we us to solve these. Hopefully through the lens of linear equations with various examples and explanations along the way. I'm hoping to cover as much ground as possible while also building a stronger and stronger understanding of these systems.

Most linear algebra courses will usually begin with linear equations and then quickly jump in to matrices. I am going to be doing something slightly different were I will save matrices till later. This is because matrices will make more sense when we have a stronger grasp of linear equations and the vector spaces that form them.

The structure of these videos will be as follows:

1. Linear Equations
   - We will begin discussing linear equations and develop a few tools and understanding surrounding them.
2. Vector Spaces
   - We will then move on to take a deeper look at vectors and their relation to general systems of equations.
3. Matrices
   - We will then look at matrices and how they relate to linear equations and vector spaces.
4. Additional Components
   - At this point we will discuss additional topics in linear algebra as well as some applications.

Outcomes

The outcomes of this booklet and series of videos are:

1. Strong Intuition for the fundamentals
   1. Connection between linear equations and determinants, matrices and vectors
   2. Understanding when to represent a problem as a linear system and solve with matrices
2. Understand a brief history of Linear Algebra and some of the work that's been done
   1. Pre-European history and Chinese origins
   2. European formulations
   3. Mathematicians and Important Time Periods
3. Understanding of the connections between Linear Algebra and other parts of mathematics
   1. Abstract Vector Spaces
   2. Linear Transformations
   3. Quadratic Forms
4. Identify **why** something works
   1. Why is RREF important/useful
   2. Where do determinants come from
   3. Why does the Gram-Schmidt Algorithm work/exist
   4. Why does the SVD exist
   5. Why can we represent a certain problem as a matrix operation
5. As well as a variety of advanced techniques
   1. Hilbert Spaces
   2. Algorithm Optimizations
      1. Linear-Least Squares (QR Decomposition)
      2. Computational Methods for Matrix solving
   3. Advanced Matrix Operations
      1. Kronecker
      2. Direct Sum
   4. Advanced Vector Space Operations
      1. Outer Product
      2. Wedge Product
   5. Basic Tensor Theory

History

Linear Systems of Equations

A simple Linear Equation

So to begin, what is a linear equation?

Well let's think about a 2D line in space. How can we describe this line algebraicly? Right away we can notice that the line crosses our y and x axis only once, and extends off to the edge of our plane and beyond on either side. We can also see that the line has unique x and y positions.

Let's take 2 points in our x axis and slightly increase them by a small factor. If you notice, the distance between our y values remains the same as it was in our original points. Additionally, there's nothing special here about the points we chose. No matter where we take these 2 x values, any changes in them will always result in thesame change in the corresponding y values since the line will never curve.

This is in contrast to what we can consider non-linear, where as we change our x, the y value will increase or decrease and the distance between two y values will vary throughout the entire line.

So given what we now know from quickly analysing the geometry we can come to form some algebraic description of our line. To do this we're going to have to create some reference for our line to begin, to make things simpler I'll draw in some numbers, and we'll say that our line will start when $x = 0$. Which in this case is 3. We now take the factor that governs our change in y for every change in x. This will be our slope and we can find it by taking the ratio of the change in both y and x. This gives us a slope of 2. Now to verify that this is the equation of our line we can take any number in x and then verify that the corresponding y wil place us on our line.

We can do this same process for multiple lines to check if this works for all lines, but essentially every line we can draw in two dimensions will have this form:

$$y = ax + b$$

We could have also done this with $y = 0$ as our origin and we would have found the equation:

$$x = \frac{y}{2} - \frac{3}{2}$$

However if we really think about it, this is still in the same form of $y = ax + b$, since the variables we pick are kind of arbitrary.

Linearity

The property that graphically represents a straight line is called linearity. A relationship between two things is called linear if it satisfies the property of linearity. Which looks something like:

$$L(x + y) = L(x) + f(y)$$
$$L(kx) = kL(x)$$

Standard and General Form

Let's take our equation again:

$$y = ax + b$$

and adjust it so all our variables are on one side:

$$-ax + y = b$$

This is called the **standard form** for our equation.

If we adjust this some more:

$$-ax + y - b = 0$$

This is called the **general form** of our equation.

Both of these equations represent the same line, and we can say more generally that any equation that any relationship that can be written like this must represent a single line through space.

Multiple Variables

Let's take a moment to think about what each variable tells us about the dimension we're working with. A single x variable kind of represents a point in space doesn't it?

$$ax + b = 0$$

There is only one solution to this problem. And that's $x = -\frac{b}{a}$, so the space of solutions to this equation represents only a single point.

Now when we add a 2nd variable y our equation looks like this:

$$ax + by + c = 0$$

And we've already shown that the space that represents the solutions to these equations gives us a line.

Now let's think about what happens if we add one more variable z:

$$ax + by + cz + d = 0$$

$$z = -\frac{a}{c}x - \frac{b}{c}y - \frac{d}{c}$$

If we set both variables to 0, we find a single point $-\frac{d}{c}$, we also should keep in mind that we want we want our c to not be 0. If we start from this point, and start only with our x we'll find that we create a line. If we then set x as 0 and traverse y we find another line. Geometrically as we travel along both lines we essentially fill in a plane in 3D space. This equation in 3 variables represents a 2D plane.

We could also continue this for 4 variables.

$$ax + by + cz + dw + e = 0$$

However the same logic will apply and it's sort of difficult to visualize, but essentially we would have some 3 dimensional object embedded in 4D space being represented here. We can continue this pattern for even more variables as well!. Taking only this basic intuition that we've just discovered, that is every linear equation is formed from smaller singular linear equations representing lines. In essence it's anything with the following form:

$$a_1x_1 + ... + a_nx_n + b = 0$$

Systems of Linear Equations

Once we start grouping multiple linear equations with the same variables, we get what is called a **system of linear equations**

This is an example of a system of linear equations with 3 equations and 3 variables in general form:

$$\begin{aligned} 2x - 4y + 5z - 5 &= 0 \\ -3x + 5y + 1z - 4 &= 0 \\ -1x - 1y + 4z - 6 &= 0 \end{aligned}$$

If you remember we can write this system in standard form like so:

$$\begin{aligned} 2x - 4y + 5z &= 5 \\ -3x + 5y + 1z &= 4 \\ -1x - 1y + 4z &= 6 \end{aligned}$$

It's important to remember that every x, y, and z in each equation refer to 3 unique numbers that are present in all those equations that will make the equations true.

Some examples of other linear systems of n equations and m variables E_{nm} in general form:

$$E_{22} = \begin{aligned} 4x + 3y + 5 &= 0 \\ 7x - 2y - 3 &= 0 \end{aligned} \qquad E_{32} = \begin{aligned} 4x + 5y - 3 &= 0 \\ 5x - 6y + 0 &= 0 \\ -2x - 4y - 6 &= 0 \end{aligned}$$

$$E_{56} = \begin{aligned} 1x - 8y + 7z - 2w + 3i + 4j + 2 &= 0 \\ 5x + 2y + 5z + 1w + 7i - 8j - 4 &= 0 \\ 3x - 5y + 2z + 4w + 3i + 1j + 5 &= 0 \\ 2x - 4y + 3z - 2w + 2i - 3j - 3 &= 0 \\ 6x + 1y + 8z - 9w + 2i - 4j + 7 &= 0 \end{aligned}$$

Try to think about what these represent geometrically!

So now that we've finished defining what a linear equation *is*, we can't really do much with them at the moment. Part of this conundrum happens because we aren't quite sure how to solve these systems, and what their solutions even represent.

Solving Systems of Linear Equations

Let's take the following system of equations:

$$\begin{aligned} x - y + 2z &= 11 \\ 2y - z &= 2 \\ 3x - 2y + 2z &= 13 \end{aligned}$$

Let's think about how we can solve this system, in other words what x, y and z satisfy this equation?

To aid us in this struggles I will be labeling each equation like so:

$$\begin{aligned} E_1(x,y,z) &= 1x - 1y + 2z = 11 \\ E_2(x,y,z) &= 0x + 2y - 1z = 2 \\ E_3(x,y,z) &= 3x - 2y + 2z = 13 \end{aligned}$$

Remember linearity, we can formulate this as a linear equation with $ax + b$, where b is just another equation.

$$E_1(x,y,z) = x + E_1(y,z) = 11$$

Notice that equation 3 also has this similar structure with a non-zero coefficient for x and addition of another linear equation in y and z.

$$E_3(x,y,z) = 3x + E_3(y,z) = 13$$

Now we're going to subtract one equation from the other. This will remove the x variables in our equation.

$$3x + E_3(y,z) - 3(x + E_1(y,z)) = 13 - 33$$

We can rewrite this in terms of our equations E_1 and E_3 as:

$$E_3(x,y,z) - 3E_1(x,y,z) = 13 - 33$$

The act of eliminating our equations like this is called **Gauss-Jordan Elimination**.

I'm going to remove the parameters to make things easier to read. It will also come in handy later to make sure we mark down the operations we are doing and their equations. So I'll put that off to the side. Now if you notice this creates a brand new set of equations, that are still just as true as the previous system.

$$\begin{aligned} x - y + 2z &= 11 \\ 2y - z &= 2 \qquad E_3 \to E_3 - 3E - 1 \\ y - 4z &= -20 \end{aligned}$$

Using what we've just used we can continue to adjust our system of equation writing down all the operations we do between those equations. We will continue this process of adding and subtracting equations until we have a single variable per line:

$$\begin{aligned} x - y + 2z &= 11 \\ 2y - z - (2y - 4z) &= 2 + 40 \qquad E_2 \to E_2 - 2E_3 \\ y - 4z &= -20 \end{aligned}$$

$$\begin{aligned} x - y + 2z &= 11 \\ 7z \div 7 &= 42 \div 7 \qquad E_2 \to E_2/7 \\ y - 4z &= -20 \end{aligned}$$

$$\begin{aligned} x - y + 2z &= 11 \\ y - 4z &= -20 \qquad E_2 \leftrightarrow E_3 \\ z &= 6 \end{aligned}$$

$$\begin{aligned} x - y + 2z &= 11 \\ y - 4z + 4z &= -20 + 24 \qquad E_2 \to E_2 + 4E_3 \\ z &= 6 \end{aligned}$$

$$\begin{aligned} x - y + 2z + y - 2z &= 11 + 4 - 12 \\ y &= 4 \qquad E_1 \to E_1 + E_2 - 2E_3 \\ z &= 6 \end{aligned}$$

At the end we've arrived at our solution:

$$\begin{aligned} x &= 3 \\ y &= 4 \\ z &= 6 \end{aligned}$$

And the operations we did to get there are:

$$\begin{aligned}
&E_3 \to E_3 - 3E_1, \\
&E_2 \to E_2 - 2E_3, \\
&E_2 \to E_2/7, \\
&E_2 \leftrightarrow E_3, \\
&E_3 \to E_3 + 4E_2 \\
&E_1 \to E_1 + E_3 - 2E_2
\end{aligned}$$

If we notice something, in our final system of linear equations, $x = 3$ was our first equation, $y = 4$ our second and $z = 6$ our third. By substituting our final versions of E_1, E_2, E_3 for x, y, z respectively we end up with the following system of equations were the variables are our original starting equations:

$$\begin{aligned}
&x = E_1 + E_3 - 3E_1 + \frac{4}{7}(E_2 - 2(E_3 - 3E_1)) - \frac{2}{7}(E_2 - 2(E_3 - 3E_1^3)) \\
&y = E_3 - 3E_1 + \frac{4}{7}(E_2 - 2(E_3 - 3E_1)) \\
&z = \frac{E_2 - 2(E_3 - 3E_1)}{7}
\end{aligned}$$

Let's organize this a little bit:

$$\begin{aligned}
x &= -\frac{2}{7}E_1 + \frac{2}{7}E_2 + \frac{3}{7}E_3 \\
y &= \frac{3}{7}E_1 + \frac{4}{7}E_2 - \frac{1}{7}E_3 \\
z &= \frac{6}{7}E_1 + \frac{1}{7}E_2 - \frac{2}{7}E_3
\end{aligned}$$

Isn't it a bit odd that this $\frac{1}{7}$ appears in every term? This isn't just a coincidence, there is a reason for this however we will discuss this much later.

One thing that becomes immediatly apparent however, is that we have been able to write x,y and z in terms of linear equations. Now keep in mind that if we look back to when we started, these linear equations actualy had values:

$$\begin{aligned}
E_1 &= 11 \\
E_2 &= 2 \\
E_3 &= 13
\end{aligned}$$

Let's try changing these values, how about $E_1 = E_2 = E_3 = 7$:

$$\begin{aligned}
x &= -\frac{2}{7}7 + \frac{2}{7}7 + \frac{3}{7}7 \\
y &= \frac{3}{7}7 + \frac{4}{7}7 - \frac{1}{7}7 \\
z &= \frac{6}{7}7 + \frac{1}{7}7 - \frac{2}{7}7
\end{aligned}
\quad \to \quad
\begin{aligned}
&x = -2 + 2 + 3x = 3 \\
&y = 3 + 4 - 1 \quad y = 6 \\
&z = 6 + 1 - 2 \quad z = 5
\end{aligned}$$

Let's plug this in to our original system just to see if this adds up:

$$\begin{aligned}
E_1 &\to & (3) - (6) + 2(5) &= 10 - 3 \\
E_2 &\to & 2(6) - (5) &= 12 - 5 E_1 = E_2 = E_3 = 7 \\
E_3 &\to & 3(3) - 2(6) + 2(5) &= 9 - 2
\end{aligned}$$

If our original system of equations E is a function that takes 3 numbers (x, y, z) to 3 new numbers (a, b, c), then this new system of equations F will take our numbers (a, b, c) and return the (x, y, z) that formed them.

$$E \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$
$$F \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

We might be tempted to consider F to be the inverse of E, E^{-1}, however we might want to wait a bit before making this assertion.

This system of equations is what we call **consistent**. That means that there is a **unique** solution for every unique x,y and

Infinite Solutions

Now one thing we had was a system of 3 equations with 3 variables. Let's spice things up and take the following system of equations:

$$\begin{aligned}
3x + 3y &= 18 \\
2x + y + z &= 9
\end{aligned}$$

I implore you to pause and solve this on your own using the methods we discussed previously on your own for practice, however I'm going to quickly walk through it.

$$\begin{aligned}
\frac{1}{3}(3x + 3y) &= 18/3 \\
2x + y + z &= 9
\end{aligned} \qquad E_1 \to E_1/3$$

$$\begin{aligned}
x + y &= 6 \\
2x + y + z - 2(x + y) &= 9 - 12
\end{aligned} \qquad E_2 \to E_2 - E_1$$

$$\begin{aligned}
x + y &= 6 \\
-(-y + z) &= -(-3)
\end{aligned} \qquad E_2 \to -E_2$$

$$\begin{aligned}
x + y - (y - z) &= 6 - 3 \\
y - z &= 3
\end{aligned} \qquad E_1 \to E_1 - E_2$$

And we are left with this irreducible system of equations:

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

Now it might seem like we have done something wrong in the solving process, however I assure you that we have solved our sytem. The issue is that we were looking for a particular solution, however there is actually an *infinite* amount of solutions to this system. We can demonstrate this by setting z to be a parameter t. This will give us a solved system that varies based on t.

$$\begin{aligned}
x &= 3 - t \\
y &= 3 + t \qquad \forall t \in \mathbb{R} \\
z &= t
\end{aligned}$$

Now we're going to run through a quick check to see if this is actually true:

$$3(3-t)+3(3+t)=9+9+3t-3t=18$$
$$2(3-t)+(3+t)+t=6+3+2t-2t=9$$

So as we can see our t terms will always vanish so this will hold true for all t.

Now let's put these together. Remember that we're also going to have to include our z here.

$$x+z=-\frac{1}{3}E_1+E_2$$
$$y-z=\frac{2}{3}E_1-E_2$$

What happens if we plug in our original numbers 18 and 9?

$$x+z=-\frac{1}{3}18+9 \quad =3$$
$$y-z=\frac{2}{3}18-9 \quad =3$$

As expected.

Let's try two new numbers 27 and 12

$$x+z=-\frac{1}{3}27+12 \quad =3$$
$$y-z=\frac{2}{3}27-12 \quad =6$$

Giving this parameterized system:

$$\begin{aligned} x&=3-t \\ y&=6+t \qquad \forall t\in\mathbb{R} \\ z&=t \end{aligned}$$

So even though we've *kind of* found some inverse equations, there's still this weird parameterization that's sitting around. In this next section we will try to peer a little deeper in to that.

let's set our equations to be the same value of $E_1=E_2=0$

$$x+z=-\frac{1}{3}0+0 \quad =0$$
$$y-z=\frac{2}{3}0-0 \quad =0$$

This leaves us with this parameterized system:

$$\begin{aligned} x&=-t \\ y&=t \\ z&=t \end{aligned}$$

so anytime we have this case we're $z=y=-x$ our linear equations **all** collapse to the same value. In essence disappearing. We have an entire subset of combinations of x, y, and z that simply go to nothing. This is called the **null set** for this system. However we will come back to explore this topic in greater detail at a later time.

So if we have a unique x_1 and y_1 and add a y_2, x_2, and include z such that $z=y_2=-x_2$ then we could almost imagine our system of linear equations as a linear function:

$$E(x_1+x_2,y_1+y_2,z)=E(x_1,y_1,0)+E(x_2,y_2,z)=E(x_1,y_1,0)+0$$

When we think of linear equations like this, these random parameters in our solution seem to make a little more sense now. The parameters tell us the specific relationship that a particular x_2, y_2 and z need to have to dissapear. They also tell us how we could construct any number that will always satisfy the equation we're looking for.

This type of system of equations is called **inconsistent**. Because there are multiple x,y,z that lead to the same solution.

No Solutions

Let's try one more example:

$$\begin{aligned} 3x+2y+1z&=13 \\ x+2y+3z&=10 \\ 4x+4y+4z&=15 \end{aligned}$$

Go ahead and try doing solving this, however you should end up with this relatively quickly.

$$\begin{aligned} 3z+2y+1x&=13 \\ x+2y+3z&=10 \\ 0x+0y+0z&=-1 \end{aligned}$$

But that's impossible, there seems to be a contradiction here. Because of this contradiction, we can only conclude that this system has **no solutions**. This system is also called **inconsistent** then as it is impossible to determine which x,y,z led to a given solution.

Inverse

From what we've just discovered, only **consistent** systems of linear equations can have this property were we can properly go backwards. So essentially, only consistent systems have inverse systems.

Consistency of Linear Systems

As we have seen, only consistent systems have unique solutions. Whereas inconsistent do not. So this begs the question, is there a way to know when a system of linear equations is consistent without having to solve it?

Determining Consistency

One thing we can notice immediately is that if there is a difference between the number of variables and the number of equations, then the system is already inconsistent. This makes sense because if we have, say, 3 variables, we will need 3 equations to represent each of their values. However if we only have 2 equations then we will only be able to properly elimante for 2 of the variables. The final one will become our parameter. The same goes for systems where there might be more equations than variables.

Let's think about a simple system of 2 equations in 2 variables:

$$\begin{aligned} ax+by&=e \\ cx+dy&=f \end{aligned}$$

One thing that becomes apparent is that if all our variables are zero then the system is inconsistent. However another observation along those lines shows that if $c = ka$ and $d = kb$ for some constant k, then our system is inconsistent, since the second equation is just the first one scaled by some factor. Once we begin our elimination we will be left with one equation and two variables. Addtionally our system will have no solutions if $f - ke \neq 0$

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned} \rightarrow \begin{aligned} ax + by &= e \\ kax + kby - k(ax + by) &= f - ke \end{aligned} \rightarrow \begin{aligned} ax + by &= e \\ 0x + 0y &= f - ke \end{aligned}$$

There is nothing special about the ordering here, we could have also said $a = kc$ and $b = kd$.

Now let's think about what we would want to do if we were solving this system.

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned} \rightarrow \begin{aligned} ax + by &= e \\ (ac - ca)x + (ad - bc)y &= f - ce \end{aligned}$$

Now if you notice, we do not want our system to look like:

$$\begin{aligned} ax + by &= e \\ 0x + 0y &= f - ce \end{aligned}$$

Since $ca - ac = 0$, we need to make sure that $ad - bc$ does not equal zero either.

So in order to find out if our 2 variable 2 equation system is consistent, we only really need to check if $ad - bc \neq 0$. We don't need to check if $f - ce \neq 0$ since whatever that value is shouldn't matter so long as the y variable has a non-zero coefficient. One thing you might notice, is that for a 2x2 system, we subtract the product of each corner. top left times bottom right minus top left times bottom right.

Let's try this again for a system of 3 equations and 3 variables:

$$\begin{aligned} ax + by + cz &= j \\ dx + ey + fz &= k \\ gx + hy + iz &= l \end{aligned}$$

$$\begin{aligned} ax + by + cz &= j \\ 0x + (ae - db)y + (af - dc)z &= k \\ 0x + (ah - gb)y + (ai - gc)z &= l \end{aligned}$$

So we need to make sure that these equations are consistent, notice that the bottom two lines are a system of linear 2 equations and 2 variables, which we already know how to check if it's consistent.

$$(ae - db)(ai - gc) - (af - dc)(ah - gb) \neq 0$$

Substituting our terms we get: $$

$$ So our full equation that we need to check is:

Determinants

This is what is called the determinant of a matrix (for those of you already familiar with matrices). It is a way in which we can determine whether or not a system is consistent.

The determinant will also come in handy later when we start talking about matrices.

Matrix Notation

So those of you who have already started a first course in linear algebra, or have been exposed to a bit of matrix notation already are probably wondering where they are.

I've decided to hold on to using matrices until I've already shown the process of solving linear systems. Since that is all that a matrix really is, it is just a way to represent a system of linear equations.

For example:

$$\begin{aligned} ax + by + cz \\ dx + ey + fz \\ gx + hy + iz \end{aligned}$$

as the matrix:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Essentially what I hope to achieve is show how linear algebra works on linear equations, and then move in to showing how talking about matrices helps us have a concrete representation of these systems and how we can manipulate them. As well as how those systems can act on arbitrary inputs, which will be our vectors. Or how vector spaces can represent all the combinations of inputs and outputs that are allowed.

So the next part of this series will be on Vector Spaces and how linear equations are actually representing operations over a vector space.

After that we will finally move on to more Matrix Theory and how to manipulate matrices, and what each manipulation means in the context of linear equations.

Complex Numbers

Motivations

$$i^2 = -1$$

Complex Number Properties

Complex numbers have a variety of additional special properties that real number do not.

Complex Conjugate

The complex conjugate of a complex number denoted $\overline{z}$ or z^* simply reverses the sign on the imaginary part of the number.

$$\overline{a+bi} \to a - bi$$

Modulus

If you look at a com

This length is called it's modulus, and it's given by the pythagorean theorem:

$$|z| = \sqrt{a^2 + b^2}$$

However if we inspect the radican we find that we can describe this in terms of multiplication with z's conjugate:

$$a^2 + b^2 = (a+bi)(a-bi) = zz^*$$

so in fact we can write the modulus in terms of z and it's conjugate:

$$|z| = \sqrt{zz^*}$$

Argument

The argument of a complex number is simply the angle it forms with the real number line in the complex plane.

We can rewrite a complex number in terms of it's modulus and argument by considering the associated x and y for z as $|z|\cos\phi$ and $|z|\sin\phi$ respectively:

$$z = |z|\big(\cos\phi + i\sin\phi\big)$$

Proving that $\mathbb{C}$ forms a field

Addition

Adding two complex numbers happens component wise like this:

$$(a+ib) + (c+id) = (a+b) + (d+b)i$$

Since this addition is component wise we know that addition is commutative since it's analogous to addition over the real numbers.

Since the result is also complex number we know this is closed under addition.

Additive Identity

Since addition is component wise, and those components are real numbers, we can basically just set our additive identity as having components made up of the additive identity in the real numbers, or 0:

$$0 + 0i + a + bi = (a+0) + (b+0)i = a + bi$$

Subtraction and Negation

Just

$$(a+ib) - (c+id) = (a-b) + (d-b)i$$

In fact this could also be described as adding the negative of $c + di$. We can also write addition and subtraction together like:

$$(a+bi) \pm (c+di) = (a \pm b) + (d \pm b)i$$

Multiplication

Since complex numbers are composed of real numbers, we know that multiplication with a real number will be distributive like so:

$$3(a+bi) = (3a) + (3b)i$$

likewise if we attempt to multiply two complex numbers by distributing a and bi we notice we also get a complex number

$$(a+bi)(c+di) = (ac-bd) + (ad+cd)i$$

Thus $\mathbb{C}$ is closed under multiplication.

It's also commutative:

$$(c+di)(a+bi) = (ca-db) + (ad+cd)i$$

We know this commutes because the multiplication on the real numbers commute

Multiplicative Identity

This one is pretty easy, we know from the real numbers that one is the multiplicative identity. Since complex numbers are composed of real numbers any multiplicative identity on $\mathbb{C}$ would have to also work on $\mathbb{R}$ so one will work here as well.

$$1(a+bi) = (1)a + (1)bi$$

Division

Right away we can define complex division of z with itself as:

$$\frac{a+bi}{a+bi} = 1$$

From this we write a general divisor as:

$$\frac{1}{c+di} = \frac{1}{c+di}\frac{c-di}{c-di} = \frac{c-di}{c^2+d^2}$$

or in a more condensed form:

$$\frac{1}{z} = \frac{z^*}{|z|^2}$$

Division between two complex numbers z and w becomes:

$$\frac{z}{w} = \frac{zw^*}{|w|^2}$$

Which will give a complex number as $\mathbb{C}$ is closed under complex multiplication.

The Meaning of Complex Multiplication

2D rotation

Cartesian and Polar Form

$$z = |z|(\cos\phi + i\sin\phi)$$

De Moivre's Formula

$$(\cos x + i\sin x)^n = \cos nx + i\sin nx$$

Euler's Formula

$$e^{i\phi} = \cos\phi + i\sin\phi$$

This means we can go back and represent our polar form as:

$$z = |z|e^{i\phi}$$

In fact it can be useful to remember that for arbitrary angles we always have multiplies of 2π that we can add or subtract that will give us the same angle.

$$z = |z|e^{i(\phi\pm 2k\pi)}, \quad k \in \mathbb{Z}$$

This is related to De Moivre's Formula like so:

$$(\cos\phi + i\sin\phi)^n = \left(e^{i\phi}\right)^n = e^{in\phi} = \cos n\phi + i\sin n\phi$$

$$e^{i\pi} + 1 = 0$$

Complex Exponentiation

Thanks to this formula, complex exponentiation *and even multiplication* can be described in a relatively simple way:

$$z^n = |z|^n e^{in\phi}$$

$$zw = |z||w|e^{i(\phi+\varphi)}$$

Euler's Formula acts as an isomorphism from $(\mathbb{C}, +)$ to $(\mathbb{C}^*, \cdot)$.

Complex Roots of Unity

Let's attempt to find the squre root of an arbitrary complex number z. This becomes simple when we represent it in polar form:

$$\begin{aligned}\sqrt{z} &= \sqrt{|z|e^{i\phi\pm 2k\pi}} \\ &= \sqrt{|z|}e^{i\frac{\phi}{2}\pm k\pi}\end{aligned}$$

However now we need to adjust k so it belongs to $\{0, 1\}$, which is $\mathbb{Z}_2$. This makes sense due to the periodic nature of 2π in angles, since adding even and odd values of k is essentially the same as adding either 0π or π respectively.

Since the values of k tell us we are adding π to our angle, it represent us flipping our angle by $180°$. When z is a real number, this basically corresponds to taking it's negative! So this confirms that there are two values for square roots of real numbers $\pm\sqrt{x}$.

Now if we try the n-th root we'll, end up with something similar:

$$\sqrt[n]{z} = \sqrt[n]{z}e^{i\frac{\phi}{n}\pm 2\frac{k}{n}\pi}$$

where $k \in \mathbb{Z}_n$.

Now we have a total of n distinct values for $\sqrt[n]{z}$, corresponding to all the possible ks. These are called the **roots of unity**.

Relation to Linear Equations

Complex Numbers can be described also in the realm of linear equations, if we treat 1 and i like variables:

$$a(1) + bi \;\rightarrow\; ax + by$$

In fact 1 and i form a basis for $\mathbb{C}$ over $\mathbb{R}$, but in order for that to make sense we need to look at how we can abstract the notion of a linear equation. Enter **Vector Spaces**.

Vector Spaces

What is a Vector?

We shall begin by asking what a vector is.

For those who've got a bit of physics you might recognize a vector as a point in space, with a magnitude and momentum. for those who've done a bit of programming and linear algebra already you know that we can also represent these as row and column matrices.

$$\begin{bmatrix} a & b & c \end{bmatrix} \qquad \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

I'm going to change the definition of what a vector is by .

A vector is an element of a Vector Space that can be added

Vector Properties

Magnitude (Norm)

The magnitude of a vector is it's length

Angle

Unit Vectors

Unit vectors are vectors with a norm of 1.

We denote this with a little hat or circumflex, $\hat{v}$, over the vector to make it stand out.

We can create a unit vector from a vector by dividing the vector by it's norm:

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

This allows us to redefine a vector in terms of it's magnitude and direction:

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

What is a Vector Space?

Axioms

Examples

This ultimately makes us asks, can we form **linear combinations** and is there a zero vector included.

Additional Vector Operations

Dot Product

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

Cross Product

$$\vec{v} \times \vec{u} = \hat{n}$$

$$ad - cb$$

Distance between two vectors

$$d(\vec{v}, \vec{w}) = ||\vec{v} - \vec{w}||$$

Vector Projection

$$\mathrm{proj}_{\vec{v}}(\vec{w}) = \frac{\vec{v} \cdot \vec{w}}{||\vec{v}||^2}\vec{v}$$

Orthogonality

$$\vec{v} \cdot \vec{u} = 0$$

If the vectors also happen to be unit vectors, then we call this set *orthonormal*.

Fourier Expansion

$$\mathrm{fourier}(\vec{x}) = \sum_{i}^{n} \mathrm{proj}_{\vec{f_i}}(\vec{x})$$

Gram-Smidt Process

$$\begin{aligned}
f_1 &= v_1 \\
f_2 &= v_2 - \mathrm{proj}_{f_1}(v_2) \\
f_3 &= v_3 - \big(\mathrm{proj}_{f_2}(v_3) + \mathrm{proj}_{f_1}(v_3)\big) \\
f_n &= v_n - \sum_{i}^{n-1} \mathrm{proj}_{f_i}(v_n)
\end{aligned}$$

Homomorphisms

A homomorphism is a structure preserving map

$$\begin{aligned} f : (G, +) &\to (H, \diamond) \\ a &\mapsto f(a) \end{aligned}$$

where the operations are preserved

$$f(a + b) = f(a) \diamond f(b)$$

In the context of a vector space that operation is vector addition.

It is called an

Vectors as Linear Equations

Spanning Sets

$$av + bw = \mathrm{span}\{v, w\}$$

Linear Independence

Basis

A basis is a linearly independent set of vectors whose span forms a subspace.

Subspaces

The dimension of a subspace is the number of linearly independent vectors in that subspace, or essentially the amount of basis vectors in U.

$$\dim(U)$$

Subspace Projection

$U = \{\vec{u}_1, \vec{u}_2, .., \vec{u}_n\}$

$$\text{proj}_U(\vec{x}) = \sum_{i=1}^{n} \text{proj}_{\vec{u}_i}(\vec{x}), \quad \forall \vec{u}_i \in U$$

Examples of Vector Spaces and Subspaces

1. $\mathbb{R}^n$
   - The standard vector space we are used to working in
2. $M_{nn}(\mathbb{F})$
   - Vector Space of $n \times n$ matrices over some field
3. $P_n(\mathbb{F})$
   - The vector space of polynomials over some field

We can also describe the complex numbers as vectors whose associated field is the real numbers:

$$\mathbb{C} = \text{span}_{\mathbb{R}}\{1, i\}$$

This extends to quaternions as well except we have two different formulations, one with complex and one with the reals:

$$\mathbb{H} = \text{span}_{\mathbb{C}}\{1, j\} = \text{span}_{\mathbb{R}}\{1, i, j, k\}$$

Linear Transformations Between Vector Spaces

Tying into homomorphisms from Abstract Algebra, linear transformation (often called linear maps), are simply homomorphisms from one vector space to another. A linear transformation is defined:

$$\begin{aligned} T : V &\to W \\ v &\mapsto T(v) \end{aligned}$$

and is called a **linear** simply because it satisfies the property of linearity:

$$T(av + bw) = aT(v) + bT(w)$$

Ther are other transformations we can define between vector spaces, however in the context of linear algebra linear transformations are fundamental.

A linear transformation is called a linear operator if it is also an endomorphism $T : V \to V$

Linear Equations and Linear Transformations

A linear transformation on V can be defined as a system of linear equations taking elements in V that retrun an element in W.

For example, take the linear $T : \mathbb{R}^3 \to \mathbb{R}^4$ that is defined by these equations:

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{aligned} 3x + 4y - 2z &= 5 \\ -x - 2y + 3z &= 4 \\ -2x + 4y - 3z &= 3 \\ x - y - z &= 5 \end{aligned}$$

blah blah blah

This means we can define the action of T on V based on what it does to the basis of V.

In order to describe linear transformations it's important to recognize and define certain structures and patterns within them. The following is a look into the properties of linear transformations that are useful.

Kernel and Image

Do you remember in the part about systems of equations where I mentioned solving a system of equations only to find that there were infinite solutions? If we recall our example, we ended up finding a vector where any linear combination of that vector would map to zero.

This happens to become an important part of that particular transformation, these vectors have the relationship $T(\vec{v}) = \vec{0}$. There may be multiple vectors, and it could be useful to understand what those vectors are.

This set of vectors is called the kernel of T, $\ker(T)$, notice that this forms a subspace of the domain V since those are the vectors the transformation is being applied to:

$$\ker(T) = \{\vec{v} \in V : T(\vec{v}) = \vec{0} \in W\}$$

This becomes useful in allowing us to check injectivity, which we'll talk about soon.

Now what about the rest of the values in our domain? What about the codomain itself. Well if we look at the vectors that are completely mapped into our codomain. We could form a set of these vectors:

$$\{T(v_1), T(v_2), .., T(v_n)\}$$

This set is called the image of T, $\text{Im}(T)$, notice that this set is a subspace of the codomain W.

$$\text{Im}(T) = \{\vec{w} \in W : T(\vec{v}) = \vec{w}, \quad \vec{v} \in V\}$$

This is useful for finding surjectivity, which we'll talk about in the next chapter.

Injectivity

A linear transformation is said to be injective if it maps every value in the domain to a uniqe value in the codomain. This is often mention as T being *one-to-one*:

$$\text{injective}(T) \iff T(\vec{v}_i) = T(\vec{v}_j) \to \vec{v}_i = \vec{v}_j, \quad \forall \vec{v}_i, \vec{v}_j \in V$$

or

$$\text{injective}(T) \iff T(\vec{v}) \neq T(\vec{u}), \quad \forall (\vec{v}, \vec{u}) \in V \quad \vec{v} \neq \vec{u}$$

and the negation is:

$$\neg(\text{injective}(T)) \iff \exists (\vec{v}, \vec{u})\ \vec{v} \neq \vec{u} : T(\vec{v}) = T(\vec{u})$$

Since the $\ker(T)$ tells us the set of vectors that map to zero, then if the kernel has values other than the standard zero vector, then we can find at least two vectors that map to the same value. Making our transformation non-injective.

This becomes a good way to look at vectors that map to the same value that isn't zero as well. Suppose there were two vectors that mapped to the same value in the codomain:

$$T(v_1) = T(v_2) = w$$

Well if we were to write v_1 and v_2 as sums of one vector u with some arbitrary vectors $x_1, x_2 \in \ker(T)$ then:

$$\begin{aligned} T(v_1) &= T(u + x_1) = T(u) + \vec{0} \\ T(v_1) &= T(u + x_2) = T(u) + \vec{0} \end{aligned}$$

So the kernel is enough to tell us whether or not T is injective.

Surjectivity

A linear transformation is **surjective** if it maps every vector in the domain to the entirety of the codomain. it is often said that T is onto.

$$\text{surjective}(T) \iff W = \{T(\vec{v}) : \vec{v} \in V\}$$

or

$$\text{surjective}(T) \iff W = \text{Im}(T)$$

and it's negation would be:

$$\neg(\text{surjective}(T)) \iff \exists \vec{w} \in W : \text{Im}(T) \not\ni \vec{w}$$

So if we can find a vector in our codomain that is not an element of T's image, we have shown that T is not surjective.

Isomorphic Maps and Inverse Transformations

A linear transformation is an **isomorphism** if it is **bijective** (injective & surjective). Since an isomorphic map maps all values in our domain to a unique value in our codomain and no two values in the domain map to each other, then we could almost imagine a reverse mapping that maps every elment in our codomain to the original element in the domain.

This would be the **inverse** of our transformation and all isomorphic maps are invertible, that is:

$$\exists (T^{-1} : W \to V) : T^{-1} \circ T(x) = x, \forall x \in V \quad \forall T(x) \in W$$

We can find the inverse the same way we found the inverse in the linear system of equations section. However we will soon find a nicer way to represent linear transformations.

Time to move on to **MATRICES**!!

Matrices

Matrices as Systems of Linear Equations

Notice how when we were solving these linear equations we didn't really need the x, y, and z variables. We only were working with their coefficients and so long as we sort of centered them properly then solving them because a lot easier.

For example take the following system of 3 equations in 3 variables:

$$\begin{aligned} x + 3y - 2z \\ y - z \\ 3x + y \end{aligned}$$

For simplicity we can write the coefficients in a nice form like this with the columns representing the positions of our x, y, and z variables:

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

This structure is called a **Matrix**. The *matrix* (meaning womb in latin) holds all of our coefficients and makes it easy for us to work through solving systems of linear equations.

Matrices can also be represented with the following index notation:

$$A = a_{ij}$$

where i is each row and j is each column.

Matrix Size

Matrices come in all sorts of shapes and sizes, however they all have a nimber of rows and a number of columns. We represent this as the matrix belonging to the set M_{mn}, where m is the number of rows and n is the number of columns.

For example: here are those linear equations from the first section again as matrices:

$$M_{22} \ni \begin{bmatrix} 4 & 3 \\ 7 & -2 \end{bmatrix} \qquad M_{32} \ni \begin{bmatrix} 4 & 5 \\ 5 & -6 \\ -2 & -4 \end{bmatrix}$$

$$M_{56} \ni \begin{bmatrix} 1 & -8 & 7 & -2 & 3 & 4 \\ 5 & 2 & 5 & 1 & 7 & -8 \\ 3 & -5 & 2 & 4 & 3 & 1 \\ 2 & -4 & 3 & -2 & 2 & -3 \\ 6 & 1 & 8 & -9 & 2 & -4 \end{bmatrix}$$

Matrix Addition and Subtraction

Now it's time to think about what it means to perform our basic arithmetic operations on matrices, starting with subtraction.

Matrix addition happens element wise:

$$A + B = (a_{ij} + b_{ij})_{ij} = c_{ij} = C$$

The best way to understand this is to simply ask, what does it mean to add two linear equations?

$$ax + bx = (a + b)x$$

For example:

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} + \begin{bmatrix}5 & 6\\7 & 8\end{bmatrix} = \begin{bmatrix}6 & 8\\10 & 12\end{bmatrix}$$

Addition is essentially done over the elements of our field, which means that we also have a matrix full of the additive identity element, which would interact with A as follows:

$$\mathbf{0}_{mn} + A = A + \mathbf{0}_{mn} = A, \quad \forall A \in M_{mn}$$

where $\mathbf{0}_{mn}$ is the $m \times n$ matrix of all zeros:

$$\mathbf{0}_{mn} = \begin{bmatrix}0 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 0\end{bmatrix}$$

Matrix Scaling

Matrix scaling happens element wise as well:

$$kA = (ka_{ij})_{ij}$$

For example:

$$5\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} = \begin{bmatrix}5 & 10\\15 & 20\end{bmatrix}$$

Matrix scaling is distributive over matrix addition for the same reasons:

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

This is because any multiplication operations on sums of linear systems of equations are linear:

$$4\begin{pmatrix}2x+3y\\-x+y\end{pmatrix} + 5\begin{pmatrix}-2x-4y\\x-2y\end{pmatrix} = \begin{matrix}(4\cdot 2 - 5\cdot 2)x + (4\cdot 3 - 5\cdot 4)y\\ (-4+5)x + (4-5\cdot 2)y\end{matrix}$$

We would also have scaling by the multiplicative identity in our field:

$$1A = A$$

As this value would leave the elements in the matrix unchanged.

Matrices and Vector Spaces

If you noticed, as I've gone through how we can go about adding matrices, these properties all satisfy the axioms of a vector space.

This means we can express linear combinations of matrices as belonging to a span of a subspace of matrices:

$$aA + bB = \operatorname{span}_{\mathbb{F}}\{A, B\}$$

We could define a basis to center these operations around.

For example, all 2×2 real matrices can be expressed as belonging to the vector space $M_{22}(\mathbb{R})$ with a basis:

$$\begin{bmatrix}a & b\\c & d\end{bmatrix} \in \operatorname{span}_{\mathbb{R}}\left\{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\}$$

Matrix Products

We're going to discuss Matrix Multiplication really briefly now.

$$AB = \sum_{i=0}^{n} A_i B_i$$

Multiplying a matrix by another one simply applies the left matrix to each column in the right matrix:

$$A\begin{bmatrix}| & | & & |\\ b_1 & b_2 & \cdots & b_n\\ | & | & & |\end{bmatrix} = \begin{bmatrix}| & | & & |\\ Ab_1 & Ab_2 & \cdots & Ab_n\\ | & | & & |\end{bmatrix}$$

We could also consider an inverse matrix B such that:

$$BA = I$$

We would usually denote this matrix as A^{-1}:

Since matrix multiplication is not commutative, we need to build the idea of a right inverse AA^{-1} and a left inverse $A^{-1}A$:

This all ties back in to linear equations in a striking way. We can represent a system of linear equations as a **matrix multiplication**. For example take the system:

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1\\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2\\ \vdots \qquad\qquad &\quad \vdots\\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_m\end{aligned}$$

Now write it as the matrix product:

$$\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\ \vdots\\ x_n\end{bmatrix} = \begin{bmatrix}y_1\\y_2\\ \vdots\\ y_m\end{bmatrix}$$

Or more compactly as:

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

This would mean that our system of equations has a unique solution **if and only if** A is invertible:

$$\begin{aligned}A\vec{x} &= \vec{y}\\ A^{-1}A\vec{x} &= A^{-1}\vec{y}\\ I\vec{x} &= A^{-1}\vec{y}\\ \vec{x} &= A^{-1}\vec{y}\end{aligned}$$

But how can we determine if a matrix is invertible?? One of the big keys is something we've already discussed in the linear equations section, which is **determinants**.

Determinants

$\det A$ or $|A|$.

$$\begin{vmatrix}a & b\\c & d\end{vmatrix} = ad - bc$$

Leibniz Formula

Laplace's Formula

This is

Minors

Adjoint Matrix

An adjoint matrix is a matrix who

Similar Matrices

Similar Matrices are matrices

The associated *space* of solutions that each matrix represents is the same, since the system of equations of both matrices represent the same relationship.

Symbolically this relationship looks like:

$$A \sim D \iff \exists P : PA = D \wedge A = P^{-1}D$$

Matrix Transpose

$(A_{ij})^T = A_{ji}$

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

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

Complex matrices come with an additional operation, the **conjugate transpose** $\left(A^T\right)^* = A^\dagger$:

$$\begin{bmatrix} 1+i & -i \\ 3+4i & 2-3i \end{bmatrix}^\dagger = \begin{bmatrix} 1-i & 3-4i \\ i & 2+3i \end{bmatrix}$$

Useful Matrix Forms

Row Echelon Form

$$\begin{bmatrix} a & * & \cdots & * \\ 0 & b & * & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & z \end{bmatrix}$$

Echelon is a french term for

Reduced Row Echelon Form

The Reduced Row Echelon Form is exactly what it sounds like, the matrix is reduced so that the first non-zero entry to every row is 1. These are called our leading ones and every entry above or below a leading one is zero

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & * \\ 0 & 1 & 0 & \vdots & \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & 1 & * \end{bmatrix}$$

Remember how this applies to our solutions for linear equations. Essentially the system of equations given by the matrix has a solution if and only if that matrix is similar to the identity matrix.

$$Ax = y$$

Can only be solved for a unique x if there exists a A^{-1} so that:

$$x = A^{-1}y$$

So what the RREF tells us, is not only whether or not a system of linear equations has a solution, but also the number of linearly independent variables in our inputs. Recall the portion about infinite solutions. When we solved that system and found a parameterized form for our variables, we essentially put our matrix into RREF only to find that it's RREF was no identical to the identity matrix. If you recall that system was inconsistent and had no general inverse system, just like the matrix representing that system would not have an associated inverse.

Putting a matrix into RREF is pretty simple, you essentially just want to put the matrix into REF with every leading term being one, then you simply subtract from the lowest leading one until all entries that are above leading ones are zero.

Column Echelon Form

There is also a column echelon form and reduced column echelon form.

To put a matrix into CEF we simply perform row operations over the *columns* of a matrix rather than it's rows. This however is equivalent to performin row operations over the matrix's transpose.

For all invertible matrices, their RREF and RCEF should be the same.

Matrix Properties

Dimension

Row and Column Space

If we go back to the relationship between matrices and systems of equations, we can remember that we can represent a linear system of equations as belonging to a subspace:

$$\begin{matrix} 2x+y+3z \\ 3x-2y-z = \\ -x+2z \end{matrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix} \in \operatorname{span}\left\{ \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} \right\}$$

But remember that the coefficients of our linear system of equations can also be represented as matrix A:

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

This means that the matrix A has the vectors in the spanning set as it's columns.

This is called the **column space** of A, denoted $\operatorname{col}(A)$.

Likewise the column space of it's transpose is the **row space** of A, denoted $\text{row}(A)$.

For example, take the following matrix

$$\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \end{bmatrix}$$

The column space is simply the set spanned by the columns of our matrix as vectors:

$$\text{col}(A) = \text{span}\left\{ \begin{bmatrix} a \\ e \\ t \end{bmatrix}, \begin{bmatrix} b \\ f \\ j \end{bmatrix}, \begin{bmatrix} c \\ h \\ k \end{bmatrix}, \begin{bmatrix} d \\ h \\ l \end{bmatrix} \right\}$$

The row space is the set of vectors spanned by the rows of our matrix

$$\text{row}(A) = \text{span}\left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \begin{bmatrix} e \\ f \\ g \\ h \end{bmatrix}, \begin{bmatrix} i \\ j \\ k \\ l \end{bmatrix} \right\} = \text{col}\left(A^T\right)$$

Rank and Linear Independence

So we've already shown how we can describe the columns of a matrix as a spanning set of vectors. Know we also know that not all spanning sets are linearly independent.

If we think about what it would mean to find out if our vectors were linearly independent we can imagine we start off with a set of linearly independent vectors β.

$$\beta = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & & | \end{bmatrix}$$

If a set of vectors is linearly dependent it would mean that at least *one* of these vectors is a linear combination of the others. If we think about this in terms of matrix multiplication, we are really just taking this linearly dependent set and multiplying it by the matrix

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & a \\ 0 & 1 & & 0 & b \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & 1 & z \end{bmatrix} \beta = \begin{bmatrix} | & & | & | & & | \\ b_1 & \cdots & b_n & ab_1 + & \cdots & +zb_n \\ | & & | & | & & | \end{bmatrix}$$

Where the final row is some linear combination of our linearly independent vectors.

So to determine the linear independence of our vectors we would want to be able to show that our matrix is equivalent to multiplying a linearly independent set by the identity.

This seem like a job for row operations, and if you recall, the RREF form of a matrix essentially tells us the number of linearly independent variables. So in this case all we actually have to do is put β into RREF where we're considering our variables as some basis.

Yet another reason why RREF is very important.

We call the number of linearly independent vectors the **rank** of the matrix. This is because that is the number of vectors that form the basis of our column space. The basis of our column space is also called the the image of A, $\text{Im}(A)$. And if you remember that the dimension of a subspace is the number of linearly independent vectors in that subspace; then finding the rank is essentially like asking for the dimension of the image:

$$\text{rank}(A) = \dim\big(\text{Im}(A)\big)$$

Kernel (Null Space) and Nullity

The nullspace of

$$A\vec{x} = \vec{0}$$

$$A\vec{x} = -A\vec{y}$$
$$A\vec{z} = 3A\vec{w}$$

$$A(\vec{x} + \vec{y}) = \vec{0}$$
$$A(\vec{z} - 3\vec{w}) = \vec{0}$$

So we can build

The dimension of the matrix kernel is called the nullity.

$$\text{nullity}(A) = \dim\left(\ker(A)\right)$$

You might want to think about how this relates to the kernel of a linear transformation as well.

Rank-Nullity Theorem

Let's think about how we can relate the image, kernel, rank and nullity together.

Let's imagine anaylizing the following set of vectors:

$$\text{span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -6 \\ 5 \\ -6 \\ 10 \end{bmatrix}, \begin{bmatrix} 4 \\ -5 \\ 3 \\ -8 \end{bmatrix}, \right\}$$

This represents the column space of the matrix:

$$\begin{bmatrix} 2 & 0 & -6 & 4 \\ 0 & 5 & 5 & -5 \\ 3 & 3 & -6 & 3 \\ -2 & 4 & 10 & -8 \end{bmatrix}$$

It's RREF looks like:

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

If we were to discuss this matrix in terms of it's column space, the RREF is essentially telling us that a span of 4 vectors has 2 linearly independent vectors, this means we can form two vectors that cancel out from the linearly dependent vectors just as we've shown when discussing how the kernel works. And 2 linearly independent vectors plus 2 linearly dependent vectors is 4 four total vectors, which is how many vectors are in our column space.

This is call the **rank-nullity theorem** and is states that the dimension of a matrix is the sum of it's rank and it's nullity:

$$\text{rank}(A) + \text{nullity}(A) = n$$

In essence all the rank-nullity theorem is stating is simply the number of linearly independent vectors plus the number of linearly dependent vectors gives us the total vector in our column space.

Types of Matrices

This next section will be discussing a variety of matrix types that are often useful. This is meant to be as comprehensive as possible so it can be used a resource for future use.

Augmented Matrices

An augmented matrix is formed when we concatenate two matrices and separate them with a line like so:

$$[A|B]$$

Any row operations we do on one side we simply apply to the other side as well.

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

$$\left[\begin{array}{ccc|ccc} 2 & 0 & 1 & 9 & 6 & 3 \\ 0 & -1 & 2 & 7 & -4 & 0 \\ -3 & -1 & 2 & 16 & 8 & 3 \end{array}\right]$$

So we started with some matrices $[A|B]$ and we ended up with matrices $[C|D]$. Since A was our identity matrix this means that our final row reductions is equivalent to multiplying CI. And since we did the exact same operations on the right side then that means that $D = CB$.

So when we write $[A|B]$ and we multiply A by some matrix P, we are also simultaniously multiplying B by that matrix as well.

To illustrate this, think of the augmented matrix representing the relation:

$$A = BC$$

and whatever we do to A we also do through P so we end up with:

$$PA = PBC$$

We can also say that our new augmented matrix represents the relation:

$$E = DC$$

This comes in handy if we want to find the inverse of a matrix, we simply set B to the identity matrix and row reduce A to the identity.

Take the matrix:

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

And now we're going to row reduce

$$\left[\begin{array}{ccc|ccc} 2 & 0 & -4 & 1 & 0 & 0 \\ 1 & 1 & 4 & 0 & 1 & 0 \\ 5 & 1 & -3 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{7}{2} & -2 & 2 \\ 0 & 1 & 0 & \frac{23}{2} & 7 & -6 \\ 0 & 0 & 1 & -2 & -1 & 1 \end{array}\right]$$

Let's investigate this a little, remember that we're representing the relation:

$$A = IC$$

in this case $C = A$, and when we apply our row reductions it is the equivalent to left multiplying by P:

$$PA = PIC$$

and since we've shown on the left hand side that $PA = I$ then $P = A^{-1}$ and the right hand represents the matrix $PI = A^{-1}$.

Say you want to solve the system:

$$Ax = y$$

We can write $[A|y]$ and if we row reduce A to the identity it is equivalent to applying the inverse to y to get:

$$x = A^{-1}y$$

Elementary Matrices and Similar Matrices

This gives rise to expressing row reductions as matrix multiplications.

1. Multiplying a row by another just means we want to form a linear combination of two rows:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & & & & & \\ \vdots & & \ddots & & & & \vdots \\ 0 & 0 & \cdots & 1 & a & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

2. Swapping two rows just means that we swap the ones associated with that row.

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & & & & & \\ \vdots & & \ddots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

3. Scaling a row just means we change the one to whatever coefficient we're scaling by:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & & & & & \\ \vdots & & \ddots & & & & \vdots \\ 0 & 0 & \cdots & a & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Two matrices A and B are said to be **similar** $A \cong B$ if there is a product of invertible elementary matrices such that:

$$\prod_i^k E_i A = B$$

Diagonal Matrices

A **Diagonal Matrix** is a matrix where all the entries are zero except for the main diagonal. These matrices are square matrices

$$\begin{bmatrix} * & 0 & & 0 \\ 0 & * & & \vdots \\ & & \ddots & \\ 0 & \cdots & & * \end{bmatrix}$$

One of the benefits to a diagonal matrix is that it's determinant is simply the product of the diagonal entries:

$$\det \begin{vmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{vmatrix} = 2 * 3 * 5 = 30$$

Triangular Matrices

A **lower** triangular matrix has zeros above the diagonal with entries below.

$$\begin{bmatrix} * & 0 & \cdots & 0 \\ * & * & & \vdots \\ & & \ddots & \\ * & \cdots & & * \end{bmatrix}$$

An **upper** triangular matrix is zero below the main diagonal with entreis above.

$$\begin{bmatrix} * & * & \cdots & * \\ 0 & * & & \vdots \\ & & \ddots & \\ 0 & \cdots & & * \end{bmatrix}$$

Like diagonal matrices, one of the benefits to a triangular matrix is that the determinant is the product of the diagonal entries:

$$\det \begin{vmatrix} 2 & 1 & 4 \\ 0 & 3 & -6 \\ 0 & 0 & 1 \end{vmatrix} = 2 * 3 * 1 = 6$$

Hessenberg Matrices

A **Hessenberg Matrix** looks like a triangular matrix but with less values.

An **upper** hessenberg matrix is zero except for the diagonal and a few entries in the subdiagonal.

$$\begin{bmatrix} * & * & \cdots & * \\ * & * & & \vdots \\ & & \ddots & \\ 0 & \cdots & & * \end{bmatrix}$$

A **lower** hessenberg matrix is zero except for the diagonal and a few entries in the superdiagonal.

$$\begin{bmatrix} * & * & \cdots & 0 \\ * & * & & \vdots \\ & & \ddots & \\ * & \cdots & & * \end{bmatrix}$$

Tridiagonal Matrices

A matrix that is both an upper and lower hessenberg matrix is called **tridiagonal**. A tridiagonal matrix looks like:

$$\begin{bmatrix} * & * & 0 & 0 & \cdots & 0 \\ * & * & * & 0 & & \vdots \\ 0 & * & * & * & & \vdots \\ 0 & 0 & * & * & & \vdots \\ \vdots & & & & \ddots & \\ 0 & \cdots & \cdots & \cdots & \cdots & * \end{bmatrix}$$

Sparse Matrix

A **Sparse Matrix** is a matrix where most of the elements are zero.

$$\begin{bmatrix} * & 0 & 0 & & \cdots & 0 \\ 0 & * & * & & & \\ * & 0 & \ddots & & & \vdots \\ \vdots & & & & 0 & \\ 0 & & \cdots & & * & 0 \end{bmatrix}$$

These matrices are great for storage in computers.

We won't see too many Hessenberg, Tridiagonal, or Sparse matrices right now as but they come in handy in the more advanced sections.

Symmetric Matrices

A **Symmetric Matrix** is an $n \times n$ square matrix A that is equal to it's own transpose $A = A^T$

For example, take the matrix:

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

Hermetian Matrix

A **Hermetian Matrix** is an $n \times n$ square matrix A with complex entries that is equal to it's own conjugate transpose $A = A^\dagger$, denoted $A = A^H$ sometimes as well.

For example, take the matrix:

$$\begin{bmatrix} 1 & -i & 3+2i \\ i & 2 & 5-6i \\ 3-2i & 5+6i & 3 \end{bmatrix}$$

Normal Matrices

Normal matrices are matrices that commute with their conjugate transpose:

$$AA^\dagger = A^\dagger A$$

Orthogonal Matrices

Orthogonal Matrices are $n \times n$ square matrices who's transpose is equal to their inverse.

$$A^T A = AA^T = I$$

Note, since the conjugate transpose of a real matrix is it's transpose, this means that real orthogonal matrices are also normal matrices.

The reason why this works is because orthogonal Matrices are made up of columns of orthogonormal vectors:

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

So multiplying by it's transpose creates a matrix of all the dot products:

$$A^T A = a_i^T a_j$$

Since the columns are orthonormal all dot products will be zero unless $a_i = a_j$ then the dot product equals itself, then it's one.

Unitary Matrices

A matrix who's conjugate transpose is equal to it's inverse is called **unitary**.

$$AA^\dagger = A^\dagger A = I$$

Since the conjugate transpose for real matrices is just the transpose then all orthogonal real matrices are unitary. Additionally, since unitary matrices commute with their conjugate transpose they are also normal.

Involutory Matrices

An involuntory matrix is a matrix that is also an involution (it's own inverse):

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

This means that even powers of A are the identity:

$$A^{2n} = I, \quad \forall n \in \mathbb{Z}$$

And odd powers are just A

For example, take the matrix:

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

if we multiply this matrix with itself we get the identity:

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

Householder Matrix

A **Householder Matrix** is an involutory matrix that is both unitary and hermitian.

$$\begin{aligned} A &= A^\dagger \\ AA^\dagger = A^\dagger A &= I \\ A^2 &= I \end{aligned}$$

These matrices arise from the **householder transform**, which you'll see much much much later when we cover numerical linear algebra and optimizations for solving matrix problems.

Matrices as Linear Transformations

Let $T : P_3 \to M_{22}$ act on a basis for P_3 as follows:

$$T\left(x^3\right) = \begin{bmatrix} 3 & 0 \\ 2 & 0 \end{bmatrix} \qquad T\left(x^2\right) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$T(x) = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix} \qquad T(1) = \begin{bmatrix} 1 & -3 \\ 0 & -2 \end{bmatrix}$$

Now let's say we want to know how this transformation act on any individual polynomial.

$$T\left(ax^3 + bx^2 + cx + d\right) = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$$

Well by the property of linearity:

$$aT\left(x^3\right) + bT\left(x^2\right) + cT(x) + dT(1) = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$$

Now by subbing in our matrices:

$$a\begin{bmatrix} 3 & 0 \\ 2 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix} + d\begin{bmatrix} 1 & -3 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 3a+d & b-3d \\ 2a+b & 3c-2d \end{bmatrix}$$
$$= \begin{bmatrix} p & q \\ r & s \end{bmatrix}$$

This forms a system of linear equations:

$$\begin{aligned} 3a - d &= p \\ b - 3d &= q \\ 2x + b &= r \\ 3c - 2d &= s \end{aligned}$$

And anytime we have a sytem of linear equations we have a matrix for the coefficients.

$$\begin{bmatrix} 3 & 0 & 0 & -1 \\ 0 & 1 & 0 & -3 \\ 2 & -1 & 0 & 0 \\ 0 & 0 & 3 & -2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix}$$

Notice that T can be represented as this matrix multiplication $Ax = y$ where $A \in M_{44}$ and $x, y \in \mathbb{R}^4$

As we have found, we can represent the linear transformation T as this 4×4 matrix A.

Relation to Span

Now if you notice, the polynomial $ax^3 + bx^2 + cx + d$ is actually $\text{span}\{x^3, x^2, x, 1\}$. So in essence what we've done is show the matrix A represents T's action on the span of our basis. And the new matrix $\begin{bmatrix} p & q \\ r & s \end{bmatrix}$ is some combination of $\left\{ \begin{bmatrix} 3 & 0 \\ 2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}, \begin{bmatrix} -3 & -1 \\ 0 & -2 \end{bmatrix} \right\}$.

Now what's interesting here is that the columns of our matrix A, look awfully familiar. If we take our matrices as vectors in $\mathbb{R}^4$ notice we get the basis:

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

Which is the $\text{col}(A)$.

Linear Transformations as Systems of Linear Equations

This has the added bonus of saying that linear transformations between vector spaces are actually a system of linear equations from **one basis** to **another**. With the unknowns or *inputs*, belonging to the basis of our *domain*, and the *outputs* belonging to the basis codomain.

Since matrices really represent the coefficients of a system of equations, then all of our linear transformations from on vector space to another have some matrix representation in terms of how they act on our initial basis.

Non-Standard Basis

Now this is all well and good, however we've forgotten to think about what would happen if we had a non-standard basis. Since we can think about our basis $\{x^3, x^2, x, 1\}$ as:

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

Let's say our starting basis was $\{x^3 + x^2, x^3 - x, x^3 + 1\}$. This can be represented as an endomorphism $T : P_3 \to P_3$ via the following matrix:

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

Notice this also acts like a transformation $F : \mathbb{R}^4 \to \mathbb{R}^3$.

Now let's say we have the following transformation $Q : P_3 \to M_{22}$

$$Q\left(x^3 + x^2\right) = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \quad Q\left(x^3 + x\right) = \begin{bmatrix} 0 & -3 \\ 0 & 3 \end{bmatrix}$$
$$Q\left(x^3 + 1\right) = \begin{bmatrix} 7 & 0 \\ 2 & -5 \end{bmatrix}$$

If look at how Q acts on our span:

$$Q\left(a\left(x^3 + x^2\right) + b\left(x^3 - x\right) + c\left(x^3 + 1\right)\right)$$

this can be represented as the following transformation:

$$\begin{bmatrix} 1 & 0 & 7 \\ 0 & -3 & 0 \\ 2 & 0 & 1 \\ 0 & 3 & -5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

and we alredy know that out vector $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ can be written as a transformation from our standard basis in P_3. $px^3 + qx^2 + rx + s$

$$\begin{bmatrix} 1 & 0 & 7 \\ 0 & -3 & 0 \\ 2 & 0 & 1 \\ 0 & 3 & -5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix}$$

Transformations Between Dimensions

Let's say there is a plane β in $\mathbb{R}^3$ spanned by two vectors:

$$\beta = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Now let's suppose there is a point on this plane:

$$\vec{x} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$$

What are the associated coordinates of $\vec{x}$ in our standard $\mathbb{R}^3$. Or that is, from the perspective of someone outside the plane β, how could you describe $\vec{x}$.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ -4 \end{bmatrix}$$

Where in this case the 2×2 identity matrix represents the basis of our plane β.

Represent β as the column space of a matrix B

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

So we simply multiply $B\vec{x}$:

Notice that as a transformation B is injective, but not surjective. Every vector on our plane maps into a vector in $\mathbb{R}^3$, but not all vectors in $\mathbb{R}^3$ are in β.

Now suppose we were given a vector $\vec{y}$ in $\mathbb{R}^3$ and we were told that it is on our plane β. How could we determine the coordinates of $\vec{y}$ in β???

Let's say $\vec{y}$ is :

$$\vec{y} = \begin{bmatrix} 10 \\ 2 \\ 8 \end{bmatrix}$$

Well we would hopefully want to find some matrix A that we could multiply to B that would return the 2×2 identity matrix:

$$AB = I_2$$

In this case the basis of the vector space we're *sitting* in would be expressed as the standard basis for something in $\mathbb{R}^n$. Again, the basis for β is **not** the *standard basis* $\{e_1, e_2\}$, but for the purposes of the vectors *living* in β it might as well be!

So that's why we want to find some matrix A that will change our perspective to be in β.

The best way to find A is to simply form an augmented matrix: $[B|I_3]$ and row reduce until we've formed I_3:

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

This gives us the rectangular matrix:

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

and in fact if we apply this to B we get the 2×2 identity:

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

WARNING!!! This is not the same as saying $A = B^{-1}$, B is **not** invertible. If B was invertible and it's inverse was A then B would commute with A: $AB = BA = I$. This is not true so we cannot call A B's inverse, nor call B invertible.

so applying A to $\vec{y}$ we get:

$$A\vec{y} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 2 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$$

Notice that A is also not injective, that means that not every vector in $\mathbb{R}^3$ will map perfectly on to the plane β, which corresponds with B not being surjective.

Let's say we have some random point $\vec{u}$ in $\mathbb{R}^3$ that we multiply to A:

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$$

Now let's find the $\ker(A)$, this is pretty easy, it's literally any vector with a z component but no x and y:

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

If you notice this vector is **orthogonal** to the vectors in β!!

So when we apply A to some arbitrary vector in $\mathbb{R}^3$ we actually **project** that vector on to β, or in other words $A\vec{x} = \text{proj}_\beta(\vec{x})$. And the $\ker(A)$ in this case means $\vec{y} \in \mathbb{R}^3$ where $A\vec{y} \notin \beta$. Which is β's **orthogonal complement** $\beta^\perp$.

This has a wider intuition that rectangular $m \times n$ matrices act as transformations *between* dimensions where if $m > n$ you are projecting from one hyperplane to a lower dimensional hyperplane, while if $m < n$ you are expressing elements of one hyperplane in higher dimensional terms.

A hyperplane is simply an abstraction of a the 2D plane for higher dimensions, a 3D hyperplane might look something like a cube, while a 2D is a regular plane, and a 1D hyperplane is simply a line. Usually the dimension of the hyperplane is one lower than the actual dimension of the geometry, so a line would be a 0-plane and a the 2D plane a 1-plane, etc.

Vector Dot Product as Transformations

Working off of our intuition from earlier, we notice that we already know of a way to express a projection from one vector to another. The **dot product**!

In fact we could sincerely write the dot product as a linear transformation:

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \vec{u}^T \vec{v}$$

In essence, the dot product is a transformation from a Vector Space to it's associate Field $T : V(\mathbb{K}) \to \mathbb{K}$. This is like we're moving down all of our dimensions to a single number line.

Let's take an example for polynomials:

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

This kind of makes sense when we think about $3x^2 - 5x + 4$ being a linear combination of our basis $\{x^2, x, 1\}$. The row matrix $[3 \quad -5 \quad 4]$ simply represents that linear combination as a transformation.

Determinants and Transformations

The determinant is the scaling coefficient of the volume between our vectors.

Change of Basis

This requires a very similar intuition to the section on rectangular transformations. Except instead of **linear transformations** between hyperplanes we are actually looking at linear **operators**.

Rather than having β be an embedded space in $\mathbb{R}^3$ we can think of β as a subspace that spans $\mathbb{R}^3$:

$$\beta = \text{span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\2\\1\end{bmatrix}\right\}$$

Now let's suppose we had another subspace that spanned $\mathbb{R}^3$ called α:

$$\alpha = \text{span}\left\{\begin{bmatrix}2\\0\\1\end{bmatrix}, \begin{bmatrix}0\\-1\\2\end{bmatrix}, \begin{bmatrix}2\\0\\3\end{bmatrix}\right\}$$

I will call the standard basis $\epsilon = \{e_1, e_2, e_3\}$, which is:

$$\epsilon = \text{span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

Notice that since α and β form a basis for $\mathbb{R}^3$, we have 3 linearly independent vectors, this means we can reduce matrix forms of β and α to the 3×3 identity. This has the intuition of taking α or β as some arbitrary *standard* basis that we can express all of $\mathbb{R}^3$ in. Since these can be reduced to the identity we know they must represent isomorphisms. So they can be **inverted**!!

But what does it even mean to invert a basis?

You see the answer is in perspective: notice that if we were some vector, living in β, we wouldn't really think about the actual vectors themselves in terms of the standard basis. To us the basis of β **is** the standard basis. I'm going to represent that as the relationship:

$$\beta^{-1} \circ \beta = \epsilon$$

where we can think of β^{-1} as sort of removing the vectors to make them appear *standard*.

If we had a vector x in our regular $\mathbb{R}^3$ and we wanted to know what someone in β would observe that vector as: well we just need to think about how someone in β would see that vector, and they would look at it like:

$$\beta^{-1}(x) = [x]_\beta$$

This translates the same with other basis. If we had a vector in α and wanted to see hows someone in β would describe that vector we can basically take:

$$\beta^{-1} \circ \alpha(x) = [x_\alpha]_\beta$$

So let's find these inverse basis.

$$\beta^{-1} = \text{span}\left\{\begin{bmatrix}-1\\2\\-1\end{bmatrix}, \begin{bmatrix}-1\\1\\0\end{bmatrix}, \begin{bmatrix}2\\-2\\1\end{bmatrix}\right\}$$

Notice this is really just inverting the matrix B where $\text{col}(B) = \beta$, and then taking $\text{col}\left(B^{-1}\right)$. We can do the same for α.

$$\alpha^{-1} = \text{span}\left\{\begin{bmatrix}\frac{3}{4}\\0\\-\frac{1}{4}\end{bmatrix}, \begin{bmatrix}0\\-1\\1\end{bmatrix}, \begin{bmatrix}-\frac{1}{2}\\0\\\frac{1}{2}\end{bmatrix}\right\}$$

Now if we think about what it means to represent a vector in β or α in terms of our standard basis ϵ, well we can simply remember that in this case ϵ is an identity operator and involutory.

In the language of matrix multiplication we're essentially looking at something like:

$$TAx = By$$

where A and B are matrices representing α and β, and T represents some general transformation. x and y represent linear combinations of α an β vectors. The term By and Ax simply refer to making a vector in β and α from the point y and x.

Notice how if we invert B we could write y as:

$$B^{-1}TAx = y$$

and now we've managed to describe a complete transformation on x. So if we had a vector $\vec{x}$ in our standard basis and wanted to see how someone in β would see that vector we can think of applying $\epsilon^{-1}\beta(\vec{x})$ which is nothing more than $\beta(\vec{x})$:

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

So the vector that represents $2e_1 + 3e_2 - e_3 \in \epsilon$ represents the vector $5b_1 + b_2 + b_3 \in \beta$.

Likewise to transform from α to ϵ we we can do the same thing.

Now what if we had a vector $\vec{x}_\beta$ that we wanted to represent in our standard basis? Well we can just take:

$$\beta^{-1} \circ \epsilon(\vec{x})$$

Again, think of the intuition of *reversing* or eliminating a particular perspective.

Now what if we wanted to represent a vector in β as a vector in α then the change of basis operator would be:

$$\alpha^{-1} \circ \beta(\vec{x})$$

which is represented through the matrix multiplication:

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

Which is the same operation as:

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

Again think of the relationship:

$$Ax = By$$

x and y are just different ways to interpret the same vector, going from one basis to another is simply multiplying by the inverse of the basis we want to go to!

Basic Spectral Theory

Eigenvectors and Eigenvalues

One thing that we might be curious about are vectors that appear unchanged under a linear transformation:

$$Ax = x$$

where $A \in M_{mn}$.

An example of a vector like this is the zero vector of the codomain of A, $\vec{0}_n$. However when we think about it, this transformation is essentially taking a vector and then scaling it by one, so a more general form of this relationship is a vector that, under a linear transformation, maps to itself times a scalar constant λ.

$$Ax = \lambda x$$

again $\vec{0}_n$ always satisfies this equation. Additionally, you may notice that if $x \in \ker(A)$ then:

$$Ax = 0x = \vec{0}_m$$

where $\vec{0}_m$ is the zero vector in the codomain of A.

The vector x that satisfies this equation is called an **eigenvector**, and the scalar λ is it's **eigenvalue**, *eigen* being the german word for characteristic. As we have already shown 0 and $\vec{0}_n$ are already trivial eigenvalues and eigenvectors of A. But how can we go about finding other eigenvectors and eigenvalues of A?

One way to think about this is by rearranging the equation:

$$Ax - \lambda x = \vec{0}_m$$

and using linearity of matrix multiplication and the fact that $\lambda x = \lambda I x$ then:

$$(A - \lambda I)x = \vec{0}_m$$

this means that $x \in \ker(A - \lambda I) \quad \forall \lambda$. So we've found a way to find x in terms of λ. However we still don't know how to find λ. Well since $\ker(A - \lambda I)$ has a vector other than $\vec{0}_n$ then that means it's non-injective, and as a result not an isomorphism. This mean that the determinant must be zero, so we can set the equation:

$$\det |A - \lambda I| = 0$$

Since the terms along the diagonal of our matrix will look like: $(a - \lambda)$ the determinant will give us a polynomial in terms λ called the characteristic polynomial, or eigenpolynomial. We can solve that polynomial to find our eigenvalues, then for each eigenvalue λ solve for $\ker(A - \lambda I)$ to get our eigenvectors.

Notice we could have done this the other way and done $\lambda I - A$, we would have found the same eigenvalues and eigenvectors.

Diagonal Matrices

$$\begin{vmatrix} a-\lambda & 0 & 0 \\ 0 & b-\lambda & 0 \\ 0 & 0 & c-\lambda \end{vmatrix} = (a-\lambda)(b-\lambda)(c-\lambda) = 0$$

The eigenvalues of a diagonal matrix are simply it's diagonal entries. The same holds true for triangular matrix, hessenberg, and tridiagonal matrices as their determinants are all products of the diagonal entries.

Example

Shortcuts for 2×2

$$\begin{aligned} \begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} &= (a-\lambda)(d-\lambda) - bc \\ &= \lambda^2 - (a+d)\lambda + (ad - dc) \end{aligned}$$

which could be simplified in terms of A's trace and determinant:

$$\lambda^2 - \operatorname{tr}(A)\lambda + \det(A)$$

and applying the quadratic formula:

$$\lambda = \frac{\operatorname{tr}(A) \pm \sqrt{\operatorname{tr}^2(A) - 4\det(A)}}{2}$$

Numerical Methods

One thing to note is that solving higher degree polynomials is not always practical, in fact it can be rather tedious and so there are other iterative methods to finding eigenvalue that are more efficient. However we will discuss those more in depth when we talk about numerical linear algebra.

Diagonalization

Diagonalizing a Matrix

$$A = P^{-1}DP$$

Orthogonally Diagonalizing a Matrix

Jordan Canonical (Normal) Form

$$A = P^{-1}JP$$

Singular Value Decomposition

One big question that should be popping into your head about diagonalization is: **what about rectangular matrices**??

$$X = U\Sigma V^T$$

$$X = \hat{U}\hat{\Sigma}V^T$$

This is known as the *economy SVD.*

Advanced Vector Spaces

At this point we are going to be touching on a lot more advanced mathematics and I will be using and talking about advanced concepts from other parts of mathematics. If things become too difficult take the time to go back and look at the other videos and pages I've done, particularly in terms of set theory and abstract algebra.

Special Vector Operations

Inner Product

The inner product is a binary operation that takes two vectors and returns a scalar element. This naturally allows us to define a *norm* for each vector $||x||$.

If V is a vector space over $\mathbb{F}$ then the inner product $\langle \cdot, \cdot \rangle$ is defined as:

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$$

The defining properties of the inner product are:

1. Linearity in the first argument

$$\langle ax.y \rangle = a\langle x, y \rangle$$
$$\langle x + y, z \rangle = \langle x, y \rangle + \langle z, y \rangle$$

2. Conjugate Symmetry

$$\langle x, y \rangle = \langle y, x \rangle^*$$

3. positive definiteness

$$\langle x, y \rangle \geq 0$$

A vector space that has an inner product is called an **inner product space**. An example of an inner product space is $\mathbb{R}^n$. The *dot product* between two vectors is the inner product.

Outer Product

$$\vec{v} \otimes \vec{u} = \vec{v}\vec{u}^T$$

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} v_1 u_1 & v_1 u_2 & \cdots & v_1 u_n \\ v_1 u_1 & v_2 u_2 & & \vdots \\ \vdots & & \ddots & \\ v_n u_1 & \cdots & & v_n u_n \end{bmatrix}$$

notice that $\mathrm{tr}(v \otimes u) = \langle v, u \rangle$

A vector space endowed with an outer product operation is called an **outer product space**.

Exterior (Wedge) Product

A vector space with a wedge product is called an **exterior product space**.

Special Vector Spaces

Normed Vector Space

Banach Space

Hilbert Space

A **Hilbert Space** is an inner product space

Essentially a hilbert space is a normed banach space.

Any inner product space can be extended into a hilbert space

Axioms

Outer Product

Exterior Product

The exterior or **wedge** product is an abstraction of the cross product

Banach Spaces

Advanced Matrices

Quadratic Forms

$$P(x) = y^T A x$$

Take a the curve:

$$(x+1)x^2 + x^2 y + x$$

This can be represented via the dot product:

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

Which can be written as the following transformation:

$$\left(\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x^2 \\ x \\ 1 \end{bmatrix}\right)^T \begin{bmatrix} y^2 \\ y \\ 1 \end{bmatrix}$$

Applying the transpose we get:

$$\begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} y^2 \\ y \\ 1 \end{bmatrix}$$

Now this is excluding any coefficients, for general coefficientes a, b, c we have:

$$\begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} y^2 \\ y \\ 1 \end{bmatrix}$$

which can be simplified as:

$$\begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} 0 & b & 0 \\ a & 0 & c \\ a & 0 & 0 \end{bmatrix} \begin{bmatrix} y^2 \\ y \\ 1 \end{bmatrix}$$

So for a general we have:

$$P(x, y) =$$

Example in 3 variables:

$$(x+1)y^2 + zx^2 y + z^2$$

$$\left(\begin{bmatrix} 1 \\ x^2 y \\ (x+1)y^2 \end{bmatrix}\right)^T \begin{bmatrix} z^2 \\ z \\ 1 \end{bmatrix}$$

The left vector we can already see as being our prev

$$\left(\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} (x+1)y^2 \\ x^2 y \\ 1 \end{bmatrix}\right)^T \begin{bmatrix} z^2 \\ z \\ 1 \end{bmatrix}$$

$$\left(\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} y^2 \\ y \\ 1 \end{bmatrix}\right)^T \begin{bmatrix} z^2 \\ z \\ 1 \end{bmatrix}$$

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

In general, a polynomial in n variables is given recursively by:

$$P(x_1, x_2, .., x_n) = P^T(\vec{x}_1, \vec{x}_2, .., \vec{x}_{n-1}) \beta C \vec{x}_n$$

Where β is our basis and C is the coefficients for the vector representing powers of the n-th variable.

Matrix Applications

Markov Matrices

Linear Dynamical Systems (LDS)

Matrix Decompositions

These are some useful and powerful optimizations and algorithms for matrices. As well as other applications

LU Decomposition

The LU decomposition is useful for calculating things like determinants for general $n \times n$ matrices. We know that the determinant for a triangular matrix is simply the product of the diagonal entries, and the determinant of the product of two matrices is the product of their determinants:

So essentially if we can decompose a matrix A into:

$$A = LU$$

Where U is *upper* triangule and L is *lower* triangular, then $\det A$ is simply:

$$\det A = \det L \cdot \det U$$

Turning A into an upper triangular matrix is pretty easy, simply put A into REF, but how can we detrmine L?

This turns out to be even more simple. Remember how we can represent row reductions as transformations on A via multiplication by elementary matrices?

$$EA = U$$

Well as it turns out, *if you used pivots to find REF*, then that matrix E is lower triangular. This makes sense when you realize that all we're really doing is taking the pivot and multiplying downwards by the lowest common factor with the next row.

$$\begin{matrix} a_1 x + b_1 y + ... \\ a_2 x + b_2 y + ... \\ a_3 x + b_3 y + ... \\ \vdots \end{matrix} \quad \rightarrow \quad \begin{matrix} a_1 x + b_1 y + ... \\ a_2 x - \left(\frac{a_2}{a_1} a_1 x\right) + b_2 y - \left(\frac{a_2}{a_1} b_1 y\right) + ... - \dfrac{a_2}{a_1}(... \\ a_3 x - \left(\frac{a_3}{a_1} a_1 x\right) + b_3 y - \left(\frac{a_3}{a_1} b_1 y\right) + ... - a_3(... \\ \vdots \end{matrix}$$

If we were to continue this and apply that transformation to the identity matrix it would simply look like:

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\frac{a_2}{a_1} & 1 & 0 & \cdots & 0 \\ -\frac{a_3}{a_1} & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ -\frac{a_n}{a_1} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

And we can basically repeat that for all the other pivots.

REMEMBER THAT THIS ONLY MAKES SENSE IF WE *DO NOT* SWAP ANY ROWS WHILE PUTTING OUR MATRIX INTO REF

Now for the next bit we need to think about

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a & 1 & 0 & \cdots & 0 \\ b & c & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ d & e & f & \cdots & 1 \end{bmatrix}$$

Well if we wanted to invert this matrix (i.e turn it into the identity), then we actually have a very simple algorithm. We can use pivots again and basically just subtract each coefficient, clean out the whole column, then repeat that on the next pivot.

So our inverse literally looks like the negative of anything below our diagonal:

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -a & 1 & 0 & \cdots & 0 \\ -b & -c & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ -d & -e & -f & \cdots & 1 \end{bmatrix}$$

So as we go about applying E to A, we can already get a look at it's inverse:

$$E^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \frac{a_2}{a_1} & 1 & 0 & \cdots & 0 \\ \frac{a_3}{a_1} & \left(\frac{b_3 - \frac{a_3}{a_1} b_1}{b_2 - \frac{a_2}{a_1} b_1}\right) & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ \frac{a_n}{a_1} & \left(\frac{b_n - \frac{a_n}{a_1} b_1}{b_2 - \frac{a_2}{a_1} b_1}\right) & \left(\frac{c_n - \frac{a_n}{a_1} b_1 - \frac{b_n - \frac{a_n}{a_1} b_1}{b_2 - \frac{a_2}{a_1} b_1}}{c_3 - \frac{a_3}{a_1} b_1 - \frac{b_3 - \frac{a_3}{a_1} b_1}{b_2 - \frac{a_2}{a_1} b_1 t}}\right) & \cdots & 1 \end{bmatrix}$$

Which while ugly will honestly make more sense as we go about applying our transformations.

So to wrap up, we put A into REF, U, through a series of transformations E, and encode the inverse transformations E^{-1}, which is lower triangular L.

$$\begin{gathered} EA = U \\ E^{-1}EA = E^{-1}U, \quad L = E^{-1} \\ \therefore A = LU \end{gathered}$$

We also found a shortcut method to find L.

So let's try to decompose the following matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 5 & 2 \\ 1 & 3 & 12 \end{bmatrix}$$

Let's row reduce into REF to find U:

$$\begin{bmatrix} 1 & 1 & 0 \\ 3 & 5 & 2 \\ 1 & 3 & 12 \end{bmatrix} \xrightarrow{R_2 - (3)R_1} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 1 & 3 & 12 \end{bmatrix} \xrightarrow{R_3 - (1)R_1} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 12 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 12 \end{bmatrix} \xrightarrow{R_3 - (1)R_2} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 10 \end{bmatrix} = U$$

And all we do is write the coefficients of each row we're subtracting in the appropriate places

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

And we've fully factored our matrix:

$$A = LU$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 3 & 5 & 2 \\ 1 & 3 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 10 \end{bmatrix}$$

Now we can tell that it's determinant is 20, just multiply the diagonals!!

This is obviously more efficient for larger matrices, and is how most computers will calculate a matrix's determinant.

Linear Least Squares

$$(ax_1 + b - y_1)^2 + (ax_2 + b - y_2)^2 + ... + (ax_n + b - y_n)^2$$

$$\sum_i^n ((ax_i + b) - y_i)^2$$

$b = 0$

$$\sum_i^n (ax_i - y_i)^2$$

This might look familiar, it is essentially the lenth of this vector:

$$||A\vec{x} - \vec{y}||^2$$

$$||\vec{b} - A\vec{x}||^2 =$$

$$A^T A \vec{x}^* = A^T \vec{b}$$

QR Decomposition

The QR Decomposition is an optimization strategy for solbing linear least squares problems. Essentially how it works is we want to decompose an $m \times n$ matrix A into the product of an $m \times n$ matrix Q and an $n \times n$ upper triangular matrix R.

$$A = QR$$

Now it's important that the $\text{col}(Q)$ forms an orthogonal basis. This means that the product:

$$Q^T Q = I_n$$

To find Q we will apply Gram-Schmidt to the $\text{col}(A)$. We then normalize the vectors and this will form $\text{col}(Q)$

Finding R becomes pretty easy as:

$$Q^T A = Q^T QR = I_n R$$

but why is R upper triangular?

thus R will be upper triangular.

So to tie this all back to least squares, remember that we want to find:

$$A^T A\vec{x}^* = A^T\vec{b}$$

When we apply the substitution $A = QR$ we can reduce this equation to:

$$\begin{aligned}(QR)^T QR\vec{x}^* &= (QR)^T\vec{b} \\ R^T R\vec{x}^* &= R^T Q^T\vec{b} \\ \vec{x}^* &= R^{-1}Q^T\vec{b}\end{aligned}$$

It will sometimes be the case where R will not be invertible, and in that case there is **no** least squares solution. So we can also check if there are least squares solutions by checking if $\det R \neq 0$.

So here's the process:

1. Apply Gram-Schmidt to $\text{col}(A)$ to find $\text{col}(Q)$
2. find $R = Q^T A$
3. check if $\det R \neq 0$, if zero we have no solutions
4. apply $R^{-1}Q^T\vec{b}$ to find $\vec{x}^*$

Applying non Linear Functions to a Matrix

$$e^A$$
$$\cos A$$
$$\sin A$$

$$e^A = A + \frac{A^2}{2!} + \frac{A^3}{3!} + ... + \frac{A^k}{k!} + ... + \frac{A^n}{n!}$$

Additional Matrix Operations

Direct Sum

The direct sum of two matrices is a diagonal *block* matrix

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

Multiple direct sums can be condensed with a large circle plus:

$$\bigoplus_{i=1}^{n} A_i$$

Kronecker Product and Sum

The kronecker product is an operation between matrices:

$$\begin{gathered}\otimes : M_{pm}(\mathbb{F}) \times M_{qq}(\mathbb{F}) \to M_{(p\times q)(m\times n)}(\mathbb{F}) \\ A \otimes B \mapsto A_{ij}B\end{gathered}$$

Essentially we make a new matrix where each element is the right matrix scaled by each element in the left matrix.

For example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1\begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} & 2\begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \\ 3\begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} & 4\begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 4 & 3 & 8 & 6 \\ 2 & 1 & 4 & 2 \\ 12 & 9 & 16 & 12 \\ 6 & 3 & 8 & 4 \end{bmatrix}$$

One interesting note is that:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 1 \end{bmatrix} \otimes \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{m1} & \cdots & & a_{mn} \end{bmatrix} = \begin{bmatrix} 1A & 0 & \cdots & 0 \\ 0 & 1A & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 1A \end{bmatrix}$$

If A is an $m \times n$ matrix, then we have the following identity:

$$I_m \otimes A = \bigoplus_{1}^{m} A$$

where $\oplus$ is the direct sum.

The **Kronecker Sum** is defined as:

$$A \oplus B = A \otimes I_m + I_n \otimes B$$

$$

$$

For example:

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

$$B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

Their kronecker sum is :

$$\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 4 & 0 \\ 0 & 3 & 0 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 6 & 0 & 0 \\ 7 & 8 & 0 & 0 \\ 0 & 0 & 5 & 6 \\ 0 & 0 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 2 & 0 \\ 7 & 9 & 0 & 2 \\ 3 & 0 & 9 & 6 \\ 0 & 3 & 7 & 12 \end{bmatrix}$$

Advanced Transformations

Householder Transformations

Let me pose a simple problem that we've already breifly discussed way back at the start of this booklet. Say we wanted to rotate a vector about some axis

Complex conjugation can already do this rotation above the real line.

However, how can we reflect any arbitrary vector? and how can we express this as a matrix transformation Px?

We get the vector

$$x + 2w$$

where $w = -\langle x, v\rangle v$

$$x - 2v(x^{\dagger} v)$$

We can write Px as the following matrix transformation:

$$\left(I - 2vv^{\dagger}\right) x$$

Which can be expressed in terms of an outer product on v:

$$(I - 2v \otimes v)x$$

Givens Rotations

One advantage givens rotations habe over householder transformations is that they can be easily parallelized. Additionally they are more efficient on sparse matrices.

Other Derivative Stuff that is incomplete

Say we want to take the derivative of $ax^2 + bx + c$, we know this to be $2ax + b$ from calculus class

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

We can also see that for 3 dimensions we have $D(ax^3 + bx^2 + cx + d) = 3a^2 + 2bx + c$ given by this transformation:

$$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

Which has the following construction

$$[3] \oplus [2] \oplus [1] \oplus [0]$$

$$\left(\bigoplus_{i}^{n} [i] \right) \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

$$\left(\bigoplus_{i=2}^{n} [i] \oplus \begin{bmatrix} 1 & 0 \end{bmatrix} \right) \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

If we were to apply the second derivative we would essentially be applying:

$$\left(\bigoplus_{i=2}^{n-2} [i] \oplus \begin{bmatrix} 1 & 0 \end{bmatrix} \right) \left(\bigoplus_{i=2}^{n} [i] \oplus \begin{bmatrix} 1 & 0 \end{bmatrix} \right) \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

$D^k P_n(x)$

$$\prod_{i}^{k} \left(\bigoplus_{j}^{n-(i+1)} [j] \right) \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Intro to Numerical Linear Algebra

Thanks to advances in modern technology we have been able to use computers to solve many of the matrix problems we face. However due to the way floating point numbers are stored it is impossible to accurately represent certain data. As we continously apply transformations the difference between the real data and the numerical approximations becomes larger and larger. Computer Linear Algebra Systems and Numerical Linear Algebra use a variety of techniques to minimize this error with an emphasis on efficiency.

Since numerical linear algebra is a subset of mathematics and computer science, this section will contain a lot of coding concepts, including many more intermediate to advanced coding concepts (your basically reading a tutorial on how to make numpy from scratch, this is **not** for beginners). So fair warning to the reader.

I will be writing examples in both Python and Rust.

Matrix Optimizations

Householder Transformation

$$x - 2vv^{\dagger}x$$

Which can be represented as the matrix multiplication Px where P is:

$$P = I - 2vv^{\dagger}$$

Givens Rotation

Finding Eigenvalues

Using Hessenberg Matrices

Iterative Methods

Arnoldi Iteration

Matrix Decompositions

LU

QR

SVD

Eigendecompositions

Solving Systems of Equations