JoYo's Linear Algebra Notes

Joseph Victor

September 23, 2013

Preface

Who This Book is For and What this Book is About

This book is for any math 51 student (or student of a comparable linear algebra course at another university). The target reader is someone who is not necessarily a huge math geek, but wants to do well in this course and is willing to put in a reasonable amount of work. This book is designed to help focus such a student's learning, fill in gaps left by the regular text book and ask the kind of questions that will stimulate the reader's understanding of linear algebra. Translation: I want to help make yourself battle-ready for exams and future courses and professional work (but secretly I want to make you think this is as cool as I do).

This book is meant to *supplement* the lectures and course text, not replace it. My hope is to explain the material in a clear way, emphasizing the connections between different parts of the book and the reasons we think linear algebra is *so cool*. I have tried to err on the side of being more verbose, since the course text is often rather terse, and I try to motivate each section with a problem or curiosity. Much of this text is devoted to conceptual exercises, as a sort of way of "Socratic Dialog". The purpose of these exercises is two-fold. The first purpose is the exams. The math 51 exams are notoriously different from the problem sets. The problem sets emphasize mechanically working with vectors and matrices, while the exams tend to emphasize conceptual understanding and synthesis. Thus if the only problems you do are on the problem sets, you will get really good at row-reducing matrices, but less good at using the ideas in new situations. The second purpose is that it is a well known fact that if you discover something for yourself, you will learn it better, so I attempt to ask you the right questions so you can really get this material into your soul.

The reader shouldn't need a whole lot of formal math knowledge or skill. I try to, as much as possible, "de-formalize" the math. The formal proofs, to me, are less important than the ideas and the connections between them.

I have been tutoring math 51 for over three years and using linear algebra in advanced mathematics and computer science course work and as an software engineer. I know this material like the back of my hand, and I've worked with such a huge number of diverse students that I feel like know where students get confused, and how students succeed. I think all that experience will help.

How to Use This Book

The best way of using this book would be to read the section after you go over it in class, making sure you can do *all* the exercises (especially the exam ones). The second best way to use this book is just to do the exercises from the section after you go over it in class, and only reading my blah-blah-blah expositions if you are having trouble. The third best way to use this book is just to, the week before the exam, go to the end of each section and do the "Exam Exercises", and then if you can't get certain exam questions, try reading the chapter.

Each chapter has an "Exam Exercises" section at the end with links to questions from the actual exams. This is *very important*! The exams are worth the vast majority of your grade in this class, but most students don't do practice exams until the week before the exam. This is actually an entirely reasonable thing to do, since if you just cracked open a practice exam a few weeks early, chances are, you wouldn't be able to do half of it since you haven't learned all the material yet. Even when the exam is coming up, just doing practice exams (though a great thing to do) can be inefficient, for several reasons. First of all, if you can't do a problem, you might not have any clue what skills/chapters you need to brush up on. Second of all, if you can't do a problem because you haven't learned skill X, and X depends on skill Y (which you haven't quite mastered either), you might waste a lot of time being confused about X when if you spent that time working on Y, you would be able to make progress and would feel much better about yourself. Third of all, its just plain intimidating. My hope organizing these exam questions in these ways will be more manageable, focused, and allow you to start working on skills which actually matter for your grade a little bit every night, rather than all in one week.

If all you do is try practice exam problems as you are learning the material, I think you will succeed.

A Word Exercises

For the most part, the exercises I provide should be learning experiences. I usually ask you to show or realize something. The thing I ask you to do will usually turn out to be a useful tool in your arsenal. In fact, most of the real "content" in this book is meant to be *discovered* through exercises, and many of the exercises rely on tools learned during other exercises. If you'd like, pretend you're playing an RPG game and you want to expand your character's skills and experience. To maximize your success in a given chapter, make sure you know how to do the exercises from the chapter before!

You shouldn't worry all too much about the formal mechanics of proofs. If I say to show something, I really just mean convince yourself it must be true. If you aren't sure whether your reasoning makes sense, try explaining it to a class mate, a dog or a tutor. It really only matters that you understand, so I try to "de-formalize" things by removing the logical "code words"; you shouldn't be tripping over set builders, "for-alls", "there exists", etc. The only time in this course when some logical chops might be required is in proving that things are subspaces, so a sort of template is provided.

There are several types of exercises in this book, which is a concept I stole from Ravi Vakil's awesome Algebraic Geometry book. Let me tell you about them

- Easy Exercises

  These should be very very short, and seem "obvious" after you do them. Usually I just want you to make a simple observation, make an easy connection, or write down the definition of something in a specific case and notice something about it. There is nothing, usually, to write down during these exercises, or if there is it will be some small thing in the margin. They are things that the textbook might take for granted and the lecturer might go over pretty fast because its so simple, but it is important that you come to the conclusions on their own. You shouldn't feel bad if one of these exercises takes a little while to do; time spent pondering now will make these observations second-nature when it counts. Just because the solution is small and "obvious-in-retrospect" doesn't mean it will be obvious the first time!

- Important Exercises

  These exercises are either to understand something very important which other things rest on, or to use a skill which is very important and other things rest on. If you are limited in time, do these. Again, like the easy exercises, they might take a little while to work out the first time, but after you do you will have internalized something that might come in useful later, and if you try to do them again it should be quick.

- Tricky Exercises

  These are usually optional, as they can take a bit longer and some more insights, but you should still try to do them.

Other Resources

Books are great because you can work with them any time and at any pace, but nothing can replace working with a person one-on-one. In addition to taking advantage of office hours with TA's and professors, you can stop by SUMO hours and work with the tutors. I was a SUMO tutor for years and I can tell you that the tutors are knowledgeable and excited to help you with math. Also, for what its worth, the SUMO tutors are never involved with grading, which makes some people more comfortable asking for help (although I assure you, the TA's and professors would never penalize you for asking for help). Anyways, the SUMO sessions and great, so if you want to go you can find the exact time and date on the math 51 website (or probably posted in the hallways).

Why Don't You have Multivariable Calculus Too?

Because, despite what you might think about somebody who writes math books in their spare time, I

Where Can I Give Feedback

PLEASE PLEASE PLEASE Tell me how I can improve this!

Shoot me an email at jvictor@stanford.edu. I would love comments like "you aren't very clear in this part", or "I want more exercises of this type", or "this exercise is too easy or too hard", or "one of your links is broken or points to a question that doesn't fit with the chapter". Or just tell me how great this is and how it helped you get an $A+$. Or tell me how much it sucks. Just tell me what you think and I'll try to improve it.

Contributing

We're on GitHub! The repo is public! Clone me with

```
git clone https://github.com/jvictor0/LinearAlegebraNotes.git
```

Contents

Part I

Vectors, Matrices and Systems of Equations

1 Vectors in $\mathbb{R}^n$

1.1 Motivation

By now, you probably feel pretty comfortable doing mathematics with a single variable. You can add, subtract, multiply, divide, calculate trigonometric functions and even take derivatives an integrals. You are probably quite used to the situation when you have one "independent variable", perhaps called x, and one "dependent variable", perhaps called $y = f(x)$. Perhaps you can even throw in a z or a w without panicking, but, unfortunately, many (if not most) problems arising naturally in every branch of science and engineering will have hundreds or thousands or millions of variables. One quickly becomes bogged down in complexity. We need a way to organize all this data which hopefully enlightens, informs and even inspires us. This will be perhaps the primary goal of a first course in linear algebra.

Back in the single variable days, we would say x is a real number, of $x \in \mathbb{R}$, and visualize x as living on the "number line" (which we will call $\mathbb{R}^1$). We might consider a pair of real numbers (x, y), and draw it as a point on a plane ($\mathbb{R}^2$), or a triple (x, y, z) and draw it as a point in space ($\mathbb{R}^3$). Obviously, we can draw these points on a plane or in space *even* if they don't represent an actual location; say if we want to plot money vs time, It still helps to think about it "geometrically". The brilliant-yet-simple observation is this: while one cannot so easily draw the point (x, y, z, w), there is absolutely nothing stopping us from writing down a 4-tuple, or a 5-tuple, or a 34-tuple, or a 3511-tuple.

Definition 1.1 (Vectors and the Vector Space $\mathbb{R}^n$)**.** *Define $\mathbb{R}^n$ to be the set of n-tuples of real numbers $(x_1, x_2, ..., x_n)$. We will often refer to $\mathbb{R}^n$ as a* vector space *and call the tuples in $\mathbb{R}^n$* vectors. *We will denote vectors by lower-case letters with arrows above them, like $\vec{x} = (x_1, x_2, ..., x_n)$. We will sometimes call a real number a* scalar, *to distinguish it from a vector. We will call the numbers in a vector the* coordinates.

Remark 1.2:
Notice I did not mention "magnitude and direction"[1]. An arrow with a certain length pointing in a certain direction is a great image to have in your head, but you might have trouble saying what you mean by "direction" when $n > 3$. Still, if the mnemonic might helps you visualize the situation at hand, then all the better.

Remark 1.3:
There isn't much of a difference between a scalar $x \in \mathbb{R}$ and a 1-dimensional vector $\vec{x} = (x_1) \in \mathbb{R}^1$. While it does not much matter, we will call it a vector $\vec{x}$ when we want to emphasize vector-like properties and a scalar x otherwise. When trying to understand $\mathbb{R}^n$ in general, always keep in mind what happens in $\mathbb{R}^1$. The situation will almost always be extremely simple, but possibly enlightening.

Remark 1.4:
Some people will tell you there's a difference between points and vectors. We will not make that distinction in these notes.

Unimportant Remark 1.5:
What is $\mathbb{R}^0$? If $\mathbb{R}^1$ is a line and $\mathbb{R}^2$ is a plane, then $\mathbb{R}^0$ should be a point. Therefor we say there is but one vector in $\mathbb{R}^0$, and we'll call it $\vec{0}$. Of course, this will never be important, so if you feel disturbed by 0-tuples, feel free to ignore $\mathbb{R}^0$.

1.2 Addition and Subtraction

We define addition of vectors as one might expect. Let $\vec{x}$ and $\vec{y}$ be two vectors in $\mathbb{R}^n$, that is,

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

[1] Like the nemesis from Despicable Me

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Then we define

$$\vec{x} + \vec{y} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

that is, we just add vectors “coordinate by coordinate”. Thus, in $\mathbb{R}^3$,

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix}$$

Subtraction is done in much the same way:

$$\vec{x} - \vec{y} = \begin{pmatrix} x_1 - y_1 \\ \vdots \\ x_n - y_n \end{pmatrix}$$

so

$$\begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}$$

Exercise 1.6:
Interpret addition of vectors in $\mathbb{R}^2$ geometrically. If $\vec{x}$ and $\vec{y}$ are two points in the plane, can you come up with a geometric rule for the position of $\vec{x} + \vec{y}$? What about $\vec{x} - \vec{y}$? (hint: what happens when you connect the dots $(0, 0)$, $\vec{x}$, $\vec{y}$ and $\vec{x} + \vec{y}$?)

Exercise 1.7:
Convince yourself that vector addition is “commutative”, that is $\vec{x} + \vec{y} = \vec{y} + \vec{x}$. Then convince yourself it is “associate”, that is $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$. If you’re unsure of how to begin, write out the right side and left side of each equation and notice they are the same. Use the fact that scalar addition is commutative and associative.

Let $\vec{0}$ be the vector with zero in each coordinate, that is

$$\vec{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Easy Exercise 1.8:
Let $\vec{x} \in \mathbb{R}^n$ be any vector. Show $\vec{x} + \vec{0} = \vec{x}$. Show that $\vec{x} - \vec{x} = \vec{0}$.

We see that addition and subtraction of vectors obey some laws we are already used to. Vector addition is just an extension of scalar addition; we’re just adding and subtracting in bulk!

1.3 Scalar Multiplication

There is nothing stopping us from defining multiplication or division “coordinate by coordinate”, but we will not do this, because we do not need to. While we don’t allow multiplying two vectors, we do allow multiplying a scalar by a vector. Let $c \in \mathbb{R}$ be a scalar and $\vec{x} \in \mathbb{R}^n$. Define $c \cdot \vec{x}$ by

$$c \cdot \vec{x} = \begin{pmatrix} c \cdot x_1 \\ \vdots \\ c \cdot x_n \end{pmatrix}$$

Important Easy Exercise 1.9:
Interpret scalar multiplication of vectors in $\mathbb{R}^2$ geometrically. If $\vec{x}$ is a vector and c a scalar, where is $c \cdot \vec{x}$? (hint: do $c > 0$, $c = 0$ and $c < 0$ separately).

Important Remark 1.10:
In the past, everything in sight has been a number, so one didn't need to worry about mixing apples and oranges. In linear algebra, there one must keep straight many different fruits: infinitely fruits many in fact! One cannot add a vector and a scalar, nor can you add a vector in $\mathbb{R}^n$ with a vector in $\mathbb{R}^m$ if $n \neq m$. A vector plus a vector is a vector of the same size. One cannot multiply vectors, but one can multiply a vector times a scalar and get a new vector of the same size. The following easy exercise should give you a chance to practice keeping track of types.

Important Easy Exercise 1.11:
The following laws should be easy to show.

a) $0 \cdot \vec{x} = \vec{0}$

b) $(a + b) \cdot \vec{x} = a \cdot \vec{x} + b \cdot \vec{x}$

c) $a \cdot (\vec{x} + \vec{y}) = a \cdot \vec{x} + a \cdot \vec{y}$

d) $(a \cdot b) \cdot \vec{x} = a \cdot (b \cdot \vec{x})$

Remark For Experts 1.12:
It turns out that 1.7 and 1.11 are actually the only properties one really needs to do linear algebra. Rather than only considering $\mathbb{R}^n$, we could consider "any object with these properties", and much of what we will say will come out exactly the same. We choose to consider only $\mathbb{R}^n$ for the sake of concreteness.

2 Spans of Vectors

Let us begin with the definition.

Definition 2.1. *Let $\vec{v_1}, \cdots, \vec{v_k}$ be vectors in $\mathbb{R}^n$. Then we say another vector $\vec{x} \in \mathbb{R}^n$ is in the span of $\vec{v_1}, \cdots, \vec{v_k}$, written $\vec{x} \in \mathrm{Span}\{\vec{v_1}, \cdots, \vec{v_k}\}$, if there are k scalars $c_1, \cdots, c_k$ such that*

$$\vec{x} = c_1\vec{v_1} + \cdots + c_k\vec{v_k}$$

This is a rather abstract definition, so lets unpack it a bit.

Easy Exercise 2.2:
Let $\vec{x} \neq \vec{0}$ be any vector in $\mathbb{R}^n$. Describe

a) $\mathrm{Span}\{\vec{x}\}$

b) $\mathrm{Span}\{\vec{x}, 2\vec{x}\}$

c) $\mathrm{Span}\left\{\vec{0}\right\}$

d) If $\vec{x}, \vec{y}$ are two vectors in $\mathbb{R}^3$ that are not multiples of each other, describe $\mathrm{Span}\{\vec{x}, \vec{y}\}$.

Easy Exercise 2.3:
Show that for *any* $\vec{x_1}, \cdots, \vec{x_k}$, $\vec{0} \in \mathrm{Span}\{\vec{x_1}, \cdots, \vec{x_k}\}$

3 Linear Independence

Definition 3.1. *We say that a set of vectors $\vec{v_1}, \cdots, \vec{v_k} \in \mathbb{R}^n$ is linearly dependent if there is a set of scalars $c_1, ..., c_k$,* at least one nonzero*, such that*

$$c_1\vec{v_1} + \cdots + c_k\vec{v_k} = \vec{0}$$

A set is linearly independent if it is not linearly dependent.

Remark 3.2:
If a set of vectors are linearly dependent, at least one of the vectors is "redundant"; you could throw out a vector without changing the span. Do you see why this is?

Unimportant Exercise 3.3:
If one of the $v_i = \vec{0}$, is $\vec{x_1}, \cdots, \vec{x_k}$ linearly independent?

Exercise 3.4:
Show that the span of any two vectors in $\mathbb{R}^1$ are linearly dependent. What about three vectors in $\mathbb{R}^2$? What about $n + 1$ vectors in $\mathbb{R}^n$ (trickier).

Important Exercise 3.5:
Make 3.2 precise, and prove it! (hint: suppose $c_1\vec{v_1} + \cdots + c_k\vec{v_k} = \vec{0}$ and $c_1 \neq 0$, so you can divide by it. Now "solve for" $\vec{v_1}$)

3.1 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 1 2a
- ☐ W13 Midterm 1 2b
- ☐ S12 Midterm 1 7c
- ☐ S12 Midterm 1 7d
- ☐ W13 Final 11b

4 Dot and Cross Products

In this chapter, we are going to talk about some ways of measuring vectors. The "magnitude-and-direction" viewpoint is especially helpful in this case.

4.1 Dot Products

Remember that we don't allow you to multiply two vectors and get another vector. The dot product may be called a product and look like multiplication, but you should think of it as a sort of measurement of size and similarity. In particular, the dot product measures how "big" and close together vectors are.

There are two different formulas for the dot product. We will start with the first

Definition 4.1 (Dot Product)**.** *Let*

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \vec{y} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

be two vectors in $\mathbb{R}^n$. Then define

$$\vec{x} \cdot \vec{y} = x_1 y_1 + \cdots + x_n y_n$$

Remark 4.2:
As always, we say the types. The dot product takes two vectors in $\mathbb{R}^n$ and gives back a scalar.

Lets do some easy properties of the dot product, before we go on to see how its useful. These exercises should be as easy as writing out both sides of the equation and noticing they are the same.

Easy Exercise 4.3:
Show that, if $\vec{e}_i$ is the standard basis vector in $\mathbb{R}^n$ (that is, all coordinates 0 except the i^{th}, which is 1), and if $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, then

$$\vec{e}_i \cdot \vec{x} = x_i$$

Easy Exercise 4.4:
Show that, if $\vec{x}, \vec{y} \in \mathbb{R}^n$, then

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

Easy Exercise 4.5:
Show that, if $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then

$$(c\vec{x}) \cdot \vec{y} = c(\vec{x} \cdot \vec{y}) = \vec{x} \cdot (c\vec{y})$$

So scalars "distribute" in a funny little way.

We begin by thinking about what happens if you dot a vector in $\mathbb{R}^2$ with itself.

Important Exercise 4.6:
Use the Pythagorean Theorem to show that if $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, then

$$\sqrt{\begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}}$$

is the "length" of vector, that is, the distance between the point $\begin{pmatrix} x \\ y \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$. (hint: if you get stuck, draw a picture!) Does this still work for $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$? How about $\vec{x} \in \mathbb{R}^{41}$?

Inspired by , we define the length of a vector

Definition 4.7 (Length of a Vector)**.** *Let $\vec{v} \in \mathbb{R}^n$. Define the length $\vec{v}$, denoted $||\vec{v}||$, by*

$$||\vec{v}|| = \sqrt{\vec{v} \cdot \vec{v}}$$

The length is also sometimes called the "magnitude" or "norm".

The next exercise makes good intuitive sense:

Easy Exercise 4.8 (Lengths Behave as they Ought to):
Show that, if $\vec{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ and $c \geq 0$, then

$$||c\vec{x}|| = c||\vec{x}||$$

(hint: 4.5)

This is great. At least when you dot a vector with itself, it tells you how "big" the vector is. Even better, if you multiply a vector by 2, its "gets twice as big".

Now let's investigate what happens if we dot any old vectors together. Again, we start by considering the case in $\mathbb{R}^2$.

Important Exercise 4.9 (Cosine Dot Product Formula):
Let $\vec{x}$ and $\vec{y}$ be vectors in $\mathbb{R}^2$, and θ the angle between them. By applying the law of cosines to the triangle with corners at $\vec{0}$, $\vec{x}$ and $\vec{y}$, show that

$$\vec{x} \cdot \vec{y} = ||\vec{x}||||\vec{y}|| \cos(\theta)$$

(hint: the length of the third side of this triangle is $||\vec{x} - \vec{y}||$. What happens when you expand that into coordinates?)

Once again, we are inspired by the situation in $\mathbb{R}^2$, so we want this to be true for any n. However, it can be hard to say exactly what we mean by "angle between two vectors in $\mathbb{R}^5$". Since we don't quite have the language to say what we mean, I will tell a somewhat fanciful story. Take two vectors in $\mathbb{R}^n$ and rotate them so that they are both in the x_1x_2-plane, which looks kind of like $\mathbb{R}^2$. Of course, I haven't said why rotating doesn't change the dot product, or even what I meant by rotating in higher dimensions[2], but bear with me. Since the two vectors are nicely aligned with the x_1x_2-plane, the law of cosines works, and we get the following theorem.

Theorem 4.10 (Dot Product Cosine Formula)**.** *Let $\vec{x}$ and $\vec{y}$ be vectors in $\mathbb{R}^n$ (for any n), and θ the "angle between" them.*

$$\vec{x} \cdot \vec{y} = ||\vec{x}||||\vec{y}|| \cos(\theta)$$

Thus, the dot product measures not only how big things are, but also how narrow the angle between them is. If the two vectors are in the same line, the angle between them is 0, and $\cos(0) = 1$, so we just get the product of the lengths. If the two vectors are perpendicular, the angle between them is $\pi/2$, and since $\cos(\pi/2) = 0$, we get the dot product is zero, no matter how big the vectors involved are. For what its worth, if two vectors are pointing in opposite directions (that is, the angle between them is more than $\pi/2$ but less than $3\pi/2$) the sign of the formula aboove will be negative.

Since mathematicians love making up words, here is a synonym for perpendicular that will be used throughout this book:

Definition 4.11 (Orthogonal)**.** *Vectors $\vec{x}$ and $\vec{y} \in \mathbb{R}^n$ are called orthogonal if*

$$\vec{x} \cdot \vec{y} = 0$$

Easy Exercise 4.12:
Show that the $\vec{0}$ is orthogonal to every other vector.

4.2 Cross Products

The cross product is a special object which only works in $\mathbb{R}^3$. Given two vectors in $\mathbb{R}^3$ which are not multiples of each other, there is a unique line orthogonal to them both (try to visualize this). The cross product will give you a vector along that line, whose length is determined not by how small the angle between the two vectors is, but by how large. It is defined as follows

Definition 4.13 (Cross Product)**.** *Define the cross product of two vectors in $\mathbb{R}^3$ by*

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix}$$

[2] If you go on to a higher level linear algebra class like math 113, they will say, roughly, that a rotation is just anything which doesn't change the dot product

Wow; that formula is a complicated looking! Remembering that we care about types, note that the cross product takes two vectors in $\mathbb{R}^3$ and returns a vector in $\mathbb{R}^3$. Remember that the goal of the cross product is to get a vector orthogonal to both of the two original vectors.

Important Exercise 4.14:
Show that if $\vec{x}, \vec{y} \in \mathbb{R}^3$, then $\vec{x} \times \vec{y}$ is orthogonal to both $\vec{x}$ and $\vec{y}$ by writing out the dot product explicitly and seeing that everything cancels.

Easy Exercise 4.15:
The cross product is not commutative. Show that

$$\vec{x} \times \vec{y} = -\vec{y} \times \vec{x}$$

Easy Exercise 4.16:
Show that

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

Easy Exercise 4.17:
Show that, if $\vec{x}, \vec{y} \in \mathbb{R}^3$ and $c \in \mathbb{R}$, then

$$(c\vec{x}) \times \vec{y} = c(\vec{x} \times \vec{y}) = \vec{x} \times (c\vec{y})$$

We will now prove one of the more important properties of the cross product in a few steps.

Exercise 4.18 (optional, but make sure you know the result):
Show that, for $\vec{x}, \vec{y} \in \mathbb{R}^3$, that

$$||\vec{x} \times \vec{y}||^2 = ||\vec{x}||^2||\vec{y}||^2 - (\vec{x} \cdot \vec{y})^2$$

This should amount to expanding out both sides and noticing they are the same. Make sure you have plenty of paper for this one, this is a good bit of number crunching.

Exercise 4.19:
Use 4.18, 4.10 and the fact that $\sin^2(x) = 1 - \cos^2(x)$ to show that, if $\vec{x}, \vec{y}$ are two vectors in $\mathbb{R}^3$ with angle θ between them, then

$$||\vec{x} \times \vec{y}|| = ||\vec{x}||||\vec{y}|| \sin(\theta)$$

Easy Exercise 4.20:
Recalling that the area of a triangle with sides of length a, b and c and angle θ between the sides of length a and b is

$$\frac{ab\sin(\theta)}{2}$$

show that $||\vec{x} \times \vec{y}||$ is the area of the parallelogram with corners $\vec{0}$, $\vec{x}$, $\vec{y}$ and $\vec{x} + \vec{y}$. (hint: 4.19)

Remark 4.21:
Cross products and dot products are geometric objects, so make sure you know when to use them. The first midterm tends to have one or two cute geometry problems, which the alert student will know to solve with dot products and cross products. Often times you will have to compute the area or some angle of a random triangle in $\mathbb{R}^3$, or find a vector perpendicular to it. If you come across such a question, start by translating the figure so that one of the interesting corners is at the origin. That is, given a triangle with corners $\vec{a}, \vec{b}$ and $\vec{c}$, define $\vec{v} = \vec{b} - \vec{a}$ and $\vec{w} = \vec{c} - \vec{a}$. The triangle with corners $\vec{0}, \vec{v}$ and $\vec{w}$ is geometrically the same as the original triangle (just moved over a bit), but the dot product, cross product and length formula can give you information about it.

4.3 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 1 3a
- ☐ W13 Midterm 1 3b
- ☐ W13 Midterm 1 3c
- ☐ W13 Midterm 1 6e
- ☐ A12 Midterm 1 3a
- ☐ A12 Midterm 1 6a
- ☐ A12 Midterm 1 6c
- ☐ S12 Midterm 1 3a
- ☐ S12 Midterm 1 7e
- ☐ S12 Midterm 1 7f
- ☐ W13 Final 11a
- ☐ W13 Final 11c

5 Systems of Linear Equations

Systems of linear equations are ubiquitous in all branches of science and engineering. Being able to identify, understand and solve them is an invaluable tool for anyone who wants to deal with data. Systems of linear equations will appear as a sub-problem in an extraordinary number of contexts and you must be able to deal with them when they appear. Luckily there are tons of computer packages which can help you deal with them when they appear in real life situations... but first you need to learn the rules! Studying linear equations–their anatomy and solutions–will be the topic of the first half of these notes.

5.1 The Problem

A linear equation is an equation of the form

$$m_1x_1 + \cdots + m_nx_n - b = 0$$

where the m_i and b are fixed numbers and the x_i are variables. If $n = 1$, this is just very familiar;

$$mx - b = 0$$

When $n = 2$, these equation are

$$ax + by = c$$

and they implicitly define a line (I am playing fast and loose with the variable names here. If n is small, you can, if you so please, call x_1 by x and x_2 by y). Notice that there could be lots of solutions for $n > 1$, for instance there are infinitely many on a line. Notice, however, that there need not be *any* solutions. For instance, consider

$$0 \cdot x + 1 = 0$$

Easy Exercise 5.1:
Let $\vec{m}, \vec{x} \in \mathbb{R}^n$, where $\vec{m}$ is some fixed vector and $\vec{x}$ is a variable vector. Show that the equation

$$\vec{m} \cdot \vec{x} = -b$$

is “the same” as a linear equation. (hint: vector dot vector = scalar)

Remark 5.2:
The nice thing about writing a linear equation in the above way is that it emphasizes the "geometric" nature of linear equations. Sure, the original definition is strictly "algebraic", but considering the set of solutions as points in $\mathbb{R}^n$ immediately gives us an intuitive picture for what would otherwise just be a mess of numbers.

Temporarily, let us restrict to the case $b = 0$ and $\vec{m} \neq \vec{0}$. Then, by 5.1, the set of solutions to the linear equation $\vec{m} \cdot \vec{x} = 0$ is merely the set of all vectors orthogonal to $\vec{m}$. If $\vec{m} \in \mathbb{R}^2$, vectors on the line perpendicular to $\vec{m}$ solve this equation. If $\vec{m} \in \mathbb{R}^3$, vectors in the plane perpendicular to $\vec{m}$ solve this equation. This "geometric" intuition remains correct when $n \geq 3$. When we talk about dimension, we will be able to make this precise, but in $\mathbb{R}^n$ the set of solutions to $\vec{m} \cdot \vec{x} = 0$ should be somehow an $(n-1)$-dimensional "plane".

If we had two vectors, $\vec{m}_1$ and $\vec{m}_2$, we could consider the set of all vectors which *simultaneously* solved $\vec{m}_1 \cdot \vec{x} = b_1$ and $\vec{m}_2 \cdot \vec{x} = b_2$.

Exercise 5.3:
Let $n = 3$ and $b_1 = b_2 = 0$. Describe, geometrically, the set of vectors $\vec{x}$ with $\vec{m}_1 \cdot \vec{x} = \vec{m}_2 \cdot \vec{x} = 0$, if

a) $\vec{m}_1$ and $\vec{m}_2$ are linearly independent.

b) $\vec{m}_1$ and $\vec{m}_2$ are linearly dependent, but both nonzero.

Definition 5.4 (System of Linear Equations)**.** *A system of linear m linear equations in n variables is a list of equations*

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & - & b_1 & = 0 \\ \vdots & & \ddots & & \vdots & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & - & b_m & = 0 \end{array}$$

where the $a_{i,j}\mathbb{R}$ are constants and the x_i are variables. A system is called homogeneous if all the $b_i = 0$, and inhomogeneous otherwise. A system is inconsistent if there are no solutions.

Remark 5.5:
We changed the name of the coefficients from m to a with a full system of linear equations because, traditionally, m is the number of equations and n is the number of variables. This convention is horribly confusing because, not only do the letters in question have little mnemonic meaning, but they rhyme! Beware of confusing the two.

Important Exercise 5.6:
Come up with an example of a system of linear equations with

a) exactly one solution.

b) more than one solutions.

c) no solutions

(hint: these can all be done with a single equation and a single variable, that is, $m = n = 1$)

Easy Exercise 5.7:
Extend 5.1 to a system of m linear equations.

Easy Exercise 5.8:
Show a homogeneous system of linear equations has at least one solution. (hint: it should feel like cheating)

Exercise 5.9:
If you add an additional equation to a system of m equations in n variables, show that the solution set either stays the same or gets smaller (hint: a solution to the new system of $m+1$ equations is also a solution to the first m equations). In the case of a homogeneous system, can you come up with a rule telling if the solution set will shrink or stay the same? (hint: write system as in 5.1 and consider the span of the first m coefficient vectors $\vec{a_i}$)

5.2 The Solution

It turns out that there is a single, end-all solution to this problem. First of all, you could try using the substitution method. The best way of seeing why this is inefficient is trying it.

Unimportant Exercise 5.10:
Pick a system of linear equations from the text and try solving it by substitution. Make sure you have plenty of paper. Do you see why this is kind of a bad idea.

The other alternative is called Gaussian Elimination or Row Reduction. Consider the system

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & - & b_1 & = 0 \\ \vdots & & \ddots & & \vdots & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & - & b_m & = 0 \end{array}$$

We will develop a group of very simple operations, each of which does not change the solution set. The idea is this: if you know modifying the system in a certain way will not change the solution set, you are of course free to do that modification. If that modification makes the system simpler, then you have made progress. It is merely a generalization of the very familiar principle that if you see "$x - 4 = 5$", you want to "subtract 4 from both sides", because you know that operation does not change the solution, but will make the equation much simpler.

The first operation is just swapping the position of two equations. Next, is multiplying through by a scalar, that is, if you had an equation

$$a_{i,1}x_1 + \cdots + a_{i,n}x_n - b_i = 0$$

and c is some nonzero scalar, then replace this equation with

$$(ca_{i,1})x_1 + \cdots + (ca_{i,n})x_n - (cb_i) = 0$$

Finally, you can replace an equation with the sum of two other equations, that is, given two equations from the system

$$\begin{array}{cccccccc} a_{i,1}x_1 & + & \cdots & + & a_{i,n}x_n & - & b_i & = 0 \\ a_{j,1}x_1 & + & \cdots & + & a_{j,n}x_n & - & b_j & = 0 \end{array}$$

replace one of them with

$$(a_{i,1} + a_{j,1})x_1 \quad + \quad \cdots \quad + \quad (a_{i,n} + a_{j,n})x_n \quad - \quad (b_i + b_j) \quad = 0$$

Exercise 5.11:
Verify that the following three operation do not change the solution set.

a) Swapping two equations

b) Multiplying an equation by a non-zero scalar

c) Replacing an equation with the sum of two equations

Your answer should be of the form "If $x_1, ..., x_n$ is some solution, and we do an operation, then the resulting system of equations has the same solution set, because <fill in the blank>". Notice how the sentence above does not depend on you knowing what that solution is, or even that one exists.

Thus, if you want to find a solution to a system of equations, you know you won't get the wrong answer by using these operations.

Important Exercise 5.12:
Try out a few examples from the text; pick a system of linear equations and apply these operations until the system is simple enough for you to solve "by hand". Invent strategies for clearing out variables and see just how simple you can make a system. Try to get as many variables as possible to appear in only one equation. How do you know when you're "done"?

5.3 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 1 4a
- ☐ W13 Midterm 1 4b
- ☐ W13 Midterm 1 7b
- ☐ W13 Midterm 1 8a
- ☐ W13 Midterm 1 8b
- ☐ A12 Midterm 1 1a
- ☐ A12 Midterm 1 4a
- ☐ A12 Midterm 1 4h
- ☐ A12 Midterm 1 6b
- ☐ S12 Midterm 1 1a
- ☐ S12 Midterm 1 3b
- ☐ S12 Midterm 1 3c
- ☐ S12 Midterm 1 5a
- ☐ W13 Midterm 2 6a (ignore P)
- ☐ S12 Final 1a
- ☐ S12 Final 2a
- ☐ S12 Final 2b

6 Matrices

You probably noticed in 5.12 a few things. First of all, you might have noticed that "leaving a blank spot" when a variable didn't appear in an equation was useful (if not, think about why it might be). For instance, it would be useful to write the system like this:

$$\begin{array}{ccccccccc} x_1 & + & 5x_2 & + & -3x_3 & - & 2 & = 0 \\ & & 2x_2 & & & - & 5 & = 0 \end{array}$$

rather than

$$\begin{array}{cccccccc} x_1 & + & 5x_2 & + & -3x_3 & - & 2 & = 0 \\ 2x_2 & - & 5 & = & 0 & & & \end{array}$$

Lining things up is helpful because it suggests row reduction strategies and allows you to check on your "progress" at a glance.

The second thing you might have noticed is that you spend an awful lot of time drawing x's and $+$'s, even though they don't really tell you anything interesting, especially if you aligned your system as mentioned above. Thus it would be just as good to only write the coefficients, so long as you kept track of the blank spaces by putting a zero. We don't even really have to keep track of the right hand side of the equation, because we know its just a zero. Thus we could write the system above "compactly" as

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

We define a matrix just like we did a vector

Definition 6.1 (Matrix)**.** *An $m \times n$* matrix *is a rectangular array of real numbers with n columns and m rows. If A is an $m \times n$ matrix, we denote the entry in column i and row j by $A_{i,j}$.*

Remark 6.2:
It is important to remember this is just a syntactic transformation: just an easier way of writing down a system of equations. It means the exact same thing, although we hope this format can yield insight about the problem at hand.

Remark 6.3:
Everybody gets the order of the entries confused. I often write down a matrix you wonder if it should be $m \times n$ or $n \times m$, I find this horribly confusing (see 5.5). Just always remember n is the number of variables and figure it out from there.

Remark For Programmers 6.4:
This translates quite literally to a programming paradigm: store the "system of an equation" in a 2-d array. This is good because you don't need to waste memory on the x_i's, has good locality and if you want to know "the coefficient of x_3 in equation 2", you just look at $A[2][3]$. Interestingly, if space is limited and you have a large system where each equation only has a few variables (called a sparse system), you end up wasting a lot of space on zeros, so a "literal" format might be preferred.

We might wish to remind ourselves that the b_i special, in the sense that they are not to be eventually multiplied by an x_i. This leads to the concept of an "augmented matrix", which is a purely syntactic concept designed to keep us from getting confused while working with matrices.

Definition 6.5 (Augmented Matrix)**.** *We will often draw a vertical dashed line separating the b_i's column from the rest of the matrix, just to visually remind elf's what the matrix represents. We call such a* matrix augmented*, and the column to the right of the line the augmented column.*

I find Gaussian Elimination is easier to think about with matrices, since the operations are just row-wise additions, multiplications and swaps.

Exercise 6.6:
What do the Gaussian Elimination rules (5.11) look like in matrix form?

In 5.12, I asked how do you know when you're done? With the system written as a matrix, this is an easy-to-check. You are done when your matrix is in "reduced row echelon form", or "rref". That is a scary sounding name (I have no idea with echelon means), but it just means the following.

Definition 6.7 (Reduced Row Echelon Form)**.** *An $m \times n$ matrix A is in* reduced row echelon form *if*

1. *Each column either contains only a single 1 (and the rest zeros), or the nonzero entries are* not *the row's leftmost nonzero entry.*

2. *The rows are sorted by their first nonzero entry.*

We call columns with only a single one pivot columns *and otherwise* free columns.

Definition 6.8 (Gaussian Elimination)**.** *Gaussian Elimination is the process of using cleverly chosen row operations until a matrix is in reduced row echelon form.*

Why is reduced row echelon form so nice?

Example 6.9:
Consider this rref'd matrix.

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

Remember that a matrix is just shorthand for a system of equations

$$\begin{array}{ccccccc} x_1 & + & & -3x_3 & - & 2 & = 0 \\ & & x_2 & + \quad x_3 & - & \frac{1}{2} & = 0 \\ & & & & & 0 & = 0 \end{array}$$

Notice that you can set x_3 to anything you want and you automatically know what x_1 and x_2 are; that is, we can rearrange the system to look like.

$$\begin{array}{ccccc} x_1 & = & 3x_3 & + & 2 \\ x_2 & = & -x_3 & + & \frac{1}{2} \\ 0 & = & 0 & & \end{array}$$

Thus for each possible value of x_3 there is exactly one solution to the system. This is about as simple a system of equations as you can ask for, as it has the property that you can "just plug in any value for x_3 and get a solution". However, we can actually write this in a slightly "slicker" form. We know that the solution to the system is a vector

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

As we saw before, we can write x_1 and x_2 in terms of x_3. In a completely trivial way, (the kind of way which is so trivial you'd never think of it), we can "write x_3 in terms of x_3". How? $x_3 = x_3$ (womp womp). Thus we can write this as

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3x_3 + 2 \\ -x_3 + \frac{1}{2} \\ x_3 \end{pmatrix}$$

We can do better still. Each of the coordinates of the right hand vector are of the form $ax_3 + b$, where $b = 0$ in the last coordinate. Thus we can split this vector up as a sum, and then factor out the x_3, like so:

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3x_3 + 2 \\ -x_3 + \frac{1}{2} \\ x_3 \end{pmatrix} = \begin{pmatrix} 3x_3 \\ -x_3 \\ x_3 \end{pmatrix} + \begin{pmatrix} 2 \\ \frac{1}{2} \\ 0 \end{pmatrix} = x_3 \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ \frac{1}{2} \\ 0 \end{pmatrix}$$

This is very cool! Writing the solution this way emphasizes the geometric nature of the solution set, and makes it is obvious that the solutions form a line! Do you see why?

Important Exercise 6.10:
Convince yourself the solution set forms a line in $\mathbb{R}^3$. Can we interpret this as "the line line going through a certain point in a certain direction"?

Important Exercise 6.11:
Go through the logic of 6.9 for the following rref'd matrix, and then do 6.10 for it.

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

What is the geometry of the solution set? (hint: the system is the same except for there is one more free variable)

Definition 6.12 (Parametric Form)**.** *We call a solution to a system of equations in parametric form if it is written a constant vector plus a sum of variables times constant vectors, as in 6.9 and 6.11.*

Example 6.13:
Consider the following matrix in rref.

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

Notice the bottom row has non nonzero entries corresponding to any x_i, but a nonzero entry in the $\vec{b}$ column. Thus when we write out the system of equations, we get

$$\begin{array}{ccccccc} x_1 & + & & & -3x_3 & & = 0 \\ & & x_2 & + & x_3 & & = 0 \\ & & & & & 1 & = 0 \end{array}$$

We know that $1 \neq 0$, so no matter what the x_i are, the bottom equation is false, and thus the whole system is false, so the system has no solutions.

Exercise 6.14:
Show that a system of linear equations is inconsistent if and only if the augmented column of the corresponding matrix is a pivot column in the rref form. Additionally, show that if there is a solution (that is, if the system is not inconsistent), we can write it in parametric form. (hint: show the method of parametrization from 6.9 and 6.11 works whenever the augmented column is not a pivot)

Exercise for Programmers 6.15 ((optional)):
Try writing a computer program (in the language of your choice) to put a matrix in rref.

We end with an easy definition:

Definition 6.16. *Matrix Addition We add two matrices of the same size by adding corresponding entries, that is*

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} + \begin{pmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{pmatrix} = \begin{pmatrix} a_{1,1} + b_{1,1} & \cdots & a_{1,n} + b_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} + a_{m,1} & \cdots & a_{m,n} + a_{m,n} \end{pmatrix}$$

We define the zero matrix as the matrix with all zero entries.

6.1 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 1 1a
- ☐ W13 Midterm 1 6b
- ☐ S12 Midterm 1 2a
- ☐ W13 Final 2a

7 Matrix Vector Products

Recall that a homogeneous linear equation is a one where the augmented column is all zeros.

Easy Exercise 7.1:
Show that if a system of linear equations is homogeneous, then the rref'd form of the matrix for that system is homogeneous. (hint: what do each of the three operations of Gaussian Elimination do to the augmented column of such a matrix?)

Since throughout our analysis the last column stays zero, we may neglect it and represent our system of equations by a non-augmented matrix. A homogeneous system of equation looks like this:

$$\begin{array}{ccccccc} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = 0 \\ \vdots & & \ddots & & \vdots & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & = 0 \end{array}$$

The corresponding matrix, which we will call A, is

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$$

Define $\vec{a_i}$ to be the vector with jth coordinate $a_{i,j}$, that is, $\vec{a_i}$ is the ith row of A. Then recall, as in 5.1, that our system of equations can also be written

$$\begin{array}{ccc} \vec{a_1} \cdot \vec{x} & = & 0 \\ \dots & & \dots ccc \\ \vec{a_n} \cdot \vec{x} & = & 0 \end{array}$$

Following the philosophy of atomizing things, it would be nice if we could write this system of many equations as some kind of single equation involving vectors. To do this though, we need to make a definitions.

Definition 7.2 (Matrix Vector Product)**.** *Let A be an $m \times n$ matrix with rows $\vec{a_i} \in \mathbb{R}^m$. If $\vec{x} \in \mathbb{R}^n$, then define the matrix-vector product*

$$A\vec{x} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} \vec{a_1} \cdot \vec{x} \\ \dots \\ \vec{a_n} \cdot \vec{x} \end{pmatrix}$$

Remark 7.3:
With each new operation it is important to remember the type. In this case:

$$\text{matrix times vector} = \text{vector}$$

Or even better

$$m \times n\text{-matrix times } n\text{-vector} = m\text{-vector}$$

Pedagogical Remark 7.4:
Students often find this definition unmotivated, as if it came out of left field. It is extremely important that you understand why the product is defined the way it is, and why we want a matrix vector product at all!

Easy Exercise 7.5:
Let A be the non-augmented matrix for a system of m homogeneous linear equations in n variables. Show that the system can be written

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

Easy Exercise 7.6:
Given a system of equations:

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & - & b_1 & = 0 \\ \vdots & & \ddots & & \vdots & & & \\ a_{m,1}x_1 & + & \cdots & + & a_{m,n}x_n & - & b_m & = 0 \end{array}$$

Let A be the non-augmented matrix associated with this system, that is,

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$$

and let

$$\vec{b} = \begin{pmatrix} b_1 \\ \cdots \\ b_m \end{pmatrix}$$

Show that the original system of linear equations is equivalent to the matrix equation

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

This generalizes 7.5.

Exercise 7.7:
Show that the following definition of matrix-vector multiplication is equivalent to the original. If A is an $m \times n$ matrix and $\vec{x} \in \mathbb{R}^n$,

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\begin{pmatrix} a_{1,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + \cdots + x_n\begin{pmatrix} a_{1,n} \\ \vdots \\ a_{m,n} \end{pmatrix}$$

In particular, $A\vec{x}$ is a linear combination of the columns with coefficients the coordinates of $\vec{x}$.

Definition 7.8 (Identity Matrix)**.** *Fix some n. Let I (sometimes called I_n) be an $n \times n$ matrix with 1 on the diagonal entries and 0 elsewhere, that is*

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

We will see that this is a very special matrix.

Exercise 7.9:
Verify the following properties:

a) $(A + B)\vec{x} = A\vec{x} + B\vec{x}$

b) $A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}$

c) $A\vec{0} = \vec{0}$

d) If Z is a $m \times n$ matrix of all zeroes, $Z\vec{0} = \vec{0}$ (the two zero vectors are different... where do they live?)

e) $I\vec{x} = \vec{x}$ (see 7.8)

(hint: recall that each entry of a matrix vector product is a dot product, and that dot products obey similar looking laws)

Exercise 7.10:
If we wish to extract a column from a matrix, there is a nice way of doing so. Show that if A is an $m \times n$ matrix and $\vec{e}_i$ is the i^{th} standard basis vector (that is, all zeros except for a 1 in the i^{th} position), then Ae_i is the i^{th} column of A. Keep this in mind, it can be a handy shortcut come exam time.

Remark 7.11:
It may seem silly at first to define all this notation, but it gives us the advantage of being able to write a system of linear equations in a few short symbols. Also, the symbols $A\vec{x} = \vec{b}$ looks an awful lot like the very simple scalar equation $ax = b$. In fact, if $m = n = 1$, it is exactly that. We will see in 13 that "multiplying by a matrix" is actually a quite reasonable generalization of "multiplying by a number" to situations with many variables.

7.1 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 1 1c
- ☐ A12 Midterm 1 4e
- ☐ S12 Midterm 1 4b
- ☐ S12 Midterm 1 5b
- ☐ S12 Midterm 1 7a
- ☐ S12 Midterm 1 7b
- ☐ W13 Final 1a
- ☐ W13 Final 1b
- ☐ W13 Final 1c

Part II

Spaces in $\mathbb{R}^n$

8 Null Space

We want to say that the equation $A\vec{x} = \vec{b}$ generalizes $ax = b$, but we have to notice that the situation *is* sightly different. Consider for instance the homogeneous case, $ax = 0$. If $a \neq 0$, then the only solution is $x = 0$. However, the situation in many variables is a big more subtle: you can multiply a nonzero matrix by a nonzero vector and get the zero vector! Actually, this shouldn't be that surprising, you may have quietly observed this phenomenon when solving some homogeneous system of linear equations and finding many solutions, but that was before we started writing the system as a multiplication. To measure this phenomenon, we introduce the null space, which is merely a special case of a something you're already used to.

Definition 8.1 (Null Space)**.** *Given an $m \times n$ matrix A, The null space of A (written $N(A)$) is the set of solutions of the system of equations*

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

That is, a vector $\vec{x} \in N(A)$ if $A\vec{x} = \vec{0}$. In some text it will be called the kernel, and written $\operatorname{Ker}(A)$*, but we will not use this notation.*

Remark 8.2:
We always want to think about the types. If A is $m \times n$, the null space vectors must be n-dimensional (why again?). Thus we observe that $N(A)$ is a subset of $\mathbb{R}^n$.

Remark 8.3:
You might have seen

$$N(A) = \left\{\vec{x} \in \mathbb{R}^n \middle| A\vec{x} = \vec{0}\right\}$$

In English this says "the null space of A is the set of all things that go to zero when you multiply by A". I prefer to say "a vector is in the null space of A if multiplying it by A gives zero", but it should be clear that these mean the same thing. The point is that, given a vector $\vec{x}$ and a matrix A, you can "test" whether or not it is in the null space by computing $A\vec{x}$ and seeing if it is zero.

Easy Exercise 8.4:
Show that for *any* matrix A, $\vec{0} \in N(A)$.

Exercise 8.5:
Show that the columns of A are linearly independent if and only if $\vec{0}$ is the *only* vector in $N(A)$. We say that $N(A)$ is *trivial* in this case. (hint: use the definition of from 7.7 and the definition of linear independence)

Easy Exercise 8.6:
Say how you would calculate $N(A)$. (hint: we've already done this) If you'd like, pick some matrices from the text and calculate their null spaces. Write the solution in parametric form.

8.1 The Null Space Parametrizes Solutions

The null space measure the failure of our ability to expect the product of nonzero things to be nonzero, but it also does something else. This is one of the most important themes in linear algebra, so make sure you fully understand the next few exercises.

Important Exercise 8.7:
Let $\vec{x}_1$ and $\vec{x}_2$ be two solutions to $A\vec{x} = \vec{b}$. Show that

$$\vec{x}_1 - \vec{x}_2 \in N(A)$$

(hint: use 7.9) Now show, similarly, that if $\vec{y} \in N(A)$ and $\vec{x}_1$ is a solution to $A\vec{x} = \vec{b}$, then

$$A(\vec{x}_1 + \vec{y}) = \vec{b}$$

In other words, if you have some solution and you add a vector in the null space, you get another solution.

To hammer home the meaning of this exercise, we introduce the following slogan:

The null space parametrizes the solutions to a system of equations

Important Exercise 8.8:
Explain what that slogan means. Make connections between 8.7 and 6.14: does the null space somehow "appear" in the parametric form of a solution? Interpret this geometrically as the solution set of an arbitrary set of linear equations, if nonempty, is just a translation of the null space.

8.2 Exam Exercises

Try the following exercises from past exams

- ☐ A12 Midterm 1 2a
- ☐ A12 Midterm 1 4f

9 Column Space

Definition 9.1 (Column Space). *The column space $C(A)$ of an $m \times n$ matrix A is the span of the columns in A.*

Remark 9.2:
Just as the null space was naturally a subset of $\mathbb{R}^n$, the column space is naturally a subset of $\mathbb{R}^m$.

Easy Exercise 9.3:
Show that $\vec{b} \in C(A)$ if and only if the system $A\vec{x} = \vec{b}$ has at least one solution.

Remark 9.4:
Column space is a horrible name! It should be called the "range" or the "image"! We could just as easily have define the row space to be the span of the rows in $\mathbb{R}^n$, or anything else equally as stupid. We don't care about the column space because we're being cute, we care about the column space because they tell us which equations are solvable, or equivalently what vectors are possible results of a matrix vector product.

Exercise 9.5:
WARNING: applying row reduction to a matrix can *change* the column space! Can you find an example?

Remark 9.6:
"Calculating" the column space is a particularly easy thing to do, because it is linearly just the span of the columns. However, the span of columns might not be the easiest way to say what the column space in, in the same sense that $\frac{2}{4}$ is not the easiest way to write $\frac{1}{2}$. If this is not clear, then thinking about the column space of

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

9.1 Exam Exercises

Try the following exercises from past exams

- □ W13 Midterm 1 5a
- □ A12 Midterm 1 1b
- □ A12 Midterm 1 2b
- □ S11 Final 1a

10 Subspaces

One might be tempted to say that $\mathbb{R}^2$ is a subset of $\mathbb{R}^3$, by identifying the vector (x, y) with $(x, y, 0)$. But, once done, it becomes obvious you made an arbitrary choice. Why not identify (x, y) with $(0, x, y)$, or $(x, 0, y)$, or even $(y, 0, x)$. In fact, if we really want to consider $\mathbb{R}^2$ as a "subspace" (whatever that is) of $\mathbb{R}^3$, we could just pick two random linearly independent vectors, say $(1, 2, 3)$ and $(4, 5, 6)$, and use the identification

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto x \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + y \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$$

Parametrize a plane with two variables is kind of like realizing a (perhaps stretched or rotated) copy of $\mathbb{R}^2$ in $\mathbb{R}^3$! The point is, there are a lot of ways to see $\mathbb{R}^2$ in $\mathbb{R}^3$.

So, if we want to recognize $\mathbb{R}^2$ as a "subspace" (whatever that is) of $\mathbb{R}^3$ (or $\mathbb{R}^m$ as a "subspace" or $\mathbb{R}^n$ with $m < n$), we are going to have to actually have an idea. So we ask, "what structure does $\mathbb{R}^m$ have that we expect it still to have when you stuff it inside $\mathbb{R}^n$?" Well, you should be able to add vectors together. Think about $(x, y) \mapsto (x, y, 0)$. For now, we'll call these vectors "our copy of $\mathbb{R}^2$ in $\mathbb{R}^3$" (this terminology is temporary. If you had two vectors, $(x_1, y_1, 0)$ and $(x_2, y_2, 0)$, and you added them together, the sum would still have zero in the last coordinate; its still in our copy of $\mathbb{R}^2$. Likewise, multiplying by a scalar c gives $c(x, y, 0) = (cx, cy, 0)$, which remains in our copy of $\mathbb{R}^2$. The key idea is that these properties are enough to recognize when a subset of $\mathbb{R}^n$ is acting like a "copy" of $\mathbb{R}^m$ for some m, and we call it a subspace.

Definition 10.1 (Subspace)**.** *A subset of S $\mathbb{R}^n$ is called a subspace if*

1. *(Zero) The zero vector is in S.*
2. *(Closed Under Sums) For any two vectors $\vec{x}, \vec{y} \in S$, $\vec{x} + \vec{y} \in S$ (sums of things in S stay in S)*
3. *(Closed Under Scalar Multiplication) For any vector $\vec{x} \in S$ and any scalar $c \in \mathbb{R}$, $c\vec{x} \in S$ (multiplying things in S by scalars leaves them in S)*

In other words, if S is a subspace and you do vector space operations with things in S, the result will still be in S.

Unimportant Remark 10.2:
We require $\vec{0}$ to be in a subspace for no reason other than we don't want the empty set to be a subspace (because $\mathbb{R}^0$ has one vector, not zero vectors). If $S \subset \mathbb{R}^n$ is nonempty, then there is some vector $\vec{x} \in S$, so if S is closed under scalar multiplication, $0\vec{x} = \vec{0} \in S$, so the first condition is automatic most of the time.

Remark 10.3:
How do we show something is a subspace? There are two main ways to do it. The first is to verify the axioms of 10.1 directly. The second is to show that S can be written either as the null space of some matrix, or as the span of some vectors (of course, this is cheating until after 10.4 and 10.5).

Lets outline how to show a subset $S \subset \mathbb{R}^n$ is a subspace directly from the axioms. Given some description of S, start by showing that zero satisfies that description (usually just a simple observation). Then say "suppose $\vec{x}$ and $\vec{y}$ are in S". This probably means, depending on the way S is described, that $\vec{x}$ and $\vec{y}$ satisfy a certain property/equation, or can be written in a certain way. Restate that property, it will almost surely be helpful. Then say why the vector $\vec{x}+\vec{y}$ *also* satisfies that property/equation or can be written in that way; this will usually be some very simple observation. Finally, repeat these steps for $c\vec{x}$, for arbitrary $c \in \mathbb{R}$.

Students sometimes find this sort of "higher order" reasoning slightly difficult, as it is usually the first time they have seen it, but fear not. After a few practice exercises you'll be a pro, which is good, because you're going to have to do something like it on the exam.

Example 10.4:
We will show (rather verbosely) directly from the axioms, that for any $m \times n$ matrix A, $N(A)$ is a subspace. Make sure you follow along fully! Recall that a vector $\vec{x}$ is in $N(A)$ when $A\vec{x} = \vec{x}$. We first observe that, by 7.9

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

and thus $\vec{0} \in N(a)$. Next suppose $\vec{x}$ and $\vec{y} \in N(A)$. This means

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

But, again by 7.9,

$$A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y} = \vec{0} + \vec{0} = \vec{0}$$

Thus $\vec{x}+\vec{y} \in N(A)$. Finally, let c be any scalar. Then, again by 7.9

$$A(c\vec{x}) = cA\vec{x} = c\vec{0} = \vec{0}$$

so $c\vec{x} \in N(A)$. Thus $N(A)$ is a subspace.

Exercise 10.5:
Let $\vec{x}_1, \cdots, \vec{x}_k \in \mathbb{R}^n$. Show from the axioms that Span$\{\vec{x}_1, \cdots, \vec{x}_k\} \subset \mathbb{R}^n$ is a subspace. Conclude that for any $m \times n$ matrix A, $C(A) \subset \mathbb{R}^m$ is a subspace.

Easy Exercise 10.6:
Show that $\mathbb{R}^n$ itself is a subspace by

a) Showing that the axioms 10.1 hold.

b) Showing that $\mathbb{R}^n$ is the null space of a particular matrix.

c) Showing that $\mathbb{R}^n$ can be written as the span of some vectors

Easy Exercise 10.7:
Show that the set $\{\vec{0}\}$ is a subspace by

a) Showing that the axioms 10.1 hold (there is only one vector in your space, so this amounts to checking things like $\vec{0} + \vec{0} = \vec{0}$).

b) Showing that $\mathbb{R}^n$ is the null space of a particular matrix.

We call this the trivial subspace.

Exercise 10.8:
Let S be the set of vectors in $\mathbb{R}^n$ satisfying the equation $A\vec{x} = \vec{x}$. Show that S is a subspace by

a) Showing that the axioms 10.1 hold.

b) Showing that S is the null space of a particular matrix. (hint: subtract $\vec{x}$ from both sides and recall that $\vec{x} = I\vec{x}$, where I is defined in 7.8)

Exercise 10.9:
Let S be the set of vectors orthogonal to a given vector $\vec{v}$. Show S is a subspace by

a) Showing that the axioms 10.1 hold.

b) Showing that S is the null space of a particular matrix.

Exercise 10.10:
Let V and W be subspaces or $\mathbb{R}^n$. Show from the axioms that $V \cap W$ (that is, the set of vectors in both V and W) is a subspace. Now let $V = N(A)$ where A is an $m_1 \times n$ matrix and $W = N(B)$ where B is an $m_2 \times n$ matrix. Find a matrix C (of course using A and B) such that $V \cap W = N(C)$. (hint: there is an $(m_1 + m_2) \times n$ matrix which does the trick)

Unimportant Remark 10.11:
There are things which are not subsets of $\mathbb{R}^n$ which are still kind of like subspaces. For instance, consider the set of all polynomials, denoted $\mathbb{R}[x]$. Recall that a polynomial is a sum

$$f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

Of course we can add two polynomials and we'll get another polynomial. We can also multiply a real number by a polynomial and get another polynomial. Finally, the constant polynomial $f(x) = 0$ behaves like zero, so for all intense and purposes $\mathbb{R}[x]$ satisfies 10.1, despite not being inside of any $\mathbb{R}^n$. Without being too specific, we call such a thing a general Vector Space.

10.1 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 1 7a
- ☐ A12 Midterm 1 4b
- ☐ A12 Midterm 1 4g
- ☐ S11 Final 2a

11 Basis For a Subspace

By 10.5, the span of some vectors is a subspace of $\mathbb{R}^n$. However, we know that if the spanning set $\vec{v}_1, \cdots, \vec{v}_k$ is linearly dependent, we can throw away one of the vectors without making the span smaller. If we wished, we could keep throwing away vectors until the set becomes linearly independent, at which point we could evict no more vectors without changing the subspace they spanned. This leads to the notion of a *basis* for a subspace.

Definition 11.1 (Basis). *Let $S \subset \mathbb{R}^n$ be a subspace. Some vectors $\vec{v}_1, \cdots, \vec{v}_k \in \mathbb{R}^n$ form a* basis *for S if*

1. Span $\{\vec{v}_1, \cdots, \vec{v}_k\} = S$.
2. *$\vec{v}_1, \cdots, \vec{v}_k$ are linearly independent.*

The first condition says that the basis should be big enough to span S, the second condition says it should be no bigger. If you are working with some subspace, having an explicit basis can make your life much easier. Often times your subspace will originally be described as those vectors which satisfy some equation or property. If you can describe that subspace with a basis $\vec{v}_1, \cdots, \vec{v}_k$, you can say "every vector in my subspace can be written $c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$. Better yet, there is exactly one way of writing a given vector as a linear combination of the basis vectors.

Important Exercise 11.2:
Let S be a subspace with basis $\vec{v}_1, \cdots, \vec{v}_n$. Show that every vector $\vec{x} \in S$ can be written as a linear combination of the $\vec{v}_i$ in only one way. (hint: if there were two different ways of doing so, say $\vec{x} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = d_1\vec{v}_1 + \cdots + d_k\vec{v}_k$, then what happens if we subtract?)

Easy Exercise 11.3:
Let $\vec{e}_i \in \mathbb{R}^n$ be the vectors with zeros in for each coordinate except coordinate i. Show the $\vec{e}_i$ form a basis for $\mathbb{R}^n$.

Warning 11.4:
There can be more than one basis for a given subspace. For instance any two nonzero vectors in $\mathbb{R}^2$ that are not multiples of each other (that is, not in the same line), form a basis for $\mathbb{R}^2$.

Unimportant Remark 11.5:
We do not speak of bases of $\{\vec{0}\}$, but if you'd like, define the empty set to be a basis for $\{\vec{0}\}$. Its true that this doesn't quite make sense (what is the span of the empty set?), but at least the empty set is linearly independent!

Exercise 11.6:
Show that writing the null space of a matrix in parametric form actually yields, with a tiny bit of work, a basis for the null space (at least if it is nontrivial).

Thus we can aways find a basis for the null space, and we already know how.

We already kind of know a (crappy) way of finding a basis for the column space $C(A)$ of a matrix A: start with all the columns and remove them one by one until the set is linearly independent. However, we can do better, and in fact, a basis can be easily found if you know the rref of A. The idea is this. We wish to pick a subset of the columns of A to use as our basis vectors. Lets say A is $m \times n$ and there is some subset of k columns which would form the basis. We could form an $m \times k$ matrix A' by just removing the other columns. Since these k columns are a basis for $C(A)$, they are linearly independent, so $N(A')$ is trivial (recall that a nonzero element of the null space is a linear dependence relationship of the columns), meaning that if you were to rref A' you would find only pivot columns. But, since each of the row reduction operations make changes only depending on one column at a time, doing some row operations and then removing columns is the same as removing columns and then doing some row operations. Thus the columns of A corresponding to the pivots of rref(A) are linearly independent. Moreover, if you add in any other column, it would be free after rrefing it, so meaning that those columns would be linearly dependent. Thus we get the following theorem:

Theorem 11.7 (Basis for Column Space)**.** *Let A be an $m \times n$ matrix. The columns of A corresponding to the pivots of rref(A) form a basis for $C(A)$.*

Warning 11.8:
This is *very* different from saying the pivots of rref(A) form a basis for $C(A)$. You need to look at the original matrix!

Exercise 11.9:
Let S be a subspace. Show that if $\vec{v}_1, \cdots, \vec{c_k} \in S$, then Span $\{\vec{v}_1, \cdots, \vec{c_k}\} \subseteq S$, or equivalently, that if for any real numbers $c_1, \cdots, c_k$,

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \in S$$

(hint: subspaces are closed under all the operations in the expression above)

The following method is a last-ditch effort for finding a basis for a subspace, if you cannot write your subspace as a null space.

Tricky Exercise 11.10:
Find a method for finding a basis for *any* (nontrivial) subspace. Start with the empty set, and then pick a vector which is in your subspace but not in the span of the previously chosen vectors. Use 11.9 to show that you have not overshot your target. Say why this process must end, and why when it ends you have found a basis for your subspace.

Using this process, we have proven the following:

Theorem 11.11. *Every subspace has a basis.*

11.1 Exam Exercises

Try the following exercises from past exams

- □ W13 Midterm 1 1b
- □ A12 Midterm 1 3b
- □ A12 Midterm 1 3c
- □ S12 Midterm 1 1b
- □ S12 Midterm 1 4a
- □ S12 Midterm 1 4b
- □ S12 Midterm 1 4c
- □ S12 Midterm 1 4d
- □ W13 Final 2b
- □ W13 Final 2c
- □ S11 Final 1b
- □ S11 Final 2b

12 Dimension of a Subspace

Intuitively, a line is "1-dimensional", a plane is "2-dimensional" and space is "3-dimensional", but we'd like to make this precise. The observation is this. A line through $\vec{0}$ has a basis with only one vector in it. A plane through $\vec{0}$ has a basis with two vectors, etc. This leads us to make the following definition.

Definition 12.1 (Dimension of a Subspace)**.** *Let S be a subspace. Then by 11.11, there is a basis for S, say $\vec{v}_1, \cdots, \vec{v}_k$. We say that the* dimension *of S,* $\dim(S)$*, is the number of vectors in that basis.*

But wait a minute! A subspace can have different bases. How do we know that this makes sense? How do we know the number of vectors in any basis for the same subspace is the same?

Tricky Exercise 12.2:
Show that if a subspace S has a basis of k vectors, then any set of $k+1$ vectors in S are linearly dependent. Conclude that any two bases of the same space have the same number of vectors, so the concept of dimension makes sense. (hint: pick a set of $k+1$ vectors. Each can be written as a linear combination of the original basis. Can you set up a $k \times (k+1)$ matrix whose null space contains linear dependence relationships among the $k+1$ vectors. Does $k \times (k+1)$ matrix always have a free column?)

Unimportant Remark 12.3:
The dimension of $\{\vec{0}\}$, is defined to be zero.

Lets calculate some dimensions.

Easy Exercise 12.4:
Show that the dimension of $\mathbb{R}^n$ is n.

The next exercise will give a remarkable relationship between the null space and column space of a matrix.

Important Exercise 12.5 (Rank-Nullity Theorem):
Let A be an $m \times n$ matrix. Show that the dimension of $N(A)$ is the number of pivot vectors. Show that the dimension of $C(A)$ is n minus the number of pivot vectors. Conclude

$$\dim(C(A)) + \dim(N(A)) = n$$

Explain how this means "more linear relationships means smaller span"

Definition 12.6 (Rank and Nullity)**.** *We call the dimension of the column space the rank of a matrix. We call the dimension of the null space the nullity of a matrix.*

Remark 12.7:
We can interpret the rank-nullity theorem geometrically. Recall that the solutions to a set of equations $A\vec{x} = \vec{b}$ are, geometrically, either the empty set or just $N(A)$ slided over in some direction. For each possible place we could translate the null space, there is a unique corresponding vector in the column space. Thus the rank nullity theorem is saying that, if the null space is k dimensional, there are $n - k$ dimensions of translation. Try to think through this remark for the matrices

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

and

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

In the first case, the null space is the z-axis and the column space is the xy-plane. Notice that when translating the null space by some vector $\vec{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, only the x and y coordinates matter, and you just get the line parallel to the z-axis through $\vec{v}$ (which meets the xy-plane at $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$, which does not depend on z). Said another way, there are only "2 dimensions of translation" for the z-axis (2 dimensions, of course, is the rank). For the second matrix, the null space is the yz-plane, and we find there is only 1 "dimension of translation". Keep these pictures in mind! The general situation is somehow like this, except multiplying by the matrix can stretch or twist space to make things slightly more obscure. By the rank-nullity theorem, no matter how much stretching and twisting happens, the situation is fundamentally the same, if slightly harder to visualize directly.

The notion of subspaces and dimension simplifies things in an amazing way! Think about it. A subspace typically has infinitely many vectors in it, but the bases of a subspace only has a handful. The shape of a set of vectors originally sounds like it could be a confusing, hard to grapple with issue, but if the shape forms a subspace you know you are dealing with a line or a plane or a higher dimensional generalization. You can use dimension to compare the size of two subspaces, even though both are hugely infinite. You see automatically that you can't fit a "bigger" subspace into a "smaller" one (where by "bigger" I mean higher dimensional), since it won't fit. The rank-nullity theorem tells us precisely what we mean by "more relationships means smaller span".

12.1 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 1 5b
- ☐ W13 Midterm 1 6a
- ☐ W13 Midterm 1 6c
- ☐ W13 Midterm 1 6d
- ☐ A12 Midterm 1 4c
- ☐ A12 Midterm 1 5a
- ☐ A12 Midterm 1 5b
- ☐ S12 Midterm 1 2b
- ☐ A12 Midterm 2 6c
- ☐ S11 Final 3a

Part III

Linear Transformations

13 Linear Transformations

Originally, we invented matrices to keep track of linear equations, and then allowed ourselves to consider solution sets geometrically. We invented matrix multiplication as a shorthand of writing a system of linear equations, but have occasionally talked about what a matrix "does" to a vector or to space. We now wish to go back on that just a little bit, and say that, in fact, a matrix is a sort of machine which takes vectors and gives back new vectors.

We recall that a function (say from calculus), is a machine which takes some number and gives back a new number. We could say like $f(x) = x^2$ or $f(x) = \sin(x) + e^{3x+\pi}\cos^{-1}(x)$ or whatever. Recall that you defined the domain, codomain and range of a function as the set where the function was defined, the set where the function took values, and the set of actual values the function could take. Typically the domain was all of $\mathbb{R}$ or $\mathbb{R}$ minus a few points (since one wished to avoid evaluating $f(x) = \frac{1}{x}$ at 0), or the positive numbers, or whatever. The codomain was typically just $\mathbb{R}$ and the range would be some restricted subset where you could actually hit (for instance the range of $f(x) = x^2$ is the non-negative numbers). The key idea is that there is no reason that a function can't take as its input a vector and return another vector.

For instance, a function could take a vector and return a scalar, like

$$f\begin{pmatrix} x \\ y \end{pmatrix} = 2x + 3y^2 + \sin(x + y)$$

Here the domain is all of $\mathbb{R}^2$, the codomain is $\mathbb{R}$ and the range is also all of $\mathbb{R}$. Notate this, we write

$$f : \mathbb{R}^2 \to \mathbb{R}$$

meaning the domain is $\mathbb{R}^2$ and the codomain is $\mathbb{R}$. We don't write the range because sometimes it can be hard to say what the range of a function is, but the domain and codomain are of course known directly from the type. Note that we could graph such a function in $\mathbb{R}^3$ if we were so inclined, but the fact that we can graph it was not needed.

We can also have a function taking a vector and returning a vector, like

$$f : \mathbb{R}^2 \to \mathbb{R}^3$$

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + y \\ ysin(x^8) \\ e^{xy} \end{pmatrix}$$

Now the range is really complicated, and graphing this function would require drawing in 5 dimensions, but it is a function all the same.

The functions above were mostly complicated, but we can have simpler functions too, like

$$f : \mathbb{R} \to \mathbb{R}$$

$$f(x) = 5x$$

Or

$$g : \mathbb{R}^3 \to \mathbb{R}^2$$

$$g(\vec{x}) = \begin{pmatrix} 543 \\ 210 \end{pmatrix}\vec{x}$$

It is a very simple idea, but very important.

Multiplying by an $m \times n$ matrix can be thought of as a function $\mathbb{R}^n \to \mathbb{R}^m$

Let A be an $m \times n$ matrix and

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

$$T(\vec{x}) = A\vec{x}$$

(we use the capitol letter T for this sort of function sometimes, but its the same as f) What properties does T have?

Easy Exercise 13.1:
Show that

a) $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$

b) $T(\vec{0}) = \vec{0}$

c) $T(c\vec{x}) = cT(\vec{x})$

(hint: 7.9)

These properties will be nice, because they can be verified from the description of a function without knowing there is a matrix behind it, yet imply the existence of a matrix.

Definition 13.2 (Linear Transformation)**.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a function. We say T is a linear transformation if*

a) $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$

b) $T(\vec{0}) = \vec{0}$

c) $T(c\vec{x}) = cT(\vec{x})$

Important Remark 13.3:
Notice how the linear transformation properties look kind of like the subspace properties. We must remember types; a subspace is a set and a linear transformation is a function. However, in a strong sense that we shall not make precise, a linear transformation is the function counterpart of a subspace.

Important Exercise 13.4:
Given a description of a linear transformation $T : \mathbb{R}^n \to R^m$, we can actually come up with an $m \times n$ matrix so that $T(\vec{x}) = A\vec{x}$? (hint: build the matrix column by column, using 7.10, and then say why the matrix you get at the end actually works). This method for taking a description of a linear transformation and coming up with a matrix is very useful in practice!

This proves

Theorem 13.5. *Every linear transformation has an underlying matrix.*

Recall that a function is one-to-one (or an injection) if each input goes to a different output, that is, if $x \neq y$ (for x, y in the domain) then $f(x) \neq f(y)$. A function is onto (or a surjection) if each element of the codomain has a preimage in the domain, that is, if y in the codomain then there is an x in the domain with $y = f(x)$. A one-to-one function is like the inclusion of a smaller thing into a bigger thing, and an onto function is like the domain fills up the range. Tests like to ask questions about injections and surjections which are free points for students who know whats going on and very hard for students who don't.

Important Easy Exercise 13.6:
Let A be an $m \times n$ matrix and $f\mathbb{R}^n \to \mathbb{R}^m$ be $f(\vec{x}) = A\vec{x}$. Show that if f is one-to-one, then $N(A) = \{0\}$. Show that if $N(A) = \{0\}$ then f is one-to-one (hint: if $x \neq y$ bu $f(x) = f(y)$, then $f(x-y) = f(x) - f(y) = 0$). Conclude that if $m < n$, f is *not* one-to-one, so "a bigger dimensional space is too big to fit inside a smaller one".

Important Easy Exercise 13.7:
Let A and f as 13.6. Show that if f is onto then $C(A) = \mathbb{R}^m$. Show that if $C(A) = \mathbb{R}^m$ then f is onto. Conclude that if $m > n$, f is *not* onto, so "a smaller dimensional space is to small to fill up a bigger one".

Important Easy Exercise 13.8:
Use the rank nullity theorem and 13.6 and 13.7 to conclude that, if A is an $n \times n$ matrix, then A is either both one-to-one *and* ontp or neither. If $m \neq n$, can an $m \times n$ matrix be both one-to-one and onto?

13.1 Exam Exercises

Try the following exercises from past exams

- □ S12 Midterm 1 1c
- □ S12 Midterm 1 6a
- □ S12 Midterm 1 6b
- □ W13 Midterm 2 5a
- □ S11 Final 3b
- □ S11 Final 4a
- □ S11 Final 4b

14 Examples of Linear Transformations

Warning 14.1:
We will use some ideas from Chapter 15 in this section. I think the course text should have put these two chapters in the opposite order, as the ideas in the next section support the ideas from this one. I leave the sections in this order for consistency with the book.

Warning 14.2:
If you find it difficult, after reading this chapter, to find matrices for non "axis-aligned" linear transformations, chapter IV may be of use.

In this chapter we will see some naturally interesting geometric transformations, and talk about how to build a matrix from them. Remember you can use 13.4 to find the matrix of a linear transformation, which is a very important skill on the exams and in life. You should have a general idea of what each of these things are and how to find their matrices.

14.1 Scalings

A scaling is a linear transformation which makes vectors bigger or smaller by some factor of their length. We could make all vectors twice as big, or all vectors half as big.

Easy Exercise 14.3:
What is the matrix for the linear transformation which scales all vectors in $\mathbb{R}^n$ by c?

Of course, we could scale a different amount in different directions. For instance, we could have a linear transformation which doubles the x coordinate and triples the y coordinate in $\mathbb{R}^2$.

Easy Exercise 14.4:
What is the matrix for the linear transformation which scales the $\vec{e}_i$ by c_i for each i?

14.2 Rotations

A rotation in $\mathbb{R}^2$ is just that, it rotates each vector about the origin by some angle θ. It is of course a linear transformation.

Exercise 14.5:
Show that the matrix for a ration by θ is

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

Exercise 14.6:
Show that the columns of a rotation matrix are orthogonal. Show that if R is a rotation matrix and $\vec{x}, \vec{y} \in \mathbb{R}^2$, then

$$\vec{x} \cdot \vec{y} = (R\vec{x}) \cdot (R\vec{y})$$

The next exercise is put here for organizational reasons, although requires one to know a bit about matrix multiplication.

Important Exercise 14.7:
Explain without doing any computation why

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}\begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} = \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix}$$

(hint: what does matrix multiplication mean again?)

You can also rotate about some axis in $\mathbb{R}^3$. For instance, if you rotate about the z axis, the same thing happens in the xy-plane, and the z axis stays fixed. Thus the matrix for such a transformation is

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Similar logic works for the other two axes.

Exercise 14.8:
Show that a rotation matrix about the x and y axis in $\mathbb{R}^3$ are given

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \qquad \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{pmatrix}$$

How do we rotate a vector across an arbitrary axis? Well, any rotation in $\mathbb{R}^3$ is made up of rotations about the standard basis vectors. Thus, since matrix multiplication is the same as "do one and then do the other", you can multiply matrices which rotate about standard axis to get arbitrary rotations. But be careful, these matrices don't commute with each other!

14.3 Reflections

Reflections are like mirrors; they take each vector and swap it with another vector on the opposite side of some hyperplane. For instance, we can reflect across the x axis in $\mathbb{R}^2$ with

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

What if we want to reflect across an arbitrary line? Well, reflecting across a line which makes an angle θ with the x axis is of course the same as rotating by $-\theta$, then reflecting across the x axis, then rotating back[3]. That is, the matrix for such a reflection is

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

Exercise 14.9:
Work out the explicitly the matrix for a reflection in $\mathbb{R}^2$ about a line making an angle θ with the x-axis.

Important Exercise 14.10:
Explain, using the methods above, how to find the matrix reflecting across an arbitrary plane in $\mathbb{R}^3$. You don't have to actually do any computations.

14.4 Projections

Projections are only slightly less intuitive than the above types of linear transformations, but fear not, they are easy enough to work with once you get the hang of it. Let S be a subspace of $\mathbb{R}^n$ (for instance, a plane or line in $\mathbb{R}^3$). Then the projection P onto S (also called the orthogonal[4] projection) takes each vector $\vec{x} \in \mathbb{R}^n$ to the closest point in S to $\vec{x}$. That is, if S is a line and $\vec{x}$ is some vector, there is a unique line perpendicular to S through $\vec{x}$, and $P\vec{x}$ is the intersection of that line and S.

MAKE A PICTURE

We can draw a similar picture in our mind for projections onto planes.

How to find the matrix for such a thing. Again, let us start with a simple example: the "floor plan" matrix

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

This matrix takes each point in $\mathbb{R}^3$ to the point directly "below" or "above" it on the "floor" (that is, the xy-plane). This is the matrix for the orthogonal projection onto the xy-plane.

We can also consider

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

This takes each point and throws away the y and z coordinates, projecting to the closest point on the x axis.

Once again, we can use rotations to get arbitrary projections, although it might be easier to wait until IV to try to write down matrices for more complicated projections onto planes or higher dimensional subspaces. However, in the case of lines, you can use dot-products and geometric intuition to come up with projection matrices. The following slightly tricky exercise is worth doing; draw a picture and see where the standard basis vectors go!

Tricky Exercise 14.11:
Let $\vec{x} \in \mathbb{R}^n$, and for simplicity assume $||\vec{x}|| = 1$ (that is, $\vec{x}$ is a unit vector). Let $L = \mathrm{Span}\{\vec{x}\}$. What is the matrix for the orthogonal projection onto L.

---

[3]We will talk more about this technique of doing something, doing an easy transformation and then undoing something in part IV

[4]there are other kinds of projections too, but orthogonal are the simplest

14.5 Exam Exercises

Try the following exercises from past exams

- □ S12 Midterm 1 7g
- □ S12 Midterm 1 7h
- □ S12 Midterm 1 7i
- □ S12 Midterm 1 7j
- □ S12 Final 3a
- □ S12 Final 3b

15 Composition and Matrix Multiplication

What sort of things can we do with functions. If we have two functions, f and g, and the codomain of g is the domain of f (or more generally the range of g is a subset of the domain of f), we can form the composition $f \circ g$. This means "do g first, and take the output and use it as the input to f", or in symbols

$$(f \circ g)(x) = f(g(x))$$

Note that, if you aren't paying attention, you can get it backwards, since f is written before g but applied after g. For that reason people sometimes pronounce the $\circ$ symbol "after", as in $f \circ g$ is pronounced "f after g".

Exercise 15.1:
Show that if $f : \mathbb{R}^m \to \mathbb{R}^p$ and $g : \mathbb{R}^n \to R^m$ are linear transformations, then so is $f \circ g$. (hint: you don't need matrices, just use the properties).

Let $f(x) = 5x$ and $g(x) = -3x$. Then we have

$$(f \circ g)(x) = f(g(x)) = f(-3x) = 5(-3x) = -15x$$

That is, for linear functions from $\mathbb{R} \to \mathbb{R}$ (that is, 1×1 matrices), function composition is just multiplication; composing two linear functions gives another linear function, whose "matrix" is just the product of the two original "matrices".

Now what happens if

$$f : \mathbb{R}^m \to \mathbb{R}^p$$
$$f(\vec{x}) = A\vec{x}$$
$$g : \mathbb{R}^n \to \mathbb{R}^m$$
$$g(\vec{x}) = B\vec{x}$$

where A and B are appropriate sized matrices. Then we have

$$(f \circ g)(\vec{x}) = f(B\vec{x}) = A(B\vec{x})$$

That is, evaluate $B\vec{x} \in \mathbb{R}^m$ and then apply the matrix A to it. But wait! We know that $f \circ g$ is linear, so there is some $p \times n$ matrix with C with $C\vec{x} = A(B\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$. This leads, quite abstractly, to the definition of matrix multiplication

Definition 15.2 (Matrix Multiplication)**.** *If A and B are $p \times m$ and $m \times n$ matrices respectively, then define their product to be the $p \times n$ matrix of the linear transformation*

$$T(\vec{x}) = A(B\vec{x})$$

Remark 15.3:
By definition, $(AB)\vec{x} = A(B\vec{x})$, so we can drop the parentheses all together.

Exercise 15.4:
We defined matrix multiplication without ever saying a formula for it! This has the advantage of making matrix multiplication easier to think about (it means just do one and then the other), but the disadvantage of not knowing, given two matrices, how do I find their product. Use 13.4 to write down a formula for matrix multiplication. (If you just look it up in the textbook, you'll have to look it up every time. Do this exercise once and I promise you'll never forget the formula).

Easy Exercise 15.5:
Show that matrix multiplication is associate using the definition given above.

Hard, Unnecessary Exercise 15.6:
Show that matrix multiplication is associative using the formula.

Easy Exercise 15.7:
Show that

$$(A + B)C = AC + BC$$

and

$$C(A + B) = CA + CB$$

(hint: what does + mean in the context of linear transformations?)

Remark 15.8:
Matrix multiplication is *not* commutative. It is not true that, if A and B are both $n \times n$ matrices, then $AB = BA$. In fact, most matrices have this property, so be careful when doing matrix algebra not to flip the order!

Warning 15.9:
Now that you've read about matrix multiplication, make sure you go back and (re)-read 14. This chapter will give you some nice examples of how to make matrix multiplication work for you.

15.1 Exam Exercises

Try the following exercises from past exams

- ☐ A12 Midterm 2 3a
- ☐ A12 Midterm 2 2a

16 Inverses

Lets think about $n \times n$ matrices. These are nice because the domain and codomain are the same. This means, by 13.8, that either f is both one-to-one and onto or neither. This is great, if we want to find inverses.

Easy Exercise 16.1:
Say why a function must be both one-to-one and onto for us to consider inverses. Say why an $m \times n$ matrix with $m \neq n$ can never have an inverse.

Exercise 16.2:
It turns out the inverse of a linear transformation is also a linear transformation. Show this. It will require considering things like

$$f(f^{-1}(\vec{x}) + f^{-1}(\vec{y})) = f(f^{-1}(\vec{x})) + f(f^{-1}(\vec{y}))$$

Easy Exercise 16.3:
This means that, in particular, if f^{-1} has a matrix. If the matrix of f is A and the matrix of f^{-1} is B, then show that

$$AB = BA = I$$

(hint: use the function composition definition). We give B the name A^{-1}, since it acts kind of like "$1/A$", except that that doesn't make sense.

Easy Exercise 16.4:
Let $A = (a)$ be a 1×1 matrix. Show that $A^{-1} = (1/a)$.

We say that a matrix is invertible if it has an inverse, which is true exactly when it is square and the null space is trivial. An invertible matrix is a sort of generalization of a nonzero number.

Exercise 16.5 (Socks on Shoes on; Shoes off Socks off):
Let A and B be invertible. Show that

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

(hint: what is $ABB^{-1}A^{-1}$?) We call this the "Socks on Shoes on; Shoes off Socks off" theorem, because in order to undo putting your socks and shoes on, you must undo each action in the *opposite order*. The same logic applies to reversing directions or operations on a Rubik's Cube: to undo something, do the opposite series of operations in the opposite order.

16.1 Computing Inverse Matrices

We wish to compute inverse matrices. Let A be an $n \times n$ matrix with $N(A) = \{\vec{0}\}$. We will start by computing each column individually, and then show a way to do all the columns at once. Lets start by computing the first column. The first column satisfies the relationship

$$AA^{-1}\vec{e}_1 = \vec{e}_1$$

(remember: we know A, but we don't know A^{-1}. We want to solve for A^{-1}!) By writing parentheses on the left hand side, this is equivalent to

$$A(A^{-1}\vec{e}_1) = \vec{e}_1$$

Since multiplying by $\vec{e}_1$ give the first column, this equations means "A (which we know) times the first column of A^{-1} (which we want to know) is $\vec{e}_1$". This is just a linear equation! Since $A^{-1}\vec{e}_1$, is unknown, we can just call it $\vec{x}$. We wish to solve

$$A\vec{x} = \vec{e}_1$$

for $\vec{x}$, and this will be the first column (notice that there is exactly one solution since $N(A)$ is trivial). Of course the same logic works for each i, so we just need to solve

$$A\vec{x} = \vec{e}_i$$

for each i from 1 to n.

Of course, naively this means we have to set up n augmented matrices:

$$(A|\,\vec{e}_i)$$

and row reduce each one. But you'll notice that the operations you do on the left side (the A side) are the same each time. It would be a shame to redo that work n times if you don't have to. Luckily, we have the following:

Exercise 16.6:
Let A be an $n \times n$ matrix with $N(A) = \{\vec{0}\}$. Show that if we start with the augmented matrix

$$(A|\, I)$$

the rref form of that matrix is

$$\left(I|\, A^{-1}\right)$$

That is, we can compute the inverse of A by setting up the matrix above and row reducing it. (hint: if you threw out all the columns on the right side of the line except for one, you'll be computing a column of A^{-1}, by the logic above.)

16.2 Inverses in Practice

Example 16.7:
Let B be any old $m \times n$ matrix and let A be an $m \times m$ invertible matrix. Then what is $A^{-1}B$? It is just the matrix whose columns are the solutions to $A\vec{x} = \vec{b}_i$ where $\vec{b}_i$ is the i^{th} column of B. Thus we just solved n systems of equations "at the same time".

Computing the inverse of a matrix can be a long and involved computation, but very useful. For instance, suppose you have an invertible $n \times n$ matrix A and you know you are going to have to solve $A\vec{x} = \vec{b}$ for a whole bunch of different $\vec{b}$. One option is to just form the augmented matrix and row reduce each time, but this is highly tedious. The better option might be to compute A^{-1} and then say your solution is $\vec{x} = A^{-1}\vec{b}$ for each $\vec{b}$, since matrix multiplication is easier than solving systems. Matrix inversion is a big hammer though, because it takes so long to compute. You should only do it when you think the benefit of solving many systems fast outweighs the cost of computing the inverse.

16.3 Exam Exercises

Try the following exercises from past exams

- □ A12 Midterm 2 4b
- □ A12 Midterm 2 4c
- □ A12 Midterm 2 2b

17 Determinants

Determinants are a notoriously opaque topic for most first time linear algebra students, but we will see that they actually make a lot of intuitive sense.

In this chapter we will investigate how the area of a shape changes under a linear transformation. That is, if R is a shape and f is a linear transformation, then $f(R)$ is some other shape. If we know the area of R, can we say anything about the area of $f(R)$?

17.1 Motivation by area in $\mathbb{R}^2$

Suppose you have two vectors in $\mathbb{R}^2$, $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$. If they are not in a line, then there is a parallelogram formed as shown

MAKEAPARALLELOGRAM

We can think of these vectors in $\mathbb{R}^3$ as $\begin{pmatrix} a \\ c \\ 0 \end{pmatrix}$ and $\begin{pmatrix} b \\ d \\ 0 \end{pmatrix}$. We get

$$\begin{pmatrix} a \\ c \\ 0 \end{pmatrix} \times \begin{pmatrix} b \\ d \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ ad - cb \end{pmatrix}$$

which means that, by 4.20, the volume of the parallelogram above is $|ad - cb|$.

Now let R be some square in $\mathbb{R}^2$, as show.

$$MAKEASQAURE$$

The corners are at

$$\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x+s \\ y \end{pmatrix}, \begin{pmatrix} x+s \\ y+s \end{pmatrix}, \begin{pmatrix} x \\ y+s \end{pmatrix}$$

or

$$\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} + s\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} + s\begin{pmatrix} 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} + s\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

If we have a matrix

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

what is AR, that is, what happens to the square R after we transform by A? Since linear transformations preserve lines (why?), we get a new quadrilateral with corners

$$A\begin{pmatrix} x \\ y \end{pmatrix}, A\begin{pmatrix} x \\ y \end{pmatrix} + s\begin{pmatrix} a \\ c \end{pmatrix}, A\begin{pmatrix} x \\ y \end{pmatrix} + s\begin{pmatrix} a \\ c \end{pmatrix} + s\begin{pmatrix} b \\ d \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} + s\begin{pmatrix} b \\ d \end{pmatrix}$$

Since obviously translation doesn't effect area, if we subtract $A\begin{pmatrix} x \\ y \end{pmatrix}$ from each corner, we get the parallelogram discussed at the beginning, scaled by s.

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, s\begin{pmatrix} a \\ c \end{pmatrix}, s\begin{pmatrix} a \\ c \end{pmatrix} + s\begin{pmatrix} b \\ d \end{pmatrix}, s\begin{pmatrix} b \\ d \end{pmatrix}$$

the area of which is

$$\text{Area}(AR) = |ad - cb|s^2 = |ad - cb|\text{Area}(R)$$

We must give that factor $ad - cb$ a name!

Definition 17.1 (2×2 Determinant)**.** *The determinant of a 2×2 matrix*

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

is $ad - cb$.

Thus we have shown that if you apply a linear transformation to some random square in $\mathbb{R}^2$, the area of the image is the area of the original square times that weird factor $ad - cb$, which we call the determinant of A. But remember from calculus that you can compute the area of shapes by dividing them into squares and taking a funny limit, we have actually shown (kind of) that if R is *any* shape, then

$$\text{Area}(AR) = |ad - cb|\text{Area}(R) = |\det(A)|\,\text{Area}(R)$$

Tricky Exercise 17.2 (optional):
Our proof that the formula $\text{Area}(AR) = |\det(A)|\,\text{Area}(R)$ works for *any* shape skipped over a lot of details. Can you fill in the details using calculus or measure theory?

Let us get used to working with determinants in the 2×2 case.

Easy Exercise 17.3:
Show that $\det(I) = 1$, where I is the 2×2 identity matrix.

Exercise 17.4:
Let A be a 2×2 matrix. Show that if $\det(A) = 0$, then $N(A) \neq \{0\}$. Show that if $\det(A) \neq 0$, then $N(A) = \{\vec{0}\}$. (hint: if the columns are linearly dependent and R is the unit square, what is the volume of AR?)

Important Exercise 17.5:
Show that if A and B are 2×2 matrices, then

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

(hint: what happens to the volume of a standard square after applying B and then A?)

Exercise 17.6:
Let A be a 2×2 invertible matrix. Can you write $\det(A^{-1})$ in terms of $\det(A)$? (hint: combine 17.5 and 17.3 to get an equation involving $\det(A)\det(A^{-1})$, and then divide). Why do I not get a contradiction if $\det(A) = 0$?

Exercise 17.7:
If c is a scalar and A a 2×2 matrix, can you write $\det(cA)$ in terms of c and $\det(A)$. (hint: $cA = (cI)A$)

17.2 Determinants in General

We will now say what a determinant is for a general $n \times n$ matrix.

Definition 17.8 (Determinant). *The definition of the determinant can be found in your course text, because I don't feel like writing it out. It is recursive: taking the determinant of a $n \times n$ matrix requires taking the determinant of many $(n-1) \times (n-1)$ matrices.*

Remark 17.9:
Determinants of 3×3 matrices obey the same intuition as 2×2 ones, except that we are tracking volume instead of area. Determinants of $n \times n$ matrices obey all the intuition that 2×2 matrices do, except we might not know what "area" or "volume" means in $\mathbb{R}^{37}$. It remains true then that the exercises 17.3 to 17.7 still hold for larger square matrices (with some obvious modifications), and you should intuitively think through why each is true, or at least plausible. When working with determinants abstractly, always keep the 2×2 case in mind and you probably won't be led astray. The multiplication law, the inversion law and the fact that $\det(A) = 0$ if and only if A is not invertible are the most important things to remember, as well as the intuition that "the determinant measures how volume of shapes change under linear transformations".

Remark 17.10:
While is true that the determinant measures how any shape's area/volume changes, it is almost always easier to think about the unit hypercube, that is, the shape enclosed by vectors with only zeros and ones. Obviously this is the standard square in $\mathbb{R}^2$ and the standard cube in $\mathbb{R}^3$. The reason is because the $\det(A)$ is (up to a sign) the "volume" of the hyper-parallelogram defined by the columns of A.

The following exercises can be done from the formula, but should be checked against geometric intuition.

Easy Exercise 17.11:
Let A be an $n \times n$ matrix. Show that swapping any two columns changes the determinant by a sign.

Exercise 17.12:
We haven't talked about matrix transposes yet in these notes, the transpose of a square matrix is what you'd get if you spun a matrix 180 degrees about the diagonal. In notation, the transpose of A is called A^T and $(A^T)_{i,j} = A_{j,i}$. Show that

$$\det(A) = \det(A^T)$$

Conclude that swapping two rows also changes the determinant by a sign.

17.3 Computing Determinants Quickly

[optional]

The next section will help you compute determinants of matrices larger than 3×3 faster. This might save you a bit of time on the exam, but mostly its just interesting.

Exercise 17.13:
We say a matrix is upper triangular if all entries below the diagonal are zero. We say a matrix is lower triangular if all entries above the diagonal are zero. Show that the determinant of an upper triangular matrix is just the product of the entries on the diagonal. Conclude that the same is true of a lower triangular matrix.

Exercise 17.14:
Suppose that A is a square matrix and you wish to perform the row reduction operation of adding c times row i to row j, and call the resulting matrix A'. Show that $\det(A') = \det(A)$. (hint: this operation is equivalent to multiplying by a certain matrix. what is the determinant of this matrix?(

This gives an easier way (easier than expanding all the alternating blocks) of computing the determinant! Using the operations of swapping rows and adding c times row one row to another row, try to clear out everything below the diagonal. Go left to right, never undoing a cleared zero. Make sure you count the number of swaps you do! You should only need to do a swap if an entry on the diagonal that you want to divide by is zero. Also, make sure not to multiply rows by scalars, as that changes the determinant. You don't need to go all the way to rref. When you have your matrix upper triangular, the determinant is just the product of the diagonal entries, plus a sign for each swap.

Exercise for Programmers 17.15:
Write a program computing the determinant of a matrix A using the algorithm above, as well as the original definition. If you don't mind destroying A during the computation, you can easily do this in constant memory! Can you think of a way to do this in only linear memory without destroying A? (hint for the last part: you can use the zeros below the diagonal to store just enough info to undo all the row reduction operations you did to get, teehee. You can use your linear memory to remember which rows you swapped)

17.4 A Very Important Tables

For matrices, we have seen that many conditions are actually the same. We will sum them up in the following list. Let A be an $m \times n$ matrix and $T(\vec{x}) = A\vec{x}$. The columns of the following tables (square, tall and fat matrices) are equivalent: if one thing in a column is true, the whole column is true, and if one thing is false, the whole column is false. A – means that that corresponding property *never* holds, for example a fat matrix never has linearly independent columns.

$m = n$ (square)	$m > n$ (tall)	$m < n$ (fat)
Columns of A linearly independent	Columns of A linearly independent	—
$N(A) = \{\vec{0}\}$	$N(A) = \{\vec{0}\}$	$\text{nullity}(A) = n - m$
$C(A) = \mathbb{R}^n$	$\text{rank}(A) = n$	$C(A) = \mathbb{R}^m$
T is one-to-one	T is one-to-one	—
T is onto	—	T is onto
A is invertible	—	—
$\det(A) \neq 0$	—	—

Make sure you learn this table.

17.5 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 2 3a
- ☐ W13 Midterm 2 3b
- ☐ W13 Midterm 2 5b
- ☐ A12 Midterm 2 3b
- ☐ A12 Midterm 2 3c
- ☐ A12 Midterm 2 4a
- ☐ A12 Midterm 2 6a
- ☐ A12 Midterm 2 6e
- ☐ S12 Midterm 2 1a
- ☐ S12 Midterm 2 1b
- ☐ S12 Midterm 2 2c
- ☐ S12 Midterm 2 4b
- ☐ W13 Final 8b
- ☐ S12 Final 1b

Part IV

Systems of Coordinates and Applications

18 Systems of Coordinates

18.1 Motivating Challenge and Definition

The idea of changing coordinate systems is one of the most powerful ideas in linear algebra. It can help you reason clearly about linear transformations which do all sorts of stretching and rotating across in all sorts of obscure directions. Here is the deal. The following exercise should, at this point, be extremely easy:

Easy Exercise 18.1:
Write down 3×3 matrix for the linear transformation which

a) Reflects vectors across the xy-plane.

b) Projects vectors onto the xy-plane.

c) Projects vectors onto the z-axis.

d) Rotates vectors 30 degrees about the z axis.

These exercises were easy since it is "obvious" what happens to the standard basis vectors. In the following exercises you solve fundamentally the same problem, but, because the lines and planes in question are not lined up with the standard basis vectors, it ends up being a bit trickier. Its worth an attempt, but don't spend too long before giving up.

Tricky Exercise 18.2:
Write down 3×3 matrix for the linear transformation which

a) Reflects vectors across the plane spanned by the vectors $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} -4 \\ 3 \\ 2 \end{pmatrix}$.

b) Projects vectors onto the plane spanned by the vectors $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} -4 \\ 3 \\ 2 \end{pmatrix}$.

c) Projects vectors onto the line spanned by $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

d) Rotates vectors 30 degrees about the line spanned by $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

(hint: Find three linearly independent vectors that it is "obvious" what happens to them and write the standard basis vectors as linear combinations of these vectors)

The takeaway point is that, given a problem, some vectors are easier to work with than others. Thus it would be helpful to invent a technique to systematically deal with these sorts of problems.

Let $\mathcal{B} = \{\vec{v_1}, \cdots, \vec{v_n}\}$ be a basis for $\mathbb{R}^n$. For instance, if the problem at hand is to project onto a line spanned by some vector in $\mathbb{R}^2$, you can let $\vec{v_1}$ be that vector and $\vec{v_2}$ be some orthogonal vector. Just pick vectors germane to the problem at hand. Of course, any vector $\vec{x} \in \mathbb{R}^n$ can be written in exactly one way as a linear combination of the $\vec{v_i}$, that is, we can find some a_i so that

$$\vec{x} = a_1\vec{v_1} + \cdots + a_n\vec{v_n}$$

We say that the a_i are the coordinates of $\vec{x}$ *with respect to* the basis $\mathcal{B}$, which we write as $[\vec{x}]_{\mathcal{B}}$, or

$$\left[\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Remark 18.3:
If $\mathcal{B} = \{\vec{e}_1, \cdots, \vec{e}_n\}$, then $[\vec{x}]_{\mathcal{B}} = \vec{x}$, since, by definition,

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$$

18.2 Working with Respect to another Basis

Lets do an example to see how this can work for you. It can all be quite disorienting the first time you see it, so go slow and make sure you follow the following example.

Example 18.4:
Lets find a matrix for the first part of 18.2 with respect to a nicely chosen basis. We want to find a matrix for T, which reflects vectors across the plane spanned by the vectors $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} -4 \\ 3 \\ 2 \end{pmatrix}$. We will start by picking three vectors to make up $\mathcal{B}$, our basis for $\mathbb{R}^3$.

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad \vec{v}_2 = \begin{pmatrix} -4 \\ 3 \\ 2 \end{pmatrix}$$

are natural choices for basis vectors, so we just need one more. We will pick as our third basis vector the vector

$$\vec{v}_3 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \times \begin{pmatrix} -4 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ -14 \\ 11 \end{pmatrix}$$

We shall calculate B, the matrix for the reflection with respect to $\mathcal{B}$. Since $\vec{e}_i = [v_i]_{\mathcal{B}}$, we have the columns of B are the where T sends the v_i, but written with respect to the basis. Now, the first two vectors in our basis actually don't change when you apply T, that is,

$$B\vec{e}_1 = [T(\vec{v}_1)]_{\mathcal{B}} = [\vec{v}_1]_{\mathcal{B}} = \vec{e}_1 \qquad B\vec{e}_2 = [T(\vec{v}_2)]_{\mathcal{B}} = [\vec{v}_2]_{\mathcal{B}} = \vec{e}_2$$

so we have that the first two columns of B are $\vec{e}_1$ and $\vec{e}_2$ respectively. We also have that the third vector is merely flipped since it is orthogonal to the plane of reflection, that is

$$B\vec{e}_3 = [T(\vec{v}_3)]_{\mathcal{B}} = -[\vec{v}_3]_{\mathcal{B}} = -\vec{e}_3$$

so the last column is $-\vec{e}_3$. Thus

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

Note that this is the same as 18.1. Why is that no coincidence?

Easy Exercise 18.5:
Pick "nice" basis $\mathcal{B}$ for each of the remaining problems of 18.2 and write the given linear transformation in terms of it, as above.

18.3 Change of Basis and the (not-so) mysterious $A = CBC^{-1}$ formula

Remember, however, that our goal was to get these matrices in terms of standard coordinates. To do this we introduce change of basis matrices.

Easy Exercise 18.6 (Change of Basis Matrix):
Let $\mathcal{B} = \{\vec{v_1}, \cdots, \vec{v_n}\}$ be a basis for $\mathbb{R}^n$ and let C be the $n \times n$ matrix whose columns are the $\vec{v_i}$. Show that

$$C[\vec{x}]_{\mathcal{B}} = \vec{x}$$

Conclude that

$$[\vec{x}]_{\mathcal{B}} = C^{-1}\vec{x}$$

Thus if we have a vector written in some coordinate system, we can multiply it by the "change of basis matrix" above to see what it is in standard coordinates, so to change from standard coordinates to some other system we multiply by the inverse! We can kind of sum this up in a funny kind of picture.

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, bend left, "C^{-1}"] \arrow[r, bend right, "C"'] & {[\mathbb{R}^n]_{\mathcal{B}}} \end{tikzcd}$$

This gives us a schema for trying to solve certain sorts of problems. Suppose we are given a description of a linear transformation, as in 18.2. We wish to write down a matrix, called A, for this linear transformation in standard coordinates. So what we do is this. We find a basis, $\mathcal{B}$, where it is easy to solve the problem, that is, where we can come up with a matrix B for the linear transformation in question with respect to $\mathcal{B}$. At this point the picture looks like this:

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, dashed, "A"] & \mathbb{R}^n \\ {[\mathbb{R}^n]_{\mathcal{B}}} \arrow[r, "B"] & {[\mathbb{R}^n]_{\mathcal{B}}} \end{tikzcd}$$

We draw A as a dashed line because we don't know what it is, even though we know B. But wait, we also know how to get from $\mathbb{R}^n$ to $[\mathbb{R}^n]_{\mathcal{B}}$ and back; you merely multiply by the change of basis matrix. Thus we can draw in vertical arrows, like so

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, dashed, "A"] \arrow[d, "C^{-1}"'] & \mathbb{R}^n \\ {[\mathbb{R}^n]_{\mathcal{B}}} \arrow[r, "B"] & {[\mathbb{R}^n]_{\mathcal{B}}} \arrow[u, "C"'] \end{tikzcd}$$

But this is great, because it means that applying A (which is unknown) is the same as applying C^{-1}, then B, then C (which are known). After taking a second to remember which order matrix multiplication happens in, we get:

Important Easy Exercise 18.7:
Conclude that $A = CBC^{-1}$.

Exercise 18.8:
Can you draw the diagram above from scratch? Remember, we want to "go to the other system of coordinates, solve the problem, and come back". If so, you won't need to memorize the formula $A = CBC^{-1}$, which is good because you can easily get confused and swap A and B or C and C^{-1}.

Important Exercise 18.9:
Do 18.2 with this technique.

Important Remark 18.10:
If we have a vector written with respect to two different bases, we really want to think of those two descriptions as merely different *names* for the *same* vector. Various intrinsic properties of vectors do not change when changing the basis. For instance, the zero vector stays the zero vector, and if there is a linear relationship between some vectors, that relationship will still hold after applying a change of coordinates. That is, if

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = \vec{0}$$

then

$$c_1[\vec{v}_1]_{\mathcal{B}} + \cdots + c_k[\vec{v}_k]_{\mathcal{B}} = [\vec{0}]_{\mathcal{B}} = \vec{0}$$

In particular, if one vector is a multiple of another or a sum of some others, that relationship will hold, that is, written somewhat strangely,

$$c[\vec{v}]_{\mathcal{B}} = [c\vec{v}]_{\mathcal{B}}$$

$$[\vec{v}]_{\mathcal{B}} + [\vec{w}]_{\mathcal{B}} = [\vec{v} + \vec{w}]_{\mathcal{B}}$$

However, some properties are not always preserved when you change basis. For instance, the length and dot product can change if you change the basis, because they somehow "depend on how we measure things"[5]. An easy example is letting the $\mathcal{B} = \{2e_1, \cdots, 2e_n\}$; changing to this basis just cuts all the coordinates in half, which cuts the length by half and the dot product by a fourth.

Exercise 18.11:
Show the things asserted in 18.10. They should follow easily from fact that you can change basis by multiplying by a matrix.

We find that matrices with the relationship above share a number of properties, so we call them similar.

Definition 18.12 (Similar Matrices)**.** *We say two $n \times n$ matrices, A and B, are similar if there is some matrix C such that*

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

Easy Exercise 18.13:
Show that, if A and B are similar, then

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

(hint: 17.5)

Exercise 18.14:
Show that if $A = CBC^{-1}$, then for any whole number k,

$$A^k = CB^kC^{-1}$$

(hint: $A^2 = CBC^{-1}CBC^{-1}$)

Easy Exercise 18.15:
Show that if $A = CBC^{-1}$, then

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

(hint: 16.5)

[5] The dot product and length will change unless the change-of-basis matrix C is "orthonormal", which means each pair of columns are orthogonal and each column has length 1. This is a very interesting class of matrices because they represent linear transformations which "do not change how we measure things". We will not discuss them much in this course.

18.4 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 2 1c
- ☐ W13 Midterm 2 5c
- ☐ W13 Midterm 2 6b
- ☐ W13 Midterm 2 6c
- ☐ A12 Midterm 2 7a
- ☐ S12 Midterm 2 4a
- ☐ S12 Midterm 2 4c

19 Eigenvectors

19.1 Definition

We saw last chapter that changing coordinate systems is great, but it only seems to work if the problem is low dimensional or you somehow know that certain vectors which behave nicely. Given some random problem, choosing a basis might be very hard. Luckily there is a very intrinsic way of saying if a vector should be part of a basis.

Definition 19.1 (Eigenvectors and Eigenvalues)**.** *Let A be an $n \times n$ matrix and $\lambda \in \mathbb{R}$ be a scalar. Then we say a nonzero vector $\vec{x} \in \mathbb{R}^n$ is an eigenvector if*

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

and we call λ an eigenvalue of A.

Remark 19.2:
This looks bizarre! Whats going on here? It seems that $\vec{x}$ is an eigenvalue of A if multiplying by A is the same as multiplying by some scalar, which has the funny Greek name λ. Since multiplying a matrix mixes up all the different coordinates, this is a very special situation indeed, most vectors are not eigenvectors. Make sure the types make sense to you!

Easy Exercise 19.3:
It is somewhat difficult to find eigenvectors, but it is at least easy to check if a vector is an eigenvector. Let

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

Check that $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ are eigenvectors by multiplying them by A. What are the eigenvalues?

Easy Exercise 19.4:
Show that $\vec{x}$ is an eigenvector of A with eigenvalue 0 if and only if $\vec{x} \in N(A)$. (hint: write the definition of eigenvector with eigenvalue 0)

Exercise 19.5 (Eigenspace):
Let A be an $n \times n$ matrix. Fix $\lambda \in \mathbb{R}$ and show that the set of solutions to $A\vec{x} = \lambda\vec{x}$, is a subspace of $\mathbb{R}^n$. Thus the eigenvectors with eigenvector λ, along with zero, form a subspace. We call this the eigenspace or λ-eigenspace and denote it $E_\lambda(A)$ or E_λ.

Exercise 19.6:
Show that if A is and $n \times n$ matrix and $\vec{x}$ is an eigenvector with eigenvalue λ, and if $B = CAC^{-1}$, then $C\vec{x}$ is an eigenvector of A. Conclude that similar matrices have the same eigenvalues.

Exercise 19.7:
Show that if A is an $n \times n$ matrix with eigenvalue λ, then

a) if $c \in \mathbb{R}$, then cA has eigenvalue $c\lambda$

b) if k is a whole number, A^k has eigenvalue λ^k

c) if A is invertible then A^{-1} has eigenvalue $\frac{1}{\lambda}$

(hint: in each case, what happens when you multiply the modified matrix by an eigenvector of the original matrix)

Important Exercise 19.8:
What are the eigenvalues (and eigenvectors) of a reflection? How about an orthogonal projection? Do rotations in $\mathbb{R}^2$ that aren't by multiplies of π have eigenvectors?

19.2 Diagonal Matrices

Why are eigenvectors so nice? Because if we write a matrix with respect to a basis of eigenvectors, it ends up diagonal!

Important Easy Exercise 19.9 (Diagonalizable Matrices):
Let A be an $n\times n$ matrix. If $\mathcal{B} = \{\vec{v}_1, \cdots \vec{v}_n\}$, where the $\vec{v}_i$ are linearly independent eigenvectors of A with eigenvalues $\lambda_1, \cdots, \lambda_n$, then show that B, the matrix with respect to $\mathcal{B}$, is diagonal (that is, it has only zeros off the main diagonal). A matrix with this property is called diagonalizable. (hint: $B\vec{e}_i = [A\vec{v}_i]_{\mathcal{B}}$)

Diagonal matrices are the simplest kind of matrix. For evidence of this, write $A\vec{x} = \vec{b}$ as a system of equations, when A is diagonal. Each equation only involves a single variable!

Easy Exercise 19.10:
If A is diagonal, show that $\det(A)$ is just the product of the diagonal entries.

Easy Exercise 19.11:
Show that if A is diagonal, then A is invertible if and only if each entry on the diagonal is nonzero.

Easy Exercise 19.12:
Show that the standard basis vectors are eigenvectors of a diagonal matrix.

19.3 Finding Eigenvalues and Eigenvectors

Suppose A is an $n \times n$ matrix and somehow we know that a specific λ is an eigenvalue. Then since $\lambda\vec{x} = \lambda I\vec{x}$ Then we can rewrite the eigenvector equation as

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

or

$$\vec{x} \in N(A - \lambda I)$$

Thus, in order for λ to be an eigenvalue, we need $\det(A - \lambda I) = 0$. So to find the eigenvectors of A, we need first to find the eigenvalues by solving $\det(A - \lambda I) = 0$ for λ and then, for each solution, finding a basis for $N(A - \lambda I)$.

Important Exercise 19.13:
Pick some actual matrices from the book and calculate their eigenvectors, eigenvalues and eigenspaces.

Exercise 19.14:
If A is an $n \times n$ matrix with eigenvalue λ, show that A^T also has eigenvalue λ. (hint: what is $\det(A^T - \lambda I)$? Use 17.12 and the fact that $A^T - \lambda I = (A - \lambda I)^T$)

Notice that $\det(A - \lambda I)$ is a polynomial expression, since if you expand it out using Kramer's rule you only add, subtract, multiply and divide. We call this the "characteristic polynomial" of A, though that terminology will not be important. Notice further that the degree of this polynomial is n. That means if $n = 2$ you can solve it using the quadratic formula. However, if $n > 2$, finding eigenvalues is a real pain, and if $n > 4$ this is actually impossible to always find exact solutions. This also means there are at most n eigenvalues, since a degree n polynomial has at most n roots.

Warning 19.15:
All the matrices you will see in this class have real eigenvalues, but that doesn't need to be the case. For instance, what are the eigenvalues of the matrix

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

We have that

$$\det(A - \lambda I) = \lambda^2 + 1$$

which has roots

$$\lambda = \pm\sqrt{-1}$$

The $\sqrt{-1}$, also known simply as i, the imaginary unit, does not live in $\mathbb{R}$ but rather in $\mathbb{C}$. If such a thing happens to you, you can either say give up or realize that every word I've said in these notes works equally well for complex numbers. Either way, since you (might) know that roots of real polynomials come in conjugate pairs, you know that if $x + iy$ is an eigenvalue, then so is $x - iy$.[6]

Remark 19.16:
Here's a tip. When calculating eigenvalues of matrices larger than 2×2, if you get stuck, look at the matrix itself. Often you can figure out one of the eigenvalues just by staring at it. For instance, if all of the rows (or, by 19.14, columns) add up to the same number c, you know the vector with all coordinates 1 is an eigenvector with eigenvalue c. Going back to the polynomial you were trying to factor, you can then divide by $\lambda - c$, using polynomial long division, since you know c is a root. Tricks like that can get you out of factoring a higher degree polynomial by hand.

Remark 19.17:
The following relationship is helpful. If λ is an eigenvalue, it is a root of the polynomial $\det(A - \lambda I)$, and say it has multiplicity r. Then

$$1 \leq \dim(E_\lambda) \leq r$$

that is, the dimension of the eigenspace for a given eigenvalue can be no more than the multiplicity of that root in the characteristic polynomial. If you know how to do induction, go ahead and try to prove it, first for diagonal matrices and then for general ones.

19.4 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 2 1a
- ☐ W13 Midterm 2 1b
- ☐ W13 Midterm 2 2b

[6]This is actually awesome! If you know some complex analysis, you know that $e^{i\theta}$ is somehow like a rotation, and the rotation matrix by θ has $e^{i\theta}$ as a complex eigenvalue.

- ☐ A12 Midterm 2 6d
- ☐ A12 Midterm 2 7b
- ☐ A12 Midterm 2 7c
- ☐ A12 Midterm 2 8a
- ☐ S12 Midterm 2 3a
- ☐ S12 Midterm 2 3b
- ☐ S12 Midterm 2 5a
- ☐ S12 Midterm 2 5b
- ☐ S12 Midterm 2 5c
- ☐ S12 Midterm 2 6a
- ☐ S12 Midterm 2 6b
- ☐ S12 Midterm 2 6c
- ☐ W13 Final 8a
- ☐ W13 Final 8c
- ☐ S12 Final 4a
- ☐ S12 Final 4b
- ☐ S12 Final 4c
- ☐ S12 Final 5a
- ☐ S11 Final 3c
- ☐ S11 Final 5a
- ☐ S11 Final 5b

20 Symmetric Matrices

The study of symmetric matrices is an interesting topic, but one which requires a tiny bit more material to fully appreciate. Symmetric matrices come up, for instance, in machine learning and graph theory. You should think of the next two sections as a teaser for future linear algebra classes; learn the statements, but don't worry too much about full understanding of the spectral theorem.

20.1 Transpose Review

Recall that the transpose of an $m \times n$ matrix A is an $n \times m$ matrix, denoted A^T, which has its (i, j) and (j, i) entries flipped. We sometimes play fast and loose with types and say the transpose of a vector in $\mathbb{R}^n$ is the $n \times 1$ matrix given by just flipping the vector on its side. Some easy things to check:

Easy Exercise 20.1: a) If $\vec{x}, \vec{y} \in \mathbb{R}^n$, then $\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y} = \vec{y}^T \vec{x}$

b) If A and B are matrices of the same size then $(A + B)^T = A^T + B^T$.

c) If A is an $m \times n$ matrix and $\vec{e}_i$ is a unit vector in $\mathbb{R}^n$ and $\vec{e}_j$ is a unit vector in $\mathbb{R}^m$, then $\vec{e}_j^T A \vec{e}_i$ is the (j, i)-entry of A.

d) If A and B are matrices which can be multiplied then $(AB)^T = B^T A^T$. (hint: $\vec{e}_j(AB)^T \vec{e}_i = \vec{e}_i AB \vec{e}_j$. This one is slightly harder)

20.2 Properties or Symmetric Matrices

An $n \times n$ matrix is called symmetric if $A^T = A$, that is, if the (i, j)-entry is equal to the (j, i)-entry for each i and j. Symmetric matrices have a funny little property which goes like this: if A is a symmetric matrix and $\vec{x}$ and $\vec{y}$ are any vectors, then

$$(A\vec{x}) \cdot \vec{y} = \vec{x} \cdot (A\vec{y})$$

Why is that the case? Recall that

$$\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y} = \vec{y}^T \vec{x}$$

Then we have

$$\begin{aligned} (A\vec{x}) \cdot \vec{y} &= (A\vec{x})^T \vec{y} \\ &= \vec{x}^T A^T \vec{y} \\ &= \vec{x}^T A \vec{y} \\ &= \vec{x} \cdot (A\vec{y}) \end{aligned}$$

where we used the formula $(AB)^T = B^T A^T$ in the second line and the fact that A is symmetric to replace A^T with A in the third. We can use this fact to show another nice fact. If $\vec{x}$ and $\vec{y}$ are eigenvalues of a symmetric $n \times n$ matrix A with different eigenvalues λ and μ, then $\vec{x} \cdot \vec{y} - 0$. Said again, eigenvectors with different eigenvalues are orthogonal. To see this, apply the formula:

$$\begin{aligned} \lambda \vec{x} \cdot \vec{y} &= (A\vec{x}) \cdot \vec{y} \\ &= \vec{x} \cdot (A\vec{y}) \\ &= \vec{x} \cdot \mu \vec{y} \\ &= \mu \vec{x} \cdot \vec{y} \end{aligned}$$

and thus, moving everything to one side

$$(\lambda - \mu)\vec{x} \cdot \vec{y} = 0$$

but since $\lambda \neq \mu$, we must have $\vec{x}$ and $\vec{y}$ are orthogonal. We end with two important theorems which we opt not to prove, since the proofs require some dealing with complex numbers and a longish induction.

Theorem 20.2. *If A is a symmetric $n \times n$ matrix, all the eigenvalues of A are real.*

Theorem 20.3 (Spectral Theorem)**.** *If A is a symmetric $n \times n$ matrix, there is a basis of $\mathbb{R}^n$ made up of eigenvectors of A (that is, symmetric matrices are diagonalizable), and furthermore, each pair of vectors in that basis are orthogonal.*

20.3 Exam Exercises

Try the following exercises from past exams

- ☐ A12 Midterm 2 8b
- ☐ S12 Midterm 2 7a

21 Quadratic Forms

Chances are, when you first started studying algebraic equations, you started with linear ones $mx + b$ and eventually moved up to quadratic ones. You might have even learned a song to help find the zeros of $ax^2 + bx + c$. If the linear equations are supposed to generalize to $\vec{m} \cdot \vec{x} + b$, what is the situation with quadratics? Lets start as always by thinking about $\mathbb{R}^2$ (for no reason but concreteness). In this case, a linear equation is of the form

$$m_1 x + m_2 y + b$$

What is our quadratic term. Well, it should have something like

$$a_1 x^2 + a_2 y^2$$

But wait, we could also have a "degree-2" term xy, and look like

$$a_1 x^2 + a_2 y^2 + a_3 xy$$

Easy Exercise 21.1:
Show that

$$\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} a_1 & \frac{a_3}{2} \\ \frac{a_3}{2} & a_2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = a_1x^2 + a_2y^2 + a_3xy$$

We use this as a jumping off point for the definition

Definition 21.2 (Quadratic Form)**.** *A quadratic form is a function $Q : \mathbb{R}^n \to \mathbb{R}$ of the form*

$$Q(\vec{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}x_ix_j$$

The matrix of a quadratic form is the symmetric *$n \times n$ matrix with entries $a_{i,j}$.*

Easy Exercise 21.3:
If Q is a quadratic form with matrix A, show

$$Q(\vec{x}) = \vec{x}^T A\vec{x}$$

(hint: this is basically just 21.1)

Exercise 21.4:
Why do you think we require the matrix of a quadratic form be symmetric?

Recall that one nice thing about quadratics is that it is easy to tell if they are always positive of always negative. A 1-dimension quadratic form is just the term ax^2, which is almost always positive if $a > 0$, almost always negative if $a < 0$, and always zero if $a = 0$. We can make a similar classification of quadratic forms

Definition 21.5 (Definiteness of a Quadratic Form)**.**

A quadratic form Q is positive definite if $Q(\vec{x}) > 0$ when $\vec{x} \neq \vec{0}$.

A quadratic form Q is positive semi-definite if $Q(\vec{x}) \geq 0$ always.

A quadratic form Q is negative (semi-)definite in the opposite case.

A quadratic form Q is indefinite otherwise.

Easy Exercise 21.6:
Sometimes it is easy to classify matrices quadratic forms. Classify the following

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

Important Remark 21.7:
Locally, derivatives are like linear terms and second derivatives are like quadratic ones. This intuition holds in many variables, the second derivative of a multi-variable function will be locally a quadratic form called the Hessian. The definiteness of this quadratic form will be the multi-variable version of the second-derivative-test for local extrema, just like in calculus.

Luckily, there is an easy way to classify the definiteness of an arbitrary matrix. First, though, we need the following exercise.

Important Exercise 21.8:
If two matrices are similar, show that they have the same definiteness.

But by the spectral theorem, any symmetric matrix is similar to a diagonal one with the eigenvalues on the diagonal. Thus

Important Exercise 21.9:
Say how to find the definiteness of a matrix given its eigenvalues. (hint: 21.6)

21.1 Exam Exercises

Try the following exercises from past exams

- ☐ W13 Midterm 2 2a
- ☐ W13 Midterm 2 2c
- ☐ A12 Midterm 2 5b
- ☐ S12 Midterm 2 7b
- ☐ S12 Final 5b