Math 308: Autumn 2019

Course Notes

Robert Won

These are course notes from a linear algebra course at the University of Washington taught during Autumn 2019. The textbook was *Linear Algebra with Applications* by Jeffrey Holt, but many other sources contributed to these notes. Jake Levinson provided me with his handwritten notes from when he taught 308 at UW. Jeremy Rouse and Jason Gaddis both gave me TeXed notes from their linear algebra courses at Wake Forest (which used the textbook by Lay). I thank Jake, Jeremy, and Jason for generously sharing their materials. I also thank Liz Wicks for contributing the images of graphs, which she made when she used these notes for her 308 class.

1. Wednesday 9/25: Systems of Linear Equations (1.1)

Introduction and Syllabus

- On board: MATH 308, Robert Won, `robwon@uw.edu`, PDL C-418, Office hours: M 1:30–2:30, W 2:30–3:30pm.
- **Q:** What is linear algebra?

  If you've taken any math before, you've probably thought a good bit about one-dimensional space and two-dimensional space. If you're one of the many students who has taken multivariable calculus, then you've probably thought a good bit about 3-dimensional space.

  In linear algebra, we will study n-dimensional space, but we will only worry about "flat" things. This theory is still extremely useful, because even if things aren't flat, if we look at them closely enough, they can look flat. (This is the essential idea in calculus.)

  The main focus of the class will be getting a coherent and somewhat abstract theory of systems of linear equations. (We'll start that today.) The main objects we will talk about are vectors, matrices, systems of equations, linear transformations and vector spaces.
- **Q:** Why is it useful?

  Linear algebra plays an important role in understanding the behavior of the US economy. Linear algebra can be used to rank sports teams, or predict the outcomes of sporting events. Linear algebra plays a big role in sabermetrics, which is used to analyze baseball. The theory of eigenvalues and eigenvectors is the main theoretical tool that makes Google as awesome as it is. (See Google quote on syllabus)[1].

[1] This is a bit of a grandiose way to start the class, since we will have to start from the very basics and it will take a long time to get to cool applications. It's like Snape starting the very first Potions lesson with: "I can teach you how to bottle fame, brew glory, even stopper death."

- **Q:** What else will I get out of this course?

  In any math class, you will probably learn to be more detail-oriented. You will learn to write proofs and think carefully and deeply about mathematics, likely to a greater extent than you have in previous classes. These will help you develop the gray matter in your head, and doing that will make it easier for you to do anything in the future.
- **Syllabus highlights:** Almost everything will be on Canvas. Two kinds of homework: WebAssign will be due Thursdays at 11pm. WebAssign representatives will have student office hours in the Math Study Center: Thursday, October 3 11am–3pm and Monday, October 7 11am–3pm in the Math Study Center **This is where you should go to figure out your WebAssign issues.**

  Conceptual problems are harder but will be graded for completeness only. They will be due on Sundays at 11:59pm, via pdf uploaded to Canvas.

  There will be two midterm exams and a final. I will give you more information as we get closer to the first midterm.

  My office hours are M 1:30–2:30pm and W 2:30–3:30pm in my office (PDL C-418). You can also get help at the Center for Learning and Undergrad Enrichment (Google UW CLUE). If you have learning disabilities, contact the DRS office.
- Linear algebra will be a math course that is quite different than what you may be used to. There will be lots of new *vocabulary* and *abstraction*. I recommend reading the book before class (it is okay to not understand everything... even 20% understanding will make lecture a lot better!).

Systems of Linear Equations (1.1)

Before we spend a significant amount of time laying the groundwork for Linear Algebra, let's talk about some linear algebra you already know.

Example 1. Consider the following systems of equations:

$$\begin{array}{lll} 2x_1 - x_2 = 4 & x_1 - x_2 = 1 & x_1 + 2x_2 = 3 \\ x_1 + x_2 = -1 & -2x_1 + 2x_2 = -2 & x_1 + 2x_2 = 4 \end{array}$$

These are all systems of two linear equations in two variables. (Note that we say "x-two" for x_2). A very basic question that we want to answer is: how many solutions does a system have?

This is already our first exercise in vocabulary and abstraction. What is a linear equation? What is a solution?

Definition 2. A linear equation is an equation of the form

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

where $a_1, a_2, \ldots, a_n$ and b are constants and $x_1, x_2, \ldots, x_n$ are unknowns.

Note that we don't allow any of the variables to be multiplied together or squared. Linear means that you are allowed to multiply by constants and add. Also note that we sometimes use x, y, and z rather than x_1, x_2, and x_3. But once you move to more variables the x_i notation is nicer.

Definition 3. A solution to a linear equation in n unknowns $x_1, x_2, \ldots, x_n$ is an ordered set $(s_1, \ldots, s_n)$ such that substituting the s_i for x_i produces a true statement.

A system of linear equations is a set of linear equations in the same variables and a solution to the system is a common solution to all the equations in the system.

Example 4. Okay with this vocabulary in place, let's find solution sets to the three systems we started with. Let's solve the first system algebraically using *elimination*.

$$2x_1 - x_2 = 4$$
$$x_1 + x_2 = -1$$

There are many ways to proceed. Let's start by eliminating x_1 from the second row. First, multiply the second equation by -2.

$$2x_1 - x_2 = 4$$
$$-2x_1 - 2x_2 = 2$$

Now replace the second equation with the sum of the two equations

$$2x_1 - x_2 = 4$$
$$-3x_2 = 6.$$

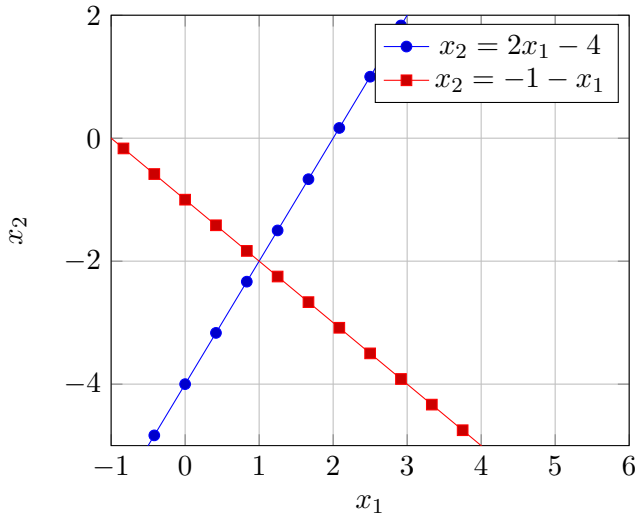
We can now solve for x_2 in the second row ($x_2 = -2$) and plug this result into the first row

$$2x_1 + 2 = 4$$
$$x_2 = -2$$

so $x_1 = 1$. What does this result tell us? It gives us the solution $(1, -2)$. You could also write this as $x_1 = 1$, $x_2 = -2$ is a solution to this system.

Is this the only solution? Would we get this solution if we did our elimination in some other order? One tool we have is to analyze this sytem geometrically.

We can graph the system $2x_1 - x_2 = 4$ by using slope-intercept form $x_2 = 2x_1 - 4$. You have likely been drawing these graphs since elementary school, but what does this graph mean? This is the picture of all of the solutions to this linear equation! The set of all points (s_1, s_2) such that (s_1, s_2) is a solution to $2x_1 - x_2 = 4$. We can graph all of the solutions to $x_1 + x_2 = -1$. And there is exactly one point that solves both linear equations, so $(1, -2)$ is the only solution.



In general, the solution set may be larger or smaller. Let's check out the second example.

Example 5. Let's try the elimination method again!

$$x_1 - x_2 = 1$$
$$-2x_1 + 2x_2 = -2$$

To eliminate the x_1 in the second equation, multiply the first by 2 and add to the second equation

$$x_1 - x_2 = 1$$
$$0x_1 + 0x_2 = 0.$$

How do we interpret this? Obviously the second row is always true. But maybe we can still try to write down all the solutions. In the first example, we found a *unique* solution for x_2. That is different here. Let's go general. Let t be any real number (sometimes denoted $t \in \mathbb{R}$). Then setting $x_2 = t$, the first equation is $x_1 - t = 1$ so $x_1 = t + 1$.

So for any real number t, we have a solution $(t + 1, t)$. The solution set is $(t + 1, t)$ for all real numbers t. There are infinitely many solutions! We call t a free parameter.

Geometrically, the two equations are both the same line. So every point on the line is a solution.

Example 6. Finally, let's look at the third example

$$x_1 + 2x_2 = 3$$
$$x_1 + 2x_2 = 4.$$

Eliminate x_1 by multiplying the first equation by -1 and adding

$$x_1 + 2x_2 = 3$$
$$0x_1 + 0x_2 = 1.$$

No choice of real numbers for x_1 and x_2 makes the second equation true! Hence, this system has no solutions.

Geometrically, these are two parallel lines. Since they do not intersect, no point is a common solution to both equations.

Definition 7. A system is called consistent if it has at least one solution. Otherwise, it is called inconsistent.

In fact, the behavior of these three examples is typical.

Theorem 8. Every system of linear equations has either no solutions (inconsistent), exactly one solution (consistent) or infinitely many solutions (consistent).

Some systems are easier to solve than others. Notice that in attempting to find solutions to the above systems, we first worked to eliminate a variable from the second equation. This gave us a new system of equations, for which it was easy to read off the solutions. We now discuss two special forms that a system of equations can have that make it easy to read off solutions.

Triangular systems

Example 9.

$$\begin{aligned} 4x_1 - 2x_2 + 3x_3 + x_4 &= 17 \\ x_2 - 2x_3 - x_4 &= 0 \\ 5x_3 + 2x_4 &= 20 \\ 3x_4 &= 15. \end{aligned}$$

This is a system of four equations in four unknowns. Henceforth for the rest of the course, we order our variables with x_1 coming first, x_2 coming second, and so on. (If there are four variables or fewer, the order is x, y, z, w.) For the first equation, x_1 is called the leading variable of the equation, since it is the first variable in the equation that occurs with a nonzero constant. Notice that each equation has a different leading variable.

Definition 10. A system of linear equations is in triangular form and is said to be a triangular system if there are n variables and n equations, and every variable is the leading variable of exactly one equation.

Triangular systems *always* have one solution. How can we find it? By a method we call *back-substitution.*

(1) Solve the last equation: $3x_4 = 15$ implies $x_4 = 5$.

(2) Substitute the value of the last variable into the previous equation to solve for the previous variable: $5x_3 + 2(5) = 20$ implies that $x_3 = 2$.

(3) Continue substituting and solving backwards through the system: $x_2 - 2(2) - 5 = 0$ implies $x_2 = 9$ and $4x_1 - 2(9) + 3(2) + 5 = 17$ implies that $x_1 = 6$.

So the solution to the triangular system is $(x_1, x_2, x_3, x_4) = (6, 9, 2, 5)$. This is the only solution.

That was nice! Of course, in general, we can't expect to be able to write a system of equations in triangular form. For one thing, the number of variables and the number of equations might be different. For another, there are some systems which have no solutions or infinitely many solutions. Therefore, we will talk about a more general "nice form" that a system of equations can have.

Echelon form

Definition 11. A system of linear equations is in echelon form if each variable is a leading variable *at most* once. Further, the equations are organized in a descending "stair step" pattern from left to right so that the indices of the leading variables increase from top to bottom.

An example would be the system

$$\begin{aligned} x_1 + x_2 \quad + \quad x_4 &= 4 \\ x_3 + x_4 &= 2 \\ x_4 &= 1 \end{aligned}$$

(Note the descending stair-step.)

Definition 12. For a system in echelon form, a variable that never occurs as a leading variable is called a free variable.

We can also use back-substitution to find solutions of echelon systems, but first we need to account for free variables.

(1) Set each free variable as a distinct free parameter: here, let $x_2 = t$ where t is a free parameter.

(2) Use back substitution to solve for the remaining variables: $x_4 = 1$, $x_3 + 1 = 2$ implies $x_3 = 1$ and $x_1 + t + 1 = 4$ implies that $x_1 = 3 - t$. Hence, $(3 - t, t, 1, 1)$ is a solution to this system for any real number t.

Example 13.

$$\begin{aligned} 2x_1 - x_2 + 5x_3 - x_4 &= -30 \\ x_3 + x_4 &= -6. \end{aligned}$$

Here, x_1 and x_3 are leading variables, x_2 and x_4 are free variables.

(1) Set each free variable as a distinct free parameter: here, let $x_2 = s$ and $x_4 = t$ be our free parameters.

(2) Use back substitution to solve for the remaining variables: $x_3 + t = -6$ so $x_3 = -6 - t$. Now $2x_1 - s + 5(-6 - t) - t = -30$ so $x_1 = 1/2s + 3t$.

So the solution set is $(x_1, x_2, x_3, x_4) = (1/2s + 3t, s, -t - 6, t)$ for any real numbers s and t.

Linear Systems and Matrices (1.2) *Key Idea:* Convert any linear system into echelon form, which we already know how to solve.

Key Idea 2: Simplify our notation!

We can arrange the system in a matrix. (A matrix is just a table of numbers[2].) First we will lay out notation for this process. Subsequently we will outline the process (Gaussian elimination).

Definition 14. An $m \times n$ matrix M is a rectangular array with m rows and n columns. We denote by M_{ij} the entry in the ith row and jth column of the matrix M.

We can represent a system of equations with a matrix.

Example 15. Consider the system

$$\begin{aligned} x_1 + 4x_2 - 3x_3 &= 7 \\ 3x_1 + 12x_2 + 2x_3 &= 0 \\ -2x_1 + x_2 + x_3 &= 1. \end{aligned}$$

Note that when we solve this system, really the only thing we need to keep track of are the coefficients in the equations. Hence, we can represent this system with the augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & 4 & -3 & 7 \\ 3 & 12 & 2 & 0 \\ -2 & 1 & 1 & 1 \end{array}\right].$$

This is called the augmented matrix because we have included the constant terms as the fourth column.

[2]One of my favorite topics in the world is irregular pluralizations! Note that the plural of the word *matrix* is *matrices*. Try not to say "matrice" when you mean "matrix" or "matrixes" when you mean "matrices."

We will simplify our systems/matrices using three types of elementary operations or row operations.

(1) Switch two equations/rows. "$R_i \leftrightarrow R_j$".

(2) Rescale an equation by a nonzero constant: "$cR_i \to R_i$".

(3) Change an equation by adding a multiple of a different equation: "$R_i + cR_j \to R_i$".

Definition 16. We say that two matrices are equivalent if one can be obtained from the other through a sequence of elementary row operations.

Let's use these row operations to get our augmented matrix into echelon form.

$$\left[\begin{array}{ccc|c} 1 & 4 & -3 & 7 \\ 3 & 12 & 2 & 0 \\ -2 & 1 & 1 & 1 \end{array}\right] \overset{R_2-3R_1\to R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 4 & -3 & 7 \\ 0 & 0 & 11 & -21 \\ -2 & 1 & 1 & 1 \end{array}\right] \overset{R_3+2R_1\to R_3}{\sim} \left[\begin{array}{ccc|c} 1 & 4 & -3 & 7 \\ 0 & 0 & 11 & -21 \\ 0 & 9 & -5 & 15 \end{array}\right] \overset{R_2\leftrightarrow R_3}{\sim} \left[\begin{array}{ccc|c} 1 & 4 & -3 & 7 \\ 0 & 9 & -5 & 15 \\ 0 & 0 & 11 & -21 \end{array}\right].$$

This corresponds to the system

$$\begin{aligned} x_1 + 4x_2 - 3x_3 &= 7 \\ 9x_2 - 5x_3 &= 15 \\ 11x_3 &= -21. \end{aligned}$$

which we can solve using back-substitution.

Question. What would happen if our echelon form had a row $\left[\begin{array}{ccc|c} 0 & 0 & 0 & c \end{array}\right]$?

This corresponds to the equation $0 = c$. If c isn't zero, this is impossible so the system is inconsistent.

Remark. There are many possible echelon forms! (Discuss). We can multiply a row by a constant and keep it in echelon form.

3. Monday 9/30: Gaussian Elimination (1.2)

The algorithm we did in the last example on Friday, where we used row operations to reduce a matrix to a matrix in echelon form is called *Gaussian Elimination*.

Example 17. Consider the system of equations

$$\begin{aligned} -y - 3z &= 1 \\ x - y - 2z &= 2 \\ 4y + 11z &= 3. \end{aligned}$$

First, we write the augmented matrix corresponding to the system.

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

Step 1 - Select the leftmost nonzero column.

Step 2 - Choose a nonzero entry in that column. Swap rows (if necessary) to move that nonzero entry to the top.

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

Step 3 - Use row replacement to make all the entries below the nonzero entry in that column equal zero.

This is already true.

Step 4 - Cover up the row which contains the nonzero entry, and keep doing steps 1-3 to the submatrix that remains.

The first nonzero column is now $\begin{bmatrix} -1 \\ 4 \end{bmatrix}$. We add four times the 2nd row to the third row and get

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

The matrix is now in echelon form.

Definition 18. The leading variables are called pivots. Their locations are called pivot positions and their columns are called pivot columns.

Example 19. We're not done yet though! *Gaussian Elimination* is the process of performing row operations until a matrix is in echelon form. But we can do more row operations to get an equivalent matrix in even nicer form. This is called *Gauss-Jordan Elimination.*

Step 5 - Scale each row so that each leading entry is equal to 1.

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

Step 6 - Use row replacement to make all the entries in a column other than a leading entry equal to 0.

We add twice the third row to the first row, and subtract three times the third row from the second row. We get

$$\left[\begin{array}{ccc|c} 1 & -1 & 0 & -12 \\ 0 & 1 & 0 & 20 \\ 0 & 0 & 1 & -7 \end{array}\right].$$

Finally, we add the second row to the first row and get

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & 20 \\ 0 & 0 & 1 & -7 \end{array}\right].$$

The matrix is now in reduced echelon form.

Definition 20. A matrix is in reduced echelon form if it is in row echelon form and

(1) The leading entry in each nonzero row is 1.

(2) Each leading 1 is the only nonzero entry in its column.

Recall that echelon form was not unique. Starting with a given matrix, you can get to many different echelon forms. However, it turns out that reduced echelon form *is* unique.

Theorem 21. Any matrix is equivalent to one and only one reduced echelon matrix.

Just so we have it all in one place

Algorithm. Row reduction algorithm (Gauss-Jordan Elimination)

(1) Begin with the leftmost nonzero column (this is a pivot column).

(2) Interchange rows as necessary so the top entry is nonzero.

(3) Use row operations to create zeros in all positions below the pivot.

(4) Ignoring the row with containing the pivot, repeat 1-3 to the remaining submatrix. Repeat until there are no more rows to modify. (The matrix is now in echelon form.)

(5) Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. Make each pivot 1 by multiplying. (The matrix is now in reduced echelon form.)

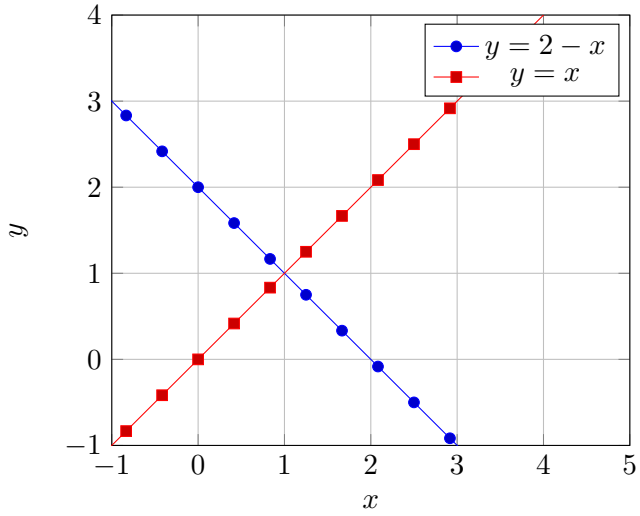
The Geometry of Gaussian Elimination.

Why does Gaussian Elimination work? That is to say, with each row operation, we are changing the equations that appear in our system of linear equations. When we change from one system to a different system via a row operation, why should we expect the new system to have the same solutions as the old one?

Let's think about this with a specific example, which should illuminate the general phenomenon. Consider the system

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

We can graph both of these equations as lines in the plane and see that they intersect at the point $(1, 1)$. Now what geometric effect do the elementary row operations have on this picture?



(1) Of course switching the two equations does not change the picture, we will still graph the exact same lines.

(2) What about scaling? Let's scale the first equation by, say, -2:

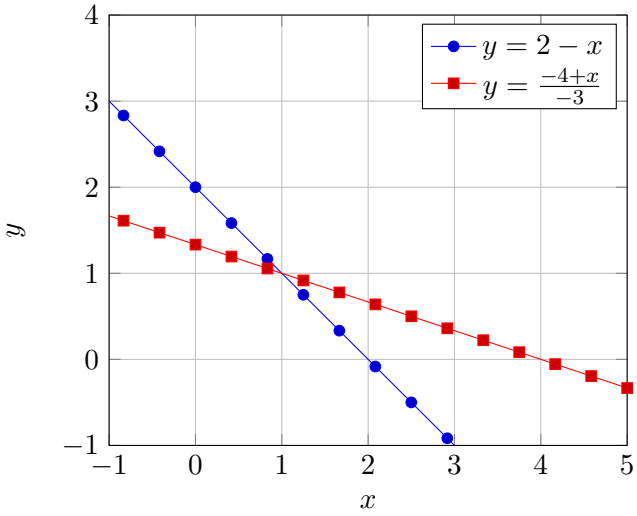
$$-2x - 2y = -4$$
$$x - y = 0.$$

Scaling also doesn't change the line (i.e., the set of solutions to the first equation).

(3) The most interesting one is adding a multiple of one equation to another equation. So let's replace the second equation with -2 times the first one plus the second one:

$$x + y = 2$$
$$-x - 3y = -4.$$

This time we have legitimately changed the second line. But geometrically we've just pivoted around the solution to the system. This new system is not the *same* as the old system but does have the *same solutions*.



We can also picture this in higher dimensions. If you have two planes in $\mathbb{R}^3$, defined by two equations in three variables, and these planes intersect in a line, then performing a row operation pivots the planes along this common line.

Homogeneous Systems.

Definition 22. A homogeneous linear equation is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

A homogeneous linear system is a system of linear equations in which every equation is homogeneous. Such a system is always consistent since $x_1 = 0$, $x_2 = 0$, $\ldots$, $x_n = 0$ is always a solution, called the trivial solution. There may be other solutions, which are called nontrivial solutions.

A Crash Course on Vectors.

Two of the key players in linear algebra are systems of linear equations, and matrices, both of which we've already talked about. Today we introduce vectors. You have already taken Math 126, but if you need a quick refresher, see pages 53–57 of your textbook.

Definition 23. A vector is an ordered list of real numbers displayed either as a column vector or a row vector:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad \text{or} \quad \mathbf{w} = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix}.$$

Each entry in $\mathbf{v}$ is called coordinate or component. Often (and in your book) vectors are denoted by boldface letters. Since this is difficult to achieve at the board, it is standard to use hats or arrows to decorate your vectors, so a vector might be denoted $\mathbf{v}$ or $\vec{v}$.

Notation. It is standard to denote the set of all real numbers by $\mathbb{R}$. The set of all vectors with n entries from $\mathbb{R}$ is written $\mathbb{R}^n$ and is called (n-dimensional) Euclidean space.

Notation. We will also sometimes use the symbol "$\in$" to denote set membership. So for example, $\pi \in \mathbb{R}$ since the number π is in the set $\mathbb{R}$. Also $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \in \mathbb{R}^2$. This notation is standard in mathematics.

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

The standard operations on vectors in $\mathbb{R}^n$ (or $\mathbb{C}^n$) are scalar multiplication and addition.

(Scalar Multiplication) For $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$,

$$c\mathbf{v} = c \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}.$$

(Addition) For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

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

The first two videos in the series by 3Blue1Brown would be good to watch here youtu.be/fNk_zzaMoSs and youtu.be/k7RM-ot2NWY.

4. Wednesday 10/2: Span (2.2)

Geometry of vectors.

We can also think about vectors geometrically. We will focus on $\mathbb{R}^2$ since this is where it is easiest to draw pictures, but the visualization can also work in $\mathbb{R}^3$ or $\mathbb{R}^n$.

We visualize a vector $\begin{bmatrix} a \\ b \end{bmatrix}$ as an arrow with endpoint at $(0,0)$ and pointing to (a,b).

Vector addition can be visualized via the tip-to-tail rule or the parallelogram rule.

Tip-to-Tail Rule. Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors. Translate the graph of $\mathbf{v}$ preserving the direction so that its tail is at the tip of $\mathbf{u}$. Then the tip of the translated $\mathbf{v}$ is at the tip of $\mathbf{u}+\mathbf{v}$.

Parallelogram Rule for Addition. If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ are represented by points in the plan, then $\mathbf{u}+\mathbf{v}$ corresponds to the fourth vertex of a parallelogram whose other vertices are $\mathbf{0}, \mathbf{u}$, and $\mathbf{v}$.

Example 24. Let $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Find $\mathbf{u}+\mathbf{v}$ geometrically and confirm that it is correct via the rule above for vector addition.

Scalar Multiplication. If $\mathbf{v}$ is a vector and c is a constant, then $c\mathbf{v}$ points in the same direction as $\mathbf{v}$. You stretch the vector by a factor of c if $c > 0$. If $c < 0$ then you flip and stretch.

Linear Combinations and Span.

Now we turn to one of the most fundamental concepts in linear algebra: linear combinations.

Definition 25. A linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ with weights $c_1, \ldots, c_m \in \mathbb{R}$ is defined as the vector $\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$.

Example 26. $\begin{bmatrix} -8 \\ 1 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} -4 \\ 1 \end{bmatrix}$ because

$$2\begin{bmatrix} 2 \\ -1 \end{bmatrix} + 3\begin{bmatrix} -4 \\ 1 \end{bmatrix} = \begin{bmatrix} -8 \\ 1 \end{bmatrix}.$$

Note that it is possible for the scalars to be negative or equal to zero.

Example 27. $-3\mathbf{v}_1$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ because $-3\mathbf{v}_1 = -3\mathbf{v}_1 + 0\mathbf{v}_2$.

Definition 28. The set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is called the span of $\mathbf{v}_1, \ldots, \mathbf{v}_m$ and is denoted $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$.

If $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = \mathbb{R}^n$ then we say "$\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ spans $\mathbb{R}^n$".

Example 29. Let $\mathbf{v}_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Is $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$?

We can draw a picture of the span of $\mathbf{v}_1$ and $\mathbf{v}_2$ and guess that the answer is yes. We can also solve algebraically. For $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ to be in the span of $\mathbf{v}_1$ and $\mathbf{v}_2$, there need to be two real numbers c_1 and c_2 such that

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

So we are asking if there are real numbers c_1 and c_2 so that

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

which is the same as asking if there are any solutions to the system

$$-2c_1 = 1$$
$$c_1 + c_2 = 1.$$

So our very natural question about spans of vectors is equivalent to the question of solving a system of linear equations. And luckily we already know how to solve a system of equations. A solution is given by $c_1 = -1/2$ and $c_2 = 3/2$.

In fact, if we think about the picture we drew, we might guess that for *any* $\begin{bmatrix} x \\ y \end{bmatrix}$ in $\mathbb{R}^n$, that $\begin{bmatrix} x \\ y \end{bmatrix}$ is in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. We can in fact verify this:

$$\left[\begin{array}{cc|c} -2 & 0 & x \\ 1 & 1 & y \end{array}\right] \sim \left[\begin{array}{cc|c} -2 & 0 & x \\ 0 & 1 & \frac{1}{2}x + y \end{array}\right]$$

so we can choose constants $c_1 = -1/2x$ and $c_2 = 1/2x + y$ to show that $\begin{bmatrix} x \\ y \end{bmatrix}$ is in the span of $\mathbf{v}_1$ and $\mathbf{v}_2$.

Example 30. Find all solutions to the system

$$-2x_2 + 2x_3 = -6$$
$$x_1 + 2x_2 + x_3 = 1$$
$$-2x_1 - 3x_2 - 3x_3 = 1.$$

Form the corresponding augmented matrix and row reduce

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

Hence, for each real number t, $x_1 = -5 - 3t, x_2 = 3 + t, x_3 = t$ is a solution.

Example 31. Is $\begin{bmatrix} -6 \\ 1 \\ 1 \end{bmatrix}$ in the span of $\begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$?

This is the same thing as asking if there exist constants x_1, x_2, x_3 such that

$$x_1 \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -6 \\ 1 \\ 1 \end{bmatrix}$$

i.e., whether $\begin{bmatrix} -6 \\ 1 \\ 1 \end{bmatrix}$ is a linear combination of the other three vectors. But this reduces to solving the same system of equations we solved above. Any of the infinitely many solutions to the linear system gives us the weights x_1, x_2, x_3. In particular, you can choose $x_1 = -5, x_2 = 3, x_3 = 0$.

The process we went through in the last examples works in general, as the following theorem says.

Theorem 32. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^n$. Then $\mathbf{b} \in \operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ if and only if the linear system with augmented matrix

$$\left[\begin{array}{cccc|c} | & | & & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m & \mathbf{b} \\ | & | & & | & | \end{array}\right]$$

has a solution.

Note. The phrase "A if and only if B" means that either both A and B are true or both A and B are false (see Math 300).

Theorem 33. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$. Suppose $\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix} \sim M$ where M is in echelon form. Then $\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = \mathbb{R}^n$ if and only if the "staircase" reaches the bottom of M (i.e., every row has a pivot).

5. Friday 10/4: More on Span (2.2)

Example 34. Determine whether $\mathbf{w} = \begin{bmatrix}1\\1\\1\end{bmatrix}$ is in the span of $\mathbf{u} = \begin{bmatrix}3\\-1\\1\end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix}2\\1\\-3\end{bmatrix}$.

Again, we are asking whether there exists c_1, c_2 such that $c_1\mathbf{u} + c_2\mathbf{v} = \mathbf{w}$, which is equivalent to

$$\begin{bmatrix}3\\-1\\1\end{bmatrix} c_1 + \begin{bmatrix}2\\1\\-3\end{bmatrix} c_2 = \begin{bmatrix}1\\1\\1\end{bmatrix} \Leftrightarrow \begin{bmatrix}3c_1 + 2c_2\\-c_1 + c_2\\c_1 - 3c_2\end{bmatrix} = \begin{bmatrix}1\\1\\1\end{bmatrix}.$$

Luckily, we know how to solve such a system. We form the corresponding augmented matrix and row reduce

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

Since the bottom row of the augmented matrix is $\left[\begin{array}{cc|c}0 & 0 & 9\end{array}\right]$, the system has no solution, so $\mathbf{w}$ is not a linear combination of $\mathbf{u}$ and $\mathbf{v}$, so is not in their span.

Another conclusion we can draw from this is that $\{\mathbf{u}, \mathbf{v}\}$ does not span $\mathbb{R}^3$.

Example 35. Does $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ span $\mathbb{R}^3$?

This is the same thing as asking, for any $\mathbf{b} \in \mathbb{R}^3$, is $\mathbf{b}$ a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$? Luckily, we've already done most of the work for this problem! The augmented matrix corresponding to the system of equations is

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

whatever vector $\mathbf{b}$ is, if we row reduce this matrix, there is some vector $\mathbf{b}'$ such that the augmented row reduces to

$$\left[\begin{array}{ccc|c}1 & -3 & 1 & |\\0 & 1 & -1 & \mathbf{b}'\\0 & 0 & 9 & |\end{array}\right].$$

This system will have a solution, so it is possible to write $\mathbf{b}$ as a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$.

Example 36. How can we picture spans?

Well Span$\{\mathbf{0}\}$ is all of the linear combinations of $\mathbf{0}$. So it is just the single point the origin.

If $\mathbf{v}$ is not $\mathbf{0}$ then Span$\{\mathbf{v}\}$ is a line through the origin. For example, Span $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ is just the x-axis in $\mathbb{R}^3$.

What about the span of two vectors? Span $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ can be pictured as the x-y plane in $\mathbb{R}^3$.

This is more or less what we said on the first day of class. Linear algebra studies "flat things" in space. So the span of a nonzero vector is a 1-dimensional flat thing through the origin (a line), the span of two vectors can be a 2-dimensional flat thing (a plane).

Of course it doesn't need to be. Span $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} \right\}$ = Span $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ is also just a line through the origin.

What happened in the previous example is that $\mathbf{u}$ and $\mathbf{v}$ determined some plane through the origin. The vector $\mathbf{w}$ was not on this plane. In fact, you can check that any vector in the span of $\{\mathbf{u}, \mathbf{v}\}$ lies on the plane $2x + 11y + 5z = 0$. If you'd like a challenge, try to solve for the plane yourself (there are many ways to do this).

The theorem from last class also has the following corollary[3].

Corollary 37. If $m < n$, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ *cannot* span $\mathbb{R}^n$.

That is to say, you need *at least* n vectors to have any hope of spanning $\mathbb{R}^n$.

[3] A corollary is a theorem that follows from another theorem. If you're British or Canadian you put the emphasis on the second syllable. Americans put the emphasis on the first.

Let's get some idea about why this theorem is true. We will give a proof. In this course, you're not expected to write proofs on homework or on an exam, but you *should* be able to understand the logic and concepts behind this proof. And when you understand a proof, you really need to understand how the different pieces fit together.

Proof. Suppose that $m < n$ and you have m vectors in $\mathbb{R}^n$. By Theorem 33, $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ spans $\mathbb{R}^n$ if and only if the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_m$ row reduces to a matrix with a pivot in every row. But this matrix will be $n \times m$, with n rows and m columns. Each column can contain at most one pivot, so in echelon form, there will be at most m pivots. Since $m < n$, there must be a row that does not have a pivot. So $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ *cannot* span $\mathbb{R}^n$. □

A geometric reason: m vectors can span at most an "m-dimensional plane" in $\mathbb{R}^n$, and $\mathbb{R}^n$ is n-dimensional.

Question. If $m \geq n$, is $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = \mathbb{R}^n$ guaranteed?

No! for example, $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \right\}$ does not span $\mathbb{R}^3$. Any linear combination of these vectors will have third coordinate equal to 0.

In that previous example, we notice that the third and fourth vector seem to be kind of extraneous. In fact, we have the following:

Theorem 38. Suppose $\mathbf{x} \in \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$. Then $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{x}\} = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ (i.e., there is no benefit to adding $\mathbf{x}$).

So in the example above, $\text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \right\} = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \right\} = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ and we already know that the span of two vectors can never be $\mathbb{R}^3$.

Matrix equations.

One new piece of notation. Suppose that $A = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$ (where since we used the same variable m, this means that the number of columns of A is equal to the number of rows of $\mathbf{x}$.

Definition 39. We define $A\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m$.

If you already know how to multiply a matrix and a vector, you can pick your favorite matrix A and your favorite vector $\mathbf{x}$ and check to see that this definition makes sense.

In fact, let's do that now:

Example 40. Let's take

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 10 \\ -1 \end{bmatrix} = 10 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - 1 \begin{bmatrix} 5 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 15 \\ 7 \\ -3 \end{bmatrix}.$$

You can also just multiply the matrix and vector and see you get the same result.

This is a compact way to express a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$ (the columns of A). Again, when you see a matrix A multiplied by a vector $\mathbf{x}$, one useful way to think about it is a linear combination of columns of A, where the coefficients are given by the entries of $\mathbf{x}$.

We can also write a system of linear equations more compactly.

Example 41. Consider the system:

$$\begin{aligned} 3x_1 + 2x_2 &= 1 \\ -x_1 + x_2 &= 3 \\ x_1 - 3x_2 &= -7. \end{aligned}$$

We now have three ways to write this! The first is the system above. It is not hard to see that this is the same thing as the vector equation

$$\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} x_2 = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}.$$

We just learned that this is the same as the matrix equation

$$\begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}.$$

The equation $A\mathbf{x} = \mathbf{b}$ is much more compact than writing an entire system of linear equations!

With this new notation, we can connect several ideas in this course into one summary theorem.

Theorem 42 (Summary theorem on span)**.** Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^n$. Then *the following are equivalent* (this means that they are either *all* true or *all* false):

(1) $\mathbf{b}$ is in $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$. (This is geometric, and you can draw a picture.)

(2) The vector equation $x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m = \mathbf{b}$ has a solution. (This is algebraic.)

(3) The system corresponding to the augmented matrix

$$\left[\begin{array}{cccc|c} | & | & & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m & \mathbf{b} \\ | & | & & | & | \end{array}\right]$$

is consistent. (This is computational.)

(4) The equation $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$. (This is really just a notational way of expressing number 2 above.)

6. Monday 10/7: Linear Independence (2.3)

In this section, we learn about linear independence. Let me motivate this for a second by recalling a problem we thought about in 2.2.

Before: Given a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ in $\mathbb{R}^n$, which vectors can we obtain as linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_m$? (I.e., what is the span of $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$?)

Now: Did we need *all* of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ or were some of them redundant?

Example 43. We already saw an example of "redundancy"[4] in the last section. $\text{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}\right\} = \text{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$ since $\begin{bmatrix}1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}1\\1\\0\end{bmatrix}$. If you took some linear combination of these three vectors, you could rewrite the third vector in terms of the first two, so it is already a linear combination of the first two vectors.

On the other hand, we could also write the first vector in terms of the other two:

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

so we could think of the first vector as "redundant" rather than the third one. Or indeed, the second. Really we should move them all to the same side and say that

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

So there is a linear combination of our three vectors which equals $\mathbf{0}$. This will be our criteria for "redundancy".

Definition 44. A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ in $\mathbb{R}^n$ is said to be linearly dependent if the vector equation

$$x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m = \mathbf{0} \tag{1}$$

has a solution other than $x_1 = \cdots = x_m = 0$ (i.e., has a nontrivial solution). So there exist weights $c_1, \ldots, c_m$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m = \mathbf{0}.$$

If the only solution is the trivial solution, then the set is said to be linearly independent.

Based on our investigation before the definition, a linearly dependent set of vectors corresponds to some of the vectors being redundant.

Question. A related question if $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent, can *every* $\mathbf{v}_i$ be expressed as a linear combination of the others?

The answer is no. There are some constants $c_1, \ldots, c_m$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m = \mathbf{0}.$$

If $c_1 \neq 0$, then we can solve for $\mathbf{v}_1$ as a linear combination of the others. But if $c_1 = 0$, then we cannot solve. Here's a concrete example:

Example 45. Let $\mathbf{v}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}2\\0\\2\end{bmatrix}$. The set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly *dependent*. We should be able to understand this via our intuition, since $\mathbf{v}_3$ is "redundant" if we already have $\mathbf{v}_1$.

But mathematically, to argue that they are linearly dependent, we should use the *definition*. We need to exhibit a nontrivial linear combination of the vectors which is equal to $\mathbf{0}$. But of course

$$2\mathbf{v}_1 + 0\mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$$

so the set is linearly dependent. On the other hand, there is no way to write $\mathbf{v}_2$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_3$.

What is the geometric idea of linear independence? Linearly *independent* vectors point in "fundamentally different" directions. On the other hand, if a set of vectors is linearly *dependent*, then at least one of the vectors points in a "redundant" direction" (i.e., is in the span of the other vectors).

Example 46. Let $\mathbf{v}_1 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}1\\1\\0\\2\end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix}2\\3\\1\\5\end{bmatrix}$. Write down the definition of the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ being linearly independent.

Here are two definitions and a non-definition. The set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent means:

(1) **Definition A.** There are no nontrivial solutions to the equation $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$.

(2) **Definition B.** If $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}$ then it must be that $c_1 = c_2 = c_3 = 0$.

(3) **Non-definition.** $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}$ and $c_1 = c_2 = c_3 = 0$.

The definition of linear independence is really an if-then statement. *If* a linear combination of the vectors equals $\mathbf{0}$, *then* it must be the trivial linear combination.

Question. How can we tell (computationally) if a set of vectors is linearly independent?

Example 47. Determine if the vectors $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 5 \end{bmatrix}$ are linearly independent.

We want to understand the solutions to $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$. If there is only the trivial solution, the vectors are linearly independent. If there are nontrivial solutions, the vectors are linearly dependent. This system corresponds to the augmented matrix

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

Since the system is homogeneous, it is automatically consistent. It has the trivial solution $x_1 = x_2 = x_3 = 0$.

If that is the *unique* solution to this system, then there is no nontrivial linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ which equals $\mathbf{0}$, so the set is linearly independent.

If there is another nontrivial solution, then there will be infinitely many solutions. If this is the case, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

So we should row reduce the augmented matrix and see if there is a unique solution or infinitely many solutions to the corresponding linear system.

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

and the system corresponding to this matrix has a free variable, namely x_3. Set $x_3 = t$. Then $x_2 = -2t$ and $x_1 = -t$. This gives a solution for any real number t.

Since there are nontrivial solutions to this system, the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent. In fact, the solutions we found above tell us which linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is equal to $\mathbf{0}$. For example, if we let $t = 1$, then $-\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$, which you can easily verify.

The previous example works as a general computational tool to determine if some given vectors are linearly independent. What did we end up doing? We ended up taking the vectors, putting them as the columns of a matrix, and reducing the matrix to echelon form. Now every homogeneous system always has the trivial solution. So if there is going to be another solution, it must be that our echelon form matrix had a free variable. The following theorem records this fact.

Theorem 48. Suppose $\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix} \sim M$ where M is in echelon form. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly independent if and only if M has no free variables (i.e., every "step" of the staircase is one column wide or equivalently, every column has a pivot).

Question. How many linearly independent vectors can you have in $\mathbb{R}^n$?

Our intuition tells us that linearly independent vectors should correspond to different "directions" so our guess should be n.

Theorem 49. If $m > n$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a set of vectors in $\mathbb{R}^n$, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent.

Proof. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be the columns of a matrix A. Reduce A to a matrix M in echelon form. The theorem above says that we can detect linear independence of the set by the existence of free variables in M. But because $m > n$, M has more columns than rows. Hence, there must be a free variable!

Therefore $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions and $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a linearly dependent set. □

Question. Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ span $\mathbb{R}^n$ *and* the set is linearly independent. What does this mean about m and n?

The theorem above says that for the vectors to be linearly independent, we must have $m \leq n$. On the other hand, if $m < n$ then we already saw that m vectors cannot span n. Hence, the only choice we have is $m = n$!

To give you a non-rigorous picture in your head "linearly independent" means "pointing in different directions." So the more vectors you have, the harder it is to be linearly independent.

On the other hand, "spanning $\mathbb{R}^n$" means "pointing in every direction." So the fewer vectors you have, the harder it is to span.

7. Wednesday 10/9: More on Linear Independence (2.3)

Recall from last time that we said "linearly independent" means "pointing in different directions." So the more vectors you have, the harder it is to be linearly independent. On the other hand, "spanning $\mathbb{R}^n$" means "pointing in every direction." So the fewer vectors you have, the harder it is to span.

A set that both spans and is linearly independent is both a *minimal* spanning set and a *maximal* linearly independent set. Such sets of vectors are very important and we will study them more in the coming weeks.

One last way to think about linear independence is in terms of *uniqueness* of solutions to equations.

Say $A = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix}$. Remember that $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ spans $\mathbb{R}^n$ if and only if *every* $\mathbf{b} \in \mathbb{R}^n$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$. That is, they span if and only if the equation $A\mathbf{x} = \mathbf{b}$ always has *at least* one solution.

There is also a way to think about linear independence in this kind of way.

Theorem 50. The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly independent if and only if the equation $A\mathbf{x} = \mathbf{b}$ has *at most* one solution for any $\mathbf{b} \in \mathbb{R}^n$.

That is, there is only one way (or no way) to represent each $\mathbf{b} \in \mathbb{R}^n$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

Before we explain why this theorem is true, let's look at an example.

Example 51. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$. Find two ways to write $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

Let's just think about the set up for a second. Since we have three vectors in $\mathbb{R}^2$, we know that they cannot possibly be linearly independent. So they are linearly dependent. And if they are linearly dependent then the theorem above says that as long as there is one way to write $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, there should be infinitely many ways.

How do we find them? Well this is same as solving the equation $A\mathbf{x} = \mathbf{b}$ where the columns of A are $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ and $\mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. So we form the augmented matrix and reduce

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

The solutions to this system are given by $x_1 = 3t + 1$, $x_2 = -2t - 1/2$, $x_3 = t$ for any real number t. So, for example $\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2 = 4\mathbf{v}_1 - \frac{5}{2}\mathbf{v}_2 + \mathbf{v}_3$.

Now let's prove Theorem 50. This proof was omitted from lecture, but it's provided here for you to take a look at.

Proof. Suppose that $A\mathbf{x} = \mathbf{b}$ has at most one solution for any $\mathbf{b}$ in $\mathbb{R}^n$. Then in particular, $A\mathbf{x} = \mathbf{0}$ has at most one solution. But $A\mathbf{x} = \mathbf{0}$ always has the trivial solution $\mathbf{x} = \mathbf{0}$ so that is the unique solution. By the definition of linear independence, the columns of A are linearly independent.

Now suppose that the columns of A are linearly independent. Suppose that for some $\mathbf{b}$ in $\mathbb{R}^n$, there are a $\mathbf{y}$ and $\mathbf{z}$ in $\mathbb{R}^m$ such that $A\mathbf{y} = \mathbf{b}$ and $A\mathbf{z} = \mathbf{b}$. If we show that $\mathbf{y} = \mathbf{z}$, then we will have proven the theorem. Let $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ and $\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$. We then have that

$$y_1\mathbf{v}_1 + \cdots + y_m\mathbf{v}_m = \mathbf{b}.$$

$$z_1\mathbf{v}_1 + \cdots + z_m\mathbf{v}_m = \mathbf{b}$$

Now subtract the second equation from the first so

$$(y_1 - z_1)\mathbf{v}_1 + \cdots + (y_m - z_m)\mathbf{v}_m = \mathbf{0}.$$

Since we assumed that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent, this must mean that $y_1 - z_1 = \cdots = y_m - z_m = 0$. Hence, $\mathbf{y} = \mathbf{z}$. □

Example 52. The proof of the theorem also shows us that if we have some set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ and two different linear combinations of these vectors are equal to the same vector $\mathbf{b}$, then we can subtract them to show that the $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly dependent.

So, for example, in Example 51, we found that

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4\mathbf{v}_1 - \frac{5}{2}\mathbf{v}_2 + \mathbf{v}_3.$$

Subtracting, we see that

$$3\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$$

which shows that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly dependent.

So, to restate our intuition from earlier. A set of vectors is linearly independent if there exists a unique way to get to any (reachable) vector by taking linear combinations. They span $\mathbb{R}^n$ if it is possible to reach any vector.

Question. Say $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a linearly independent set and $\mathbf{w}$ is *not* in $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$. What can we say about $\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{w}\}$?

The answer is that $\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{w}\}$ should be linearly independent. Why? Well suppose

$$c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m + d\mathbf{w} = \mathbf{0}.$$

If $d \neq 0$, then we can write

$$\mathbf{w} = -\frac{1}{d}(c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m)$$

so $\mathbf{w}$ is in $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$. But this is impossible since we assumed that it was not true.

Therefore $d = 0$. But now since $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent, this means that $c_1 = c_2 = \cdots = c_m = 0$. Hence, the only linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{w}$ which equals $\mathbf{0}$ is the trivial one, so the vectors are linearly independent.

Since Theorem 50 identifies exactly when a set of vectors is linearly *independent*, it should also identify exactly when a set of vectors is linearly *dependent*. We record the restatement of THeorem 50 here.

Theorem 53. The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ in $\mathbb{R}^n$ is linearly dependent if and only if it is possible to find some $\mathbf{b}$ in $\mathbb{R}^n$ so that $A\mathbf{x} = \mathbf{b}$ has more than one solution.

Let's summarize our knowledge about linearly (in)dependent sets of vectors into one summary theorem:

Theorem 54 (Summary theorem for linear (in)dependence)**.** Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be vectors in $\mathbb{R}^n$ and let A be the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_m$. The following are equivalent:

(1) The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent.

(2) The equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution (i.e., there is a nontrivial $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$ so that $x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m = \mathbf{0}$).

(3) Some $\mathbf{v}_i$ is in the span of the other $\mathbf{v}_j$'s.

(4) The echelon form of A has a non-pivot column (i.e., the corresponding echelon system has a free variable).

(5) There exists a choice of $\mathbf{b}$ so that $A\mathbf{x} = \mathbf{b}$ has more than one solution (namely, any $\mathbf{b}$ in $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$).

8. Friday 10/9: The Unifying Theorem (2.3)

Example 55. An example to make sure we know how to apply the computations that we learned. Consider the set

$$\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -3 \\ 7 \\ 14 \end{bmatrix} \right\}.$$

Answer the following:

(1) Is the set linearly independent?

(2) Does the set span $\mathbb{R}^3$?

(3) Is $\begin{bmatrix} 1 \\ 11 \\ 30 \end{bmatrix} \in \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\mathbf{v}_3\}$?

(4) How many ways can $\begin{bmatrix} 1 \\ 11 \\ 30 \end{bmatrix}$ be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$?

The computational tool to answer these questions is to put $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ as the columns of a matrix, reduce the matrix to echelon form, and use our theorems. This will tell us, for example, that the vectors are not linearly independent and that they do not span $\mathbb{R}^3$.

But thinking geometrically, it seems like these two concepts are related in this case. Once we know that the vectors are not linearly independent, we know that they span some plane in $\mathbb{R}^3$ (since the first two vectors are not scalar multiples of each other, they span more than just a line). Since they span a plane, they can't possibly span $\mathbb{R}^3$! Of course, in general, the concepts of spanning and linear independence need not be that related.

Example 56. For each of the following, give an example of a set of vectors in $\mathbb{R}^n$ that is:

(1) linearly independent but does not span $\mathbb{R}^n$,

(2) spans $\mathbb{R}^n$ but is not linearly independent,

(3) both spans $\mathbb{R}^n$ and is linearly independent,

(4) neither spans $\mathbb{R}^n$ nor is linearly independent.

Some possible example solutions are

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

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

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

(4) $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$.

For example (3) above, we *needed* to take 3 vectors in $\mathbb{R}^3$. When you have exactly n vectors in $\mathbb{R}^n$, the notions of spanning $\mathbb{R}^n$ and linear independence are very closely related.

Theorem 57 (Unifying theorem)**.** Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$ (Note that the n's match; there are the same number of vectors as the dimension of the space). Let A be the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ (so a square matrix). The following are equivalent

(1) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans $\mathbb{R}^n$ ("enough vectors")

(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are linearly independent ("nothing redundant")

(3) The equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for *every* $\mathbf{b} \in \mathbb{R}^n$.

One way to prove of this theorem is to put the vectors as the columns of a matrix A and reduce to echelon form. Since A is a square matrix, there will be a pivot in every row if and only if there is a pivot in every column. The third part of the unifying theorem follows from the two summary theorems. If the set spans $\mathbb{R}^n$, then for every $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ has at least one solution. Also if the set spans, then it must be linearly independent and so for every $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ has at most one solution.

Introduction to Linear Transformations (3.1)

In this chapter, we introduce yet another of the key players in Linear Algebra: *linear transformations.*

Key Idea: A linear transformation will be a *function* that inputs vectors and outputs other vectors (with some additional nice properties).

Linear transformations are important functions in geometry (rotating and resizing shapes). But they also come up in optimization, statistics/data analysis, economics, ...

Example 58. Say we decide to open up a bookstore and coffeeshop on UW's campus.

Supppose that if x_1 UW students come to our store, we expect to sell $\frac{4}{5}x_1$ coffees, $\frac{1}{4}x_1$ muffins (you should try one, they're delicious), and $\frac{1}{20}x_1$ books.

If x_2 UW faculty/staff visit, we expect to sell x_2 coffees, $\frac{3}{5}x_2$ muffins (they know how good the muffins are), and $\frac{1}{10}x_2$ books.

We can represent the people coming to our shop by the vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. We can also represent our sales of coffee, muffins, and books by the vector $\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} \#\text{ coffees} \\ \#\text{ muffins} \\ \#\text{ books} \end{bmatrix}$. Then clearly

$$\mathbf{s} = x_1 \begin{bmatrix} 4/5 \\ 1/4 \\ 1/20 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 3/5 \\ 1/10 \end{bmatrix}.$$

Given the "input vector" of customers $\mathbf{x} \in \mathbb{R}^2$, we have the "output vector" of sales $\mathbf{s} \in \mathbb{R}^3$.

So we'd like to study nice functions that take input vectors from some $\mathbb{R}^m$ and output vectors to some $\mathbb{R}^n$.

In order to do this, we should first do a brief review of functions.

Review of Functions.

The notation $T : \mathbb{R}^m \to \mathbb{R}^n$ means "T is a function from $\mathbb{R}^m \to \mathbb{R}^n$" or "$T$, a function from $\mathbb{R}^m$ to $\mathbb{R}^n$", depending on context.

The name of the function is T.

The *domain* of T is $\mathbb{R}^m$ (The set of all possible inputs).

The *codomain* of T is $\mathbb{R}^n$.

The *range* of T is the subset of $\mathbb{R}^n$ consisting of all vectors $\mathbf{w}$ such that $\mathbf{w} = T(\mathbf{x})$ for some $\mathbf{x} \in \mathbb{R}^m$.

The codomain and range may be different. The range consists of all values that the function T actually outputs. The codomain may be bigger; it is just some space where all the outputs live.

Example 59. A familiar function from calculus. We can consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. The domain and codomain are both $\mathbb{R}$. But the range of f is only $\mathbb{R}^{\geq 0}$, the non-negative real numbers.

You could define a similar function $g : \mathbb{R} \to \mathbb{R}^{\geq 0}$ by $g(x) = x^2$. Now the domain is $\mathbb{R}$, and the codomain and range are both $\mathbb{R}^{\geq 0}$.

Example 60. In the bookstore/coffeeshop example, we had the function $T : \mathbb{R}^2 \to \mathbb{R}^3$ where vectors in $\mathbb{R}^2$ represented customers and the vectors in $\mathbb{R}^3$ represented sales. In particular, the function was given by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \frac{4}{5}x_1 + x_2 \\ \frac{1}{4}x_1 + \frac{3}{5}x_2 \\ \frac{1}{20}x_1 + \frac{1}{10}x_2 \end{bmatrix}.$$

Example 61. We can also define another function

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 x_2 \\ x_1 + x_2 + 1 \end{bmatrix}.$$

for example, $T\left(\begin{bmatrix} 5 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$. The domain of T is $\mathbb{R}^2$ and so is the codomain. What is the range? It is hard to say, but we do know that $\begin{bmatrix} 5 \\ 7 \end{bmatrix}$ is in the range of T.

Definition 62. A function $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation if

(1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$.

(2) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $c \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^m$.

Linear transformations are very special functions[5]. They behave nicely with respect to both addition and scalar multiplication.

[5]There's a well-known joke that classifying objects in mathematics as "linear" or "nonlinear" is like classifying everything in the universe as "bananas" or "non-bananas". The reason why it's useful to try to understand linear things is that if you zoom in close enough to a "nice" mathematical object, it looks approximately linear. As your classmate Michael Cunetta pointed out, if you zoom in close enough to an object in the universe, it rarely looks approximately like a banana.

The videos on linear transformations would be good to watch here https://youtu.be/kYB8IZa5AuE, https://youtu.be/v8VSDg_WQlA, and https://youtu.be/rHLEWRxRGiM

9. Monday 10/14: More on Linear Transformations (3.1)

Example 63. Let's show that $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \frac{4}{5}x_1 + x_2 \\ \frac{1}{4}x_1 + \frac{3}{5}x_2 \\ \frac{1}{20}x_1 + \frac{1}{10}x_2 \end{bmatrix}$ is a linear transformation.

(1) Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$. So $\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$. Then

$$T(\mathbf{u} + \mathbf{v}) = T\left(\begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}\right) = \begin{bmatrix} \frac{4}{5}(u_1 + v_1) + (u_2 + v_2) \\ \frac{1}{4}(u_1 + v_1) + \frac{3}{5}(u_2 + v_2) \\ \frac{1}{20}(u_1 + v_1) + \frac{1}{10}(u_2 + v_2) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{4}{5}u_1 + u_2 \\ \frac{1}{4}u_1 + \frac{3}{5}u_2 \\ \frac{1}{20}u_1 + \frac{1}{10}u_2 \end{bmatrix} + \begin{bmatrix} \frac{4}{5}v_1 + v_2 \\ \frac{1}{4}v_1 + \frac{3}{5}v_2 \\ \frac{1}{20}v_1 + \frac{1}{10}v_2 \end{bmatrix} = T(\mathbf{u}) + T(\mathbf{v}).$$

This verifies the first condition.

(2) Now also let $c \in \mathbb{R}$ so $c\mathbf{u} = \begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}$. Then

$$T(c\mathbf{u}) = T\left(\begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}\right) = \begin{bmatrix} \frac{4}{5}cu_1 + cu_2 \\ \frac{1}{4}cu_1 + \frac{3}{5}cu_2 \\ \frac{1}{20}cu_1 + \frac{1}{10}cu_2 \end{bmatrix} = \begin{bmatrix} c(\frac{4}{5}u_1 + u_2) \\ c(\frac{1}{4}u_1 + \frac{3}{5}u_2) \\ c(\frac{1}{20}u_1 + \frac{1}{10}u_2) \end{bmatrix} = c\begin{bmatrix} \frac{4}{5}u_1 + u_2 \\ \frac{1}{4}u_1 + \frac{3}{5}u_2 \\ \frac{1}{20}u_1 + \frac{1}{10}u_2 \end{bmatrix} = cT(\mathbf{u}).$$

Since T satisfies both of these properties, T is a linear transformation.

Example 64. Our other example $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1x_2 \\ x_1 + x_2 \end{bmatrix}$ is *not* a linear transformation. (We could maybe guess this, since the x_1x_2 makes it look not very linear.) We just need to show that it fails one of the properties.

$$T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

but

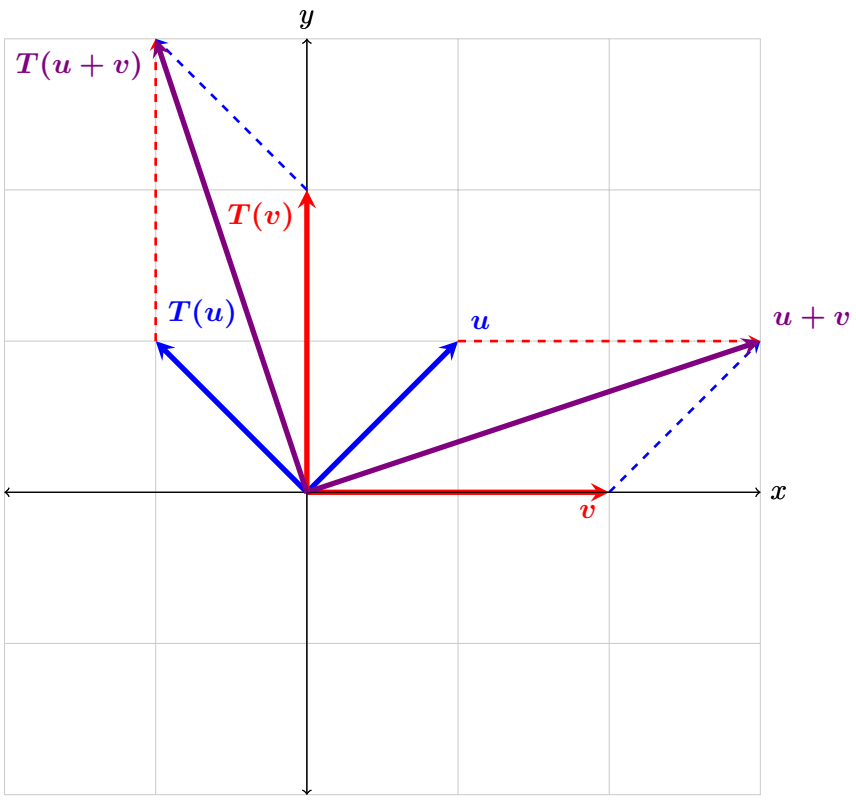
$$T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) + T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}.$$

Hence, this function is not a linear transformation.

Example 65. Here is an example of an interesting linear transformation. It is possible to show the properties computationally, but we will just motivate it geometrically.

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the function that rotates every vector counterclockwise by an angle of $\pi/4$ (or any fixed angle θ you wish).

If we take some $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$, then we can visualize the sum $\mathbf{u} + \mathbf{v}$ as the diagonal of the parallelogram determined by $\mathbf{u} + \mathbf{v}$. Now $T(\mathbf{u})$ is just $\mathbf{u}$ rotated by $\pi/4$ and $T(\mathbf{v})$ is just $\mathbf{v}$ rotated by $\pi/4$. So $T(\mathbf{u}) + T(\mathbf{v})$ is the diagonal of the parallelogram determined by $T(\mathbf{u})$ and $T(\mathbf{v})$, which is the same thing as the original diagonal rotated by $\pi/4$. (The picture makes this explanation much clearer).



Further, if $c \in \mathbb{R}$ then $c\mathbf{u}$ is just $\mathbf{u}$ appropriately scaled by a factor of c. Then $T(c\mathbf{u})$ is this scaled vector rotated by $\pi/4$. On the other hand, $cT(\mathbf{u})$ is the rotated vector scaled by a factor of c. Hence, $T(c\mathbf{u}) = cT(\mathbf{u})$. This function is a linear transformation.

Definition 66. A matrix with n rows and m columns is an $n \times m$ matrix. If $n = m$ then we call it a square matrix.

Recall. If A is an $n \times m$ matrix and $\mathbf{x} \in \mathbb{R}^m$, then $A\mathbf{x}$ is a linear combination of the columns of A (with weights coming from $\mathbf{x}$). So $A\mathbf{x}$ is a vector in $\mathbb{R}^n$.

So multiplying by the $n \times m$ matrix A gives us a function from $\mathbb{R}^m$ to $\mathbb{R}^n$. In fact:

Theorem 67. Let A be an $n \times m$ matrix. Define the function $T : \mathbb{R}^m \to \mathbb{R}^n$ by $T(\mathbf{x}) = A\mathbf{x}$.

Then T is a linear transformation.

As a consequence, one way to show that a function $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation is to find an $n \times m$ matrix A such that $T(\mathbf{x})$ is just $A\mathbf{x}$.

Example 68. Our favorite function this lecture:

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \frac{4}{5}x_1 + x_2 \\ \frac{1}{4}x_1 + \frac{3}{5}x_2 \\ \frac{1}{20}x_1 + \frac{1}{10}x_2 \end{bmatrix} = x_1 \begin{bmatrix} \frac{4}{5} \\ \frac{1}{4} \\ \frac{1}{20} \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ \frac{3}{5} \\ \frac{1}{10} \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & 1 \\ \frac{1}{4} & \frac{3}{5} \\ \frac{1}{20} & \frac{1}{10} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Before, we showed that T was linear "by hand" from the definition. But now, we found a matrix A such that $T(\mathbf{x}) = A\mathbf{x}$, so we know that T is linear by Theorem 67.

Corollary 69. For any $n \times m$ matrix A and any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ and scalar $c \in \mathbb{R}$, we have $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ and $A(c\mathbf{u}) = c(A\mathbf{u})$.

So every matrix gives a linear transformation. We can ask the reverse question: can *every* linear transformation be described this way?

I think expecting the answer to this question to be yes is quite optimistic. There are many many linear transformations out there in the world. We just learned that multiplication by a matrix gives a linear transformation. It is very hopeful to hope that these are actually *all* of the linear transformations.

By analogy, in calculus, you learn about continuous functions $\mathbb{R} \to \mathbb{R}$. You also learn that polynomial functions are continuous. But hoping that all continuous functions are polynomials is

extremely optimistic. Indeed, we know lots of continuous functions that are not polynomials! E.g., $\sin(x)$ or e^x.

For another example, we showed that rotation by an angle θ is a linear transformation. Can we actually write that as a matrix multiplication? I would say that those of you who guessed "yes" are truly bright-eyed and optimistic while those of you who guessed "no" are the sober realists.[6]

The answer to the question, though, is surprisingly yes! The optimists take this round, and we'll see why next lecture.

[6]And if you didn't guess one way or the other... I should mention that in Dante's *Inferno*, the Vestibule of Hell is occupied by the souls of people who in life took no sides...

10. Wednesday 10/16: Even More on Linear Transformations (3.1)

Key Idea: We introduce the standard basis vectors

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

Notice that if $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$ then $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_m\mathbf{e}_m$.

Let T be *any* linear transformation. How do we write the matrix A that represents T?

$$T(\mathbf{x}) = T(x_1\mathbf{e}_1) + \cdots + T(x_m\mathbf{e}_m) = x_1T(\mathbf{e}_1) + \cdots + x_mT(\mathbf{e}_m) = A\mathbf{x}$$

where $A = \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_m) \\ | & | & & | \end{bmatrix}$.

So to get the matrix A, just let the ith column be $T(\mathbf{e}_i)$. Let's record this result as a theorem.

Theorem 70. Suppose $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation, and let $\mathbf{v}_1 = T(\mathbf{e}_1), \ldots, \mathbf{v}_m = T(\mathbf{e}_m)$. Then in fact $T(\mathbf{x}) = A\mathbf{x}$ where A is the matrix $\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_m \\ | & | & & | \end{bmatrix}$.

Consequence. Even if a linear transformation isn't initially defined by multiplication by a matrix, we can use the theorem to convert it to a matrix.

Example 71. In $\mathbb{R}^2$ let $T(\mathbf{x})$ be the linear transformation that gives counterclockwise rotation by an angle θ about the origin.

Earlier, we showed that T is linear by arguing geometrically. This means that there should be a matrix A so that $T(\mathbf{x}) = A\mathbf{x}$. We compute it using the above theorem.

The first column of A should be $T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}$ (draw a picture). While the second column is $T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}$.

So $A = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$. This is called a "rotation matrix". Note that the entries of A are *numbers*, which depend on θ, not functions. So, if T is rotation by $\pi/3$, then

$$T(\mathbf{x}) = \begin{bmatrix}1/2 & -\sqrt{3}/2\\ \sqrt{3}/2 & 1/2\end{bmatrix}\mathbf{x}.$$

So $T(\mathbf{x})$ *doesn't* apply nonlinear functions $(\cos, \sin, \text{etc})$[7] to x_1 and x_2.

Properties of linear transformations.

In this section we investigate some special properties that a linear transformation may possess.

Definition 72. A function (not necessarily linear) $T : \mathbb{R}^m \to \mathbb{R}^n$ is said to one-to-one if, for each $\mathbf{b} \in \mathbb{R}^n$ there exists *at most* one $\mathbf{x} \in \mathbb{R}^m$ such that $T(\mathbf{x}) = \mathbf{b}$.

Other ways to phrase one-to-one:

- If $T(\mathbf{x}) = T(\mathbf{y})$ then $\mathbf{x} = \mathbf{y}$.
- If $\mathbf{x} \neq \mathbf{y}$ then $T(\mathbf{x}) \neq T(\mathbf{y})$.

Note. The words injective is what professional mathematicians say.

Example 73. First a familiar example (that is not a linear transformation, it is just a function) from calculus. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. This function is not one-to-one. Why not? Because $f(1) = f(-1) = 1$. That is, f maps two different elements of the domain $\mathbb{R}$ to the same element of the codomain.

Example 74. Another example from calculus. The function $g : \mathbb{R} \to \mathbb{R}$ be given by $g(x) = e^x$ is one-to-one. For different values x and y, $e^x \neq e^y$. You can see this by drawing a graph of the function e^x. Each horizontal line (representing an element of the codomain) crosses the function at most once. So each possible output comes from at most one possible input.

Definition 75. A function (not necessarily linear) $T : \mathbb{R}^m \to \mathbb{R}^n$ is said to be onto if, for each $\mathbf{b} \in \mathbb{R}^n$ there exists *at least* one $\mathbf{x} \in \mathbb{R}^m$ such that $T(\mathbf{x}) = \mathbf{b}$.

Note. The words surjective is what professional mathematicians say.

Example 76. The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is also not onto since there is no x such that $f(x) = x^2 = -1$.

Example 77. The function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = e^x$ is also not onto since there is no x such that $g(x) = e^x = 0$.

We then drew pictures of functions by drawing their domains, codomains, and picturing the function by arrows pointing from elements of the domain to elements of the codomain. Our pictures only had finitely many elements of the domain or the codomain, so this is not what a linear transformation actually "looks" like (since $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3, ...$ are all infinite sets). Nevertheless, these pictures can help us understand the concepts of one-to-one-ness and onto-ness.

We drew pictures of functions that were one-to-one, onto, both, and neither.

Material covered on your first exam ends here.

11. Friday 10/18: One-to-one and Onto (3.1)

Example 78. The function $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} b \\ a \end{bmatrix}$ is both one-to-one and onto.

To see it is one-to-one: suppose $f(\mathbf{x}) = f(\mathbf{y})$. Write $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Then since $f(\mathbf{x}) = f(\mathbf{y})$ we have $\begin{bmatrix} x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} y_2 \\ y_1 \end{bmatrix}$. Hence, $x_2 = y_2$ and $x_1 = y_1$. Therefore, $\mathbf{x} = \mathbf{y}$, so f is one-to-one.

To see that f is onto: given an arbitrary $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \in \mathbb{R}^2$, we can find some $\mathbf{x} = \begin{bmatrix} b_2 \\ b_1 \end{bmatrix}$ so that $f(\mathbf{x}) = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$.

Example 79. The function $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a+b \\ 2a+3b \\ -3a-4b \end{bmatrix}$$

is not onto.

This is because if we set $\mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and we are searching for some $\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$ so that $T(\mathbf{x}) = \mathbf{b}$, we need

$$\begin{aligned} a + b &= 0 \\ 2a + 3b &= 0 \\ -3a - 4b &= 1. \end{aligned}$$

The corresponding augmented matrix is

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

The reduced echelon form is

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

Since there is a pivot in the last column, the system is inconsistent, so there is no $\mathbf{x}$ so that $T(\mathbf{x}) = \mathbf{b}$.

However, for every $\mathbf{b}'$ for which $T(\mathbf{x}) = \mathbf{b}'$ *does* have a solution, the solution will be unique. Hence, T *is* one-to-one.

This example suggests that the way to understand whether a linear transformation T is one-to-one or onto is to write down the matrix A corresponding to T. One-to-one-ness should be related to uniqueness of solutions or linear independence of the columns. Onto-ness should be related to existence of solutions or the columns spanning.

Using this example as motivation, we have the following theorem, which tells us exact conditions that determine the one-to-one-ness or onto-ness of a linear transformation T.

Fact. If T is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$. This is because $T(c\mathbf{v}) = cT(\mathbf{v})$ for all c and all $\mathbf{v}$. Setting $c = 0$ shows that $T(\mathbf{0}) = \mathbf{0}$.

So any linear transformation takes $\mathbf{0}$ to $\mathbf{0}$. We can detect whether a transformation is one-to-one if this is the *only* vector that gets mapped to $\mathbf{0}$.

Theorem 80. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

Proof. ($\Rightarrow$) Assume T is one-to-one. Since T is linear, $T(\mathbf{0}) = \mathbf{0}$. Because T is one-to-one, $T(\mathbf{x}) = \mathbf{0} = T(\mathbf{0})$ implies $\mathbf{x} = \mathbf{0}$. Hence $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

($\Leftarrow$) Assume $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution. Suppose $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{x} - \mathbf{y}) = \mathbf{0}$ so $\mathbf{x} - \mathbf{y} = \mathbf{0}$, so $\mathbf{x} = \mathbf{y}$ and T is one-to-one. □

This gives us an easy way to determine if a linear transformation is one-to-one. It is part of the following theorem.

Theorem 81. Let A be an $n \times m$ matrix and let $T : \mathbb{R}^m \to \mathbb{R}^n$ be defined by $T(\mathbf{x}) = A\mathbf{x}$. Then

(1) T is one-to-one if and only if the columns of A are linearly independent (i.e., $A\mathbf{x} = \mathbf{0}$ has only the trivial solution).

(2) T is onto if and only if the columns of A span $\mathbb{R}^n$ (i.e., for all $\mathbf{w} \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{w}$ has at least one solution).

(3) If $m > n$, then T is not one-to-one.

(4) If $m < n$, then T is not onto[8].

Note. You also know how to computationally verify parts (1) and (2) of the previous theorem. The columns of A are linearly independent if and only if the row echelon form of A has a pivot in every column. The columns of A span $\mathbb{R}^n$ if and only if the row echelon form of A has a pivot in every row.

Example 82. If $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 3 & -3 \end{bmatrix}$ then $T : \mathbb{R}^2 \to \mathbb{R}^3$ is not onto. The columns cannot span $\mathbb{R}^3$. To determine if it is one-to-one, we put A in echelon form $A \sim \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ so the columns are linearly independent and so T is one-to-one.

We can also add to our "Unifying theorem" which dealt with the case of n vectors in $\mathbb{R}^n$.

Theorem 83 (Unifying theorem, version 2)**.** Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$. Let A be the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. The following are equivalent

(1) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans $\mathbb{R}^n$.

(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are linearly independent.

(3) The equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for *every* $\mathbf{b} \in \mathbb{R}^n$.

(4) T is onto.

(5) T is one-to-one.

Geometrically: A linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ maps $\mathbb{R}^m$ to some "flat space" through the origin in $\mathbb{R}^n$ which could be m dimensional or smaller.

For example, $T : \mathbb{R}^2 \to \mathbb{R}^3$ could map $\mathbf{e}_1$ and $\mathbf{e}_2$ to linearly independent vectors in $\mathbb{R}^3$, in which case the image of T is the plane which is the span of $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ in $\mathbb{R}^3$.

It might also map them to two vectors which are linearly dependent, in which case the image of T would be either a line or just the origin.

Matrix Algebra (3.2)

In this section, we will learn to work with matrices. Right now our biggest tool is to take one matrix (or augmented matrix) and perform Gaussian elimination of Gauss-Jordan elimination to reduce the matrix to an equivalent one in echelon form or reduced echelon form.

There are also notions of taking two different matrices and adding them or multiplying them by a scalar.

Basics: Let's take

$$A = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 2 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 9 & 10 & 6 \\ -1 & 0 & 1 \end{bmatrix}.$$

We define $A + B$ to be the matrix

$$A + B = \begin{bmatrix} 4+9 & 0+10 & 1+6 \\ 2-1 & 2+0 & 2+1 \end{bmatrix} = \begin{bmatrix} 13 & 10 & 7 \\ 1 & 2 & 3 \end{bmatrix}.$$

Similarly, if $c \in \mathbb{R}$ we define

$$cA = \begin{bmatrix} 4c & 0c & 1c \\ 2c & 2c & 2c \end{bmatrix}.$$

Lastly, the transpose of A, denoted A^T is what you get when you exchange the rows and columns of A

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

Here we've only defined these operations by example, but the definition works in general. You add two matrices component-wise and you scalar multiply a matrix by multiplying the scalar in to every entry of the matrix.

Note.

- You can only add matrices of the same size. So if C is some 3×2 matrix, then $A + C$ is not defined.
- The zero matrix of size $n \times m$ is denoted "$0_{n,m}$" (or just "0" if it is clear from context) is the $n \times m$ matrix whose entries are all 0.

  It is the *additive identity:* for any $n \times m$ matrix M, we have

$$M + 0_{n,m} = 0_{n,m} + M = M.$$

Other basic algebraic properties are in your book as Theorem 3.11 (and transpose properties in Theorem 3.15). More or less, these properties say that addition and scalar multiplication of matrices behave the way that you expect them to.

The video on matrix multiplication and composition https://youtu.be/XkY2DOUCWMU and the video on inverses https://youtu.be/XkY2DOUCWMU would both be good to watch.

12. Monday 10/21: Exam 1

Monday was your first midterm exam.

13. Wednesday 10/23: Matrix Algebra (3.2)

Since I was out of the country attending a conference, the next two lectures were recorded as YouTube videos. Check your e-mail for a link.

We now move on to a more interesting operation.

Matrix Multiplication and Composition of Linear Transformations.

Suppose $T : \mathbb{R}^m \to \mathbb{R}^n$ and $S : \mathbb{R}^n \to \mathbb{R}^k$ are linear transformations. The composition $Q : \mathbb{R}^m \to \mathbb{R}^k$ is the function $Q(\mathbf{x}) = S(T(\mathbf{x}))$.

Notation. We write the composition as $Q = S \circ T$ and read it "S composed with T". Note that the order the functions are performed in is right-to-left. $S \circ T$ means first apply T then apply S to the result.

Fact. If T and S are linear transformations then $Q = S \circ T$ is a linear transformation.

Proof.

$$\begin{aligned} Q(\mathbf{u}+\mathbf{v}) &= S(T(\mathbf{u}+\mathbf{v})) = S(T(\mathbf{u})+T(\mathbf{v})) \text{ since } T \text{ is linear} \\ &= S(T(\mathbf{u})) + S(T(\mathbf{v})) \text{ since } S \text{ is linear} \qquad\qquad = Q(\mathbf{u}) + Q(\mathbf{v}). \end{aligned}$$

Similarly, you can prove that if c is a constant then $Q(c\mathbf{u}) = cQ(\mathbf{u})$. □

Now since Q is a linear transformation $\mathbb{R}^m \to \mathbb{R}^k$, a previous theorem tells us that Q must correspond to some matrix. Which matrix?

Well suppose $S(\mathbf{x}) = A\mathbf{x}$ and $T(\mathbf{x}) = B\mathbf{x}$ for some matrices A and B (such matrices exist since S and T are linear transformations). In particular, A is a $k \times n$ matrix and B is an $n \times m$ matrix.

We will define the product $A \cdot B$ to be the matrix corresponding to Q (of size $k \times m$).

Example 84. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $T(\mathbf{x}) = \begin{bmatrix} 1 & 1 \\ 0 & 2 \\ -1 & 0 \end{bmatrix} \mathbf{x}$ (call the matrix B) and $S : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $S(\mathbf{x}) = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{x}$ (call the matrix A).

We will find the matrix $A \cdot B$ for the linear transformation $S \circ T : \mathbb{R}^2 \to \mathbb{R}^2$. How do you find the matrix? We learned how to last class! We simply need to evaluate $S \circ T$ on the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$.

The first column of $A \cdot B$ should be

$$S \circ T(\mathbf{e}_1) = S(T(\mathbf{e}_1)) = S\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right) = A \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

Similarly, the second column of $A \cdot B$ should be

$$S \circ T(\mathbf{e}_2) = S(T(\mathbf{e}_2)) = S\left(\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right) = A \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

So we define

$$A \cdot B = \text{ matrix for } S \circ T = \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix}.$$

This example suggests how to define matrix multiplication in general.

Definition 85. Version 1 of matrix multiplication (one column at a time)

If $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_m \end{bmatrix}$ (an $n \times m$ matrix) and A is a matrix of size $k \times n$ then

$$A \cdot B = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_m \end{bmatrix}.$$

We can also compute $A \cdot B$ one entry at a time.

Definition 86. Version 2 of matrix multiplication (one entry at a time)

To get the entry in the ith row and jth column of $A \cdot B$, take the dot product of the ith row of A with the jth column of B.

This is best looked at with the example above

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

To compute the 2 in the first row and first column of $A \cdot B$, we take the first row $\begin{bmatrix} 3 & 0 & 1 \end{bmatrix}$ of A and take the dot product with the first column $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ of B. So $3 \cdot 1 + 0 \cdot 0 + 1 \cdot (-1) = 2$.

Similarly, the 3 in the first row and second column of $A \cdot B$ can be computed by dotting the first row of A with the second column of B.

Note. If you take the product of an $n \times m$ matrix with an $m \times k$ matrix, then the result is an $n \times k$ matrix.

Warnings about matrix multiplication.

(1) Sometimes $A \cdot B$ will be defined but $B \cdot A$ will not be defined. You can only multiply A and B if the number of columns of A is equal to the number of rows of B.

(2) If A is $n \times m$ and B is $m \times n$, then both AB and BA are defined, but if $n \neq m$, then AB and BA are not even the same size.

(3) If A and B are both $n \times n$, then both AB and BA are defined and the same size, but it is possible that $AB \neq BA$. Take $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Then $AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ but $BA = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

(4) It is possible for $AB = 0$ even though $A \neq 0$ and $B \neq 0$. The previous example is an example of this.

(5) It is possible that $A \neq 0$ and $B \neq C$ but $AB = AC$. Again, take $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Then $A \neq 0$ and $B \neq C$ but $AB = AC$.

This shows that you can't "divide both sides by A" even if $A \neq 0$. (In section 3.3, we will study matrices that you can "divide by", these are the invertible matrices!)

14. Friday 10/25: Matrix Algebra (3.2) and Inverses (3.3)

Things that do work well for matrix multiplication.

(1) **Associativity:** $(AB)C = A(BC)$ (assuming that the products are defined).

(2) **Distributivity:** $A(B+C) = AB + AC$ and $(A+B)C = AC + BC$.

(3) **Multiplication by 0:** $0 \cdot A = 0$.

(4) **Powers of a matrix:** When k is a positive integer and A is a *square* matrix, we can define

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

But note that $(A+B)^2 = A^2 + AB + BA + B^2 \neq A^2 + 2AB + B^2$!

(5) Define the $n \times n$ identity matrix $I_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 0 & \dots & & \dots & 1 \end{bmatrix}$.

Then $IA = A$ and $AI = A$ (for A of the appropriate size).

Inverses (3.3).

Suppose that $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation that is *both* one-to-one and onto. By our theorem on linear transformations, this is only possible if $m = n$, so really $T : \mathbb{R}^n \to \mathbb{R}^n$.

Since T is onto, for every $\mathbf{y} \in \mathbb{R}^n$, there exists a $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{y}$. But since T is one-to-one, this $\mathbf{x}$ is unique.

So this process gives us a function going in the other direction! For every $\mathbf{y} \in \mathbb{R}^n$ (the codomain), we can find one and only one $\mathbf{x} \in \mathbb{R}^n$ (the domain) such that $T(\mathbf{x}) = \mathbf{y}$.

Definition 87. The inverse function $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is the function defined by

$$T^{-1}(\mathbf{y}) = \mathbf{x} \quad \text{if and only if } \mathbf{y} = T(\mathbf{x}).$$

The inverse function T^{-1} "undoes" T. That is to say:

$$T^{-1}(T(\mathbf{x})) = \mathbf{x} \quad \text{and} \quad T(T^{-1}(\mathbf{y})) = \mathbf{y}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Fact. If T is a linear transformation, then T^{-1} is also linear.

Definition 88. Consequence and Definition. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation that is both one-to-one and onto. Write $T(\mathbf{x}) = A\mathbf{x}$ for an $n \times n$ matrix A. Then T^{-1} corresponds to a unique matrix, denoted A^{-1} called the inverse of A.

This is the unique $n \times n$ matrix such that $A \cdot A^{-1} = A^{-1} \cdot A = I$. We say that A (and A^{-1}) are invertible matrices.

Sometimes an invertible matrix is also called nonsingular. If A is not invertible, we say that it is singular.

Example 89. There is a quick formula to write the inverse of a 2×2 matrix.

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

So if $A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}$ then

$$A^{-1} = \frac{1}{(-14 + 15)} \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}.$$

Multiply these together (in both orders) to verify that you get the identity matrix.

Fact. If A and B are square matrices such that $AB = I$ then $B = A^{-1}$ (i.e., you don't need to check that $BA = I$, it is automatic).

Properties/Uses of the Inverse.

(1) If $AB = 0$ and A is invertible, then $B = 0$.
Why? If $AB = 0$ then you can multiply by A^{-1} on the *left* (remember, multiplying on the left and on the right are different, in general). So

$$\begin{aligned} A^{-1} \cdot AB &= A^{-1} \cdot 0 = 0 \\ (A^{-1}A)B &= 0 \\ IB &= 0 \\ B &= 0. \end{aligned}$$

(2) If $AB = AC$ and A is invertible, then $B = C$.

Why? Similar to the above, simply multiply on the left by A^{-1}:

$$\begin{aligned} A^{-1} \cdot AB &= A^{-1} \cdot AC \\ (A^{-1}A)B &= (A^{-1}A)C \\ IB &= IC \\ B &= C. \end{aligned}$$

(3) The *unique* solution to the equation $A\mathbf{x} = \mathbf{y}$ is given by $\mathbf{x} = A^{-1}\mathbf{y}$.
Why? Same trick again:

$$\begin{aligned} A^{-1} \cdot A\mathbf{x} &= A^{-1}\mathbf{y} \\ (A^{-1}A)\mathbf{x} &= A^{-1}\mathbf{y} \\ \mathbf{x} &= A^{-1}\mathbf{y}. \end{aligned}$$

Example 90. If we want to solve the system

$$\begin{aligned} 2x + 5y &= a \\ -3x - 7y &= b \end{aligned}$$

This corresponds to $A\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$ where $A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}$. But we already computed A^{-1} above. Hence, the solution to this system is given by

$$\mathbf{x} = A^{-1} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -7a - 5b \\ 3a + 2b \end{bmatrix}.$$

(4) $(A^{-1})^{-1} = A$. This actually just follows from the definition.

(5) If A and B are both invertible matrices of the same size then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$ (note the switch in the order).
Why? We can just check that the matrix $B^{-1}A^{-1}$ satisfies the properties of the inverse of AB. Namely:

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$

Hence, $B^{-1}A^{-1}$ is the inverse of AB.

Okay, so inverses are nice and useful. And we know a formula to write down the inverse of a 2×2 matrix. The question still remains: how to find the inverse of an invertible matrix when it's not 2×2? Next week, we'll discuss the computational algorithm.

15. Monday 10/28: Computing Inverses (3.3) and Subspaces (4.1)

Okay, so inverses are nice and useful. And we know a formula to write down the inverse of a 2×2 matrix. The question still remains: how to find the inverse of an invertible matrix when it's not 2×2? Let's derive an algorithm to compute the inverse!

Computing Inverses.

We know that if T is an invertible linear transformation given by $T(\mathbf{x}) = A\mathbf{x}$ for the invertible matrix A, then A^{-1} is the matrix for T^{-1}.

Recall that the first column of A^{-1} is given by $T^{-1}(\mathbf{e}_1)$, the second column of A^{-1} is given by $T^{-1}(\mathbf{e}_2)$, etc.

So how do we find $T^{-1}(\mathbf{e}_1)$? Call it $\mathbf{b}_1$. Well, by definition $T^{-1}(\mathbf{e}_1) = \mathbf{b}_1$ if and only if $\mathbf{e}_1 = T(\mathbf{b}_1)$ which is the same thing as $\mathbf{e}_1 = A\mathbf{b}_1$.

Hence, the vector we are looking for, $\mathbf{b}_1$ is the *solution* to $A\mathbf{x} = \mathbf{e}_1$! (Remember, since T is invertible, for every $\mathbf{y}$, there is a unique solution to $T(\mathbf{x}) = \mathbf{y}$.)

So the first column of A^{-1} is the solution to $A\mathbf{x} = \mathbf{e}_1$. How do we solve that?

$$\left[\begin{array}{c|c} & \\ A & \mathbf{e}_1 \\ & \end{array}\right] \overset{\text{reduce}}{\sim} \left[\begin{array}{ccc|c} 1 & & 0 & \\ & \ddots & & \mathbf{b}_1 \\ 0 & & 1 & \end{array}\right]$$

where since A is invertible, its reduced echelon form is the identity matrix.

Whatever is left in the augmented column is the solution, which we called $\mathbf{b}_1$, which is the first column of A^{-1}.

Similarly, to compute the second column of A^{-1}, you would

$$\left[\begin{array}{c|c} & \\ A & \mathbf{e}_2 \\ & \end{array}\right] \overset{\text{same row operations}}{\sim} \left[\begin{array}{ccc|c} 1 & & 0 & \\ & \ddots & & \mathbf{b}_2 \\ 0 & & 1 & \end{array}\right].$$

Since we have to do the same row operations in each step, we may as well combine them all into one step.

$$\left[\begin{array}{c|ccc} & & & \\ A & \mathbf{e}_1 & \dots & \mathbf{e}_n \\ & & & \end{array}\right] \overset{\text{same row operations}}{\sim} \left[\begin{array}{ccc|ccc} 1 & & 0 & & & \\ & \ddots & & \mathbf{b}_1 & \dots & \mathbf{b}_n \\ 0 & & 1 & & & \end{array}\right].$$

Which we can write as

$$\left[\begin{array}{c|c} A & I \end{array}\right] \overset{\text{same row operations}}{\sim} \left[\begin{array}{c|c} I & A^{-1} \end{array}\right].$$

The punchline: To compute the inverse of a matrix A, augment with the identity matrix, row reduce A to the identity matrix. What remains on the right-hand (augmented) side is A^{-1}.

Algorithm (Computing the inverse of a matrix)**.** $\left[\begin{array}{c|c} A & I \end{array}\right] \sim \left[\begin{array}{c|c} I & A^{-1} \end{array}\right]$.

Example 91. Let $A = \begin{bmatrix} 1 & 4 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. Let's find A^{-1}.

$$\left[\begin{array}{ccc|ccc} 1 & 4 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \overset{R_1 \leftrightarrow R_2}{\sim} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 4 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \overset{R_2 - R_1 \to R_2}{\sim} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 4 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\overset{R_1 - R_3 \to R_1}{\sim} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 4 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \overset{1/4 R_2 \to R_2}{\sim} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1/4 & -1/4 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right].$$

Hence, you can read off A^{-1} from the right-hand side: $A^{-1} = \begin{bmatrix} 0 & 1 & -1 \\ 1/4 & -1/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

You should check that this is indeed the inverse by multiplying.

Since an invertible matrix is one that corresponds to a linear transformation that is both one-to-one and onto, we can also add to our unifying theorem:

Theorem 92 (Unifying theorem, version 3)**.** Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$. Let A be the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. The following are equivalent

(1) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans $\mathbb{R}^n$.

(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent.

(3) The equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for *every* $\mathbf{b} \in \mathbb{R}^n$.

(4) T is onto.

(5) T is one-to-one.

(6) A is invertible.

Introduction to subspaces.

Now we explore *subspaces*, which are special subsets of $\mathbb{R}^n$.

Definition 93. A subspace of $\mathbb{R}^n$ is a subset S of $\mathbb{R}^n$ such that

(1) $\mathbf{0} \in S$;

(2) If $\mathbf{u}, \mathbf{v} \in S$, then $\mathbf{u} + \mathbf{v} \in S$;

(3) If $\mathbf{u} \in S$ and $c \in \mathbb{R}$, then $c\mathbf{u} \in S$.

The two latter conditions are sometimes stated "S is closed under addition and scalar multiplication."

(Another way of saying this is that if $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are vectors in S, then any linear combination of those vectors is also in S.)

Example 94. Both $\{\mathbf{0}\}$ (the zero subspace) and $\mathbb{R}^n$ (the improper subspace) are always subspaces of $\mathbb{R}^n$. Any other subspace is called proper.

Example 95. Suppose $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^n$ and let $S = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$. We should check each condition.

(1) Is $\mathbf{0} \in S$? Yes, because $\mathbf{0} = 0\mathbf{a}_1 + 0\mathbf{a}_2$ so it is indeed a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$.

(2) If $\mathbf{u}, \mathbf{v} \in S$ then is $\mathbf{u} + \mathbf{v} \in S$? If $\mathbf{u}, \mathbf{v} \in S$ that means that we can find constants $x_1, x_2, y_1, y_2 \in \mathbb{R}$ so that

$$\mathbf{u} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2$$

$$\mathbf{v} = y_1\mathbf{a}_1 + y_2\mathbf{a}_2.$$

And therefore

$$\mathbf{u} + \mathbf{v} = (x_1 + y_1)\mathbf{a}_1 + (x_2 + y_2)\mathbf{a}_2$$

so $\mathbf{u} + \mathbf{v} \in \text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$.

(3) If $\mathbf{u} \in S$ and $c \in \mathbb{R}$, then is $c\mathbf{u} \in S$? Again, write $\mathbf{u} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2$. Then

$$c\mathbf{u} = (c \cdot x_1)\mathbf{a}_1 + (c \cdot x_2)\mathbf{a}_2$$

so $c\mathbf{u} \in \text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$.

Hence, the span of two vectors is a subspace.

In fact, in the example above, we could have used more than two vectors. The same reasoning works for three, four, or any number of vectors. We record this as a theorem.

Theorem 96. If $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ and $S = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ then S is a subspace.

Example 97. Let D be the unit disk in $\mathbb{R}^2$ (the filled in unit circle).

Then $\mathbf{0} \in D$. But D is not a subspace since, for example, $\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix} \notin D$ but $\begin{bmatrix}1\\0\end{bmatrix} \in D$.

Geometrically: A subspace is a line or a plane (of any dimension) through the origin.

16. Wednesday 10/30: More on Subspaces (4.1)

Example 98. Let S be the set of solutions $\mathbf{x} \in \mathbb{R}^3$ to the system of equations

$$4x_1 + 5x_2 - x_3 = 10$$
$$-x_1 + x_2 + x_3 = 5.$$

Then S is **not** a subspace of $\mathbb{R}^3$ since $\mathbf{0} \notin S$.

Geometrically, each equation gives a plane in $\mathbb{R}^3$ (neither of which pass through the origin). Their intersection is a line that does not pass through $\mathbf{0}$.

Okay but maybe we could fix the system in the previous example by making it homogeneous so that $\mathbf{0} \in S$. Let's see what happens by looking at a different, smaller homogeneous system.

Example 99. Let $S = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x + y + 2z = 0 \right\}$.

Is S a subspace?

(1) Is $\mathbf{0} \in S$? If $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, then $x + y + 2z = 0 + 0 + 2 \cdot 0 = 0$. So $\mathbf{0} \in S$.

(2) Is S closed under addition? Suppose that

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

are both in S. This means that $u_1 + u_2 + 2u_3 = 0$ and $v_1 + v_2 + 2v_3 = 0$. Then,

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix}.$$

This is in S because

$$(u_1 + v_1) + (u_2 + v_2) + 2(u_3 + v_3) = (u_1 + u_2 + 2u_3) + (v_1 + v_2 + 2v_3) = 0 + 0 = 0.$$

Thus, $\mathbf{u} + \mathbf{v} \in S$.

(3) Finally, we have to show that if $\mathbf{u} \in S$ then $c\mathbf{u} \in S$ too. This follows because if $u_1 + u_2 + 2u_3 = 0$, then $c(u_1 + u_2 + 2u_3) = c \cdot 0$ and therefore $(cu_1) + (cu_2) + 2(cu_3) = 0$. Hence, $c\mathbf{u} \in S$. Thus, S is a subspace.

In fact the previous example also generalizes to any homogeneous system. So let's record this as a theorem and prove it..

Theorem 100. Let S be the set of solutions to a *homogeneous* system of linear equations, i.e., fix a matrix A and consider the set of all $\mathbf{x} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{0}$.

Then S is a subspace of $\mathbb{R}^m$.

Proof. Clearly $\mathbf{0} \in S$ since $A\mathbf{0} = \mathbf{0}$.

If $\mathbf{u}, \mathbf{v} \in S$ then $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. Hence $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0}$. Thus, $\mathbf{u} + \mathbf{v} \in S$.

Finally, if $c \in \mathbb{R}$ then $A(c\mathbf{u}) = cA\mathbf{u} = c\mathbf{0} = \mathbf{0}$. So $c\mathbf{v} \in S$. □

Example 101. If S is the set of solutions to

$$\begin{aligned} 4x_1 + 5x_2 - x_3 &= 0 \\ -x_1 + x_2 + x_3 &= 0 \end{aligned}$$

i.e., the set of all $\mathbf{x} \in \mathbb{R}^3$ such that

$$\begin{bmatrix} 4 & 5 & -1 \\ -1 & 1 & 1 \end{bmatrix} \mathbf{x} = \mathbf{0}$$

then S is a subspace of $\mathbb{R}^3$.

Note that this example is different from Example 98 since that example did not involve a homogeneous system.

Example 102. Define a set S in $\mathbb{R}^2$ by the following property: $\mathbf{v} \in S$ if and only if $\mathbf{v}$ has exactly one nonzero entry. Is S a subspace of $\mathbb{R}^2$?

No, since $\mathbf{0} \notin S$.

So let's change our definition a bit. Suppose we define S by $\mathbf{v} \in S$ if and only if at least one entry of $\mathbf{v}$ is zero. Then $\mathbf{0} \in S$. This set is not closed under addition since $\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix} \in S$ but $\begin{bmatrix}1\\1\end{bmatrix} \notin S$ However, S *is* closed under scalar multiplication. So S fails to be a subspace only because it is not closed under addition. To picture S geometrically, it is the union of the two coordinate axes.

Subspaces associated to a matrix.

Theorem 100 gives a certain subspace of $\mathbb{R}^m$ associated to an $n \times m$ matrix. We will define several of these.

Let A be an $n \times m$ matrix. There are several very important subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ which are naturally associated to A.

Definition 103. The column space of A, $\text{col}(A) = \text{Span}\{\text{columns of } A\} \subseteq \mathbb{R}^n$.

The row space of A, $\text{row}(A) = \text{Span}\{\text{(transposes of) rows of } A\} \subseteq \mathbb{R}^m$.

The null space of A, $\text{null}(A) = \{\text{solutions } \mathbf{x} \text{ to } A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{R}^m$.

As with any definitions, the first thing you should do when you see a definition is think about an example (then the definition will become more clear, and you will understand it better).

Example 104. Let $A = \begin{bmatrix}1 & -1 & 0\\2 & 4 & 3\end{bmatrix}$.

Then $\text{col}(A) = \text{Span}\left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-1\\4\end{bmatrix}, \begin{bmatrix}0\\3\end{bmatrix}\right\} \subseteq \mathbb{R}^2$ (in fact, you can verify for yourself that $\text{col}(A) = \mathbb{R}^2$).

And $\text{row}(A) = \text{Span}\left\{\begin{bmatrix}1\\-1\\0\end{bmatrix}, \begin{bmatrix}2\\4\\3\end{bmatrix}\right\} \subseteq \mathbb{R}^3$ (some plane in $\mathbb{R}^3$).

What is $\text{null}(A)$? The set of all $\mathbf{x} \in \mathbb{R}^3$ such that

$$\begin{bmatrix}1 & -1 & 0\\2 & 4 & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \mathbf{0}.$$

Hopefully you did not forget how to find solutions to an equation like this!

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

We have free variable $x_3 = t$, $x_2 = -1/2t$ and $x_1 = -1/2t$. Hence if $\mathbf{x}$ is a solution to this equation, we have that

$$\mathbf{x} = t \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix}.$$

So $\text{null}(A) = \text{Span}\left\{ \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix} \right\}$. This is some line in $\mathbb{R}^3$.

Subspaces associated to a linear transformation.

Since linear transformations are closely related to matrices, we expect the general story to be the same.

Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, given by, say, $T(\mathbf{x}) = A\mathbf{x}$ for some $n \times m$ matrix A.

Then $\text{range}(T)$ is a subspace of $\mathbb{R}^n$. In fact, $\text{range}(T) = \text{col}(A)$.

Definition 105. The kernel of T is the set

$$\ker(T) = \{\mathbf{x} \mid T(\mathbf{x}) = \mathbf{0}\} \subseteq \mathbb{R}^m.$$

This is a subspace. In fact, $\ker(T) = \text{null}(A)$.

We can draw a rough picture of the $\ker(T)$ and $\text{range}(T)$, to visualize what is happening (of course, this picture is not at all geometric).

Example 106. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$, "projection onto the xy-plane."

We can draw a picture of $\text{range}(T)$, which is the plane $z = 0$, and $\ker(T)$, which is the z-axis.

17. Friday 11/1: Basis and Dimension (4.2)

We now turn to the concept of the basis of a subspace. A subspace will usually contain infinitely many vectors (it will as long as it is not just $\{\mathbf{0}\}$), so working with subspaces can be tricky.

Example 107. Say $S = \text{Span}\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 3 \\ -1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix} -2 \\ 2 \\ 4 \\ 2 \end{bmatrix} \right\} \subseteq \mathbb{R}^4$.

S is some subspace of $\mathbb{R}^4$. So S could be a line, or a 2-dimensional plane, or a 3-dimensional plane, or all of $\mathbb{R}^4$ (or just $\mathbf{0}$, but it is clear that S contains more than just $\mathbf{0}$).

How can we tell what S is? The idea is to find a *minimal* spanning set for S. The set above is not minimal because $\mathbf{v}_4 = 2\mathbf{v}_2$, so $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. So in fact, $S \neq \mathbb{R}^4$.

Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ minimal? This is the same thing as asking: is it a linearly independent set? We know how to figure out if it is: make them the columns of a matrix, row reduce, and look at the pivot. We see:

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

so the set was not linearly independent. In particular, $\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2$. Hence, $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and by looking at the first two columns of our row reduction, $\mathbf{v}_1$ and $\mathbf{v}_2$ *are* linearly independent.

Hence, $S = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ is a 2-dimensional plane in $\mathbb{R}^4$.

The above example shows that in trying to understand a subspace, we really would like to have a set of vectors that is linearly independent and spans the subspace. This leads to the following definition.

Definition 108. Let S be a subspace of $\mathbb{R}^n$. A basis for S is a set of vectors $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ such that

(1) $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = S$ and

(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly independent.

Equivalently, a basis is a "*minimal* spanning set" or a "*maximal* linearly independent set."

Remark. A subspace can have many different bases[9]. In our previous example. we could have also taken $\{\mathbf{v}_1, \mathbf{v}_3\}$ as a basis for S. Or, indeed, $\{\mathbf{v}_2, \mathbf{v}_3\}$. Or any nonzero scalar multiples $\{c_1\mathbf{v}_1, c_2\mathbf{v}_2\}$. There are infinitely many possible bases of S. However, all of these bases share a feature: they all contain exactly two vectors. In fact, this is a theorem.

Theorem 109. Let S be a subspace. Then any basis of S contains the same number of vectors.

Since any basis of a subspace contains the same number of vectors this is a reasonable notion of the "size" of a subspace. This theorem leads to the following definition.

Definition 110. The dimension of a subspace S, denoted $\dim(S)$, is the number of vectors in any basis for S.

Remark. This is nice because even though subspaces generally have infinitely many vectors, dimension gives us a way to understand that a plane is "bigger" than a line. It has higher dimension.

Remark. If $S = \{\mathbf{0}\}$, then by convention, a basis for S is given by the empty set so $\dim(S) = 0$.

Remark. We finally understand why the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ were called the standard basis vectors. They formed a (particularly nice) basis of $\mathbb{R}^n$. And $\mathbb{R}^n$ has dimension n since there are n vectors in a basis.

So, finding a basis for S tells us the size (dimension) of S. Also, a basis gives us a precise way to write elements of S.

Theorem 111. Let S be a subspace. If $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a basis of S, and $\mathbf{u} \in S$ is any vector in S, then there is a *unique* $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$ such that

$$\mathbf{u} = x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m.$$

That is, every vector in S can be written *uniquely* as a linear combination of the basis vectors.

[9]Note that I get another opportunity to talk about irregular pluralizations. The plural of *basis* is *bases*, pronounced "bay-seas". How delightfully nautical.

Proof. Let's give a quick argument here. Since $\mathcal{B}$ spans S, therefore you can write $\mathbf{u}$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$. Why is it unique? Suppose you have two ways

$$\mathbf{u} = x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m$$

$$\mathbf{u} = y_1\mathbf{v}_1 + \cdots + y_m\mathbf{v}_m.$$

Then you can subtract to get

$$\mathbf{0} = (x_1 - y_1)\mathbf{v}_1 + \cdots + (x_m - y_m)\mathbf{v}_m$$

and since $\mathcal{B}$ is linearly independent, this means that $x_1 - y_1 = 0$, $x_2 - y_2 = 0$, etc. Hence, the two linear combinations were actually the same. □

Now that we have new vocabulary for a linearly independent spanning set, we can also add to our unifying theorem. It's useful to think through all the parts again periodically, as you learn something new every time you try to understand a theorem.

Theorem 112 (Unifying Theorem, version 4)**.** Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$. Let A be the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. The following are equivalent

(1) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans $\mathbb{R}^n$.

(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent.

(3) The equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for *every* $\mathbf{b} \in \mathbb{R}^n$.

(4) T is onto.

(5) T is one-to-one.

(6) A is invertible.

(7) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$.

Okay great! So together we came up with the definition of *basis* and *dimension*, and we have some conceptual and geometric understanding of what they mean. What we *haven't* discussed yet is any computational tools for working with these concepts. Let's discuss one computational tool now.

Common situation: Suppose $S = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ where the set is not necessarily linearly independent. How do we find a basis for S?

Method 1: Find a subset of $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ that is linearly independent (but still spans).

(1) Form the matrix $A = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_m \end{bmatrix}$.

(2) Reduce to echelon form $A \sim B$

Then keep the vectors $\mathbf{v}_i$ which *correspond* to pivot columns of B. These vectors will form a basis for S. Note that you need to take a subset of the *original* vectors. Not the pivot columns of B!

Why this works: Any linear dependence among the columns of A is *also* a dependence among the columns of B. The echelon form just makes the dependences easy to see.

Method 2[10]**:**

(1) Form A^T i.e., $A^T = \begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_m^T & - \end{bmatrix}$.

(2) Reduce A^T to echelon form, $A^T \sim C = \begin{bmatrix} - & \mathbf{u}_1^T & - \\ & \vdots & \\ - & \mathbf{u}_m^T & - \end{bmatrix}$. Then the nonzero rows (pivot rows) of C form a basis for S.

Why this works: Doing row operations on A^T means adding/subtracting the $\mathbf{v}_i$'s with each other, i.e., trying to eliminate redundant ones and simplify the others. For example, if we have that $\mathbf{v}_3 = 2\mathbf{v}_1 + \mathbf{v}_2$, then row operations can get us to

$$\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} \underset{\sim}{R_3 - 2R_1 - R_2 \to R_3} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ 0 \cdots 0 \end{bmatrix}.$$

When you have finished row reducing, the leftover rows are linearly independent by the definition of echelon form. For example, we might have a situation like this:

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

and there is no way to write a nontrivial linear combination of these (row) vectors to get $\mathbf{0}$, because in order to get 0 in the first coordinate, you need to take 0 of the first row.

Sometimes you have some vectors in a subspace S and would like to modify that set to get a basis for S. The following theorem gives two cases when this is possible.

Theorem 113. Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set of vectors in a subspace S of $\mathbb{R}^n$.

(1) If $\mathcal{U}$ is linearly independent, then either $\mathcal{U}$ is already a basis for S or you can add vectors to $\mathcal{U}$ to form a basis for S.

[10] Probably Method 1 is what you will actually use in practice, but it is a bit easier to see why Method 2 works.

(2) If $\mathcal{U}$ spans S, then either $\mathcal{U}$ is already a basis for S or you can remove vectors to $\mathcal{U}$ to form a basis for S.

18. Monday 11/4: More on Basis and Dimension (4.2) and Rank-Nullity (4.3)

At the start of class, we spent ten minutes working on the following problem in groups.

Example 114. Let $A = \begin{bmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}$.

- Compute null(A).
- Find a basis for null(A).
- Find a basis for null(A) that contains the vector $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -3 \\ -2 \\ -6 \\ 2 \end{bmatrix}$.

So we first find vectors that span null(A) which is just the set of solutions to $A\mathbf{x} = \mathbf{0}$. Luckily for you, someone already gave it to you in row echelon form, so we can find that

$$\text{null}(A) = \text{Span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ -1 \\ -3 \\ 1 \end{bmatrix} \right\}.$$

To find a basis including $\mathbf{v}_1$, we can throw $\mathbf{v}_1$ into the set as the first vector, place them as the columns of a matrix, and pick out two linearly independent ones.

$$\begin{bmatrix} 2 & -2 & -2 \\ -3 & 1 & 0 \\ -2 & 0 & -1 \\ -6 & 0 & -3 \\ 2 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & -1 \\ 0 & 2 & -3 \\ 0 & -2 & 3 \\ 0 & -6 & -9 \\ 0 & 2 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & -1 \\ 0 & 2 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so the first two columns are linearly independent. Hence, a basis for the null(A) is given by $\left\{ \begin{bmatrix} 2 \\ -3 \\ -2 \\ -6 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}$.

We also have some generalizations of theorems for $\mathbb{R}^n$ to subspaces. The first is an analogue of the first part of the unifying theorem.

Theorem 115. Suppose $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is a set of m vectors in a subspace S of dimension m.

(1) If $\mathcal{U}$ is linearly independent, then $\mathcal{U}$ is a basis for S.

(2) If $\mathcal{U}$ spans S, then $\mathcal{U}$ is a basis for S.

Proof. Suppose $\mathcal{U}$ is linearly independent. Then by Theorem 113, either $\mathcal{U}$ is already a basis or we can extend $\mathcal{U}$ to a basis. But if we add vectors to get a basis, we will have a basis of S with more than m vectors, and any basis of S has m vectors. Hence, $\mathcal{U}$ is already a basis.

Similarly, if $\mathcal{U}$ spans, then either it is already a basis or we can remove some vectors to make it a basis. But since a basis of S must have m elements, this means $\mathcal{U}$ is already a basis. □

Hence, just like in the unifying theorem, if you have exactly m vectors in a space of dimension m, then being linearly independent and spanning are equivalent. You only need to check one. This theorem in fact generalizes the first part unifying theorem since you recover the unifying theorem if you let $S = \mathbb{R}^m$.

We also have the following theorem which generalizes a theorem about $\mathbb{R}^n$ to any subspace.

Theorem 116. Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set of vectors in a subspace S of dimension k.

(1) If $m < k$ then $\mathcal{U}$ does not span S.

(2) If $m > k$, then $\mathcal{U}$ is not linearly independent.

Finally, another handy fact about subspaces, that you guessed when prompted.

Theorem 117. If S_1 and S_2 are two subspaces of $\mathbb{R}^n$ and $S_1 \subseteq S_2$ then $\dim(S_1) \leq \dim(S_2)$. Further, $\dim(S_1) = \dim(S_2)$ if and only if $S_1 = S_2$.

Row and Column Spaces (aka The Rank-Nullity Theorem).

Recall these definitions from 4.1. Let A be an $n \times m$ matrix and let $T : \mathbb{R}^m \to \mathbb{R}^n$ be the linear transformation $T(\mathbf{x}) = A\mathbf{x}$.

Definition 118. The column space of A, $\text{col}(A) = \text{Span}\{\text{columns of } A\} \subseteq \mathbb{R}^n$. This is the same as the range of T, $\text{range}(T)$.

The null space of A, $\text{null}(A) = \{\text{solutions } \mathbf{x} \text{ to } A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{R}^m$. This is the same as the kernel of T, $\ker(T)$.

Definition 119. The rank of a matrix A is the dimension of the column space of A:

$$\text{rank}(A) = \dim \text{col}(A) = \dim \text{range}(T).$$

Definition 120. The nullity of A is the dimension of the null space of A:

$$\text{nullity}(A) = \dim \text{null}(A) = \dim \ker(T).$$

Let's make some observations. If $\text{rank}(A)$ is large, then the dimension of the column space (and range) is large, so T is closer to being onto. In particular, T is onto if and only if $\text{rank}(A) = n$.

If $\text{nullity}(A)$ is large, then this means $\ker(T)$ is large, so T is "closer" to being the zero function. The zero function is very far from being one-to-one (since it sends everything to 0). So the *smaller* $\ker(T)$ is, the closer T is to being one-to-one. In particular, T is one-to-one if and only if $\text{nullity}(A) = 0$.

We talked about the Netflix problem https://en.wikipedia.org/wiki/Netflix_Prize. In 2006, Netflix offered a prize of \$1,000,000 to any team that could provide an algorithm to predict user ratings that improved on their own algorithm by 10%. We now have the vocabulary and concepts to talk a little bit about the idea behind this problem.

As of today, Netflix has roughly 150 million users and roughly 2,000 films (and many more episodes of TV shows). Netflix would like to be able to suggest movies to you that you will enjoy. They have data about user ratings of movies (which used to be 1 to 5 stars, but is now just thumbs up or thumbs down.)

We can imagine this data as occupying a humongous matrix with 150 million columns and 2,000 rows. Each user is a column, and each film is a row. Netflix would like to predict how every user will rate every film—in other words, they would like to know this entire matrix. But what they actually have is very few ratings (most of us only rate a handful of the movies that we've seen). Hence, the problem is to *complete* this matrix to a matrix that approximates the "true" matrix of user ratings.

If posed in this way, this question does not have a solution. In principle, any user can rate any movie with any rating. However, the idea that makes this problem tractable is that user ratings are *not* just random vectors in $\mathbb{R}^{2000}$. For example, we expect that if two users have similar tastes, then their columns should be pretty similar, In other words, we expect there to be *many relationships between the columns*. By this, we mean that the column space should have low dimension. User ratings

don't live randomly in $\mathbb{R}^{2000}$, rather, they should more or less lie in a relatively low-dimensional subspace!

So a good way to predict user ratings is to try to complete the matrix to a matrix with as *low rank* as possible (the smallest-dimensional column space). This is a hard problem: see https://en.wikipedia.org/wiki/Matrix_completion, with many computational and implementational challenges.

How did Netflix judge the algorithms? So they have some relatively small number of user ratings for films. What they can do is publicly give an even smaller subset of these ratings. Then, they can ask you to predict the ratings that they didn't give you (but they know). They feed this smaller data set to train their own algorithm, you feed the smaller data set to train your algorithm, and in the end you can compare performance by seeing what the two algorithms predict for the unpublished known values.

In 2009, a team successfully improved on Netflix's algorithm by 10% and claimed the million dollar prize.

19. Wednesday 11/6: The Rank-Nullity Theorem (4.3)

At the start of class, we spent ten minutes working on the following problems in groups.

Example 121. Let A be the matrix

$$A = \begin{bmatrix} 3 & -6 & 9 & 0 \\ 2 & -4 & 7 & 2 \\ 3 & -6 & 6 & -6 \end{bmatrix}.$$

So A gives a linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$. Find $\text{rank}(A)$, a basis for $\text{col}(A)$, $\text{null}(A)$, and a basis for $\text{null}(A)$. Is T one-to-one? Is T onto?

The first thing we ought to do is reduce A to reduced echelon form

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

So $\text{rank}(A) = 2$ and a basis for $\text{col}(A)$ is given by the first and third columns of A: $\left\{ \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 9 \\ 7 \\ 6 \end{bmatrix} \right\}$.

To find $\text{null}(A)$, we must find the set of vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ that satisfy $A\mathbf{x} = \mathbf{0}$. That is, we seek all solutions to the homogeneous system.

So augment with the zero vector and reduce to obtain:

$$\left[\begin{array}{cccc|c} 1 & -2 & 0 & -6 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Thus, $x_1 - 2x_2 - 6x_4 = 0$, $x_3 + 2x_4 = 0$. So we can write each of the basic variables in terms of the free variables and get

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2x_2 + 6x_4 \\ x_2 \\ -2x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ x_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 6x_4 \\ 0 \\ -2x_4 \\ x_4 \end{bmatrix}$$

$$= x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 6 \\ 0 \\ -2 \\ 1 \end{bmatrix}.$$

This shows that the null space has the basis

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 0 \\ -2 \\ 1 \end{bmatrix} \right\}.$$

(Well, we know from the calculation above that these two vectors span the null space.)

They will also be linearly independent. The reason for this is that the first vector corresponds to x_2 and the second vector corresponds to x_4. This means that the 2nd entry in the first vector is 1 and the fourth entry in the first vector is 0. Also, the 4th entry in the first vector is 0 and the fourth entry in the second vector is 1. By looking at those entries, we can see that they are linearly independent. Hence, $\text{nullity}(A) = 2$.

What did we learn from this example?

Given some matrix A, to find $\text{rank}(A)$, reduce to echelon form. Then $\text{rank}(A)$ is the number of pivot columns. On the other hand, $\text{nullity}(A)$ is equal to the number of free variables. So together, the rank and the nullity of A account for all of the columns of A!

Theorem 122 (The Rank-Nullity Theorem)**.** If A is an $n \times m$ matrix, then

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

Okay, so we have some computational idea of why the Rank-Nullity Theorem is true. But why should we expect it to be true geometrically?

Idea: "Conservation of dimension"

In the example, we had a linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ with rank 2 and nullity 2. So the domain of T is $\mathbb{R}^4$, but the dimension range(T) is only 2. So T "flattens" $\mathbb{R}^4$ into something 2-dimensional. That is, T "lost" 2 dimensions. This corresponds to the fact that $\ker(T)$ is 2-dimensional. T mapped "2 dimensions worth of vectors" in $\mathbb{R}^4$ to $\mathbf{0}$.

Basically, the more vectors you send to $\mathbf{0}$, the smaller your range gets, while the fewer vectors you send to $\mathbf{0}$, the larger your range gets. Altogether, the dimensions should sum to the dimension of the domain.

Before we get to some examples, let's add to our unifying theorem, since we can now talk more precisely about dimensions of subspaces assocated to matrices:

Theorem 123 (Unifying theorem, version who knows?)**.** Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$. Let A be the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. The following are equivalent

(1) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans $\mathbb{R}^n$.

(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent.

(3) The equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for *every* $\mathbf{b} \in \mathbb{R}^n$.

(4) T is onto.

(5) T is one-to-one.

(6) A is invertible.

(7) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$.

(8) $\text{col}(A) = \text{range}(T) = \mathbb{R}^n$.

(9) $\text{row}(A) = \mathbb{R}^n$.

(10) $\text{rank}(A) = n$.

(11) $\text{null}(A) = \ker(T) = \{\mathbf{0}\}$.

(12) $\text{nullity}(A) = 0$.

Even though the Rank-Nullity Theorem seems to be a trivial computational fact, it actually can give you a lot of information, because it relates the range of a linear transformation to its kernel. Let's see how, in a few examples.

Example 124. Let $T : \mathbb{R}^7 \to \mathbb{R}^{10}$ be a linear transformation such that $\ker(T)$ is 4-dimensional.

What is $\dim(\text{range}(T))$? Rank-Nullity says that

$$\dim(\text{range}(T)) + \dim(\ker(T)) = 7$$

where 7 is the dimension of the domain. Hence, $\dim(\text{range}(T)) = 3$.

Example 125. Suppose $T : \mathbb{R}^5 \to \mathbb{R}^3$ is an onto linear transformation. What is $\dim(\ker(T))$? Rank-Nullity says that

$$\dim(\text{range}(T)) + \dim(\ker(T)) = 5.$$

Since T is onto, $\dim(\text{range}(T)) = 3$. Hence, $\dim(\ker(T)) = 2$.

Example 126. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation given by multiplication by the matrix A. You notice that $T\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \mathbf{0}$ and that $\begin{bmatrix}2\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\3\end{bmatrix} \in \text{range}(T)$.

Determine $\text{rank}(A)$ and $\text{nullity}(A)$.

Since $T\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \mathbf{0}$, you know that $\ker(T)$ is *at least* 1-dimensional. Hence, $\text{nullity}(A) \geq 1$.

Similarly, since $\begin{bmatrix}2\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\3\end{bmatrix}$ are linearly independent, you know that $\text{rank}(T)$ is *at least* 2-dimensional and so $\text{rank}(A) \geq 2$.

By Rank-Nullity,

$$\text{rank}(A) + \text{nullity}(A) = 3$$

and so $\text{rank}(A) = 2$ and $\text{nullity}(A) = 1$.

In fact, this also implies that $\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$ is a basis for $\ker(T) = \text{null}(A)$.

Fun Fact. The Rank-Nullity Theorem is really a theorem that says that the alternating sums dimensions of certain subspaces is equal to 0. This is related to a very interesting topological invariant

called the *Euler characteristic*: https://en.wikipedia.org/wiki/Euler_characteristic. We discussed this at the end of class[11].

[11] And this counts as a practical application, because if you ever find yourself stranded on an alien planet, you have a tool to determine what shape the planet is.

20. Friday 11/8: Determinants: Definition and first properties (5.1)

We now turn to the study of the determinant. The determinant will be a number that we associate to a square matrix (or, equivalently, a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$).

We already know some determinants. The determinant of a 1×1 matrix $[a]$ is just given by a.

The determinant of a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $ad - bc$. This is the expression in the denominator when computing the inverse of a 2×2 matrix. Hence, we saw that the matrix is invertible if and only if this value is nonzero. The same will be true for determinants of $n \times n$ matrices.

For an $n \times n$ matrix A, we denote by A_{ij} the $(n-1) \times (n-1)$ submatrix obtained by deleting the ith row and jth column of A.

Example 127. Let $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. Then $A_{23} = \begin{bmatrix} 2 & -4 \\ 1 & 4 \end{bmatrix}$ and $A_{33} = \begin{bmatrix} 2 & -4 \\ 3 & 1 \end{bmatrix}$.

The definition of determinant is *recursive*. That means, in order to compute the determinant of an $n \times n$ matrix we first need to know how to compute the determinant of an $(n-1) \times (n-1)$ matrix. This is ok because we already know how to compute the determinant of a 2×2 matrix. Hence, we will know how to compute the determinant of a 3×3 and therefore a 4×4 and so on.

Definition 128. For $n \geq 2$, the determinant of an $n \times n$ matrix $A = (a_{ij})$ is the sum of n terms of the form $\pm a_{1j} \det(A_{ij})$ with alternating $\pm$ signs:

$$|A| = \det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + \cdots + (-1)^{n+1} a_{1n} \det(A_{1n}) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1j}).$$

Example 129. Compute the determinant of the matrix in Example 127.

The rule above says that

$$\begin{aligned} \det(A) &= 2 \cdot \begin{vmatrix} 1 & 2 \\ 4 & -1 \end{vmatrix} + 4 \cdot \begin{vmatrix} 3 & 2 \\ 1 & -1 \end{vmatrix} + 3 \cdot \begin{vmatrix} 3 & 1 \\ 1 & 4 \end{vmatrix} \\ &= 2(-1-8) + 4(-3-2) + 3(12-1) \\ &= -18 - 20 + 33 = -5. \end{aligned}$$

The definition of the determinant we gave is one of many (equivalent) definitions. In particular, the choice of using the first row is completely arbitrary.

The (i,j)-cofactor is $C_{ij} = (-1)^{i+j}\det(A_{ij})$. Our earlier formula is then

$$\det(A) = \sum_{j=1}^{n} a_{1j}C_{1j}.$$

Theorem 130 (Cofactor Expansion)**.** The determinant of any $n \times n$ matrix can be determined by cofactor expansion along any row or column. In particular,

$$(i\text{th row}) \qquad \det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \sum_{j=1}^{n} a_{ij}C_{ij}$$

$$(j\text{th col}) \qquad \det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} = \sum_{i=1}^{n} a_{ij}C_{ij}.$$

Example 131. Compute the determinant of the matrix A in Example 127 using cofactor expansion along the first column.

Note: The pattern of signs $(-1)^{i+k}$ or $(-1)^{j+k}$ is

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Example 132. Let

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

What is $\det(M)$?

Observe that the fourth column only contains one nonzero entry. So if we do cofactor expansion down the fourth column, we get

$$\begin{aligned}\det(M) &= -2\begin{vmatrix} 2 & 3 & 0 \\ -5 & 1 & 2 \\ 3 & 2 & 1 \end{vmatrix} \\ &= (-2)\left[2\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} - 3\begin{vmatrix} -5 & 2 \\ 3 & 1 \end{vmatrix}\right] \\ &= (-2)\left[2\cdot(-3) - 3(-5-6)\right] \\ &= (-2)\left[-6 - 3(-11)\right] = (-2)(33-6) = -54.\end{aligned}$$

First Properties of the Determinant.

We now develop some properties of the determinant.

Our entire motivation in deriving the 2×2 and 3×3 determinant was to have the determinant detect the invertibility of matrices. Indeed, the following theorem establishes this fact.

Theorem 133. Let A be an $n \times n$ matrix. Then A is invertible if and only if $\det(A) \neq 0$.

Which also means we get to add to our unifying theorem!

Theorem 134 (Unifying theorem, version who knows? +1)**.** Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$. Let A be the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. The following are equivalent

(1) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans $\mathbb{R}^n$.

(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent.

(3) The equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for *every* $\mathbf{b} \in \mathbb{R}^n$.

(4) T is onto.

(5) T is one-to-one.

(6) A is invertible.

(7) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$.

(8) $\text{col}(A) = \text{range}(T) = \mathbb{R}^n$.

(9) $\text{row}(A) = \mathbb{R}^n$.

(10) $\text{rank}(A) = n$.

(11) $\text{null}(A) = \ker(T) = \{\mathbf{0}\}$.

(12) $\text{nullity}(A) = 0$.

(13) $\det(A) \neq 0$.

Theorem 135. For $n \geq 1$, we have that $\det(I_n) = 1$.

Proof. The proof is a proof by induction. We won't explain mathematical induction super rigorously. But the idea is the following.

Clearly if $n = 1$, the 1×1 identity matrix is just $[1]$ and has determinant 1. For the 2×2 case, do a cofactor expansion along the first row so that

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

For 3×3, again do a cofactor expansion along the first row:

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

Continue this process to see that it works for the $n \times n$ identity matrix. □

Theorem 136. If A is triangular, then $\det(A)$ is the product of the entries on the main diagonal.

Theorem 137. If A is a square matrix, then $\det(A) = \det(A^T)$.

Both theorems can be proved using induction, but as this is not a focus of this class, we won't prove them. If you're interested, learning proof by induction is a very fun part of Math 300!

Theorem 138. Let A be a square matrix.

(1) If A has a row or column of zeros, then $\det(A) = 0$.

(2) If A has two identical rows or columns, then $\det(A) = 0$.

This theorem follows from the unifying theorem, since in either case, either the rows or the columns cannot possibly be linearly independent or span $\mathbb{R}^n$, so the matrix cannot be invertible.

Also, determinants behave well with respect to matrix multiplication[12].

Theorem 139. Let A and B be $n \times n$ matrices. Then

$$\det(AB) = \det(A)\det(B).$$

We will see a reason why we should expect this when we talk about the geometry of the determinant in two sections.

Corollary 140. If A is invertible then

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Proof. This is a corollary: a result that follows from another result. Assuming the above theorem is true, then since $AA^{-1} = I_n$ we have

$$\det(A)\det(A^{-1}) = \det(I_n) = 1$$

and so $\det(A^{-1}) = 1/\det(A)$. □

[12]Just based on the definition of the determinant, it is non-obvious that the determinant should be multiplicative. Consider what you would have to do to prove it: take A and B and write the entries of AB in terms of the entries of A and B (and the entries in a product have some not-so-beautiful formula). Then do cofactor expansions to compute $\det A$ and $\det B$, multiply them, and show that it's the same as doing a cofactor expansion on AB to compute $\det(AB)$. The fact that this nice result holds is a *eucatastrophe*. Completely unexpectedly, something good happens.

3Blue1Brown has a nice video on determinants here https://youtu.be/Ip3X9LOh2dk.

21. Monday 11/11: Veteran's Day

No class today!

22. Wednesday 11/13: Properties and Geometry of the Determinant (5.2, 5.3)

For a large matrix (say 4×4 or larger), the "shortcut method" provided in your book does not work, and using cofactor expansion involves many computations. In this section, we will develop some more properties of the determinant in order to give a computationally fast method to compute determinants.

The idea is simple. If we row reduce our matrix to a triangular one, then we can compute the determinant by just multiplying the diagonal entries. We just need to know how row reductions affect the determinant.

Let's think about this through an example.

Example 141. Let $A = \begin{bmatrix} 2 & 4 & 3 \\ 3 & 1 & 2 \\ 1 & 0 & -1 \end{bmatrix}$. Compare $\det(A)$ with $\det(B)$ where B is the matrix we get from A after performing the given row operation.

(1) Interchange rows 2 and 3.

(2) Multiply row 1 by 2.

(3) Add 2 times row 1 to row 2.

First, we should compute the determinant of A, perhaps by cofactor expansion along the third row. This gives us

$$\det(A) = 1 \det \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} - 1 \det \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} = 5 - (-10) = 15.$$

(1) Interchange rows 2 and 3.

Then $B = \begin{bmatrix} 2 & 4 & 3 \\ 1 & 0 & -1 \\ 3 & 1 & 2 \end{bmatrix}$. Now if we do cofactor expansion along the second row, you see that we will get the exact same thing as before, except our signs will be switched. So

$$\det(B) = -1 \det \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} + 1 \det \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} = -5 + (-10) = -15.$$

(2) Multiply row 1 by 2.

Then $B = \begin{bmatrix} 4 & 8 & 6 \\ 3 & 1 & 2 \\ 1 & 0 & -1 \end{bmatrix}$. Cofactor expansion along the third row gives

$$\det(B) = 1 \det \begin{bmatrix} 8 & 6 \\ 1 & 2 \end{bmatrix} - 1 \det \begin{bmatrix} 4 & 8 \\ 3 & 1 \end{bmatrix} = 10 - (-20) = 30.$$

(3) Add 2 times row 1 to row 2.

Then $B = \begin{bmatrix} 2 & 4 & 3 \\ 7 & 9 & 8 \\ 1 & 0 & -1 \end{bmatrix}$. Cofactor expansion along third row gives

$$\det(A) = 1 \det \begin{bmatrix} 4 & 3 \\ 9 & 8 \end{bmatrix} - 1 \det \begin{bmatrix} 2 & 4 \\ 7 & 9 \end{bmatrix} = 5 - (-10) = 15.$$

In fact, what we saw in this example is true in general

Theorem 142. Let A be a square matrix.

(1) If two rows of A are interchanged to produce B, then $\det A = -\det B$.

(2) If one row of A is multiplied by c to produce B, then $\det A = \frac{1}{c} \cdot \det B$.

(3) If a multiple of one row of A is added to another row to produce a matrix B, then $\det A = \det B$.

So the strategy for a large matrix is: use row reductions to get to a triangular one. Keep track of how your row reductions affect the determinant.

Example 143. Use row reduction to compute the determinant of $A = \begin{bmatrix} 2 & 4 & 3 \\ 3 & 1 & 2 \\ 1 & 0 & -1 \end{bmatrix}$ (of course we already know the determinant and cofactor expansion isn't too bad here, but for larger matrices this method is much faster computationally).

We have

$$A = \begin{bmatrix} 2 & 4 & 3 \\ 3 & 1 & 2 \\ 1 & 0 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 3 & 1 & 2 \\ 2 & 4 & 3 \end{bmatrix} = A_1$$

and $\det(A) = -\det(A_1)$. Now adding multiples of the first row to the second and third gives

$$A_1 = \begin{bmatrix} 1 & 0 & -1 \\ 3 & 1 & 2 \\ 2 & 4 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 5 \\ 0 & 4 & 5 \end{bmatrix} = A_2$$

and $\det(A) = -\det(A_1) = -\det(A_2)$. Again subtracting a multiple of row 2 from row 3 gives

$$A_2 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 5 \\ 0 & 4 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 5 \\ 0 & 0 & -15 \end{bmatrix} = A_3$$

And $\det(A) = -\det(A_2) = -\det(A_3) = -(-15) = 15$.

Remark. For a large matrix, the row reduction method of computing a determinant is much faster than cofactor expansion. For an $n \times n$ matrix, since cofactor expansion is recursive, the runtime is bounded below by $O(n!)$. On the other hand, Gaussian elimination is $O(n^3)$. This is *much* faster.

Geometry of the determinant.

We know the definition of the determinant, some basic properties, and an efficient way to compute them. One natural question is: what is the *meaning* of the determinant?

Geometrically, if $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ then $\det(A)$ is the volume (up to a sign) of the "parallelogram in $\mathbb{R}^n$" determined by $\mathbf{v}_1, \dots, \mathbf{v}_n$.

In $\mathbb{R}^2$, it is easy to picture the parallelogram P determined by $\mathbf{v}_1$ and $\mathbf{v}_2$ and then $\det \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \pm\text{area}(P)$.

In $\mathbb{R}^3$, it is a bit harder to draw, but you can form the parallelepiped determined by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_2$ and $\det \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \pm\text{volume}(P)$.

You can also imagine a "parallelogram" in n-dimensions, but I don't know about drawing it.

In terms of linear transformations, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T(\mathbf{x}) = A\mathbf{x}$ for an $n \times n$ matrix A, then T takes the "standard cube" with edges $\mathbf{e}_1, \dots, \mathbf{e}_n$ to the n-dimensional "parallelogram" with edges $\mathbf{v}_1, \dots, \mathbf{v}_n$ (the columns of A).

So $\det(A)$ is the *scaling factor* of T (times ± 1). In fact, all volumes get rescaled by $\det(A)$, not just the cubes.

(Why? Calculus! Cut your region into small "cubes" and take the limit. Each cube gets rescaled by a factor of $\det(A)$.)

Fact. Rescaling a column $\mathbf{v}_i$ of A by a factor of c results in rescaling $\det(A)$ be c.

We can visualize this in $\mathbb{R}^3$ by drawing the parallelepiped P determined by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ and scaling one of them by a factor of, say, 2. The resulting parallelepiped is 2 stacked copies of P so has twice the volume.

Fact. $\det(AB) = \det(A)\det(B)$.

If A and B are $n \times n$ matrices, If we think of A and B as giving the linear transformations S and T, then AB corresponds to $S \circ T$. When you start with a volume in $\mathbb{R}^n$, T scales the volume by a factor of $\det(B)$. Then applying S scales the volume by a factor of $\det(A)$. Altogether, performing $S \circ T$ scales the volume by a vector of $\det(A)\det(B)$. Hence, $\det(AB) = \det(A)\det(B)$.

Chapter 6: Eigenvectors and Eigenvalues.

Our next chapter is on eigenvectors and eigenvalues. Eigenvectors and eigenvalues are important tools in both pure and applied mathematics. They might actually be the most widely-applied concept from linear algebra: e.g., Google PageRank, principal component analysis, spectral graph theory, etc[13].

[13]This is a bit like explaining to a young student that learning Greek will be great because eventually she will be able to read the epics of Homer, the *Elements* of Euclid, and the dialogues of Plato... and then starting by sitting down and learning Greek grammar. There are many cool applications but first we have to learn the basics, and it won't seem that enlightening or exciting at first.

3Blue1Brown has a video on eigenvectors and eigenvalues: https://youtu.be/PFDu9oVAE-g.

23. Friday 11/15: Introduction to Eigenstuff (6.1)

Example 144. Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$. What do you notice about $A\mathbf{u}$? Interpret this geometrically.

Observe that $A\mathbf{u} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\mathbf{u}$. Geometrically, we can interpret this as A stretching the vector $\mathbf{u}$. This does not happen with $\mathbf{v}$.

This is a somewhat interesting phenomenon. Usually, given an $n \times n$ matrix A (or equivalently a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$), the vector $T(\mathbf{u})$ is completely different from $\mathbf{u}$. But we have just seen an example where the vector $T(\mathbf{u})$ was similar to the vector $\mathbf{u}$ in that it pointed in the same direction. We give vectors with this property a special name: eigenvectors[14].

Definition 145. An eigenvector of an $n \times n$ matrix A is a **nonzero** vector $\mathbf{u}$ such that $A\mathbf{u} = \lambda\mathbf{u}$ for some scalar λ. A scalar λ is called an eigenvalue of A if there is a nontrivial solution of $A\mathbf{u} = \lambda\mathbf{u}$. We say $\mathbf{u}$ is the eigenvector corresponding to λ.

In the previous example, $\mathbf{u}$ is an eigenvector of A corresponding to the eigenvalue -4.

We could ask of the matrix A: does A have any other eigenvectors with eigenvalue -4? It turns out that it does and that the set of all such vectors is very nice.

Theorem 146. Let A be a square matrix and suppose that $\mathbf{u}$ is an eigenvector of A with eigenvalue λ. Then for any scalar $c \neq 0$, $c\mathbf{u}$ is also an eigenvector of A with eigenvalue λ.

Proof. We simply check

$$A(c\mathbf{u}) = cA\mathbf{u} = c\lambda\mathbf{u} = \lambda(c\mathbf{u})$$

and so $c\mathbf{u}$ is in fact an eigenvector of A with eigenvalue λ. □

[14]The prefix "eigen" comes from German where it means something like "own", "private", "innate", or "natural". You should think of an eigenvector or eigenvalue as being strongly associated to the matrix or linear transformation, so the eigenvalues are "innate". The eigenvalues carry a lot of information about the matrix, so can be thought of as the matrix's "own".

Indeed, we actually have a much stronger theorem.

Theorem 147. Let A be a $n \times n$ matrix with eigenvalue λ. Let $E_\lambda(A)$ be the set of *all* eigenvectors of A with eigenvalue λ, together with the zero vector $\mathbf{0}$. Then $E_\lambda(A)$ is a subspace of $\mathbb{R}^n$.

Proof. We already checked that $E_\lambda(A)$ is closed under scalar multiplication. By definition, $\mathbf{0} \in E_\lambda(A)$. So we need only check that $E_\lambda(A)$ is closed under addition. Suppose $\mathbf{u}$ and $\mathbf{v}$ are two eigenvectors of A with eigenvalue λ. Then

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} + \lambda\mathbf{u} + \lambda\mathbf{v} = \lambda(\mathbf{u} + \mathbf{v})$$

so $\mathbf{u} + \mathbf{v}$ is an eigenvector of A with eigenvalue λ. Hence, $E_\lambda(A)$ is a subspace. □

Note. Note that it is very important in the above theorem that we are only considering a single eigenvalue of A at a time. If $\mathbf{u}$ and $\mathbf{v}$ were eigenvectors with *different* eigenvalues, the step in the proof where we factored out the eigenvalue would not work.

Another way to write $E_\lambda(A)$ is

$$E_\lambda(A) = \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \lambda\mathbf{v}\}$$

since this includes all of the eigenvectors as well as automatically including the zero vector.

Definition 148. Let A be a square matrix with eigenvalue λ. The subspace $E_\lambda(A)$ of $\mathbb{R}^n$ is called the eigenspace of λ.

Now onto the big computational question: How do you find eigenvalues and eigenvectors?

Finding Eigenvalues.

Given an $n \times n$ matrix A, if λ is an eigenvalue, then it must have at least one nonzero eigenvector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$ so $(A - \lambda I)\mathbf{v} = \mathbf{0}$. In other words, $A - \lambda I$ must have a nontrivial null space. This means that $A - \lambda I$ must not be invertible. Which means that $\det(A - \lambda I)$ must be zero. In fact, this train of logic basically proves this theorem:

Theorem 149. Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Example 150. Use the theorem to find the eigenvalues of $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$.

We need to find all λ so that $\det(A - \lambda I_2) = 0$. That is, we want

$$\det \begin{bmatrix} 1-\lambda & 6 \\ 5 & 2-\lambda \end{bmatrix} = (1-\lambda)(2-\lambda) - 30 = 0.$$

But this means that $\lambda^2 - 3\lambda - 28 = 0$ or $(\lambda - 7)(\lambda + 4) = 0$. The only values of λ that makes this true are -4 and 7. So A has eigenvalues -4 and 7.

As we saw in the previous example, finding eigenvalues comes down to solving for the roots of a polynomial (in the variable λ).

Definition 151. Let A be an $n \times n$ matrix. The polynomial

$$p_A(\lambda) = \det(A - \lambda I)$$

is a polynomial of degree n (in λ). This polynomial is called the characteristic polynomial of A. The eigenvalues of A are given by the roots of the characteristic polynomial.

Note. The eigenvalues of A **cannot** be read off the row echelon form of A, unlike most other information about A.

Example 152. Let $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$. Find the eigenvalues of A.

We compute the characteristic polynomial:

$$p_A(\lambda) = \det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = (2-\lambda)(-6-\lambda) - 9 = \lambda^2 + 4\lambda - 21 = (\lambda+7)(\lambda-3).$$

Since the roots of the characteristic polynomial are 3 and -7, these are the eigenvalues.

Finding Eigenvectors.

Okay so we know how to find eigenvalues. Assuming that you know some eigenvalue for a matrix, how do you find the eigenspace? Well, if you know that λ is an eigenvalue for A, then the

eigenvectors will be the vectors $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. This means that

$$A\mathbf{v} - \lambda\mathbf{v} = (A - \lambda I)\mathbf{v} = \mathbf{0}.$$

So the eigenspace of λ is the same thing as the null space of $A - \lambda I$!

The upshot: Given an $n \times n$ matrix A.

(1) First find the eigenvalues of A by solving $\det(A - \lambda I) = 0$.

(2) For each eigenvalue λ, find the eigenspace of λ by $E_\lambda(A) = \text{null}(A - \lambda I)$.

Example 153. We know that $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$ has eigenvalues 3 and -7. What are bases of the corresponding eigenspaces?

For the eigenvalue 3, the eigenspace is $\text{null}(A - 3I) = \text{null} \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix}$ which has basis $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

For the eigenvalue -7, the eigenspace is $\text{null}(A - 7I) = \text{null} \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}$ which has basis $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

3Blue1Brown has a video on eigenvectors and eigenvalues: https://youtu.be/PFDu9oVAE-g.

24. Monday 11/18: More on Eigenstuff

We began class by working on the following example:

Example 154. Let

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

What are the eigenvalues of A? What are bases for the eigenspaces?

We have $A - \lambda I = \begin{bmatrix} -\lambda & -5 & 0 \\ 0 & 5-\lambda & 0 \\ -3 & -5 & 3-\lambda \end{bmatrix}$. When we take the determinant, we get

$$p_A(\lambda) = \det(A - \lambda I) = (-\lambda)(5-\lambda)(3-\lambda).$$

The roots of the characteristic polynomial are 0, 3 and 5, so these are the eigenvalues. For $\lambda = 0$, when we row reduce $A - 0I = A$ we get

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

We see that a basis of the eigenspace is $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

For $\lambda = 3$, $A - 3I$ row reduces to

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

so a basis of the eigenspace is given by $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Finally, for $\lambda = 5$, $A - 5I$ row reduces to

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

and $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ is a basis of this eigenspace.

It is easy to check that $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ are actually eigenvectors.

Remark. Let's make some remarks about this example (which hopefully reminded you how to compute eigenvalues and eigenvectors).

Note that even 0 can definitely be an eigen*value* of a matrix. In our definition of eigen*vector* we specified that an eigenvector can never be $\mathbf{0}$.

Indeed, if 0 is an eigen*value*, then that means there is a nonzero vector $\mathbf{u}$ such that $A\mathbf{u} = 0\mathbf{u} = \mathbf{0}$. In other words, 0 is an eigenvalue of A if and only if A has a nontrivial null space! Which means that we can add something to our unifying theorem.

Theorem 155 (Unifying theorem (continued))**.** Assume the setup of the previous unifying theorem. Then the following are equivalent:

(1) $\det A \neq 0$

(2) $\lambda = 0$ is *not* an eigenvalue of A.

Proof. We know that 0 is an eigenvalue of A if and only if $\det(A - 0I) = 0$. But this is the same as $\det(A) = 0$. □

Remark. Another thing to remark is that so far, in all three of our examples (two from last week and one from today), A has been an $n \times n$ matrix which had n disticint eigenvalues. Every one of the eigenspaces we computed we was one-dimensional.

Of course, since the eigenvalues of a matrix are given by the roots of the characteristic polynomial, we might not have n distinct eigenvalues. This motivates the following definition.

Definition 156. The multiplicity of an eigenvalue is its multiplicity as a root of the characteristic polynomial. For example, if the characteristic polynomial is $(\lambda - 3)^2(\lambda + 2)^5(\lambda - 7)$ then 3 is an eigenvalue of multiplicity 2, -2 is an eigenvalue with multiplicity 5, and 7 is an eigenvalue with multiplicity 1.

With this language, in every example we have done so far, the multiplicity of the eigenvalue λ has been 1, and the dimension of the corresponding eigenspace $E_\lambda(A)$ was also equal to 1. We should consider what might happen if the multiplicity of an eigenvalue is bigger than one. Let's look at the simplest case:

Example 157. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Then A has a single eigenvalue, 1, with multiplicity 2. The corresponding eigenspace is all of $\mathbb{R}^2$ (since the identity matrix scales every vector in $\mathbb{R}^2$ by a factor of 1, i.e., it fixes every vector).

In this case, the multiplicity of the eigenvalue was 2, and we got a 2-dimensional eigenspace. We might therefore hope:

Hope. The dimension of $E_\lambda(A)$ is equal to the multiplicity of λ as an eigenvalue.

Unfortunately, this does not always happen. Consider the following example.

Example 158. Compute the eigenvalues and find bases for the eigenspaces for the matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$.

To find the eigenvalues, we take

$$\det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 2 \\ 0 & 1-\lambda \end{bmatrix} = (1-\lambda)^2 = 0.$$

So 1 is the only eigenvalue (with multiplicity 2). The corresponding eigenspace is given by

$$(A - I) = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$$

which is one-dimensional (spanned by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$).

So it is possible for the eigenspace to have dimension *smaller* than the multiplicity of the eigenvalue. We can therefore update our hope!

A New Hope[15]**.** The dimension of $E_\lambda(A)$ is always less than or equal to the multiplicity of λ.

[15] The second best *Star Wars* movie.

It turns out that this new hope is true!

Theorem 159. Let A be a square matrix with eigenvalue λ. Then

$$1 \leq \dim E_\lambda(A) \leq \text{the multiplicity of } \lambda.$$

The way that this theorem should be remembered is:

"Sometimes there aren't enough eigenvectors[16]."

We will see in the next section that the *nice* cases are the ones in which every eigenspace is as big as possible.

Diagonalization (6.2)

Example 160. Let $B = \begin{bmatrix} 7 & 0 \\ 0 & 1 \end{bmatrix}$. Find a formula for B^k. What if you let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$? Can you find a formula for A^k?

It is not hard to see that $B^k = \begin{bmatrix} 7^k & 0 \\ 0 & 1^k \end{bmatrix}$. It is unclear how to write down a formula for A^k.

Why might you care? If you have some vector $\mathbf{v}$ representing some real-life data, then $A\mathbf{v}$ might represent the output after one time-step. Then if you want to know the data after k steps, you want to compute $A^k\mathbf{v}$. For example, maybe we can represent the US economy by breaking it down into 100 sectors and assessing the value of each sector. We might then have some matrix that predicts how the US economy will change from one year to the next (there are lots of interactions between different sectors of the economy). This matrix would be 100×100. So if $\mathbf{v} \in \mathbb{R}^{100}$ represented the economy in 2019, we could make predictions about the economy in 2020 by computing $A\mathbf{v} \in \mathbb{R}^{100}$. If we wanted to make predictions about the economy in 2029, we would want to multiply by A nine more times. In other words, we would want $A^{10}\mathbf{v}$.

Definition 161. An $n \times n$ matrix is diagonalizable if there is an $n \times n$ invertible matrix P and an $n \times n$ diagonal matrix D so that $A = PDP^{-1}$

This definition says that A is closely related to a diagonal matrix. A diagonal matrix is already diagonalizable, because you can take $P = I$ the identity matrix.

[16] To paraphrase the line from *Forrest Gump*: "Sometimes, I guess there just aren't enough rocks."

And if A were diagonalizable, then

$$A^k = (PDP^{-1})^k = PDP^{-1}PDP^{-1}\dots PDP^{-1} = PD^kP^{-1}.$$

So we would be able to take powers of A!

25. Wednesday 11/20: Exam 2

On Wednesday you took your second exam.

26. Friday 11/22: More on Diagonalization (6.2)

Example 162. Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Verify that A is diagonalizable by checking that $A = PDP^{-1}$ with $P = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$. Use this fact to find a formula for A^k

The key thing to notice here is that

$$A^k = (PDP^{-1})^k = PDP^{-1}PDP^{-1}\dots PDP^{-1} = PD^kP^{-1}.$$

Hence,

$$\begin{aligned} A^k &= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \cdot 5^k & 5^k \\ 3^k & 3^k \end{bmatrix} \\ &= \begin{bmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ -2 \cdot 5^k + 2 \cdot 3^k & -5^k + 2 \cdot 3^k \end{bmatrix} \end{aligned}$$

Some questions are just begging to be asked. Is it always possible to diagonalize A? If A is diagonalizable, how do you find P and D? If you can diagonalize A, are P and D unique?

Observation. Well, suppose first that A has n linearly independent eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Set $P = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ and $D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$.

Since the $\mathbf{u}_i$'s are eigenvectors for A, we have that $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$ and so

$$AP = A \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{u}_1 & \cdots & A\mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{u}_1 & \cdots & \lambda_n\mathbf{u}_n \end{bmatrix}.$$

But also

$$PD = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{u}_1 & \cdots & \lambda_n\mathbf{u}_n \end{bmatrix}.$$

Since the columns of P are linearly independent, P is invertible. Therefore

$$AP = PD \Rightarrow A = PDP^{-1}.$$

Example 163. Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$ and use this to diagonalize A (this is how I came up with Example 162).

So now we know that *if* a matrix has n linearly independent eigenvectors, then it is diagonalizable. In fact, this is exactly the condition for diagonalizability.

Theorem 164. An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

Here are the steps to diagonalize an $n \times n$ matrix A.

Algorithm (Diagonalization)**.** (1) Find the eigenvalues of A.

(2) Find n linearly independent eigenvectors of A. (If there are not n of them, then by Theorem 164, A is not diagonalizable.)

(3) Construct P with columns from eigenvectors.

(4) Construct D from eigenvalues in order corresponding to P.

Of course this process is not unique. You can choose a different order for your eigenvalues as the diagonal entries in D. You just need to make sure that the eigenvectors in P are written in the same order as the eigenvalues in D. Also, you could scale your eigenvectors in P (or, if the eigenspace has dimension higher than 1, you could take linear combinations of eigenvectors with the same eigenvalue).

We now turn to the question of: when is A diagonalizable? Of course, by the above theorem, this happens when A has n linearly independent eigenvectors. But how can we tell when eigenvectors are going to be linearly independent?

Theorem 165. If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent.

Corollary 166. An $n \times n$ matrix with n distinct real eigenvalues is diagonalizable.

Proof. If A has n distinct real eigenvalues, then they must all have multiplicity 1, since the characteristic polynomial of A is a polynomial of degree n. Hence, the dimension of each eigenspace is 1, so we can choose one eigenvector from each of the n eigenspaces. By the previous theorem, these are linearly independent, so we have n linearly independent eigenvectors of A whence it is diagonalizable. □

The next theorem provides a set of conditions required for a matrix to be diagonalizable.

Theorem 167. Suppose that an $n \times n$ matrix A has only real eigenvalues. Then A is diagonalizable if and only if the dimension of each eigenspace is equal to the multiplicity of the corresponding eigenvalue.

Proof. Here is the idea of the proof. We know that A will be diagonalizable as long as it has n linearly independent eigenvectors.

We also know that each eigenspace has dimension less than or equal to the multiplicity of its associated eigenvalue, and that the multiplicities sum to n (since the degree of the characteristic polynomial is n and we know that A has only real eigenvalues).

So in order to have n linearly independent eigenvectors, we need each eigenspace to have dimension as large as possible. And if you have enough eigenvectors in each eigenspace, then since eigenvectors with different eigenvalues are linearly independent, you will have n linearly independent eigenvectors. □

Example 168. Diagonalize the following matrix.

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

The eigenvalues for this matrix are 1 and -2. We compute a basis for $E_\lambda(A)$ for each eigenvalue.

$$A-(1)I \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\},$$

$$A-(-2)I \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Thus, A is diagonalizable by the matrices

$$P = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

3Blue1Brown has a video on change of basis here: https://youtu.be/P2LTAUO1TdA.

27. Monday 11/25: Change of Basis (4.4)

We started class with the following example to review diagonalization.

Example 169. Determine if the following matrix is diagonalizable. Do as little work as possible.

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

The characteristic polynomial of A is $(\lambda - 1)(\lambda + 2)^2$ so the eigenvalues of this are 1 and -2. The multiplicity of -2 is 2. We know that $E_1(A)$ will have dimension 1. We just need to check the dimension of $E_{-2}(A)$. If the dimension is 2, then A will be diagonalizable, otherwise it will not be.

So we compute

$$\text{null}(A + 2I) = \text{null} \begin{bmatrix} 4 & 4 & 3 \\ -4 & -4 & -3 \\ 3 & 3 & 3 \end{bmatrix} = \text{null} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

which has dimension 1.

Hence, A is *not* diagonalizable.

We now return to chapter 4 to finish up section 4.4 on change of basis.

Change of Basis.

Recall the following theorem from when we learned about bases.

Theorem 170. Let S be a subspace. If $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of S, and $\mathbf{x} \in S$ is any vector in S, then there is a *unique* set of scalars $c_1, c_2, \ldots, c_n$ such that

$$\mathbf{x} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n.$$

If we all agree on a basis $\mathcal{B}$, then to describe a vector $\mathbf{x}$, I can just give you the scalars (think of these as the directions to $\mathbf{x}$ in terms of the agreed-upon basis $\mathcal{B}$). Since we have agreed upon the basis, the scalars tell you all of the information about $\mathbf{x}$.

Definition 171. Suppose $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of $\mathbb{R}^n$. For $\mathbf{x} \in \mathbb{R}^n$, express $\mathbf{x}$ as a linear combination of the basis vectors (which you can do in only one way):

$$\mathbf{x} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n.$$

We write

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}_{\mathcal{B}}$$

for the coordinate vector of $\mathbf{x}$ with respect to $\mathcal{B}$.

Remark. If $\mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ -1 \\ 2 \end{bmatrix}$, then we all know that this means that $\mathbf{x} = 3\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. So w we write a vector, we are implicitly working in the standard basis $\mathcal{S} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of $\mathbb{R}^n$! That is, $[\mathbf{x}]_{\mathcal{S}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{S}}$. But other bases are just as good. We can, for example, choose to work in a new basis $\mathcal{B} = \{\mathbf{e}_1, -\mathbf{e}_2, \mathbf{e}_3, -\mathbf{e}_4\}$. If we lived in a world where this was the normal basis, we would probably call $\mathbf{x}$ instead $\begin{bmatrix} 3 \\ -4 \\ -1 \\ -2 \end{bmatrix}$, or in other words $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ -4 \\ -1 \\ -2 \end{bmatrix}_{\mathcal{B}}$.

Observation. Notice that if we let

$$U = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{bmatrix}$$

be the $n \times n$ matrix whose columns are our basis vectors, then

$$\mathbf{x} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = U[\mathbf{x}]_{\mathcal{B}}.$$

So the matrix U tells you how to go from coordinates with respect to $\mathcal{B}$ to the standard basis. Also, since the columns of U are a basis, this means that U is invertible, and we can rewrite the above

equation as

$$[\mathbf{x}]_{\mathcal{B}} = U^{-1}\mathbf{x}.$$

So given a vector written in the standard basis, if we would like to express it in terms of our new basis $\mathcal{B}$, we multiply by U^{-1}. (This also shows that converting $\mathbf{x}$ from the standard basis to coordinates with respect to $\mathcal{B}$ is a linear transformation!) We should record this as a theorem.

Theorem 172. Let $\mathbf{x} \in \mathbb{R}^n$ and let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be any basis of $\mathbb{R}^n$. If $U = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{bmatrix}$ then

(1) $\mathbf{x} = U[\mathbf{x}]_{\mathcal{B}}$

(2) $[\mathbf{x}]_{\mathcal{B}} = U^{-1}\mathbf{x}$.

Definition 173. The matrix U is called a change of basis matrix.

Let's do a small example.

Example 174. Let $\mathcal{B} = \left\{ \mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ which is a basis of $\mathbb{R}^2$, since they are two linearly independent vectors in $\mathbb{R}^2$. Consider the vector $\mathbf{x} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$. We should be able to express $\mathbf{x}$ uniquely as a linear combination of the basis vectors. Indeed,

$$\mathbf{x} = 5\mathbf{u}_1 - 3\mathbf{u}_2.$$

This means that

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}_{\mathcal{B}}.$$

How could we have computed that with a change of basis matrix? Well the matrix

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

has the property that $U[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$. And

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

has the property that $U^{-1}\mathbf{x} = [\mathbf{x}]_{\mathcal{B}}$. Therefore,

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}_{\mathcal{B}},$$

as we computed above.

Why would we ever want to change basis? There are actually many reasons. Sometimes whatever you are trying to do is naturally or more easily done in some special basis (other than the standard one).

We talked about some examples in class. One example was about rendering a scene with perspective. If you want to put a (rectangular) painting on the wall with perspective, you actually want to perform a linear transformation to make your painting into a parallelogram. Basically what this is doing is changing basis to the natural basis for the wall (in perspective).

We also talked a bit about JPEG image compression and how it is related to change of basis. When an image is compressed using JPEG, it is first cut into 8×8 chunks. We could record each of the RGB values in an 8×8 matrix to represent each of these 8×8 chunks of the image. We could then view these matrices as vectors in $\mathbb{R}^6 4$. JPEG makes use of the fact that images are not just random vectors in $\mathbb{R}^6 4$. It chooses a very special basis using the *discrete cosine transform* (which is related to the Fourier transform). To compress the image, you can just forget very small coefficients in this basis, and treat them as 0. Most of the information is contained in the large coefficient terms.

Being able to change basis is very useful. We talked a bit about how you might want to classify Netflix users by projecting onto a smaller-dimensional space. Basically, you want to change basis to include a basis of the smaller-dimensional space, which contains most of the information about a users preferences.

28. Wednesday 11/27 and Friday 11/29: Thanksgiving

There was no class on Wednesday and Friday due to the Thanksgiving holiday. If you went home for the holiday, your assignment was to tell your family about something that you're learning this quarter that you find cool. There are no shortage of things to say about linear algebra...

Enjoy the break!

29. Monday 12/2: Markov Chains (3.5)

Change of Basis Perspective on Diagonalization.

Recall that given a basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$, you can form the matrix $P = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{bmatrix}$ whose columns are your basis vectors. We saw (in section 4.4) that this is a *change of basis matrix* such that $\mathbf{x} = P[\mathbf{x}]_{\mathcal{B}}$, that is, P takes vectors written with respect to the basis $\mathcal{B}$ to their representation in the standard basis.

Similarly, P^{-1} is the matrix that takes a vector written in the standard basis and expresses it in terms of the basis $\mathcal{B}$.

When we diagonalize an $n \times n$ matrix A, we find a basis of eigenvectors for $\mathbb{R}^n$ and make them the columns of a matrix P and write $A = PDP^{-1}$. So this is clearly related to some kind of change of basis. How?

The idea is that the linear transformation given by A acts in a very simple way on eigenvectors! If $\mathbf{v}_i$ is an eigenvector with eigenvalue λ_i, then $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. The linear transformation acts *diagonally* on this basis.

So to figure out what A does to a vector $\mathbf{x}$, it is convenient to change basis to the basis of eigenvectors, i.e., write

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

(we can do this via the matrix P^{-1}). Then it is clear that what A does to this vector is it takes it to

$$A\mathbf{x} = c_1\lambda_1\mathbf{v}_1 + \cdots + c_n\lambda_n\mathbf{v}_n$$

(which we can do via the matrix D). But this is still written in terms of the basis of eigenvectors. In order to write it with respect to the standard basis, we need to change back to the standard basis. We can do this via the matrix P.

Altogether, we have $A = PDP^{-1}$.

Markov Chains (3.5).

Our next goal is to understand Google's PageRank algorithm. In order to give ourselves a firm footing to do this, we first learn about Markov chains.

Example 175. Suppose that students at UW behave in the following way:

(1) If they are in class at one time, then one hour later, there is a 50 percent chance they are still in class, a 30 percent chance they will be in the library, and a 20 percent chance they will be at the HUB.

(2) If they are at the library at one time, then one hour later, there is a 30 percent chance they will be in class, a 50 percent chance they will still be at the library, and a 20 percent chance they will be at the HUB.

(3) If they are at the HUB at one time, then one hour later, there is a 10 percent chance they will be in class, a 20 percent chance they will be at the library, and a 70 percent chance they will still be at the HUB.

Question: Suppose all the students start out in class. After a long time, what fraction of them will be in class, at the library, and at the HUB?

Suppose that at a certain hour, the fraction of students that are in class, at the library, and at the HUB are v_C, v_L and v_H respectively. Let v'_C, v'_L and v'_H be the fractions one hour later.

Then

$$\begin{bmatrix} v'_C \\ v'_L \\ v'_H \end{bmatrix} = \begin{bmatrix} 0.5v_C + 0.3v_L + 0.1v_H \\ 0.3v_C + 0.5v_L + 0.2v_H \\ 0.2v_C + 0.2v_L + 0.7v_H \end{bmatrix}$$

and so

$$\begin{bmatrix} v'_C \\ v'_L \\ v'_H \end{bmatrix} = \begin{bmatrix} 0.5 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} v_C \\ v_L \\ v_H \end{bmatrix}.$$

We can use this to easily calculate what the fractions will be. At $t = 0$, we have $v_C = 1$, $v_L = 0$ and $v_H = 0$. Then we get the following table.

t	v_C	v_L	v_H	t	v_C	v_L	v_H
0	1	0	0	6	0.279	0.327	0.394
1	0.5	0.3	0.2	7	0.277	0.326	0.397
2	0.36	0.34	0.30	8	0.276	0.326	0.398
3	0.312	0.338	0.350	9	0.276	0.325	0.399
4	0.292	0.333	0.375	10	0.275	0.325	0.400
5	0.283	0.329	0.388	11	0.75	0.325	0.400

It appears that in the long run, 27.5% of the students are in class, 32.5% of the students are at the library, and 40% of the students are at the HUB. We'll develop some theory that shows that's true.

What is a Markov chain?

Definition 176. A Markov chain is a probabilistic process where you can reliably make predictions of the future based *only* on the present state (without knowing the full history).

A Markov chain has n states, and for each pair (i, j), there is a probability of moving from state i to state j, p_{ij}. We make this into a matrix P where the ijth entry in P is p_{ji}. (This might be a bit weird. What it means is that we put all the probabilities of moving out of state i in the ith column. That's what we did in the example above.)

The matrix in P is called the transition matrix of the Markov chain.

Definition 177. A probability vector is a vector $\mathbf{x}$ in $\mathbb{R}^n$ with nonnegative entries whose entries add up to 1.

We will use a probability vector keeps track of the state of the Markov chain (the ith entry of $\mathbf{x}$ is the probability of being in state i).

Definition 178. A stochastic matrix is a square matrix whose columns are probability vectors.

In our UW example, the transition matrix $P = \begin{bmatrix} 0.5 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.7 \end{bmatrix}$ is a stochastic matrix.

We let $\mathbf{x}_0$ be the starting state. For $k \geq 0$, define $\mathbf{x}_k = P\mathbf{x}_{k-1}$.

So $\mathbf{x}_k$ encodes the state after k "steps".

What are good examples of Markov chains?

(1) Speech analysis: Voice recognition, keyboard word prediction.

(2) Brownian motion: You know the probability of the particle moving to a certain position given its current position.

(3) Create randomly generated words or sentences that look meaningful. (The website http://projects.haykranen.nl/markov/demo/ is quite fun. You can generate random phrases by making the next word be chosen randomly based on the previous 4 words, according to the distribution of words in the source. As a source, you can use the Wikipedia article about Calvin and Hobbes, Alice and Wonderland, or Kant.)

(4) Board games: You could consider the Markov chain whose states, say, are the 40 spaces on a Monopoly board. If you know what space you are currently on, then because you know the

probabilities of dice rolls and the probabilities of chance cards, you can compute the probability that you move from your current square to any other square. This is why you have heard that in Monopoly you should aim to obtain the orange and light blue properties. In the long run, these squares give a good return on investment, because players are more likely to be on those squares.

Markov chains are quite important—they are the main objects you study in Math 491.

How can we determine the long-term behavior?

Definition 179. A stochastic matrix P is regular if there is some $k > 0$ for which all the entries in P^k are positive.

The matrix P from the UW example is regular, because all the entries of P are positive.

Definition 180. A steady-state vector $\mathbf{q}$ is a probability vector for which $P\mathbf{q} = \mathbf{q}$. Note that this is the same thing as an eigenvector with eigenvalue 1!

The following is an amazing theorem which we will not prove (see Math 491).

Theorem 181. If P is an $n \times n$ regular stochastic matrix, then P has a unique steady-state vector. Further, if $\mathbf{x}_0$ is *any* initial state and $\mathbf{x}_{k+1} = P\mathbf{x}_k$, then over time $\mathbf{x}_k$ converges to $\mathbf{q}$, i.e.:

$$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{q}.$$

(Also, this convergence happens relatively quickly.)

Another way to state this theorem is that any $n \times n$ regular stochastic matrix has eigenvalue equal to 1 with a one-dimensional eigenspace. And every vector in $\mathbb{R}^n$ converges to this eigenvector as P is repeatedly applied!

30. Wednesday 12/4: Google PageRank as a steady-state vector

This lecture is adapted from http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html.

Back to the UW example.

The matrix

$$P = \begin{bmatrix} 0.5 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.7 \end{bmatrix}$$

is a regular stochastic matrix, so the theorem above applies. We just need to find the eigenvector $\mathbf{q}$ with eigenvalue 1 so that $P\mathbf{q} = \mathbf{q}$.

Solving for the nullspace of $P - I$, row reduction tells us that the null space is one-dimensional and is spanned by

$$\mathbf{v} = \begin{bmatrix} 1 \\ \frac{13}{11} \\ \frac{16}{11} \end{bmatrix}.$$

We take a scalar multiple of $\mathbf{v}$ to make it a probability vector. To do this, we let $c = \frac{1}{1+\frac{13}{11}+\frac{16}{11}} = \frac{11}{40}$. Then

$$\mathbf{q} = c\mathbf{v} = \begin{bmatrix} \frac{11}{40} \\ \frac{13}{40} \\ \frac{2}{5} \end{bmatrix} = \begin{bmatrix} 0.275 \\ 0.325 \\ 0.4 \end{bmatrix}.$$

We will now talk about Google PageRank, the algorithm invented by the Larry Page and Sergey Brin while they were graduate students at Stanford. PageRank is an algorithm to rank webpages by importance, and it is essentially the implementation of one very good idea with some basic linear algebra (which you are capable of understanding completely!).

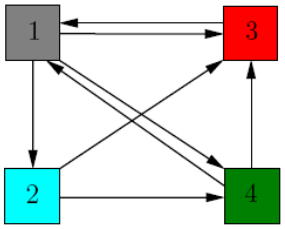
Let's think about the problem that they were facing. How can we create a search engine? One first idea is to have our search engine keep an index of all web pages. When a user performs a search, we can simply look through the index and count the occurrences of the key word in each web page. We can then return the web pages in order by the number of occurrences.

This naive approach has some major problems. Simply listing a key word multiple times is no guarantee of relevance of the web page. If a user searches for "University of Washington", we want our search engine's first hit to be uw.edu. There may be many pages on the Internet that include the phrase "University of Washington", perhaps many more times than www.uw.edu does. Indeed, we could make a web page that simply has the phrase "University of Washington" listed millions of times and nothing else. We really don't want this web page to be a high hit.

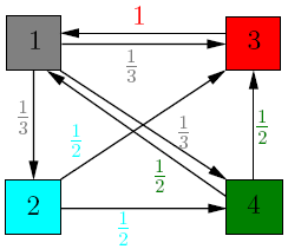
The idea is we want to return web pages related to the phrase "University of Washington" that are particularly relevant or authoritative. The idea behind PageRank is that the hyperlinks between web pages give some clue to the importance of the page. If page A links to page B, then page A considers page B to be important or relevant to its topic. If lots of other web pages point to B, this means that there is a common belief that page B is important.

On the other hand, if few pages link to B, but if the pages that link to B are important or relevant themselves (`google.com`, `uw.edu`, `cnn.com`), then B still be considered important. So the importance of a page has something to do with the number of pages linking to B and the importance of those pages.

To this aim, we begin by picturing web pages and the links between them as a directed graph, with nodes each web page being represented by a node. We put an arrow from i to j if i links to j. Suppose we have four sites on the Internet which are related to the phrase "University of Washington". We want to know how to rank them. Suppose they are given by the following directed graph:



In our model, each page should transfer its importance evenly to the pages that it links to. Node 1 has 3 outgoing edges, so it will pass on 1/3 of its importance to each of the other 3 nodes. Node 3 has only one outgoing edge, so it will pass on all of its importance to node 1. In general, if a node has k outgoing edges, it will pass on $1/k$ of its importance to each of the nodes that it links to. Let us better visualize the process by assigning weights to each edge.



Let us denote by A the transition matrix of the graph, $A = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$.

Notice that this is a stochastic matrix! What does it represent? We can think of a *random surfer*, who starts at some web page, and then clicks some link on the web page at random. So if you start at page 1, the surfer has a 1/3 probability of landing at pages 2, 3, or 4.

Since A is a stochastic regular matrix, there is a unique steady-state vector. For our random surfer, this vector gives the probability that he will be at each web page (after surfing randomly for a long time). This vector is our PageRank vector[17].

We know that for a stochastic regular matrix, the steady-state vector is given by the an eigenvector with eigenvalue 1 that is a probability vector. So we simply solve $\text{null}(A - I) = \text{Span}\left\{\begin{bmatrix}12\\4\\9\\6\end{bmatrix}\right\}$. Of course this isn't the steady-state vector yet because it is not a probability vector. We have to normalize so that we choose the eigenvector whose entries sum to 1. The PageRank vector is $\begin{bmatrix}12/31\\4/31\\9/31\\6/31\end{bmatrix} \approx \begin{bmatrix}.38\\.12\\.29\\.19\end{bmatrix}$

The PageRank vector we have computed indicates that page 1 is the most relevant page. This might seem surprising since page 1 has 2 backlinks, while page 3 has 3 backlinks. If we take a look at the graph, we see that node 3 has only one outgoing edge to node 1, so it transfers all its importance to node 1. Equivalently, once a web surfer that only follows hyperlinks visits page 3, he can only go to page 1. Notice also how the rank of each page is not trivially just the weighted sum of the edges that enter the node. Intuitively, at step 1, one node receives an importance vote from its direct neighbors, at step 2 from the neighbors of its neighbors, and so on.

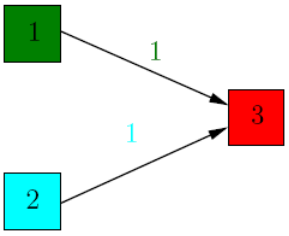
And this is really the entire idea! Think of a random surfer surfing the pages related to the query, and then rank them by the probability that the random surfer is there after a long time.

Fixing Some Problems.

We do need to patch our idea because certain graphs will yield funny PageRank results. This is because not every directed graph formed this way gives a regular stochastic matrix. Let me give you some examples.

Example 182. Suppose some page has no outgoing links.

[17] Almost, we need to change it a bit in order to avoid the problems that arise in the next section of the notes

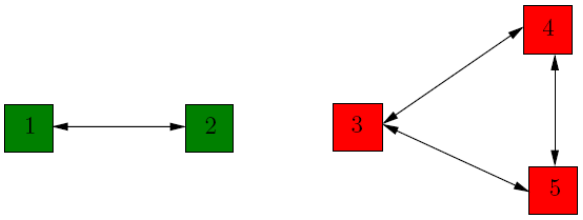


Then the transition matrix is given by $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$ which is *not* a stochastic matrix. As a result, 1 is not an eigenvalue of this matrix.

An easy fix for this problem would be to replace the column corresponding to the dangling node 3 with a column vector with all entries 1/3. In this way, the importance of node 3 would be equally redistributed among the other nodes of the graph, instead of being lost.

Now we have the transition matrix $\begin{bmatrix} 0 & 0 & 1/3 \\ 0 & 0 & 1/3 \\ 1 & 1 & 1/3 \end{bmatrix}$. This matrix is now a regular stochastic matrix.

Example 183. Or perhaps there are disconnected components of the web graph.



A random surfer that starts in the first connected component has no way of getting to web page 5 since the nodes 1 and 2 have no links to node 5 that he can follow. Linear algebra fails to help as well. The transition matrix for this graph is $A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}$. This matrix has an eigenspace of dimension 2 associated to eigenvalue 1, spanned by $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ and

$\mathbf{u} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. The reason why this happened is that A was not a **regular** stochastic matrix. If you continue taking powers of A, some of the entries will always be 0. So, both in theory and in practice, the notion of ranking pages from the first connected component relative to the ones from the second connected component is ambiguous.

The web is very heterogeneous by its nature, and certainly huge, so we do not expect its graph to be connected. Likewise, there will be pages that are plain descriptive and contain no outgoing links. What is to be done in this case? We need a non ambiguous meaning of the rank of a page, for any directed Web graph with n nodes.

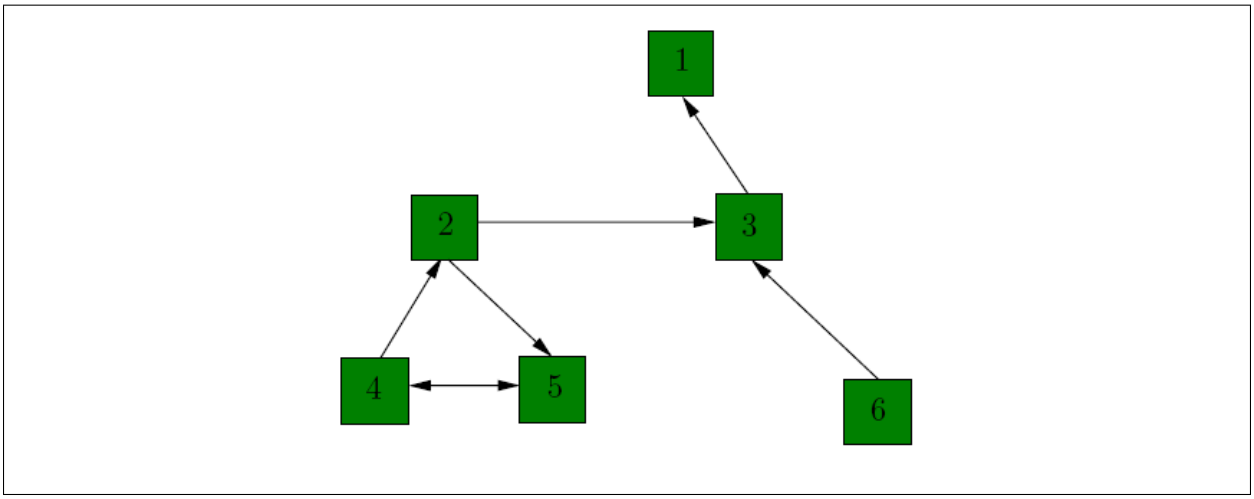
The solution is also a simple one. Rather than modeling a random surfer who only clicks links, we imagine a random surfer who sometimes clicks links, but sometimes just goes to a random page on the Internet. This will fix the problem because now it is possible to go from one component to another. How do we model this?

We fix a positive constant p between 0 and 1, called a *damping factor* (a typical value is 0.15). We imagine that with probability p, the random surfer picks one of the n pages on the Internet and teleports there. With the remaining probability $1 - p$, the surfer clicks one of the links on the current page at random. Hence, we can take the matrix

$$M = (1 - p)A + pB$$

where $B = \frac{1}{n} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$.

Example 184. Compute the PageRank vector of the following graph, considering the damping constant to be $p = 0.15$.


1
2
3
4
5
6

31. Friday 12/6: Examples of Singular Value Decomposition

We will not cover singular value decomposition in this course, but the idea is this. If A is not diagonalizable (maybe not even square!), how close can you come to diagonalizing A? That is, if A can't be written as PDP^{-1}, can you do something similar?

Given an $m \times n$ matrix A, you can look at A^TA, which is a symmetric square matrix! You can then diagonalize A^TA. The square roots of the eigenvalues of A^TA are called the *singular values* of A. It is then possible to decompose A into what is called its singular value decomposition (which you should think of as analogous to diagonalization).

Theorem 185 (Singular Value Decomposition)**.** Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ where D is an $r \times r$ diagonal matrix for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. and there exists an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that $A = U\Sigma V^T$.

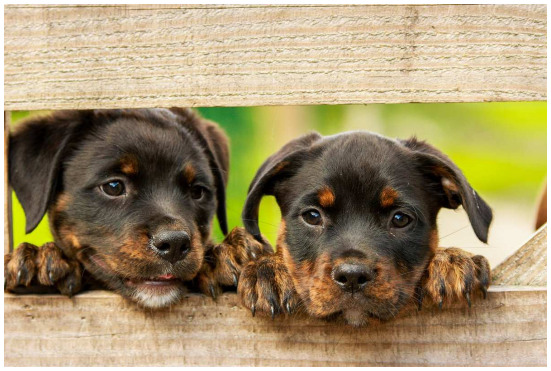
31.1. **Image Compression.** It turns out that "most" of the information about the matrix A is actually contained in the first few singular values (the largest ones). So rather than taking all r singular values, you can take the first k singular values and approximate the matrix A using these k singular values and the first k singular vectors.

But we can actually visualize what I mean when I say "most of the information". Let's look at an example using Mathematica (You can follow along using the file `imageCompression.nb`[18]).

First we import an image which we call `img`[19].

```
img = Import[''U:\\SVD Mathematica\\image.jpg'']
```

Out[ ]=



The original image is 800×1200 pixels (I've resized it in this document to save space). How can you store the data? Well you can store the RGB values in three separate 800×1200 matrices. We will compress the image by computing the SVD using the first few singular values for each of

[18] Thanks to Frank Moore at Wake Forest for showing me how to do this in Mathematica

[19] This procedure should work for all animals, but I've only tested it on pictures of puppies.

these matrices. Let's start by trying 100 singular values. This will decompose an 800×1200 matrix M into the product of three matrices $U\Sigma V^T$ where U is 800×100, Σ is 100×100, and V is 1200×100.

At the end we will compute how much less data we need to store. You can also play around with the number of singular values and see what effect it has visually.

```
numSing = 100;
```

We now separate our original image into the three color channels

```
imgs = ColorSeparate[img]
```

Let's take a look at the first channel, which is the red channel.

```
img1Data = ImageData[imgs[[1]];
height = Dimensions[img1Data][[1]]
width = Dimensions[img1Data][[2]]
```

```
800
1200
```

As expected, this is an 800×1200 matrix. Let's take a closer look at it

```
MatrixRank[img1Data]
img1Data[[1,1;;5]]
```

```
800
{0.976471,0.992157,0.996078,0.980392,0.968627}
```

The matrix has rank 800 (we should have expected this, since there's no reason for there to be any linear relationships between the columns of this image). Also looking at the first five entries of the matrix, we can see how Mathematica is storing this image. It stores each pixel as a number between 0 and 1, representing how much red occurs at that pixel.

We now compute the SVD of the matrix, using the first `numSing` singular values.

```
{u1, s1, v1} = SingularValueDecomposition[img1Data, numSing];
newImg1Data = u1.s1.(Transpose[v1]);
```

The matrices in the SVD are `u1`, `s1`, and `v1`. So the product of these matrices which we call `newImg1Data` is an 800×1200 matrix that should approximate the matrix for the red channel. Let's compare them!

```
imgs[[1]]
```

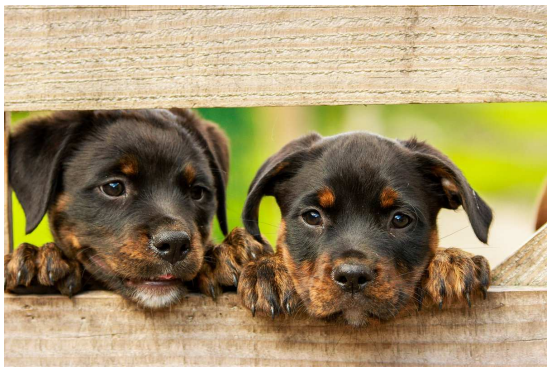
```
newImg1 = Image[newImg1Data]
```

Not bad! The image is not as crisp as before, but it is definitely still very recognizable. Let's do the same thing for the green and blue channels.

```
img2Data = ImageData[imgs[[2]];
{u2, s2, v2} = SingularValueDecomposition[img2Data, numSing];
newImg2Data = u2.s2.(Transpose[v2]);
newImg2 = Image[newImg2Data];
img3Data = ImageData[imgs[[3]];
{u3, s3, v3} = SingularValueDecomposition[img3Data, numSing];
newImg3Data = u3.s3.(Transpose[v3]);
newImg3 = Image[newImg3Data];
```

Let's now compare our original image to the image that we get using the SVDs for the three color channels combined into one image. The original image:

```
img
```



And the new compressed image obtained from combining the compressed color channels.

```
newImg = ColorCombine[{newImg1, newImg2, newImg3}, ‘‘RGB’’]
```

Okay, now back to the question of how much less data do we need to store for the SVDs? Well like I said, we decomposed the 800×1200 into the product of three matrices $U\Sigma V^T$ where U is 800×100, Σ is 100×100, and V is 1200×100. Let's just make sure that's true

```
{Dimensions[u1] ,Dimensions[s1], Dimensions[v1]}
```

```
{{800, 100}, {100, 100}, {1200,100}}
```

Since the matrix `s1` is diagonal, we only need to store the diagonal entries. So that gives us `numSing` entries. The matrices `u1` and `v1` are $800\times$`numSing` and $1200\times$`numSing`. Altogether, we need to store `numSing`$(800 + 1200 + 1)$ entries. How does this compare to $800 \cdot 1200$?

```
(numSing * (height + width + 1.))/(height * width)
```

```
0.208438
```

So we only need to store $\approx 20\%$ as much data as for the original image. As you change the number of singular values, you can see how this number changes, as well as the compressed image.

31.2. **Principal Component Analysis.** We also looked at an amazing figure from a *Nature* paper in 2008 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2735096/.

The authors of this paper used Principal Component Analysis (PCA) to find the first two principal components in their dataset (consisting of genetic data from 3000 Europeans). This involves a singular value decomposition and is an example of unsupervised learning. The amazing thing is that when graphing the individuals with respect to the first two principal components, you can see a recognizable map of Europe. In some sense, these people have a map of Europe hidden within their genes!