Matrices and Linear Transformations (21-241)
Spring 2023 Lecture Notes

Elisa Bellah

Contents

Preface

"Mathematics is the art of reducing any problem to linear algebra."

"If you can reduce a problem to linear algebra, you've won. If you reduce it to combinatorics, you're screwed."

I'm unsure who these quotes are attributed to, but they are widely known in the mathematics community because they highlight two important features:

(1) Linear algebra is a tool that captures many fundamental ideas in mathematics;

(2) The basic theory of linear algebra (unlikely many other areas, such as combinatorics) is complete.

In this class, we'll learn the fundamental objects of linear algebra: matrices and linear transformations. We'll introduce these objects by building off of the intuitive problem of solving systems of linear equations (which we will see in the first section). My hope for this semester is for you to leave with a conceptual understanding of linear algebra and an appreciation of its strength both in mathematics and in the sciences.

These lecture notes will follow Alayont and Schlicker's "Linear Algebra and Applications: An Inquiry-Based Approach" and will be accompanied by in-class activities. Please keep in mind that these notes are only to be used for Lecture 3 of 21-241 in Spring 2023 at CMU, and are not to be distributed. Let me know if you catch any mistakes or typos along the way.

List of Notation.

The following list will be updated as new notation appears in the notes.

$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers
$\mathbb{Q}$	the set of rational numbers
$\mathbb{R}^n$	n-dimensional Euclidean space
$\in$	is an element of
$\forall$	for all
$\exists$	there exists
T_A	the linear transformation corresponding to the matrix A, given by $T_A(\vec{x}) = A\vec{x}$.
A_T	the defining matrix of a linear transformation T, as constructed in Theorem 1.46
$[\vec{x}]_{\mathcal{B}}$	the coordinates of a vector $\vec{x}$ with respect to the basis $\mathcal{B}$
$A^\top$	the matrix transpose
$\mathrm{Nul}(A)$	the null space of A
$\mathrm{Col}(A)$	the column space of A
E_λ	the λ-Eigenspace
$\det(A)$	the determinant of A
χ_A	the characteristic equation of A
$\vec{x} \cdot \vec{y}$	the dot product
$\lVert\vec{x}\rVert$	the norm

Chapter 1

Systems of Linear Equations

1.1. Introduction to Systems of Linear Equations

The following definition is meant to formalize our intuitive understanding of systems of linear equations.

Definition 1.1. A LINEAR EQUATION in variables $x_1, x_2 \ldots, x_n$ is an equation of the form

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b,$$

for constants $a_i, b \in \mathbb{R}$. A SYSTEM OF LINEAR EQUATIONS is a collection of one or more linear equations in the same variables. A tuple $(s_1, \ldots, s_n) \in \mathbb{R}^n$ is a SOLUTION to a system of linear equations if $(s_1, \ldots, s_n)$ is a solution to every linear equation in the system.

Example 1.2. The equation $2x + y = z$ is linear, while the equations

$$x^2 - y = 3, \;\; e^x + \sqrt{y} = z^3, \;\; \frac{x}{y} + 2 = zw$$

are not linear.

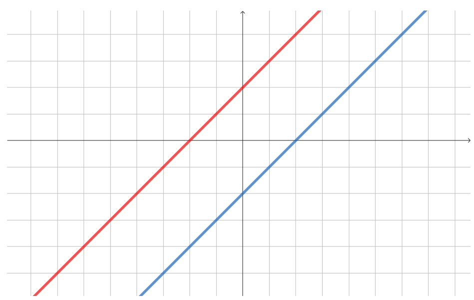
In Worksheet 0, you encountered a few systems of linear equations in three variables. My guess (which we'll discuss later) is that you used some ad-hoc methods to find their solutions. Our goal in the next section will be to introduce matrices as the appropriate bookkeeping in order to solve systems of linear equations algorithmically (so that we can program a computer to do this hard work for us).

This bookkeeping will also help us build a theory behind what our solution sets can look like. But it's difficult to make conjectures with bookkeeping alone (this is certainly not how the theory was developed). Let's first build some geometric intuition before we dive into the theory.

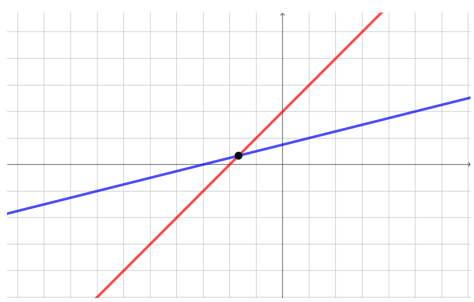
1.1.1. Some Geometry of Solution Sets of Linear Equations. Let's look at some geometry of systems linear equations in two and three variables.

Two Variables. Since the graph of a linear equation in two variables is a line, solving systems of linear equations in two variables is equivalent to finding intersection points of lines in $\mathbb{R}^2$. If our system has two equations, then there are three possible situations:

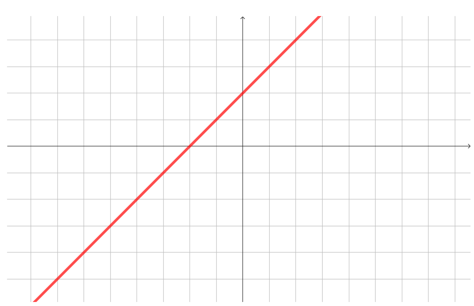
(i) lines are parallel and distinct:



(ii) lines are not parallel:



(iii) lines are equal:



In this case, our system of equations either has no solutions, one solutions, or infinitely many solutions, respectively.

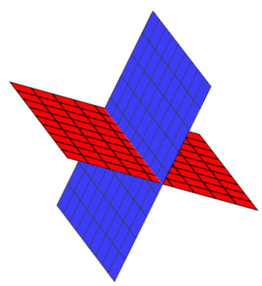
Three Variables. Since the graph of a linear equation in three variables is a plane, solving systems of linear equations in two variables is equivalent to finding intersection points of planes in $\mathbb{R}^3$.

A system of **two** linear equations in three variables then corresponds to the number of points two planes in $\mathbb{R}^3$ may intersect at. We have the following cases:

(i) If the planes are parallel and distinct, then our system has no solutions.

(ii) If the planes are equal, then our system has infinitely many solutions.

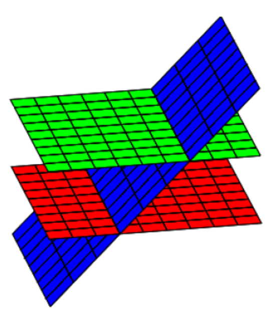
(iii) If the planes are not parallel, then our system has infinitely many solutions, and the solution set is given by a line as below



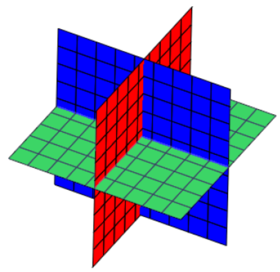
A system of **three** linear equations in three variables corresponds to the number of points three planes in $\mathbb{R}^3$ may intersect at. We have the following cases:

(i) One or more of the planes are equal, in which case we're back to the previous situation (zero or infinitely many solutions).

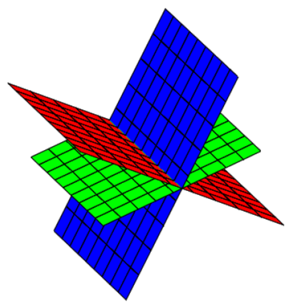
(iv) All planes are distinct, and two or more of the planes are parallel, such as in the picture below (in which case we have zero or infinitely many solutions)



(ii) Planes intersect at a point, as in the picture below



(iii) Planes intersect in a line, as in the picture below



1.1.2. Consequence and Limitations of our Geometric Intuition. In all of our previous examples, we saw that our geometric objects either intersected at no points, one point, or infinitely many points. This leads to the following conjecture.

Conjecture 1.3. Any system of linear equations either has no solutions, one solution, or infinitely many solutions.

Using our geometric intuition helped us form this conjecture, and will be used throughout the course to help us get our barrings. Learning how to reason through what's going on (through examples, geometry, etc) is an invaluable part of the mathematical process. But this process is not complete until we can provide rigorous proofs of our claims. The machinery we'll need to prove Conjecture 1.3, and to generate algorithms to find our solutions, is the matrix.

1.2. The Matrix Representation of a Linear System

Recall our system of equations from Worksheet 0

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

Since our variables must all be the same in any system of linear equations, it's a bit redundant to write them every time. Instead, we could just write down the important pieces of this system in an array

$$\begin{matrix} 1 & -1 & -1 & 1 \\ 2 & -3 & -1 & 3 \\ -1 & 1 & -1 & -3 \end{matrix}$$

The first three columns of this array correspond to our three independent variables, and the last column corresponds to the constant on the right-hand side of our linear equations. This array is a matrix and will be our basic bookkeeping device. Let's give some formal definitions.

Definition 1.4. A MATRIX is any rectangular array of quantities or expressions. The quantities or expressions in a matrix are called its ENTRIES. If a matrix has n rows and m columns, then we call this an $n \times m$ matrix.

In this class, our matrices will generally typically contain real number entries (or sometimes variables working as placeholders for real number entries). Matrices are often written using soft brackets, such as in the 2×2 matrix below

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

or by using hard brackets, such as in the 3×2 matrix below

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

I'll be using soft brackets throughout these notes, since that's what I'm used to, but either one is perfectly fine.

Definition 1.5. Consider a general system of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= b_n. \end{aligned}$$

The MATRIX OF COEFFICIENTS corresponding to this system is given by

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

The AUGMENTED MATRIX of this system is

$$\left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & b_n \end{array}\right).$$

Note that the augmented matrix accounts for the constants on the right-hand side of our equations, while the coefficient matrix does not.

Let's develop our ad-hoc methods of solving systems of linear equations into something more algorithmic, which will be easier to keep track of with the use of matrices.

Definition 1.6. We call two systems of linear equations EQUIVALENT if they have the same solution set.

Observation 1.7. Given a system of linear equations, we can perform the three ELEMENTARY OPERATIONS to obtain an equivalent system:

(E1) Replace one equation by the sum of that equation and a scalar multiple of another equation;

(E2) Interchange two equations;

(E3) Replace an equation by a nonzero multiple of itself.

Example 1.8. Let's solve the following system of equations (which we first encountered in Worksheet 0) using the elementary operations above

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

Since we're trying to find a way to do this algorithmically, let's be systematic. First, we'll try to remove two of the variables from the third equations, one variable from the second equation, and then use back substitution.

(1) Replace third equation with the sum of the first and the third equation:

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

(2) Next, replace the second equation with -2 times the first equation plus the second:

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

(3) Finally, we apply back substitution to solve our system. From the third equation we get $z = 1$ and so

$$-y + 1 = 1 \Rightarrow y = 0$$

which gives

$$x - 0 - 1 = 1 \Rightarrow x = 2.$$

So, this equation has the unique solution

$$(x, y, z) = (2, 0, 1).$$

Our next goal will be to understand the method of back substitution in the language of matrices. This will help us to develop an algorithmic method to find solutions to systems of linear equations, and to show that this method always works.

In Activity 1.2, you kept track of the corresponding augmented matrices at each step. We should have found out that the elementary operations given above correspond to the following ELEMENTARY ROW OPERATIONS

(ER1) Replace one row by the sum of that row and a scalar multiple of another row;

(ER2) Interchange two rows;

(ER3) Replacing a row by a nonzero scalar multiple of itself.

Definition 1.9. We say that two matrices are ROW EQUIVALENT if one can be obtained by some sequence of elementary row operations from another. Observe that if two augmented matrices are row equivalent, then the system of linear equations that they represent are also equivalent.

1.3. Echelon Forms of a Matrix

Example 1.8 continued... Let's keep track of the augmented matrices at each step of our solution using back substitution to Example 1.8. The augmented matrix of our original equation is

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 2 & -3 & -1 & 3 \\ -1 & 1 & -1 & -3 \end{array}\right)$$

Step (1) of our solution corresponds to replacing the third row of this matrix with the sum of the first and third row, which gives the matrix

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

Step (2) coressponds to replacing the second row with -2 times the first row plus the second, which gives the matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -2 & -2 \end{array}\right).$$

Once we're at this step, we applied back substitution. Observe that the matrix we ended up with has zeros in every entry below the diagonal.

Definition 1.10. The PIVOT (aka the LEADING ENTRY) of a row in a matrix is the leftmost nonzero entry.

Definition 1.11. A matrix is said to be in ROW ECHELON FORM if

(1) all rows consisting only of zeros are at the bottom, and

(2) the pivot of each row is in a column to the right of the leading entry of the row above it.

A matrix which is in row echelon form is also called UPPER TRIANGULAR, because the only nonzero entries form a triangle in the upper right-hand corner of the matrix.

We have the following key observations.

Observation 1.12. Once an augmented matrix is in row echelon form, we can use back substitution to find the solutions to the corresponding system of linear equations.

In fact, we can use our matrices to perform the back substitution process as well.

Example 1.8 continued... Recall we had the augmented matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -2 & -2 \end{array}\right).$$

Our next step was to note the final row corresponded to the equation

$$-2z = -2$$

and so we canceled -2 on both sides. Instead, we could do this as an elementary row operations to the augmented matrix above: replace R_3 by $-\frac{1}{2}R_3$ to get the row equivalent matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

This tells us $z = 1$. Next, we plugged $z = 1$ into the equation $-y + z = 1$ to get

$$-y + 1 = 1 \Rightarrow -y + 0 = 0.$$

This corresponds to the elementary row operation: replace R_2 by $R_2 - R_3$ to get the row equivalent matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right)$$

and then replace R_2 with $-R_2$ to get

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right)$$

telling us $y = 0$. Finally, we plugged in $z = 1, y = 0$ to the equation $x - y - z = 1$ from the top row to solve for x. This corresponds to using the second and third rows to cancel the coefficients of y and z in the first row: replacing R_1 by $R_1 + R_3$ gives the row equivalent matrix

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

and finally replacing R_1 with $R_1 + R_2$ gives the row equivalent matrix

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

Now, we can just read off our solution from the matrix above: just as before, we get the unique solution

$$(x, y, z) = (3, 0, 1).$$

Definition 1.13. A matrix is said to be in REDUCED ROW ECHELON FORM if the matrix is in echelon form and

(1) the leading term (aka the pivot) in each nonzero row is 1, and
(2) each leading term is the only nonzero entry in its column.

The process of applying elementary row operations to obtain the reduced row echelon form of a matrix is called GAUSS-JORDAN ELIMINATION.

In Activity 1.3, you practiced finding row echelon and reduced row echelon forms of further matrices. My hope is that you began to convince yourself of the following observations.

Theorem 1.14. *Every matrix is row equivalent to a matrix in row echelon form. The method of finding an equivalent matrix in row echelon form using the elementary row operations (ER1)-(ER3) is often called* GAUSSIAN ELIMINATION.

Theorem 1.15. *Every matrix is row equivalent to a matrix in reduced row echelon form. Furthermore, the reduced row echelon form of any matrix is unique.*

On your next homework, I'll ask you to give a proof of these two results in the 3×3 case (we'll practice some formal proof writing before you need to turn these in).

The important takeaway for now is that every matrix can be written in (reduced) row echelon form (and in fact there are efficient algorithms to do so), and once a matrix is in one of these forms, we can easily find the solutions to our system. This completely resolves our problem of finding solutions to linear equations.

In practice, if you have a system of linear equations you'd like to solve, this is what you should do:

(1) Write down the augmented matrix of your system
(2) Ask a computer to find the reduced row echelon form
(3) Read off the solutions

Let's look at the systems of equations you solved in Activity 1.3.

Example 1.16. Suppose we have a system of linear equations with the following augmented matrix

$$\left(\begin{array}{cccc|c} 1 & 2 & -2 & 3 & 1 \\ -2 & -4 & 4 & 2 & 6 \\ 0 & 0 & 4 & 8 & 6 \end{array}\right).$$

Just for practice, let's perform Gauss-Jordan elimination by hand to find the reduced row echelon form of this matrix. Note that this method is not unique (you may swap the order of some steps), but the matrix we end up with is.

(1) Replace R_2 with $R_2 + 2R_1$ to get

$$\left(\begin{array}{cccc|c} 1 & 2 & -2 & 3 & 1 \\ 0 & 0 & 0 & 8 & 8 \\ 0 & 0 & 4 & 8 & 6 \end{array}\right).$$

(2) Next, swap R_2 and R_3, and then replace R_3 with $\frac{1}{8}R_3$

$$\left(\begin{array}{cccc|c} 1 & 2 & -2 & 3 & 1 \\ 0 & 0 & 4 & 8 & 6 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right).$$

(3) Now, replace R_1 with $R_1 - 3R_3$ and R_2 with $R_2 - 8R_1$ to get

$$\left(\begin{array}{cccc|c} 1 & 2 & -2 & 0 & -2 \\ 0 & 0 & 4 & 0 & -2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right).$$

(4) Finally, we replace R_1 with $R_1 + \frac{1}{2}R_2$, and then replace R_2 with $\frac{1}{4}R_2$ to get our matrix in reduced row echelon form

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & -3 \\ 0 & 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 1 & 1 \end{array}\right).$$

Now, let's try to read off our solution as before. Let's use variables x, y, z, w to get soloution

$$z = -1/2,\ w = 1$$

from the second and third columns. But we have

$$x + 2y = -3$$

from the first column. Note that this does not have a unique solution! We can choose either of our variables to be FREE to describe the solution set. Let $y = t$ be free. Then we get

$$x = -3 + 2t$$

and so all solutions to this system of linear equations can be described by

$$(x, y, z, w) = (-3 + 2t, t, -1/2, 1)$$

where t is any real number.

Example 1.17. Suppose that you have a system of linear equations with the following augmented matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 1 & 1 & -3 & 1 \\ 3 & -1 & -1 & 6 \end{array}\right).$$

Asking a computer to perform Gauss-Jordan elimination for you gives that the reduced row echelon form of this matrix is

$$\left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

But then the final row of this matrix says that

$$0 \cdot z = 1 \Rightarrow 0 = 1.$$

This is nonsense, and so our system of linear equations must have no solution.

It turns out that the (reduced) row echelon form of a matrix can always tell us how many solutions a system of linear equations may have, and will help us resolve Conjecture 1.3

1.3.1. The Number of Solutions to Systems of Linear Equations. Recall that we conjectured a system of linear equations either has no solutions, one solution, or infinitely many solutions.

Definition 1.18. We say that a system of linear equations is CONSISTENT if there exists **at least** one solution. Otherwise, we call a system INCONSISTENT.

Theorem 1.19. *Suppose that a system of linear equations has augmented matrix written in reduced row echelon form as the matrix A.*

(1) The system is inconsistent if and only if there is a pivot in the last column of A.

(2) The system of equations has exactly one solution if and only if there is a pivot in each column of A, but no pivot in the last column of A.

(3) The system has infinitely many solutions if and only if it is consistent and there is a column with no pivot.

Note that Conjecture 1.3 follows as a Corollary to Theorem 1.19, since A must either have a pivot in the last column or not.

The process of mathematics happens in two steps: first, you have to convince yourself what's happening. After that, you have to convince other people you know what's happening. So far, we've been focusing on the first step. As we progress through the course, we'll start to practice the formal method of mathematical argument: the mathematical proof. Before we move on to further course material, we take a brief interlude to cover some of the basics of constructing formal mathematical proofs.

Keep in mind that formal proofs are not just a formality. Oftentimes we can trick ourselves into thinking a false statement is true (or vice versa) unless we write down all of the details and make sure our logic is sound. Sometimes (oftentimes?) examples we think will generalize will not and geometric intuition fails us. The rigor of mathematical proof saves us from making these mistakes. However, if we don't work with examples and intuition, we won't be able to generate statements to prove. Both aspects of the mathematical process are necessary, and we will aim to practice them equally throughout the course.

Interlude: Introductory Remarks on Proofs

Generally, in mathematics we are trying to prove or disprove whether certain statements are true or false. Not all statements fall into the binary of true or false. For example, if someone said "working as a postdoc at Carnegie Mellon is a great experience" I would say that is true in some ways, but not in others. In mathematics, we are typically looking at something which has a definite answer of "true" or "false".

Definition 1.20. A PROPOSITION is a declarative sentence that is either true or false.

Example 1.21. The statement "2+2=4" is a true proposition. The statement "2+2=5" is a false proposition.

Definition 1.22. A PROOF is a logical argument that demonstrates a given proposition has a truth value of "true". A DISPROOF is a logical argument that demonstrates a given proposition has a truth value of "false".

Note that to prove anything, we must make some assumptions. Formally, what we assume in mathematics are the axioms (most mathematicians work with ZFC). Going all the way back to the axioms to prove a statement is wildly impractical. In reality, we allow ourselves to start with "common knowledge", which is not particularly well defined. Unfortunately what we exchange for practicality is ambiguity. Good mathematical writing takes a bit of practice, with time you'll become more comfortable with what is reasonable for your reader to assume.

Mathematicians typically use different language for different propositions depending on their importance or use. Here are some synonyms and when they're typically used:

> PROPOSITION: the default label for a result that stands on its own but isn't the main result of a work;
>
> THEOREM: this typically refers to an important proposition that much of your work is building up to;
>
> LEMMA: these are propositions which help to prove larger results like Theorems;
>
> COROLLARY: these are proposition which follow as a direct consequence of a previous proposition or theorem.

My opinion is, the best way to learn how to write mathematical proofs is to write mathematical proofs. If you'd like a primer on some of the logic behind proof writing, see the proof writing resources in the "Pages" tab on our Canvas page. My plan for the course will be to give you the information needed to construct a proof when it comes up.

1.4. Vector Representations

Definition 1.23. A VECTOR in $\mathbb{R}^n$ is a list of n real numbers in a specified order, which we'll write in the form

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

We call the real numbers v_i the ENTRIES or COMPONENTS of the vector $\vec{v}$.

Addition of two vectors and multiplication of a vector times as "scalar" (that is, a constant in $\mathbb{R}$) is performed component-wise. That is, if

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \text{ and } \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

are vectors in $\mathbb{R}^n$ and $c \in \mathbb{R}$ is a scalar, then we have

$$\vec{u} + \vec{v} := \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix} \text{ and } c\vec{u} := \begin{pmatrix} cu_1 \\ cu_2 \\ \vdots \\ cu_n \end{pmatrix}.$$

Example 1.24. Here is one way to solve Scenario One from Activity 1.6. Let

$$\vec{h} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \text{ and } \vec{m} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Let x denote the number of times you travel along the vector $\vec{h}$ with the hoverboard, and y the number of times you travel along the vector $\vec{m}$ with the magic carpet. Observe that it doesn't matter whether we switch between the two vehicles mid trip, or just stay on the hoverboard for its full distance and then switch to the magic carpet for the remaining time y (the paths will be different, but we will end up in the same location! can you convince yourself of this fact?). So it suffices to find real numbers x and y satisfying

$$x \begin{pmatrix} 3 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 107 \\ 64 \end{pmatrix}. \tag{1.1}$$

Adding our vectors on the left gives

$$\begin{pmatrix} 3x + y \\ x + 2y \end{pmatrix} = \begin{pmatrix} 107 \\ 64 \end{pmatrix}.$$

For these vectors to be equal, their components must also be equal. That is, we need to solve the system of linear equations

$$3x + y = 107$$

$$x + 2y = 64.$$

This system has augmented matrix

$$\left(\begin{array}{cc|c} 3 & 1 & 107 \\ 1 & 2 & 64 \end{array}\right)$$

which is row equivalent to the matrix

$$\left(\begin{array}{cc|c} 1 & 0 & 30 \\ 0 & 1 & 17 \end{array}\right).$$

That is, $x = 30, y = 17$, which tells us that **yes** we can reach Gauss's cabin using our two modes of transportation.

Equation 1.1 is an example of a LINEAR COMBINATION of the vectors $\vec{h}$ and $\vec{m}$.

Definition 1.25. A LINEAR COMBINATION of vectors $\vec{v}_1, \vec{v_2}, \ldots, \vec{v}_m$ in $\mathbb{R}^n$ is any vector of the form

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m,$$

where $c_1, c_2, \ldots, c_m$ are scalars (i.e. constants in $\mathbb{R}$) which we will refer to as the WEIGHTS.

The following result generalizes our strategy from Activity 1.6.

Theorem 1.26. *The vector equation*

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{w}$$

has the same solution set as the linear system represented by the augmented matrix

$$\left(\begin{array}{cccc|c} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m & \vec{w} \end{array}\right).$$

In particular, the system has a solution if and only if $\vec{w}$ is a linear combination of the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$.

Example 1.27. In Scenario Two of our carpet ride problem (in Activity 1.7) we asked whether there are places Gauss can move so that you cannot reach him using your two modes of transportation. Let's say Gauss is located x miles East and y miles North from your home. Then, we would like to know if the vector $\begin{pmatrix} x \\ y \end{pmatrix}$ is a linear combination of $\vec{h}$ and $\vec{m}$. By the Theorem above, this is equivalent to asking the question: for what values of x and y does the system of linear equations with augmented matrix

$$\left(\begin{array}{cc|c} 3 & 1 & x \\ 1 & 2 & y \end{array}\right)$$

have a solutions? Observe that this matrix is row equivalent to

$$\left(\begin{array}{cc|c} 1 & 0 & * \\ 0 & 1 & * \end{array}\right)$$

where $*$ denotes some real number. Since this system is consistent and each row has a pivot, we know that there is in fact a unique solution for any value of x and y. That is, there is nowhere Gauss can hide on land.

Definition 1.28. The SPAN of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ in $\mathbb{R}^n$ is the set

$$\operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m) = \{c_1\vec{v}_1 + \cdots + c_n\vec{v}_n \mid c_1, c_2, \ldots, c_m \in \mathbb{R}\}.$$

That is $\operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m)$ is the set of all linear combinations of vectors $\vec{v}_1, \ldots, \vec{v}_m$.

Using this terminology, our solution to P1 from Activity 1.7 is equivalent to saying that

$$\operatorname{Span}(\vec{h}, \vec{m}) = \mathbb{R}^2.$$

Example 1.29. Continuing our carpet ride problem, P2 of Activity 1.7 asks if we can reach Gauss's hovercabin using our two modes of transportation in flying mode. Let's say Gauss parks his cabin x miles East, y miles North, and z miles above ground. Then, we would like to know which vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ are in the span of

$$\vec{H} = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \text{ and } \vec{M} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

By our Theorem above, this is equivalent to asking the question: for what values of x, y, z does the system of linear equations with augmented matrix

$$\left(\begin{array}{cc|c} 3 & 1 & x \\ 1 & 2 & y \\ 1 & 3 & z \end{array}\right)$$

have a solution? Observe that this matrix is row equivalent to

$$\left(\begin{array}{cc|c} 1 & 0 & * \\ 0 & 1 & * \\ 0 & 0 & (!) \end{array}\right)$$

where $*$ denotes some real number. Note that we cannot guarantee this system is consistent, because it may be possible that the last entry in the last row (which we've labeled as "(!)") is nonzero. We have a few options here to find a point (x, y, z) making the system inconsistent: we can either go through our row operations carefully to get an expressions for "(!)" and find when this expression is zero. Or, we can guess-and-check. For purposes of laziness, I used the latter method to find that the matrix

$$\left(\begin{array}{cc|c} 3 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 3 & 1 \end{array}\right)$$

is row equivalent to

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)$$

which is inconsistent. So, we cannot reach Gauss's cabin if he parks it 1 mile East, 1 mile North, and 1 mile above ground.

Note that we could also argue geometrically: since the linear combination of two vectors in $\mathbb{R}^2$ forms either a line or a plane (we'll talk more about this formally later), we're going to miss "most" of the points in $\mathbb{R}^3$. So there should be plenty of space for Gauss to hide in his hovercabin.

1.5. The Matrix-Vector Form of a Linear System

Recall from Activity 1.6, to solve Scenario One from the carpet ride problem we decided that we needed to solve the vector equation

$$x\begin{pmatrix}3\\1\end{pmatrix}+y\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}107\\64\end{pmatrix}$$

Here we introduce some new notation, which in short order will lead to our next topic in the course. This new notation will treat the coefficient matrix of the corresponding linear system

$$3x+y=107$$
$$x+2y=64$$

as an object that *transforms* the vector $\begin{pmatrix}x\\y\end{pmatrix}$ into the vector $\begin{pmatrix}107\\64\end{pmatrix}$. We'll write

$$\begin{pmatrix}3&1\\1&2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}:=x\begin{pmatrix}3\\1\end{pmatrix}+y\begin{pmatrix}1\\2\end{pmatrix}.$$

In general, we have the following.

Definition 1.30. Let A be an $n\times m$ matrix and $\vec{x}$ a vector in $\mathbb{R}^m$. Write

$$A=\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1m}\\a_{21}&a_{22}&\cdots&a_{2m}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nm}\end{pmatrix}\quad\text{and }\vec{x}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_m\end{pmatrix}$$

The matrix-vector product $A\vec{x}$ is the vector in $\mathbb{R}^m$ defined by

$$A\vec{x}:=x_1\begin{pmatrix}a_{11}\\a_{21}\\\vdots\\a_{n1}\end{pmatrix}+x_2\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{n2}\end{pmatrix}+\cdots+x_m\begin{pmatrix}a_{1m}\\a_{2m}\\\vdots\\a_{nm}\end{pmatrix}.$$

Example 1.31. Let

$$A=\begin{pmatrix}1&2\\0&1\\-1&1\end{pmatrix},\vec{v}=\begin{pmatrix}2\\1\end{pmatrix}\text{ and }\vec{w}=\begin{pmatrix}0\\1\\3\end{pmatrix}.$$

Then,

$$A\vec{v}=\begin{pmatrix}1&2\\0&1\\-1&1\end{pmatrix}\begin{pmatrix}2\\1\end{pmatrix}=2\begin{pmatrix}1\\0\\-1\end{pmatrix}+1\cdot\begin{pmatrix}2\\1\\1\end{pmatrix}=\begin{pmatrix}4\\1\\-1\end{pmatrix}.$$

Note that $A\vec{w}$ is undefined, since $\vec{w}\in\mathbb{R}^3$ but A only has two columns. In fact, the vector product $A\vec{u}$ is **only** defined for vectors $\vec{u}$ in $\mathbb{R}^2$.

Example 1.32. In Activity 1.8, we looked at the matrix-vector equation

$$\begin{pmatrix}1&2&-1\\2&1&3\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}2\\6\end{pmatrix}.$$

Observe that finding a vector solution to this matrix equation is equivalent to finding a solution x, y, z to the system of linear equations

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

$$2x + y + z = 6.$$

This linear system has augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 2 \\ 2 & 1 & 1 & 6 \end{array}\right)$$

which is row equivalent to the matrix in reduced row echelon form

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 10/3 \\ 0 & 1 & -1 & -2/3 \end{array}\right).$$

So all solutions to our system of linear equations above can be written as

$$(x, y, z) = (10/3 - t, -2/3 + t, t).$$

Finding solutions to this system of linear equations is also equivalent to solving the vector equation

$$x \begin{pmatrix} 1 \\ 2 \end{pmatrix} + y \begin{pmatrix} 2 \\ 1 \end{pmatrix} + z \begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}.$$

So, any solution (x, y, z) to our linear system should also satify the vector equation above. For example, if we choose the solution $(x, y, z) = (10/3, -2/3, 0)$ this shows that

$$\begin{pmatrix} 2 \\ 6 \end{pmatrix} = \frac{10}{3} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

That is, the vector $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$ is in the span of the vectors

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \text{ and } \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

Observe that we now have multiple ways to represent and study our original problem of solving systems of linear equations. We should aim to become comfortable with all of the equivalent representations below, as we can learn something using each of these perspectives (as demonstrated in the previous example). The following representations are all equivalent.

Representation 1: Systems of Linear Equations

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m = b_2$$

$$\vdots$$

$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = b_n.$$

Why? This representation is familiar, and can model situations from different disciplines we may need to solve.

Representation 2: Augmented matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1m} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2m} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} & b_m \end{array}\right)$$

Why? This representation is convenient for computation. We've developed theory to solve systems of linear equations algorithmically by manipulating the corresponding augmented matrix.

Representation 3: Vector equation

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Why? Vector equations arise as another intuitive problem (recall our magic carpet activities from last week). Being able to understand vector equations as equivalent to systems of linear equations helped us solve vector equation problems. In turn, understanding systems of linear equations as vector equations can help us understand possible solutions to linear equations geometrically (as the span of a set of vectors).

Representation 4: Matrix-vector equation

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Why? In the following section, we'll see how matrices can be interpreted as "linear transformations". The representation above shows us that this matrix-vector operation sends a vector in $\mathbb{R}^m$ to another vector in $\mathbb{R}^m$.

Remark 1.33. The remaining problems in Activity 1.8 asked you to move between these perspectives to see how the concepts of solving matrix-vector equations, showing an element is in the span of some set of vectors, and solving systems of linear equations, are all the same. My hope is that you came up with some parts of the following result.

Theorem 1.34. *Let A be an $n \times m$ matrix. The following statements are equivalent.*

(1) The matrix-vector equation $A\vec{x} = \vec{b}$ has a solution for every vector $\vec{b}$ in $\mathbb{R}^n$;

(2) The system of equations with augmented matrix

$$\left(A \mid \vec{b}\right)$$

is consistent for any vector $\vec{b}$ in $\mathbb{R}^n$.

(3) The reduced row echelon form of the matrix A has a pivot in each row;

(4) The span of the columns of A is equal to $\mathbb{R}^n$;

Note: When we say the statements above are "equivalent", we mean that statement (i) is true if and only if statement (j) is true, for all combinations of $i, j \in \{1, 2, 3, 4\}$. More intuitively, we should think of this theorem as saying: if you want to show a matrix A satisfies any one of the properties itemized above, you can instead show that it satisfies any of the other properties itemized above. Let's look at a quick example before sketching this proof.

Example 1.35. Recall again the matrix-vector equation from Activity 1.8

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}.$$

Since the matrix

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

is row equivalent to

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

and this matrix has a pivot in every row, we know that the matrix-vector equation

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

has a solution for any vector $\vec{b}$ in $\mathbb{R}^n$. We also know that

$$\text{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix}\right) = \mathbb{R}^2$$

and that the system of linear equations

$$x + 2y - z = b_1$$

$$2x + y + 3z = b_2$$

has a solution for any $b_1, b_2 \in \mathbb{R}$.

Interlude: Conditional Statements

The most common type of proposition we'll see in the class is the CONDITIONAL STATEMENT, which takes the form

"if blah then blah".

For example, "if x is even, then x^2 is even" is a true conditional statement, and "if $x = 1$ then $x + 1 = 5$" is a false conditional statement. To prove a conditional statement DIRECTLY we do the following:

(1) Assume the hypothesis is true

(2) Use some direct logical reasoning

(3) Deduce that the conclusion is also true.

Hopefully this proof method feels intuitively true, but if you'd like a breakdown for why this is logically sound, visit the Proof Writing Guides page in the "Pages" tab on Canvas. Let's look at an example.

Proposition 1.36. If x is even, then x^2 is even.

To prove this proposition, it's important that we recall the formal definition of an even integer.

Definition 1.37. An integer x is EVEN if there exists an integer k so that $x = 2k$.

We now prove our proposition directly.

Proof. Suppose that x is even *(Here we assume the hypothesis is true)*. Then, there exists an integer k so that

$$x = 2k$$

(Here we appealed to the formal definition of even integers). So we have

$$x^2 = (2k)^2 = 2(2k^2).$$

Letting $\ell = 2k^2$ we see that

$$x^2 = 2\ell,$$

for $\ell \in \mathbb{Z}$. So, x^2 is even. *(Using our definition of an even integer, we deduce the conclusion is true by using direct logical reasoning.).* $\square$

We will also see the BICONDITIONAL STATEMENT, which takes the form

"blah if and only if BLAH"

and means "if blah then BLAH **and** if BLAH then blah". To prove a biconditional statement, we need to prove that both conditional statements are true. For example, if we wanted to prove the proposition:

"x is even if and only if x^2 is even"

we would need to prove the two conditional statements are true:

(1) "if x is even then x^2 is even" **and**

(2) "if x^2 is even then x is even".

1.5 Continued

We now have the tools to prove our result. Here's my writeup of this proof without the scaffolding from Activity 1.10.

Proof of Theorem 1.34. Note that (1) $\Leftrightarrow$ (2) follows directly from translating our matrix-vector product in Representation 4 into an augmented matrix as in Representation 2. Next, by Theorem 1.19 we know that the system represented by the augmented matrix

$$\left(A \mid \vec{b}\right)$$

is consistent if and only if there is no pivot in the last column. If A has a pivot in every row already, then by definition there cannot be a pivot in the last column of $\left(A \mid \vec{b}\right)$. If A does not have a pivot in the kth row, then letting $\vec{b}$ be any vector with a nonzero kth entry would make our system inconsistent. This proves (2) $\Leftrightarrow$ (3). Finally, we show (4) $\Leftrightarrow$ (1). Recall that solutions

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

to the matrix-vector equation $A\vec{x} = \vec{b}$ (as in Representation 4) are identical to the solutions $x_1, x_2, \ldots, x_m$ to the vector equation

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix} = \vec{b}$$

(as in Representation 3). So, the matrix-vector equation $A\vec{x} = \vec{b}$ having a solution for every $\vec{b} \in \mathbb{R}^n$ is equivalent to every vector $\vec{b}$ in $\mathbb{R}^n$ being in the span of the columns of A. Observe that our proof is complete, since we've shown (4) $\Leftrightarrow$ (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3), and so all biconditional statements follow by transitivity.

□

1.6. Linear Independence and Bases

In Activity 1.11, we looked at the span of vectors

$$\vec{v}_1 = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \text{ and } \vec{v_3} = \begin{pmatrix} 1 \\ -1 \\ -6 \end{pmatrix}.$$

Note that we can write

$$\vec{v}_3 = 2\vec{v}_1 - 3\vec{v}_2.$$

So, if we take any $\vec{c} \in \mathrm{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ then we could write

$$\begin{aligned} \vec{c} &= a\vec{v}_1 + b\vec{v}_2 + c\vec{v}_3 \\ &= a\vec{v}_1 + b\vec{v}_2 + c(2\vec{v}_1 - 3\vec{v}_2) \\ &= (a + 2c)\vec{v}_1 + (b - 3c)\vec{v}_2. \end{aligned}$$

So, $\vec{c} \in \mathrm{Span}(\vec{v}_1, \vec{v}_2)$. Since every element of $\mathrm{Span}(\vec{v}_1, \vec{v}_2)$ is in $\mathrm{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$, we get the equality

$$\mathrm{Span}(\vec{v}_1, \vec{v}_2) = \mathrm{Span}(\vec{v}_1, \vec{v}_2).$$

So, it was redundant to include the vector $\vec{v}_3$ in our definition. We next build some machinery so that we don't waste our time describing spanning sets with redundant vectors.

Definition 1.38. A set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ in $\mathbb{R}^n$ is called LINEARLY DEPENDENT if there exists constants $c_1, \ldots, c_m \in \mathbb{R}$ not all equal to zero so that

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

Otherwise, the set S is called LINEARLY INDEPENDENT. That is, the set of vectors $\{\vec{v}_1, \ldots, \vec{v}_m\}$ are linearly independent if the only solution to

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

is $c_1 = c_2 = \cdots = c_m = 0$. Note that sometimes we'll say the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ are linearly (in)dependent if the set of vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ is.

Example 1.39. Since

$$2\vec{v}_1 + (-3)\vec{v}_2 + \vec{v}_3 = \vec{0},$$

we see that the set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly dependent. To see that the set $\{\vec{v}_1, \vec{v}_2\}$ is linearly independent, observe that finding solutions to

$$x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0} \tag{1.2}$$

is equivalent to solving the system of linear equations with augmented matrix

$$\left(\begin{array}{cc|c} 2 & 1 & 0 \\ 1 & 1 & 0 \\ -3 & 0 & 0 \end{array}\right)$$

which is row equivalent to

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

So, the vector equation (1.2) only has the solution $x_1 = x_2 = 0$. Hence, $\vec{v}_1, \vec{v}_2$ are linearly dependent.

The following theorem tells us that the notion of linear dependency precisely measures whether we've included redundant vectors in our definition of a spanning set. Our proof will work similarly to the examples we saw in Activity 1.11.

Theorem 1.40. *A set of vectors $\{\vec{v}_1, \ldots, \vec{v}_m\}$ in $\mathbb{R}^n$ is linearly dependent if and only if at least one of the vectors in the set can be written as a linear combination of the others.*

Proof. Suppose that $\vec{v}_1, \ldots, \vec{v}_m$ are linearly dependent. Then, there exists nonzero $c_1, \ldots, c_m \in \mathbb{R}$ not all equal to zero so that

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

Without loss of generality, assume that $c_1 \neq 0$ (that is, just relabel everything so you've listed the first vector as the one having a nonzero coefficient in the equation above). Then, we can write at least one of our constants c_i is nonzero, which means we can write

$$\vec{v}_1 = \frac{c_2}{c_1}\vec{v}_2 + \cdots + \frac{c_m}{c_1}\vec{v}_m,$$

making $\vec{v}_1$ a linear combination of $\vec{v}_2, \ldots, \vec{v}_m$. Conversely, suppose that $\vec{v}_i$ is a linear combination of the other vectors. Again, for convenience, we relabel so that $\vec{v}_1$ is a linear combination of $\vec{v}_2, \ldots, \vec{v}_m$. Then, there exist constants $a_2, \ldots, a_m \in \mathbb{R}$ so that

$$\begin{gathered} \vec{v}_1 = a_2\vec{v}_2 + \cdots + a_m\vec{v}_m \\ \Rightarrow (-1)\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_m\vec{v}_m = \vec{0}. \end{gathered}$$

Since -1 is nonzero, this shows our vectors $\vec{v}_1, \ldots, \vec{v}_m$ are linearly dependent. $\square$

The following result should generalize your conjectures from Activity 1.12.

Theorem 1.41. *A set $S = \{\vec{v}_1, \ldots, \vec{v}_m\}$ of vectors in $\mathbb{R}^n$ is linearly independent if and only if then there is only one way to write $\vec{b} \in Span(S)$ as a linear combination of the vectors $\vec{v}_1, \ldots, \vec{v}_m$.*

Proof. Suppose that the set S is linearly independent, and that we could write

$$\vec{b} = a_1\vec{v}_1 + \cdots + a_m\vec{v}_m$$

and

$$\vec{b} = b_1\vec{v}_1 + \cdots + b_m\vec{v}_m.$$

Subtracting these two equations gives

$$\vec{0} = (a_1 - b_1)\vec{v}_1 + \cdots + (a_m - b_m)\vec{v}_m.$$

But since our vectors $\vec{v}_1, \ldots, \vec{v}_m$ are linearly independent, we must have

$$a_i - b_i = 0 \Rightarrow a_i = b_i$$

for all i. So, our two representations of $\vec{b}$ as a linear combination of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ are the same.

Conversely, if S is a linearly dependent set, then there is a nonzero solution to the equation

$$x_1\vec{v}_1 + \cdots + x_m\vec{v}_m = \vec{0}.$$

This means there must be infinitely many solutions to the matrix-vector equation above, and so the matrix A with columns given by $\vec{v}_i$ has a column without a pivot. □

Theorems 1.40 and 1.41 tell us two things: (1) if we describe a span by a linearly independent set we don't have any redundancy, and (2) if we describe a span by a linearly independent set, every element in that span can be written uniquely. This means, sets of linearly independent vectors are the "best" way to describe spans $\mathbb{R}^n$. Let's give a name to these nice sets.

Definition 1.42. Let S be a set of vectors in $\mathbb{R}^n$ and $V = \text{Span}(S)$. A subset B of V is a BASIS for V if B is linearly independent and $\text{Span}(B) = V$.

On your next homework, you'll show the following for vectors in $\mathbb{R}^3$. Note that the proof holds identically for the more general statement below.

Theorem 1.43. *For vectors $\vec{v}_1, \ldots, \vec{v}_n$ in $\mathbb{R}^n$,*

$$Span(\vec{v}_1, \ldots, \vec{v}_n) = \mathbb{R}^n$$

if and only if $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a linearly independent set.

In a future section, we'll use the notions of spans and bases to define certain subsets of $\mathbb{R}^n$ called VECTOR SPACES. Before we do that, we pause this discussion and turn to a new perspective of the matrix-vector product.

1.7. Linear Transformations

Let A be an $n \times m$ matrix. Recall that the matrix-vector product $A\vec{x}$ is defined for a vector $\vec{x}$ in $\mathbb{R}^m$ and yields a vector $\vec{y} = A\vec{x}$ in $\mathbb{R}^n$. So, this product defines a function $T_A : \mathbb{R}^m \to \mathbb{R}^n$, given by

$$T_A(\vec{x}) := A\vec{x}.$$

In P4, I asked you to observe that T_A satisfy a *linearity* property. Let

$$A = \begin{pmatrix} \vec{a}_1 & \cdots & \vec{a}_m \end{pmatrix}$$

where $\vec{a}_i$ are the columns of an $n \times m$ matrix A, which we recall can be thought of as vectors in $\mathbb{R}^n$. Then, for any vectors

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \text{ and } \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

' in $\mathbb{R}^m$ and constants $c, d \in \mathbb{R}$, we have

$$\begin{aligned} T_A(c\vec{x} + d\vec{y}) &= A(c\vec{x} + d\vec{y}) \\ &= (cx_1 + dy_1)\vec{a}_1 + (cx_2 + dy_2)\vec{a}_2 + \cdots + (cx_m + dy_m)\vec{a}_m \\ &= c(x_1\vec{a}_1 + \cdots + x_m\vec{a}_m) + d(y_1\vec{a}_1 + \cdots + y_m\vec{a}_m) \\ &= cA\vec{x} + dA\vec{y} \\ &= cT_A(\vec{x}) + dT_A(\vec{y}). \end{aligned}$$

That is, T_A distributes over addition and scalar multiplication. Functions satisfying this property make up a special family.

Definition 1.44. A function $F : \mathbb{R}^m \to \mathbb{R}^n$ is called LINEAR if the following property holds

$$F(c\vec{x} + d\vec{y}) = cF(\vec{x}) + dF(\vec{y})$$

for all $c, d \in \mathbb{R}$ and vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^m$.

Note that sometimes people call functions TRANSFORMATIONS, and so oftentimes we'll refer to functions with the property above as LINEAR TRANSFORMATIONS.

Example 1.45. The function $F : \mathbb{R}^2 \to \mathbb{R}^2, \vec{x} \mapsto 2\vec{x}$ is linear (can you show this?), but the function

$$G : \mathbb{R}^2 \to \mathbb{R}^2, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^2 \\ y^2 \end{pmatrix}$$

is not linear, since for example

$$G(4) = \begin{pmatrix} 16 \\ 16 \end{pmatrix}$$

but

$$G(2) + G(2) = \begin{pmatrix} 8 \\ 8 \end{pmatrix}$$

so $G(4) \neq G(2) + G(2)$.

From our discussion above, every function T_A defined by an $n \times m$ matrix is a linear transformation. We'll see next that the converse also holds.

Theorem 1.46. *Let $F : \mathbb{R}^m \to \mathbb{R}^n$ be a linear function. Then, there exists an $n \times m$ matrix A so that $F = T_A$.*

First, we need a definition. In P2 and P3 of Activity 1.13, you should have been able to find the matrix A given T_A defined on some basis elements. These special basis elements will be given a name.

Definition 1.47. Let

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \dots, \vec{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

The set $\{e_1, e_2, \dots, e_n\}$ is called the STANDARD BASIS for $\mathbb{R}^n$.

Note that in fact this is a basis for $\mathbb{R}^n$, because it is a linearly independent spanning set. We are now ready to prove our result.

Proof. Suppose that $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear function. Define the $n \times m$ matrix A to be the matrix with column vectors equal to $F(\vec{e}_i)$ for $i = 1, \dots, n$. We claim that $F = T_A$. We have

$$F(\vec{e}_i) = T_A(\vec{e}_i) \tag{1.3}$$

for every $i \in \{1, \dots, m\}$. Now, take any $\vec{x} \in \mathbb{R}^m$. Since $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_m\}$ form a basis for $\mathbb{R}^m$ we can write $\vec{x}$ uniquely as

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_m\vec{e}_m.$$

So, we have

$$\begin{aligned}
F(\vec{x}) &= F(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_m\vec{e}_m) \\
&= x_1F(\vec{e}_1) + x_2F(\vec{e}_2) + \cdots + x_mF(\vec{e}_m) && \text{by linearity of } F \\
&= x_1T_A(\vec{e}_1) + x_2T_A(\vec{e}_2) + \cdots + x_mT_A(\vec{e}_m) && \text{by (1.3)} \\
&= T_A(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_m\vec{e}_m) && \text{by linearity of } T_A \\
&= T_A(\vec{x}).
\end{aligned}$$

So, we have $F(\vec{x}) = T_A(\vec{x})$ for every vector $\vec{x}$ in $\mathbb{R}^m$, which means $F = T_A$ as functions. □

The previous results allow us to define the following.

Definition 1.48. Given a matrix A, we call T_A the MATRIX TRANSFORMATION corresponding to the matrix A. Given a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$, we call the $n \times m$ matrix A_T constructed in Theorem 1.46 the DEFINING MATRIX of the transformation T.

Next, we see how what the theory we've developed about matrices so far says about linear transformations. First, we recall some definitions.

Definition 1.49. Let $F : \mathbb{R}^m \to \mathbb{R}^n$ be a function.

(1) F is called ONE-TO-ONE (or INJECTIVE) if the following property holds: for every vector $\vec{b}$ in $\mathbb{R}^n$, there is *at most* one vector $\vec{x}$ in $\mathbb{R}^m$ so that $F(\vec{x}) = \vec{b}$. We often use the arrow $F : \mathbb{R}^m \hookrightarrow \mathbb{R}^n$ to indicate when a function is injective.

(2) F is called ONTO (or SURJECTIVE) if the following property holds: for every vector $\vec{b}$ in $\mathbb{R}^n$ there is *at least* one vector $\vec{x}$ in $\mathbb{R}^m$ so that $F(\vec{x}) = \mathbb{R}^n$. We often use the arrow $F : \mathbb{R}^m \twoheadrightarrow \mathbb{R}^n$ to indicate when a function is surjective.

Example 1.50. This may be familiar from your calculus courses for functions between one-dimensional Euclidean space. For example, the function

$$f : \mathbb{R} \to \mathbb{R}, x \mapsto 2x + 1$$

is one-to-one and onto. The function

$$g : \mathbb{R} \to \mathbb{R}, x \mapsto x^2$$

is neither one-to-one nor onto.

In Activity 1.14, you investigated what we can say about matrix transformations (and hence linear transfromations) by looking at the shape of the corresponding matrix. My hope is that you were able to come up with some version of the following.

Theorem 1.51. *Let A be an $n \times m$ matrix. Then*

(1) T_A is injective if and only if every column in the reduced row echelon form of A has a pivot.

(2) T_A is surjective if and only if every row in the reduced row echelon form of A has a pivot.

Proof. This follows directly from Theorem 1.19. $\square$

We introduce one more definition

Definition 1.52. Given a function $F : \mathbb{R}^m \to \mathbb{R}^n$, the KERNEL of F is the subset of $\mathbb{R}^m$ given by

$$\ker(F) := \{\vec{x} \in \mathbb{R}^m \mid F(\vec{x}) = \vec{0}\}.$$

Example 1.53. In Activity 1.15, you looked at $\ker(T_A)$ where

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & 8 \\ 2 & 1 & 12 \end{pmatrix}.$$

Observe that A is row equivalent to

$$\begin{pmatrix} 1 & 0 & 10 \\ 0 & 1 & -8 \\ 0 & 0 & 0 \end{pmatrix},$$

and so all vectors $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \ker(T_A)$ satisfy

$$x = -10z, y = 8z.$$

That is,

$$\ker(T_A) = \text{Span}\left(\begin{pmatrix} -10 \\ 8 \\ 1 \end{pmatrix}\right).$$

In the next chapter, we'll explore what $\ker(T_A)$ can look like in general, but for now let's observe that T_A is not one-to-one since there is more than one vector being mapped to $\vec{0}$.

In fact, the kernel precisely determines when a linear function is one-do one. We have the following.

Theorem 1.54. *For a linear function $F : \mathbb{R}^m \to \mathbb{R}^n$, $\ker(F) = \{\vec{0}\}$ if and only if F is one-to-one.*

Proof. Recall, this is a *biconditional* statement, so we need to prove both implications. Let A be the defining matrix of F, so that $F = T_A$ (which we showed exists in Theorem 1.46).

Suppose first that $\ker(F) = \{\vec{0}\}$ and that $F(\vec{x}) = F(\vec{y})$ for some vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^m$. Then we have

$$\begin{aligned} F(\vec{x}) &= F(\vec{y}) \\ \Rightarrow A\vec{x} &= A\vec{y} \\ \Rightarrow A(\vec{x} - \vec{y}) &= \vec{0}. \end{aligned}$$

So, $\vec{x} - \vec{y} \in \ker(F) = \{\vec{0}\}$ which gives $\vec{x} = \vec{y}$.

Conversely, suppose that F is one-to-one. By definition, we have

$$F(\vec{0}) = A\vec{0} = \vec{0}$$

so $\vec{0} \in \ker(F)$. Since F is one-to-one, there cannot be another vector $\vec{x}$ satisfying $F(\vec{x}) = \vec{0}$. So $\ker(F) = \{\vec{0}\}$. □

Chapter 2

Matrix Operations

2.8. The Matrix Product

Given sets A, B, C, D and functions $f : A \to B$ and $g : B \to C$, recall that the composite function $f \circ g : A \to C$ is defined by

$$(f \circ g)(a) = f(g(a)).$$

Let's look at how function composition behaves on linear transformations. Given an $n \times k$ matrix A and an $\ell \times m$ matrix B, we have the associated linear transformations

$$T_A : \mathbb{R}^k \to \mathbb{R}^n, \vec{x} \mapsto A\vec{x}$$

$$T_B : \mathbb{R}^m \to \mathbb{R}^\ell, \vec{x} \mapsto B\vec{x}.$$

So, for the composition $T_A \circ T_B$ to be defined, we need $\mathbb{R}^k = \mathbb{R}^\ell$. That is, $k = \ell$. In this case, we have $T_A \circ T_B : \mathbb{R}^m \to \mathbb{R}^n$ defined by

$$(T_A \circ T_B)(\vec{x}) = A(B\vec{x}). \tag{2.1}$$

In Activity 1.16, you showed that the composition of two linear functions is linear. So, by Theorem 1.46, there exists an $n \times m$ matrix C so that

$$T_A \circ T_B = T_C.$$

Also in Activity 1.16, you found the matrix C for various examples of A and B in order for equation (2.1) to be satisfied. Let's derive what the matrix C looks like in general, and learn some tools to speed up this computation. Write

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

and suppose B has column vectors

$$\vec{b}_1, \vec{b}_2, \dots, \vec{b}_m.$$

Then we have

$$\begin{aligned} A(B\vec{x}) &= A\left(x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_m\vec{b}_m\right) \\ &= x_1A\vec{b}_1 + x_2A\vec{b}_2 + \cdots + x_mA\vec{b}_m \\ &= C\vec{x}, \end{aligned}$$

where the second equality follows by linearity of T_A. So, if we let C be the matrix with column vectors

$$A\vec{b}_1, A\vec{b}_2, \ldots, A\vec{b}_m$$

then we have $T_A \circ T_B = T_C$. We'll call the matrix C the MATRIX PRODUCT of A and B. Let's write this definition down formally.

Definition 2.1. Let A by an $n \times k$ matrix and B be a $k \times m$ matrix. Write

$$B = \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_m \end{pmatrix}$$

where $\vec{b}_i$ are vectors in $\mathbb{R}^k$. Then, the MATRIX PRODUCT AB is the $n \times m$ matrix defined by

$$AB := \begin{pmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_m \end{pmatrix}.$$

From our discussion above, note that AB satisfies $T_A \circ T_B = T_{AB}$.

Let's look at some examples.

Example 2.2. Let

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & -1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 2 \\ 0 & 3 \\ -1 & 1 \end{pmatrix}$$

as in P2 of Activity 1.16. Since A is 2×3 and B is 3×2, the matrix product AB is a 2×2 matrix. We have

$$AB = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 \end{pmatrix}$$

where

$$\vec{a}_1 = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \text{ and } \vec{a}_2 = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

So,

$$AB = \begin{pmatrix} -2 & 5 \\ 2 & 4 \end{pmatrix},$$

which should match what you found in P1(c). With practice, we can perform this computation a bit more quickly. Let's perform the steps above by just keeping track of how we're generating each entry:

$$\begin{pmatrix} \color{red}{1} & \color{red}{0} & \color{red}{3} \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \color{red}{1} & 2 \\ \color{red}{0} & 3 \\ \color{red}{-1} & 1 \end{pmatrix} = \begin{pmatrix} \color{red}{-2} & \\ & \end{pmatrix}$$

$$\begin{pmatrix} \color{red}{1} & \color{red}{0} & \color{red}{3} \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & \color{red}{2} \\ 0 & \color{red}{3} \\ -1 & \color{red}{1} \end{pmatrix} = \begin{pmatrix} -2 & \color{red}{5} \\ & \end{pmatrix}$$

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

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

Remark 2.3. Note that if A is an $m \times k$ matrix and B is $\ell \times m$ matrix, the matrix product AB is only defined when $k = \ell$. So, in the example above, the matrix BA also happens to be defined and gives a 3×3 matrix.

Let's practice a few more examples.

Example 2.4. Let

$$A = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 2 & 3 & 1 \end{pmatrix}, B = \begin{pmatrix} 0 & -1 \\ 3 & 0 \\ 1 & 0 \\ -1 & 4 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & -2 & 1 \end{pmatrix}.$$

Then,

$$AB = \begin{pmatrix} 2 & -5 \\ 8 & 4 \end{pmatrix}, BA = \begin{pmatrix} 0 & -2 & -3 & -1 \\ 3 & 0 & 3 & -3 \\ 1 & 0 & 1 & -1 \\ -1 & 8 & 11 & 5 \end{pmatrix}, \text{ and } CB = \begin{pmatrix} 5 & -1 \\ 6 & 4 \\ -3 & 4 \end{pmatrix}.$$

Note that the matrix products $BC, AC,$ and CA are undefined, because they do not have compatible dimensions.

2.10. The Inverse of a Matrix

In Activity 1.17, you were introduced to the following definitions.

Definition 2.5. The IDENTITY MATRIX I_n is the $n \times n$ matrix with 1s on the diagonal, and zeros everywhere else. For example,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \text{ and } I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We call I_n the identity matrix because it "does nothing" under matrix multiplication. That is, $AI_n = A$ for any $m \times n$ matrix A. This should feel similar to how 1 is the identity element with respect to multiplication: for any nonzero $a \in \mathbb{R}$ we have $a \cdot 1 = 1$.

Definition 2.6. Let A be an $n \times n$ matrix. Then INVERSE MATRIX OF A, if it exists, is the $n \times n$ matrix B satisfying

$$AB = BA = I_n$$

If such a matrix B exists, we say that the matrix A is INVERTIBLE, and we write $B = A^{-1}$.

Note that if $B = A^{-1}$ then we have

$$(T_A \circ T_B)(\vec{x}) = AB\vec{x} = I_n\vec{x} = \vec{x},$$

$$(T_B \circ T_A)(\vec{x}) = BA\vec{x} = I_n\vec{x} = \vec{x}$$

and so T_B is the inverse transformation of T_A; that is,

$$T_{A^{-1}} = T_A^{-1}.$$

Observe, as you did in P2 of Activity 1.17, that the converse also holds; that is, the defining matrix of the inverse of a linear transformation is the inverse of the defining matrix of the linear transformation. So, finding matrix inverses is the same problem as finding inverses of linear transformations. In this section, we look at one tool to compute the inverse of a matrix, and determine when a matrix is invertible.

Example 2.7. Let

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

Observe that if A^{-1} exists, it must satisfy

$$AA^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so A^{-1} must have column vectors $\vec{b}_1, \vec{b}_2, \vec{b}_3$ satisfying the matrix-vector equations

$$A\vec{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, A\vec{b}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \text{ and } A\vec{b}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

We can use row reduction to solve the first matrix-vector equation, as below

$$\begin{aligned} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right) &\sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2 \\ &\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1 \\ &\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3 \end{aligned}$$

This gives

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Observe that we can use exactly the same row operations to solve for $\vec{b}_2$ and $\vec{b}_3$:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2$$

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1$$

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3$$

so that $\vec{b}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ and similarly we compute

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2$$

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1$$

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3$$

so that $\vec{b}_3 = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$, which gives

$$A^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

Note that, since we performed the same row operations for every matrix-vector equation, we could performed the same operations as above by instead looking at the augmented matrix $\big(A \mid I_3\big)$, as follows

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & -1 & 2 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3$$

and observe that the matrix on the right is A^{-1}. The following Theorem tells us that in fact this always works as a general strategy.

Theorem 2.8. *Let A be an $n \times n$ matrix. If $\begin{pmatrix} A \mid I_n \end{pmatrix}$ is row equivalent to $\begin{pmatrix} I_n \mid B \end{pmatrix}$ for an $n \times n$ matrix B, then then A is invertible with $A^{-1} = B$.*

Proof. Suppose that $\begin{pmatrix} A \mid I_n \end{pmatrix}$ is row equivalent to $\begin{pmatrix} I_n \mid B \end{pmatrix}$ for an $n \times n$ matrix B, and and write

$$B = \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{pmatrix}$$

Let $\vec{u}_i$ be the vector in $\mathbb{R}^n$ with 1 in the ith component and 0s everywhere else. From above, we have that $\begin{pmatrix} A \mid \vec{u}_i \end{pmatrix}$ is row equivalent to $\begin{pmatrix} I_n \mid \vec{b}_i \end{pmatrix}$, and so $\vec{b}_i$ is a solution to the matrix-vector equation $A\vec{x} = \vec{u}_i$. This gives $AB = I_n$. Next, observe that $\begin{pmatrix} A \mid I_n \end{pmatrix}$ being row equivalent to $\begin{pmatrix} I_n \mid B \end{pmatrix}$ implies that $\begin{pmatrix} B \mid I_n \end{pmatrix}$ is row equivalent to $\begin{pmatrix} I_n \mid A \end{pmatrix}$ (this is not immediate, you may try convincing yourself of this with some examples). Following the same reasoning as above gives $BA = I_n$ and so $B = A^{-1}$. $\square$

If you and your neighbor found different answers from A^{-1} in P3 of Activity 1.17, there must have been a computational error. We have the following.

Proposition 2.9. If a matrix A is invertible, then its inverse A^{-1} is unique.

Proof. Suppose that B and C are two matrices satisfying

$$AB = BA = I_n$$

$$AC = CA = I_n.$$

This gives

$$B = BI_n = B(AC) = (BA)C = I_nC = C$$

and so $B = C$. $\square$

We can use this process to find a general formula for the inverse of a 2×2 matrix.

Proposition 2.10. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

If $ad - bc \neq 0$ then A is invertible with

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

Proof. We have

$$\begin{aligned}
&\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \\
&\sim \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & \dfrac{ad-bc}{a} & -\dfrac{c}{a} & 1 \end{array}\right) && R_2 - \frac{c}{a}R_1 \\
&\sim \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & 1 & -\dfrac{c}{ad-bc} & \dfrac{a}{ad-bc} \end{array}\right) && \frac{a}{ad-bc}R_2 \\
&\sim \left(\begin{array}{cc|cc} a & 0 & \dfrac{ad}{ad-bc} & \dfrac{-ab}{ad-bc} \\ 0 & 1 & \dfrac{-c}{ad-bc} & \dfrac{a}{ad-bc} \end{array}\right) && R_1 - bR_2 \\
&\sim \left(\begin{array}{cc|cc} 1 & 0 & \dfrac{d}{ad-bc} & \dfrac{-b}{ad-bc} \\ 0 & 1 & \dfrac{-c}{ad-bc} & \dfrac{a}{ad-bc} \end{array}\right) && \frac{1}{a}R_1
\end{aligned}$$

□

Next class, we'll show that the converse of Theorem 2.8 also holds, and look at some equivalent conditions to help us determine when a matrix is invertible.

2.11. The Invertible Matrix Theorem

In Theorem 2.8 we saw that an invertible $n \times n$ matrix A must be equivalent to the identity matrix I_n. The converse also turns out to be true. The main theorem in this section gives equivalent conditions to determine when an $n \times n$ matrix is invertible. We first need a definition.

Definition 2.11. An $n \times n$ matrix is called ELEMENTARY if it can be obtained by performing exactly one row operation to the identity matrix.

Since we have three elementary row operations, there should be three types of elementary matrices. In Activity 1.18, you looked at the following:

Row-switching matrices: let S_{ij} be the matrix which is obtained by swapping the ith and jth rows of the identity matrix. For example, for 3×3 matrices we have

$$S_{1,3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Observe that multiplying a matrix on the left by $T_{1,3}$ swaps the first and third row:

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{pmatrix}.$$

In general, $\boxed{\text{multiplying a matrix on the left by } T_{i,j} \text{ swaps the } i\text{th and } j\text{th rows}}$.

Row-multiplying matrix: Let $M_i(c)$ be the matrix which is obtained by multiplying the ith row of the identity matrix by a constant c. For example, for 3×3 matrices we have

$$M_2(5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Observe that multiplying a matrix on the left by $M_2(5)$ multiplies the 2nd row of that matrix by 5:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 5a_{21} & 5a_{22} & 5a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

In general, $\boxed{\text{multiplying a matrix on the left by } M_i(c) \text{ multiplies the } i\text{th row of that matrix by c}}$.

Row-addition matrix: Let $A_{i,j}(c)$ be the matrix with 1's on the diagonal, c in the (i, j) entry, and zeros everywhere else. That is, $A_{i,j}(c)$ is the matrix which is obtained by adding c times the jth row to the ith row of the identity matrix. For example, for 3×3 matrices we have

$$A_{1,2}(5) = \begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Observe that multiplying a matrix on the left by $A_{1,2}(5)$ adds 5 times the 2nd row to the first row:

$$\begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} + 5a_{21} & a_{12} + 5a_{22} & a_{13} + 5a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

In general, $\boxed{\text{multiplying a matrix on the left by } A_{i,j}(c) \text{ adds c times the } j\text{th row to the ith row}}$.

Observation 2.12. Note that every elementary matrix is invertible. Indeed, we have

$$S_{ij}S_{ij} = I_n$$

$$M_i(c)M_i(1/c) = M_i(1/c)M_i(c) = I_n$$

$$A_{i,j}(c)A_{i,j}(-c) = A_{i,j}(-c)A_{i,j}(c) = I_n$$

Theorem 2.13 (The Invertible Matrix Theorem)**.** *Let A be an $n \times n$ matrix. The following are equivalent:*

(1) A is invertible;

(2) The matrix-vector equation $A\vec{x} = \vec{b}$ has a unique solution for any vector $\vec{b}$ in $\mathbb{R}^n$;

(3) The reduced row echelon form of A is I_n;

(4) A is a product of elementary matrices.

Note that further equivalent conditions also hold (see Theorem 11.1 in your text), but my opinion is that we should only remember the conditions that take some work to prove as part of this theorem. The rest we can easily figure out when we need them. We need one lemma before we prove our main result.

Lemma 2.14. *For $n \times n$ matrices A, B we have*

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

Proof. In Activity 1.18 you saw that matrices do not necessarily commute. So, the order we write the product of our inverses in is important! We have

$$(AB)(B^{-1}A^{-1}) = AI_nA^{-1} = AA^{-1} = I_n$$
$$(B^{-1}A^{-1})(AB) = B^{-1}I_nB = B^{-1}B = I_n$$

so $(AB)^{-1} = B^{-1}A^{-1}$. □

We are now prepared to prove the Invertible Matrix Theorem.

Proof of 2.13. We'll prove (1) ⇒ (2) ⇒ (3) ⇒ (4) ⇒ (1).

(1) ⇒ (2): Suppose that A is invertible. Observe that

$$A(A^{-1}\vec{b}) = (AA^{-1})\vec{b} = I_n\vec{b} = \vec{b},$$

and so $A^{-1}\vec{b}$ is a solution to the matrix-vector equation $A\vec{x} = \vec{b}$. To see this solution is *unique*, suppose that we have another solution $\vec{y}$. Then,

$$A\vec{y} = \vec{b} \text{ and } A(A^{-1}\vec{b}) = \vec{b}$$
$$\Rightarrow A\vec{y} = A(A^{-1}\vec{b}).$$

Multiplying the above equation on both sides by A^{-1} gives

$$A^{-1}A\vec{y} = A^{-1}A(A^{-1}\vec{b})$$
$$\Rightarrow \vec{y} = A^{-1}\vec{b},$$

showing uniqueness.

(2) ⇒ (3): Since the matrix-vector equation $A\vec{x} = \vec{b}$ has a unique solution, we must have a pivot in every column of the RREF of A. But since A is $n \times n$, this means we must also have a pivot in every row of the RREF of A. The only $n \times n$ matrix with a pivot in every column and row is I_n.

(3) ⇒ (4): Since A is row equivalent to I_n, there is a series of elementary row operations which transform A to I_n. This is equivalent to the equality

$$I_n = AE_1 \cdots E_\ell$$

where E_i are elementary matrices. Since elementary matrices are invertible, we get

$$A = E_1^{-1} \cdots E_\ell^{-1}$$

and so the result follows because the inverse of an elementary matrix is an elementary matrix.

(4) ⇒ (1): If $A = E_1 E_2 \cdots E_m$ for elementary matrices E_i. then by Lemma 2.14

$$A^{-1} = E_m^{-1} \cdots E_1^{-1},$$

and so A is invertible (since the inverse exists!). □

Chapter 3

Vector Spaces

3.12. The Structure of $\mathbb{R}^n$

A major task in mathematics is to take a concrete idea and generalize it to other contexts. Distilling our object of interest down to its essential parts allows us to study a broader class of concepts all at once, and often leads to a deeper understanding of our original object by placing it in a broader context. In this chapter, we'll aim to do this with one of our main objects in this course: n-dimensional Euclidean space. First, we'll look at the defining properties of $\mathbb{R}^n$. Then, we'll define any object satisfying these properties as a VECTOR SPACE. Finally, we'll look for correspondences between these new general objects and concrete example of Euclidean space.

In Section 1.4, we defined two operations on vectors in $\mathbb{R}^n$: vector addition and scalar multiplication. Observe that these operations satisfy the following properties.

Vectors $\vec{u}, \vec{v}, \vec{w}$ $\mathbb{R}^n$ satisfy

(1) **Closure under addition**: $\vec{u} + \vec{v} \in \mathbb{R}^n$;

(2) **Commutativity**: $\vec{u} + \vec{v} = \vec{v} + \vec{u}$;

(3) **Associativity**: $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$;

Furthermore, scalars $c, d \in \mathbb{R}$ and vectors $\vec{u}, \vec{v}$ in $\mathbb{R}^n$ satisfy

(4) **Closure under scalar multiplication:** $c\vec{u} \in \mathbb{R}^n$;

(5) **Compatability of scalar multiplication:** $c(d\vec{u}) = (cd)\vec{u}$

(6) **Distributivity of scalar multiplication:** $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$;

(7) **Disributivity of scalar addition:** $(c + d)\vec{u} = c\vec{u} + d\vec{u}$.

Finally, the space $\mathbb{R}^n$ contains

(8) **An additive identity**: $\vec{v} + \vec{0} = \vec{v}$ for any vector $\vec{v}$ in $\mathbb{R}^n$;

(9) **Additive inverses**: $\vec{v} + (-\vec{v}) = \vec{0}$ for any vector $\vec{v}$ in $\mathbb{R}^n$;

(10) **Scalar identity**: $1\vec{v} = \vec{v}$ for any $\vec{v}$ in $\mathbb{R}^n$.

We'll see in this chapter that these are the defining properties of Euclidean space (in a way we'll formalize rigorously later on). Any other set which satisfies these defining properties will be called a VECTOR SPACE.

To write out the most general definition of a vector space, we would need to know what a FIELD is. Since we won't be doing things quite so generally in this course, I'm going to sweep this definition under the rug. Instead, you can think of a field as one of the following sets with the usual operations of addition and multiplication: $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{Q}$.

Definition 3.1. Let $\mathbb{F}$ be a field and V a set whose elements we'll call VECTORS. Then, a VECTOR SPACE V OVER $\mathbb{F}$ is any set satisfying all of the following axioms: Vectors $\vec{u}, \vec{v}, \vec{w}$ in V satisfy

(1) **Closure under addition**: $\vec{u} + \vec{v} \in V$;

(2) **Commutativity**: $\vec{u} + \vec{v} = \vec{v} + \vec{u}$;

(3) **Associativity**: $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$;

Furthermore, scalars $c, d \in \mathbb{F}$ and vectors $\vec{u}, \vec{v}$ in V satisfy

(4) **Closure under scalar multiplication:** $c\vec{u} \in V$;

(5) **Compatability of scalar multiplication:** $c(d\vec{u}) = (cd)\vec{u}$

(6) **Distributivity of scalar multiplication:** $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$;

(7) **Disributivity of scalar addition:** $(c + d)\vec{u} = c\vec{u} + d\vec{u}$.

Finally, the space V contains

(8) **An additive identity** $\vec{0}$: for any vector $\vec{v}$ in V we have $\vec{v} + \vec{0} = \vec{v}$;

(9) **Additive inverses** $-\vec{v}$: for any vector $\vec{v}$ in V we have $\vec{v} + (-\vec{v}) = \vec{0}$;

(10) **Scalar identity**: for any $\vec{v}$ in V we have $1\vec{v} = \vec{v}$.

Example 3.2. The following are all vector spaces over the indicated field. To prove this, we would need to check each of these sets with the indicated operations satisfies all 10 axioms above, but we omit this for the sake of time. (Note: you'll prove the last example is a vector space on your next Worksheet).

(1) The set of polynomials with coefficients in $\mathbb{C}$ is a vector space over $\mathbb{C}$;

(2) The set of continuous function $f : \mathbb{R} \to \mathbb{R}$ is a vector space over $\mathbb{R}$;

(3) The set of $n \times m$ matrices with entries in $\mathbb{Q}$ is a vector space over $\mathbb{Q}$;

Linear algebra as a discipline studies general vector spaces, and allows us to assign structure to strange examples such as those given above. In this class, we'll mainly focus on getting our footing in this general context by studying one particular family of vector spaces: vector space over $\mathbb{R}$ which are contained in Euclidean space. We define the following.

Definition 3.3. A SUBSPACE W of $\mathbb{R}^n$ is any subset of $\mathbb{R}^n$ that is also a vector space over $\mathbb{R}$. In general, a VECTOR SUBSPACE W of a vector space V is any subset of V which is also a vector space.

The following theorem will make our lives much simpler when verifying a given set is a subspace. Rather than checking all 10 axioms above, we can instead do the following.

Theorem 3.4. *A subset W of a vector space V is a vector space (and hence a vector subspace of V) if and only if W is*

(1) closed under addition,

(2) closed under scalar multiplication, and

(3) contains the zero vector $\vec{0}$.

We omit the details of the proof here, but this isn't meant to be too challenging. Looking through the axioms, as long as W satisfies the above three properties, the remaining axioms are satisfied because W is defined under the same operations as the vector space V.

The following Theorem tells us that in fact we have already been studying the vector subspaces of $\mathbb{R}^n$.

Theorem 3.5. *A subset W is a vector subspace of $\mathbb{R}^n$ if and only if exists vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ so that*

$$W = Span\left(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\right).$$

Proof. Suppose first the $W = \text{Span}\left(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\right)$. Observe that $\vec{0} \in W$ since we can write

$$\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_m.$$

Furthermore, for $\vec{x}, \vec{y} \in W$ we can write

$$\vec{x} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m$$

$$\vec{y} = d_1\vec{v}_1 + \cdots + d_m\vec{v}_m$$

for scalars $c_i, d_i \in \mathbb{R}$ and so

$$\vec{x} + \vec{y} = (c_1 + d_1)\vec{v}_1 + \cdots + (c_m + d_m)\vec{v}_m \in \text{Span}(\vec{v}_1, \ldots, \vec{v}_m)$$

and for any $a \in \mathbb{R}$ we have

$$a\vec{x} = (ac_1)\vec{v}_1 + \cdots + (ac_m)\vec{v}_m \in \text{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

So, by Theorem 3.4, W is a subspace of $\mathbb{R}^n$.

Conversely, suppose that W is a vector space of $\mathbb{R}^n$. By Theorem 1.43, any subset of $\mathbb{R}^n$ contains between 0 and n linearly independent vectors. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ be a linearly independent subset of W where $m \leq n$ is maximal. Since W is a vector space, we know that

$$\text{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m) \subseteq W.$$

For the opposite set inclusion, take any $w \in W$. If $w = v_i \in B$ then certainly $w \in \text{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$ by writing

$$w = 0\vec{v}_1 + \cdots + 1 \cdot \vec{v}_i + \cdots + 0\vec{v}_m.$$

So, suppose that $w \notin \text{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$. Since m is maximal, the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m, \vec{w}\}$ must be linearly **dependent**. So, there exist scalars $c_1, \ldots, c_{m+1} \in \mathbb{R}$ not all equal to zero so that

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m + c_{m+1}\vec{w} = \vec{0}.$$

Furthermore, $c_{m+1} \neq 0$ since the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ is linearly independent. So we have

$$w = \frac{1}{c_{m+1}}\left(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m\right) \in \text{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m).$$

□

In particular, we obtain the following important Corollary.

Corollary 3.6. Every vector subspace of $\mathbb{R}^n$ contains a basis.

In fact, we have the more general result.

Theorem 3.7. *Any vector space V over a field $\mathbb{F}$ contains a basis.*

The proof of Theorem 3.7 is beyond the scope of the course, but we state it here since we'll see later on that this is the key step to proving one of the fundamental structure theorems on vector spaces.

3.15. Bases and Dimension

In Activity 1.21, I asked you to find two distinct bases of some of the sets from P1. There are actually infinitely many different bases you could have chosen from, but what I hoped you notice is that the *size* of your two bases (that is, the number of elements your basis contains) for a given vector space are the same. Let's give a formal proof of this fact.

Theorem 3.8. *Let W be a vector subspace of $\mathbb{R}^n$. Then, the size of any basis for W is unique.*

The proof of this fact uses a proof method called PROOF BY CONTRADICTION. We take another short interlude to introduce this method formally, and then return to the proof of this theorem.

Interlude: Proof by Contradiction

Recall that a conditional statement is a statement of the form "if P then Q". We learned how to prove a conditional statement directly in a previous interlude, by assuming the hypothesis P is true, and using some logical arguments to conclude that the conclusion Q must also be true.

Sometimes, proving a statement directly is out of reach. Instead, we have the following crafty strategy. Note that this works to prove any type of statement (not just conditional statements).

Suppose that we want to prove statement X is true.

(1) Assume that X is **false**.
(2) Use some logical reasoning to show that this implies a contradiction (that is, some nonsense that certainly isn't true, like $2 = 4$ or π is an integer, etc).
(3) Conclude that statement X must have been true.

Hopefully, it feels intuitive that the only way for a statement to imply something false is if the statement itself was false (note that the statement "X is false" being false means that X is true). But if you're interested in why proof by contradiction is valid using formal logic, I would look at Section 3.3 Sundstrum's text (also linked in the Proof Writing Guides page on the "Pages" tab on Canvas).

Let's look at a quick example of proof by contradiction.

Proposition 3.9. The sum of a rational number and an irrational number is irrational.

Proof. Suppose that x is rational and y is irrational. Then we can write $x = a/b$ for integers a and b with $b \neq 0$. For a contradiction, suppose that $x + y$ is rational. Then, there exist integers c and $d \neq 0$ so that $x + y = c/d$. We get

$$\frac{a}{b} + y = \frac{c}{d}$$
$$\Rightarrow y = \frac{c}{d} - \frac{a}{b} = \frac{cb - ad}{bd},$$

which implies that y is rational. Since a real number cannot be both rational and irrational, we have a contradiction. So, it must be the case that $x + y$ is irrational. $\square$

3.15 Continued

We are now prepared to prove our result.

Proof of Theorem 3.8. Suppose that W is a vector subspace of $\mathbb{R}^n$ with bases

$$B = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}, \text{ and } C = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m\}.$$

We need to show that $k = m$.

For a contradiction, suppose that $k < m$. We'll show this implies the set C is linearly **dependent** (similar to the examples you showed in Activity 3.22), contradicting the fact that it's a basis. For each $i \in \{1, \dots, m\}$ write

$$\vec{u}_i = a_{i1}\vec{v}_1 + a_{i2}\vec{v}_2 + \cdots + a_{ik}\vec{v}_k. \tag{3.1}$$

Now, consider the vector equation

$$\vec{0} = x_1\vec{u}_1 + \cdots + x_m\vec{u}_m \tag{3.2}$$

with unknowns $x_1, \dots, x_m$. Replacing each $\vec{u}_i$ with its representation in terms of the basis elements from B as in Equation (3.1) and collecting coefficients, we obtain

$$\begin{aligned}\vec{0} = \quad & (x_1a_{11} + x_2a_{21} + \cdots + x_ma_{m1})\vec{v}_1 \\ & + (x_1a_{12} + x_2a_{22} + \cdots + x_ma_{m2})\vec{v}_2 \\ & \vdots \\ & + (x_1a_{1k} + x_2a_{2k} + \cdots + x_ma_{mk})\vec{v}_k.\end{aligned}$$

Now, since the $\vec{v}_i$ are linearly independent, each of the coefficients above must be equal to zero. This yields the system of linear equations

$$\begin{gathered}x_1a_{11} + x_2a_{21} + \cdots + x_ma_{m1} = 0 \\ x_1a_{12} + x_2a_{22} + \cdots + x_ma_{m2} = 0 \\ \vdots \\ x_1a_{k1} + x_2a_{k2} + \cdots + x_ma_{mk} = 0,\end{gathered}$$

which is equivalent to the matrix-vector equation

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

where

$$A = \begin{pmatrix} a_{ji} \end{pmatrix} \text{ and } \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

Since A is a $k \times m$ matrix and $k < m$, there are more columns than rows, which means we have a column with no pivot. Since a homogeneous system is always consistent, then Theorem 1.19 tells us that this system has infinitely many solutions. Namely, we have a nontrivial solution to (3.2), which means our set $C = \{\vec{u}_1, \dots, \vec{u}_m\}$ is linearly dependent, a contradiction.

Note that if $k > m$, then we could follow the argument above to show that B is linearly dependent. Since k is neither smaller nor larger than m, we must have $k = m$ as desired.

□

This gives rise to the following definition.

Definition 3.10. Let V be a vector subspace of $\mathbb{R}^n$. Then, the DIMENSION of V, denoted $\dim V$, is equal to the size of any basis for V. We define the dimension of the trivial subspace $\{\vec{0}\}$ to be 0.

In the next section, we'll look at strategies to compute bases and the dimension of vector spaces. First, let's prove the key structure theorem I've been alluding to for a few weeks now (hopefully this will clear up questions many of you have had as well!). First, we need a definition.

Definition 3.11. Let V and W be vector spaces. An ISOMORPHISM between V and W is a linear map $T : V \to W$ that is both one-to-one and onto. In this case, we write $V \cong W$.

Theorem 3.12. *Let V be a vector subspace of $\mathbb{R}^n$ of dimension m. Then $V \cong \mathbb{R}^m$.*

Proof. Let $B = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_m\}$ be a basis for V. Define the function $T : V \to \mathbb{R}^m$ by sending $T(\vec{b}_i) = \vec{e}_i$ and extend linearly. That is,

$$T(a_1\vec{b}_1 + \cdots + a_m\vec{b}_m) = a_1\vec{e}_1 + \cdots + a_m\vec{e}_m.$$

Observe that T is linear. To see that T is one-to-one, suppose that $\vec{v} \in \ker T$. Writing $\vec{v} = a_1\vec{b}_1 + \cdots + a_m\vec{b}_m$ this gives

$$T(\vec{v}) = \vec{0} \Rightarrow a_1\vec{v}_1 + \cdots + a_m\vec{b}_m \equiv \vec{0}.$$

But since B is a basis, it's linearly independent, so we must have $a_i = 0$ for all i. Hence, $\vec{v} = \vec{0}$ and so $\ker T = \{\vec{0}\}$. For surjectivity, note that any $\vec{x} \in \mathbb{R}^m$ can be written in the form $\vec{x} = a_1\vec{e}_1 + \cdots + a_m\vec{e}_m$ and so

$$T(a_1\vec{b}_1 + \cdots + a_m\vec{b}_m) = \vec{x}.$$

Hence, T is an isomorphism, and we have $V \equiv \mathbb{R}^m$. □

Remark 3.13. The Theorem above tells us that the defining feature of a vector space is its dimension. Geometrically, every vector subspace of $\mathbb{R}^n$ of dimension m "looks like" a copy of $\mathbb{R}^m$ sitting inside of $\mathbb{R}^n$. For example, 2-dimensional vector spaces are all planes through the origin in $\mathbb{R}^3$, and 1-dimensional vector spaces are all lines through the origin.

3.13. The Null and Column Space of a Matrix

Example 3.14. In Activity 3.23 I asked you to find a basis (that is, a linearly independent generating set) for the vector subspace $V = \operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4)$ of $\mathbb{R}^4$ where

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 1 \\ 3 \\ -1 \\ 1 \end{pmatrix}, \vec{v}_3 \begin{pmatrix} 0 \\ 2 \\ -1 \\ 1 \end{pmatrix}, \text{ and } \vec{v}_4 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}.$$

We can observe by hand that

$$\vec{v_3} = 2\vec{v}_1 - \vec{v}_2 \text{ and } \vec{v_4} = -\vec{v}_1 + \vec{v}_2$$

and so $\vec{v}_3, \vec{v}_4 \in \operatorname{Span}(\vec{v}_1, \vec{v}_2)$. Furthermore, we can check that $\{\vec{v}_1, \vec{v}_2\}$ is a linearly independent set, and so we can take this to be a basis. This method is perfectly fine, but is a bit lengthy and depends quite a lot on our choice of generating set (imagine if we were looking at something like 100 vectors in $\mathbb{R}^{57}$!). Let's use this example to develop a more systematic way to find bases.

Let A be the matrix with column vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$. Then A has reduced row echelon form

$$X = \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Observe that columns 3 and 4 of X have no pivot, so there's a nontrivial solution to the vector equation

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 + x_4\vec{v}_4 = \vec{0}.$$

which implies the set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ is linearly dependent. So, we need eliminate one of our vectors. **BUT** we don't want to change our vector space, so we need to make sure the vector we remove is in the span of the other three. I claim that we can remove any vector that's not in a "pivot column" of A and not change the span. Let's understand why this works in general before continuing this example.

Lemma 3.15. *Let A be the $m \times n$ matrix*

$$A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{pmatrix}$$

and suppose that the reduced row echelon form of A is given by the matrix

$$X = \begin{pmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \end{pmatrix}.$$

If the column $\vec{x}_n$ of X does not have a pivot, then

$$\mathrm{Span}(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n) = \mathrm{Span}(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{n-1}).$$

We call a vector $\vec{v}_i$ a PIVOT COLUMN *of A if the column $\vec{x}_i$ contains a pivot in X. In general, we can remove any column of A that is not a pivot column and not change the span of its column vectors.*

Proof. Since there is no pivot in the column $\vec{x}_n$, we know that the system of linear equation with augmented matrix

$$\left(\begin{array}{cccc|c} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n-1} & \vec{v}_n \end{array}\right)$$

has a solution. So, $\vec{v}_n \in \mathrm{Span}(\vec{v}_1, \dots, \vec{v}_{n-1})$. □

Example 3.16 (Example 3.14 continued...)**.** Since $\vec{v}_4$ is not a pivot column of A, then by Lemma 3.15 we have

$$V = \mathrm{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3).$$

Now, the matrix with columns $\vec{v}_1, \vec{v}_2, \vec{v}_3$ has reduced row echelon form

$$X = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

which still has a column with no pivot. So our set is still linearly dependent, which means we should remove another vector. Since there is only one column with no pivot, we should remove $\vec{v}_3$ By Lemma 3.15 again, we have

$$V = \mathrm{Span}(\vec{v}_1, \vec{v}_2).$$

Furthermore, the reduced row echelon form of the matrix with columns $\vec{v}_1, \vec{v}_2$ is given by

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and so the set $\{\vec{v}_1, \vec{v}_2\}$ is also linearly independent, making it a basis for V.

We have the following.

Theorem 3.17 (Finding Bases)**.** *Let V be a vector subspace of $\mathbb{R}^n$ written in the form*

$$V = \operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

If A is the matrix with column vectors $\vec{v}_1, \ldots, \vec{v}_m$ then the pivot columns of A will form a basis for V.

Let's omit the proof of this theorem for the sake of time, and believe that our Example 3.14 generalizes.

3.13.1. The Column Space.

Observe that finding bases of a vector space had to do with finding the space spanned by a certain matrix. Sometimes we might want to reverse this problem (as we'll see later on when we look for images of matrix transformations) and look for the span of the columns of a matrix A. We define the following.

Definition 3.18. Let A be an $n \times m$ matrix with column vectors $\vec{v}_1, \ldots, \vec{v}_m$. Then the COLUMN SPACE of A is the vector subspace $\operatorname{Col}(A)$ of $\mathbb{R}^n$ defined by

$$\operatorname{Col}(A) := \operatorname{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m).$$

The RANK of A is defined to be the dimension of the column space of A. That is,

$$\operatorname{rank}(A) = \dim(\operatorname{Col})(A).$$

We say that an $n \times m$ matrix A has FULL RANK if $\operatorname{rank}(A) = m$.

Remark 3.19. The column space of a matrix is particularly useful when studying linear transformations: if $T_A : \mathbb{R}^m \to \mathbb{R}^n$, observe that the image of T_A is precisely equal to the column space of A. The final main theorem of this Chapter gives a relation between the image of a linear transformation and its kernel. First, we have the following important observation.

Lemma 3.20. *Let $T : V \to W$ be a linear transformation between vector spaces V and W. Then,*

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

is a vector subspace of W.

Proof. By definition, $\ker T$ is a subset of the vector space W, so we just need to check the three properties in Theorem 3.4. Since T is linear, we must have $T(\vec{0}) = \vec{0}$. Now, take any $\vec{u}, \vec{v} \in \ker T$. Then, using the linearity property for T we have

$$T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}) = \vec{0} + \vec{0} = \vec{0}.$$

Hence, $\vec{u} + \vec{v} \in \ker T$. Now, for any scalar $c \in \mathbb{R}$ and vector $\vec{v} \in V$, again using the linearity property for T we have

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

Hence, $\ker T$ is a vector space. □

3.13.2. The Null Space. Note that, for an $n \times m$ matrix A, we have

$$\ker T_A = \{\vec{x} \in \mathbb{R}^m \mid A\vec{x} = \vec{0}\}.$$

This gives rise to the following definition.

Definition 3.21. The NULL SPACE of an $n \times m$ matrix A is the vector subspace of $\mathbb{R}^m$ defined by

$$\mathrm{Nul}(A) := \{\vec{x} \in \mathbb{R}^m \mid A\vec{x} = \vec{0}\}.$$

The NULLITY of A is defined to be the dimension of the null space of A. That is,

$$\mathrm{nullity}(A) := \dim(\mathrm{Nul}(A)).$$

Example 3.22. Let $V = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ be the vector space in Example 3.14, and A the matrix with column vectors $\vec{v}_i$. To find $\mathrm{Nul}(A)$, we need to find all vector solutions $\vec{x}$ to the homogeneous system

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

which are equal to the solution to

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

where X is the reduced row echelon form of A. Since

$$X = \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

then we have two free variables, call them z, w. So, all solutions to $A\vec{x} = \vec{0}$ can be written in the form

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} z - w \\ -z + w \\ z \\ w \end{pmatrix}.$$

Observe that this gives

$$\mathrm{Nul}(A) = \mathrm{Span}\left(\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right)$$

and so $\mathrm{nullity}(A) = 2$.

In the example above, we used the number of free variables to say something about the nullity of A. In general, we have the following.

Theorem 3.23. *Let A be an $n \times m$ matrix with r pivot columns. Then,*

$$\mathrm{rank}(A) = r \text{ and } \mathrm{nullity}(A) = m - r.$$

Proof. The fact that the rank of A is equal to the number of pivot columns follows from Theorem 3.17. Now, suppose that A has reduced row echelon form equal to X. If X contains r columns with pivots, then that means X will have $m-r$ free variables. To make our notation simpler, let's assume the pivots are in the first r columns (note that the argument follows identically by an appropriate relabeling of the column vectors if not). So, if

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \in \mathrm{Nul}(A)$$

then we can write $x_1, \ldots, x_r$ in terms of the remaining $m-r$ free variables $x_{r+1}, \ldots, x_m$. That is, there are real numbers a_i with

$$\begin{aligned} x_1 &= a_{1,r+1}x_{r+1} + a_{1,r+2}x_{r+2} + \cdots + a_{1,m}x_m \\ x_2 &= a_{2,r+1}x_{r+1} + a_{2,r+2}x_{r+2} + \cdots + a_{2,m}x_m \\ &\vdots \\ x_r &= a_{r,r+1}x_{r+1} + a_{r,r+2}x_{r+2} + \cdots + a_{r,m}x_m. \end{aligned}$$

So, as in Example 3.22, we have

$$\mathrm{Nul}(A) = \mathrm{Span}\left(\begin{pmatrix} a_{1,r+1} \\ a_{2,r+1} \\ \vdots \\ a_{r,r+1} \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} a_{1,r+2} \\ a_{2,r+2} \\ \vdots \\ a_{r,r+2} \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} a_{1,m} \\ a_{2,m} \\ \vdots \\ a_{r,m} \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \right).$$

Finally, observe that the $m-r$ vectors above are linearly independent, since the matrix with the vectors above as its columns has a pivot in every column. □

We have the following consequence to Theorem 3.23, often referred to as the RANK-NULLITY THEOREM.

Corollary 3.24 (The Rank-Nullity Theorem)**.** Let A be an $n \times m$ matrix. Then,

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

Equivalently, for a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ we have

$$\dim(\mathrm{im}\, T) + \dim(\ker T) = m.$$

Chapter 4

Eigenvalues and Eigenvectors

4.16. The Determinant Part I: Motivation

Recall, in Proposition 2.10 we found that the inverse of a 2×2 matrix is given by

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

This tells us that a 2×2 matrix is invertible precisely when $ad - bc \neq 0$. Our goal in this section is to generalize this straightforward computation in order to gain a new tool to determine whether any $n \times n$ matrix is invertible. Using geometric reasoning we'll develop a function, called the DETERMINANT, which inputs a matrix and outputs a real number that is nonzero precisely when that matrix is invertible.

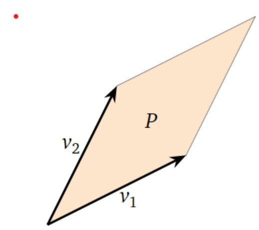
(Note: much of the content in the remainder of this section, including the pictures, are pulled from the notes for similar course taught at Georgia Tech (see this link)).

Let A be an $n \times n$ matrix. Recall that A is invertible if and only if $\mathrm{Col}(A) = \mathbb{R}^n$. Another way to say this is that A is *not* invertible if and only if $\mathrm{Col}(A)$ is a PROPER SUBSPACE of $\mathbb{R}^n$ (that is, $\mathrm{Col}(A) \neq \mathbb{R}^n$). For example, in $\mathbb{R}^3$, all of the matrices that are not invertible either have column space equal to a plane (i.e. an isomorphic copy of $\mathbb{R}^2$), a line (i.e. an isomorphic copy of $\mathbb{R}^1$) or a point. We'll capture this idea by computing the volume of a certain object, which will be zero exactly when we're "missing" a dimension. We have the following definition.

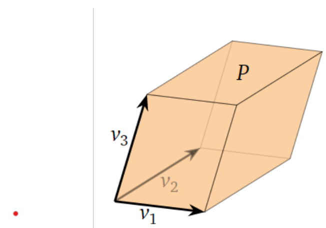
Definition 4.1. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ be a subset of $\mathbb{R}^n$. Then, the PARALLELEPIPED of the set B is defined by

$$P_B := \{a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_m\vec{v}_m \mid 0 \leq a_i \leq 1\}.$$

Example 4.2. If B contains two linearly independent vectors, the parallelepiped $P = P_B$ is just a parallelogram



If B contains three linearly independent vectors, the parallelepiped $P = P_B$ looks like a skewed cube, as below.

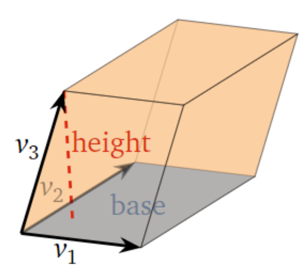


Observation 4.3. Let A be an $n \times n$ matrix and B the set containing the column vectors of A. Then, A is not invertible if and only if the volume of P_B is equal to 0.

Notation. For notational convenience, for an $n \times n$ matrix A, we'll let P_A be the parallelepiped defined by the columns of A.

This gives us an idea: why don't we just define the determinant of a matrix to be equal to the volume of the parallelepiped defined by the columns of A? This would capture the property we want, but the trouble is it's not easy to find a formula for the volume of these objects. So we'll have to be a bit more crafty. First, we note the following.

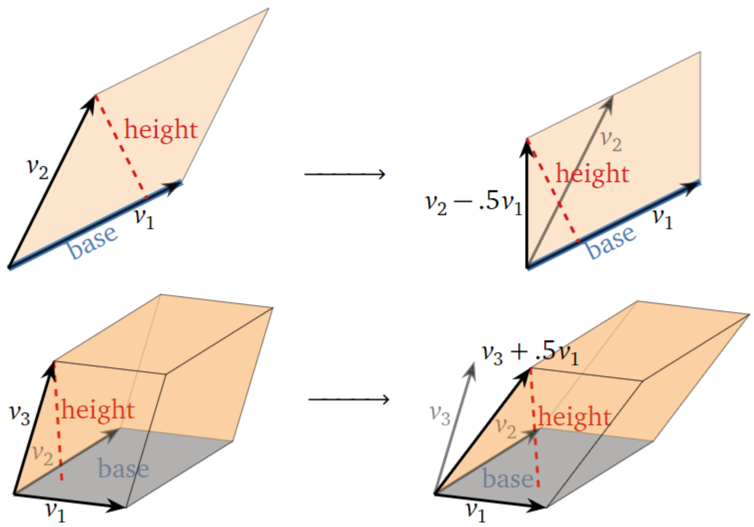
Lemma 4.4. *Let $B = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be a subset of $\mathbb{R}^n$ and let P_B be the parallelepiped defined by the set B. Let $C = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{n-1}\}$ be the subset of B with $\vec{v}_1$ removed. Then, the volume of P_B is equal to the area of the "base" P_C times the height of the vector $\vec{v}_n$ from the "base" P_C, as in the example below*



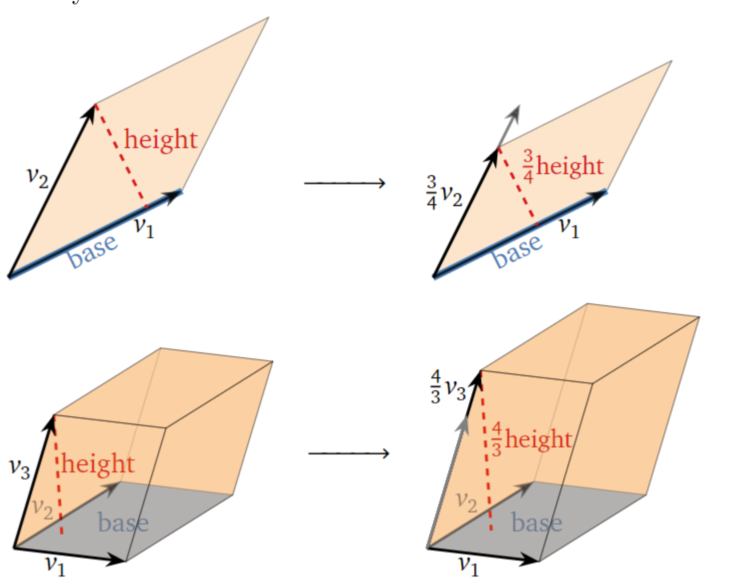
We'll skip the proof of this Lemma, but note it could be done using calculus by representing the volume of P_B as an integral. Lemma 4.4 allows us to make the following observations.

(1) If one of the vectors $\vec{v}_i$ is equal to 0, then the volume of P_B is equal to zero.

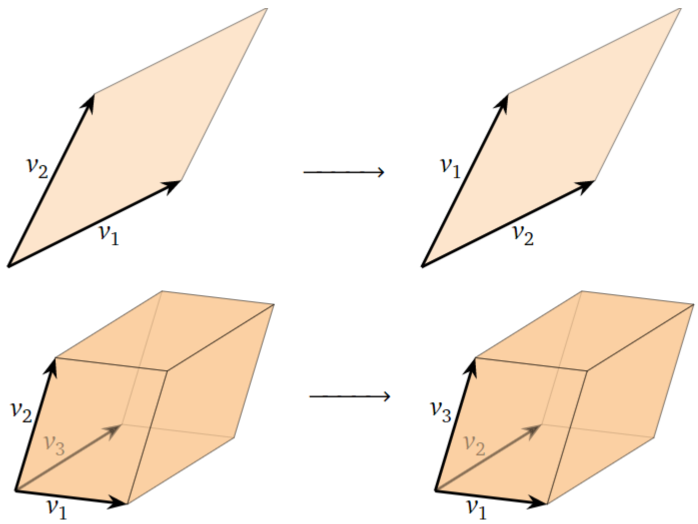
(2) If we replace $\vec{v}_n$ by $\vec{v}_n + c\vec{v}_i$, this just translates the vector $\vec{v}_n$ parallel to the base, which does not change the height, and so does not change the volume.



(3) Scaling a vector $\vec{v}_n$ by $c > 0$ changes the height by a factor of c which multiplies the volume by c.



(4) We can reorder the vectors and not change the volume



(5) If B is the standard basis for $\mathbb{R}^n$, that is if $B = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ then the volume of P_B is equal to 1.

Example 4.5. Using the properties above, we can now easily compute the area of any parallelepiped. For example, let $B = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ where

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \vec{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 5 \end{pmatrix}.$$

We can obtain P_B by taking the parallelepiped formed by the standard basis, replacing $\vec{e}_1$ by $\vec{e}_1 + \vec{e}_2$ (which does not change the volume) and the by scaling $\vec{e}_3$ by 5. Using the properties above, this gives us a parallelepiped of volume 5.

Example 4.5 should look familiar from yesterday's activity. Observe that Properties (2)-(4) look a lot like row reducing a matrix!

In the second part of this section, we'll show how the properties in Lemma 4.4 can give us a closed formula for the volume of any fundamental parallelepiped. Note that in theory we could just use the method in the example above, but we'd like to find a closed formula in order to show this method always works, and to help us come up with better strategies to make this computation more efficient.

We'll turn our attention back to defining this operation on matrices. Since the operations in Lemma 4.4 looked like *row operation*, we'll instead define the determinant to match the volume of the parallelepiped defined by the *rows* of A (rather than the columns), but we'll get a fix for this at the end of the section.

4.16. The Determinant Part II: Existence and Uniqueness

Definition 4.6. The DETERMINANT of an $n \times n$ matrix is a function

$$\det : \mathrm{Mat}_n(\mathbb{R}) \to \mathbb{R}$$

that satisfies the following properties:

(1) For any matrix B obtained from A by the elementary row operation that replaces R_i with $R_i + cR_j$, we have $\det(B) = \det(A)$

(2) For any matrix C obtained from A by the elementary row operation that replaces R_i with cR_i, we have $\det(C) = c\det(A)$.

(3) For any matrix D obtained from A by the elementary row operation that swaps two rows of A, we have $\det(D) = -\det(A)$.

(4) $\det(I_n) = 1$.

To show that this function is well-defined (that is, this function exists and is unique) we'll need to show the determinant satisfies some further properties. First, we need some definitions.

Definition 4.7. Let A be an $n \times n$ matrix

$$A = \begin{pmatrix} {\color{red}a_{11}} & a_{12} & \cdots & a_{1n} \\ a_{21} & {\color{red}a_{22}} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & {\color{red}a_{nn}} \end{pmatrix}.$$

Then, the DIAGONAL ENTRIES of A are the entries $a_{11}, a_{22}, \ldots, a_{nn}$ (highlighted in red in the matrix above). We say that A is UPPER TRIANGULAR if the only nonzero entries are those on or above the diagonal entries of A; that is

$$a_{ij} = 0 \text{ for all } i > j.$$

We say A is LOWER TRIANGULAR if the only nonzero entries are those on or below the diagonal entries of A; that is,

$$a_{ij} = 0 \text{ for all } i < j.$$

Example 4.8. The matrix

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

is upper triangular with diagonal entries $1, 2, 1$, and the matrix

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

is lower triangular with diagonal entries $1, 0, 2$.

Properties of the Determinant. Below we derive several properties of the determinant by using Definition 4.6. You will be asked to fill in the missing details for each of the following six proofs on Homework 6.

Proposition 4.9 (Determinants of Triangular Matrices)**.** If A is upper or lower triangular with diagonal entries $d_1, d_2, \dots, d_n$ then

$$\det(A) = d_1 d_2 \cdots d_n.$$

Proof. Let's prove this in the 3×3 case, and note that the general case follows similarly. Suppose first that A is upper triangular. Then we can write

$$A = \begin{pmatrix} d_1 & a_{12} & a_{13} \\ 0 & d_2 & a_{23} \\ 0 & 0 & d_3 \end{pmatrix}.$$

(a) Show that A is row equivalent to a matrix of the form

$$\begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

using only row operations of the form: replace R_i with $R_i + cR_j$ for $i \neq j$.

So, we have $\det(A) = \det \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$.

(b) Use what we've shown above and Definition 4.6 of the determinant from our course lecture notes to show that $\det(A) = d_1 d_2 d_3$.

Next, if A is lower triangular then we can write

$$A = \begin{pmatrix} d_1 & 0 & 0 \\ a_{21} & d_2 & 0 \\ a_{31} & a_{32} & d_3 \end{pmatrix}.$$

(c) Use a similar method to what we did in the upper triangular case to show that $\det(A) = d_1 d_2 d_3$.

$\square$

Proposition 4.10 (Determinants Detect Invertibility)**.** A matrix A is invertible if and only if $\det(A) \neq 0$.

Proof. Suppose first that A is invertible.

(a) Show that if A is invertible, then there exists a nonzero constant c with

$$\det(A) = c \det(I_n).$$

By Property (4) in our definition of the determinant (Definition 4.6) we know that

$\det(I_n) \neq 0$. Since $c \neq 0$ as well, then by what you've shown in part (c) we have

$$\det(A) \neq 0.$$

Next, suppose that A is *not* invertible.

(b) Let X be the reduced row echelon form of A. Explain why at least one of the diagonal entries of X must be equal to 0 when A is not invertible.

(c) Show that there exists a nonzero constant c with

$$\det(A) = c \det(X).$$

Since the reduced row echelon form of a square matrix is upper triangular, then by part (b) and Proposition 4.9 we have

$$\det(A) = c \cdot 0 = 0.$$

So, we've shown if A is *not* invertible, then $\det(A) = 0$. Note that this is equivalent to the statement: if $\det(A) \neq 0$ then A is invertible. Formally, this is called the CONTRAPOSITIVE (my hope is that this logic feels intuitive, but you can also read about this in our proof writing resources). □

Proposition 4.11 (The Determinant is Multiplicative)**.** For matrices A and B we have

$$\det(AB) = \det(A)\det(B).$$

Proof. Suppose first that A is an elementary matrix, as defined in Definition 2.11.

(a) For each of the elementary matrices E given in Definition 2.11, find $\det(E)$.

Recall that multiplication on the left by an elementary matrix corresponds to an elementary row operation.

(b) Show that $\det(EB) = \det(E)\det(B)$ for an elementary matrix E.

Now, let A be an arbitrary matrix. If A is invertible, then by Invertible Matrix Theorem (Theorem 2.13) we know that A is a product of elementary matrices, say

$$A = E_1 E_2 \cdots E_m.$$

(c) Use what we've shown above to prove that

$$\det(AB) = \det(A)\det(B)$$

when A is invertible.

Next, suppose that A is not invertible.

(d) Show that AB is not invertible. *(Hint: look at Problem 9 from Homework 4.)*

By Proposition 4.10 we have

$$\det(AB) = 0 \text{ and } \det(A) = \det(B) = 0,$$

and so $\det(AB) = \det(A)\det(B)$ as desired. □

Proposition 4.12 (Invariance Under the Transpose)**.** For any $n \times n$ matrix A we have $\det(A) = \det(A^\top)$.

Proof. Suppose first that A is not invertible. Then $\det(A) = 0$ by Proposition 4.10.

(a) Show that $\det(A^\top) = 0$ when A is not invertible.

So, $\det(A) = \det(A^\top)$ in this case.

Next, suppose that A is invertible. By the Invertible Matrix Theorem (Theorem 2.13) we can write A as a product of elementary matrices, say

$$A = E_1 E_2 \cdots E_m.$$

So by Problem 1(c) of Worksheet 4 we have

$$A^\top = E_m^\top E_{m-1}^\top \cdots E_1^\top.$$

(b) For an elementary matrix E, show that $\det(E) = \det(E^\top)$.

(c) Use part (b) and Property 4.11 to show that $\det(A^\top) = \det(A)$ when A is invertible.

□

Proposition 4.13 (Reduction)**.** If A is an $n \times n$ matrix of the form

$$A = \left(\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline * & & & & \\ * & & & & \\ \vdots & & & B & \\ * & & & & \end{array}\right),$$

for an $(n-1) \times (n-1)$ matrix B and where $*$ denotes any real number (possibly zero, but possibly not), then $\det(A) = \det(B)$.

Proof. Let A and B be as above. Suppose first that B is invertible. Then, by the Invertible Matrix Theorem, there exist elementary matrices $E_1, \ldots, E_m$ so that

$$E_m \cdots E_2 E_1 B = I_n. \tag{4.1}$$

For an $(n-1)\times(n-1)$ elementary matrix E define the $n\times n$ matrix

$$\widetilde{E} := \left(\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline 0 & & & & \\ 0 & & & & \\ \vdots & & & E & \\ 0 & & & & \end{array}\right).$$

(a) Show that $\widetilde{E}$ is also an elementary matrix with $\det(\widetilde{E}) = \det(E)$.

Next, observe the following identity, which can be verified just by chasing through the definition of matrix multiplication carefully (I won't make you do this, but I do suggest you make sure to understand why this holds)

$$\widetilde{E}A = \left(\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline * & & & & \\ * & & & & \\ \vdots & & & EB & \\ * & & & & \end{array}\right) \tag{4.2}$$

Applying equation (4.2) repeatedly, we get

$$\widetilde{E}_m \cdots \widetilde{E}_2\widetilde{E}_1 A = \left(\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline * & & & & \\ * & & & & \\ \vdots & & & E_m\cdots E_2E_1B & \\ * & & & & \end{array}\right) = \left(\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline * & & & & \\ * & & & & \\ \vdots & & & I_{n-1} & \\ * & & & & \end{array}\right)$$

where the second equality follows by definition of the E_i in Equation (4.1).

(b) Use the equality above to show that

$$\det(A) = 1/\det(\widetilde{E}_m \cdots \widetilde{E}_2\widetilde{E}_1).$$

(Hint: the matrix on the right-hand side is diagonal).

Using Proposition 4.11 repeatedly along with what you showed in part (a) we get

$$\begin{aligned}\det(\widetilde{E}_m \cdots \widetilde{E}_2\widetilde{E}_1) &= \det(\widetilde{E}_m)\cdots\det(\widetilde{E}_2)\det(\widetilde{E}_1) \\ &= \det(E_m)\cdots\det(E_2)\det(E_1) \\ &= \det(E_m\cdots E_2E_1).\end{aligned}$$

(c) Conclude that

$$\det(A) = \det(B).$$

(Hint: use the equality above, Equation (4.1), and part (b)).

$\square$

Proposition 4.14 (Row Additivity)**.** The determinant is additive in the rows of a matrix. That is,

$$\det\begin{pmatrix} x_{11}+y_{11} & x_{12}+y_{12} & \cdots & x_{1n}+y_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix} = \det\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix} + \det\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}$$

Proof. Let's prove this in the 2×2 case, and note that the general case follows similarly (if you'd like these details, checkout the multilinearity property in these notes). I don't have a good geometric intuition for this argument unfortunately (there should be one), so I'm going to leave reading this proof as a suggested exercise.

We'll show that

$$\det\begin{pmatrix} x_1+y_1 & x_2+y_2 \\ a & b\end{pmatrix} = \det\begin{pmatrix} x_1 & x_2 \\ a & b\end{pmatrix} + \det\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix}.$$

First, suppose that

$$\begin{pmatrix} x_1 \\ x_2\end{pmatrix} \in \operatorname{Span}\left(\begin{pmatrix} y_1 \\ y_2\end{pmatrix}, \begin{pmatrix} a \\ b\end{pmatrix}\right).$$

Then, there are constants c_1, c_2 so that

$$\begin{pmatrix} x_1 \\ x_2\end{pmatrix} = c_1\begin{pmatrix} y_1 \\ y_2\end{pmatrix} + c_2\begin{pmatrix} a \\ b\end{pmatrix}$$

which gives

$$\begin{aligned} x_1 &= c_1 y_1 + c_2 a \\ x_2 &= c_1 y_2 + c_2 b. \end{aligned}$$

So, we have

$$\begin{aligned}
\det\begin{pmatrix} x_1+y_1 & x_2+y_2 \\ a & b\end{pmatrix} &= \det\begin{pmatrix} c_1y_1+c_2a+y_1 & c_1y_2+c_2b+y_2 \\ a & b\end{pmatrix} \\
&= \det\begin{pmatrix} (c_1+1)y_1+c_2a & (c_1+1)y_2+c_2b \\ a & b\end{pmatrix} \\
&= \det\begin{pmatrix} (c_1+1)y_1 & (c_1+1)y_2 \\ a & b\end{pmatrix}, \text{Replacing } R_1 \text{ with } R_1 - c_2R_2 \\
&= (c_1+1)\det\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix}, \text{Replacing } R_1 \text{ with } R_1/(c_1+1) \\
&= c_1\det\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix} + \det\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix} \\
&= \det\begin{pmatrix} c_1y_1 & c_1y_2 \\ a & b\end{pmatrix} + \det\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix}, \text{ Replacing } R_1 \text{ with } c_1R_1 \text{ in the first matrix.}
\end{aligned}$$

Now, since

$$\begin{aligned} c_1y_1 &= x_1 - c_2a \\ c_1y_2 &= x_2 - c_2b \end{aligned}$$

then we can compute the determinant of the first matrix in our final equality above as follows

$$\begin{aligned}\det\begin{pmatrix} c_1y_1 & c_1y_2 \\ a & b\end{pmatrix} &= \det\begin{pmatrix} x_1 - c_2a & x_2 - c_2b \\ a & b\end{pmatrix} \\ &= \det\begin{pmatrix} x_1 & x_2 \\ a & b\end{pmatrix}, \text{ Replacing } R_1 \text{ with } R_1 + c_2R_2.\end{aligned}$$

Combining everything above, we get

$$\begin{aligned}\det\begin{pmatrix} x_1+y_1 & x_2+y_2 \\ a & b\end{pmatrix} &= \det\begin{pmatrix} c_1y_1 & c_1y_2 \\ a & b\end{pmatrix} + \det\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix} \\ &= \det\begin{pmatrix} x_1 & x_2 \\ a & b\end{pmatrix} + \det\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix}.\end{aligned}$$

Finally, suppose that

$$\begin{pmatrix} x_1 \\ x_2\end{pmatrix} \notin \operatorname{Span}\left(\begin{pmatrix} y_1 \\ y_2\end{pmatrix}, \begin{pmatrix} a \\ b\end{pmatrix}\right).$$

Then, there is a vector in $\mathbb{R}^2$ that's not contained in this span, and so we must have

$$\operatorname{Span}\left(\begin{pmatrix} y_1 \\ y_2\end{pmatrix}, \begin{pmatrix} a \\ b\end{pmatrix}\right) \neq \mathbb{R}^2.$$

This tells us that the set

$$\left\{\begin{pmatrix} y_1 \\ y_2\end{pmatrix}, \begin{pmatrix} a \\ b\end{pmatrix}\right\}$$

is linearly dependent (because any set of two linearly independent vectors forms a basis for $\mathbb{R}^2$!) and so we get that the matrix

$$\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix}$$

is not invertible. So, by Property 4.10 we have

$$\det\begin{pmatrix} y_1 & y_2 \\ a & b\end{pmatrix} = 0.$$

Furthermore, since the set

$$\left\{\begin{pmatrix} y_1 \\ y_2\end{pmatrix}, \begin{pmatrix} a \\ b\end{pmatrix}\right\}$$

is linearly dependent, one of these vectors is a multiple of the other, say that

$$\begin{pmatrix} y_1 \\ y_2\end{pmatrix} = c\begin{pmatrix} a \\ b\end{pmatrix}.$$

So, $y_1 = ca$ and $y_2 = cb$, and so we can get

$$\det\begin{pmatrix} x_1+y_1 & x_2+y_2 \\ a & b\end{pmatrix} = \det\begin{pmatrix} x_1 & x_2 \\ a & b\end{pmatrix}$$

by replacing R_1 with $R_1 - cR_2$, proving our equality in this case. $\square$

4.16.1. Existence and Uniqueness. We are now ready to show that the determinant is a well-defined function. That is, we need to show for any n there exists a unique function satisfying the properties of Definition 4.6. We'll show this by deriving an explicit formula for the determinant of an $n \times n$ matrix. First, let's look at the case when $n = 1$ and $n = 2$.

Lemma 4.15. *The determinant of a 1×1 matrix is well-defined and equal to*

$$\det\begin{pmatrix} a \end{pmatrix} = a.$$

The determinant of a 2×2 matrix is well-defined and equal to

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Proof. First, let's look at the 1×1 case. We have

$$\det\begin{pmatrix} a \end{pmatrix} = a \det\begin{pmatrix} I_1 \end{pmatrix}$$

using Property (2) of Definition 4.6, and since $\det(I_1) = 1$ by Property (4) then we must have

$$\det\begin{pmatrix} a \end{pmatrix} = a.$$

For the 2×2 case, we have

$$\begin{aligned}
\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \det\begin{pmatrix} a & 0 \\ c & d \end{pmatrix} + \det\begin{pmatrix} 0 & b \\ c & d \end{pmatrix}, \text{by row additivity (Prop 4.14)} \\
&= a\det\begin{pmatrix} 1 & 0 \\ c & d \end{pmatrix} - b\det\begin{pmatrix} 1 & 0 \\ d & c \end{pmatrix}, \text{ by Properties (1), (4) and Proposition 4.12} \\
&= a\det\begin{pmatrix} d \end{pmatrix} - b\det\begin{pmatrix} c \end{pmatrix}, \text{ by the reduction property (Prop 4.13)} \\
&= ad - bc.
\end{aligned}$$

Fortunately, this is the formula we derived using geometric reasoning, so all is on track so far! □

Next, let's look at an example of how we might use this strategy to compute the determinant of 3×3 matrices and 4×4 matrices.

Example 4.16. Let's look at the examples in P4 of Activity 3.28. Let

$$X = \begin{pmatrix} a & b & c \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

By row additivity of the determinant from (Proposition **??**) we can write

$$\det(X) = \det\begin{pmatrix} a & 0 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \det\begin{pmatrix} 0 & b & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \det\begin{pmatrix} 0 & 0 & c \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

Furthermore, since $\det(A) = \det(A^\top)$ for any matrix A, then we can perform the operations outlined in Definition 4.6 on the *columns* of A and keep track of how the determinant changes. For example,

$$\begin{aligned}
\det\begin{pmatrix}0 & b & 0\\ 1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} &= \det\begin{pmatrix}0 & 1 & 4\\ b & 2 & 5\\ 0 & 3 & 6\end{pmatrix}, \text{ since } \det(A^\top) = \det(A)\\
&= -\det\begin{pmatrix}b & 2 & 5\\ 0 & 1 & 4\\ 0 & 3 & 6\end{pmatrix}, \text{ swapping rows 1 and 2}\\
&= -\det\begin{pmatrix}b & 0 & 0\\ 2 & 1 & 3\\ 5 & 4 & 6\end{pmatrix}, \text{ again taking the transpose,}
\end{aligned}$$

but this is the same as just swapping the first and second columns. Let's do this with the third matrix in the summand above. Note that we could just swap the first and third column, but (for reasons we'll see later) I'm going to first swap the second and third column, and then after that swap the first and second column, as below

$$\det\begin{pmatrix}0 & 0 & c\\ 1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} = -\det\begin{pmatrix}0 & c & 0\\ 1 & 3 & 2\\ 4 & 6 & 5\end{pmatrix} = +\det\begin{pmatrix}c & 0 & 0\\ 3 & 1 & 2\\ 6 & 4 & 5\end{pmatrix}.$$

So, we have

$$\begin{aligned}
\det(X) &= \det\begin{pmatrix}a & 0 & 0\\ 1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} + \det\begin{pmatrix}0 & b & 0\\ 1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} + \det\begin{pmatrix}0 & 0 & c\\ 1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix},\\
&= \det\begin{pmatrix}a & 0 & 0\\ 1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} - \det\begin{pmatrix}b & 0 & 0\\ 2 & 1 & 3\\ 5 & 4 & 6\end{pmatrix} + \det\begin{pmatrix}c & 0 & 0\\ 3 & 1 & 2\\ 6 & 4 & 5\end{pmatrix}, \text{ from above}\\
&= a\det\begin{pmatrix}1 & 0 & 0\\ 1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} - b\det\begin{pmatrix}1 & 0 & 0\\ 2 & 1 & 3\\ 5 & 4 & 6\end{pmatrix} + c\det\begin{pmatrix}1 & 0 & 0\\ 3 & 1 & 2\\ 6 & 4 & 5\end{pmatrix}, \text{ from above}\\
&= a\det\begin{pmatrix}2 & 3\\ 5 & 6\end{pmatrix} - b\det\begin{pmatrix}1 & 3\\ 4 & 6\end{pmatrix} + c\det\begin{pmatrix}1 & 2\\ 4 & 5\end{pmatrix}, \text{ by the reduction property (Property 4.13)}\\
&= -3a + 6b - 3c,
\end{aligned}$$

where we used our definition of the 2×2 determinant in the last step. Let's use this same method for our 4×4 example. Let

$$Y = \begin{pmatrix}a & b & c & d\\ 1 & 1 & 3 & 5\\ 2 & 4 & 5 & 6\\ 1 & -1 & -1 & 2\end{pmatrix}.$$

Then,

$$\begin{aligned}
\det(Y) &= \det\begin{pmatrix} a & 0 & 0 & 0 \\ 1 & 1 & 3 & 5 \\ 2 & 4 & 5 & 6 \\ 1 & -1 & -1 & 2 \end{pmatrix} + \det\begin{pmatrix} 0 & b & 0 & 0 \\ 1 & 1 & 3 & 5 \\ 2 & 4 & 5 & 6 \\ 1 & -1 & -1 & 2 \end{pmatrix} + \det\begin{pmatrix} 0 & 0 & c & 0 \\ 1 & 1 & 3 & 5 \\ 2 & 4 & 5 & 6 \\ 1 & -1 & -1 & 2 \end{pmatrix} + \det\begin{pmatrix} 0 & 0 & 0 & d \\ 1 & 1 & 3 & 5 \\ 2 & 4 & 5 & 6 \\ 1 & -1 & -1 & 2 \end{pmatrix} \\
&= \det\begin{pmatrix} a & 0 & 0 & 0 \\ 1 & 1 & 3 & 5 \\ 2 & 4 & 5 & 6 \\ 1 & -1 & -1 & 2 \end{pmatrix} - \det\begin{pmatrix} b & 0 & 0 & 0 \\ 1 & 1 & 3 & 5 \\ 4 & 2 & 5 & 6 \\ -1 & 1 & -1 & 2 \end{pmatrix} + \det\begin{pmatrix} c & 0 & 0 & 0 \\ 3 & 1 & 1 & 5 \\ 5 & 2 & 4 & 6 \\ -1 & 1 & -1 & 2 \end{pmatrix} - \det\begin{pmatrix} d & 0 & 0 & 0 \\ 5 & 1 & 1 & 3 \\ 6 & 2 & 4 & 5 \\ 2 & 1 & -1 & -1 \end{pmatrix} \\
&= a\det\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 3 & 5 \\ 2 & 4 & 5 & 6 \\ 1 & -1 & -1 & 2 \end{pmatrix} - b\det\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 3 & 5 \\ 4 & 2 & 5 & 6 \\ -1 & 1 & -1 & 2 \end{pmatrix} + c\det\begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 1 & 5 \\ 5 & 2 & 4 & 6 \\ -1 & 1 & -1 & 2 \end{pmatrix} - d\det\begin{pmatrix} 1 & 0 & 0 & 0 \\ 5 & 1 & 1 & 3 \\ 6 & 2 & 4 & 5 \\ 2 & 1 & -1 & -1 \end{pmatrix} \\
&= a\det\begin{pmatrix} 1 & 3 & 5 \\ 4 & 5 & 6 \\ -1 & -1 & 2 \end{pmatrix} - b\det\begin{pmatrix} 1 & 3 & 5 \\ 2 & 5 & 6 \\ 1 & -1 & 2 \end{pmatrix} + c\det\begin{pmatrix} 1 & 1 & 5 \\ 2 & 4 & 6 \\ 1 & -1 & 2 \end{pmatrix} - d\det\begin{pmatrix} 1 & 1 & 3 \\ 2 & 4 & 5 \\ 1 & -1 & -1 \end{pmatrix}.
\end{aligned}$$

Now we've reduced our problem of finding a 4×4 determinant to a problem of finding a 3×3 determinant. We know how to do this (just as we did in the last example) by "expanding" along the first row. Once we do this to each matrix, we should get

$$\det(Y) = -21a + 13b - 14c + 10d.$$

This gives an iterative process to compute the determinant, which we'll call the COFACTOR EXAPANSION. We need a definition.

Definition 4.17. For an $n \times n$ matrix $A = (a_{ij})$, the ij-MINOR of A is defined to be the $(n-1) \times (n-1)$ matrix A_{ij} with the *ith* row and jth column deleted.

Theorem 4.18 (Cofactor Expansion). *The determimant of an $n \times n$ matrix $A = (a_{ij})$ is well-defined and equal to*

$$\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots + (-1)^{n+1}a_{1n}\det(A_{1n}).$$

Proof. We'll skip the details of this proof, since it's messy to write in general. But the basic idea is a generalization of what we saw in Lemma 4.15 and the example above. Since we can recursively apply this strategy to reach a 2×2 matrix, which we know is well-defined, then the determinant of any $n \times n$ matrix is well-defined. □

Corollary 4.19. Let A be an $n \times n$ matrix. Then, the volume of the parallelepiped formed by the columns (or rows) of A is equal to $|\det(A)|$.

Proof. Define a function

$$\mathrm{vol} : \mathbf{Mat}_n(\mathbb{R}) \to \mathbb{R}$$

by letting $\mathrm{vol}(A)$ be the volume of the fundamental parallelepiped formed by the *rows* of A. By Lemma 4.4, vol satisfies all of the properties of Definition 4.6, except for Property (3), and so by Theorem 4.18 we get $\mathrm{vol}(A) = |\det(A)|$ (noting that we need an absolute value to account for the negatives introduced in our determinant).

Since $\det(A^\top) = \det(A)$ then we know that $\mathrm{vol}(A)$ is equal to the volume of the parallelepiped formed by the *columns* of A. $\square$

Example 4.20. Let
$$A = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 0 & 1 \\ 1 & -1 & 2 \end{pmatrix}.$$
To find the determinant of A, we can use cofactor expansion written as above. Or, we can be slightly more crafty and first swap the first and second rows (since the zero in the second row is going to simplify our computation). We have
$$\det(A) = -\det \begin{pmatrix} 2 & 0 & 1 \\ 1 & 3 & 5 \\ 1 & -1 & 2 \end{pmatrix}.$$
Now, we can use the cofactor expansion formula, as below
$$\begin{aligned}
\det \begin{pmatrix} 2 & 0 & 1 \\ 1 & 3 & 5 \\ 1 & -1 & 2 \end{pmatrix} &= 2 \cdot \det \begin{pmatrix} \boxed{2} & 0 & 1 \\ 1 & 3 & 5 \\ 1 & -1 & 2 \end{pmatrix} - 0 \cdot \det \begin{pmatrix} 2 & \boxed{0} & 1 \\ 1 & 3 & 5 \\ 1 & -1 & 2 \end{pmatrix} + 1 \cdot \det \begin{pmatrix} 2 & 0 & \boxed{1} \\ 1 & 3 & 5 \\ 1 & -1 & 2 \end{pmatrix} \\
&= 2 \cdot \det \begin{pmatrix} 3 & 5 \\ -1 & 2 \end{pmatrix} - 0 \cdot \det \begin{pmatrix} 1 & 5 \\ 1 & 2 \end{pmatrix} + 1 \cdot \det \begin{pmatrix} 1 & 3 \\ 1 & -1 \end{pmatrix} \\
&= 2 \cdot (6+5) - 0 \cdot (2-5) + 1 \cdot (-1-3) \\
&= 22 - 4 = 18,
\end{aligned}$$
and so $\boxed{\det(A) = -18}$.

4.17. The Characteristic Equation

The following definitions will turn out to have some major applications in understanding the structure of square matrices.

Definition 4.21. Let A be an $n \times n$ matrix. A non-zero vector $\vec{x}$ is an EIGENVECTOR of A if there is a scalar λ such that $A\vec{x} = \lambda\vec{x}$. The scalar λ is called an EIGENVALUE of A.

We'll pause as a class to watch the first few minutes of this video to help build some geometric intuition for these objects before diving into computation. Note that for the speaker of this video
$$\hat{\imath} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \hat{\jmath} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$
That is, $\hat{\imath}$ and $\hat{\jmath}$ denote the standard basis vectors. Let's look at one of the examples from Activity 3.30.

Example 4.22. Let
$$A = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix} \text{ and } \vec{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$
Then we have
$$A\vec{x} = \begin{pmatrix} -4 \\ 2 \end{pmatrix} = 2\vec{x}.$$

So, $\vec{x}$ is an eigenvector of A, with corresponding eigenvalue $\lambda = 2$. As we saw in the video, geometrically this tells us that the matrix transformation T_A stretches the vector $\vec{x}$ by a factor of 2.

Next, let's work on finding a method to compute eigenvalues and their corresponding eigenvectors.

Proposition 4.23. For an $n \times n$ matrix A, the set of eigenvectors of A corresponding to an eigenvalue λ is equal to $\mathrm{Nul}(A - \lambda I_n)$.

Proof. Observe that any $\vec{x} \in \mathrm{Nul}(A - \lambda I_n)$ if and only if

$$\begin{aligned} &(A - \lambda I_n)\vec{x} = \vec{0} \\ \Leftrightarrow\ &A\vec{x} - \lambda\vec{x} = \vec{0} \\ \Leftrightarrow\ &A\vec{x} = \lambda\vec{x}. \end{aligned}$$

□

Definition 4.24. We call the space $\mathrm{Nul}(A - \lambda I_n)$ the λ-EIGENSPACE of A, denote E_λ. Note that we can think of the λ-eigenspace as the vectors $\vec{x} \in \mathbb{R}^n$ so that the matrix transformation T_A stretches $\vec{x}$ by a factor of λ.

Example 4.25. Let A be as in Example 4.22, and recall that we showed $\lambda = 2$ is an eigenvalue of A. To find the 2-eigenspace of A we need to find the null space of $A - 2I_2$. We have

$$A - 2I_2 = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix},$$

which is row equivalent to

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

So, if $\begin{pmatrix} x \\ y \end{pmatrix} \in E_2$ then we have

$$\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \vec{0},$$

which gives $x = -2y$. So, the 2-eigenspace is given by

$$E_2 = \mathrm{Span}\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right).$$

This gives a complete characterization of finding the eigenvectors corresponding to a given eigenvalue. So the natural next question is: how do we determine the eigenvalues of a matrix A? We have the following Lemma from your activity.

Lemma 4.26. *A scalar λ is an eigenvalue of A if and only if there is a nonzero solution to $(A - \lambda I_n)\vec{x} = \vec{0}$.*

Proof. Suppose first that λ is an eigenvalue of A. Recall that eigenvectors are *nonzero*, so there must be a corresponding nonzero vector $\vec{x}$ in $\mathbb{R}^n$ so that

$$A\vec{x} = \lambda\vec{x} \Rightarrow (A - \lambda I_n)\vec{x} = \vec{0}.$$

Conversely, if there's a nonzero solution to $(A - \lambda I_n)\vec{x} = \vec{0}$ then rearranging this equation gives $A\vec{x} = \lambda\vec{x}$, and so λ is an eigenvalue corresponding to the eigenvector $\vec{x}$. $\square$

Example 4.27. Let

$$A = \begin{pmatrix} 6 & -2 \\ 2 & 1 \end{pmatrix}$$

as in P6 of Activity 3.30. By the lemma above, to determine whether $\lambda = 1$ is an eigenvalue of A, we need to check whether

$$(A - \lambda I_n)\vec{x} = \vec{0} \tag{4.3}$$

has a nonzero solution when $\lambda = 1$. We have

$$\det(A - 1 \cdot I) = \det \begin{pmatrix} 5 & -1 \\ 1 & 1 \end{pmatrix} = 7 \neq 0$$

so $A - 1 \cdot I_n$ is invertible. But this means that equation (4.3) has a unique solution! That is, there does *not* exist a nontrivial solution to (4.3), and so $\lambda = 1$ is *not* an eigenvalue. If we instead look at equation (4.3) when $\lambda = 5$ we have

$$\det(A - \lambda I_2) = \det \begin{pmatrix} 1 & -2 \\ 2 & -4 \end{pmatrix} = 0.$$

So, $A - \lambda I_2$ is *not* invertible, which means equation (4.3) *does* have a nontrivial solution, and so $\lambda = 2$ is an eigenvalue.

In general, the homogeneous equation

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

has a solution precisely when $A - \lambda I_n$ has a column without a pivot. That is, $A - \lambda I_n$ is not invertible. By Proposition 4.11, this occurs precisely when $\det(A - \lambda I_n) = 0$. This motivates the following definition.

Definition 4.28. For an $n \times n$ matrix A,

$$\chi_A(x) = \det(A - xI_n).$$

is called the CHARACTERISTIC POLYNOMIAL of A.

Observe that $\chi_A(x)$ is a polynomial of degree n. The proof of this in general would use an inductive argument, along with the cofactor expansion formula for the determinant. Instead of worrying about understanding this formally, note that:

(1) If we look at the cofactor formula for the determinant, we see that the only operations happening are addition and multiplication, and so we end up with some algebraic expression made up of sums and products of real numbers and our unknown x, which precisely defines a polynomial.

(2) If an $n \times n$ matrix A has diagonal entries $d_1, d_2, \ldots, d_n$, then the highest degree term coming out of the cofactor exapansion will be $(d_1 - x)(d_2 - x) \cdots (d_n - x)$ (convince yourself of this in the 3×3 case). So, the degree of $\chi_A(x)$ will be at most n, and in fact equal to n when none of the diagonal entries are equal to 0.

From our discussion above, we have the following.

Proposition 4.29. The set of eigenvalues of an $n \times n$ matrix A is equal to the roots of the characteristic polynomial χ_A of A. In particular, there are at most n eigenvalues of A.

Note that the fact that A has n eigenvalues follows from the Fundamental Theorem of Algebra, which states that any degree n polynomial has at most n roots.

Example 4.30. Let's look at the matrices B and C from Activity 4.31. First, let's find $\chi_B(x)$. We have

$$\begin{aligned}
\chi_B(x) &= \det \begin{pmatrix} 3-x & -2 & 5 \\ 1 & -x & 7 \\ 0 & 0 & 1-x \end{pmatrix} \\
&= \det \begin{pmatrix} 0 & 0 & 1-x \\ 3-x & -2 & 5 \\ 1 & -x & 7 \end{pmatrix}, \text{ swapping } R_2 \leftrightarrow R_3, \text{ then } R_1 \leftrightarrow R_2 \\
&= (1-x) \det \begin{pmatrix} 3-x & -2 \\ 1 & -x \end{pmatrix} \\
&= (1-x)\left(x(x-3)+2\right) \\
&= -x^3 + 4x^2 - 5x + 2.
\end{aligned}$$

So, $\chi_B(x) = -x^3 + 4x^2 - 5x + 2$. To find the eigenvalues of A, we need to solve

$$-x^3 + 4x^2 - 5x + 2 = 0.$$

Observe that we can factor the left hand side to get

$$-(x-2)(x-1)^2 = 0$$

and so the eigenvalues are $\lambda = 2$ and $\lambda = 1$.

Next, recall that

$$C = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

So,

$$\chi_C(x) = \begin{pmatrix} 1-x & -1 \\ 1 & 1-x \end{pmatrix} = (1-x)^2 + 1.$$

So, $\chi_C(x) = (1-x)^2 + 1$ (let's leave it in this form), which means the eigenvalues of C will be solutions to the equation

$$\begin{aligned}
&(1-x)^2 + 1 = 0 \\
\Rightarrow\ &(1-x)^2 = -1.
\end{aligned}$$

There are no real solutions to this equation, but there are complex ones. In particular, we have the complex eigenvalues

$$\lambda = i - 1, -i - 1.$$

Note that in general, eigenvalues can be complex. This story is a bit tricky, because the corresponding eigenvectors will also be complex, and we've only been working with *real* vector spaces in this class. We're going to sweep this story under the rug for now, and only focus on examples where are eigenvalues are real. Time permitting, we'll return to this story (in particular, we'll spend time thinking about what this means geometrically).

The following example illustrates how eigenvalues can help us understand how a given matrix transforms a vector.

Example 4.31. Consider the matrix

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

Let's use some eigentheory to understand the matrix transformation T_A. First, we compute

$$\chi_A(x) = \det \begin{pmatrix} 1-x & 0 & 1 \\ 0 & 1-x & 1 \\ 0 & 0 & 2-x \end{pmatrix} = (1-x)^2(2-x)$$

and so $\chi_A(x) = (1-x)^2(2-x)$. This tells us that A has eigenvalues $\lambda = 1$ and $\lambda = 2$. (Note: on your homework you'll show that the eigenvectors of a triangular matrix are always equal to the diagonal entries). One we've found our eigenvalues, we can now compute the set of eigenvectors as the null space of $A - \lambda I_3$. With a bit of work, we compute $E_1 = \mathrm{Nul}(A - I_3) = \mathrm{Span}\,(\vec{v}_1 . \vec{v}_2)$, where

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

and $E_2 = \mathrm{Nul}(A - 2 \cdot I_3) = \mathrm{Span}\,(\vec{v}_3)$, where

$$\vec{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Observe that $\vec{v}_1, \vec{v}_2, \vec{v}_3$ is a linearly independent set, and so for any $\vec{v} \in \mathbb{R}^3$ we can write

$$\vec{v} = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3.$$

Using linearity of the matrix transformation T_A gives

$$\begin{aligned} T_A(\vec{x}) &= x_1 T_A(\vec{v}_1) + x_2 T_A(\vec{v}_2) + x_3 T_A(\vec{v}_3) \\ &= (1)(x_1\vec{v}_1 + x_2\vec{v}_2) + 2(x_3\vec{v}_3). \end{aligned}$$

where the second equality follows by recalling that $\vec{v}_1, \vec{v}_2 \in E_1$ and $\vec{v}_3 \in E_2$. This tells us that, if instead of plotting our vectors on the xyz-plane as usual, we plot points on the "$\vec{v}_1, \vec{v}_2, \vec{v}_3$"-plane, we could more easily understand this matrix transformation (it's the transformation that does nothing to the first two coordinates, and scales the third coordinate by 2). This is the perspective we'd like to build to in general. Let's add a bit more rigor to build on this idea.

4.18. Diagonalization

In the previous example, we saw that it was convenient to view vectors in terms of a particular basis. Let's develop this idea further.

4.18.1. Coordinate Systems.

Definition 4.32. Let $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be a basis for a vector space V. Recall that every vector $\vec{v}$ in V can be written in the form

$$\vec{v} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n.$$

The COORDINATES of $\vec{x}$ with respect to the basis $\mathcal{B}$ is given by

$$[\vec{x}]_{\mathcal{B}} := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Remark 4.33. Observe that if $B = \{\vec{e}_1, \dots, \vec{e}_n\}$ is the standard basis for $\mathbb{R}^n$, then

$$\left[\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right]_B = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

That is, when we talk about the coordinates of a vector (without referencing any specific basis), we really mean the coordinates of that vector *with respect to the standard basis.*

Example 4.34. Let

$$\mathcal{B} = \left\{ \begin{pmatrix} -1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}. \right\}$$

Observe that $\mathcal{B}$ is a basis for $\mathbb{R}^2$. Let's find the coordinates of the vector $\vec{x} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$ with respect to $\mathcal{B}$. Since $\mathcal{B}$ is a basis, we can write

$$\vec{x} = x_1 \begin{pmatrix} -1 \\ 4 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Now, let

$$C = \begin{pmatrix} -1 & 2 \\ 4 & 0 \end{pmatrix}.$$

Note that C is invertible with

$$C^{-1} = \frac{1}{8} \begin{pmatrix} 0 & 2 \\ 4 & 1 \end{pmatrix}.$$

So, we have

$$\vec{x} = C \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = C^{-1}\vec{x}$$

which gives

$$\left[\begin{pmatrix} 0 \\ 6 \end{pmatrix}\right]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 3/4 \end{pmatrix}.$$

The following lemma generalizes this strategy.

Lemma 4.35 (Changing Basis)**.** *Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$. Then, for any vector $\vec{x} \in \mathbb{R}^n$ we have*

$$[\vec{x}]_{\mathcal{B}} = C^{-1}\vec{x},$$

where $C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}$.

Proof. Since $\mathcal{B}$ is a basis for $\mathbb{R}^n$, we can write any vector $\vec{x}$ in the form

$$\vec{x} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n = C\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

where C is the matrix with column vectors $\vec{v}_1, \ldots, \vec{v}_n$. Since $\mathcal{B}$ forms a basis for $\mathbb{R}^n$ the column vectors of C are linearly independent, which implies that C is invertible. So we can write

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = C^{-1}\vec{x}.$$

The vector on the left is precisely the definition of the coordinates of $\vec{x}$ with respect to the basis $\mathcal{B}$, and so $[\vec{x}]_{\mathcal{B}} = C^{-1}\vec{x}$ as desired. □

Remark 4.36. As we saw in lecture, I find this notation difficult to work with. For the rest of the lecture notes, I'll abandon this notation, and instead just write our vectors as linear combinations of the desired basis. Conceptually, we should remember that different bases define different coordinate systems on $\mathbb{R}^n$. Our goal in this chapter is to figure out what coordinate systems help us best understand a given linear transformation.

4.18.2. Matrix Similarity.

Example 4.37. Let A and $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ be as in Example 4.31. Recall that

$$T_A\left(x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3\right) = x_1\vec{v}_1 + x_2\vec{v}_2 + 2x_3\vec{v}_3.$$

Let $C = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{pmatrix}$. Then the above equality is identical to

$$\begin{aligned} T_A\left(C\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) &= C\begin{pmatrix} x_1 \\ x_2 \\ 2x_3 \end{pmatrix} \\ \Leftrightarrow AC\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= C\begin{pmatrix} x_1 \\ x_2 \\ 2x_3 \end{pmatrix} \\ \Leftrightarrow C^{-1}AC\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} x_1 \\ x_2 \\ 2x_3 \end{pmatrix} \\ \Leftrightarrow T_{C^{-1}AC}\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) &= \begin{pmatrix} x_1 \\ x_2 \\ 2x_3 \end{pmatrix}. \end{aligned}$$

That is,

$$T_A(x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3) = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3$$

if and only if

$$T_B(x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3) = x_1\vec{e}_1 + x_2\vec{e}_2 + 2x_3\vec{e}_3,$$

where $B = C^{-1}AC$. That is, the matrices A and B behave similarly with respect to different bases for $\mathbb{R}^3$. This gives rise to the following definition.

Definition 4.38. Two $n \times n$ matrices X and Y are called SIMILAR if there exists an invertible matrix C so that

$$X = C^{-1}YC.$$

In this case, we write $X \sim Y$. Note that if X is similar to Y, then Y is similar to X, since we can write

$$Y = (C^{-1})^{-1}X(C^{-1}).$$

The following Proposition generalizes what we observed in the previous example.

Proposition 4.39. Let A and B be similar matrices, so that $B = C^{-1}AC$ for an invertible matrix C. Let $\mathcal{B}$ be the basis for $\mathbb{R}^n$ defined by the columns of C. Then, for any $\vec{x}$ in $\mathbb{R}^n$ we have

$$T_A(x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n$$

if and only if

$$T_B(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) = a_1\vec{e}_1 + \cdots + a_n\vec{e}_n$$

where $\{\vec{e}_1, \ldots, \vec{e}_n\}$ is the standard basis for $\mathbb{R}^n$. That is,

$$T_A\left([\vec{x}]_{\mathcal{B}}\right) = \left[T_A(\vec{x})\right]_{\mathcal{B}}.$$

We should read this equality as saying that the matrix tranformations T_A and T_B act the same with respect to different coordinate systems.

Proof. This proof follows similarly to our computation in the previous example. We have

$$\begin{aligned}
T_A(x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) &= a_1\vec{v}_1 + \cdots + a_n\vec{v}_n. \\
\Leftrightarrow T_A\left(C\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) &= C\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \\
\Leftrightarrow AC\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &= C\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \\
\Leftrightarrow C^{-1}AC\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &= \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \\
\Leftrightarrow T_B\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) &= C\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \\
\Leftrightarrow T_B(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) &= a_1\vec{e}_1 + \cdots + a_n\vec{e}_n.
\end{aligned}$$

$\square$

In Example 4.31, we observed that the matrix A was similar to

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

Matrices of this form define particularly nice matrix transformations, since the transformation just scales each coordinate of our vector. We have the following definition.

Definition 4.40. An $n \times n$ matrix D is called DIAGONAL if the only nonzero entries lie on the diagonal of D. That is,

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}.$$

In this case, we write $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$.

4.18.3. Diagonalization. Note that a matrix transformation defined by a diagonal matrix is particularly simple to understand: if

$$D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$$

then

$$D \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \\ d_n x_n \end{pmatrix}.$$

This leads the question: what matrices A are similar to a diagonal matrix? We have the following definition.

Definition 4.41. An $n \times n$ matrix A is called DIAGONALIZABLE if it is similar to a diagonal matrix.

By Proposition 4.39, if a matrix is diagonalizable, then we can find a coordinate system to conveniently view our matrix transformation, just like we saw in Example 4.31. The remaining results in this section give a characterization of such matrices. The following result generalizes what we observed in Example 4.31.

Theorem 4.42 (The Diagonalization Theorem)**.** *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. Furthermore, if A is diagonalizable with linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$ then $D = C^{-1}AC$ where*

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \text{ and } C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

Proof. First, suppose that A has n linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$ and let C be the matrix

$$C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

Since the columns of C are linearly independent, we know that C is invertible. To compute the matrix $C^{-1}AC$, we'll multiply by the standard basis vectors $\vec{e}_i$, which exactly pick out the ith column, as you observed in Activity 4.34. So, if $\vec{e}_i$ is the ith standard basis vector, note that

$$C\vec{e}_i = \vec{v}_i \Rightarrow \vec{e}_i = C^{-1}\vec{v}_i. \tag{4.4}$$

We have

$$\begin{aligned} C^{-1}AC\vec{e}_i &= C^{-1}A\vec{v}_i, \text{ by Equation (4.4)} \\ &= C^{-1}\lambda_i\vec{v}_i, \text{ since } \vec{v}_i \text{ is an eigenvector of } \lambda_i \\ &= \lambda_i C^{-1}\vec{v}_i \\ &= \lambda_i\vec{e}_i, \text{ by Equation (4.4).} \end{aligned}$$

Since the ith column of $C^{-1}AC = \lambda_i\vec{e}_i$ we have that

$$C^{-1}AC = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

as desired. Conversely, suppose that A is diagonalizable, so that

$$C^{-1}AC = D$$

for an invertible matrix C and diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$. Suppose that C has column vectors equal to $\vec{c}_i$. Since C is invertible, we know that $\{\vec{c}_1, \ldots, \vec{c}_n\}$ is a linearly independent set. So, we just need to show that $\vec{c}_i$ is an eigenvector with eigenvalue d_i. As we observed before, note that

$$C\vec{e}_i = \vec{c}_i \Rightarrow \vec{e}_i = C^{-1}\vec{c}_i. \tag{4.5}$$

So we have

$$\begin{aligned} A\vec{c}_i &= CDC^{-1}\vec{c}_i \\ &= CD\vec{e}_i, \text{ by Equation (4.5)} \\ &= Cd_i\vec{e}_i, \text{ since the } i\text{th column of } D \text{ is } d_i\vec{e}_i \\ &= d_iC\vec{e}_i \\ &= d_i\vec{c}_i, \text{ by Equation (4.5)} \end{aligned}$$

So, $\vec{c}_i$ is an eigenvector of A with eigenvalue d_i as desired. □

Remark 4.43. The Diagonalization Theorem along with Proposition 4.39 tells us that when A is diagonalizable, then if we define coordinates for $\mathbb{R}^n$ in terms of any n linearly independent eigenvectors, then the matrix transformation T_A behaves like a diagonal matrix on this coordinate system. There is one other useful application of diagonalization you observed in your activity today, which I'll state as a Proposition below.

Proposition 4.44. Let A be a diagonalizable matrix, with diagonalization $D = C^{-1}AC$. Then,

$$A^n = CD^nC^{-1}$$

for any integer n.

Proof. We have

$$\begin{aligned}
A^n &= (CDC^{-1})^n \\
&= \underbrace{(CDC^{-1})(CDC^{-1})\cdots(CDC^{-1})}_{n \text{ times}} \\
&= \underbrace{CD(C^{-1}C)DC^{-1}\cdots CDC^{-1}}_{n \text{ times}} \\
&= C\underbrace{DD\cdots D}_{n \text{ times}}C^{-1} \\
&= CD^nC^{-1}.
\end{aligned}$$

□

Example 4.45. Consider the matrix

$$A_1 = \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}$$

from Activity 4.34. Observe that A_1 has eigenvalues $\lambda = -1$ and $\lambda = 2$ and

$$E_{-1} = \text{Span}\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right), E_2 = \text{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right).$$

From the Diagonalization Theorem, we can write $D = C^{-1}AC$ where

$$D = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}.$$

So, we have

$$A^{10} = C\begin{pmatrix} (-1)^{10} & 0 \\ 0 & 2^{10} \end{pmatrix}C^{-1}.$$

We compute

$$C^{-1} = \frac{1}{3}\begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}$$

and so

$$A^{10} = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1024 \end{pmatrix}\begin{pmatrix} 2/3 & -1/3 \\ 1/3 & 1/3 \end{pmatrix} = \begin{pmatrix} 342 & 341 \\ 682 & 683 \end{pmatrix}.$$

To use the Diagonalization Theorem, we need a method to determine whether A has enough linearly independent eigenvectors. We have the following Proposition, which gives a special case.

Proposition 4.46. Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of a matrix A, and suppose that $\vec{v}_i \in E_{\lambda_i}$ for each $i \in \{1, \ldots, k\}$. Then $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is a linearly independent set.

Proof. On your next homework, you'll show that any two eigenvectors coming from distinct eigenvalues are linearly independent. The rest of this proof follows by mathematical induction, similar to what you've shown. I'll omit the details of the proof here, since this isn't a method we've discussed together. If you'd like to see the details, feel free to stop by office hour. □

Corollary 4.47. If an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.

Proof. If A has n distinct eigenvalues, then by Proposition 4.44 A must then have n distinct eigenvectors, and so A is diagonalizable by the Diagonalization Theorem. $\square$

Remark 4.48. Note that the converse of Proposition 4.44 does not hold. That is, it's not the case that if an $n \times n$ matrix A is diagonalizable then A must have n distinct eigenvalues. For example, if we let A be the matrix defined in example 4.31 is diagonalizable but only has two eigenvalues. The following definition will help us characterize $n \times n$ diagonalizable matrices with less than n eigenvalues.

Definition 4.49. For an eigenvalue λ of a matrix A, the GEOMETRIC MULTIPLICITY of λ is defined to be the dimension of the λ-eigenspace E_λ.

Example 4.50. Let A be as in Example 4.31. Recall that A has eigenvalues $\lambda = 1$ and $\lambda = 2$ and that $\dim E_1 = 2$ and $\dim E_2 = 1$. So the geometric multiplicity of the eigenvalue $\lambda = 1$ is equal to 2, and the geometric multiplicity of the eigenvalue $\lambda = 2$ is equal to 1.

We can then rephrase the Diagonalization Theorem as follows.

Theorem 4.51 (Diagonalization Theorem, rephrasing)**.** *An $n \times n$ matrix A is diagonalizable if and only if the sum of the geometric multiplicities of the eigenvalues of A is equal to n.*

Note that we really done anything in this rephrasing but give new terminology, so there's nothing new to prove.

Example 4.52. Recall the setting of Example 4.31. If

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

then we found that

$$\chi_A(x) = (1-x)^2(2-x),$$

and that E_1 had dimension 2 and E_2 had dimension 1. In this case, observe that the dimension of the λ-eigenspace exactly matches the power on the term $(\lambda - x)$ in χ_A. This turns out to always be the case for diagonalizable matrices. We have the following definition.

Definition 4.53. Suppose that a matrix A has eigenvalue λ. The ALGEBRAIC MULTIPLICITY of λ is the largest integer m so that $(x-\lambda)^m$ divides the characteristic polynomial χ_A of A.

Example 4.54. In Example 4.31, the algebraic multiplicity of $\lambda = 1$ is 2, and the algebraic multiplicity of $\lambda = 2$ is 1.

We can now add the our Diagonalization Theorem. We have the following.

Theorem 4.55 (Diagonalization Theorem, final version)**.** *Let A be an $n \times n$ matrix. The following are equivalent.*

(1) A is diagonalizable;

(2) The sum of the geometric multiplicities of A is equal to n;

(3) The geometric multiplicity of every eigenvalue λ is equal to the algebraic multiplicity of λ.

Furthermore, if A is diagonalizable with linearly independent eigenvectors $\vec{v}_1, \dots, \vec{v}_n$ corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$ then $D = C^{-1}AC$ where

$$D = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \text{ and } C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

All that's left to show in this final version is (2) $\Leftrightarrow$ (3). We first need a Lemma.

Lemma 4.56. *The geometric multiplicity of any eigenvalue is less than or equal to its algebraic multiplicity.*

We'll omit the proof of this Lemma (there are ways to prove this with the information we have, but they're all a bit complicated... I'll keep searching for a better alternative). Let's see how this proves our Theorem.

Proof of Theorem 4.55. Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \dots, \lambda_k$ and let λ_i have geometric multiplicity d_i and algebraic multiplicity m_i. By the previous lemma, we know that $d_i \le m_i$.

(2) $\Rightarrow$ (3): Suppose that $d_1 + \cdots + d_n = n$. Note that the characteristic polynomial of A is of degree n, and since we can write

$$\chi_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

this gives $m_1 + \cdots + m_k = n$. So, if $d_i < m_i$ for any i we would have

$$d_1 + \cdots + d_k < m_1 + \cdots + m_k < n,$$

contradicting our assumption. Since by our Lemma $d_i \le m_i$, we must then have the equality $d_i = m_i$ for every i.

(3) $\Rightarrow$ (2): Conversely, suppose that $m_i = d_i$ for all i. Then we have

$$n = m_1 + \cdots m_k = d_1 + \cdots + d_k,$$

as desired. □

Remark 4.57. Note that the addition of part (3) to our Theorem doesn't quite simplify our computation when our matrix *is* diagonalizable: no matter what we do, we still need to compute the dimension of the λ-eigenspace for each value of λ. But it does give the potential to simplify our justification that a matrix *is not* diagonalizable: all we need to do is find ONE eigenvalue where the geometric and algebraic multiplicities do not agree.

Example 4.58. Consider the matrix

$$A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{pmatrix}$$

from Activity 4.34. The characteristic polynomial of A_2 is given by

$$\chi_{A_3}(x) = -(x-2)(x-1)^2$$

and so $\lambda = 1$ has algebraic multiplicity 2. However,

$$A_2 - 1 \cdot I_3 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 2 & -5 & 3 \end{pmatrix}$$

which is row equivalent to a matrix with 2 pivots. So, $\text{nullity}(A - I_3) = 1$ which tells us that the geometric multiplicity of $\lambda = 1$ is equal to 1. So, by our final version of the Diagonalization Theorem, we know that A_2 is not diagonalizable.

Chapter 6

Orthogonality

So far, we've focused on building an algebraic understanding of $\mathbb{R}^n$ as a vector space, by equipping the set of vectors in $\mathbb{R}^n$ with the algebraic operations of vector addition and scalar multiplication. While this perspective has helped us to understand many of the algebraic properties of Euclidean space, we've left out two important features: angles and distances. In this chapter, we see how the dot product gives $\mathbb{R}^n$ the structure of an "inner product space", which will fill in these missing features. We will then see how our old algebraic understanding of $\mathbb{R}^n$ interacts with these new measurements.

6.27. The Dot Product

Definition 6.1. Let

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \text{ and } \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

be vectors in R^n. The DOT PRODUCT of $\vec{u}$ and $\vec{v}$ is the scalar

$$\vec{u} \cdot \vec{v} := u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

On your next homework, you'll show the dot product satisfies the following properties.

Proposition 6.2. Let $\vec{u}, \vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then,

(1) *Commutativity:* $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
(2) *Distributivity with Addition:* $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$
(3) *Distributivity with Scalar Multiplication:* $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v})$

This innocent looking operation is all we need to give some geometric structure to our understanding of $\mathbb{R}^n$. In fact, it turns out that the dot product defines something called an INNER PRODUCT on $\mathbb{R}^n$. We'll omit these details here, but inner products are the operations needed to define notions of distance and angles on more general vector spaces.

Definition 6.3. The NORM of $\vec{u}$ is defined by

$$\|\vec{u}\| := \vec{u} \cdot \vec{u}.$$

Observe that, for any vector $\vec{u}$, we have

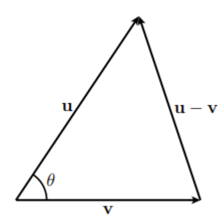
$$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

and so the norm measures the length of a given vector. This allows us to define the distance between vectors.

Definition 6.4. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^n$. Then, the DISTANCE between $\vec{u}$ and $\vec{v}$, denoted $d(\vec{u}, \vec{v})$ is equal to the length of the vector $\vec{u} - \vec{v}$. That is,

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

Definition 6.5. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^n$. Then, the ANGLE between vectors $\vec{u}$ and $\vec{v}$ is defined as the angle θ in the triangle below



where this picture is taking place in the 2-dimensional plane spanned by $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$

We have the following.

Proposition 6.6. Let $\vec{u}$ and $\vec{v}$ be nonparallel vectors in $\mathbb{R}^n$. Then, the angle between $\vec{u}$ and $\vec{v}$ is the value $\theta \in (0, \pi]$ given by

$$\cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}$$

Proof. By Law of Cosines we have

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\| \cos\theta.$$

Furthermore, observe that for any vector $\vec{x}$ we have

$$\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}.$$

So, we get

$$(\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2\|\vec{u}\|\|\vec{v}\| \cos\theta.$$

Using Proposition 6.2 we have

$$\vec{u} \cdot \vec{u} - 2\vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{v} = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2\|\vec{u}\|\|\vec{v}\| \cos\theta$$
$$\Rightarrow -2\vec{u} \cdot \vec{v} = -2\|\vec{u}\|\|\vec{v}\| \cos\theta$$

$$\Rightarrow \cos\theta = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}. \qquad \square$$

Corollary 6.7. Vectors $\vec{u}$ and $\vec{v}$ are perpendicular (that is, the angle between them is equal to $90°$) if and only if $\vec{u}\cdot\vec{v} = 0$.

In the following section, we'll see that this notion of orthogonality is precisely what's needed for our algebraic and geometric notions to blend. =

6.28. Orthogonal Bases

In the previous chapters, we saw that the fundamental object needed to understand a vector space is a basis. We learned that real vector spaces of dimension n are all isomorphic ("the same" algebraically) to $\mathbb{R}^n$, and we saw how different bases define coordinate systems on our vector spaces which can help us better understand certain linear transformations (this is the eigen-story from the previous chapter). Let's look at how the dot product interacts with different bases for $\mathbb{R}^n$.

Example 6.8. Consider the basis $\mathcal{B}_1 = \{\vec{v}_1, \vec{v}_2\}$ where

$$\vec{v}_1 = \begin{pmatrix}1\\1\end{pmatrix}, \text{ and } \vec{v}_2 = \begin{pmatrix}2\\0\end{pmatrix}$$

and consider a vector $\vec{x}$ written in terms of this basis, say

$$[\vec{x}]_{\mathcal{B}_1} = \begin{pmatrix}-1\\1\end{pmatrix}.$$

Since $\mathcal{B}$ is not an orthogonal basis, if we wanted to compute something like $\|\vec{x}\|$ our best option would be to just rewrite our vector in terms of the standard basis and use our formulas above. The following Proposition will tell us which bases make our geometric computations similarly easy to computations with the standard basis.

Definition 6.9. A basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is ORTHOGONAL if $\vec{v}_i \cdot \vec{v}_j = 0$ for every $i \neq j$. If it's also the case that $\|\vec{v}_i\| = 1$ for every i we call $\mathcal{B}$ an ORTHONORMAL BASIS for $\mathbb{R}^n$.

Example 6.10. The basis

$$\mathcal{B}_1 = \left\{ \begin{pmatrix}2\\2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\-2\end{pmatrix}, \begin{pmatrix}-4\\5\\-2\end{pmatrix} \right\}$$

from Activity 5.35 is an orthogonal basis for $\mathbb{R}^3$, but it is not orthonormal. The standard basis for $\mathbb{R}^n$ is an orthonormal basis.

Proposition 6.11. Let $\mathcal{B}$ be an orthonormal basis for $\mathbb{R}^n$ and take any vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^n$. Then

$$[\vec{x}]_{\mathcal{B}} \cdot [\vec{y}]_{\mathcal{B}} = \vec{x}\cdot\vec{y}.$$

In particular, we have $\|\vec{x}\| = \|[\vec{x}]_{\mathcal{B}}\|$.

Proof. Suppose that $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ and write

$$[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ and } [\vec{y}]_{\mathcal{B}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

That is,

$$\vec{x} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$$
$$\vec{y} = y_1\vec{v}_1 + \cdots + y_n\vec{v}_n.$$

Then we have

$$\vec{x} \cdot \vec{y} = (x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) \cdot (y_1\vec{v}_1 + \cdots + y_n\vec{v}_n)$$

Using the distributive property of the dot product, we'll end up with a sum of terms of the form

$$x_i y_j \vec{v}_i \vec{v}_j.$$

But, since we know that $\vec{v}_i \cdot \vec{v}_j = 0$ whenever $i \neq j$ then we have

$$\vec{x} \cdot \vec{y} = x_1 y_1 \vec{v}_1 \cdot \vec{v}_1 + \cdots x_n y_n \vec{v}_n \cdot \vec{v}_n.$$

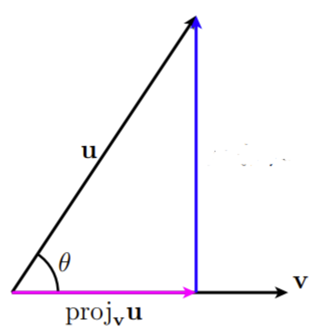
But we also know that $\vec{v}_i \cdot \vec{v}_i = \|\vec{v}\| = 1$, since our basis is orthonormal. So, we have

$$\vec{x} \cdot \vec{y} = x_1 y_1 + \cdots + x_n y_n$$

as desired. □

Proposition 6.11 tells us that orthonormal bases are the "right" kinds of bases to use when we want to understand $\mathbb{R}^n$ as an inner product space. So, the next natural question is, how can we find an orthonormal basis? To answer this question, we first need to make an observation.

Definition 6.12. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^n$. Then, the ORTHOGONAL PROJECTION of $\vec{u}$ onto the vector $\vec{v}$ is the vector parallel to $\vec{v}$ obtained by dropping a perpendicular line from $\vec{u}$ as in the picture below



We have the following.

Proposition 6.13. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^n$. Then,

$$\text{proj}_{\vec{v}}(\vec{u}) = \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}\,\vec{v}$$

Proof. Since $\text{proj}_{\vec{v}}\vec{u}$ is in the same direction as $\vec{v}$ we must have

$$\text{proj}_{\vec{v}}\vec{u} = c\vec{v},$$

for a positive scalar c and so $\|\text{proj}_{\vec{v}}\vec{u}\| = c\|\vec{v}\|$. We have

$$\cos(\theta) = \frac{\|\text{proj}_{\vec{v}}\vec{u}\|}{\|\vec{u}\|} = \frac{c\|\vec{v}\|}{\|\vec{u}\|}$$

and from Proposition 6.6 we know that

$$\cos(\theta) = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}.$$

This gives

$$\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|} = \frac{c\|\vec{v}\|}{\|\vec{u}\|} \Rightarrow c = \frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|^2}.$$

Since $\vec{v}\cdot\vec{v} = \|\vec{v}\|^2$, the desired equality follows. □

We need one more observation.

Proposition 6.14. For any vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$ we have $\vec{u} - \text{proj}_{\vec{v}}\vec{u}$ is orthogonal to $\vec{v}$.

Proof. We have

$$\begin{aligned}(\vec{u} - \text{proj}_{\vec{v}}\vec{u})\cdot\vec{v} &= (\vec{u} - \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v})\cdot\vec{v}\\ &= \vec{u}\cdot\vec{v} - \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}\cdot\vec{v}\\ &= \vec{u}\cdot\vec{v} - \vec{u}\cdot\vec{v}\\ &= 0.\end{aligned}$$

□

We can use the Proposition above to generate orthogonal basis. Note that the obtain an orthonormal basis, we just divide each orthogonal basis vector by its norm.

Algorithm(The Gram-Schmidt Process). Let V be a vector subspace of $\mathbb{R}^n$ with basis $\{\vec{u}_1, \ldots, \vec{u}_m\}$. Define the vectors

$$\begin{aligned}&\vec{v}_1 = \vec{u}_1\\ &\vec{v}_2 = \vec{u}_2 - \text{proj}_{\vec{v}_1}\vec{u}_2\\ &\vec{v}_3 = \vec{u}_3 - \text{proj}_{\vec{v}_1}\vec{u}_3 - \text{proj}_{\vec{v}_2}\vec{u}_3\\ &\vdots\\ &\vec{v}_m = \vec{u}_m - \text{proj}_{\vec{v}_1}\vec{u}_m - \text{proj}_{\vec{v}_2}\vec{u}_m - \vdots - \text{proj}_{\vec{v}_{m-1}}\vec{u}_m.\end{aligned}$$

Then, $\{\vec{v}_1, \ldots, \vec{v}_m\}$ is an orthogonal basis for V.

Remark 6.15. The proof of this is a bit messy, so I'm not going to type this up. We'll talk about the idea for this in class, so you can go back to that lecture recording to review this. Or, you can look at pages 532-533 of our text.

Remark 6.16. Suppose that we have an orthogonal basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for $\mathbb{R}^n$, and let A be the matrix with column vectors $\vec{v}_i$. Then $A^\top$ is the matrix with *rows* $\vec{v}_i$, and so observe that

$$A^\top A = \begin{pmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 & \cdots & \vec{v}_1 \cdot \vec{v}_n \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 & \cdots & \vec{v}_2 \cdot \vec{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_n \cdot \vec{v}_1 & \vec{v}_n \cdot \vec{v}_2 & \cdots & \vec{v}_n \cdot \vec{v}_n \end{pmatrix}.$$

But since $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is an orthonormal basis, we know that

$$\vec{v}_i \cdot \vec{v}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

So, from above we have $A^\top A = I_n$. This proves the following.

Proposition 6.17. Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$ and let A be the matrix with column vectors $\vec{v}_1, \ldots, \vec{v}_n$. Then $\mathcal{B}$ is orthonormal if and only if $A^{-1} = A^\top$.

This gives rise to the following (annoying) definition.

Definition 6.18. We call a matrix A ORTHOGONAL if $A^{-1} = A^\top$.

Remark 6.19. This definition is annoying, because orthogonal matrices aren't just those matrices with orthogonal column vectors, but rather with orthonormal column vectors. I don't know why we don't just call them orthonormal matrices. My guess is because matrices with column vectors that are orthogonal, but not orthonormal, don't have many nice properties so they don't get their own name.

Just for fun and in case you're interested, there are the Hadamard matrices, which are matrices with orthogonal (but not orthonormal) column vectors, and entries equal to ± 1. It can be shown that for an $n \times n$ Hadamard matrix H we have $HH^\top = nI_n$ so that $H^{-1} = (1/n)H^\top$.

Remark 6.20. Note that Proposition 6.17 doesn't give a more convenient method of checking whether a set of vectors is orthonormal – in fact, it's often the case that computing the inverse of a matrix is *more* computationally expensive than just computing dot products – but it does tell us that if you know a matrix is orthonormal, its inverse is easy to compute.

We have the following observation.

Corollary 6.21. Let A be an orthogonal matrix. Then $\det(A) = \pm 1$.

Proof. Since $AA^\top = I_n$ then $\det(AA^\top) = \det(I_n) = 1$ but we know that

$$\det(AA^\top) = \det(A)\det(A) = (\det(A))^2.$$

So, $(\det(A))^2 = 1 \Rightarrow \det(A) = \pm 1$. $\square$

6.29. Orthogonally Diagonalizable Matrices

Note that nothing in this section will be needed to complete your final exam, but I think it's a nice story to end the semester with.

Example 6.22. Consider the matrix
$$A = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}.$$
Observe that A has eigenvalues $\lambda_1 = -3$ and $\lambda_2 = 2$ and we have
$$E_{-3} = \mathrm{Span}(\vec{v}_1) \text{ and } E_2 = \mathrm{Span}(\vec{v}_2)$$
where
$$\vec{v}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \text{ and } \vec{v}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$
Since $\|\vec{v}_1\| = \sqrt{5}$ and $\|\vec{v}_2\| = \sqrt{5}$ we can normalize these vectors to obtain a basis $\mathcal{B} = \{\vec{u}_1, \vec{u}_2\}$ for $\mathbb{R}^2$ where
$$\vec{u}_1 = \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix} \text{ and } \vec{u}_2 = \begin{pmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}.$$
Furthermore, since $\vec{u}_1 \in E_{-3}$ and $\vec{u}_2 \in E_2$ then the basis $\mathcal{B}$ is **both** orthonormal (which is good for preserving distances and angles) and consists of eigenvectors (which is good for understanding the linear transformation T_A).

We have the following.

Definition 6.23. A square matrix A is ORTHOGONALLY DIAGONALIZABLE if there exists an orthogonal matrix Q and a diagonal matrix A so that $Q^\top AQ = D$.

Note that, an $n \times n$ matrix A being orthogonally diagonalizable is equivalent to the existence of an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A. This is a particularly nice situation, since orthonormal bases preserve the dot product, and bases consisting of eigenvectors help us understand the linear transformation T_A. The following result completely characterizes when we're in this situation. Recall that a matrix A is SYMMETRIC when $A = A^\top$.

Theorem 6.24 (The Spectral Theorem)**.** *An $n \times n$ matrix A is orthogonally diagonalizable if and only if it is symmetric.*

Proof. Suppose first that A is orthogonally diagonalizable. Then, there exists an orthogonal matrix Q and diagonal matrix D so that
$$Q^\top AQ = D.$$
This gives
$$QDQ^\top = QQ^\top AQQ^\top = A,$$
recalling that for orthogonal matrices we have $Q^\top = Q^{-1}$. But then we have
$$A^\top = (QDQ^\top)^\top = (Q^\top)^\top D^\top Q^\top = QDQ^\top = A$$
noting that $D^\top = D$ since diagonal matrices are always symmetric.

Our text has a proof for the converse of this result, but this requires a bit more machinery than we had time to cover. There's a clean proof by using an inductive argument I quite like from Poole's text. I'll go ahead and post that to our Canvas Page's tab, since the exposition in Poole is quite nice. □